Ultra-Large Large-Scale Structures from Intervening Mg II Absorbers; Possible Implications for Cosmology

Alexia Marie Lopez

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jeremiah Horrocks Institute for Mathematics, Physics and Astronomy
University of Central Lancashire

June 2025

Declaration

Type of Award: Doctor of Philosophy

School: Engineering and Computing

The work presented in this thesis was carried out at the Jeremiah Horrocks Institute for Mathematics, Physics and Astronomy, University of Lancashire (formerly Central Lancashire). I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work. The only exception to this statement is that the original discovery of the Giant Arc was reported in my MSc thesis, but its analysis here is new work. Data and analysis used in this thesis that are not my own are clearly cited in the text.

This thesis contains work that has been published in the following papers.

1. Lopez, A. M., Clowes, R. G., & Williger, G. M. 2022, *Mon. Not. Roy. Astron. Soc.,* 516, 1557
2. Aluri, P. k., Cea, P., Chingangbam, P., ... Lopez, A. M., et al. 2023, *Class. and Quantum Gravity,* 40, 094001

3. Lopez, A. M., Clowes, R. G., & Williger, G. M. 2024, *J. Cosmo. Astroparticle Phys.,* 2024, 055
4. Lopez, A. M., Clowes, R. G., & Williger, G. M., 2025, *Phil. Trans. A Roy. Soc.,* 383, 2290

Following the PhD viva (April 2025), an additional section was added to the thesis (Section 3.4) which partially contains work that has been published in arXiv and submitted to the *Journal of Cosmology and Astroparticle Physics (JCAP).*

5. Lopez, A. M. & Clowes, R. G., 2025, [arXiv:2504.14940]

No proof-reading service was used in the compilation of this thesis.

Alexia Marie Lopez
June 2025

Abstract

The two ultra-large large-scale structure (uLSS) discoveries of the Giant Arc (GA) and the Big Ring (BR) are presented in this thesis. The GA (~ 1 Gpc) and BR (~ 400 Mpc) both exceed the often-cited Yadav et al. (2010) scale of homogeneity ($R_h \sim 370$ Mpc), and add to an accumulating set of large-scale structure (LSS) discoveries that potentially challenge the Cosmological Principle (CP), on which the standard model of cosmology (ΛCDM) is founded. In cosmological terms, the GA and BR are very close; they are separated by only $\sim 12^\circ$ on the sky and are both in the same redshift slice at $z \sim 0.8$.

The observational properties of the GA and BR are presented, and they are assessed statistically using methods based on: (i) Single-Linkage Hierarchical Clustering; (ii) the Convex Hull of Member Spheres — GA: 4.53σ, and BR: $(3.65 \pm 1.13)\sigma$; (iii) the Minimal Spanning Tree — BR: $(4.10 \pm 0.45)\sigma$; (iv) the Cuzick-Edwards test — GA: 3.00σ, and BR: 2.0σ; (v) power spectrum analysis — GA: 4.8σ; (vi) and the FilFinder algorithm. Each of these methods has distinctive attributes and powers, and it is advised to consider the evidence from the ensemble. The approaches to mitigating any *post-hoc* aspects of analysing significance after discovery is discussed. The CHMS-overdensity and the MST-overdensity of the GA is $\delta\rho/\rho \simeq 0.9 \pm 0.6$ and $\delta\rho/\rho \simeq 1.3 \pm 0.3$, respectively. The alpha-hull overdensity of the BR(-all) is $\delta\rho/\rho \simeq 0.75$.

The discoveries were made using the novel method of mapping intervening Mg II absorbers in the spectra of background quasars. This method allows for the detection of low-luminosity, distant matter which might otherwise go unseen with standard methods. The quasar spectra are obtained from the Sloan Digital Sky Survey (SDSS) and the Mg II catalogues are provided by independent authors.

Currently, the huge sizes and intriguing, circular-type morphologies of the GA and BR are not easily explained within ΛCDM. In addition, there have been hints of a second arc to the north of the GA and BR that appears to be an extension of the GA. If this is indeed the case, the GA might be extending into a Giant Ring (GR) which would then encompass the BR. The implications that these structures might have for cosmology could be important; but what they are telling us is not yet clear. Perhaps these uLSS discoveries may even be hinting at a direction of new physics within the standard model, or alternative theories in cosmology such as cosmic strings, pre-inflationary particles, or Conformal Cyclic Cosmology.

Contents

Declaration **ii**

Abstract **iv**

Acknowledgements **ix**

1 Introduction **1**

1.1 Large-Scale Structure of the Universe 4

1.2 A Review of the Cosmological Principle 9

1.3 Cosmological Parameters 11

2 The Physics and Astrophysics of Mg II Absorbers **12**

2.1 The Physics of the Mg II Absorption Doublet 13

2.2 The Physical Environment of Mg II Absorbers 15

3 Data and Method **22**

3.1 The SDSS Quasar Catalogues 23

3.2 The Mg II catalogues 26

3.3 Constructing the Mg II Images 40

3.3.1 A Note on Redshift Errors 41

3.4 Statistical Tests 42

3.4.1 Minimal Spanning Tree 43

3.4.2 Cuzick-Edwards Test 45

4 **A Giant Arc on the Sky** **46**

4.1 Discovery of the Giant Arc 46

4.2 Observational properties of the Giant Arc 48

4.3 Connectivity and statistical properties 52

4.3.1 SLHC / CHMS (Minimal Spanning Tree) 54

4.3.2 Selecting a Linkage Scale 64

4.3.3 Cuzick-Edwards test 67

4.3.4 Power Spectrum Analysis 72

4.4 Overdensity 75

4.5 Comparisons with Other Data 77

4.6 Discussion and Conclusions 85

5 **A Big Ring on the Sky** **89**

5.1 Discovery of the Big Ring 89

5.2 Initial checks of the data 91

5.3 Statistical analysis 97

5.3.1 Single-Linkage Hierarchical Clustering algorithm 99

5.3.2 Significance: CHMS and MST 113

5.3.3 Filament identification algorithm 118

5.3.4 Cuzick and Edwards test 125

5.4 Observational properties 127

5.4.1 Corroboration with independent data 127

5.4.2 Viewing the BR from other angles 131

5.4.3 Equivalent widths 139

5.5 Discussion and conclusions 142

6 Discussion and Conclusions 147

6.1 Discussion: Implications for Cosmology 147

6.2 Future Work 157

6.3 Conclusions 160

Acknowledgements

I began my part-time PhD in 2020: the start of the pandemic. Reflecting on the highlights of my PhD journey I am able to appreciate how fortunate I am to have had so many incredible experiences, especially when considering the many difficulties we all had to face during the Covid years.

Prior to the PhD, I had just finished my MSc by research here at UCLan (I also completed my BSc here too). In that way, the transition to the PhD was almost seamless; I was to continue working with Dr Roger Clowes on cosmological large-scale structure and the method of Mg II absorbers tracing faint matter. So, just short of a decade of being at UCLan I have been able to enjoy watching the University grow and change. I think here, at my final stage of education, it is an appropriate time to reminisce on these years.

It is an absolute privilege to do a PhD in Cosmology: to spend time asking questions and searching for the answers, to wonder about the Universe, to converse with like-minded individuals who share the same passions, to attend conferences, eat pastries, and be gifted free pens at said conferences. I am incredibly fortunate to have found myself in a position to do the things that I love, and to this, I owe my first, foremost and greatest thanks to my supervisor Dr Roger Clowes.

I first met Roger in the 3^{rd} year of my BSc working with him on my project on alternative theories in cosmology (a literature review). If anyone knows both myself and Roger personally, they might chuckle at the stark differences in our personalities. However, I count myself incredibly lucky to have found precisely the

type of person I would want by my side in my academic career. Roger has been consistent (routine?), inspiring (yes, I mean that one), patient, dependable, reliable and unarguably the most wonderful supervisor I could be so honoured to work with; he has undoubtedly helped shape my academic career into something I could have never imagined possible. In the 7 years I have known and worked with Roger I have had countless belly laughs, many intriguing discussions, hundreds of self-doubts (met with the calm logic of suggesting I should read a good textbook), and immeasurable support and encouragement. Thank you *so* much Dr Roger Clowes, truly and wholeheartedly, and I look forward to hopefully many opportunities to continue working with you.

My next thanks goes to my second supervisor Dr Gerard Williger for the many helpful supervisory meetings, useful discussions and advice, but most importantly for his continual kindness and support. In addition, I would like to thank my previous secondary supervisors Prof. Luis Campusano and Dr Srinivasan Raghunathan for their help and advice when it was needed. Finally, I would like to thank Dr Mark Norris for volunteering to fulfil the administrative commitments following the retirement of Dr Roger Clowes.

A second thanks goes to Prof. Luis Campusano and his wife Sophie who looked out for me during my two-month research visit in Chile (Cerro Calán, Department of Astronomy, Universidad de Chile). Luis and Sophie made me feel very welcome in Chile: helping me find my way around and taking me to new places to enjoy the culture. I would also like to thank those at Calan, as well as those at the Beauchef (Department of Physics and Mathematics), who welcomed me with open arms, quickly became my good friends, and made my visit the most memorable. I am equally grateful to those at ESO and Paranal Observatory, including Dr Gabriel Marinello who oversaw my visit to the ESO (Chile) headquarters, and Dr Joseph Anderson, who I shadowed and worked with for three nights at the VLT.

My gratitude extends to two academic individuals who, although outside my own research team, have made a commendable impact on my research career. Firstly, thank you to Prof. Subir Sarkar for inviting me to present my work at the Royal Society Meeting held in London in April, 2024 on *"Challenging the Standard Cosmological Model".* It was a huge honour to be invited to this prestigious meeting, and to discuss major cosmological themes amongst the greatest cosmological minds. Secondly, thank you to Prof. Sir Roger Penrose. I am incredibly grateful, and honoured, to have been contacted by Sir Roger Penrose who commended my work on the Giant Arc and Big Ring discoveries. Additionally, I am grateful for the fruitful discussions that followed on the potential link between his Conformal Cyclic Cosmology (alternative) model and the giant ring-type structures that I have discovered. I look forward to pursuing some of these interesting ideas.

In January 2024 I attended the 243^{rd} American Astronomical Society (AAS) Meeting in New Orleans, USA, to present my work on the Big Ring. I am incredibly fortunate to have been awarded an AAS FAMOUS grant that helped to cover my travel and accommodation costs, allowing me to attend this conference.

I would like to acknowledge the University of Central Lancashire and the Jeremiah Horrocks Institute for awarding me the part-time PhD funding that allowed me to carry out my PhD research.

A PhD would be a lonely journey if it was not for the friends you make along the way. Having been at UCLan for almost a decade, and being a post-graduate for six of those years, I have had the pleasure of working alongside many wonderful people. I would like to explicitly name two individuals who have made an impact on my academic and/or personal life. Dr Jordan J. Thirlwall: for your constant reminders to take breaks; for being my routine cuppa-tea companion (even over Skype!); and for continuing to be a friend after you finished your PhD and moved away. Secondly, Dr Janik Karoly: for your immeasurable moral support (particularly in the final, and

most difficult, year of my PhD); for helping me fix broken code; for your friendship, companionship and company; and for the ability to always make me laugh when I needed it the most.

In my personal life I am incredibly lucky to be surrounded by so many loved ones who have constantly encouraged and believed in me. Unequivocally, my first and foremost thanks is extended to Chang Hung Nelson Tsui (Nelly). I will forever be grateful to Nelly for helping me pick up the pieces in my personal life when things reached a low. He has supported me in more ways than one can count. Most of all, he helped me find my way back to the person that I am and the things that I love.

It goes without saying that I have an immense gratitude to my family; the people in my life who have cheered me on every step of the way since the beginning — my father (Nestor Lopez), my mother (Lisa Lopez), and my two sisters (Victoria Lopez and Kirstie Lopez). In particular, to my father, who was, and always will be, my original inspiration for pursuing an academic (and musical) career.

My musical life has always travelled in parallel with my academic life, remaining an important passion and joy. I am very fortunate to have found a violin teacher who has helped me to grow musically, and also take a great interest in my academic life. So, I would like to thank my violin teacher Fryderyk Andrasz for his tutorship and friendship. In addition, I would like to send out a special thanks to one of my long-standing violin students: Simon Ziegloser. Simon has been a loyal student, but more importantly, a wonderful friend. He has always provided dependable advice and offered moral support on many occasions over the years.

Finally, my acknowledgements would simply be incomplete without sending out a warm, loving thank you to my wonderful, furry friend, Diego (Figure 1). Thank you to Diego for all of the times you warmed my lap while I worked, snuggled in close when I needed consoling, and reminded me of the importance of cat-naps!

Figure 1: Picture of me with Diego (my cat) resting on my arm while I (attempt to) work from home at my desk.

Chapter 1

Introduction

To begin at the beginning... or on second thought, to begin at the present. It is commonly accepted that the standard cosmological model, the ΛCDM model, is the closest approximation to our observable Universe. Since we are advantageously living through an era of precision cosmology, we have, and continue to, accumulate observational data that allow *precise* (notably, independent of *accurate*) comparison with theory.

The two main themes of this thesis are the large-scale structure (LSS) of the Universe and the Cosmological Principle (CP) — later in this chapter there is an overview and discussion on both of these subjects. The CP provides cosmologists with a simpler view of the Universe; it assumes that on large scales the matter in the Universe is homogeneous and isotropic. Throughout the last century, this assumption has held, but the observational support for it has wavered, opinions of cosmologists may have even become somewhat polarised, and a likely root cause for this is the vague description and ill-established understanding of CP (e.g., what are 'large scales'?). Additionally, observational cosmology, like any observational science, is not without its complications; regardless of the advantages of modern technology, there can always be calibration errors or systematic oversights that ultimately lead to inaccurate conclusions. So, we are at a stage in cosmology where we are presented with both observational results that continue to support the ΛCDM model, and observational results that appear to challenge the ΛCDM model, either by challenging the assumptions of the CP or by being anomalous to predictions made by the standard model. For any science, I think it is sensible to push the boundaries of theory with observation, and even take the approach of 'prod until it hurts' to search for weaknesses. It is known that the ΛCDM model is just an approximation of our Universe, and to find the weaknesses of this model — where the ideas and theories are not yet complete — we must intuitively follow the subtle clues and hints presented by observations. In doing so, we continue to improve our understanding, and our theories, of the Universe.

In this thesis I present the observational analysis of two ultra-large large-scale structure (uLSS) discoveries that potentially challenge the CP: the Giant Arc (GA) and the Big Ring (BR). Being individual candidate structures, they do not possess the calibre to overturn cosmology as we know it, but as constituents of a growing list of LSSs (and uLSSs) that potentially challenge the CP, they do merit further investigation and the possibility that they are hinting towards new physics beyond the standard model. These ideas will be explored, particularly, the possibility that such uLSSs could arise from the presence of cosmic strings — topological defects formed in the early Universe. Admittedly, there does exist a large pool of alternative ideas in cosmology, some more fanciful than others, so we have been selective and cautious in our speculative suggestions regarding the origin of the two uLSSs (the GA and BR) presented here. We briefly consider an origin of the BR in Baryon Acoustic Oscillations (BAOs), and we passingly mention two alternative theories of cosmology: Conformal Cyclic Cosmology (CCC) and Modified Newtonian Dynamics (MOND). However, the speculations and suggestions are just that, and the bulk of this thesis is to present the observational results and analysis. (Indeed, a whole PhD could be spent investigating the validity, theoretical groundwork, and observational predictions of any one of those alternative ideas!) Nonetheless, I have included a brief discussion on cosmic strings and other ideas at the end of Chapter 6.

The reader will find there is emphasis on the observational anomalies found in cosmology and the implications they might have for the standard model of cosmology. The many successes of the ΛCDM model are well documented in the literature, and thus they will be left to stand as they are. But, it must not go unnoticed that there are other avenues in cosmology, such as alternative theories and new physics that extend the standard model, that may provide explanations to the observational anomalies, and in this sense it might be fruitful to consider them more seriously.

The results presented in this thesis are focused primarily on the analysis of LSS using the method of intervening Mg II absorbers in the spectra of distant quasars, i.e., mapping faint, intervening matter at intermediate redshifts. The Mg II method of analysing LSS is a somewhat novel method in cosmology, which was first presented in this way in my MSc thesis (Lopez, 2019); the method will be reiterated in chapter 3.

Finally, a sizeable portion of the work presented in this thesis has already been published in scientific journals (appended to the back), but additional analysis has in places been included for depth and clarity; the reader may compare the work here with the appended papers.

The rest of this chapter has been written to introduce the two key topics that will feature throughout the rest of this thesis: large-scale structure studies and the Cosmological Principle. General cosmology knowledge is assumed, so the purpose of the following introductory sections is to provide a current status review of the notable topics.

CHAPTER 1

1.1 Large-Scale Structure of the Universe

Large-scale structure (LSS) studies in cosmology map and analyse matter in the Universe on the largest scales. In this section, I discuss the current state of LSS studies including the relevant background, the motivations, and a summary of the largest LSSs in the literature.

LSS studies are motivated by the need for observational data to confirm the predictions of the Standard Cosmological Model (ΛCDM). In particular, from studying the LSS of matter on the very largest scales, one can learn about the growth of cosmic structure (Huterer, 2023) and about the Universe's dynamical history (Chow-Martínez et al., 2014), thus allowing comparison with ΛCDM. Furthermore, LSS studies can test the assumption of large-scale homogeneity, a fundamental aspect of the Cosmological Principle (CP; see the next section for a review), and hence of the theoretical framework in cosmology.

Before advancing further, I would like to address the term 'structure' in the context relevant to this work. In cosmology, the largest scale of coherent, ubiquitous structure is seen in the cosmic web. The cosmic web, a term coined by Bond et al. (1996), is the intricate network of filaments, walls, and sheets that are made up of gas, galaxies, and dark matter, which surround low-density regions, such as voids. The typical scale of the cosmic web is on the order of tens to hundreds of megaparsecs (van de Weygaert et al., 2016) and has been observed in many galaxy redshift surveys, e.g., the Two-degree Field Galaxy Redshift Survey (2dFGRS, Colless, 2004), the Sloan Digital Sky Survey (SDSS, Einasto et al., 2016), and the Two Micron All Sky Survey (2MASS, Huchra et al., 2012). Additionally, the cosmic web is reproduced in advanced cosmological simulations, such as the FLAMINGO project (Schaye et al., 2023). In these simulations the cosmic web is prevalent on huge, megaparsec scales (see Figure 1.1). The cosmic web is seeded by small quantum fluctuations and early-Universe processes; the high-density peaks and low-density

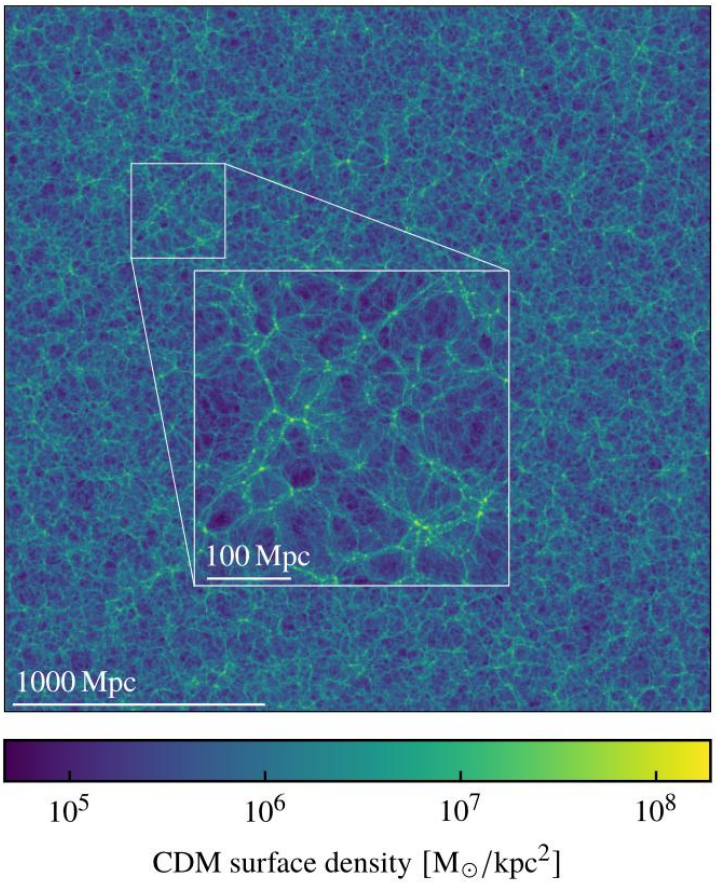


Figure 1.1: Figure taken from Schaye et al. (2023). This figure shows the intricate network of interconnected filaments, walls, and voids known as the cosmic web, as traced by the cold dark matter (CDM) using the FLAMINGO cosmological simulation. The figure shows a single snapshot of the CDM in a 20 Mpc thick box at $z = 0$.

voids become almost ‘fixed’ due to rapid expansion of the Universe during inflation (Einasto et al., 2016). These peaks and voids grow with the expansion, and matter accumulates along the high-density walls and filaments. On these huge scales ($\sim 10^1$ – 10^2 Mpc) we can see this ‘structure’, but note that here, the matter is not expected to be virialised; that is, cosmological structures are *not* expected to be gravitationally bound. So the term ‘structure’ in cosmology is referring to the geometrical patterns that can be observed in the matter distribution which can be assessed with a variety of statistical techniques. In cosmology, individual LSSs or even superclusters are generally established *algorithmically* and defined as a structure, but they are certainly not expected to be gravitationally bound.

Beyond the scales of the cosmic web ($\gtrsim 10^2$ Mpc) LSSs have been observed, and it is of interest for cosmology to understand them. In particular, in the CP we expect to observe a homogeneous Universe above some large scale; i.e., the scale of homogeneity (R_h). Yadav et al. (2010) gave $\sim$ 370 Mpc as an ideal or upper limit to the scale of homogeneity in concordance cosmology, beyond which departures from homogeneity should not be evident. As Yadav et al. state, above this scale it should not be possible to distinguish a given point distribution from a homogeneous distribution. We can therefore take $\sim$ 370 Mpc as an indication of the size (and, incidentally, separation also) beyond which LSS becomes cosmologically interesting. Table 1.1 lists numerically (by present-epoch size) the largest LSSs discovered to date. Of particular interest for cosmology are the ultra-large large-scale structures (uLSS) [1] with sizes $\gtrsim$ 370 Mpc. Many of the sizes quoted in this table will be somewhat uncertain, usually because of uncertainty in the boundaries, but sometimes because of uncertainty in what is being quoted in the papers (e.g. cosmological model and parameters, cosmological epoch).

It is not yet agreed whether the growing list of LSS discoveries poses a problem

---

[1] We have introduced the new term ‘ultra-large LSS’ (uLSS) to denote those structures that exceed the Yadav estimated $\sim$ 370 Mpc upper limit to the scale of homogeneity (Yadav et al., 2010).

Name	**Mean z**	**Size (Mpc)**	**References**
HCB Great Wall	~ 2	2000–3000	Horváth et al. (2014); Horvath et al. (2020)[1]
Giant GRB Ring	0.82	1720	Balázs et al. (2015)
Correlated LQG orientations	1.0–1.8	1600	Friday et al. (2022)
U1.27, Huge-LQG	1.27	1240	Clowes et al. (2013)[2]
Giant Arc	0.802	1000	Lopez et al. (2022)
Coherent quasar polarisation[3]	1–2	1000	Hutsemékers (1998); Hutsemékers & Lamy (2001); Hutsemékers et al. (2005)
U1.11	1.11	780	Clowes et al. (2012)
U1.28, CCLQG	1.28	630	Clowes & Campusano (1991); Clowes et al. (2012)
Sloan Great Wall	0.073	450	Gott et al. (2005)
South Pole Wall	0.04	420	Pomarède et al. (2020)
King Ghidorah Supercluster	0.55	400	Shimakawa et al. (2023)
Big Ring	0.802	400	Lopez et al. (2024)
Blazar LSS	~ 0.35	350	Marchã & Browne (2021)
Local void	< 0.07	300	Keenan et al. (2013); Whitbourn & Shanks (2016)
BOSS Great Wall (BGW)	0.47	250	Lietzen et al. (2016)
Great Wall	0.029	240	Geller & Huchra (1989)
Saraswati supercluster	0.28	200	Bagchi et al. (2017)

Further references.
[1] See also: Christian (2020)
[2] See also: Nadathur (2013); Marinello et al. (2016); Hutsemékers et al. (2014)
[3] See also: Marchã & Browne (2021)

Table 1.1: Some of the largest LSSs reported in the literature. The columns are: the name of the LSS; the mean redshift; the reported size in Mpc (present epoch); and references. This table has been adapted from Table 1 in Lopez et al. (2022).

for the ΛCDM model or CP. For the ΛCDM model, Poisson statistics and more complex mock catalogues have been used to draw comparisons of observations with theoretical expectations. For the CP, the question of whether the uLSSs are posing a challenge is down to our particular interpretation of what the CP really means (discussed in the next section).

With regards to the above, on testing the ΛCDM model with Poisson statistics and mock catalogues, consider the following examples. In a study from Park et al. (2012), the authors looked at SDSS-like mock samples of galaxies using the Horizon Run 2 N-body cosmological simulation and found that structures comparable to the Sloan Great Wall (SGW) are easily reproduced. However, the authors also point out that the comparable structures were always amongst the six largest and richest structures in their 200 mocks. The results indicate that uLSSs may be consistent with ΛCDM predictions provided they are the most-extreme, rare cases. On the contrary, N-body cosmological simulations were found to be lacking in the power-law behaviour that is observed in the probability distribution of supercluster size (De Marzo et al., 2021; Chow-Martínez et al., 2014), and thus there is scepticism on whether the N-body cosmology simulations can accurately represent the largest-scale features of the Universe. In another example, work done by Nadathur (2013) showed that structures comparable to the Huge Large Quasar Group (Huge-LQG) were easily reproduced in a simple Poisson point distribution. Then, counter to this argument, Marinello et al. (2016) showed that while structures comparable to the Huge-LQG are reproducible using Horizon Run 2 N-body cosmological simulations, a structure as large as the Huge-LQG would be expected to be the largest structure in a survey sample five times their investigated sample — in this case, the authors looked at an SDSS-like footprint in the redshift range $1.2 \leq z \leq 1.6$. These examples showcase the complications of LSS studies, and the opposing (polar) views on whether uLSSs are compatible with the ΛCDM model.

It is worth noting however, that LSS studies and uLSS discoveries are of more interest for cosmology than simply the statistical probability of their existence and the compatibility of that with ΛCDM. Since LSSs are not expected to have reached virialisation, they may carry important information about the primordial initial conditions, which makes them worth studying (De Marzo et al., 2021; Chow-Martínez et al., 2014; Oort, 1983). The most interesting of all is their formation, and whether such ultra-large LSS are seeded in the initial conditions, or whether they require more exotic solutions for their existence, such as cosmic strings (e.g., Peebles, 2023).

1.2 A Review of the Cosmological Principle

A fundamental aspect of the ΛCDM standard cosmological model is the cosmological principle (CP). The CP is an assumption that states that on 'large scales' our Universe looks both homogeneous and isotropic. In reality, it was not (and is still not) known if this was true, but in assuming it to be true (thought to be a reasonable assumption at the time) the equations of general relativity then become much simpler to solve analytically. The equations of general relativity are the foundation of the standard cosmological model; if the CP is disproved, what does this then mean for cosmology? Fortunately, it is quite possible that there is a simple solution to this. One interpretation of the CP (see below) states that there should be similarity everywhere. Since the term 'large scales' is never explicitly defined, this gives us the freedom to adapt the definition based on what is observed. That is, where we find deviations from homogeneity in the matter distribution (and isotropy, but here we are interested in only the homogeneity in particular) then clearly homogeneity must emerge *beyond* this scale; the expectation of what is a plausible scale of homogeneity seems to have increased by at least a factor of ten over the years.

Consider the following three interpretations of the CP, and specifically what is meant by homogeneity. (i) There exists some large scale, known as the scale of homogeneity on which the Universe can be smoothed and the distribution of matter would then be well represented by a stationary random process, e.g., Peebles (1993); Yadav et al. (2010); Aluri et al. (2023). This is what I would consider the ‘classical’ definition of the CP as it is prevalent in the earliest works and discussions. (ii) The power spectrum suggests that there can always be some large scale at which statistically-significant deviations might be found in the matter distribution, but such deviations on large scales should be rare, e.g., Park et al. (2012); Chow-Martínez et al. (2014). (iii) There should be similarity everywhere (maximally symmetric). Any observed large-scale structuring indicates that the scale of homogeneity, if it exists, must then be larger than these scales. The occurrence of a particular LSS, even the largest known, does not imply that the *probability* of finding a comparable LSS elsewhere is any different. Notice that points (ii) and (iii) could be contradictory, given that point (ii) suggests that the largest structures should be rare, and point (iii) suggests that the largest structures need not be a problem for the CP if their probabilities are homogeneous and isotropic. (How would we know that?) A useful overview of the various interpretations of the CP can be found in Schwarz (2010).

The different interpretations of homogeneity in the CP has led to opposing views on whether the observed matter supports a homogeneous Universe. Large-survey analysis has frequently supported a homogeneous Universe, e.g., in luminous red galaxies and in quasars (Andrade et al., 2022; Gonçalves et al., 2021, 2018). Yet, the uLSS discoveries listed in Table 1.1 might indicate that homogeneity is not supported from the classical understanding of the CP (see point (i) above). So in this sense, the accumulating set of uLSSs appear to challenge the CP. (Note that the word ‘challenge’ is not synonymous with ‘contradict’, but it does imply something to be investigated further.) In practice, it will be more important to settle on a universal definition of the CP before any real conclusions can be made regarding the homogeneity of the observed matter distribution. But, as stated earlier, the study of the LSS of the Universe, and uLSSs in particular, will provide hints towards our current theories of the Universe and thus are worthy of further investigation.

1.3 Cosmological Parameters

Throughout this thesis the concordance model is adopted for cosmological calculations, with $\Omega_{T0} = 1$, $\Omega_{M0} = 0.27$, $\Omega_{\Lambda 0} = 0.73$, and $H_0 = 70$ kms^{-1}Mpc^{-1}. All sizes given are proper sizes at the present epoch.

Chapter 2

The Physics and Astrophysics of Mg II Absorbers

Singly-ionised magnesium (Mg II) exists in the roughly spherical, low-ionised, gaseous haloes hosting galaxies and clusters. The Mg II absorption doublet, with rest wavelengths of 2796 Å and 2803 Å, present in the spectra of distant quasars, is detectable in optical wavelengths for a broad range of redshifts ($0.4 \lesssim z \lesssim 1.7$). Together, these two properties make the Mg II doublet a useful and reliable tracer of the low-luminosity circumgalactic medium that traces galaxies and galaxy clusters.

The Mg II haloes are generally quite patchy and extended. Figure 2.1 is taken from Huang et al. (2021a) and shows 15 randomly-selected galaxy-quasar pairs from the Sloan Digital Sky Survey (SDSS) that have spectroscopic Mg II counterparts. The figure aims to demonstrate how the azimuthal angle of the galaxy-quasar orientation might affect the Mg II equivalent widths. (In Section 2.2 I review the topic of Mg II absorbers and their physical environment.) Here it provides a useful example of the typical projected distances for galaxy-quasar pairs from the on-sky perspective. Similarly, Figure 2.2 from Guha et al. (2024) shows a selection of (DESI) Dark Energy Camera Legacy Survey (DECaLs) quasar field r-band images hosting ultra-strong Mg II absorbers. Again, the figure provides additional examples of

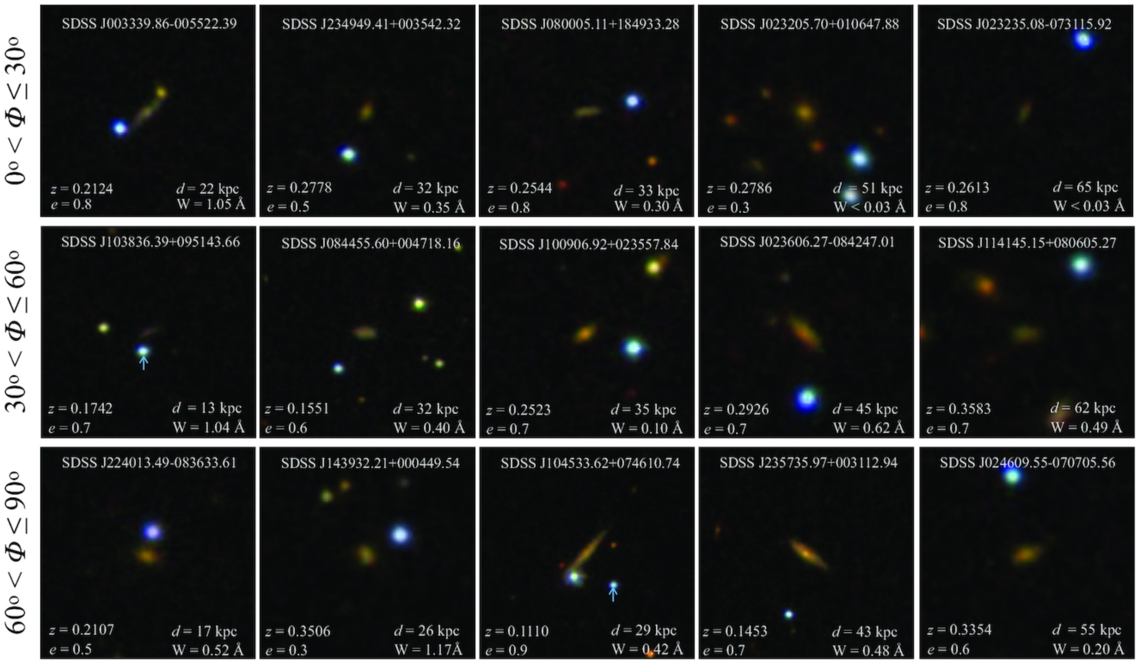


Figure 2.1: Figure taken from Huang et al. (2021a) showing 15 randomly-selected galaxy-quasar pairs in the SDSS DR6 catalogue. Each panel contains the Mg II host-galaxy in the centre, the associated quasar which appears as a blue source (in some cases, labelled with a blue arrow for clarity), and labels of: the galaxy redshift (z), the ellipticity of the galaxy (e), the projected distance (d), and the Mg II rest equivalent width (W).

the typical projected distances of galaxy-quasar pairs from the on-sky perspective. (Note, in ten of the small subfigures there was no optically-identified galaxy pair within the search parameters of that quasar field.)

In this chapter I will cover the background physics of the Mg II absorption doublet and then the astrophysical and cosmological applications of the Mg II absorbers.

2.1 The Physics of the Mg II Absorption Doublet

Absorption lines in the spectra of quasars tell us about the intervening matter along the sight-line of the quasar. The shape, profile, intensity and wavelength of the absorption lines tell us about the properties of the intervening matter. In the context of this work, we are interested in only the Mg II absorption doublet in particular.

The electronic configuration of MgII is the same as neutral sodium: $1s^2 2s^2 2p^6 3s^1$.

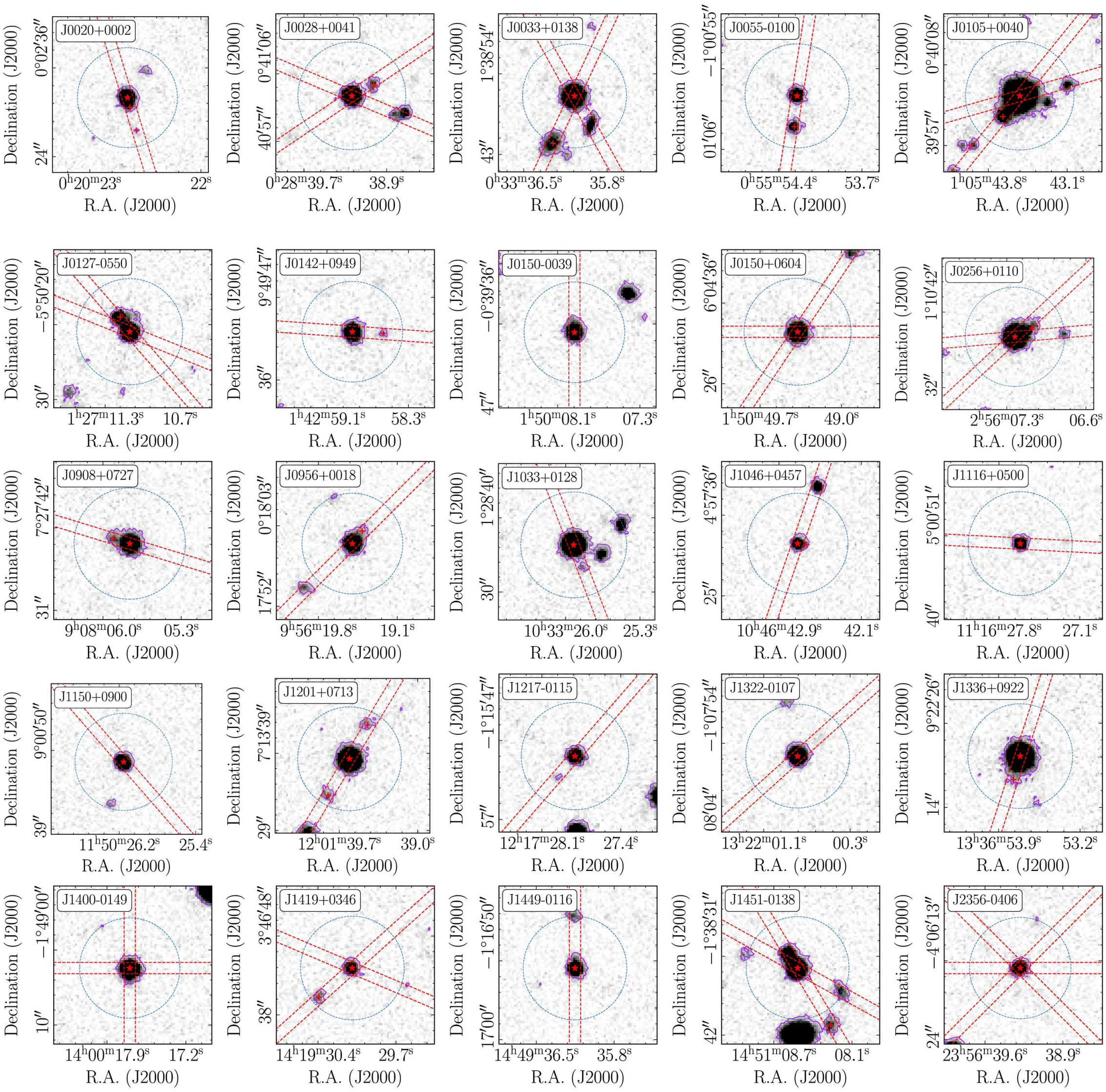


Figure 2.2: Figure taken from Guha et al. (2024) showing a selection of (DESI) Dark Energy Camera Legacy Survey (DECaLs) quasar field r-band images hosting ultra-strong Mg II. Each panel contains: the background quasar in the centre of the field, marked with a red star shape; the identified host-halo galaxy marked with a red plus sign; a dashed circle representing an impact parameter of 50 kpc at the redshift of the absorber; and the slit orientation for the spectroscopic observations of the host-galaxies (using the Southern African Large Telescope; SALT) marked with straight, red, dashed lines.

When the Mg II gas absorbs light at a specific frequency, the unpaired outer electron in the $3s$ orbital can be excited to the $3p$ orbital. Due to magnetic interaction between the orbital angular momentum $\vec{L}$ and the spin angular momentum $\vec{S}$ of the electron, the $3p$ orbital is split into two states. The total angular momentum of the fine-split $3p$ states can be calculated $\vec{J} = \vec{L} + \vec{S}$, where $\vec{L} = 1$ in the $3p$ state and $\vec{S} = \frac{1}{2}$. The two fine-split states in the $3p$ orbital have total angular momentum $\vec{J} = \frac{1}{2}$ (denoted $3p_{\frac{1}{2}}$) and $\vec{J} = \frac{3}{2}$ (denoted $3p_{\frac{3}{2}}$). The electron transition occurs between the $3s$ orbital and both of the fine-split $3p$ states resulting in the doublet feature of the Mg II absorption doublet. The statistical weight corresponding to the $3p$ orbital states are calculated $g = 2J + 1$, meaning: the $3p_{\frac{1}{2}}$ state, corresponding to the 2803 Å line, has a statistical weight of $g = 2$ and the $3p_{\frac{3}{2}}$ state, corresponding to the 2796 Å line, has a statistical weight of $g = 4$. Since the statistical weight of the 2796 Å line is twice as large as the statistical weight of the 2803 Å line, their line strengths will also follow the same ratio of 2 : 1. Of course, in practice, when measuring the line strengths of the Mg II absorption doublet, one will more often find that the line-strength ratios of the Mg II doublet are much closer to unity due to saturation; the intrinsic line-strength ratio can only be observed if the absorption lines are optically thin.

2.2 The Physical Environment of Mg II Absorbers

Mg II absorbers (and other sharp, metal absorption lines generally) have been studied in depth for their astrophysical and cosmological applications; for some of the pioneering work see Bergeron (1988); Steidel & Sargent (1992); Churchill et al. (2005); Kacprzak et al. (2005, 2008); Barnes et al. (2014). However, the precise physical nature of the Mg II absorbers still remains unclear. It is thought that the characteristics of the Mg II absorbers (e.g., strength, velocity dispersion, covering fraction) could be affected by, or correlated to, the specific host-halo environment of the Mg II gas. Potential contributing factors to the Mg II absorption-line properties could be: the impact parameter, inclination and position angle; the galaxy morphology; if the host environment is an isolated galaxy, group or cluster; host halo radius; star-formation rate (SFR) and if the galaxy is star-forming or quiescent; and stellar/galaxy mass and luminosity. Analysing and understanding the connection between the Mg II gas and its host-halo environment will help to constrain the theories of the general evolution and dynamics of galaxies.

In early works, it was thought that Mg II gas could be characterised by a simple spherical halo extending up to ~ 40 kpc from the galaxy centre. However, this view rapidly changed with observational follow-up of the associated Mg II absorbing galaxies (Churchill et al., 2000a,b, 2005; Kacprzak et al., 2005). In some cases the follow-up observations revealed examples of strong Mg II absorbers associated with galaxies well beyond the expected absorber halo radius ($D > 40$ kpc), suggesting that the Mg II haloes are highly extended. Churchill et al. (2005) confirmed that the Mg II systems are patchier and more complex than previously thought. The complexity of the Mg II gas distribution is largely responsible for the difficulty in determining the contributing factors of the Mg II-line profiles. In Churchill et al. (2000b), the authors performed 114 tests for correlations between the galaxy properties and the Mg II absorption line characteristics; seven of these tests showed tentative significant correlation. However, in most of these tests, the results were found to be heavily skewed by the inclusion or removal of the damped Lyman-alpha (DLA) systems, which are generally not drawn from a homogeneous set. The authors did find that the strongest correlation was that of the Mg II gas kinematics versus the position of the spectrum on a diagram of the Mg II rest equivalent width (EW_{MgII}) versus the C IV rest equivalent width (EW_{CIV}). That is, the position of a Mg II doublet feature on a EW_{MgII} vs EW_{CIV} diagram will be closely correlated with the kinematic complexity of the Mg II doublet (see Figure 2.3).

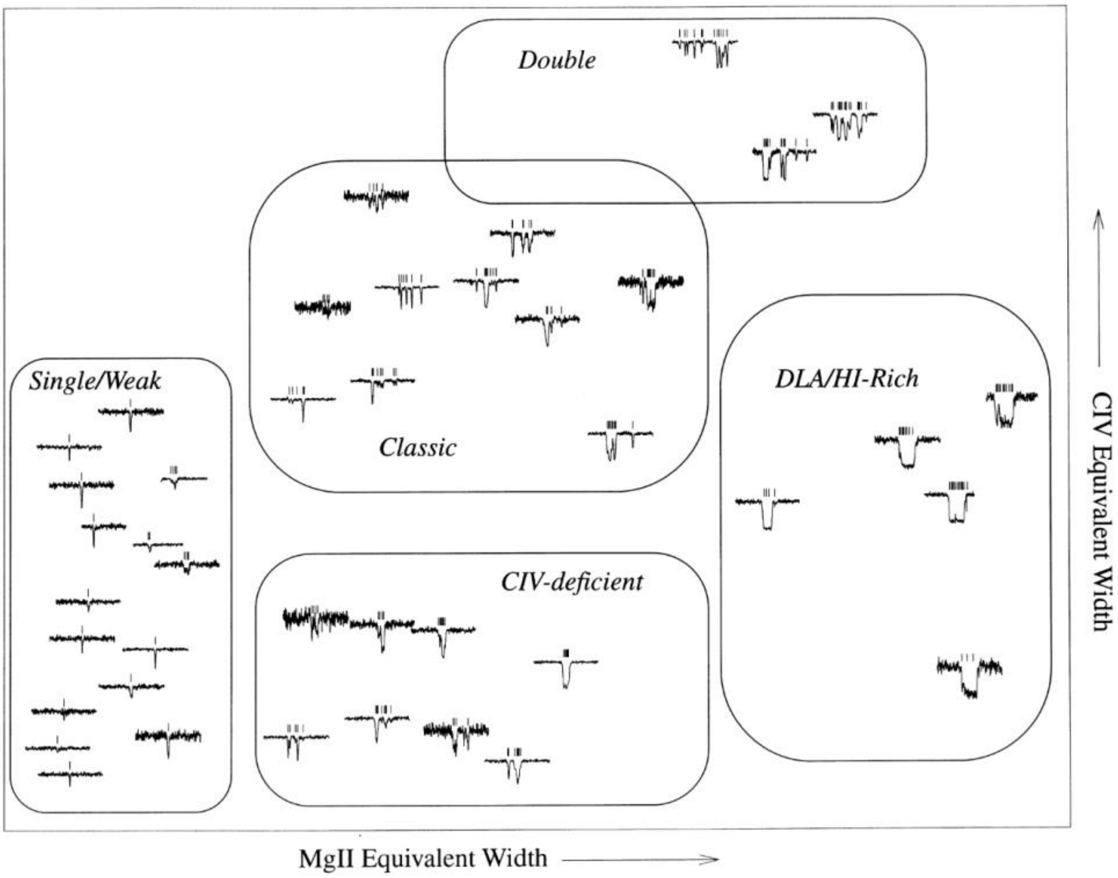


Figure 2.3: Figure taken from Churchill et al. (2000b) showing the Mg II gas kinematics on a diagram of Mg II (rest) equivalent width (EW) versus C IV EW. The different classes of absorbers appear to occur in specific loci on the diagram represented by the outlines. This figure demonstrates that the kinematic 'complexity' of the Mg II absorption profiles can be categorised into 5 different classes: Single/Weak, C IV-deficient, Classic, Double, and DLA/HI-rich. Each class of absorber can be defined in terms of where it falls in the Mg II EW versus the C IV EW diagram. In general, the kinematic complexity appears to be the greatest when the C IV EW is large.

It is worth noting that in the Mg II literature one will often find examples and counter-examples of correlations between the MgII-line profile and MgII host-halo environment. For example, in Lanzetta & Bowen (1990) the authors found a significant anti-correlation between the EW of the Mg II line and the impact parameter. This led to the generally-accepted simplified spherical view of the host halo of a galaxy (see Figure 3 from Churchill et al., 2005, for an adaptation). It was not immediately realised that the damped Lyman-alpha (DLA) systems were in fact dominating this correlation. Similarly, weak Mg II absorbers were thought to be associated with a particular class of galaxies such as low-surface brightness galaxies or dwarf galaxies (Churchill et al., 1999). Then, in Churchill et al. (2005) the authors utilised high-quality images taken from the Hubble Space Telescope (HST) of absorption-selected galaxies with confirmed spectroscopic redshifts, combined with the associated Mg II absorption spectra, and concluded that the weak Mg II absorbers do not in fact select a particular class of galaxies. Instead, the authors found normal galaxies directly associated with the weak Mg II absorbers. Their work also showed that the weak Mg II systems were detected over a wide range of impact parameters ($0 < D < 80\ h^{-1}$kpc) reinforcing that (i) the strength of the absorber is not simply anti-correlated to the impact parameter and (ii) Mg II environments are far more complex and patchy than previously thought. Finally, however, later works showed that the relationship between Mg II EW and impact parameter might *again* hold, but for specifically isolated star-forming galaxies (not groups or clusters) (Huang et al., 2021b). This to and fro of the general consensus in Mg II studies highlights the complexity in understanding the Mg II environments. It is important, but extremely complicated, to disentangle the Mg II absorption-line properties and the contributing environmental factors.

At present, many efforts continue to be made to understand the connection between the Mg II absorption-line characteristics and the host-halo environment. The cool ($T \sim 10^4$ K) gas around galaxies and galaxy clusters, known as the circumgalactic medium (CGM), can be studied for understanding the formation and evolution of galaxies. For example, the cool gas can trace the intricate gas outflows and show where gas is accreting onto the galaxy. The extremely complex nature of the Mg II haloes is the primary challenge to understanding the Mg II / host-halo link.

In the current Mg II literature, one will find discussions relating to: the impact parameter; isolated vs non-isolated galaxies (and quiescent or star-forming); and host-halo radius, mass and luminosity. Studies have frequently shown that the strength of the Mg II absorber is anti-correlated with impact parameter, as discussed above and also in more recent studies (Cherrey et al., 2024; Huang et al., 2021a). However, a recent study from Guha et al. (2024) found that the host-galaxies of ultra-strong Mg II absorbers did not show any anti-correlation of the Mg II EW against impact parameter. The authors suggest that the strong anti-correlation usually seen is a consequence of sample bias. However, the authors did find that normalising the impact parameter to the virial radius reduced the scatter in their results, which was also found in Churchill et al. (2013); consequently, an anti-correlation was seen between the Mg II EW and the normalised impact parameter.

In agreement with Guha et al. (2024) concerning possible sample bias, Huang et al. (2021a) found that the anti-correlation of the Mg II EW and impact parameter was seen in only isolated, star-forming ($\mathrm{EW}_{H\alpha} > 5$ Å) galaxies, but no trends were found in the isolated, quiescent galaxies, and non-isolated galaxies (see Figure 2.4). Additionally, the authors found that the Mg II EWs (in isolated, star-forming galaxies) are dependent on the host-halo radius, galaxy luminosity and stellar mass.

In another study, Cherrey et al. (2024) study the Mg II counterparts of galaxy groups and conclude that the Mg II gas must trace primarily the outflows of individual star-forming galaxies rather than indicating simply the abundance of low-ionised

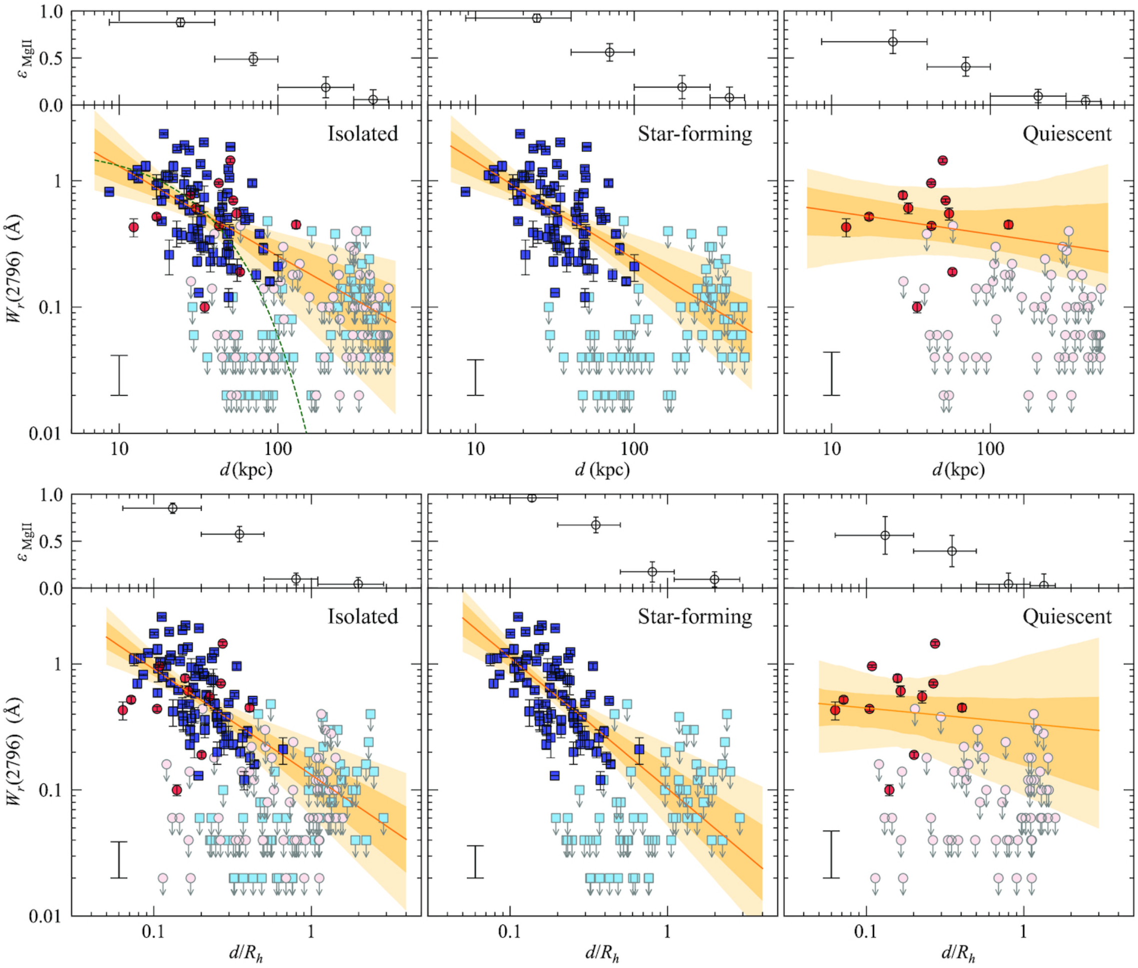


Figure 2.4: Figure taken from Huang et al. (2021a) showing the strong anti-correlation of the Mg II equivalent width (EW) versus the impact parameter (top row) and normalised impact parameter (bottom row) for their isolated galaxy sample. The left-hand side panels are the full isolated galaxy sample. The star-forming galaxies ($\mathrm{EW}_{H\alpha} > 5$ Å) are shown in blue squares and the quiescent galaxies are shown in red circles. Galaxies with detected Mg II absorbers are shown in solid symbols with error bars representing measurement uncertainties. Galaxies without detected Mg II are displayed as 2σ upper limits (points with downward arrows) and shown in a lighter colour. The orange solid lines represent the best-fit power-law models. The dark and light shaded bands represent the 68% and 95% confidence intervals from the Markov chain Monte Carlo realisations. The best-fit intrinsic scatter is marked in the lower left corner. At the top of each panel, the authors show the best-fit results of the 4 non-parametric mean covering fractions from the maximum likelihood analysis. The horizontal bars mark the full range of projected distance within each bin and vertical error bars represent the 68% confidence interval. See the full paper for more details. It can be clearly seen that the strong anti-correlation is a result of the star-forming galaxies, seen in the middle panels. In the right-hand side panels, showing the isolated, quiescent galaxies, no trend can be seen for the EWs.

gas, since the Mg II EWs were not found to be typically stronger in groups. In agreement with this, Lee et al. (2021) investigated the rate of Mg II absorption in and around clusters of galaxies. They found that, although the detection rate per quasar is higher inside the clusters, the rate is in fact quite low when considering the number of galaxies in the clusters — that is, the absorber-to-galaxy ratio is lower inside clusters, presumably because the environment within clusters modifies the galaxy haloes. A review of the Mg II literature shows that there is still no clear consensus on the connection between host-halo environment and Mg II-line properties, which is in large due to the highly-complex nature of the Mg II gas distribution.

For the purpose of the work presented in this thesis, we are concerned with only tracing galaxies and the gas associated with their environments on large scales, which can readily be interpreted from the Mg II absorbers. But, conceivably, understanding the precise physical nature of the Mg II systems should then lead to a better understanding of the physical environments, or properties, of the LSSs traced by Mg II. Therefore, the reader will find in Chapters 4 and 5 some preliminary investigations of the Mg II EW distributions in the small fields containing the Giant Arc and the Big Ring.

Chapter 3

Data and Method

The work presented in this thesis makes use of the method of intervening Mg II absorbers in the spectra of quasars to trace faint matter at intermediate redshifts, e.g., Lopez (2019); Lopez et al. (2022, 2024). The Mg II method relies on both the spectroscopic measurement of luminous, high redshift quasars from the Sloan Digital Sky Survey (SDSS) (Schneider et al., 2010; Pâris et al., 2017; Lyke et al., 2020), and the precise, spectroscopic redshifts of the intervening Mg II absorption doublets present in the quasar spectra, documented by independent authors (Zhu & Ménard, 2013; Anand et al., 2021). The intervening Mg II absorption doublet indicates the presence of galaxies and galaxy clusters (Bergeron, 1988; Churchill et al., 2005; Steidel, 1995). Together, with the quasars and Mg II absorbers, we have the information of the on-sky position of intervening matter (the same RA-Dec coordinate as the corresponding quasar position), and the redshift of the intervening matter (obtained from the measured wavelength of the Mg II doublet), so in mapping the 3D distribution of the Mg II absorption features in the spectra of quasars, we can infer the LSS of intermediate-to-high redshift, faint matter.

In this Chapter I discuss the data sources of the Sloan Digital Sky Survey (SDSS) quasars, the corresponding Mg II catalogues, and how those data are presented in the Mg II images that are used for LSS analysis. A brief overview of each of the statistical methods that are used in the analysis of the Giant Arc (Chapter 4) and Big Ring (Chapter 5) is introduced at the end of this Chapter.

3.1 The SDSS Quasar Catalogues

The Sloan Digital Sky Survey is currently in its fifth edition, SDSS-V. Each edition of the SDSS is complete with a final quasar catalogue; they are as follows.

DR7QSO (Schneider et al., 2010) compiles all of the quasar observations from (primarily) the Legacy survey of the SDSS-I (2000–2005) and SDSS-II (2005–2008). Additional quasars were obtained on a series of ‘special plates’ from two other surveys: the Extensions of the Legacy target selection algorithms along the Fall equatorial region, and the Sloan Extension for Galactic Understanding and Explorations (SEGUE). There are a total of $105,783$ quasars in this catalogue (see Figure 3.1).

DR12Q (Pâris et al., 2017) compiles all the quasar observations from the Baryon Oscillation Spectroscopic Survey (BOSS) of the SDSS-III (2008–2014) for which there are a total of $297,301$ quasars (see Figure 3.2). Of these, $272,026$ were new and unique discoveries since the beginning of SDSS-III and $25,416$ were re-observed from the SDSS-I/II (with $80,367$ of the $105,783$ DR7QSO quasars not included in this DR12Q catalogue).

Finally, DR16Q (Lyke et al., 2020) compiles all of the known quasars from Legacy, BOSS and the extended Baryon Oscillation Spectroscopic Survey (eBOSS) of the SDSS-IV (2014–2020) to create the largest catalogue of known quasars (see Figure 3.3). There are a total of $750,414$ quasars in this final catalogue, of which $225,082$ were appearing for the first time in an SDSS data release. (Note, other intermediate quasar catalogues that will not be discussed here are DR9Q, DR10Q and DR14Q.)

In figures 3.1 to 3.3 the large-scale inhomogeneities in the RA-Dec positions of the quasars can be seen which originate primarily from inhomogeneous survey sampling.

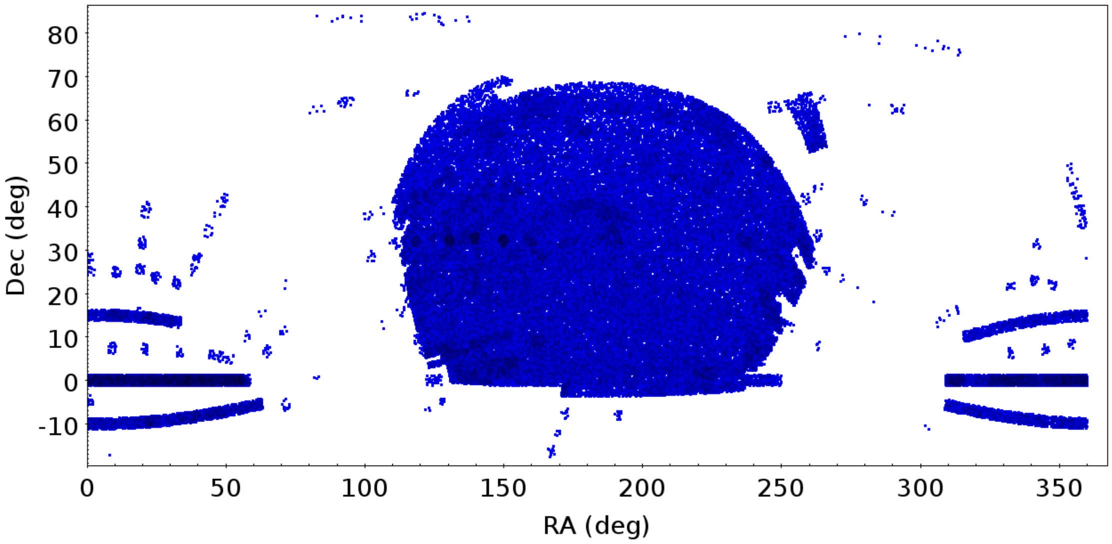


Figure 3.1: The RA-Dec positions of 105, 783 SDSS DR7QSO quasars from the Legacy survey of SDSS-I/II.

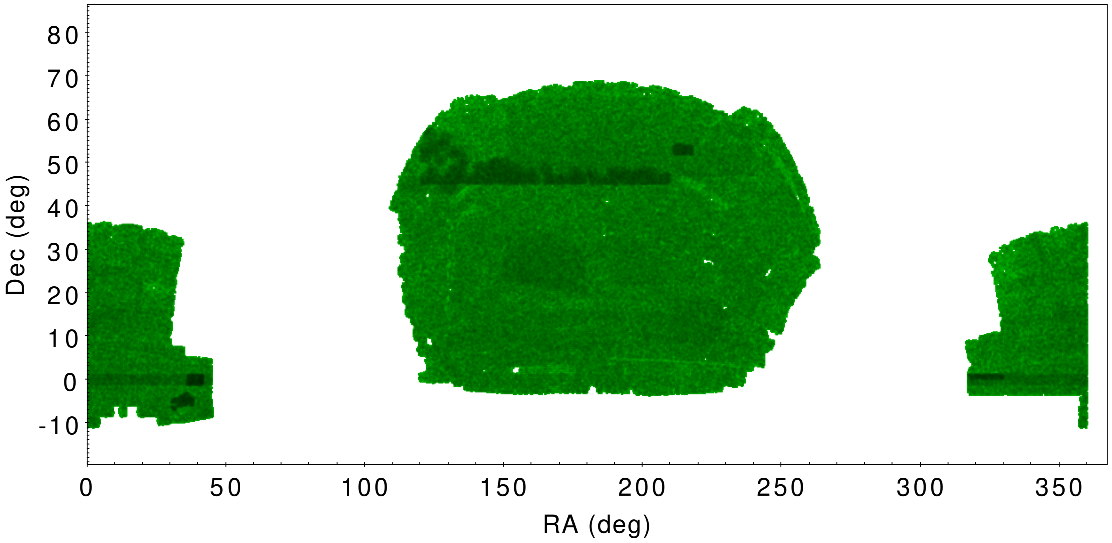


Figure 3.2: The RA-Dec positions of 297, 301 SDSS DR12Q quasars from BOSS of SDSS-III.

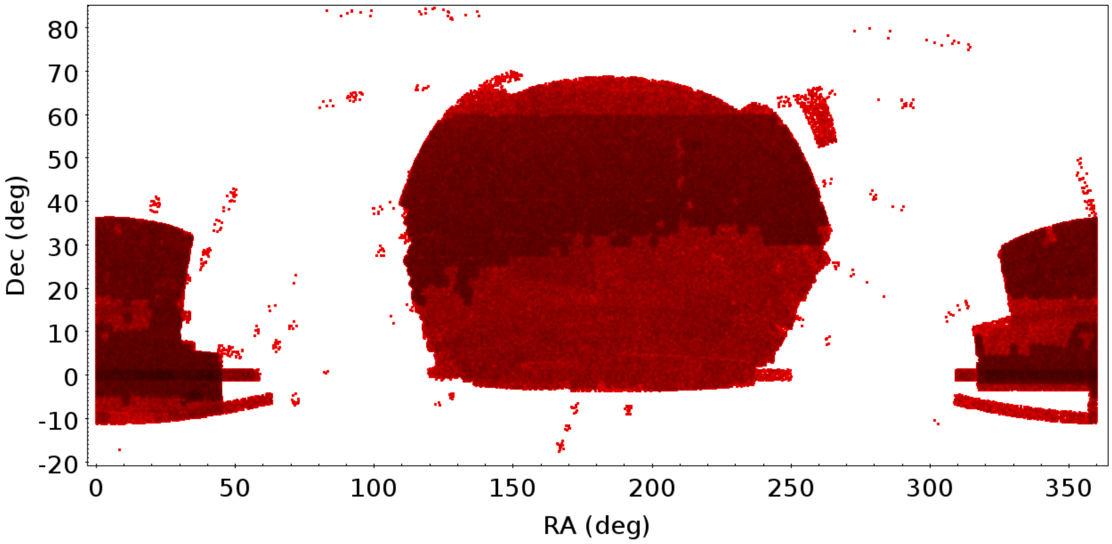


Figure 3.3: The RA-Dec positions of 750, 414 SDSS DR16Q quasars from Legacy, BOSS and eBOSS of SDSS-IV.

There will also be small-scale inhomogeneities which could be intrinsic to the data or from survey bias. In particular, in Figure 3.3 of the DR16Q quasars there are very hard borders above and below the high-density band in the north of the main (centre) sampled area. Since the background quasars are used as probes for the intervening matter, and we are assessing the large-scale structure of intervening faint matter with the Mg II absorbers, it is important that the quasar inhomogeneities are taken into account. Additionally, where there are fewer quasars as background probes we are limited to detecting the intervening matter there (at that sky coordinate). In reality, dealing with the data is very tricky and complicated, but these limitations need not be a restriction on investigating LSS.

Between SDSS-I/II and SDSS-III, the spectrographs used for taking quasar observations were upgraded. The old SDSS-I/II used two spectrographs, each with a blue and red channel, split with a dichroic filter at ~ 6000Å. Each spectrograph had 320 fibres with fibre entrance aperture sizes of $3''$. The SDSS-III (and later) used the two BOSS spectrographs which were rebuilt from the older spectrographs. The two BOSS spectrographs also have a red and blue channel, with a splitting around

6000 Å. The spectrographs were improved with an increased number of 500 fibres per spectrograph and with smaller fibre entrance apertures of $2''$. The wavelength coverage increased from (3800–9200) Å on the old spectrographs to (3600–10, 500) Å on the newer BOSS spectrographs, each with a spectral resolution of $\lambda/\Delta\lambda \approx 2000$.

In the next section I discuss the Mg II catalogues that are derived from the above quasar catalogues.

3.2 The Mg II catalogues

There are two independent authors that created the Mg II catalogues that we use: Zhu & Ménard (2013) whose Mg II catalogues are derived from the DR7QSO and DR12Q quasars, and Anand et al. (2021) whose Mg II catalogues are derived from the DR16Q quasars.

Previously, we were using the Zhu & Ménard (2013) Mg II catalogues (Z&M hereafter), and it was with these data that the discovery and analysis of the Giant Arc (GA; Chapter 4) was made and presented. Note that, although Z&M's 2013 paper only refers to the DR7QSO quasars, they applied their absorption pipeline to the later DR12Q quasars, which was then made publicly available. Eventually, the SDSS DR16Q quasar database became available, and the second independent authors Anand et al. (2021) provided a corresponding Mg II catalogue (Anand21 hereafter). We then switched from the older Z&M data to the newer Anand21 data; the discovery and analysis of the Big Ring (BR; Chapter 5) was made and presented with these newer data.

The GA and BR are contained within the same small field, at the exact same redshift. Since the GA was found in the Z&M (older) catalogue, we would expect to see the GA again in the Anand21 (newer) catalogue. Yet, the GA appeared markedly different when using the newer Anand21 data (see Figure 3.4). Intuitively, this might be placed on the huge increase in quasar observations from DR7QSO/DR12Q

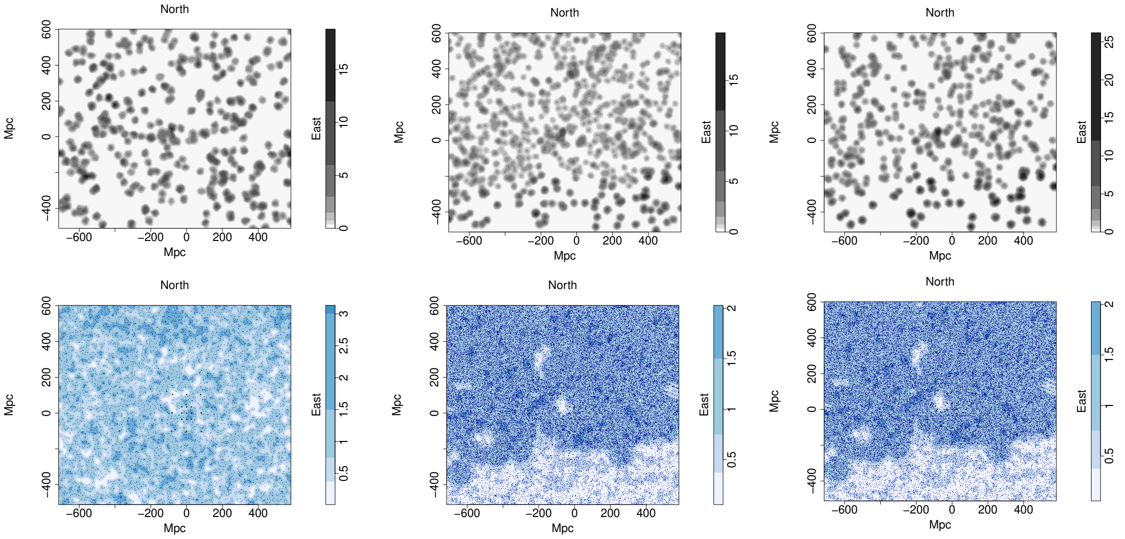


Figure 3.4: Top panels: the Giant Arc (GA) field in the Z&M data (left), the Anand21 data (middle) and the Anand21 data with the signal-to-noise (SN) cuts of $4, 2, 4$ applied to the $\lambda\lambda$ $2796, 2803$ lines and the quasar continuum, respectively (right). i magnitude cuts were made in the initial read-in stage of the data such that $10.0 \leq i_{DR16} \leq 20.5$, $10.0 \leq i_{DR12} \leq 20.2$ and $10.0 \leq i_{DR7} \leq 19.1$ (see below). The SN that are applied to the Anand data help to reduce the noise in the field, given the much greater number density of quasars and absorbers in the field (see Chapter 5). Bottom panels: the corresponding background quasars (probes) in the GA field for each of the three datasets in the top panels. The GA in the original data was visually very obvious, but in the new data it appears markedly different. One reason for the visual difference between the datasets is due to the increase in the number density of absorbers and quasars in the new Anand21 data. By reducing the Anand21 data with SN cuts, the GA on the right-hand side begins to resemble the original GA. Importantly, despite the visual impression, the GA is still detected as the most significant, and largest (by membership) candidate structures in the Single-Linkage Hierarchical Clustering and Convex Hull of Member Spheres algorithms — see Section 5.3.1 and Figure 5.7. Note, the Big Ring (BR) is just north of this figure.

to DR16Q, and thus associated Mg II absorbers. However, we spotted that a substantial fraction of Z&M absorbers in the GA were not identified by Anand21, even though visual inspection[1] confirmed each of the GA absorber members were real, positive detections.

I have taken two practical approaches to comparing the two Mg II catalogues from the independent authors: (i) investigating the general properties of the catalogues and (ii) investigating a small field and comparing the final selection of quasars and absorbers in both catalogues.

For the Z&M data, we constructed our Mg II absorber database from the publicly available catalogues[2]. We have used their DR7 and DR12 'Trimmed' catalogues (not the 'Expanded' versions). DR7 and DR12 indicate that the sources of the quasars that have been used as background probes are from data releases 7 and 12 of the SDSS. Further details on the detection of the absorbers and the construction of the catalogues is described in Zhu & Ménard (2013) and the catalogue website. We paired the Z&M absorber catalogues on RA, Dec to the 'cleaned' quasar databases DR7QSO (Schneider et al., 2010) and DR12Q (Pâris et al., 2017). Thus the absorbers can all be associated subsequently with either DR7QSO or DR12Q. We removed entries for repeat spectra within the Z&M DR12(Q) absorber catalogue, thus avoiding duplication of absorbers. There were no entries for repeat spectra within the Z&M DR7(QSO) absorber catalogue. When a particular absorber (RA, Dec, z) appeared in both the Z&M DR7(QSO) and DR12(Q) catalogues, we removed the DR7(QSO) entry, thus giving preference to DR12(Q) parameters. The final database has $63,876$ Mg II absorbers. We also produced a corresponding database of probes from the Z&M 'Quasars searched' catalogues, similarly restricting them to

---

[1] For each of the GA absorber members, we checked each of the corresponding quasar spectra and searched for the Mg II absorption doublet (rest wavelengths $\lambda\lambda$ $2796, 2803$ Å) at the reported redshifts. We found positive detections for each of the Mg II absorbers in the GA with the correct wavelengths and wavelength ratios.

[2] https://www.guangtunbenzhu.com/jhu-sdss-metal-absorber-catalog

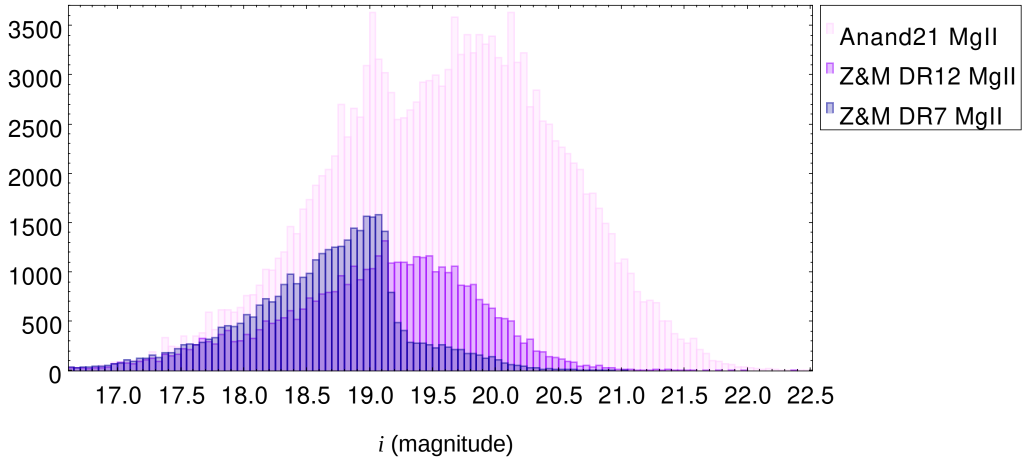


Figure 3.5: The i magnitude distribution of the background probes (quasars) of the Anand21, Z&M-DR7, and Z&M-DR12 Mg II catalogues. When applying our preliminary cuts to the raw catalogues shown here, we choose to make i magnitude cuts according to where the number counts begin to drop off in these histograms.

those that appear in either DR7QSO or DR12Q. This database has $123,351$ member quasars. Next, we standardise the tables in our process of reading in the catalogues, and apply some preliminary cuts. For the absorbers corresponding to DR7, we apply an i magnitude limit to the quasar continuum of $10 \leq i \leq 19.1$, and for the absorbers corresponding to DR12, we apply an i magnitude limit to the quasar continuum of $10 \leq i \leq 20.2$; this ensures a roughly uniform sample selection (see Figure 3.5). For the 'Quasars searched' (combined) catalogue that we produced, we then applied an i magnitude limit of $10 \leq i \leq 20.2$ to match the criteria of the DR12-associated absorbers. There are totals of $56,535$ Mg II absorbers and $117,300$ quasars (probes) in our version of the combined, standardised (processed) Z&M Mg II catalogues and corresponding quasar databases.

For the Anand21 data, we constructed our Mg II absorber database from the publicly available catalogues[3]. The Anand21 Mg II catalogue corresponds to the SDSS DR16Q quasars which contains a total of $750,414$ unique quasar observations, of which $225,082$ are new observations. For constructing the corresponding Mg II

---

[3] https://wwwmpa.mpa-garching.mpg.de/SDSS/MgII/

catalogue, Anand21 downloaded all of the spectra for objects classified as quasars, for which there were $983,317$ objects. For further details of the detection of the absorbers and the construction of the catalogues see Anand et al. (2021) and the catalogue website. We downloaded the SDSS QSO-based Mg II catalogue, for which there are $159,524$ Mg II absorbers. In addition, we downloaded all of the SDSS DR16Q quasars (object database; for which there are $750,414$ quasars). Then, we paired the Anand21 database with the SDSS DR16Q quasars, restricting to one spectrum per quasar. Thus, all the spectra will correspond to those in the DR16Q listing. As before, we standardised the catalogue tables in our process of reading in the Mg II and quasar data and apply i magnitude cuts of $10 \leq i \leq 20.5$ to both the Mg II absorbers and quasars to select a roughly uniform sample. A short summary of the standardised Anand21 and Z&M catalogues is presented in Table 3.1.

With our processed versions of Mg II catalogues from both authors, I compare the following general properties: RA-Dec area of quasar positions and density of quasars; redshift distribution; i magnitude of the background probes; S/N of the probe (quasar) continuum; the EW of the $\lambda 2796$ line ($W_{r,2796}$); and the error in the EW of the $\lambda 2796$ line (see figures 3.8 – 3.12). I discuss each of the figures below.

Figure 3.6 shows the Anand21 Mg II RA-Dec positions. As would be expected, the absorber density mirrors that of the DR16Q quasars in Figure 3.3: in particular, the dense region of Mg II absorbers to the north of the main (centre) sampled area is still apparent even with the i magnitude cuts, suggesting the overdensity is largely a result of survey-sample bias (simply re-observing that particular area much more than elsewhere). By contrast, the RA-Dec positions of the Mg II absorbers in the Z&M catalogue in Figure 3.7 appears much more uniform after the i magnitude cuts compared with the DR7QSO and DR12Q quasars in figures 3.1 and 3.2.

The redshift distributions in the Anand21 and Z&M catalogues in Figure 3.8 follow the same general shape with a steep rise in absorbers from $0.4 \lesssim z \lesssim 1.0$, and

	Anand21	**Z&M combined**
# Mg II absorbers	98, 747 (159, 534)	56, 545 (63, 876)
# probes	356, 350 (750, 414)	117, 300 (123, 351)
abs / area (deg^{-2})	∼11	∼6
# unique probes	239, 621	571

Table 3.1: The numbers of Mg II absorbers and searched quasars in the processed Anand21 and Z&M combined catalogues. Note, these values are *after* we applied our own basic criteria to the raw catalogues (see the main text), but I have included the raw numbers in parentheses. The rows from top to bottom are: the number of Mg II absorbers in the catalogues; the number of searched quasars (probes) in the corresponding cleaned quasar database; the number density of absorbers, assuming an approximate footprint area of 9400 square degrees from SDSS; and the number of probes that are unique to those authors' catalogue.

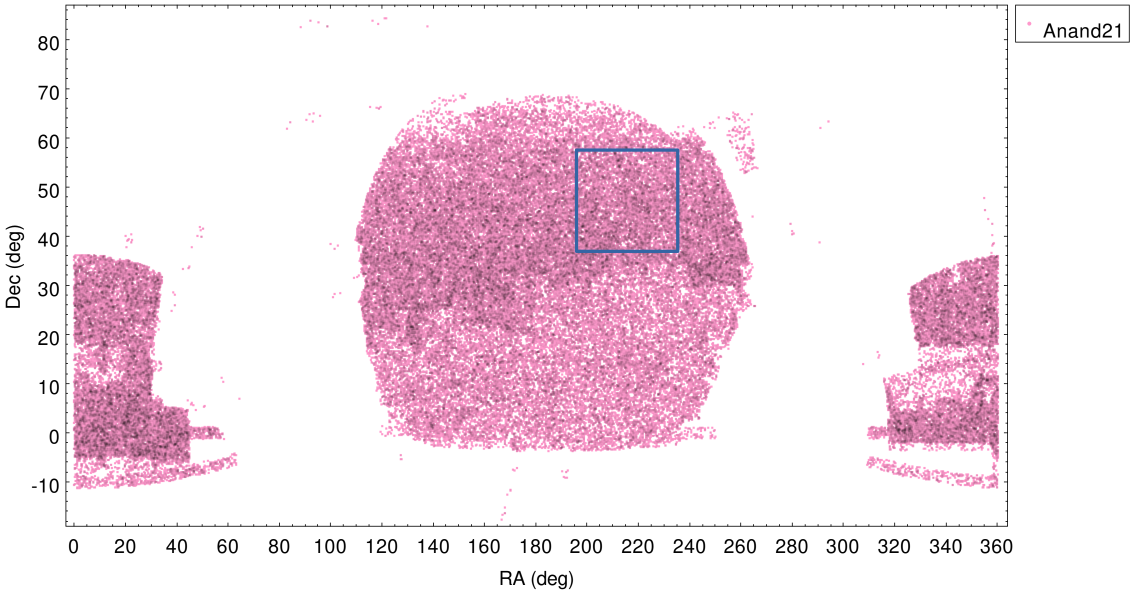


Figure 3.6: RA-Dec positions of the Mg II absorbers in the Anand21 catalogue. Notice the high-density region to the north of the main (centre) sampled area compared with the low-density region to the south of the sampled area, as was seen in the DR16Q quasars in Figure 3.3. The blue square outlines the approximate area of the Giant Arc/Big Ring field of view for future reference.

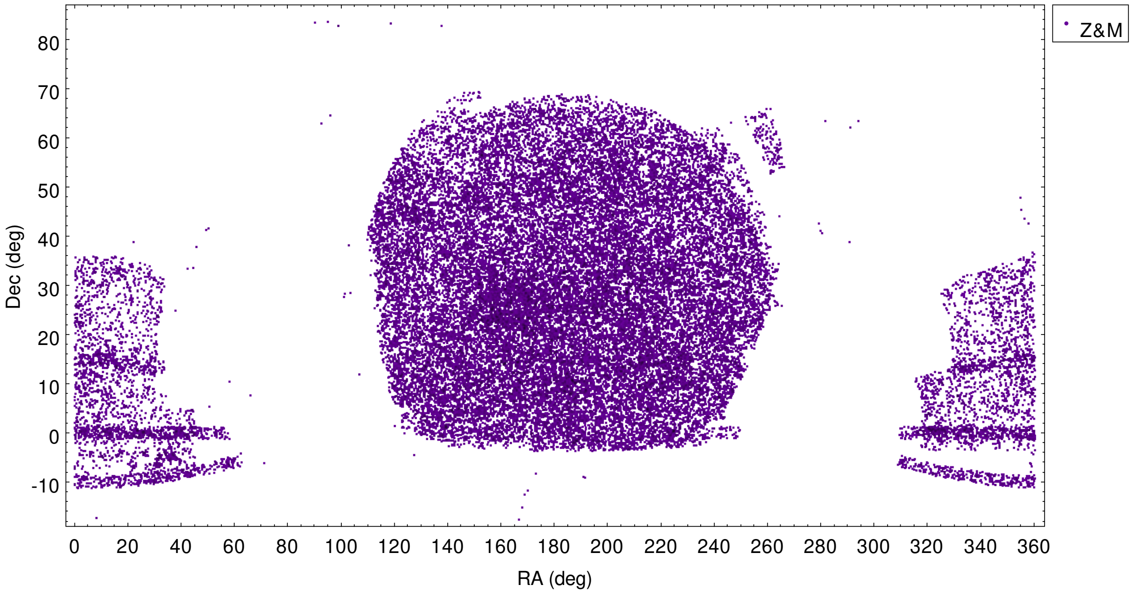


Figure 3.7: RA-Dec positions of the Mg II absorbers in the Z&M catalogue. The i magnitude cuts that we made to the quasars in this catalogue have made a close-to uniform selection, but large-scale density artefacts can still be seen, e.g., roughly centre of the main (centre) sampled area.

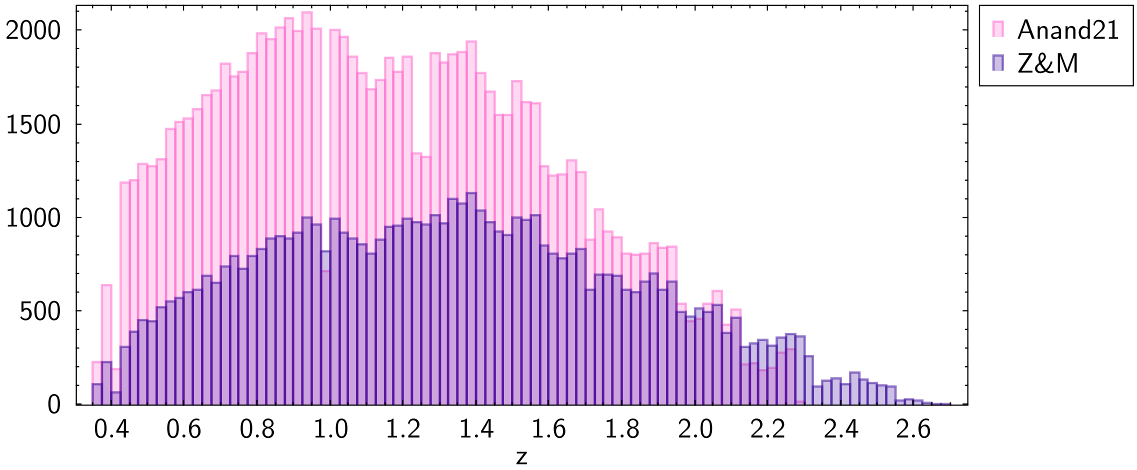


Figure 3.8: The redshift distribution of the $\lambda2796$ line in the Anand21 and Z&M catalogues.

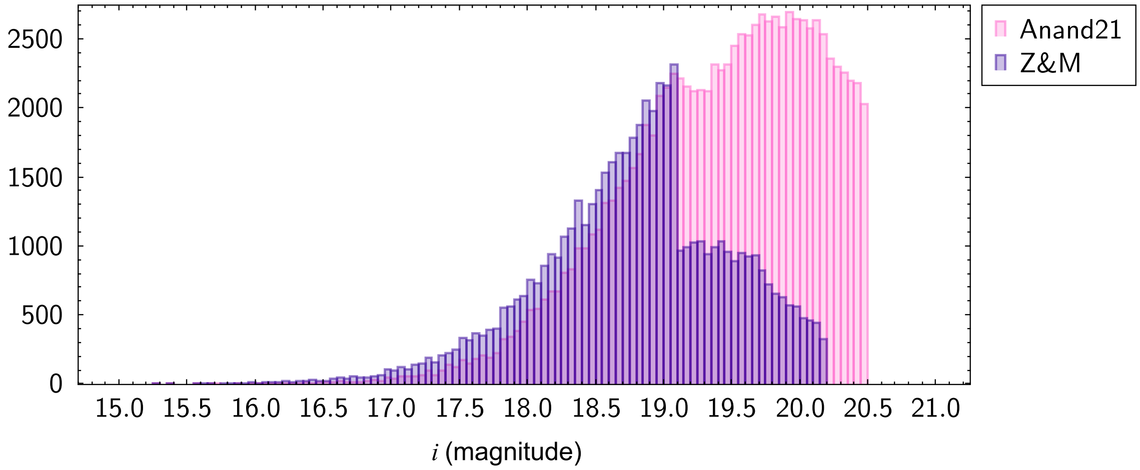


Figure 3.9: The i magnitude distribution of the background quasars (probes) in the Anand21 and Z&M catalogues. Note, the sharp, end cut-offs in both catalogues show where we have applied our i magnitude cuts which were applied based on the decline in coverage seen in Figure 3.5.

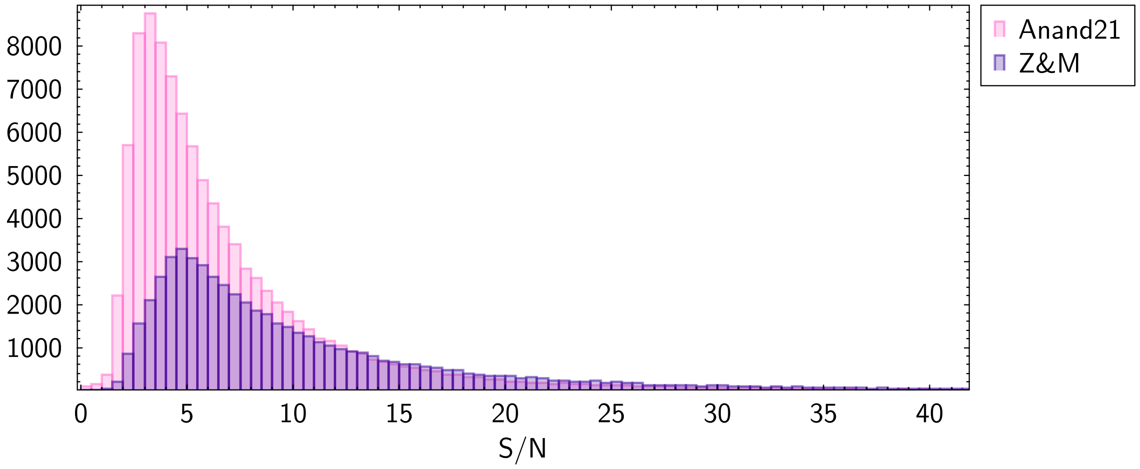


Figure 3.10: The continuum signal-to-noise (S/N) distribution of the background quasars (probes) in the Anand21 and Z&M catalogues.

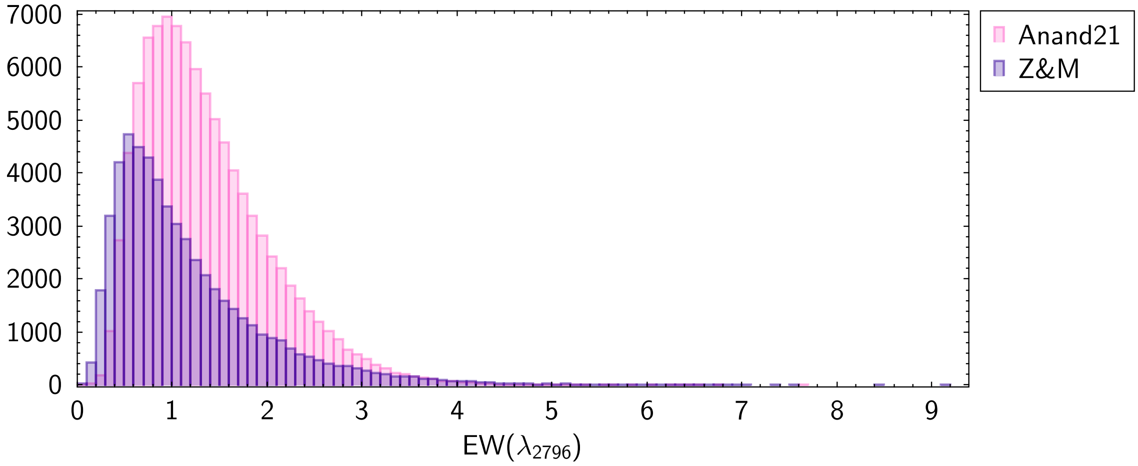


Figure 3.11: The rest equivalent width (EW) distribution of the λ2796 line in the Anand21 and Z&M catalogues.

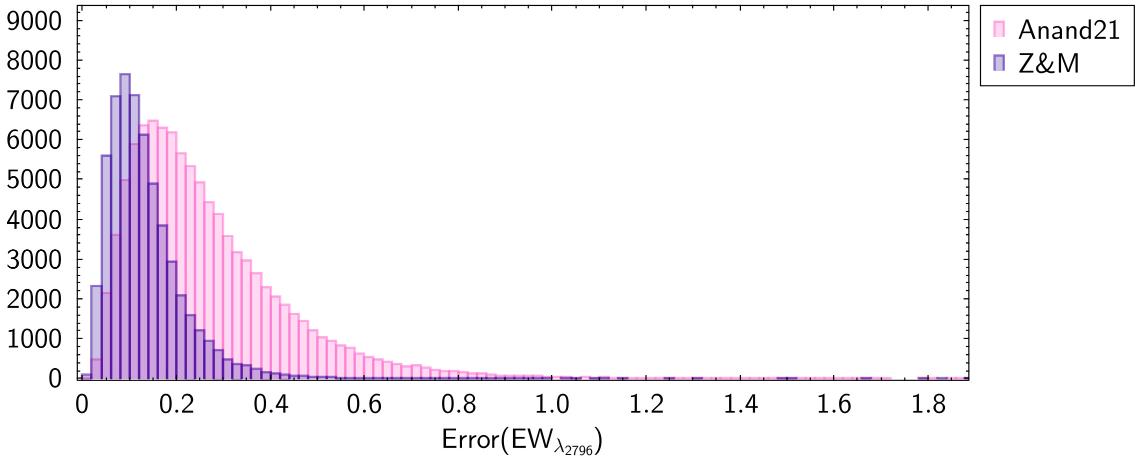


Figure 3.12: The error in the rest equivalent width (EW) distribution of the λ2796 line in the Anand21 and Z&M catalogues.

	Anand21			**Z&M**		
	Min	*Max*	*Mean*	*Min*	*Max*	*Mean*
$\mathbf{z}_{qso}$	0.428	4.656	1.871	0.415	4.653	2.029
$\mathbf{z}_{2796}$	0.360	2.277	1.172	0.359	2.678	1.332
i	15.252	20.5	19.409	15.259	20.2	18.798
$\mathbf{S/N}_{cont.}$	1.350	75.662	10.094	3.369	76.258	14.444
$W_{r,2796}$	0.048	7.645	1.366	$\sim$0	9.104	1.166

Table 3.2: A comparison of the catalogues. The row names, from the top to bottom, are: quasar (probe) redshift; Mg II λ2796 redshift; i magnitude of the quasars; signal-to-noise (S/N) of the quasar continuum; and rest equivalent width ($W_{r,2796}$) of the λ2796 line.

a gradual drop-off from around $z \sim 1.4$. Briefly, it is worth noting here the redshift ranges over which the independent authors applied their line-detection algorithms. Around a quasar's C IV emission line there tends to be a high covering fraction of C IV, so absorption doublets detected near the C IV emission line are more likely to be C IV doublets and not Mg II doublets. Due to the similar wavelength ratios of the C IV doublet and Mg II doublet, it can be tricky to discern the two doublets. Consequently, the Z&M authors begin their search window $\Delta z = 0.02$ redward of the C IV emission line to avoid false positives and similarly, the Anand21 authors begin their search window $\Delta z = 0.018$ redward of the C IV emission line. For the Z&M catalogue, the authors chose to include quasar-associated absorbers, so they extend their search window $\Delta z = 0.04$ redshifted from the quasar's Mg II emission line. Conversely, Anand21 choose to detect only the intervening Mg II absorbers, not the quasar-associated absorbers too, so they end their search window to $\Delta z = 0.03$ blueward of the Mg II emission line. Sharp drops in coverage can be seen in the Anand21 catalogue which are due to masking the various sky lines such as O I λ 5577 (corresponding to absorbers at $z \sim 1.0$), O I λ 6300 (corresponding to absorbers at $z \sim 1.25$) and the high-pressure sodium bump at λ 5900 (corresponding to absorbers at $z \sim 1.1$). The Z&M raw catalogues also mask the same sky lines, but the additional magnitude cuts (discussed above) help to create a more homogeneous catalogue.

The i magnitude distribution of the background quasars is seen again in Figure 3.9 but this time showing our processed versions of the Z&M combined catalogue (DR7 and DR12) and the Anand21 catalogue. The sharp cut-offs in both catalogues show where we have applied our i magnitude cuts which were applied based on the decline in coverage seen in Figure 3.5.

The continuum signal-to-noise (S/N) distribution is seen in Figure 3.10. The Z&M S/N distribution peaks at around ~ 5 which is slightly higher compared with the Anand21 S/N distribution; this is as expected since Anand21 probe to fainter quasar magnitudes. Otherwise, the two distributions have similar shapes.

The rest equivalent width (EW) distribution and the corresponding errors are seen in Figures 3.11 and 3.12. Here, in both figures, the two distributions are quite clearly offset. The EW in the Z&M catalogue peaks at $W_r < 1.0$, whereas the Anand catalogue EW distribution peaks at $W_r > 1.0$, indicating that the Z&M catalogue probes to lower EWs than Anand21. The errors in the Z&M catalogue are lower than those in Anand21, which is expected given that Anand21 have access to more lower-luminosity quasars from SDSS DR16Q. A summary of the above catalogue properties is presented in Table 3.2.

Although Anand21 use the Z&M approach as a starting point for their own Mg II absorption detection pipeline, their final catalogue is visually quite different. Here we investigate where some of these differences might originate. For the large majority of the work presented in this thesis, we have been interested in only the small field containing the Giant Arc and the Big Ring. For this reason we perform the small sample comparison between the Mg II catalogues documented by Anand21 and Z&M

using the BR/GA field. The following text describes how we performed our small sample comparison. Note, without access to the independent authors' absorption detection pipelines, it may not become definitively clear why the catalogues differ; our goal here is simply to highlight some commonalities and differences between the catalogues.

The sample comparison field is centred approximately on RA, Dec coordinates $(215, 45)$ deg, covering an area of ~ 1200 deg^2, in the redshift interval $z \simeq 0.80 \pm 0.06$. When running the Mg II scripts with each database, no additional limits or cuts to the data were made (e.g., magnitude (i) limits, signal-to-noise (S/N) cut-offs etc.). However, the input data do have our i limits applied to the quasars and Mg II absorbers corresponding to each SDSS data release such that: $10.0 \leq i_{DR16} \leq 20.5$, $10.0 \leq i_{DR12} \leq 20.2$ and $10.0 \leq i_{DR7} \leq 19.1$.

There are three important data files produced from the Mg II scripts:

1. MGII_4: Mg II absorbers in the FOV.
2. MGII_2: Background quasars, from the Mg II catalogues, in the FOV with redshifts greater than the far edge of the Mg II slice.
3. PRBS_2_B: Background quasars, from the SDSS quasar catalogues that were initially searched for Mg II absorbers, in the FOV, again with redshifts greater than the far edge of the Mg II slice.

Using Topcat (Taylor, 2005), the data files were paired in various combinations for analysis. Ultimately, there are three questions we sought to answer: (i) how many unmatching absorbers are there (those found by one author but not the other); (ii) how many of these absorbers that were missed were simply due to missing quasars (from DR7/DR12 to DR16 there were $\sim 860,000$ new quasar observations); (iii) how many of the unmatched absorbers are either *real* or *false positives*. The results of the first two questions are presented in table 3.3, but the third question would require individual inspection of the 154 quasar spectra containing unique absorbers (i.e., the spectra in rows ID 13 and 14 of table 3.3). Note, the number of objects present in the MGII_2 data files, and combinations thereof, have been left out of table 3.3 for conciseness. A summary of the results are as follows.

Using the Anand21 catalogue, we begin with $25,315$ probes in the field of view (FOV) which belong to the 'quasars searched' database (PRBS_2_b) — these are the quasars that were originally searched for Mg II. Of those $25,315$ in the FOV there are 8017 ($\sim 32\%$) within the FOV with detected Mg II in these spectra (MGII_2; not necessarily at the specific redshift we are here investigating). Using the Z&M catalogue, we begin with 7294 probes in the FOV which were searched for Mg II. Of those 7294 probes there are 4314 ($\sim 59\%$) with detected Mg II absorption in their spectra.

Of the searched quasars (PRBS_2_b), Anand21 and Z&M matched on 7257 quasars in common, meaning 37 of the searched quasars from the DR7QSO / DR12Q catalogues were not included in the newer DR16Q catalogue. Since the DR16Q catalogue is constructed from a superset of all of the quasars observed by SDSS we can assume that the small percentage of quasars that have not been included in the DR16Q catalogue have been excluded for not having good or standard quasar spectra. For the Mg II absorbers, there were 852 Mg II absorbers out of $25,315$ checked quasars ($\sim 3.4\%$) in the Anand21 database (for the FOV), and 359 Mg II absorbers out of 7294 checked quasars ($\sim 4.9\%$) in the Z&M database (again, for the FOV). From the Mg II absorbers, there were 255 Mg II absorbers in common between the Anand21 and Z&M databases; so there were 597 unique Mg II absorbers in the Anand21 database and 104 unique absorbers in the Z&M database. However, of the 597 Mg II absorbers that were unique to the Anand21 database, 530 of these absorbers were detected in quasars that were not present in the Z&M quasar (PRBS_2_b) database.

ID	Topcat pairing	Description	Number
1	PRBS_2_b (Anand21)		25 315
2	PRBS_2_b (Z&M)		7 294
3	PRBS_2_b (both)	Searched quasars from quasars available to both authors.	7 257
4	PRBS_2_b (Anand21 not Z&M)	Searched quasars from quasars only available to Anand21	18 058
5	PRBS_2_b (Z&M not Anand21)	Searched quasars from quasars only available to Z&M.	37
6	MGII_4 (Anand21)		852 705 with i <20.2 234 with i <19.1
7	MGII_4 (Z&M)		359
8	MGII_4 (both)	Mg II absorbers found by both authors.	255
9	MGII_4 (Anand21 not Z&M)	Mg II absorbers only found by Anand21.	597 (30 of which are multiples per probe here occurring in 15 spectra)
10	MGII_4 (Z&M not Anand21)	Mg II absorbers only found by Z&M.	104 (2 of which are in the same spectrum)
11	MGII_4 (Anand21 not Z&M) and PRBS_2_b (Anand21 not Z&M)	Mg II absorbers found by Anand21 that were from quasars not available to Z&M.	530
12	MGII_4 (Z&M not Anand21) and PRBS_2_b (Z&M not Anand21)	Mg II absorbers found by Z&M that were from quasars not available to Anand21.	1
13	MGII_4 (Anand21 not Z&M) and PRBS_2_b (both)	Mg II absorbers found by Anand21 but not by Z&M even though the quasars were searched by both.	52
14	MGII_4 (Z&M not Anand21) and PRBS_2_b (both)	Mg II absorbers found by Z&M but not by Anand21 even though the quasars were searched by both.	102 (2 of which are paired in the same spectrum)

Table 3.3: Table showing the number of objects in the quasar and Mg II absorber data files in a sample field in the older Z&M database versus the newer Anand21 database. The column ‘Topcat pairing’ refers to the ‘match selection’ function in Topcat to pair the tables in various ways, and the ‘Description’ column describes the output file. The column ‘Number’ simply gives the total number of objects in that data file.

The last two rows in table 3.3 are the Mg II absorbers of particular interest. These are the Mg II absorbers detected by only one author, from the subset of corresponding probes that were searched by both authors. Visual inspection of each of these quasar spectra would be onerous. However, based on the catalogue comparisons from earlier we might suggest that one of the reasons that Z&M found so many additional Mg II absorbers that Anand21 missed was that Z&M probed to much lower equivalent widths (see Figure 3.11). Similarly, it might be that, generally, Anand21 remove a higher fraction of spurious absorbers at the cost of losing real absorbers, whereas Z&M might be providing a catalogue with higher completion at the cost of including some spurious absorbers. Furthermore, in Chapter 5, we discuss how in the catalogues of Z&M, the Giant Arc (GA) appears markedly different than in the catalogues of Anand21. After we visually-inspect the Mg II absorbers belonging to the GA (all of which are real detections) that were missed by the Anand21 catalogues, we find a potential trend for the missed absorbers to have generally complex, broad or weak profiles. This justifies the interpretation of the more conservative Anand21 catalogue versus the more liberal Z&M catalogue.

3.3 Constructing the Mg II Images

The Mg II density images seen throughout this thesis (and in the corresponding published papers) are intended to give a useful impression of the connectivity of the absorbers (see Figure 3.13). The images are constructed by smoothing the 2D distributions of the Mg II absorbers and the background probes (quasars) with a Gaussian kernel, with the same smoothing scale for both. In a process of 'flat-fielding', the absorber image is then divided by the normalised probe image to correct for non-uniformities in the distribution of the probes on that smoothing scale. The grey contours in the Mg II density images increase by a factor of two.

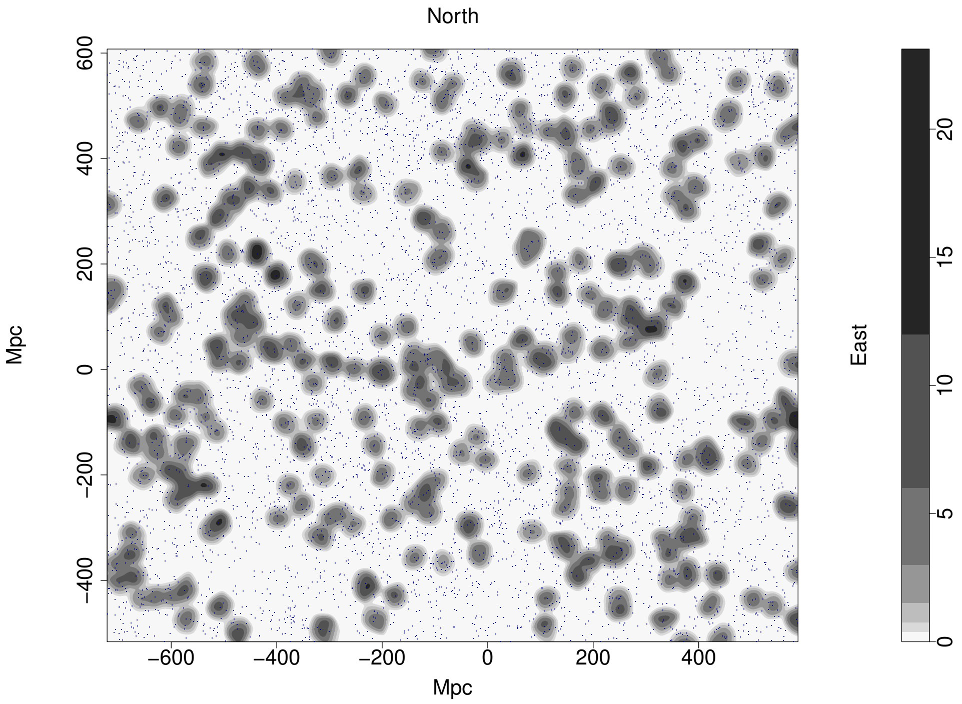


Figure 3.13: The tangent-plane distribution of Mg II absorbers from the Mg II catalogues of Zhu & Ménard (2013) centred in the redshift interval $z = 0.802 \pm 0.060$. The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes. The dark blue dots represent the background probes (quasars). The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top. The Giant Arc (discussed next in Chapter 4) can be seen running west-east in the centre of the figure, spanning ~ 1 Gpc.

We use tangent-plane coordinates, scaled, using the central redshift, to present-epoch proper coordinates in Mpc. All of the figures that are presented in this thesis, and the corresponding published papers, are orientated with east towards the right and north towards the top.

3.3.1 A Note on Redshift Errors

The use of Mg II absorbers for analysing LSS has the particular advantage of providing very precise redshifts. A disadvantage of course, which can require very careful handling, is that one must take the background probes — the quasars — where they are given by the catalogues. The catalogues now have suitably dense coverage on the sky, but generally they are affected by variations in selection criteria.

At $z_{abs} \sim 0.8$, the parameter in the Anand21 catalogues for the redshift error (Z_ABS_ERR) indicates a median error of $\sigma_{z_{abs}} \approx 4.2 \times 10^{-5}$. (The emission-redshift error for a quasar catalogue might be about two orders of magnitude larger at ~ 0.004.) This $\sigma_{z_{abs}}$ corresponds to a velocity difference of ~ 7 km s^{-1}. A comparison of repeated observations in the basic Anand21 database suggests that a practical estimate of the redshift error at $z_{abs} \sim 0.8$ is a little larger at $\sigma_{z_{abs}} \approx 1.7 \times 10^{-4}$, corresponding to a velocity difference of ~ 28 km s^{-1}.

When considering the finer details of the morphology of individual LSSs, any blurring will then be due to peculiar velocities, for which plausible values might be ~ 400 km s^{-1}, corresponding to $\sigma_{z_{pec}} \sim 0.0024$ at $z \sim 0.8$, or ~ 7 Mpc in proper distance for the present epoch. We therefore expect that any blurring effects should be minor.

3.4 Statistical Tests

In the following chapters I present the observational and statistical analysis of the two uLSS discoveries: the Giant Arc (GA; Chapter 4) and the Big Ring (BR; Chapter 5). The discoveries were made visually, so the statistical analysis was unavoidably post-hoc. Nevertheless, the statistical assessments provide quantitative descriptions of the uLSSs, and in the context of the rest of the field.

In this final section, I briefly introduce two of the statistical tests that are applied to both the GA and BR, they are: the Single-Linkage Hierarchical Clustering algorithm combined with the Convex Hull of Member Spheres, and the Cuzick and Edwards test. The purpose of this section is to act as a preliminary introduction to the statistical tests, so that later on the reader might find themselves more familiar with the statistical techniques being applied to the data. In any case, an overview of the statistical tests mentioned here will be reiterated in the relevant chapters, and other statistical tests that are not mentioned here (such as the Power Spectrum Analysis or FilFinder) are also introduced in their relevant chapters.

3.4.1 Minimal Spanning Tree

Minimal Spanning Tree (MST) methods are useful statistical tools for identifying candidate large-scale structures (LSSs). (In Section 1.1, I discussed the definition of 'structure' for the context of this thesis.) Our specific application of the MST-type method is the Single-Linkage Hierarchical Clustering (SLHC) and the Convex Hull of Member Spheres (CHMS) algorithms which we apply in $3D$ (RA, Dec and z) to the Mg II points (a full description of the methods can be found in Clowes et al., 2013). Note that the SLHC method should only be applied to data with no intrinsic spatial variations, which is not the case for the Mg II absorbers given the spatial variations in the background probes. However, with careful selections (e.g. avoiding obvious artefacts arising in the background quasars), and bearing in mind the caveats of the complicated data, approximations can be made and the complications need not be a restriction to investigating the LSS. The SLHC detects candidate LSSs, and the CHMS assesses the likelihood of the convex hull volume arising from random Poisson expectations.

The SLHC requires first generating a Minimal Spanning Tree (MST). A linkage scale is specified (based on the field density of points considered), and the MST is then separated at lengths exceeding the linkage scale. A minimum membership of candidate member points is specified (typically, $N > 10$), and groups below this threshold are discarded. With the remaining set of candidate structures, above some minimum membership and below some specified linkage scale, we then apply the CHMS to the candidate structures.

The CHMS draws a unique volume around the candidate structure (a group of points), and each member point is extended by a cloud of points with radius equal to half the mean MST separation for that group. The unique volume is compared with the convex hulls of randomly distributed points in a cube with an equal number of points to the candidate structure and a field density equal to that of the control field. The CHMS measures the likelihood that the candidate structure volume is drawn from a random distribution; thus, the CHMS provides a statistical assessment of the reality of a candidate structure (without this assessment, we cannot know whether the candidate structure is just a feature of patterns in noise; see Lopez & Clowes, 2025).

The CHMS, due to its method of calculation, is best applied to small, clumpy structures, where the volume is a fair representation of the clustering of points. However, for structures with strong curvature, such as arcs or rings, the CHMS method will likely overestimate the volume of the candidate structure by incorporating the empty volumes encompassed by the curved structures.

We include a second method for calculating the significance of candidate structures from Pilipenko (2007) which we find is more appropriate for dealing with curved, filamentary candidate structures. The Pilipenko method is based on the MST edge lengths cubed, so we refer to this method as MST-significance or MST-overdensity. The MST-significance is calculated by comparing the occurrence of MST edge-lengths in randomly distributed points — with an equal number of points as the candidate structure, and at a field density equal to that of the control density — that are shorter than the MST edge-lengths in the candidate structure. The Pilipenko MST calculations are best applied to filamentary structures with strong curvature, where the CHMS volume would be overestimated.

3.4.2 Cuzick-Edwards Test

The Mg II data that we are working with for LSS analysis is complicated; the absorbers rely on the availability of background probes, which themselves are subject to survey bias and intrinsic inhomogeneities. Presumably, the inhomogeneities of the quasars, on both small and large scales, contribute in some complicated way to the inhomogeneities of the Mg II absorbers. For the majority of our statistical analyses, we are dependent on making approximations for the quasars so that we can continue to make statistical assessments of the Mg II absorbers.

The Cuzick-Edwards (CE) test (Cuzick & Edwards, 1990) is a $2D$ case-control k nearest-neighbour statistical test that is designed for assessing the field clustering of spatial data with respect to inhomogeneities in the source data; the nature of our problem is exactly this.

The CE test was originally designed to assess the geographical clustering of medical diseases within a population. The population density for any given geographical region will be different, so to assess the clustering of individuals with a particular medical illness, you must first account for the 'control' size of that particular region. In this example, the individuals with the medical illness are labelled with a 'case' value and the individuals without the medical illness are labelled with a 'control' value (i.e., case-control). For our data, we can label the Mg II absorbers with a case value, and the background quasars with a control value.

The CE test output provides the p-values for the chosen k values of nearest neighbours; the k value is in essence providing a 'scale' of clustering, but as this is dependent on the ratio of cases to controls (see Section 4.3.3), it is not providing a *physical* scale of clustering (instead we use the Power Spectrum Analysis for this, see Section 4.3.4).

Chapter 4

A Giant Arc on the Sky

In this chapter I present the Giant Arc (GA): an ultra-large LSS (uLSS) at $z \sim 0.8$ found in the Mg II catalogues of Zhu & Ménard (2013). The GA spans $\sim$ 1 Gpc (proper size, present epoch), and appears to be intriguingly symmetrical on the sky. It is one of the largest of a steadily accumulating set of uLSSs that may (cautiously) challenge the Cosmological Principle (CP), upon which the 'standard model' of cosmology is founded. The original discovery of the GA was presented for the first time in my MSc thesis (Lopez, 2019), but the statistical and observational analysis of the GA was presented in Lopez et al. (2022), which can be found appended to the back of this thesis. The published work in the paper is presented in this chapter, but has been adapted for the thesis. For reference, the discovery of the GA, as in Lopez (2019), is briefly summarised below.

4.1 Discovery of the Giant Arc

During my MSc by Research at the University of Lancashire (formerly, Central Lancashire), I began testing the Mg II method for analysing cosmological LSS. I chose six known examples of high redshift structures to assess whether intervening MgII absorbers can highlight the LSS. The candidates were chosen based on various criteria as follows. (i) Redshift — Mg II can be observed in SDSS quasar spectra for redshifts $0.3 \lesssim z \lesssim 2.2$, but in particular, redshifts $z > 0.6$ were more desirable since we were primarily interested in the higher redshift regions; (ii) data availability; (iii) quality and availability of images; (iv) additional information, e.g., follow-up analysis; (v) qualitative assessment on interesting additional features that may make the structure worth investigating. The results were: in five of the six target fields, there were detections of dense regions of Mg II absorbers which were believed to be in association with the target structures, while the sixth target remained less clear due to low completeness.

One of the test fields investigated that fit the criteria above was a small field containing a Sunyaev-Zeldovich (SZ) cluster candidate PSZ2 G069.39+68.05 (Burenin et al., 2018, subsequently B18 when referring to this specific SZ candidate) at $z = 0.763$. The SZ cluster candidates are identified through the SZ effect (hence their name). The SZ effect is a result of inverse Compton scattering, whereby the hot electrons of the intergalactic gas around galactic clusters transfer their energy to background cosmic microwave background (CMB) photons, resulting in a detectable temperature increase against the background CMB (Sunyaev & Zeldovich, 1972). The SZ cluster candidate we investigated (B18) was one of the detected sources from the second Planck catalogue of Sunyaev-Zeldovich sources (Planck Collaboration et al., 2016a, PSZ2); Burenin et al. (2018) then provided optical identifications and spectroscopic redshifts of the cluster counterparts for some of the SZ sources.

For the analysis of B18 (and the 5 other targets selected for analysis in Lopez, 2019) we plotted the Mg II present in the target field. Curiously, for B18, the Mg II maps indicated a very large LSS extending on the sky, as a long, dense, thin band of Mg II absorbers, roughly symmetrically to both sides of the cluster. In Lopez (2019) the central redshift and the redshift interval of the GA were estimated by stepping through thin redshift slices and visually inspecting the density and connectivity of the Mg II absorbers. We have since refined a little the estimates of the central redshift and the redshift interval — see Section 4.3.1 for the details. In Figure 3.13 the GA can be seen stretching ~ 1 Gpc (proper distance, present epoch) horizontally across the centre of the field. Visually, the GA appears densely concentrated and with the distinctive shape of a giant arc.

4.2 Observational properties of the Giant Arc

We investigate the observational properties of the GA, in a visual manner, including: rest-frame equivalent width (EW) distribution ($W_{r,2796}$); signal-to-noise ratio (S/N) of the $\lambda 2796$ line; S/N of the continuum of the spectra; the i magnitude (i) of the probes (background quasars); and the redshift distribution of the $\lambda 2796$ line (z_{2796}). As mentioned in the earlier chapters of this thesis, it is thought that the EW distribution, or other line-profile properties, could be related to the host-halo galactic properties (morphology, luminosity, impact parameter, galaxy inclination etc.), but these aspects are still not fully understood. While it may not yet be clear what the EW distribution within the GA indicates, future studies of the Mg II-absorber / host-halo connection could lead to an understanding of the origins of the GA and its environment.

The values of the EW are often classed as 'strong' or 'weak', although there seems to be no agreement on what defines 'strong' and 'weak'. For example, in the literature one might find strong EW variously defined as $W_{r,2796} \geq 0.3\,\text{Å}$, $W_{r,2796} \geq 0.6\,\text{Å}$, and $W_{r,2796} \geq 1.0\,\text{Å}$ — see for example Churchill et al. (2005), Dutta et al. (2017), Evans et al. (2013) and Williger et al. (2002). We shall follow the example set by Zhu & Ménard (2013) and use the definitions of strong and weak EWs as $W_{r,2796} \geq 0.6\,\text{Å}$ and $W_{r,2796} < 0.6\,\text{Å}$ respectively.

We divide the $W_{r,2796}$ EWs into four bins with boundaries at $(0.0, 0.3, 0.6, 1.0,$

10.0) Å. (The boundaries were chosen to reflect the above diversity of what corresponds to 'strong' in the literature.) The on-sky spatial coordinates of the absorbers in the GA and its immediate field are then plotted, with colour-coding according to the four EW bins — see Figure 4.1. The shade of the blue dots in the figure represents the EW bin, with the lightest shade representing the first bin $0.0 < W_{r,2796} \leq 0.3$ Å, and the darkest shade representing the last bin $1.0 < W_{r,2796} \leq 10.0$ Å.

Similarly, with the same set of four blue shades, we show the i magnitude of the probes, the S/N of the $\lambda 2796$ line, and the S/N of the continuum (Figure 4.1). The boundaries of the bins are as follows: (i) i magnitude of the background quasars — $16.0, 17.8, 18.7, 19.6, 21.0$; (ii) S/N of the $\lambda 2796$ line — $0, 3, 6, 12, 37$; and (iii) S/N of the continuum — $0, 8, 16, 24, 47$. The colour-coding again represents the smallest values by the lightest shade of blue, and the largest values by the darkest shade. Note, that for i, the lightest shade thus represents the brightest probes.

The EW of the $\lambda 2796$ line should correlate with the S/N of the $\lambda 2796$ line; the top-left and bottom-left panel in Figure 4.1 show that this is indeed the case. The brightness (i) of the background quasar should correlate with the S/N of the quasar continuum; the top-right and bottom-right panel in Figure 4.1 show that this is indeed the case. Note that brighter quasars, having a higher continuum S/N, can detect absorbers to a lower threshold EW.

An asymmetry is apparent in the distribution of Mg II EWs within the GA: the EWs tend to be stronger on the LHS (lower RA), and in the centre of the GA, than on the RHS (higher RA). Conversely, there is a tendency for the RHS to have brighter probes and higher continuum S/N, so, on the RHS, the threshold EW for detection should tend to be lower and certainly able to detect stronger absorbers, should they be present. These two observations can be confirmed by simply counting the frequency of the two darkest-coloured points (or of the two lightest-coloured points) for points belonging to the visually-identified GA (points

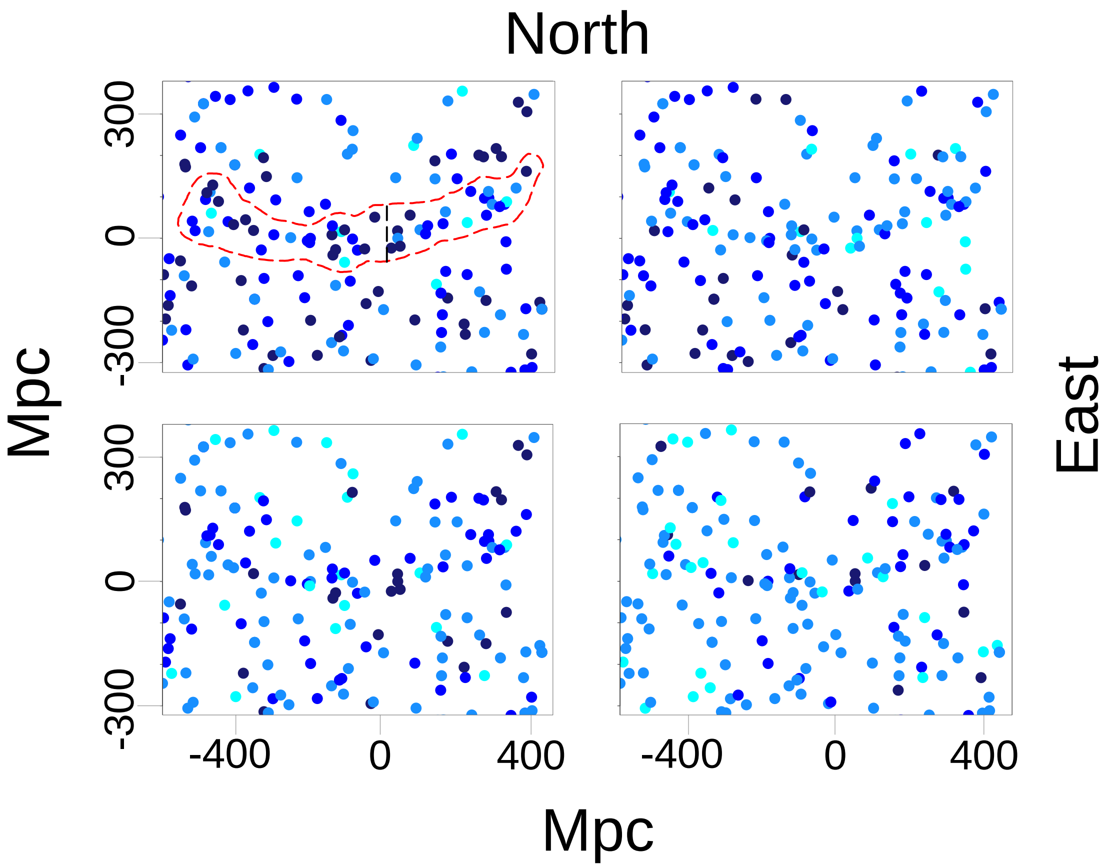


Figure 4.1: The on-sky spatial coordinates of the quasars containing Mg II absorbers in the GA and its immediate field, with colour-coding according to: (top-left) the equivalent width (EW) of the λ2796 line; (top-right) the i magnitude of the background quasar; (bottom-left) the signal-to-noise (S/N) ratio of the λ2796 line; (bottom-right) and the S/N ratio of the quasar continuum. See the main text for the bin divisions. The red, dashed line in the top-left panel indicates what we visually identified as the GA. The black, dashed line in the top-left panel divides the visually-identified GA into a left- and right-hand side.

inside the red, dashed line in the top-left panel of Figure 4.1), and comparing the results (normalised to the number of absorbers) either side of the $x = 0$ vertical line (see the black, dashed line in the top-left panel of Figure 4.1). To demonstrate, there are 52 absorbers in the visually-identified GA; 30 on the LHS and 22 on the RHS. For the EWs in the GA on the LHS, there are 24 out of 30 absorbers (80%) which are of the two darkest blue shades corresponding to strong EWs, whereas on the RHS there are 14 out of 22 absorbers (64%) which are of the darkest blue shades. Similarly, for the i magnitudes of the corresponding background quasars (top-right panel in Figure 4.1), there are 16 out of 30 absorbers (53%) in the LHS of the visually-identified GA which are of the two lightest blue shades corresponding to the brightest quasars, whereas on the RHS there are 14 out of 22 absorbers (64%) which are of the two lightest blue shades. The collected observations are consistent with the reality of the observed asymmetry of stronger absorbers on the LHS.

There have been many attempts to understand the relationship between Mg II EWs (and other line properties) and the host-halo galactic properties (morphology, luminosity, impact parameter, galaxy inclination etc.), but currently there is no clear consensus. Conceivably, the asymmetry in the EW distribution could arise from the details of the geometry of the GA and the orientations or morphologies of the galaxies within it. Future sky surveys and targeted observations seem likely to be necessary for progress on these details.

We note that there appears to be a preference for the strongest ($W_{r,2796}$) Mg II absorbers in the GA to clump together into groups of a few. See the dark blue points in the top-left panel of Figure 4.1, and note in particular those on the LHS of the GA (lower RA), the centre of the GA, and the group just above the tip of the RHS (higher RA) of the GA. As the GA is denser than the rest of the field, we can speculate that the occurrence of the strongest EWs in proximity is not accidental but is connected with the origin and environment of the GA.

Recall that the SZ cluster B18, $z = 0.763$ (Burenin et al., 2018), is at the centre of the GA. (It is what led to the discovery of the GA.) At the centre of the GA is a small, circular 'hole', and surrounding this hole is a group of the stronger absorbers. SZ clusters create a highly-ionised environment, but Mg II absorption occurs in low-ionised regions. Possibly, a region of high ionisation can account for the hole, but an origin, in environment, of the enveloping group of stronger absorbers is not then obvious.

The investigation of small ($\Delta z = 0.030$), overlapping (by 50 per cent) redshift slices reveals a noticeable difference between the left- and right-hand sides of the GA. For example, the LHS (lower RA) of the GA appears concentrated in the small redshift slice located farthest away ($z = 0.832 \pm 0.030$), whereas the RHS (higher RA) of the GA appears spread diffusely through the larger redshift slice. Interestingly, the LHS of the GA has both a narrower redshift distribution and a preference for stronger Mg II EWs.

Finally, the investigation of the redshift distribution suggests that, if the GA is represented as a segment of a cylindrical shell, then the LHS would be tilted away along the line of sight. That is, if the GA can indeed be represented as a segment of a cylindrical shell, then it is not precisely orthogonal to the line-of-sight but is rotated with respect to a north-south axis.

4.3 Connectivity and statistical properties

The GA was discovered visually, from a Mg II density image (e.g. Figure 3.13). Albeit after the event, we now discuss its connectivity and statistical properties. The Mg II absorbers can, of course, be found only where there are background quasars to act as probes, and those probes may themselves be subject to spatial variations arising from large-scale structure and, in particular, from artefacts in the surveys.

We apply three different statistical methods for assessing the GA, as follows.

(i) SLHC / CHMS — see Clowes et al. (2012). This method depends first on constructing the 3D Minimal Spanning Tree (MST), and then separating it at some specified linkage scale. At this stage it is equivalent to the Single-Linkage Hierarchical Clustering (SLHC) algorithm. The statistical significance of a candidate structure is then assessed using its volume obtained as the volume of the 'Convex Hull of Member Spheres' (CHMS). Note the important feature that this method assesses the significance of *individual* candidate structures.

(ii) The Cuzick-Edwards (CE) test — see Cuzick & Edwards (1990). It is a 2D 'case-control' method that is designed to correct the incidence of cases for spatial variations in the controls (the underlying population). It depends on the number of cases that occur within the k nearest neighbours. The CE test can detect the presence of clustering in the field, while correcting for variations in the background, and can assess its statistical significance. It cannot, however, assess the physical scale of the clustering.

(iii) 2D Power Spectrum Analysis (2D PSA) — see Webster (1976a). It is a powerful Fourier method for detecting clustering in the field. It can be effective even for detecting weak clustering. The 2D PSA can detect the scale of clustering and assess the statistical significance of the clustering at that scale.

Each of these tests has distinctive attributes, and the reader should judge the evidence provided by the ensemble. Only the SLHC / CHMS method assesses the significance of individual candidate structures, whereas the CE test and the 2D PSA address the clustering in the field. We shall describe below the 'polygon approach', in which we assess the contribution that the GA makes to the results from the CE test and the 2D PSA for the field. Only the CE test can correct for spatial variations in the underlying population. However, we shall describe below, again using the polygon approach, that the 2D PSA has more power to discriminate than the CE test.

Finally, we emphasise again, that given the nature of the discovery the statistical analysis is necessarily performed *post-hoc*. The reader will find that we have used techniques to compare the field containing the GA with other, unrelated fields (within the same Mg II dataset). This of course has its limitations due to the non-uniformity of the background quasars (probes) and potential survey artefacts. We have also compared with randomised simulations, in which we attempt to preserve these subtleties of the Mg II data.

4.3.1 SLHC / CHMS (Minimal Spanning Tree)

The Minimal Spanning Tree (MST) is a widely-used algorithm for assessing large-scale structure in astronomy and cosmology. When the MST is separated at some specified linkage scale it is equivalent to the algorithm for Single-Linkage Hierarchical Clustering (SLHC). An approach to assessing the statistical significance of the agglomerations found in this way was introduced by Clowes et al. (2012): the Convex Hull of Member Spheres (CHMS) method. It was further used by Clowes et al. (2013) in the analysis of the Huge-LQG, the Huge Large Quasar Group that they discovered.

Here, we apply the sequence of SLHC and CHMS to the Mg II absorbers in the GA field. By specifying a linkage scale and a minimum membership, the SLHC identifies the 3D agglomerations or groups within the coordinates of the absorbers. Within each identified group the CHMS constructs a sphere around each member point with a radius of half the mean linkage separation for that group. A volume for the group is then computed as the volume of the convex hull of its member spheres (and note that the convex hull is a unique construction). An expected density of absorbers is determined from a control field and the observed redshift interval of a group. (Here, the control field is specified as the same field that is being assessed.)

The observed number of member points within a group are then scattered randomly within a cube at the expected density, and their CHMS volume is calculated; this is done 1000 times. The significance of the group is calculated by the rate of occurrence of randomly-generated CHMS volumes that are smaller than the observed volume. See Clowes et al. (2012) for full details of the CHMS method.

In principle, this SLHC / MST approach should be applied only to surveys that have no intrinsic spatial variations. The background quasars — the probes of the Mg II absorbers — are drawn from a merger of the SDSS DR7QSO and DR12Q databases. While a reasonably spatially-uniform subset can be extracted from DR7QSO, DR12Q is much more strongly affected by spatial artefacts arising from deeper areas. Thus the distribution of the background quasars can conceivably affect the distribution of the Mg II absorbers in some, possibly complicated, way. However, if the distribution of the background quasars appears to be reasonably homogeneous in the area of interest, then we can assume that the distribution of Mg II absorbers is predominantly a product of the LSS and not the availability of background quasars. Of course, the distribution of background quasars can still have some effects — such as occasional gaps in connectivity — on the Mg II absorbers even in such reasonably homogeneous regions.

Figure 4.2 shows the kernel-smoothed distribution ($\sigma = 11$ Mpc, present epoch) of the background probes (quasars) in the area of the GA for $z > 0.862$. It is clear that there are denser areas, less dense areas, and even empty patches, across the whole image, indicating the spatial non-uniformity of the background probes. There is a particularly dense band in approximately the northern third, which arises from a deeper area of the DR12Q survey. However, there are evidently no artefacts that correspond to the dimensions and orientation of the GA.

We are now taking the GA to be predominantly concentrated in the redshift interval 0.802 ± 0.060, so $0.742 \rightarrow 0.862$. Its (present-epoch) depth is then ~ 340 Mpc.

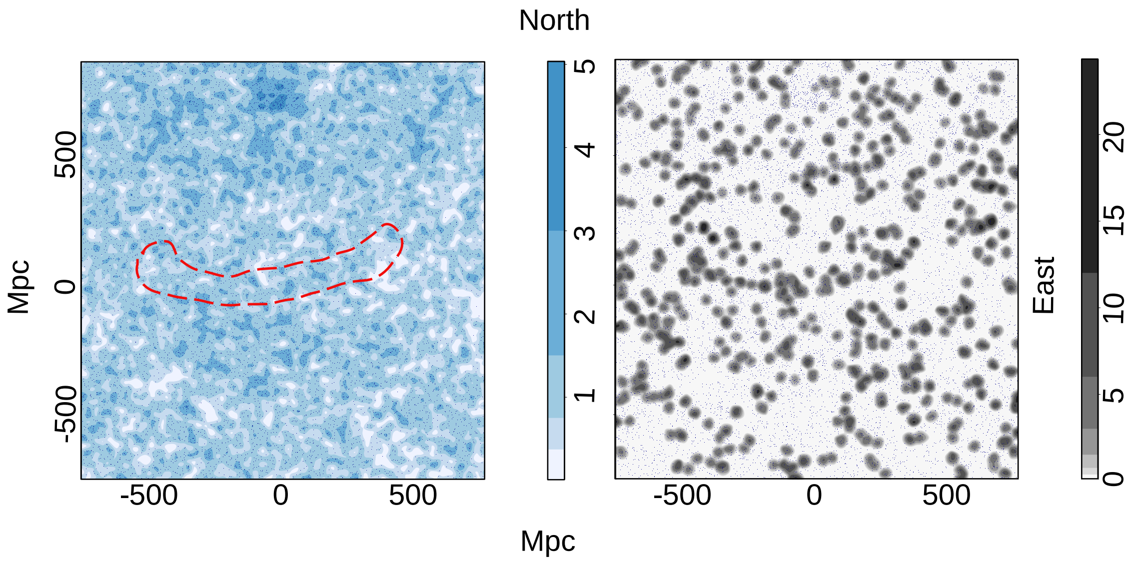


Figure 4.2: Left panel: the tangent-plane distribution of background probes (quasars) in the GA field with redshifts $z > 0.862$, represented by the dark blue points. The blue contours, increasing by a factor of two, represent the density distribution of the quasars which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc. It is clear that there are denser areas, less dense areas, and even empty patches, across the whole image, indicating the spatial non-uniformity of the background probes. There is a particularly dense band in approximately the northern third. However, there are evidently no severe artefacts that correspond to the dimensions and orientation of the GA (inside the red, dashed line). Right panel: the corresponding tangent-plane distribution of Mg II absorbers in the redshift interval $z = 0.802 \pm 0.060$. The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes. The small dots represent the background probes (quasars). The Giant Arc can be seen running west-east in the centre of this figure, spanning ~ 1 Gpc, and corresponds to the added red, dashed line in the left panel. In both figures the axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top.

This redshift interval appears to be the optimum, following a heuristic process of stepping through a range of redshift intervals and determining the membership and significance of the GA through the SLHC / CHMS method. Redshift intervals of thickness $\Delta z = \pm 0.050$ between $0.760 < z < 0.810$ were tested, using various linkage scales, and a clear peak signal for the GA was seen around the redshift $z = 0.810$ using a linkage scale of 95 Mpc. (See Section 4.3.2 for the details of choosing the optimum linkage scale.) Further, finer-scale, testing revealed the greatest number of connected GA members to be more precisely located at $z = 0.802$, again for the linkage scale of 95 Mpc. The significance of the GA (via the SLHC / CHMS algorithms) in the redshift slice $z = 0.802 \pm 0.050$ is 4.15σ. Widening the redshift interval to $z = 0.802 \pm 0.060$ gives a slightly higher significance of 4.30σ, and the number of member Mg II absorbers is then increased from 42 to 44.

A second, smaller, agglomeration made up of 11 (10) absorber members at $z = 0.802 \pm 0.060$ ($z = 0.802 \pm 0.050$), although not formally significant, i.e., 2.04σ (1.75σ), is clearly also part of what we identified visually as the Giant Arc — see the red points in Figure 4.4 (note that the points in parentheses are not shown). Conceivably, just one further background probe would be sufficient to yield one further absorber that would then connect both agglomerations as one significant unit. We have emphasised the limitations of the SLHC / CHMS method for this dataset, and we might here be seeing their consequences.

As noted above, the estimation of the CHMS significance requires a control field from which the expected average density is calculated. For the CHMS significances given above, we used the same field as that containing the GA. This was a deliberate choice, given the spatial variations of the wider survey. Clearly, the GA must then represent a small fraction of the total area and number of absorbers (~ 7 per cent of the absorbers are from the GA).

Even so, the distribution of background probes (quasars) in the field of the GA is not uniform — notably the denser band in the northern third. The non-uniformities could affect the CHMS calculations of significance, either by overestimating or underestimating, depending on whether the probes are generally under-populated or over-populated in the control field. A second estimate of the significance can be calculated from the CHMS method by increasing the field-of-view (FOV) containing the GA, and using it as a new control field. Figure 4.3 shows the background probes in the field containing the GA with the western and southern boundaries extended. Note that the eastern and northern boundaries were not extended because of proximity to the edge of the survey area.

Using the *larger* FOV in the redshift interval $z = 0.802 \pm 0.060$, the CHMS method calculates a significance of 4.53σ for the principal agglomeration of the GA. As noted previously, the GA is split into two agglomerations by the SLHC algorithm, shown in Figure 4.4. For this entire, larger, FOV, there are 35 agglomerations in total, with the principal agglomeration of the GA being the largest and most significant, and the only agglomeration with a significance $> 3.5\sigma$. Figure 4.4 shows the GA as identified by the SLHC / CHMS method, with the principal agglomeration represented by the black points; the red points indicate the smaller, separate agglomeration of much lower significance, but visually it can clearly be seen as part of the GA.

The SLHC / CHMS algorithm is next applied to three lower redshift slices — $z = 0.682 \pm 0.060$, $z = 0.562 \pm 0.060$, and $z = 0.442 \pm 0.060$ — in the same (larger) FOV as the GA (see Figure 4.3). Using lower redshift slices and the same FOV means that we can conveniently compare the Mg II absorbers arising from the same probes as those in the GA field by restricting the probes to $z > 0.862$. We can also apply the SLHC / CHMS method to the same three lower redshift slices *without* this redshift restriction on the probes. Recall, however, that the probes may show density artefacts with a large FOV. The SLHC / CHMS method can be problematic

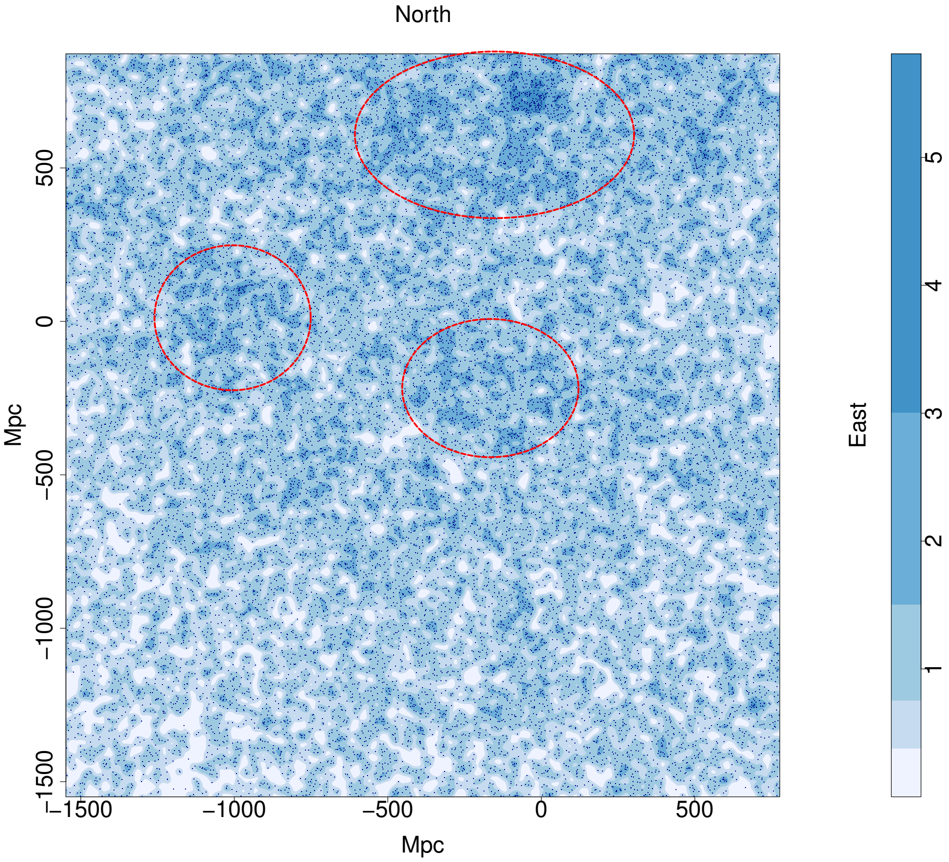


Figure 4.3: The tangent-plane distribution of background probes (quasars), in the same region of sky as the GA for redshifts $z > 0.862$, represented by the dark blue points. The field-of-view has here been increased by extending the western and southern boundaries. The blue contours, increasing by a factor of two, represent the density distribution of the quasars which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc. Red, dashed ellipses have been added to indicate the approximate areas of particularly overdense regions of background probes (quasars), which is based on having high density contours (greater than level 3) and spanning more than $\sim 100^2$ Mpc^2 The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top.

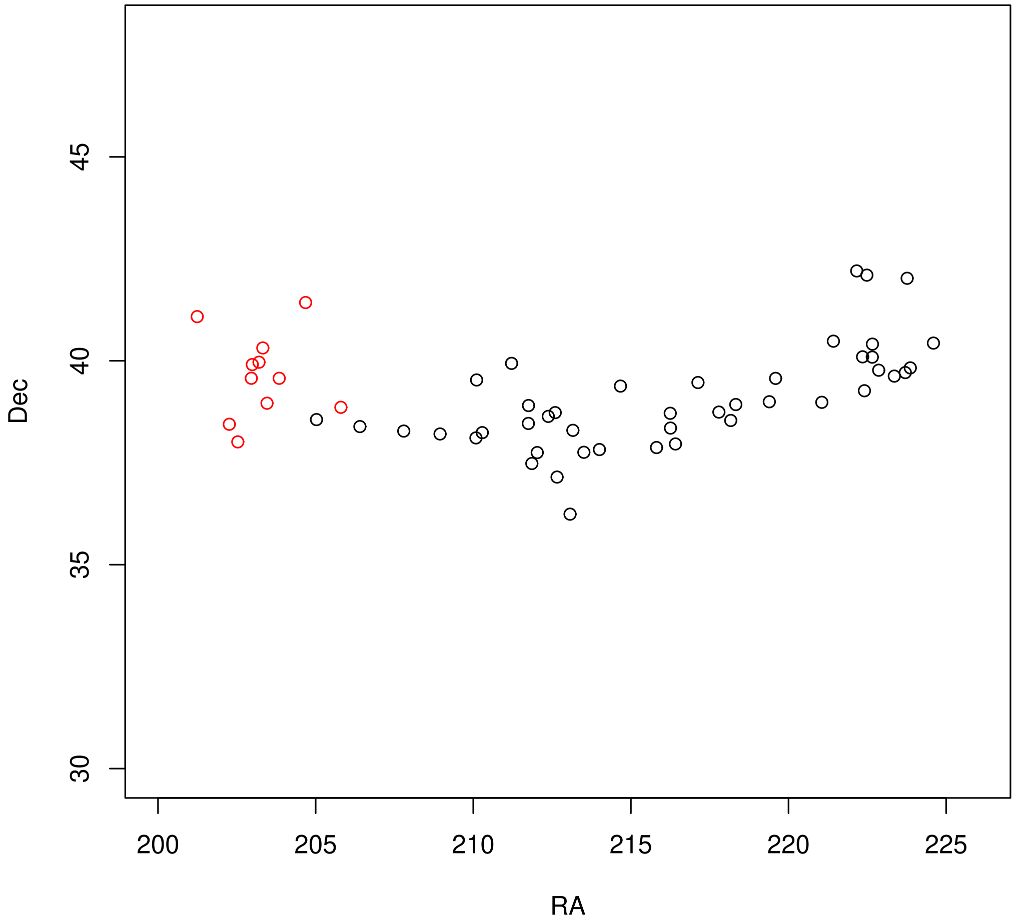


Figure 4.4: The GA connected via the SLHC algorithm. There are 44 black points which indicate the principal agglomeration of the Giant Arc — the largest, and most significant agglomeration in the field —, which comprises the majority of the GA. There are 11 red points which indicate the secondary agglomeration of the GA, which, although by itself not statistically significant, is clearly part of the GA. The axes are labelled RA and Dec where east is towards the right and north is towards the top.

for the Mg II analysis because of these artefacts. Therefore, superficially significant structures that correspond to particularly dense patches of probes[1] are likely to be discarded. The results are summarised in Table 4.1.

As can be seen in Table 4.1 there are significant ($\sim 3\sigma$) structures. At a more cautious limit of 4σ, however, there were only two candidate structures which did not reside in an artefact (in redshift slices $z = 0.682$ and $z = 0.562$). We shall in due course investigate them further, starting with optimisation of the redshift intervals.

Finally, we introduce a random-simulation aspect to the SLHC / CHMS analysis. We have carried out 1000 random simulations as follows. (i) We consider the large, extended area that corresponds to Figure 4.3. (ii) We consider only the probes at higher redshift than the redshift slice of the GA — that is, we continue (as with the slices of redshift lower than that of the GA) to use only the probes appropriate to the GA, so that density artefacts in the probes remain identical. (iii) We reassign at random Mg II absorbers of any redshift to the probes, while not splitting occurrences of multiple absorbers per line of sight. (Note that splitting absorbers would have the undesirable effect of changing the total number of probes with absorbers.) (iv) We then analyse the random-simulated data as for the actual GA slice, selecting absorber redshifts for the redshift slice.

Within the simulations, we looked for "structures" that had properties comparable to, or more extreme than, the observed properties of the GA (precisely, of GA-main — see below). The properties considered were the set of: number of absorber members; SLHC / CHMS significance; and overdensity. In all cases (roughly one occurrence per two simulations), we found that these "comparable structures" were in the regions of the visually-obvious density artefacts (i.e., the red, dashed

---

[1] The red, dashed ellipses in Figure 4.3 outline the regions of particularly dense patches of background probes (quasars). They were selected based on having high density contours (exceeding level 3 on the contour grade-scale on the right-hand side of the figure) and covering an area larger than $\sim 100^2$ Mpc^2

Table 4.1: Results from applying the SLHC / CHMS algorithm to three lower redshift slices in the same (larger) FOV as the GA (see Figure 4.3). The columns are: the redshift interval of the Mg II absorbers (z); the redshift of the background probes (quasars; z_{probes}); the number of structures found by the SLHC algorithm (N); the number of structures exceeding 3σ significance ($N_{3.0\sigma}$); the number of structures exceeding 4σ significance ($N_{4.0\sigma}$); the number of structures exceeding 3σ significance after removing those that reside in artefacts of the probes (N_{real}). Note that the three lower redshift slices have two sets of probes with different redshifts: one for the probes corresponding to those in the GA field, and one for the probes corresponding to the redshift slice.

z	z_{probes}	N	$N_{3.0\sigma}$	$N_{4.0\sigma}$	N_{real}
$z = 0.802 \pm 0.060$	$z > 0.862$	35	4	1 (GA)	2 (incl. GA)
$z = 0.682 \pm 0.060$	$z > 0.862$	31	3	1	3
	$z > 0.742$	34	6	2	4
$z = 0.562 \pm 0.060$	$z > 0.862$	17	6	3	2
	$z > 0.622$	19	7	4	3
$z = 0.442 \pm 0.060$	$z > 0.862$	9	1	0	0
	$z > 0.502$	12	2	0	0

Notes.
(1) For each redshift slice the overall density of absorbers in the field varies, so we modify the linkage scale according to the relation $s = (\rho_0/\rho)^{1/3} \times 95\,\mathrm{Mpc}$, where s is the linkage scale, ρ_0 is the density for the GA field and ρ is the density for the new field concerned.
(2) The number of structures reduces in each successively lower redshift slice, because (i) the number of Mg II absorbers reduces as the lower-wavelength limit of detectability is approached (~ 4000Å), and (ii) the size of objects in the field appear larger at smaller redshifts.

ellipses in Figure 4.3), and never in the region occupied by the real GA. The occurrence of the "comparable structures" in the density artefacts is as expected: for those artefacts, the linkage scale and the control density would clearly not be appropriate. We can infer that the probability of the real GA (precisely, GA-main) occurring as a random event is < 0.001.

The SLHC algorithm easily identifies the GA, with 44 connected Mg II absorbers, and the CHMS method estimates a significance of $\sim 4.5\sigma$ using the central redshift $z = 0.802$. In every redshift interval investigated, the GA appears in two parts: (i) the principal agglomeration, which is large in both physical size and membership, and statistically very significant; and (ii) the secondary agglomeration, small in size and membership, and by itself statistically not significant. As mentioned earlier, the Mg II absorbers depend on the availability of background probes (quasars), and without those, Mg II would not be detected. Thus an artefact in the distribution of probes — i.e. a gap, perhaps of just one missing probe — could lead to an artefact of apparent splitting into two agglomerations.

We have seen previously, in Section 4.2, that there is a noticeable difference between the LHS and RHS of the GA with regards to redshift distribution. Investigating small ($\Delta z = 0.030$), overlapping (by 50 per cent) redshift slices has highlighted the sub-structure of the GA along the redshift axis. We find that the larger agglomeration of the GA is distributed more evenly and widely along the redshift axis, while the smaller agglomeration is concentrated in a narrower redshift slice. It becomes clear from the central redshift slice and below ($z < 0.802$) that there are no Mg II absorbers available that can connect the small agglomeration to the large agglomeration.

4.3.2 Selecting a Linkage Scale

The linkage scale that is set in the SLHC / CHMS method determines both the number of agglomerations and their memberships. It was set at 95 Mpc for the GA. This setting was partly guided by the linkage scale that was known to be effective for the Clowes-Campusano LQG (Clowes & Campusano, 1991, hereafter CCLQG), and which, when used subsequently, led to the discovery of the Huge-LQG (Clowes et al., 2012, 2013). Clearly, the linkage scale must be adjusted for field density. In this case, starting from the linkage scale that was effective for LQGs (i.e. for LSS in quasars), and scaling with density, we calculate a linkage scale of 85 Mpc for Mg II absorbers in the GA field. From here we followed the heuristic process described in Section 4.3.1 to identify an optimum linkage scale of 95 Mpc for the Mg II absorbers.

One must remember that Mg II absorbers are distinctly different from quasars and therefore cannot be treated in quite the same way. For example, the linkage scale that works for quasars will not necessarily work for the Mg II absorbers since the latter is a case of inhomogeneities (the absorbers) superimposed on inhomogeneities (the quasars and survey artefacts). Future work will address the development of a clustering analysis that is specifically addressed to the requirements of Mg II absorbers.

It is in the nature of discoveries that there will be a *post-hoc* aspect to the analysis. What turned out to be effective for the CCLQG led to the discovery of the Huge-LQG: an initial discovery, followed by a heuristic process, followed by an entirely objective *a-priori* new discovery. It is in this spirit that we present the discovery of the GA: one for which the techniques and parameters used to assess and characterise it can subsequently be applied to the whole Mg II catalogue.

For completeness, we can briefly mention what results from instead setting different linkage scales for the adopted redshift slice. In five runs, using linkage scales of 85 Mpc, 90 Mpc, 95 Mpc, 100 Mpc, and 105 Mpc, there are totals of 4, 25, 35, 43, and 3 agglomerations found respectively, with the GA always being the most significant in all except the 105 Mpc run. The middle three runs split the GA into two parts: one large and significant, and one smaller and less significant. For the following comments, we concentrate only on the large, significant agglomeration of the GA identified in the middle three runs — (90, 95 and 100) Mpc — which makes up the majority of what we visually identified as the GA. (1) It has a significance greater than 3.8σ in all of the three runs. (2) It is the only agglomeration that has a significance greater than 3.5σ in each run, with only up to a maximum of three additional agglomerations above 3σ (all others being below a 3σ threshold). (3) In both the 90 Mpc and 95 Mpc runs, it is the largest agglomeration by membership, and is the second largest by membership in the 100 Mpc run. Lastly, we mention what arises from setting the linkage scale to 85 Mpc and 105 Mpc. With these linkage scales we begin to approach the lower and upper limits that the SLHC / MST method can perform with (productively). There are only 4 and 3 candidate structures identified by the SLHC algorithm with the linkage scales set to 85 Mpc and 105 Mpc, respectively. Example candidates for each linkage scale, respectively, have memberships of 20 and 133 with significances of 3.6σ and 0.8σ, respectively. This demonstrates that increasing or decreasing the linkage scale much further would yield nothing of consequence: either no structures, or one large structure containing almost everything. It is worth noting that, although the 85 Mpc linkage scale is the minimum at which structures can be found in the Mg II data at this redshift / density, the GA is still mostly detected, still the largest agglomeration in the field, and the only structure detected with a significance over 3σ — see Figure 4.5

Note that for future work, involving the remainder of the Mg II database, we can adopt this scale of 95 Mpc as a standard, with scaling according to the density of absorbers (i.e., $s = (\rho_0/\rho)^{1/3}s_0$) in the volumes of interest.

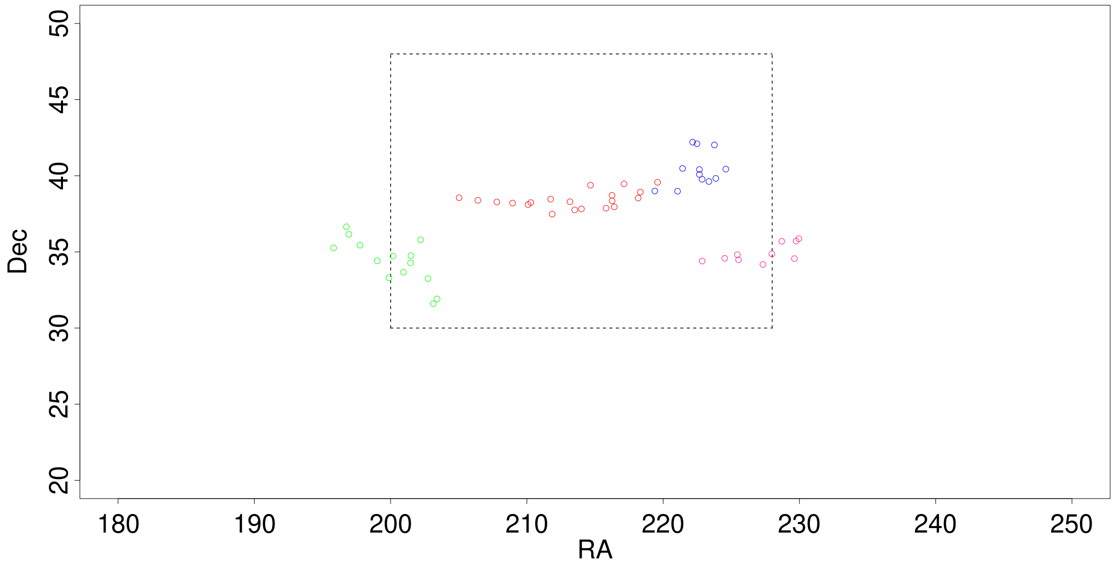


Figure 4.5: The four structures found via the SLHC / CHMS method using a reduced linkage scale of 85 Mpc. The black dashed rectangle corresponds to Figure 4.4 for comparison. The GA is still mostly detected in what appears to be two parts. The largest agglomeration, containing 20 Mg II absorber members represented by the red points in the figure, makes up the majority of what we visually identify as the GA. The second largest agglomeration, containing 14 Mg II absorber members represented by the blue points in the figure, is also part of what we visually identified as the GA and can be seen as an extension of the red points. The axes are labelled RA and Dec where east is towards the right and north is towards the top.

4.3.3 Cuzick-Edwards test

Strictly, the SLHC / MST approach should be applied only to surveys that have no intrinsic spatial variations. The Cuzick-Edwards test (Cuzick & Edwards, 1990) is a statistical test for two-dimensional spatial data that was designed to manage spatial variations in the source data; we apply it here.

The Cuzick-Edwards test (hereafter CE test) has been used mainly in medical research to assess the clustering patterns of diseases within unevenly populated geographical regions. The essential character of our problem is the same. It adopts a 'case-control' approach to a k nearest-neighbour (NN) analysis. Several papers have compared the properties of the CE test amongst various spatial clustering analyses and assert that the CE test is powerful and sensitive in estimating clustering significance within a point dataset — see, for example, Song & Kulldorff (2003), Meliker et al. (2009) and Hinrichsen et al. (2009). In Song & Kulldorff (2003) the authors note that the CE test is used more appropriately if the level of clustering is known beforehand. Inevitably, for our problem, the statistical properties of the GA are tested after the event of discovery (i.e. the level of clustering is known).

We used the CE test that is coded in the application *qnn.test* in the R package SMACPOD (French, 2020). The probes (i.e. the background quasars) are labelled as 'controls' and the Mg II absorbers in the redshift interval are labelled as 'cases'. The *qnn.test* then uses a NN algorithm to find the q (or k) NNs of any case to another case.

For a control-case dataset containing i points, z_i is the location of each point. Therefore, the subscript i is the identifier of any particular point (regardless of whether it is a case or control) and the superscript k refers to the k nearest neighbours. The test statistic is then calculated

$$T_k = \sum_{i=1}^{n} \delta_i d_i^k$$

where:

$$\delta_i = \begin{cases} 1 \text{ if } z_i \text{ is a case} \\ 0 \text{ if } z_i \text{ is a control,} \end{cases}$$

$$d_i^k = \begin{cases} 1 \text{ if the NN to } z_i \text{ is a case} \\ 0 \text{ if the NN to } z_i \text{ is a control.} \end{cases}$$

The p-value from *qnn.test* is calculated from simulations under the random-labelling hypothesis (French, 2020) for nsim = 2000 simulations.

The choice of maximum q (k) value that is adopted for the test will depend on the control-case ratio, as can be seen from the test statistic calculation. There are $\sim$ 20 times as many probes (controls) as Mg II absorbers (cases) in the redshift interval of the GA. Cuzick & Edwards (1990) examine the power of the CE test with varying control-case ratios and find that a control-case ratio of between 4 and 6 is optimum (see their Figure 5). Therefore, we choose to use a control-case ratio of $\sim$ 5:1. To achieve this we randomly select 25 per cent of the probes, for each of 100 runs of *qnn.test*. (Randomly-selected controls that duplicate the coordinates of the cases in a given run are removed.) The 100 runs also allow us to assess how robust are the estimates of significance for the Mg II absorbers.

We use a set of q (k) values: 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, and 64. We start by applying the *qnn.test* to the basic GA field. Then, to assess the (presumed) dominance of the GA itself we apply the test to a succession of smaller fields (smaller in the north-south direction), all centred on the GA. In Figure 4.6 the median p-value over 100 runs of 2000 random simulations is shown plotted against the chosen q values, with the corresponding flat-fielded Mg II images shown in the top panels.

The process of zooming into the GA allows the GA to become the dominating

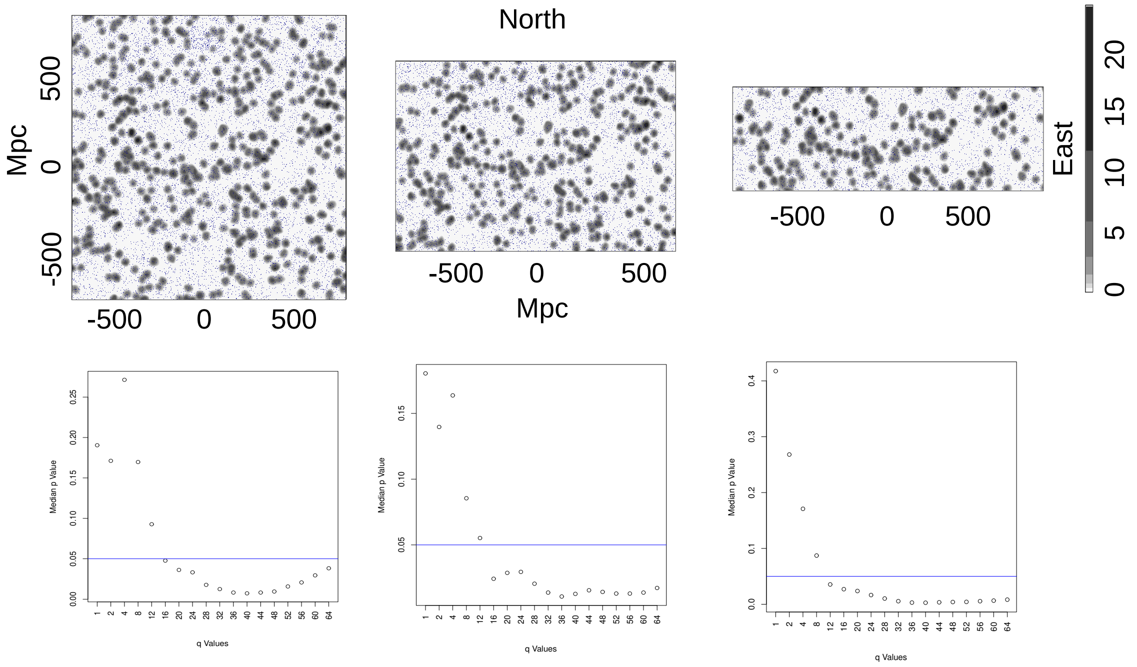


Figure 4.6: Top panels: density distribution of the flat-fielded Mg II absorbers represented by the grey contours (increasing by a factor of two) which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flatfielded with respect to the background probes (quasars). Small dots represent the background probes (quasars). From left to right, the GA field of view has been reduced in the north-south direction to 'zoom' in on the GA. (Left) GA field as usual. (Middle) First 'zoom' of the GA field. (Right) Second 'zoom' of the GA field.
Bottom panels: the median p-values over 100 runs of 2000 simulations as a function of chosen q (k) value in the GA field corresponding to the Mg II images in the top panels. The values along the x axis are: 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, and 64. The blue, horizontal line is set to $p = 0.05$. (Left) The p-value is at a minimum of 0.0072 when q is 40. (Middle) First 'zoom' on the GA field. The p-value is at a minimum of 0.0107 when q is 36. (Right) Second 'zoom' on the GA field. The p-value is at a minimum of 0.0027 when q is 40.

feature in the field, which, as a result, increases the significance of clustering (i.e. smaller p-value). In the first (standard) Mg II field (top-left panel of Figure 4.6), the corresponding CE result (bottom-left panel of Figure 4.6) has a minimum p-value of 0.0072 at $q = 40$, which (assuming a normal distribution) is equivalent to a significance of $\sim 2.68\sigma$. Whereas in the third Mg II field (second zoom; top-right panel of Figure 4.6), the corresponding CE result (bottom-right panel of Figure 4.6) has a minimum p-value of 0.0027 at a $q = 40$, which is equivalent to a significance of 3.00σ. In this way we can judge that the GA is the dominant, contributing factor to the significant level of clustering in the field.

The heuristic process of 'zooming' into the GA was next applied to three other fields at lower redshift slices (z: 0.682, 0.562, 0.442) centred on the sky coordinates of the GA. The background probes are kept the same in the three new fields as those in the GA field, allowing a direct comparison of clustering in just the Mg II absorbers (as in Section 4.3.1). Figure 4.7 shows the results of the CE test for the three lower redshift fields, paired next to the corresponding Mg II field as before, using the smallest field size (i.e. the second 'zoom'). The p-value profiles as a function of q-value in each of the lower redshift fields appear more scattered and varied compared with the GA results.

The key difference between the p-value profiles for the GA field and the p-value profiles for three lower redshift fields is that there is no sign of any significant results (p-value < 0.05) in any of the three lower redshift fields for any of the chosen q-values. The background probes were the same in all four fields (GA field and the three lower redshift fields) indicating that the Mg II absorbers are responsible for the different p-value profiles. From this we can assert that the GA field is markedly distinct, with significant clustering attributable to the GA.

As a further test of the dominance of the GA in the CE statistics we have applied our 'polygon-approach'. Visually, we draw a polygon around what we identify

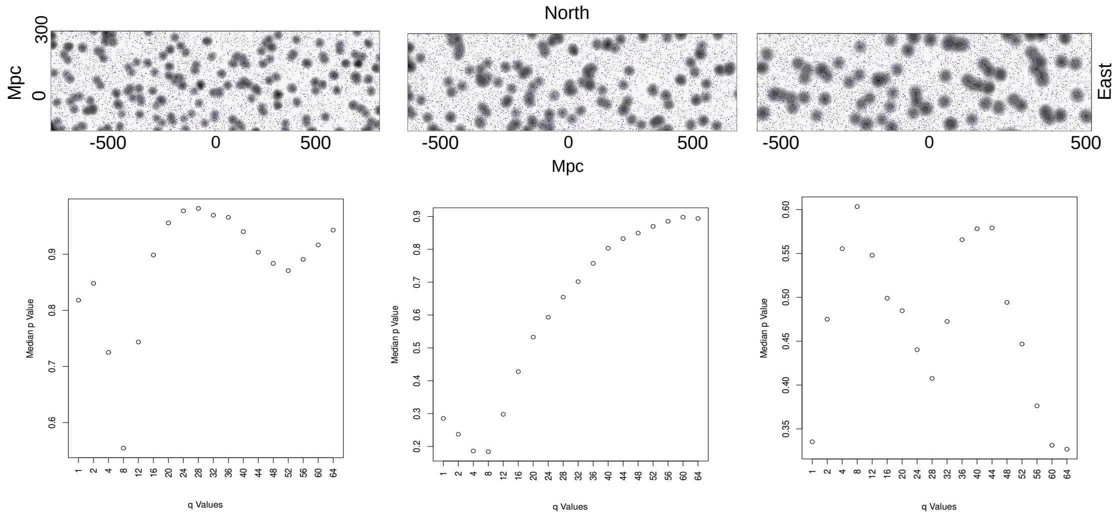


Figure 4.7: Top panels: density distribution of the flat-fielded Mg II absorbers represented by the grey contours (increasing by a factor of two) which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flatfielded with respect to the background probes (quasars). The background probes (quasars), represented by the small dots, have been restricted to be the same selection that arise in the standard GA field, i.e., $z_{qso} \geq 0.862$. In each figure, the field is centred on the GA field (as in the top-right panel of Figure 4.6), but at different central redshifts. Note, the tangent-plane y axis is not precisely accurate, but the precision is inconsequential: we are here concerned only with the CE test results, thus the Mg II images are simply providing some context. (Left) Mg II absorbers in the redshift interval $z = 0.682 \pm 0.060$. (Middle) Mg II absorbers in the redshift interval $z = 0.562 \pm 0.060$. (Right) Mg II absorbers in the redshift interval $z = 0.446 \pm 0.060$.
Bottom panels: the median p-values over 100 runs of 2000 simulations as a function of chosen q (k) value corresponding to the Mg II images in the top panels. The values along the x axis are: 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, and 64. The blue, horizontal line is set to $p = 0.05$. Evidently, there is no significant clustering detected in these fields.

visually as the member absorbers of the GA. We leave the absorbers in the polygon untouched but reassign (i.e. shuffle) at random the y-coordinates of absorbers outside the polygon, while avoiding the area within the polygon. We apply this process to the data corresponding to the top-right panel in Figure 4.6. In this way we can compare the CE statistics arising from the original data with those in which the GA points inside the polygon are unchanged but those outside the GA polygon are randomised. We find that the range of p-values for the original data (p-values $\sim$ 0.002–0.003) is very similar to that of the GA + randomised data (p-values $\sim$ 0.001–0.003), suggesting that the GA is indeed the dominant source of the clustering signal.

4.3.4 Power Spectrum Analysis

Power Spectrum Analysis (PSA) — see mainly Webster (1976a), but also Webster (1976b) and Webster (1982) — is a powerful Fourier method for assessing the presence and significance of clustering in rectangular (2D PSA) or cuboidal fields (3D PSA). PSA was designed to be effective for the detection of clustering that may be weak and escape detection by other methods; it is, however, not a case-control method. A brief summary of the theory of PSA may be found in section 5 of Clowes (1986).

We apply 2D PSA to the same rectangular field, illustrated in the top-right panel of Figure 4.6, that was used for the CE analysis above. Figure 4.8 shows the plot of the intermediate PSA statistic Q' against $1/\lambda$. The (six) high points towards the left of the plot allow a clustering scale of $\lambda_c \sim 270\,\mathrm{Mpc}$ to be identified. The final PSA statistic Q for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ.

We have applied the polygon-approach here also. As in the discussion above for the CE method, we leave the GA absorbers in the polygon untouched but reassign

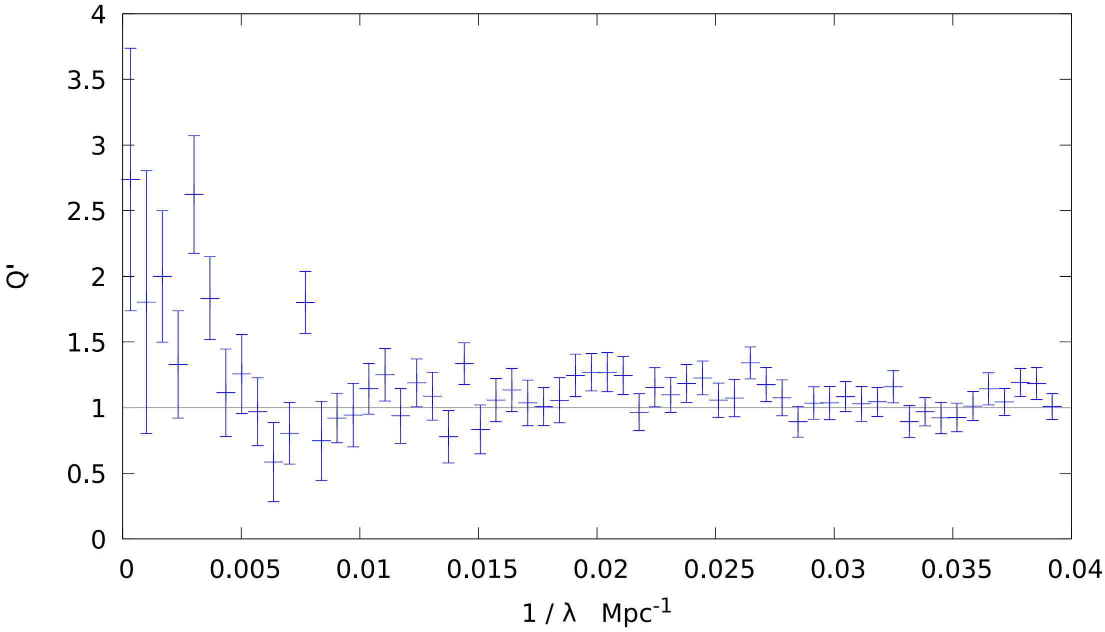


Figure 4.8: The plot shows the PSA statistic Q' plotted against $1/\lambda$, with λ in Mpc, for the 2D PSA. The bin size is $6.7 \times 10^{-4}\,\mathrm{Mpc}^{-1}$ and the error bars are $\pm\sigma$. The horizontal line $Q' = 1$ indicates the expectation value in the case of no clustering. The (six) high points towards the left of the plot allow a clustering scale of $\lambda_c \sim 270\,\mathrm{Mpc}$ to be identified. The final PSA statistic Q for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ.

(i.e. shuffle) at random the y-coordinates of absorbers outside the polygon, while avoiding the area within the polygon. In this way we can establish that the significance from the 2D PSA of the GA absorbers *alone* — i.e. with other absorbers randomised — has a mean value $\sim 3.5\sigma$ (with a range 3.0–4.4σ) for the original scale $\lambda_c \sim 270\,\mathrm{Mpc}$. In fact, the value of λ_c for the polygon-approach varies too, with the mean significance at the actual values of λ_c being $\sim 3.8\sigma$.

From this polygon-approach, it appears likely that, while the GA is the dominant contributor to the PSA result, a smaller contribution from other absorbers in the field is detected too. This outcome might be expected, given the power of the PSA method. The failure to detect a contribution from the other absorbers with the CE method could be because the CE method is intrinsically less sensitive, or because its case-control correction has successfully eliminated artefacts from the background probes (the controls).

The polygon-approach can also be used to assess the relative power of the 2D PSA and the CE test. For example, we reduced the number of GA absorbers in the polygon from 52 to 42 by random selection (with the points outside the polygon being randomised as usual but unchanged in total number). In that case, the GA is generally not detected by the CE test, at a significance level of 0.01 (2.3σ), but is generally still detected by the PSA, at $> 2.7\sigma$. The power of a statistical test to discriminate is an important factor, and an uninteresting p-value does not necessarily mean nothing interesting in the data. Webster (1976a) demonstrates that the PSA has more power to detect clustering than a simple nearest-neighbour test. The CE test uses multiple neighbours, and so can be expected to have more power than a nearest-neighbour test, but, as our tests with the polygon-approach suggest, still has less power than the PSA. Of course, the CE test has the useful feature of case-control comparisons, whereas the PSA does not.

4.4 Overdensity

The SLHC / CHMS method splits the GA into two agglomerations - one large, statistically significant portion which makes up the majority of what we visually identified as the GA, and one smaller, statistically not significant portion which makes up the remainder of what we visually identified as part of the GA. We will refer to these agglomerations as GA-main and GA-sub, respectively, for simplicity. The overdensity of the GA can be calculated using the CHMS approach as described earlier (Section 4.3.1). However, in the case of a strongly curved structure such as the GA (and GA-main), the MST-based method of Pilipenko (2007) can have some advantages. We shall refer to these two methods as CHMS-overdensity and MST-overdensity, respectively. The MST-overdensity does not consider the physical volume of the structure being assessed. Instead, it calculates the overdensity based on the MST edge-lengths: $\delta = \langle l_0^3 \rangle / \langle l^3 \rangle - 1$ where l is the MST edge-length for the structure and l_0 is that for a control field. Given the curvature of GA-main, the CHMS-volume and CHMS-overdensity refer to a volume that encloses both GA-main and some empty, unrelated space around it, particularly above the GA where there is positive curvature enveloping empty space. Therefore, the CHMS method is likely to overestimate the volume and underestimate the overdensity. In contrast, the MST-overdensity, which is an internal measure that considers only the points belonging to the group, and no additional space arising from curvature, is likely to be a better estimate of the overdensity. Conversely, GA-sub is a globular shape, so it is possible to construct a unique volume enclosing the space surrounding *only* the absorbers attached to GA-sub and not additional empty space. Therefore, we would expect the CHMS-overdensity calculation for GA-sub to be a fair estimate.

GA-main has a CHMS-significance of 4.5σ, while GA-sub has a much smaller CHMS-significance of 2.3σ. By splitting the usual control field (using the *larger* field-of-view, see Figure 4.3) into eight portions — four quarter segments and four half segments — we repeatedly calculate the significance and the overdensities of GA-main and GA-sub using different control fields. An uncertainty can then be estimated for the significance and overdensities for both portions of the GA. Our results are as follows: (1) GA-main, containing 44 Mg II absorbers, has a CHMS-significance of $(4.5 \pm 0.6)\sigma$; a CHMS-overdensity of $\delta\rho_{CHMS}/\rho_{CHMS} = 0.9 \pm 0.6$; and an MST-overdensity of $\delta\rho_{MST}/\rho_{MST} = 1.3 \pm 0.3$; (2) GA-sub, containing 11 Mg II absorbers, has a CHMS-significance of $(2.1 \pm 0.9)\sigma$; a CHMS-overdensity of $\delta\rho_{CHMS}/\rho_{CHMS} = 1.5 \pm 0.3$; and an MST-overdensity of $\delta\rho/\rho = 1.3 \pm 0.3$. As expected, the CHMS-overdensity is lower than the MST-overdensity for GA-main, indicating that the CHMS unique volume encapsulating GA-main is likely to be an overestimate because of the curvature of the arc. In contrast, for GA-sub, which has a globular-shape, the CHMS-overdensity and the MST-overdensity have similar values, as expected when there is no marked curvature. In addition, the CHMS-overdensity has a much larger error than the MST-overdensity which suggests giving preference to the latter. Notice here that both GA-main and GA-sub have the same MST-overdensity, which supports their belonging to the same structure.

A final method of calculating the number overdensity is to simply draw a rectangle around the visually-selected Mg II absorbers in the GA and compare the number of absorbers per unit area in the rectangle to the number of absorbers in the whole field. The method will underestimate the GA overdensity for three reasons: (i) the GA contributes to the density of the whole field, although only by a small fraction; (ii) the rectangular shape around the GA overestimates the area encompassing the GA as the GA is curved, therefore having a large portion of ‘empty’ space (with non-GA absorber members); (iii) this method encompasses the *whole* GA from visual inspection, rather than separating it into two agglomerations like the CHMS method, thus reducing the overall density in the GA rectangle. Using this method we calculate an overdensity of $\delta\rho/\rho \sim 0.93$.

In addition to the number overdensity, we can estimate the mass excess by assuming $\delta_n = \delta_m$, where δ_n is the MST-overdensity and δ_m is the mass overdensity. We use here the MST-overdensity, rather than the CHMS-overdensity. We are here taking the critical density of the Universe to be $9.2 \times 10^{-27}\,\mathrm{kg\,m^{-3}}$, as calculated using the cosmological parameters used throughout this thesis, and the matter-energy density parameter to be $\Omega_{M0} = 0.27$. The mass excesses for GA-main and GA-sub are then $1.8 \times 10^{18} M_\odot$ and $3.4 \times 10^{17} M_\odot$ respectively. Note that the mass excess of GA-main + GA-sub is comparable to that of the Huge-LQG (Clowes et al., 2013).

4.5 Comparisons with Other Data

Independent corroboration of a very large LSS by an independent tracer can provide compelling support. In the case of the Huge-LQG (Clowes et al., 2013), a $\sim$ Gpc structure of quasars, independent corroboration was provided by Mg II absorbers. Here, we can invert this approach and look for corroboration of the GA, a $\sim$ Gpc structure of Mg II absorbers, in quasars. We use the SDSS DR16Q database (Lyke et al., 2020). In addition, we look at the databases of DESI galaxy clusters from Zou et al. (2021).

We are concerned at this stage with simple visual inspection, and will leave the subtleties of correcting for possible artefacts in the DR16Q quasars and the DESI clusters to future work. Our approach here will be simply to superimpose contours for the spatial distribution of the quasars (in blue) and the clusters (in green) onto the Mg II density images (grey, as previously).

We begin with the quasars, selected for the same redshift interval as the GA. We show two cases, one for quasars with $i \leq 20.0$ (Figure 4.9) and one for $i \leq 19.5$ (Figure 4.10). We anticipate that we should then be restricting to 'traditional' high-luminosity quasars. In both cases, it is immediately clear that the quasars follow the same general trajectory as the GA. The quasars are entirely unrelated to the probes of the GA, and so we have in these plots quite striking independent corroboration of the GA. Furthermore, the tendency of the Mg II absorbers in general and the quasars to share common paths and voids is apparent, especially so in Figure 4.10.

Note that there is a density boundary in the distribution of the DR16Q quasars: in roughly the lower third of the plots the density of the quasars is lower than above. This artefact, however, is well separated from the GA and does not affect our visual assessment.

We continue with the DESI clusters, again selected for the same redshift interval as the GA — Figure 4.11. Note that the redshifts for the DESI clusters are photometric, with redshift errors ~ 0.024 at $z \sim 0.9$ (Zou et al., 2021). (In contrast, we might expect the redshift errors for the quasars to be ~ 0.003.)

There is no compelling association of the DESI clusters and the GA, although there is perhaps a hint on the RHS. Possibly the substantial errors in the photometric redshifts are a factor in diluting any correspondence that might exist. An interesting feature in Figure 4.11 is, however, the 'cluster of clusters' in the centre of the GA, largely coinciding with the central small gap in the Mg II absorbers of the GA. It could be a large supercluster, with the SZ cluster B18 as one of its member clusters. We previously mentioned, in Section 4.2, that there appears to be a set of strong Mg II absorbers enveloping a circular hole in the centre of the GA. It seems likely that these enveloping strong absorbers and the central hole are related to this putative supercluster.

The median richness limit for the DESI clusters is 22.5 (Zou et al., 2021). Figure 4.12 shows the relationship between the Mg II absorbers and DESI clusters with richness $R \leq 22.5$. It suggests that there could be some association of the low-richness clusters with the Mg II absorbers, both for the GA and in general.

Finally, we compare the DESI clusters with the DR16Q quasars — Figure 4.13. As with the Mg II absorbers and DESI clusters, there appears to be no compelling

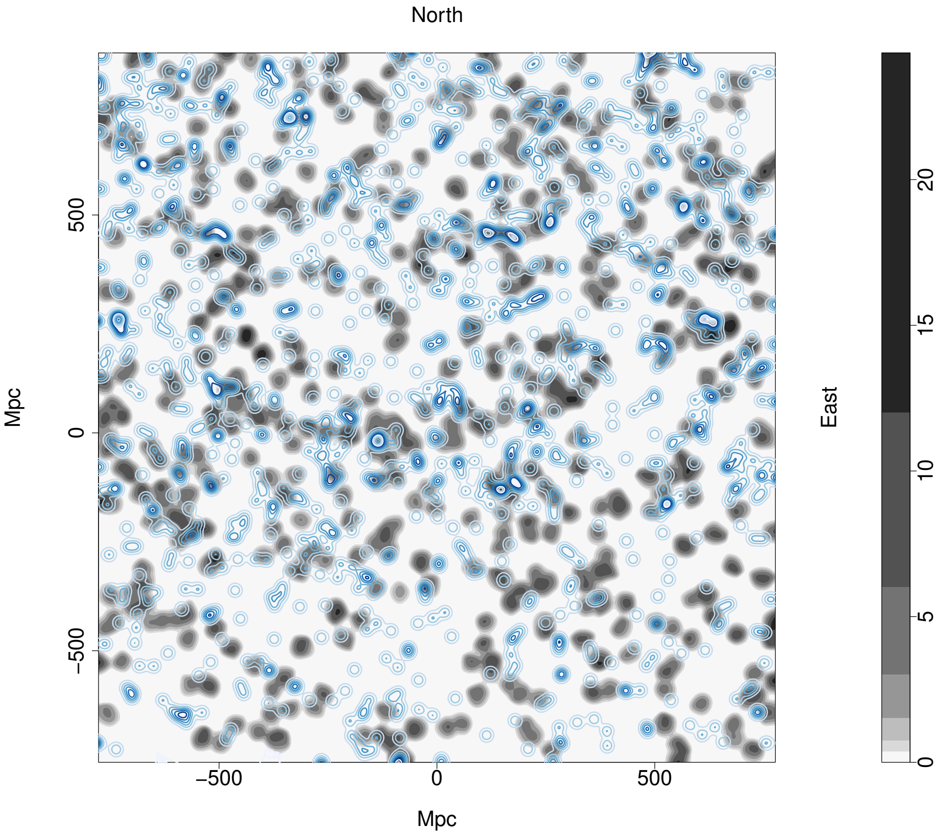


Figure 4.9: Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of two. Blue contours represent the DR16Q quasars with $i \leq 20.0$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. The GA can be seen stretching across ~ 1 Gpc in the centre of the figure (at tangent-plane y-coordinate ~ 0 Mpc). Visually, many of the blue contours seem to follow the same general trajectory as the grey contours, e.g., along the GA and along the filament above the GA located around (250, 450) Mpc.

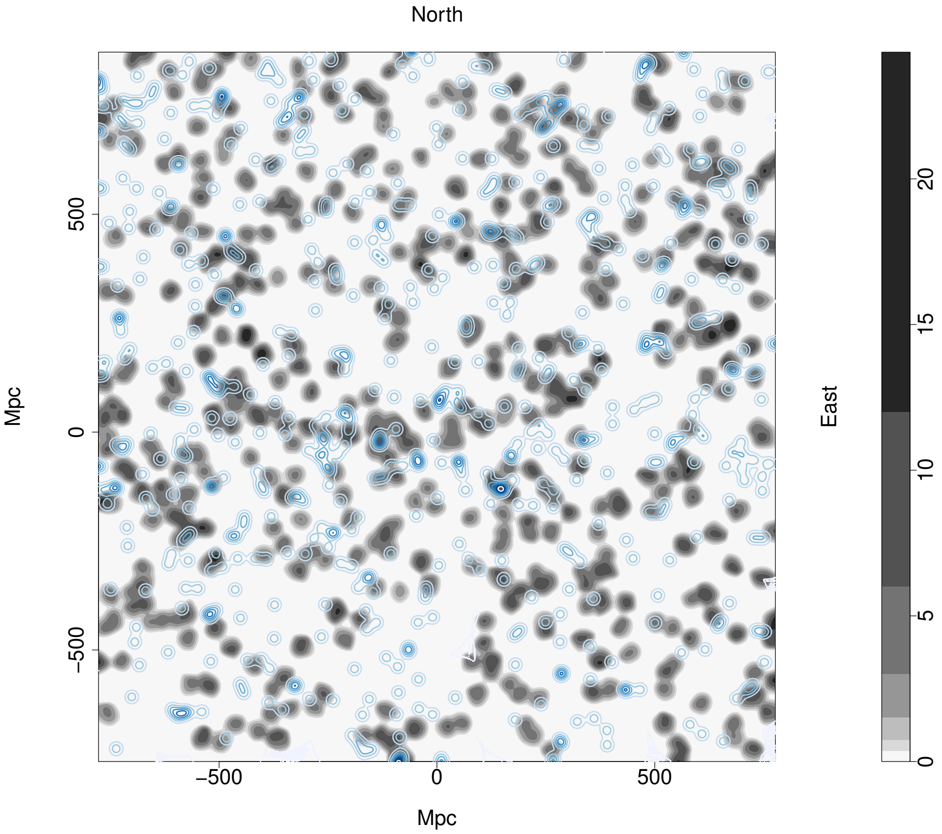


Figure 4.10: Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of two. Blue contours represent the DR16Q quasars, with $i \leq 19.5$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. The GA can be seen stretching across ~ 1 Gpc in the centre of the figure (at tangent-plane y-coordinate ~ 0 Mpc). Visually, the blue contours seem to follow the same general trajectory as the grey contours, as before — notably, the blue contours along the GA and the small filament above the GA located around (250, 450) Mpc — even with the reduced number of quasars from the larger magnitude cut.

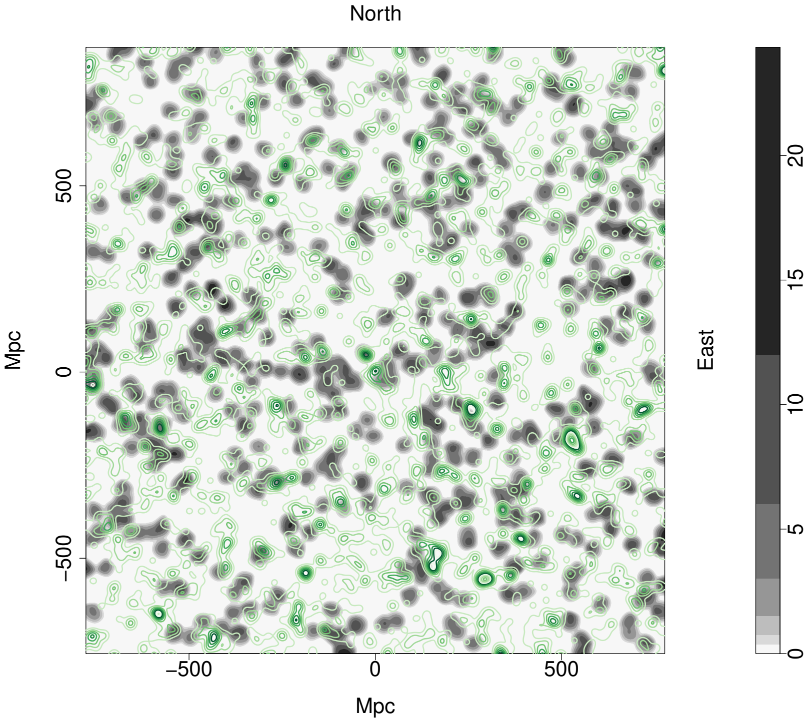


Figure 4.11: Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of two. Green contours represent the DESI clusters, of all richnesses, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. The GA can be seen stretching across ~ 1 Gpc in the centre of the figure (at tangent-plane y-coordinate ~ 0 Mpc). There are no immediately compelling connections between the DESI clusters and the Mg II absorbers in this figure.

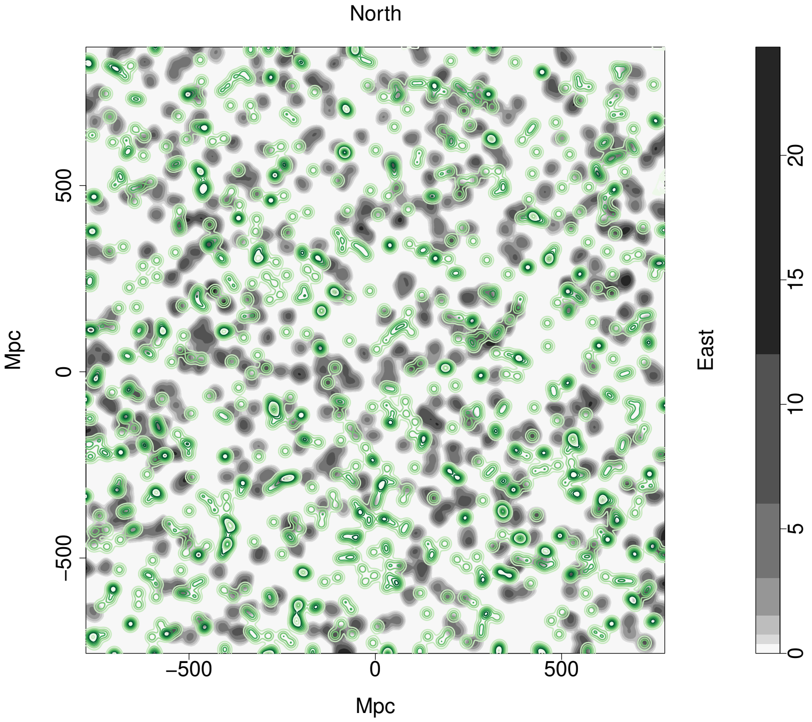


Figure 4.12: Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of two. Green contours represent the DESI clusters, with the richness limit $R \leq 22.5$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. The GA can be seen stretching across ~ 1 Gpc in the centre of the figure (at tangent-plane y-coordinate ~ 0 Mpc). There are a few occurrences of the green contours following the grey contours, e.g., the RHS tip of the GA is traced by a strong filament of DESI clusters.

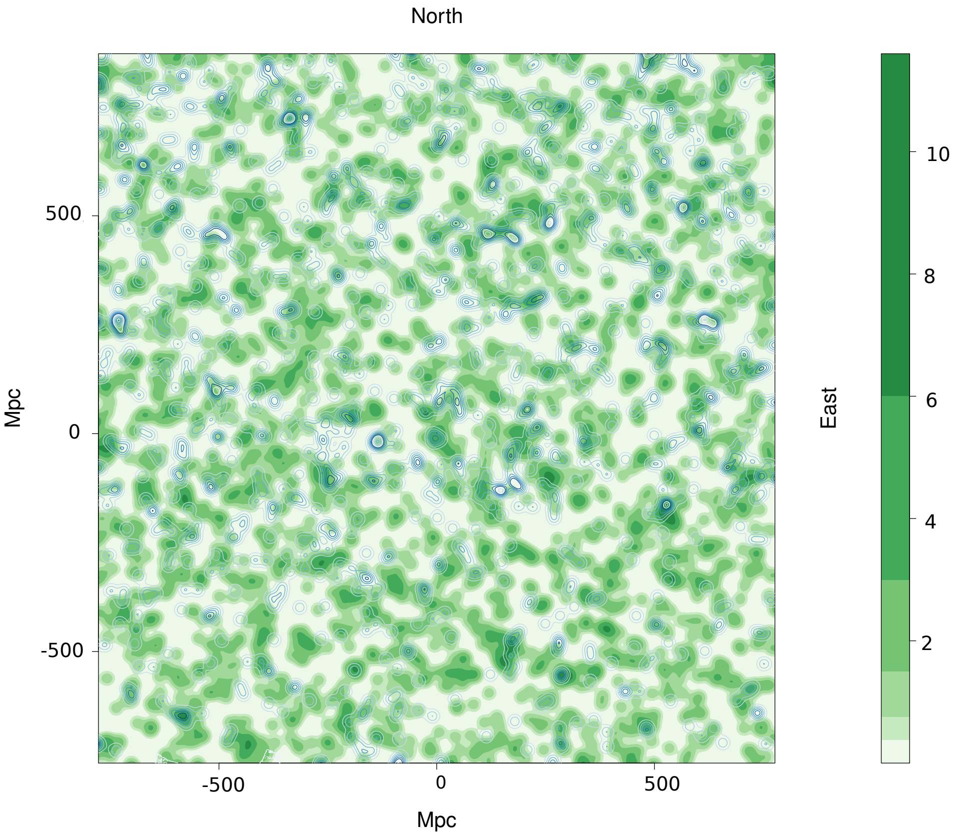


Figure 4.13: Density distribution of the DESI clusters, of all richnesses, in the redshift slice $z = 0.802 \pm 0.060$ represented by the green contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of two. Blue contours represent the DR16Q quasars, with the magnitude limit $i \leq 20.0$, in the same redshift slice as the DESI clusters, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. There are no compelling connections between the DR16Q quasars and the DESI clusters.

association. However, again, the lowest-richness clusters suggest some association — Figure 4.14.

From the independent corroboration above, we suggest that the GA, and the Mg II absorbers in general, are associated with luminous quasars but not strongly with DESI clusters. However, there is potentially an association of the Mg II absorbers and the quasars with the low richness clusters. More statistical details of the relationship between Mg II absorbers, quasars and clusters will be investigated in our future work.

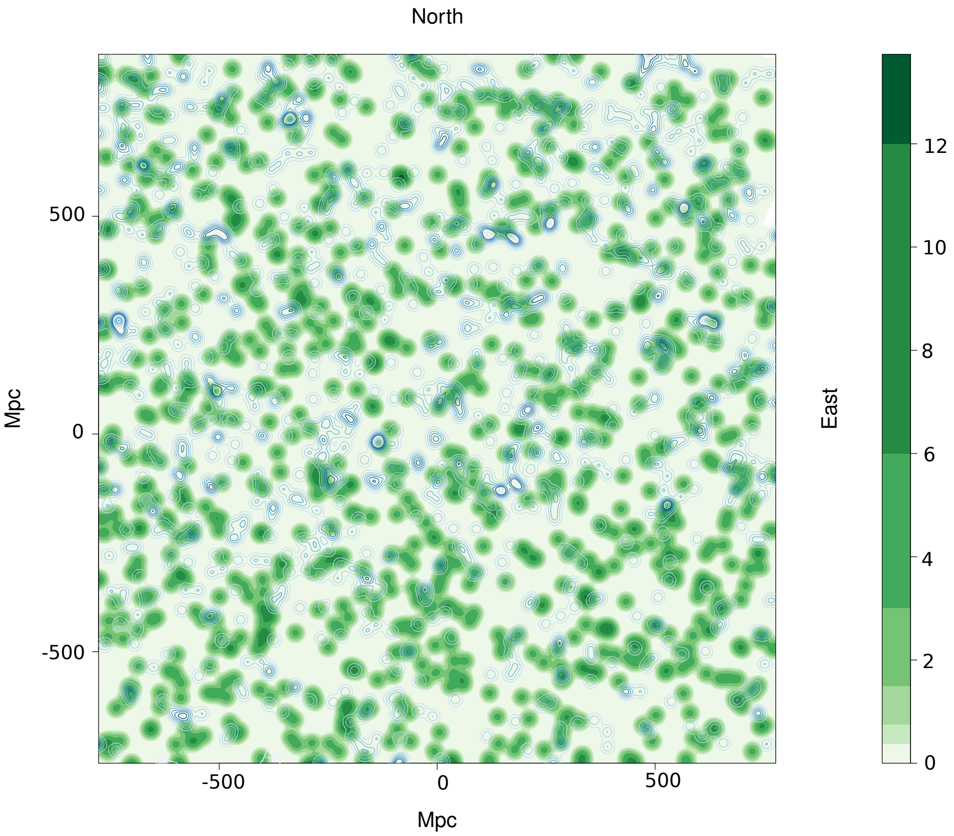


Figure 4.14: Density distribution of the DESI clusters in the redshift slice $z = 0.802 \pm 0.060$ represented by the green contours, increasing by a factor of two, which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and limited to show only those clusters with a richness $R \leq 22.5$. Blue contours represent the DR16Q quasars, with the magnitude limit $i \leq 20$, in the same redshift slice as the DESI clusters, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of two. There are a few occurrences of the blue contours following the green contours indicating that there might be some association of DR16Q quasars with low richness clusters.

4.6 Discussion and Conclusions

In this chapter I have presented the discovery of the Giant Arc (GA); the work pertaining to this discovery was carried out during the first two years of my (part-time) PhD. The work was first presented at the # 238th American Astronomical Meeting (AAS; online) in June 2021 and later it was written-up, peer-reviewed and published in the Monthly Notices of the Royal Astronomical Society (Lopez et al., 2022), which can be found appended to the back of this thesis.

The GA is a $\sim$ Gpc uLSS at $z \sim 0.8$, mapped by Mg II absorption systems in the spectra of background quasars. The GA forms a large crescent shape on the sky that appears almost symmetrical. However, deeper analysis reveals some asymmetries in the GA, in the redshift and equivalent width (EW) distributions. The GA spans $\sim$ 1 Gpc on the sky and has a redshift depth of $\sim$ 340 Mpc (both proper sizes, present epoch). Visually, we determine the GA as a single unit, but using a Minimal Spanning Tree (MST) type algorithm (Section 4.3.1) it splits into two portions: a large portion (GA-main) and a small portion (GA-sub). We proposed in Section 4.3.1 that the two portions of the GA could in fact be connected in reality, since potentially one more background probe could lead to one more Mg II absorber that would connect the two portions. On its own, GA-main is a statistically-significant clustering of Mg II absorbers, with a membership of 44 Mg II absorbers, an MST-overdensity of 1.3 ± 0.3, and a mass excess of $1.8 \times 10^{18} M_{\odot}$. In these respects, the GA is comparable to the Huge-LQG (Clowes et al., 2013).

Three different statistical tests were applied to the GA to assess the significance of connectivity and clustering. The results of each are summarised here. (i) The SLHC / CHMS method calculates the significance of clustering between points of close proximity by comparing the volumes of the CHMS for each structure to the CHMS of structures in randomly distributed points in a cube. GA-main, containing 44 Mg II absorbers, has a significance of $(4.5 \pm 0.6)\sigma$. GA-sub, containing 11 Mg II absorbers, has a significance of $(2.1 \pm 0.9)\sigma$. Both GA-main and GA-sub have the same MST-overdensity of $\delta\rho/\rho = 1.3 \pm 0.3$. This fact could indicate, as we suspect, that both agglomerations are connected in reality. (ii) The CE test is a case-control k nearest-neighbour algorithm that assesses the p-value of clustering in the field within an unevenly distributed population. A process of 'zooming' into the GA field allows the GA to become increasingly dominant. In this way, we detect a p-value of 0.0027 from the field seen in top-right panel of Figure 4.6, equivalent to a significance of 3.0σ. Applying this process of 'zooming' to lower redshift slices at the same sky coordinates of the GA, we do not detect any significant clustering. We then use our polygon approach that randomises points outside the GA while keeping the visually selected absorbers contained within the GA the same. The CE test detects a similar p-value of clustering with the polygon approach, indicating that the GA is the true, dominating feature causing significant clustering. (iii) The PSA is a Fourier method of detecting clustering in the field on a physical scale. We apply the 2D PSA to the 'zoomed' GA field, top-right panel of Figure 4.6, and find significant clustering at $\lambda_c \sim 270\,\mathrm{Mpc}$ with a significance of 4.8σ. As with the CE test, we use our polygon approach and detect similar significant clustering scales. However, a small contribution from other absorbers in the field is also detected. We do expect this given the power of the PSA test, and it is clear that the GA is still the dominant contributor to the PSA result.

Clearly, the analysis of the GA is after the event of its discovery. We have applied several different approaches to mitigating any *post-hoc* aspects of analysing the statistical significance of the GA after discovery. We have performed techniques that aim to assess the GA unbiassedly, such as the polygon approach, varying redshift slices, zooming into the GA field, and randomised simulations. In the future we can apply the same techniques used for the GA field to the whole of the Mg II dataset. In addition, the Mg II dataset is quite complex, with features that need careful attention: for example, the inhomogeneities of the Mg II absorbers are superimposed on the inhomogeneities of the quasars (background probes) and of the survey. Finally, there are different Mg II databases available from different authors, each using different detection processes. We intend eventually to produce our own databases of Mg II detections that can be used consistently with past and future quasar-survey data releases.

The GA is now amongst several other very large LSS discoveries with sizes that exceed the theoretical upper-limit (~ 370 Mpc) scale of homogeneity of Yadav et al. (2010). Potentially, there are other such significant structures in the rest of the Mg II database. We discuss that there are challenges in fairly characterising the population of structures due to the inhomogeneities in the background probes (quasars). However, the challenges can be managed with suitable care, allowing for the Mg II method of studying LSS to be fully exploited.

In standard cosmology we expect to find evidence for a homogeneous and isotropic Universe on large scales. However, the accumulated set of LSSs (and uLSSs) now seems sufficient to constitute a prima facie challenge to the assumption of the Cosmological Principle (CP). A single anomaly, such as the GA on its own, could be expected in the standard cosmological model. For example, Marinello et al. (2016) find that the Huge-LQG (Clowes et al., 2013), a structure comparable in size to the GA, is, by itself (there are others), compatible with the standard cosmological model. However, Marinello et al. (2016) state that this is on the condition that only one structure as large as the Huge-LQG is found in a field ~ 5 times the sample survey, in this case, the DR7QSO quasar database for $1.2 \leq z \leq 1.6$. Note that the GA is found in the combined footprint from DR7QSO and DR12Q (the combined footprint being almost the same area as the individual footprints), in a narrow redshift interval, so its challenge to the CP seems likely to be exacerbated. Of course, the GA is now the fourth largest LSS, so there are, at minimum, four

LSSs comparable to the size of the Huge-LQG, plus several other LSSs exceeding the Yadav-estimated scale of homogeneity ($R_h \sim 370$ Mpc). We suggest that there is a need to explore other avenues within cosmology that could explain multiple, very large LSSs and uLSSs.

We bring attention to the Sloan Great Wall (SGW) (Gott et al., 2005), which is a large, wall-like filament in the relatively local Universe. The SGW is ~ 450 Mpc in its longest dimension, which is ~ 0.5 times the length of the GA. One can note some of the similarities between the SGW and the GA, such as the general shape and comoving size — they are both long, filamentary and curved walls made up of galaxies and galaxy clusters (for the GA, the galaxies and galaxy clusters are implied by the Mg II absorbers) —, and so perhaps also envision a LSS such as the GA as a precursor to the SGW. The GA is at a redshift of ~ 0.8 which means we are seeing it when the Universe was only half its present age. Perhaps the SGW, at an earlier epoch, initially looked more like the GA. At this point, these ideas are speculative only; simulations (possibly even with alternative cosmological models) could conceivably elucidate any such potential connections between structures like the GA and the SGW.

Chapter 5

A Big Ring on the Sky

In this chapter I present the Big Ring (BR); the *second* ultra-large large-scale structure (uLSS) found in the Mg II catalogues (here from Anand et al., 2021). The BR is a striking circular, annulus-like, structure of diameter $\sim$ 400 Mpc (proper size, present epoch). In cosmological terms the BR is close to the previously-discussed GA — at the same redshift $z \sim 0.8$ and with a separation on the sky of only $\sim 12°$. Two extraordinary uLSSs in such close configuration raises the possibility that together they form an even more extraordinary cosmological system; this will be discussed in more detail in Chapter 6. The discovery and analysis of the BR is presented in Lopez et al. (2024), which can be found appended to the back of this thesis. The published work in the paper is presented in this chapter, but has been adapted for the thesis.

5.1 Discovery of the Big Ring

During the discovery of the Giant Arc (GA; Lopez et al., 2022, hereafter Lopez22), the SDSS Data Release 16 quasar database (DR16Q) became available (Lyke et al., 2020). Then, independent authors created the most up-to-date Mg II database from the DR16Q quasars (Anand et al., 2021, hereafter Anand21). Following the discovery of the GA — made using the older Mg II database from Zhu & Ménard (2013) (hereafter Z&M), with the corresponding DR7QSO and DR12Q quasar catalogues from Schneider et al. (2010) and Pâris et al. (2017), respectively — we are now in a position to continue LSS investigations with the new DR16Q database and corresponding DR16Q Mg II database. We have found an interesting ring shape in the Mg II absorbers, indicating a LSS of galaxies and galaxy clusters, that spans a diameter of ~ 400 Mpc scaled to the present epoch. Incidentally, the estimated size of the BR is close to that which could be expected in a detection of an individual Baryon Acoustic Oscillation (BAO), $r \sim 150$ Mpc (Tully et al., 2023; Einasto et al., 2016; Planck Collaboration et al., 2016b; Anderson et al., 2014; Eisenstein et al., 2005), but we later suggest that the BR is unlikely to have its origins in BAOs. The BR shape and size are both hard to understand in our current theoretical framework. Additionally, the BR is in the same redshift slice as the GA and to the north of the GA by $\sim 12°$, which raises further questions about their origin both together and independently.

For the work that led to the original discovery of the GA, using the Z&M database, we had looked at only a few small areas of sky and redshift slices (Lopez, 2019), essentially to test the viability of the Mg II approach itself. Following the discovery of the GA, and now using the Anand21 database, we have so far concentrated on the GA field and redshift slice because we immediately made the further discovery of the BR there; the only exceptions to this statement are (i) the use of adjacent redshift slices to test that the BR was not arising from artefacts and to test for extensions of the BR into adjacent redshift slices, and (ii) the use of neighbouring fields at the same redshift as the GA/BR field for comparing the spatial clustering results of the Cuzick and Edwards test (see section 5.3.4). Consequently, a ‘look-elsewhere’ effect on the statistical assessments should not be a factor. In future, of course, we intend to explore both databases in their entireties.

The Mg II data we use here are complicated and quite difficult to manage. The advantage here is the precise Mg II redshifts, with the concomitant possibility of discovering intriguing structures such as the Big Ring and the Giant Arc. In future, the Mg II approach to LSS should be enhanced by DESI spectra (Napolitano et al., 2023), taken with the KPNO 4m telescope, allowing detection of Mg II to lower equivalent widths, and hence allowing the exploration of finer detail in the morphology of structures. In the following sections the BR is assessed both regarding its reality and statistical significance with respect to the assumed, homogeneous large-scale distribution of matter.

In Figure 5.1, we are seeing the BR, which is the visually overdense ring shape of Mg II absorbers centred at approximately $x = 0$ Mpc and $y = 240$ Mpc. (The large 'void' to the south-west of the BR is also particularly striking.) In this figure, as usual, the grey contours, increasing by a factor of two, represent the density distribution of Mg II absorbers in the specified redshift slice and field-of-view (FOV). The Mg II contours have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes (quasars). The smoothing gives a useful impression of the connectivity. The background probes (quasars) are represented by the small, blue points. The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top. From Figure 5.1 we can estimate the BR diameter is $\sim 300 - 400$ Mpc, which would make its circumference comparable to the extent of the GA.

5.2 Initial checks of the data

First, we check that the Mg II absorbers belonging to the visually-identified BR are real (not false positive detections). We visually inspected spectra of 56 DR16Q quasars that are the probes that correspond to the visually-identified BR and inner filament absorbers (see Figure 5.2). Since 6 of the Mg II absorbers are multiples per

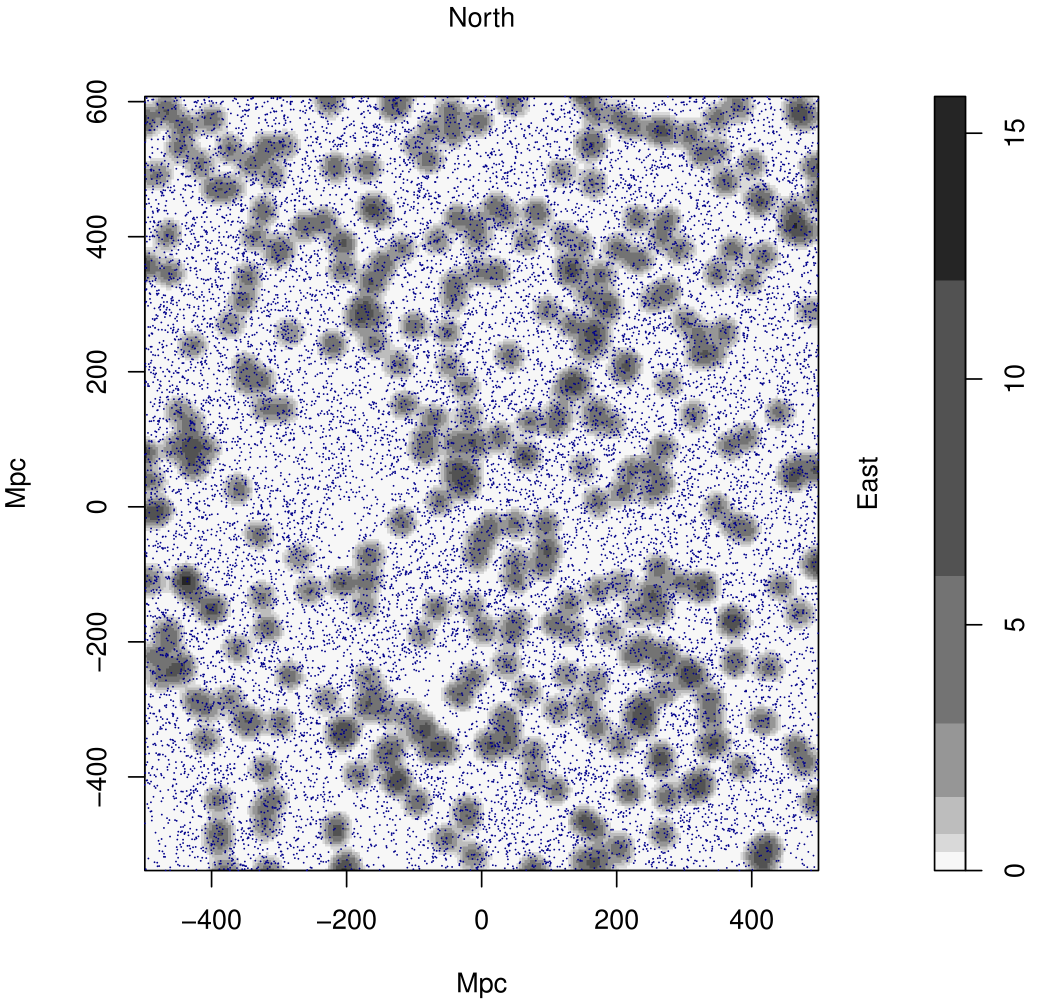


Figure 5.1: The tangent-plane distribution of Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$. The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes (quasars). The dark blue dots represent the background probes. S/N limits of: $4, 2$ and 4 were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively (details of S/N cuts are discussed in Section 5.3.1). The BR can be seen to the north of the centre point spanning ~ 400 Mpc in diameter. The field-of-view corresponds to the small, pink area seen in Figure 5.5.

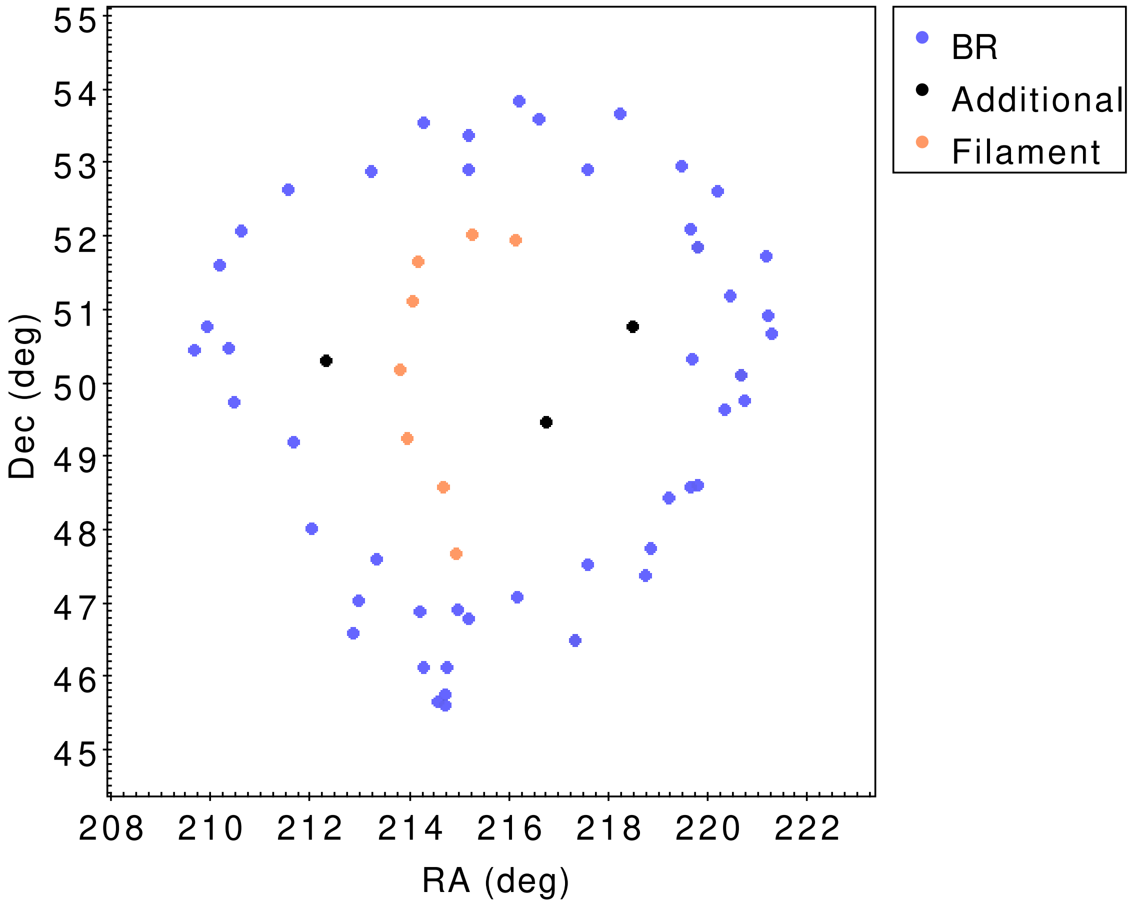


Figure 5.2: The Mg II absorbers in the visually-identified BR (blue), the inner filament (orange), and the additional absorbers inside the BR (black) from the on-sky perspective. Three of the blue points (BR) correspond to six absorbers occurring in three quasar spectra (two absorbers per probe, see Figure 5.8).

probe occurring in three quasar spectra, we were checking for a total of 59 Mg II doublet systems. Each Mg II absorption doublet that we searched for was visually confirmed and in agreement with the documented redshifts from Anand21, so we can confirm that 100% of the Mg II absorber members of the BR and inner filament are real, physical Mg II absorbers indicating the presence of intervening matter.

There are two unusual Mg II systems, occurring in the same spectrum, for which the Anand21-documented redshifts suggest that the $\lambda2796$ line of the lower-z absorption doublet appears at the same wavelength as the $\lambda2803$ line of the higher-z absorption doublet. This is indeed the case: the two Mg II absorption doublets appear as 3 absorption lines in the spectrum, which is a rare oddity. Although Anand21 recognise the 3 absorption lines as two systems, they appear not to have disentangled the EWs of the centre absorption line of the triplet (the higher-z $\lambda2796$ and the lower-z $\lambda2803$ EWs being the same).

Secondly, we investigate if the visually obvious BR is an artefact of the probes. This can be done in two ways: simply checking the density distribution of background probes and checking for obvious artefacts; and looking at the next redshift slice down from the BR field (on the near side) and checking for repeating Mg II features that correspond to any obvious artefacts in the probes. For the former, see the left-hand panel of Figure 5.3.

Many overdensities and underdensities can be seen clearly in the left-hand panel of Figure 5.3. In particular, there are a few overdense regions (small, dark clumps) centred at 0 Mpc on the x-axis and between roughly 100 Mpc to 400 Mpc on the y-axis, a few of which appear to coincide with the inner filament of BR. This could imply that the inner filament of the BR is suspect, so we will need to be sure that the filament is not an artefact of the probes. We can visually check the rate of occurrence of Mg II absorbers at the position of the overdense artefacts (dense blobs) by blinking the image of the probes with the image of the absorbers (left- and

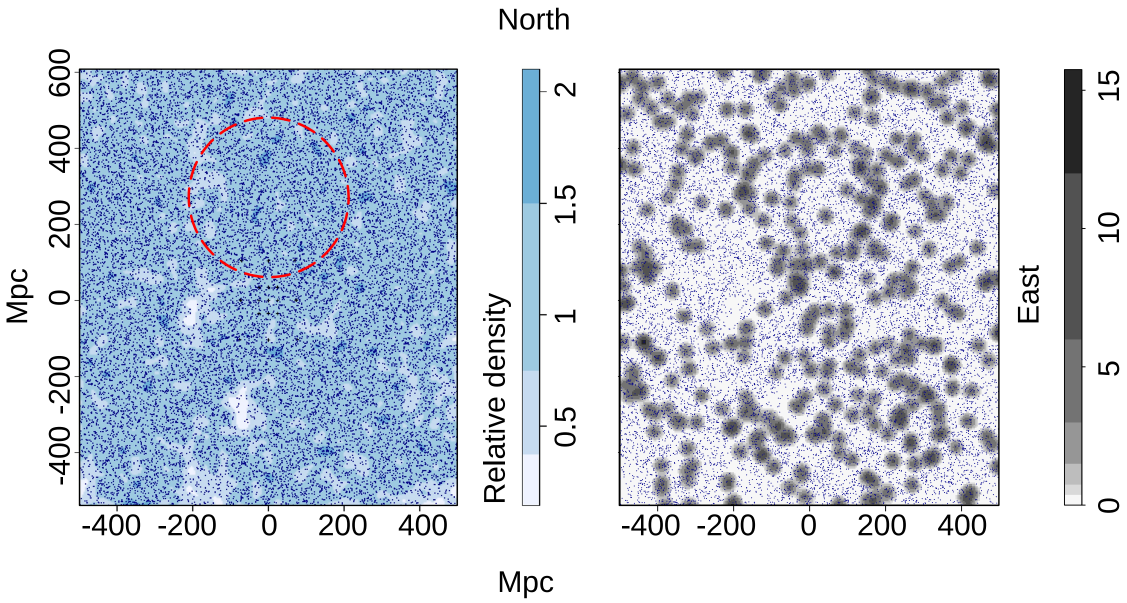


Figure 5.3: (Left) The tangent-plane distribution of probes (background quasars) in the BR FOV with the redshift condition $z > 0.862$ — i.e., the probes that are responsible for the Mg II absorbers arising in the BR field. The blue contours, increasing by a factor of two, represent the density distribution of the probes which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc. S/N limits were applied to the quasar continuum such that S/N$_{con} \geq 4$ (details of S/N are discussed in Section 5.3.1). A red, dashed circle has been added to outline the approximate location of the BR. The figure shows many areas of overdensities and underdensities. In particular, there are a few overdense regions (small, dark clumps), centred at 0 Mpc on the x-axis and between roughly 100 Mpc to 400 Mpc on the y-axis, a few of which might coincide with the inner filament of BR. There is also a much larger region of underdense probes spanning -100 Mpc to 400 Mpc in the y-axis and centred at -200 Mpc in the x-axis which coincides with the LHS of the BR.
(Right) The corresponding Mg II image for reference, as in Figure 5.1.

right-hand panels of Figure 5.3). Doing this shows that 8 out of 20 randomly-selected overdense artefacts had Mg II absorbers present, so less than half of the artefacts. Incidentally, we could also confirm that most of the artefacts that appeared to coincide with parts of the BR (including the inner filament) were in fact offset, so not responsible for the Mg II absorbers arising there. Checking the artefacts shows that there is no particular association of the artefacts with the Mg II absorbers present in the BR field. There is also a much larger region of underdense probes spanning -100 Mpc to 400 Mpc in the y-axis and centred at -200 Mpc in the x-axis which partially coincides with the LHS of the BR. The fact that part of the BR is located in an underdense region of probes is noteworthy.

For the latter way to test if the MgII absorbers are artefacts of the probes, we can check the absorbers arising from the same set of probes corresponding to the BR field in the next, non-overlapping redshift slice on the near side of the BR. To do this we keep the probes of the BR field the same (having $z > 0.862$ — i.e., a redshift greater than the far edge of the BR Mg II redshift slice) and map the Mg II absorption in the nearest, non-overlapping redshift slice (i.e., $z = 0.682 \pm 0.060$). In this way we are able to search for any obvious artefacts of the probes that could be responsible for the specific distribution of Mg II absorbers in the BR field by comparing the Mg II image in the usual BR field with the neighbouring redshift slice. We apply the Single-Linkage Hierarchical Clustering (SLHC) and the Convex Hull of Member Spheres (CHMS) algorithms (see Section 4.3.1, previous, or Section 5.3.1, ahead), and the FilFinder algorithm (see Section 5.3.3) to the field centred at $z = 0.682 \pm 0.060$ and corresponding to the usual BR FOV. In the redshift slice $z = 0.682 \pm 0.060$ we find with the SLHC / CHMS method that there are no structures detected by the SLHC algorithm corresponding to the BR, and in addition, no structures at all that are statistically significant. We also find with the FilFinder method, no filaments detected in the field that correspond to the BR (Figure 5.4). Therefore, we conclude that the BR is not an artefact of the probes.

The BR appears in the same redshift slice and FOV as the previously documented GA, but we are now using the new Anand21 databases and not the previously used Z&M databases. Previously, in Chapter 3, I discussed some differences between the 'old' and the 'new' Mg II catalogues. There was a $\sim$ 60% overall agreement in the background probes between the two datasets. Consequently, the GA appears somewhat different in this new dataset (see Figure 3.4). The GA is still the most significant, most numerous and overdense structure detected in the field despite the slight change in appearance, so quantitatively, there is very little difference in the GA in the new dataset. However, qualitatively, there are two main reasons the GA appears different: (1) the field overall is much more dense with the new dataset due to many more quasar observations; (2) Anand21 miss several GA absorbers. We investigate the second point by manually checking each of the quasar spectra that are probes to the Mg II absorbers in the GA that Anand21 missed. There were 16 from 51 absorbers that Anand21 missed (none was due to the small fraction of removed/missing quasars in DR16Q that was mentioned in Chapter 3). In each of the quasar spectra (corresponding to the GA) the Mg II doublets were visually confirmed, but the 16 absorbers missed by Anand21 had profiles that were generally complex, broad or weak. This suggests that the Anand21 Mg II detection algorithm has a narrower detection window at the cost of losing some real absorbers. Conversely, the Z&M Mg II detection algorithm could conceivably contain more spurious absorbers while likely managing to detect a higher percentage of the real absorbers.

5.3 Statistical analysis

Here I present the statistical analysis of the BR. We have taken a similar approach to the GA analysis that was presented in Chapter 4 of this thesis. Given the nature of

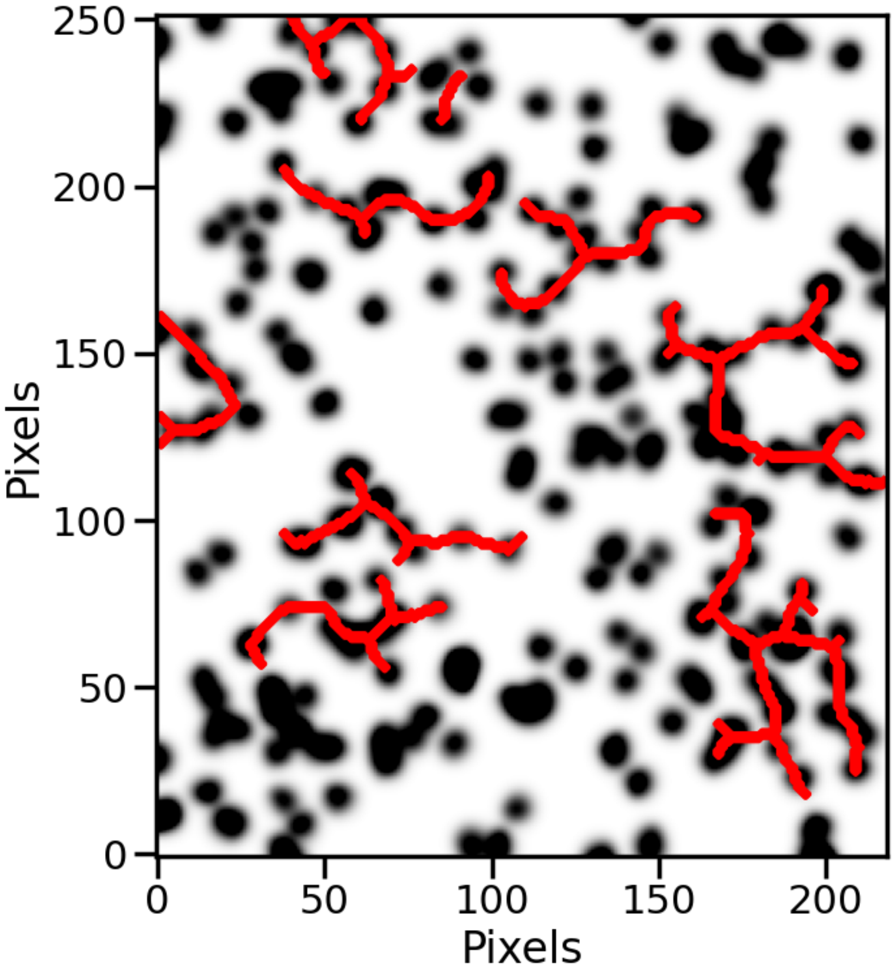


Figure 5.4: The FilFinder algorithm applied to the tangent-plane distribution of Mg II absorbers in the closest, non-overlapping redshift slice to the BR field — i.e., the field centred at $z = 0.682 \pm 0.060$ and corresponding to the usual BR field-of-view. Axes are labelled in pixels, where 1 pixel $= 4^2$ Mpc^2. S/N limits of: $4, 2$ and 4 were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively (details of S/N cuts are discussed in Section 5.3.1). The probes (quasars) responsible for the Mg II arising here have redshifts $z > 0.862$ so that they are the same probes responsible for the Mg II absorbers in the BR field centred on the usual redshift slice ($z = 0.802 \pm 0.060$). The figure shows that there are no filaments correlating to the BR, indicating that the BR is not a result of artefacts in the probes.

the discovery, the analysis is necessarily post-hoc. However, given the previous work of the GA that was then published in Lopez22, we are able to follow the guidelines set there, consequently alleviating many of the problems associated with post-hoc analysis.

This section is divided into the following subsections. (5.3.1) We assess the BR from a 'first look' perspective by using a heuristic process of stepping through redshift slices, applying the Single-Linkage Hierarchical Clustering (SLHC) algorithm, and determining the optimum redshift for the BR. (5.3.2) We use the Convex Hull of Member Spheres (CHMS) and the Minimal Spanning Tree (MST) significance calculations for assessing the significance of the BR. We apply these two methods of significance calculations to four sets of BR absorber-member estimates: the SLHC groups; the visually-identified BR absorbers (both including and excluding the inner absorbers); and the FilFinder-identified absorbers. (5.3.3) The 2D FilFinder algorithm is applied to the pixel image containing the BR to objectively identify filaments in the field. (5.3.4) Finally, we apply the 2D Cuzick and Edwards test to the BR field to determine the significance of clustering in the field (not the candidate structure itself).

5.3.1 Single-Linkage Hierarchical Clustering algorithm

The Single-Linkage Hierarchical Clustering (SLHC) algorithm is equivalent to a Minimal Spanning Tree (MST in a generic sense, not to be confused with the MST *significance* calculation in Section 5.3.2) when separated at a specified linkage scale. Our particular application of the algorithm was first described in Clowes et al. (2012), in combination with the Convex Hull of Member Spheres (CHMS) algorithm which assesses the significance of a specified structure. The SLHC / CHMS method has been used to identify and assess candidate LSSs in both quasars and Mg II absorbers (Clowes et al., 2012, 2013; Lopez et al., 2022).

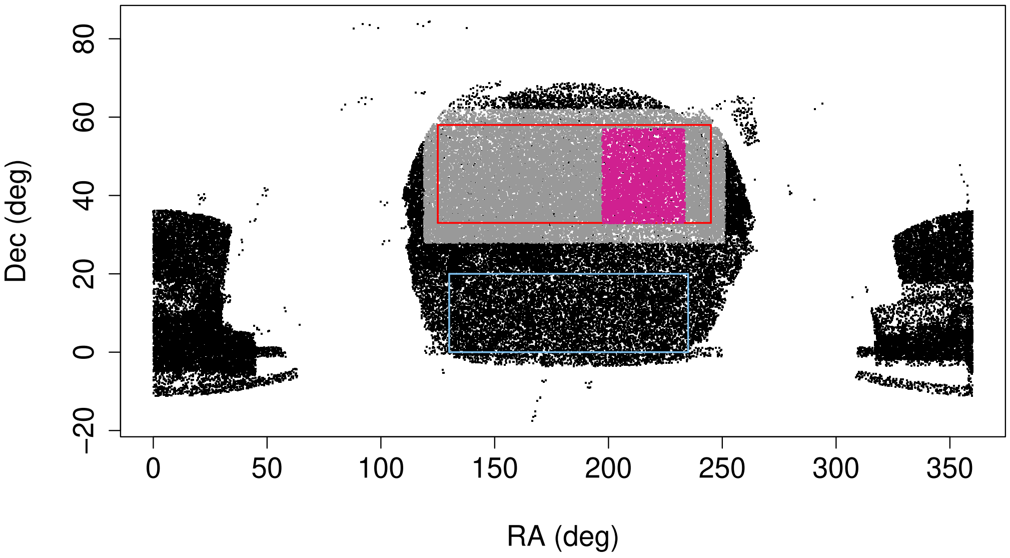


Figure 5.5: The SDSS DR16Q quasar positions. The grey points are the input quasars for the control field and the pink points are the quasars that correspond to the Mg II absorbers in the field of interest. The control fields are outlined by lines in red (overdense region, version-1) and blue (underdense region, version-2).

CHAPTER 5

The inherent difficulties of the Mg II method have been discussed throughout this thesis, especially in relation to applying statistics on an intrinsically inhomogeneous dataset. The quasars act as probes of the intervening matter, so the varying availability of background quasars leads to an incomplete image or map of the intervening matter, in this case, the Mg II absorbers. The inhomogeneity of the quasars, both intrinsic and of the survey, contribute in a complicated way to the inhomogeneity of the Mg II absorbers. Of course, with large data and large survey areas, approximations can be made. Before applying the SLHC algorithm we define the field-of-view (FOV) for the analysis presented here.

First, the BR appears in the north of the usual GA field, so we re-centre the Mg II images accordingly. Second, the field containing the BR is close to the northern SDSS footprint border as well as a southern border arising from much lower quasar coverage. We choose to shrink slightly the typical size of a Mg II image to avoid these areas. Third, given the generally patchy DR16Q quasars, and thus the corresponding Mg II absorbers in the Anand21 databases, we apply signal-to-noise (S/N) limits to the Mg II absorption lines and the continuum. We calculate the S/N of the Mg II lines by:

$$S/N = W_r/W_{err} \tag{5.1}$$

where Wr is the equivalent width of the Mg II line, and W_{err} is the corresponding error in the Mg II line. For the continuum, Anand21 provide the median quasar S/N. Applying S/N limits has the effect of removing spurious and potentially false positive absorbers as well as generally reducing the patchiness in the data. Setting additional[1] limits to the magnitude of the quasars would also have a similar, desirable effect. However, not all quasars will have the same integration time, so faint quasars could have good S/N due to long exposures. Following the example by Z&M, we apply a S/N limit of 4 and 2 for the $\lambda\lambda$ 2796, 2803 lines respectively. Since the

---

[1] A base-line magnitude limit of $i <= 20.5$ is applied to the DR16Q quasars as part of the read-in process since the completeness declines steeply for fainter quasars — see Section 3.2.

S/N of the quasar continuum necessarily has equal or higher S/N than the $\lambda2796$ line we apply a S/N limit to the quasar continuum of 6. Applying the condition of S/N ≥ 6 to the quasar continuum could be too restrictive, and later we will apply a less conservative condition to the continuum for comparison.

The CHMS significance is calculated by the rate of occurrence of volumes smaller than the CHMS volume of the structure by randomly distributing the absorbers belonging to the structure at a density equal to the control field density — the simulations are repeated 1000 times. Previously (for the GA work), the control field was chosen to be that of the field being assessed. However, this would mean that a small (but not insignificant) fraction of the field absorbers are those belonging to the structures of interest, e.g., the GA and BR in the GA/BR FOV. In addition, the FOV we have defined containing the BR is small, so small-scale inhomogeneities have a much larger effect on the average density. Conversely, choosing a control field that is too large will lead to problems involving the large-scale inhomogeneities of the SDSS survey. Subsequently, we have designed two versions of control fields accounting for the northern portion of the SDSS footprint (overdense region) and the southern portion of the SDSS footprint (underdense). Each version can be chosen depending on the location of the field of interest. So here, we will be using version-1 of the control field — Figure 5.5.

The SLHC is equivalent to an MST when separated at a specified linkage scale, thus the choice of linkage scale will determine the maximum distance between points that would be considered 'joined', or a candidate structure. The term 'structure' [2] here is not be confused with a gravitationally-bound system, but is instead referring to a grouping of more than 10 (a specified minimum) members (in this case, Mg II

---

[2] For a working definition, in investigating LSS, we often consider a candidate structure to be a set of N connected tracers, the containing volume of which is a $n\sigma$ significance of departure from the containing volume expected for a uniform, random distribution. We might choose to consider further only those candidates for which the amplitude $n\sigma$ exceeds some threshold. 'Connected' and 'containing' volume will often be determined *algorithmically*; for both, there is an implicit assumption of a uniform host survey, which might be approximately true only in restricted areas.

absorbers) that have an MST with distances smaller than the linkage scale. LSSs are not expected to be gravitationally bound, as, indeed, superclusters and great walls are not expected to be gravitationally bound (see Section 1.1). Similar usages and definitions of 'structure' are common in LSS studies (Pomarède et al., 2020; Balázs et al., 2015; Horváth et al., 2014; Gott et al., 2005). The candidate structures are then assessed by the CHMS to determine whether their volumes are statistically significant.

In Section 4.3.2 of the previous chapter, the effects of varying the linkage scale were discussed, and we decided to scale the linkage scale according to the Mg II number density of the control field and the number density of the GA field — the GA field being the candidate field from which to scale all other fields, i.e.,

$$s = (\rho_0/\rho)^{1/3} s_0$$

where s and s_0 are the linkage scale for the control field and the GA field respectively, and similarly ρ and ρ_0 are the densities of the corresponding fields. The linkage scale is to be taken as a guideline; if the linkage scale is too small then potentially interesting candidate structures will be missed, and if the linkage scale is too high then too many points will be grouped as one seemingly coherent structure, but that would of course reflect in the CHMS significance. We find that the linkage scale set for the GA field (the field which we now use as the base-line) was an appropriate choice for the specific field density. However, we saw that even with the chosen linkage scale the SLHC algorithm identified the GA as two, individual, overlapping candidate structures. We reason that, when using a smaller linkage scale, candidate structures that are overlapping or adjacent could reasonably belong to the same structure. It is important to recognise that multiple candidate structures overlapping and adjacent to each other will still need to be assessed with the CHMS or MST significance, which would then objectively determine whether their agglomeration is

z	s (Mpc)	N	$N_{3.5\sigma}$	σ_{max} (σ)	n_{max}
0.682	79.3	4	0	3.3	17
0.742	79.3	5	1	3.5	16
0.802	79.5	8	1	4.5	28
0.862	79.6	6	2	4.7	42
0.922	81.1	4	1	4.1	26

Table 5.1: Results from the SLHC / CHMS on five, overlapping redshift slices to determine the optimum redshift slice for the BR signal. The FOV of each redshift slice corresponds to the small, pink area seen in Figure 5.5. S/N limits (defined in Equation 5.1) of $4, 2$ and 6 were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively. All redshift slices have a thickness of $\Delta z = \pm 0.060$. The columns from left to right are: the central redshift of the field being assessed (z); the linkage scale used for the field being assessed, calculated as $s = (\rho_0/\rho)^{1/3} s_0$ (see the main text); the number of candidate structures identified in the field (N); the number of candidate structures identified in the field with a CHMS significance equal to or exceeding 3.5σ ($N_{3.5\sigma}$); the maximum CHMS significance calculated from the candidate structures (σ_{max}); the Mg II absorber membership identified from the largest (by membership) candidate structure (n_{max}).

statistically significant (consider if the whole field was joined as one structure then this would of course not be statistically significant).

As with the GA analysis, we step through overlapping redshift slices and use the SLHC / CHMS algorithms to determine the redshift of the peak signal of the BR. Given the much larger Mg II database from Anand21, even after S/N limits are applied, it is expected that the field density containing the BR is much higher than the GA field in the Z&M databases, and therefore the calculated linkage scale will be correspondingly lower. The five redshift slices assessed are centred at: 0.682, 0.742, 0.802, 0.862 and 0.922, each with a redshift thickness of $\Delta z = 0.060$; the results are shown in Table 5.1.

The BR that was originally identified visually appears almost fully (in a partly-open ring) in only the central redshift slice $z = 0.802$ indicating that this is the optimum redshift slice for the BR, as it was for the GA. (Note, the GA is also

z	s (Mpc)	N	$N_{3.5\sigma}$	σ_{max} (σ)	n_{max}
0.682	76.7	6	0	3.2	16
0.742	76.2	7	0	3.3	19
0.802	75.8	10	1	4.1	30
0.862	75.7	7	1	3.7	27
0.922	77.1	6	0	3.4	15

Table 5.2: As with Table 5.1. Slightly relaxed S/N limits of $4, 2, 4$ were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively. All redshift slices have a thickness of $\Delta z = \pm 0.060$. The columns from left to right are: the central redshift of the field being assessed (z); the linkage scale used for the field being assessed, calculated as $s = (\rho_0/\rho)^{1/3} s_0$ (see the main text); the number of candidate structures identified in the field (N); the number of candidate structures identified in the field with a CHMS significance equal to or exceeding 3.5σ ($N_{3.5\sigma}$); the maximum CHMS significance calculated from the candidate structures (σ_{max}); the Mg II absorber membership identified from the largest (by membership) candidate structure (n_{max}).

identified, and is statistically significant in the central redshift slice). The four structures contributing to the visually-identified BR are adjacent or overlapping on the sky, indicating that the separate structures plausibly belong to the same structure. The apparent splitting of a seemingly coherent structure was also seen with the GA, which was made up of two overlapping SLHC groups; the splitting of structures is an example of the limitations of applying the SLHC algorithm to an essentially incomplete dataset. Interestingly, the SLHC group corresponding to the bottom portion of the BR appears also to extend into the two higher redshift slices, $z = 0.862$ and $z = 0.922$, since there are similarly-shaped arcs (corresponding to the bottom portion of the BR) appearing at the same on-sky position in all three redshift slices ($z = 0.802, 0.862, 0.922$). (Note, the probes are not here restricted to be identical.)

The condition of S/N ≥ 6 for the quasar continuum could be too restrictive. Anand21 calculate the S/N over the whole quasar continuum, rather than the local continuum at the point of an absorber. Therefore, the whole quasar continuum could have lower S/N overall compared with local S/N at the position of an absorber. Accordingly, we slightly relax the S/N conditions to $4, 2, 4$ for the $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively and repeat the above described analysis (see Table 5.2).

We see again that the full BR is detected — this time as a full, closed ring, and with the inclusion of the inner filament — in the central redshift slice $z = 0.802 \pm 0.060$ (Figure 5.6). The BR is located north of the centre point, spanning ~ 10 degrees in the RA and Dec axes (x and y axes respectively), and is made up of a collection of five structures identified by the SLHC algorithm.

Note the very high similarity of the two tables of SLHC / CHMS results (Tables 5.1 and 5.2), where the only change is reducing the S/N limit of the quasar continuum from 6 to 4. For reference, the first set of results with the more restrictive S/N limits (≥ 6) will be referred to as results-1, and the second set of results with the less restrictive S/N limits (≥ 4) will be referred to as results-2. We deduce that setting the more restrictive S/N limits (≥ 6) to the quasar continuum, after already applying S/N limits to the Mg II lines, is of no great consequence, and possibly adds to the incompleteness of our data in a detrimental way. An overview of the results is as follows.

(1) In the two lowest redshift slices (corresponding to both results-1 and results-2), there is only one statistically-significant ($> 3.5\sigma$) structure detected in total. The statistically-significant structure belongs to the redshift slice centred at $z = 0.742 \pm 0.060$ (corresponding to results-1); it is a small group of absorbers located at the lower LHS of the Mg II image, and of no relevance to the BR. In addition, other than a possibility of a thin filament forming in the Mg II image in the redshift slice centred at $z = 0.742$, there is no strong indication or detection of the BR in the two lowest redshift slices.

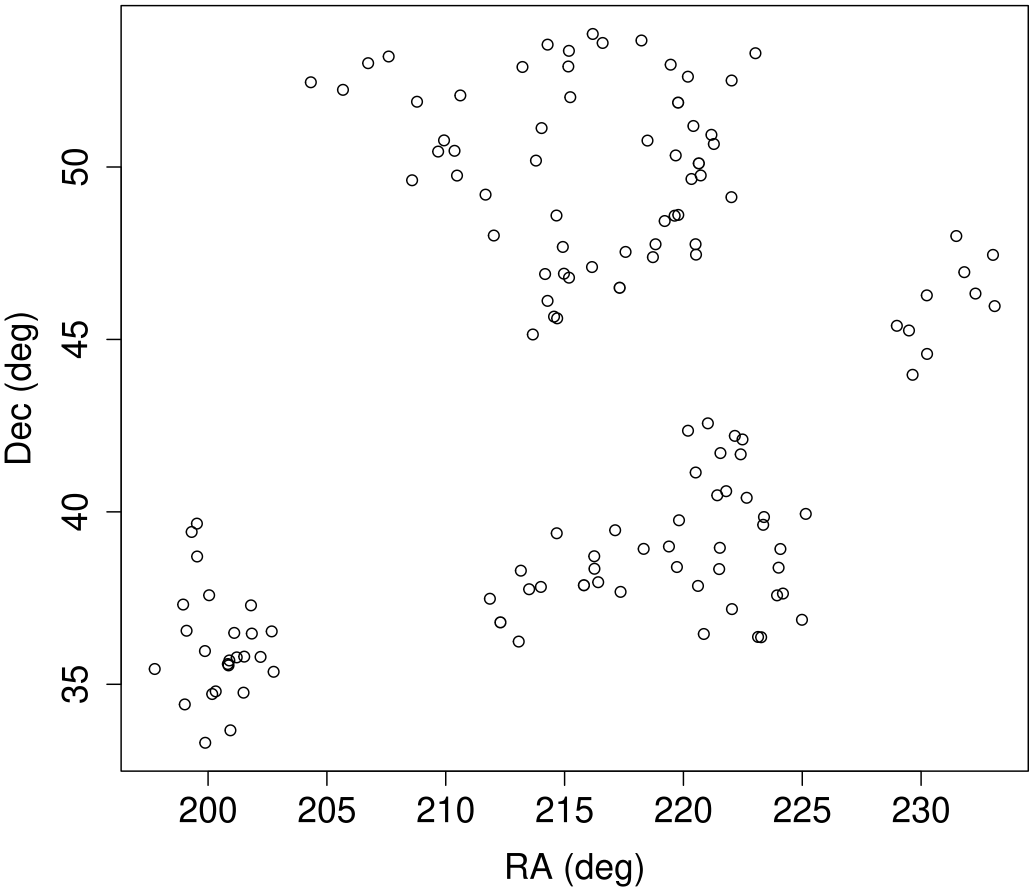


Figure 5.6: All the Mg II absorbers belonging to a candidate structure identified by the SLHC algorithm (corresponding to results-2) in the redshift slice centred at $z = 0.802 \pm 0.060$. The field-of-view seen here corresponds to the pink points in Figure 5.5. S/N limits of: $4, 2, 4$ were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively. The visually-identified BR can been seen to the north of the centre point, with an additional extension heading towards to the north-west direction, and the visually-identified inner filament is the central line cutting through the BR here. The BR spans ~ 10 degrees in RA and Dec coordinates. The large structure south of the BR belongs to the previously identified GA.

(2) In the central redshift slice, $z = 0.802$, there was one statistically-significant structure identified corresponding to the GA (corresponding to both results-1 and results-2, so a total of two statistically-significant structures). The BR is separated into four and five (respective to results-1 and results-2) individual, adjacent or overlapping, structures, that are statistically insignificant on their own (see Figure 5.7). We could again be seeing here the limitations of the Mg II method, as was seen with the GA, since the individual structures are overlapping or adjacent. From results-2, the visually-identified BR and inner filament absorber members are mostly detected (46 out of 59 absorbers, 78%) and form the full BR shape (see Figure 5.8).

(3) In the two highest redshift slices there are totals of 3 and 1 significant structures identified corresponding to results-1 and results-2, respectively. From results-1, the most significant structure in both redshift slices is an arc occurring at the same on-sky position as the lower portion of the BR (see the top-left panel in Figure 5.9). However, from results-2, we find the same arc that was detected in results-1, but across multiple candidate structures, highlighting again the nuances of applying the SLHC algorithm to an essentially incomplete dataset.

To clarify point (3) above, reducing the S/N limits in the quasar continuum from 6 to 4 initially appears inconsequential — i.e., their Mg II images appear on the whole unchanged, and there is only a 15% increase of absorbers in the whole field from results-1 to results-2. But, when the SLHC algorithm is applied, then this small increase in absorbers decreases the required (density-scaled) linkage scale. Importantly, the reduced linkage scale that is appropriate for the whole control field field, will not necessarily be the appropriate choice for a small FOV within the control field due to small scale inhomogeneities. In the case where the linkage scale is in fact too small for the small FOV, we will see previously-whole candidate structures broken into multiple candidate structures. We scale the linkage scale of the SLHC algorithm to the density for a general approach to the wide-varying

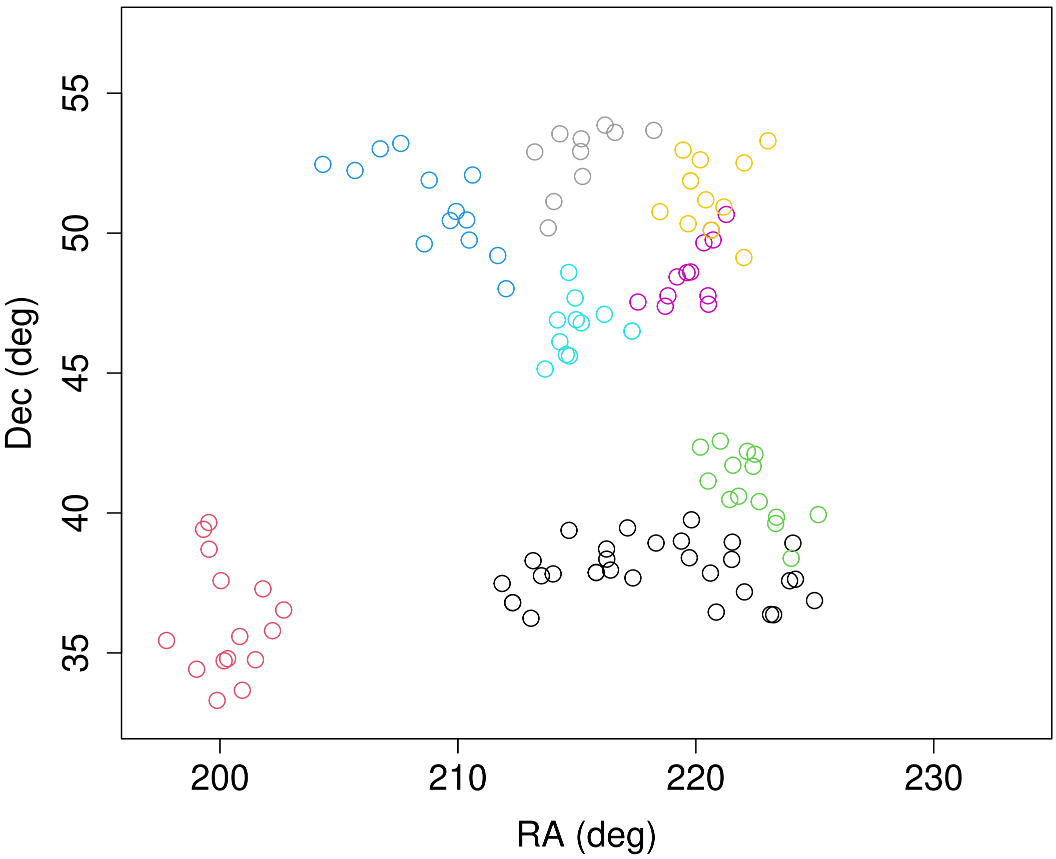


Figure 5.7: Eight of the 10 highest-membership candidate structures identified by the SLHC / CHMS algorithms (corresponding to results-2) in the redshift slice centred at $z = 0.802 \pm 0.060$. S/N limits of: $4, 2, 4$ were applied to the Mg II $\lambda\lambda$ $2796, 2803$ lines and quasar continuum, respectively. The colours represent the memberships which are ordered from high to low in the following way: black, red, green, blue, turquoise, pink, yellow, grey. The field-of-view here corresponds to the pink points in Figure 5.5. The BR and inner filament are detected, but separated into five structures, that can visually be seen adjacent to each other or overlapping. In this figure, only the black points, representing absorbers belonging to the GA, are statistically significant.

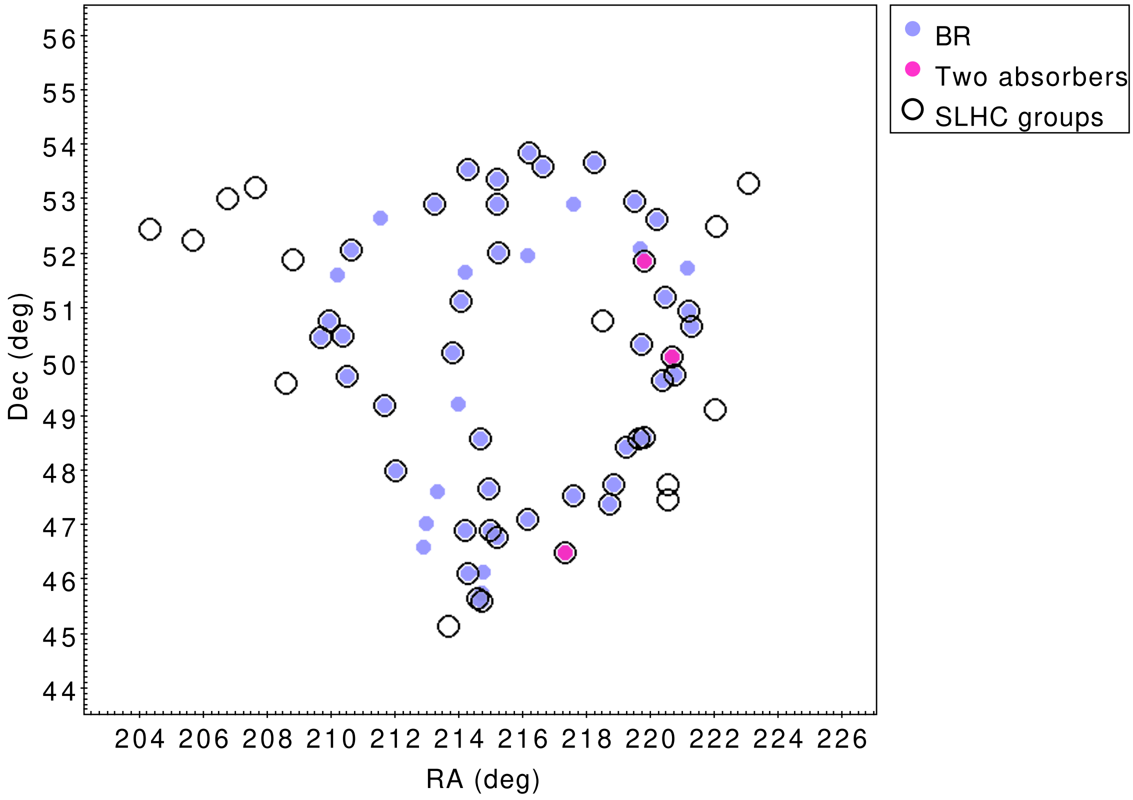


Figure 5.8: The visually-identified BR absorbers (lilac) and the SLHC-identified absorbers (corresponding to results-2) belonging to the 5 candidate structures at $z = 0.802 \pm 0.060$ (black circles) that correspond to the visually-identified BR. The pink points indicate the positions where there are two absorbers occurring in a single spectrum. There are 46 out of 59 absorbers in common to the SLHC-identified and visually-identified BR and inner filament absorbers. Of the 13 absorbers in the visually-identified BR and inner filament absorbers that were not connected by the SLHC algorithm, 9 of these occur at the most extreme edges of the redshift range, possibly explaining their exclusion.

densities in the dataset (mostly due to survey bias). We find that the absorbers belonging to (multiple) candidate structures corresponding to the arc in the redshift slice $z = 0.862 \pm 0.060$ from results-2, are very similar to the absorbers belonging to the (one) candidate structure corresponding to the arc in the redshift slice $z = 0.862 \pm 0.060$ from results-1. To further clarify this point, compare the top-left panel with the top-right panel in Figure 5.9, and see also the bottom panel of this figure.

There are 42 absorbers in the SLHC-identified arc in the redshift slice $z = 0.862 \pm 0.060$ corresponding to results-1 (the black points in the top-left panel of Figure 5.9), and a combined total of 40 absorbers in the three SLHC-identified candidate structures belonging to the arc in the redshift slice $z = 0.862 \pm 0.060$ corresponding to results-2 (the red, blue and turquoise points in the top-right panel of Figure 5.9). Of these absorbers, 35 are in common to both results-1 and results-2, confirming that the SLHC algorithm identified a high fraction (67% overlap) of the same absorbers between the two results. In fact, all 7 of the additional absorbers connected in the single candidate structure in results-1 all occur on the LHS of the arc, indicating that the increased linkage scale in results-1 (due to lower field density) is only *extending* the arc, and not responsible for the central components of the arc. Due to the above, we reason that individual, SLHC-identified candidate structures that are overlapping or adjacent could reasonably be connected as one structure if given a more complete dataset.

In conclusion, the BR appears most obvious in the original, central redshift slice, $z = 0.802 \pm 0.060$, both visually and by the presence of individual SLHC structures overlapping or adjacent on the sky that comprise the visually-identified BR. The strongly-detected arc corresponding to the lower portion of the BR located in a slightly higher redshift slice ($z = 0.862 \pm 0.060$) is also of particular interest in relation to the BR+GA system.

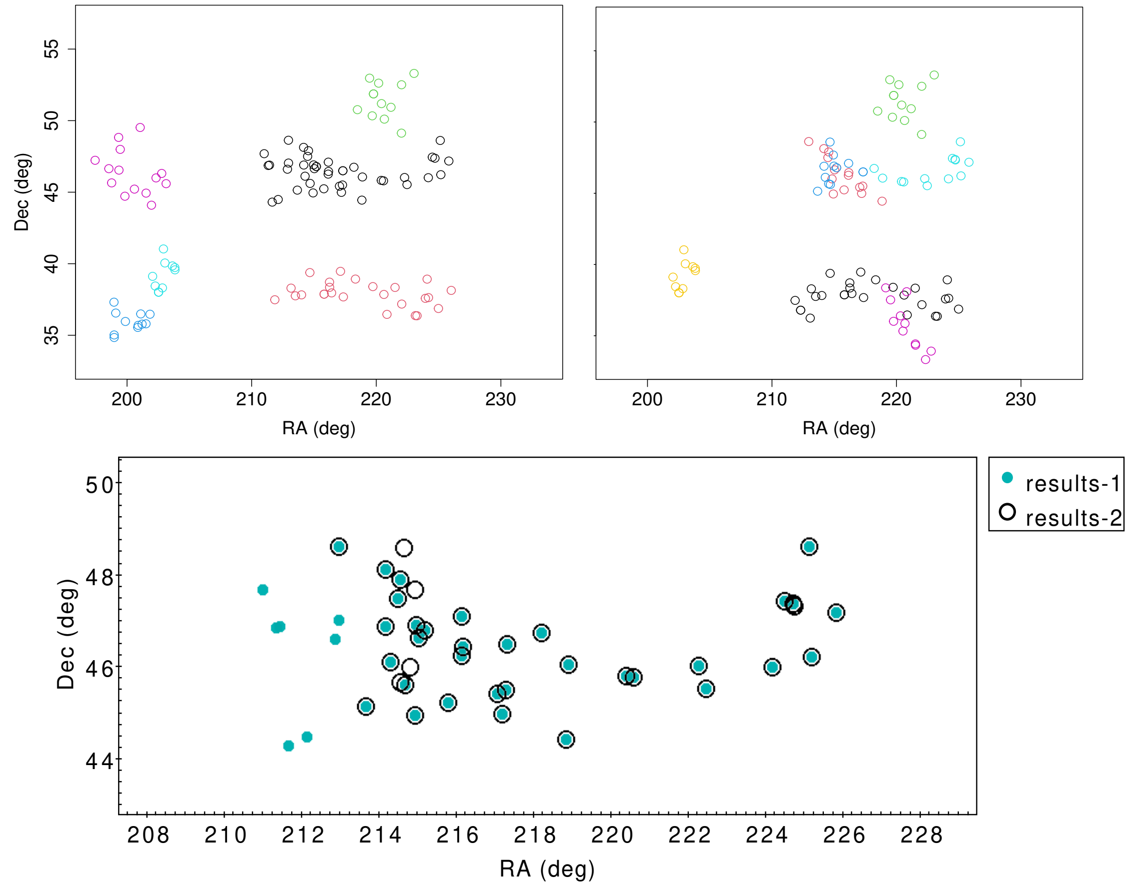


Figure 5.9: A comparison of the candidate structures identified in the redshift slice centred at $z = 0.862 \pm 0.060$ corresponding to results-1 (top-left panel) and results-2 (top-right panel). The different colours in the top panels represent the significances, and are ordered from high to low in the following way: black, red, green, blue, turquoise, pink.
(Top Left) The two most significant structures belong to the BR and GA respectively, both which are also statistically significant.
(Top Right) By comparing this figure to the figure in the left panel we can see that the absorbers identified by the 3 individual, overlapping or adjacent structures, coloured red, blue and turquoise, are clearly of the same absorbers identified as *one* full, statistically-significant structure in results-1, highlighting the complications of applying the SLHC algorithm to an essentially incomplete dataset. The absorbers belonging to the GA are statistically significant, shown in black.
(Bottom) Mg II absorbers corresponding to the arc in results-1 (turquoise), and Mg II absorbers corresponding to the arc in results-2 (black circles). There are 42 absorbers belonging to results-1, four of which appear as multiples per probe in two quasars, and 40 absorbers belonging to results-2, the same four of which appear as multiples per probe in two quasars. Both results connect many of the same absorbers, but in results-1, where the linkage scale is higher, there are additional absorbers connected by the SLHC algorithm all appearing on the LHS of the arc. The additional absorbers in results-2 occur in the middle of the arc, and there are much fewer additional absorbers than for results-1, suggesting that these absorbers were identified due to the lowered restriction on the S/N on the continuum.

5.3.2 Significance: CHMS and MST

The CHMS and MST significance calculations are applied to: the SLHC-identified absorbers; the visually-identified absorbers; and the FilFinder-identified absorbers. In addition, the Alpha Hull algorithm is applied to the visually-identified absorbers for an estimate of the volume, overdensity and significance using simple Poisson statistics.

SLHC-identified Mg II absorbers

To determine the significance of the BR in its entirety, we take the SLHC-identified absorbers (corresponding to results-2) that make up the BR (i.e., the BR in Figure 5.6), and apply the CHMS algorithm. Remember that the CHMS has the ability to assess the significance of a structure by comparing the observed convex-hull volume with the volumes that would be expected for a set of random distributions of those same absorbers at the control density of absorbers for the same redshift interval. Based on the definition of the CHMS volume and significance calculation, the algorithm is optimal when applied to clumpy structures, with no obvious gaps, holes or curvature that would lead to an overestimation of the volume. Clearly, this is not the case with the BR, having a large volume mostly unoccupied by absorbers in its centre. Instead, the MST-significance calculation introduced by Pilipenko (2007) could be more appropriate, which uses the mean MST edge-length between neighbouring data points. Using both methods we then find that the SLHC-identified BR in its entirety has a CHMS significance of 3.6σ, and an MST significance of 4.7σ. Both tests indicate statistical significance, but we can see that CHMS has likely overestimated the BR volume leading to a much lower significance compared with the MST-significance test.

	# Mg II absorber members	CHMS signif. (σ)	MST signif. (σ)
BR-only	51	3.3	4.0
BR-all	62	5.2	4.1

Table 5.3: The CHMS and MST significances for the BR-all and BR-only absorbers. The CHMS significance is dependent on the volume and number of absorbers so, clearly, removing the inner absorber members and keeping the volume the same will reduce the CHMS significance. However, for the MST significance, the volume is not directly related, but instead it is related to the MST mean edge lengths. This is clearly shown in the results as the MST significance stays mostly the same, at $\sim 4\sigma$, and the CHMS significance decreases from 5.2σ to 3.3σ by excluding the absorbers enveloped by the BR.

Visually-identified Mg II absorbers

We then take the visually-selected absorbers of the BR, and everything within the BR, for which there are a total of 62 absorbers (hereafter BR-all, see Figure 5.2), and apply the CHMS to these absorbers only. The BR-all absorbers have a CHMS significance of 5.2σ. The significance calculated here is likely the upper limit for the BR, as the algorithm was applied to those absorbers that were visually selected, including the BR inner absorbers (the absorbers enveloped by the visually-identified BR annulus), so that we were applying the CHMS algorithm to a more globular structure. We can similarly repeat this work for the BR-only absorbers, as well as compare the MST significance with the CHMS significance. Clearly, removing the BR inner absorbers will reduce the significance calculations for the CHMS, since the volume will remain the same but the number of absorbers will be reduced. In contrast, for the MST significance, the mean MST edge length may not be greatly affected by removing the BR inner absorbers, as seen in Table 5.3. We find that on both occasions, using the BR-all and BR-only absorbers, the MST significance is roughly the same, at $\sim 4\sigma$. In contrast, for the CHMS calculation, the significance drops from 5.2σ to 3.3σ after removing the absorbers contained within the BR.

However, even after removing all of the BR inner absorbers the CHMS significance is still greater than 3.0σ, but not quite reaching 3.5σ (the usual standard we apply for comparing structures). Of course, the CHMS calculation on the BR-*only* absorbers gives a drastic under-representation of the true significance since we have forcefully removed absorbers contained within the BR while keeping the unique volume the same.

For comparison, we can estimate the volume of the BR using the 2D Alpha Hull algorithm. First, the Alpha Hull area of the BR is repeatedly calculated 500 times from the BR by drawing a cloud of points in a circle with radius equal to half the mean MST edge length around each of the Mg II absorbers in the BR. Then, the area is multiplied by the physical size of the redshift range of the BR absorber members. In this manner, we are calculating the volume of a somewhat cylindrical, tube shape. The benefits of this method versus the CHMS is that we can eliminate the central region of the BR where there are very few absorbers. The downside of this method is that, although the absorbers on the sky make up a ring shape, we later see that the 3D distribution of the absorbers is more of a coil shape, so we are again overestimating the volume of the BR (and underestimating the overdensity and significance of the structure). Nevertheless, using this method we obtain a volume of 21.8×10^6 Mpc^3, an overdensity of 0.75 and a significance of 4.0σ for the number of absorbers in this volume based on Poisson statistics. The significance calculated from the Alpha Hull here, although simple, agrees with the MST significances of the SLHC-identified absorbers and the visually-identified absorbers, which adds further confidence to the statistical assessment of this structure.

FilFinder-identified Mg II absorbers

Finally, we have seen the CHMS and MST algorithms applied to the SLHC-identified and visually-identified absorbers, so now we will apply both algorithms to the

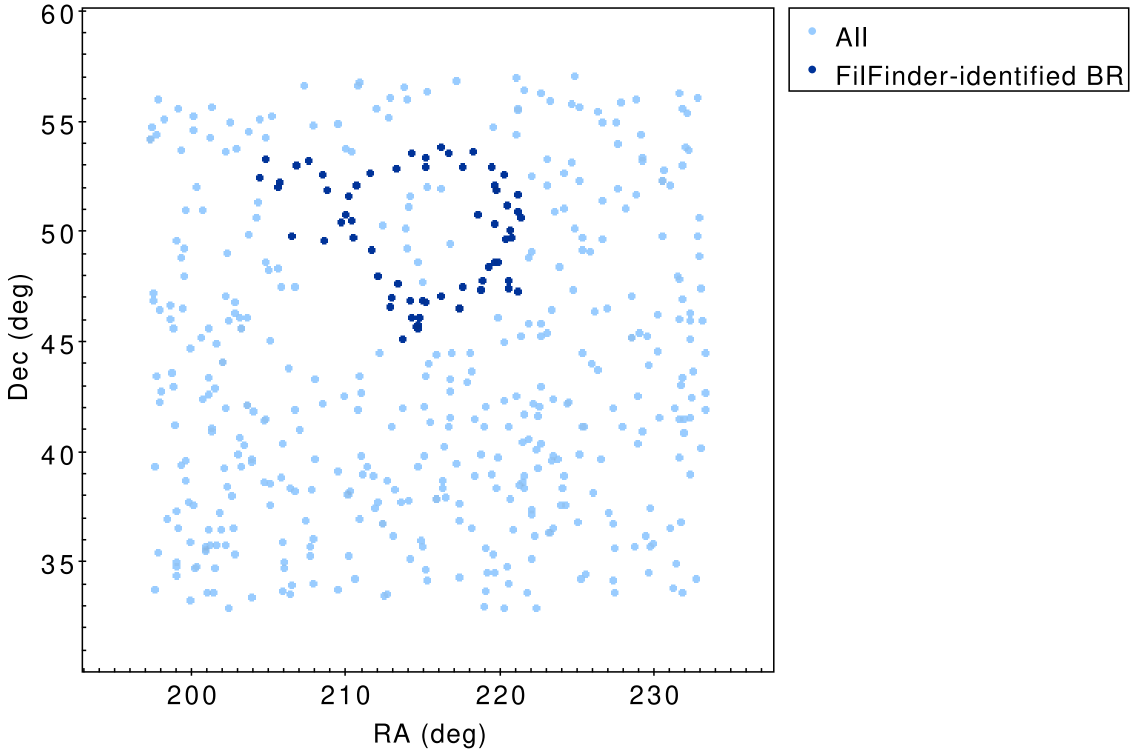


Figure 5.10: The Mg II absorbers belonging to the whole BR field, represented by light blue points, and the FilFinder-identified BR absorbers represented by dark blue points.

FilFinder-identified absorbers. The FilFinder work is discussed next, in Section 5.3.3, and we are referencing the bottom-right panel in Figure 5.11 for the work in this section.

The FilFinder algorithm is applied to 2D pixel images, so the physical absorber points are irrelevant to the algorithm. However, we can estimate the absorbers identified by FilFinder by referencing the bottom-right panel in Figure 5.11 — the absorber members connected to (touching) the FilFinder filament are considered FilFinder-identified absorbers. Our estimation of the absorbers connected by the FilFinder algorithm is shown in Figure 5.10. We can intuitively see that a volume containing all of the FilFinder-identified absorbers will include many large, empty sub-volumes — e.g., the centre of the BR and the volume between the 'extended filaments' connected to the BR. The BR volume was also overestimated with previous identifications of the BR (excluding the visually-identified BR-all absorbers),

	SLHC	Visual BR-all	Visual BR-only	FilFinder	Mean
CHMS (σ)	3.6	5.2	3.3	2.5	3.65 ± 1.13
MST (σ)	4.7	4.1	4.0	3.6	4.10 ± 0.45
				Total (σ)	3.88 ± 0.83

Table 5.4: The CHMS and MST significances of the BR calculated for each set of identified absorbers (SLHC, visual BR-all, visual BR-only and FilFinder). The separate cell called 'Total' is the total mean significance for all of the results in the table.

which can be confirmed with the comparison of: (i) the BR-all with the BR-only absorbers, and (ii) the CHMS with the MST significances. So, when the CHMS algorithm is applied to the FilFinder-identified absorbers, we calculate a 2.5σ significance. In contrast, using the MST-significance calculation, we estimate a 3.6σ significance for the FilFinder-identified absorbers. Generally, the MST significance here is in agreement with previous estimates of the BR significance using the MST calculation and the Alpha Hull Poisson statistics. So, again, the BR volume is considerably overestimated, this time using the FilFinder-identified absorbers, and the CHMS significance is therefore underestimated as a consequence, which can be confirmed with the use of the MST-significance test.

Summary of significances

We have shown our application of the CHMS and MST significance calculations to four sets of uniquely-identified BR absorber members. Table 5.4 summarises all of the above significance calculations. The results are: the total mean significance of the BR (all results from Table 5.4 considered) is $(3.88 \pm 0.83)\sigma$; the mean CHMS significance is $(3.65 \pm 1.13)\sigma$; and the mean MST significance is $(4.10 \pm 0.45)\sigma$. From these results we can see that the variation of the CHMS significance is a considerable fraction of the average, indicating its results are to be taken with caution. On the other hand, the MST significance has much lower spread, indicating that for the purpose of analysing the BR, the MST-significance test may be more appropriate, given the difficulty of defining a volume around a ring-like structure without incorporating over-estimations of the volume (as is the case with the CHMS algorithm).

5.3.3 Filament identification algorithm

FilFinder is a filament identification algorithm created by Koch & Rosolowsky (2015). It uses mathematical morphology to identify filaments ranging in size, shape and brightness on a 2D pixel image. The algorithm was intended for use in small (10^1 — 10^3 pc), gaseous areas, such as star-formation regions and the interstellar medium (Mookerjea et al., 2023; Zhang et al., 2023; Meidt et al., 2023). So, applying it to cosmological LSS is new, and we have had to make adaptations in the parameter settings to suit the data. We use the following input parameter settings.

(1) Adaptive threshold — the expected width of a typical filament. We choose a value (in number of pixels) equivalent to a filament that is only one absorber wide. Since the Mg II absorbers in the Mg II images span a diameter of approximately 12 pixels, the adaptive threshold is set to 12.

(2) Smooth size — scale on which to smooth the data. Our data are already smoothed and flat-fielded, so it is possibly unnecessary to smooth beyond the size of a Mg II absorber: i.e., we shall implement a smoothing size that does not affect the already smoothed Mg II absorbers. Therefore, we set smooth size to a value of 12 pixels.

(3) Size threshold — the smallest area to be considered a filament. Using again the estimated value for the number of pixels across a single smoothed Mg II absorber, we set this value at $4 \times 12^2 = 576$ pixels. That is, the minimum area to be considered a filament is made up of 4 Mg II absorbers (with the absorbers imagined as squares rather than as circles, for simplicity).

We use the FilFinder package available in Python to objectively identify the filaments present in the BR field. For the above parameters we find that there are no visually-obvious consequences for relatively small changes (e.g., within a few percent of the chosen value). However, much larger changes in the parameter choices render noticeably different, and consequential, results. For this reason we experiment with changing the standard parameter values above to gain a better understanding of the FilFinder algorithm when applied to cosmological data. The FilFinder algorithm is applied multiple times, with different parameter settings, on the Mg II image containing the BR (BR image; see Figure 5.1), and the results are discussed below.

First, the adaptive threshold and smoothing size are related in a way that increasing one has an almost indistinguishable effect from decreasing the other, for reasonable values (e.g., $<$ 40 pixels, since larger values flag a warning within the FilFinder algorithm). More specifically, the most meaningful and interesting effects can be found when one focuses on the ratio between the values. This can be understood intuitively since smoothing size is the scale on which to smooth the filaments and adaptive threshold is the typical width of those filaments, the size of which is affected by the smoothing size. When the adaptive threshold and smoothing size ratio is $1:2$ there is an undesirable effect of creating ‘blurred’ masking borders over the filaments; this is when the borders of the mask around individual filaments overlap. However, the opposite is true when the adaptive threshold and smoothing size ratio is $2:1$, which has the effect of creating more concise borders around the filaments, but it also drastically reduces the number of filaments that can be detected. With this in mind, we choose to slightly increase the adaptive threshold from 12, a single smoothed Mg II absorber, to 18, 1.5 times a smoothed Mg II absorber, while keeping the smoothing size at 12 as usual, thus creating a ratio of $3:2$ for the adaptive threshold and smooth size.

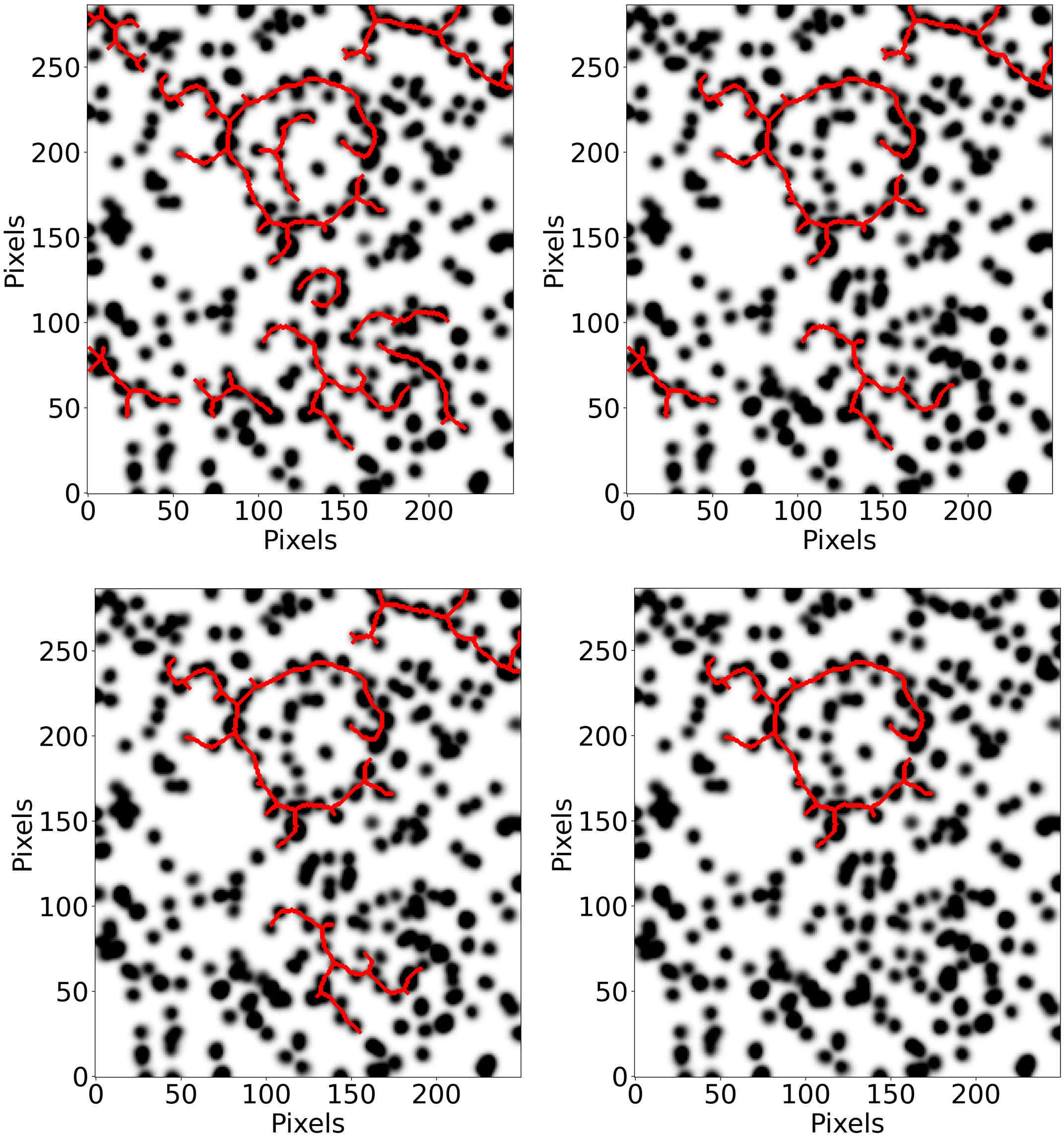


Figure 5.11: Applying the FilFinder algorithm to the standard BR field and increasing the size threshold incrementally to show the elimination process of small filaments. Axes are labelled in pixels, where 1 pixel = 4^2 Mpc^2. The standard parameters are set, where adaptive threshold is 18 pixels and smooth size is 12 pixels. (Top Left) Size threshold is the standard 576 pixels. (Top Right) Size threshold is 800 pixels. (Bottom Left) Size threshold is 1000 pixels. (Bottom Right) Size threshold is 2000 pixels — only the filament corresponding to the Big Ring remains.

Secondly, the size threshold simply determines the minimum size to be considered a filament. Increasing the size threshold incrementally and comparing results demonstrates the elimination process of small filaments. So now using the updated parameter values, where the adaptive threshold is set at 18 pixels and the smoothing size is kept at 12 pixels, we incrementally increase the size threshold and show the filaments that survive the elimination process (see Figure 5.11).

Immediately it becomes clear that the BR is the largest, and most dense, filament in the field. The BR filament is the only filament remaining after incrementally increasing the size threshold. In fact, the BR filament is not eliminated until the size threshold exceeds 4200 pixels. Using the elimination process with the FilFinder algorithm has given an impressive indication for the size and uniqueness of the BR compared with the rest of the field.

Next, we apply the same method of analysis to the SDSS DR16Q quasars in the same field and the same redshift slice as the Mg II absorbers containing the BR for comparison with an independent data source. We want to reduce the noise of the high number density of quasars so we apply an i-magnitude (i) limit of $i \leq 20.0$, such that only the intrinsically very bright quasars are included. The field size and redshift interval for producing the quasar image is the same as that for the Mg II absorbers from Figure 5.1: that is, the quasars are chosen to be those that lie in the *same field* as Mg II absorbers, not to be confused with the background quasars responsible for the Mg II absorbers. The quasar image is seen in Figure 5.12.

The FilFinder algorithm is applied to the quasar image in the same manner as described above with the Mg II absorbers, with the following parameter settings: adaptive threshold = 18, smoothing size = 12 and size threshold = 576 (see the top-left panel in Figure 5.13). Given the much higher density of the field quasars compared with the Mg II absorbers, even after i magnitude cuts have been made, we can see that there are generally more filaments identified. Incrementally increasing

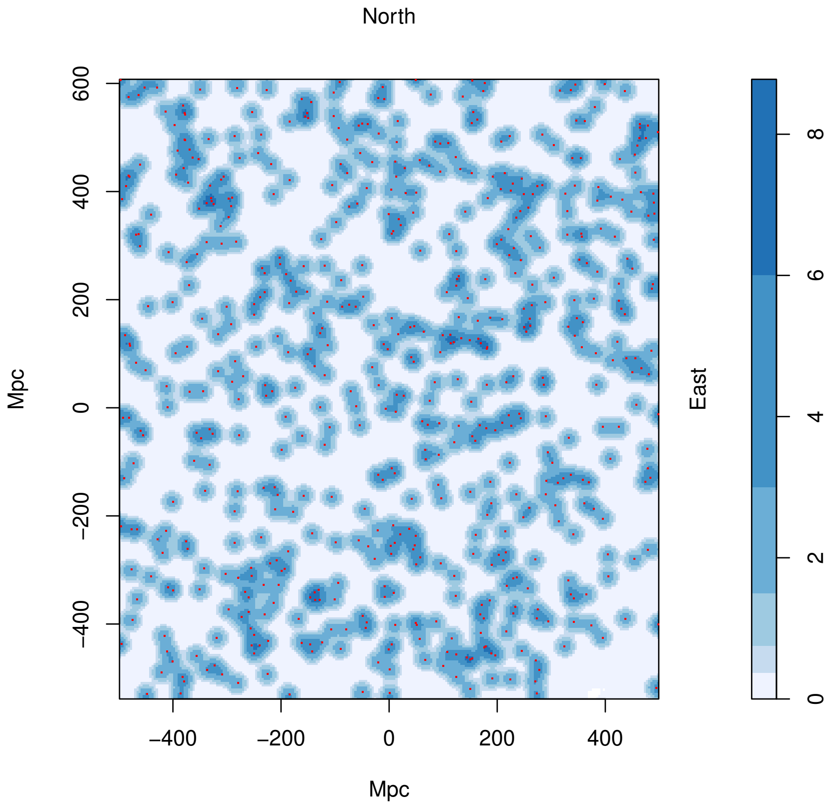


Figure 5.12: The tangent-plane distribution of quasars in the redshift slice $z = 0.802 \pm 0.060$. The blue contours, increasing by a factor of two, represent the density distribution of the field quasars which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc. Magnitude limits have been applied to the quasars such that $i \leq 20.0$. The field-of-view corresponds to the small, pink area seen in Figure 5.5. There is no strikingly obvious structure by eye, but when the FilFinder algorithm is applied to this field it detects a filamentary ring-like shape that almost coincides with the BR.

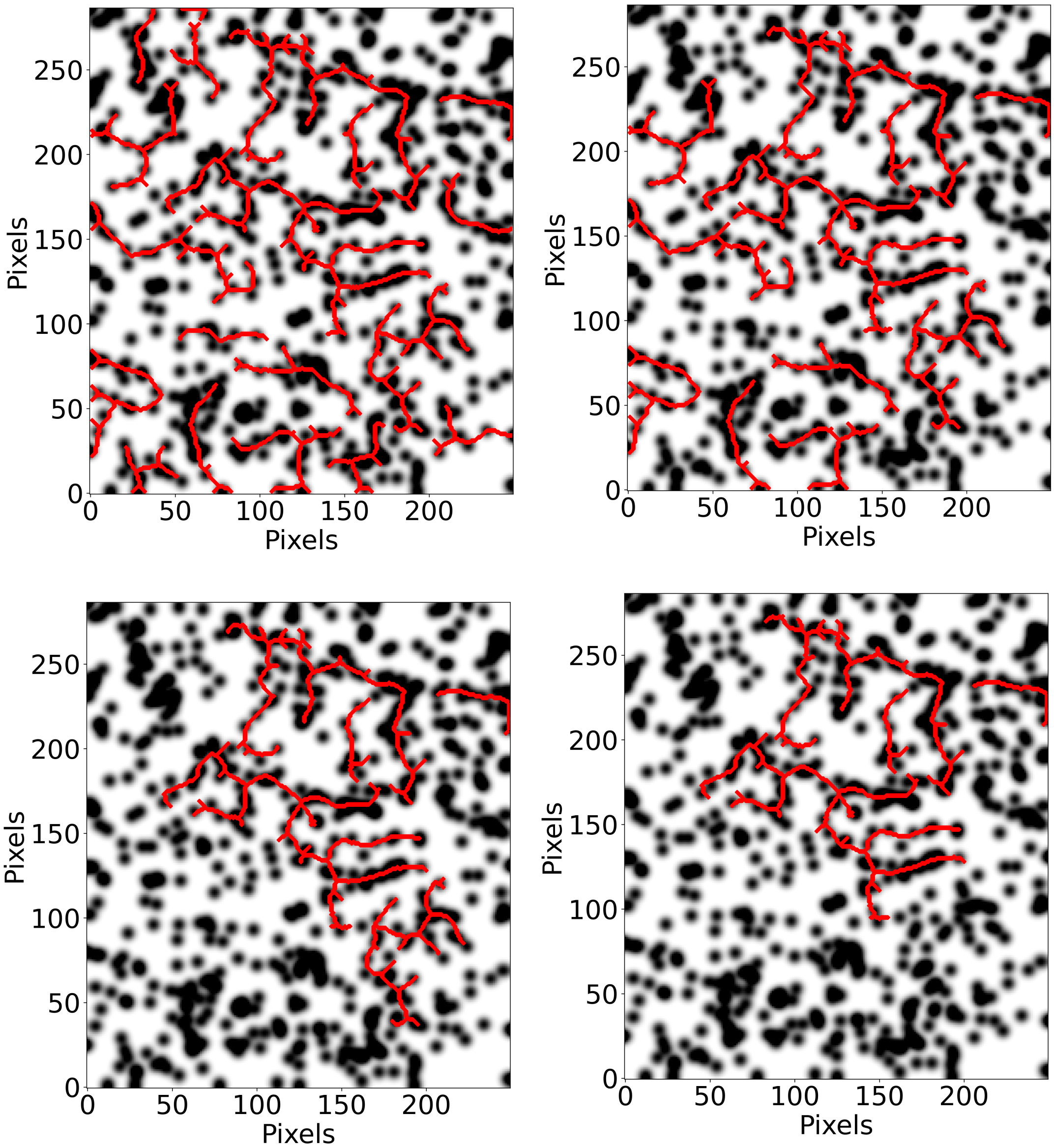


Figure 5.13: Applying the FilFinder algorithm to the *field* quasars in the standard BR field and increasing the size threshold incrementally to show the elimination process of small filaments. Axes are labelled in pixels, where 1 pixel = 4^2 Mpc^2. The standard parameters are set, where adaptive threshold is 18 pixels and smooth size is 12 pixels. (Top Left) Size threshold is the standard 576 pixels. (Top Right) Size threshold is 1000 pixels. (Bottom Left) Size threshold is 2000 pixels. (Bottom Right) Size threshold is 4000 pixels — only the filament corresponding to a ring-like feature, slightly offset from the Big Ring, remains.

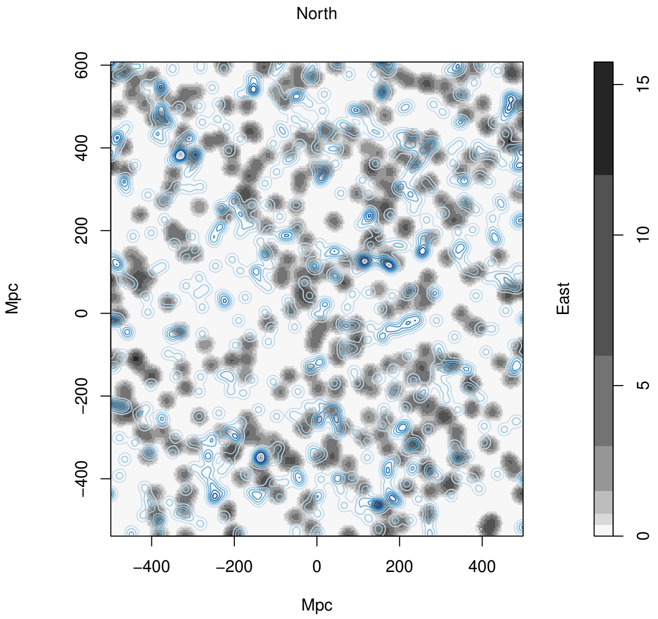


Figure 5.14: The tangent-plane distribution of Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ superimposed with the tangent-plane distribution of quasars in the same redshift slice (not to be confused with the background probes). The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes (quasars). The blue contours, increasing by a factor of two, represent the density distribution of the field quasars which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc. In the Mg II image, S/N limits of $4, 2$ and 4 were applied to the Mg II $\lambda\lambda\, 2796, 2803$ lines and quasar continuum, respectively. In the quasar image, magnitude limits were applied such that $i \leq 20.0$. Visually, it appears that the blue contours follow the grey contours, which may be suggesting an association between the two independent datasets.

the size threshold will remove the small filaments and leave only those able to survive the cuts — so just the large filaments of interest. Figure 5.13 shows the FilFinder results of the field quasars.

When the FilFinder algorithm is applied to the field quasars, and the size threshold is incrementally increased, we are left with only a ring-like filament that almost coincides with the BR (see Figure 5.15). This is particularly interesting as we are here using independent data that also yields a ring-like structure at the location of

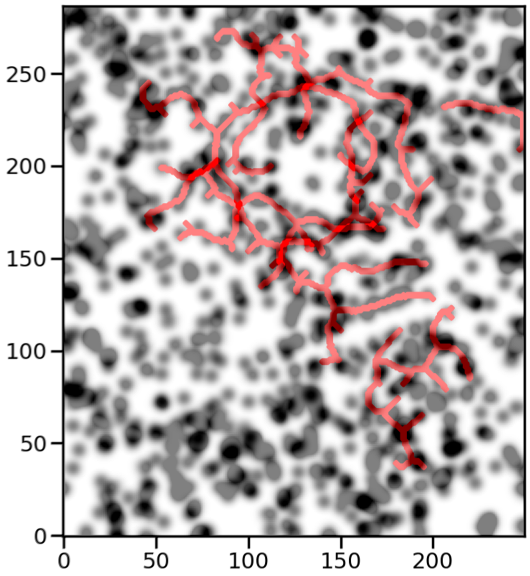


Figure 5.15: An overlay of the two largest FilFinder filaments identified in the Mg II absorbers (bottom-right panel of Figure 5.11) and the DR16Q quasars (bottom-right panel of Figure 5.13). The BR filament and the ring-like filament identified in the quasars are very closely aligned, with a slight left-sided offset; it seems likely that the two features are related.

the BR (further supporting the existence of the BR.) We can also show the independent corroboration of the quasars by superimposing the quasar image onto the Mg II image (see Figure 5.14). In Section 5.4.1 we further investigate the observational properties of both field quasars and DESI clusters in the BR field.

5.3.4 Cuzick and Edwards test

The Cuzick-Edwards (CE) test (Cuzick & Edwards, 1990) was created to assess the clustering of cases in an inhomogeneous population. It is a 2D, case-control statistical method that deals with variations in spatial populations. This method has shown to be useful for the particular data that we are working with: the Mg II are the cases to be assessed, which are affected by the inhomogeneous spatial distribution of the background probes (quasars). This method assesses the statistical significance of clustering by assessing the occurrence of cases within the k nearest neighbours, but it cannot, however, assess the significance of individual candidate structures. Therefore, it is a useful additional method to analyse the field containing the BR.

The CE test was first applied to cosmological (and to our knowledge, astrophysical) data when analysing the GA (Section 4.3.3); before this, the method had mainly been used in the context of medical research (e.g. for patterns in diseases across varying population sizes). We can now follow the steps presented in the published work of Lopez22 (and in the previous chapter of this thesis) and apply it to the BR. We are using the application *qnn.test* in the R package SMACPOD (French, 2020).

The power of the CE test is affected by the control-case ratio, and a ratio between 4 and 6 is found to be optimal (Cuzick & Edwards, 1990). For this reason, we continue (as with the GA) to use a control-case (i.e., probes to absorbers) ratio of $5:1$, which, for the BR field, requires removing $\sim 90\%$ of the probes. The $\sim 10\%$ remaining probes are randomly selected for each of the 200 runs of the *qnn.test*, which itself is computed for 499 simulations (the default setting).

We begin by assessing the BR field, and then we successively shrink the FOV to 'zoom in' on the BR, which is presumably the main feature contributing to any clustering patterns detected in the field. Unlike the GA, the BR is a full ring shape, so the successive zooming is limited in both axes. Therefore, we apply only two successive zooms to the BR, beyond which the BR would extend beyond the FOV. We then apply the CE test on four other fields that are unrelated and detached from the BR field, for comparison with the BR field. The four unrelated fields are chosen to be at the same redshift as the BR field ($z = 0.802 \pm 0.060$), the same field size (equivalent to the second 'zoom' of the BR field) and the same declination ($\delta = 50°$), so that the only change from the BR field to the unrelated fields is to move the right ascension (RA) centre point — the unrelated fields are then outside the BR field but within version-1 of the control area (see Figure 5.5). For the BR field, we repeat the CE test (with 200 runs and 499 simulations) 10 times to investigate how the random selection of probes (controls) affect the CE statistics.

The distribution of p-values calculated from applying the CE test to the BR field indicates tentative significant clustering ($p \leq 0.05$) for a range of chosen q (k) values. The median p-value dropped to a minimum of $p = 0.0238 \pm 0.0015$ over the 10 repetitions of the CE test at $q = 61$, corresponding to a significance of 2.0σ. However, in the four other unrelated fields there was no (with the exception of exactly two data points, see below) significant (significant is here defined as $p \leq 0.05$) clustering detected, suggesting that the clustering seen in the BR field is special (see Figure 5.16); this was similarly found with the GA previously.

In one of the unrelated fields there were two points, at $q = 2$ and $q = 8$, where the median p-value dropped below the $p = 0.05$ threshold, which is here considered tentative significant clustering. However, since it was at a much lower q value, the CE test was likely here detecting very small-scale clustering, and not related to the same type of clustering seen in the BR (and GA) field. Overall, the clustering seen in the BR field mimicked the clustering that was also seen in the GA field, but here at a tentative ($p \leq 0.05$) significance level; the clustering in the BR field was inconclusive at the $\sigma > 3.0$ significance level.

5.4 Observational properties

5.4.1 Corroboration with independent data

In Section 5.3.3, we saw that the bright field quasars have identified filaments similar to those of the Mg II absorbers. It was also shown visually and more generally that there were occurrences of the field quasars tracing the Mg II absorbers from mapping their density contours in $2D$ images in thin redshift slices. Previously it has been found that quasars follow the large-scale structure traced by galaxy clusters at low redshift ($z \leq 0.3$) and with a small offset ($\sim 3\ h^{-1}$Mpc) (Söchting et al., 2002,

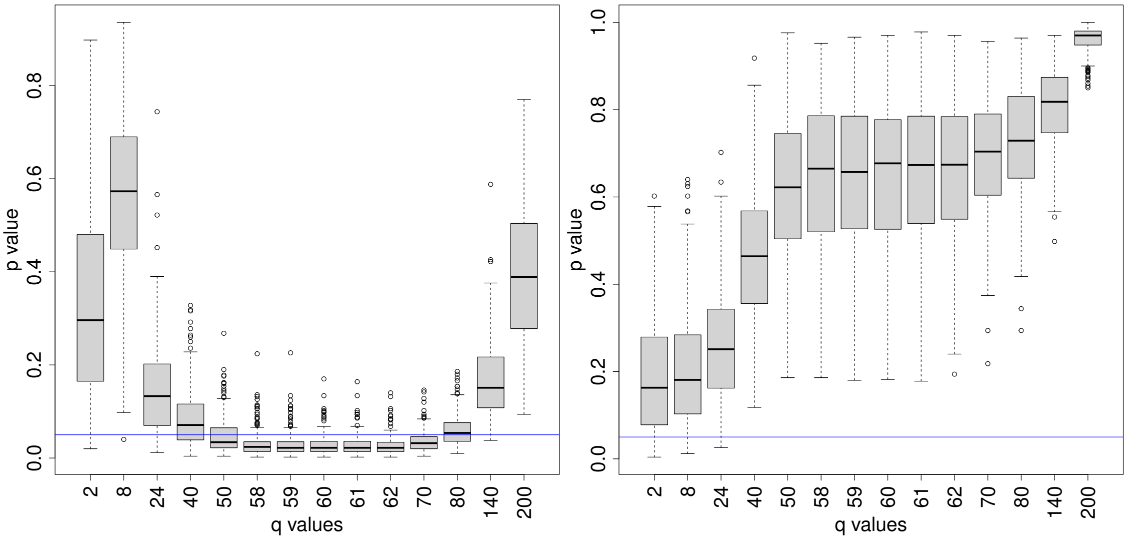


Figure 5.16: The Cuzick and Edwards (CE) test applied to two different fields. The figures show the distribution of p-values over 200 runs of 499 simulations as a function of chosen q (k). (Left) CE results corresponding to the second 'zoom' of the BR field. The BR field shows tentative significant clustering ($p \leq 0.05$) between $q = 50$ and $q = 70$, reaching a minimum of $p = 0.022$ between $q = 59$ and $q = 62$ corresponding to a significance of 2.0σ. Note, this is *one* of the figures produced from the 10 repeated tests. (Right) The CE results applied to a field unrelated to the BR field for comparison. The unrelated field shows no significant clustering on any scale.

2004). While we do not yet know whether this trend will hold at higher redshifts, we can speculate that perhaps the Mg II absorbers here, indicating galaxies and galaxy clusters, could also have an association with field quasars with some offset. Of course, the offset could be in any of the three coordinate axes, so it would be interesting to investigate in future work whether there is a measurable offset in the redshift axis for the quasar contours that appear to trace exactly the Mg II absorbers (e.g., along the Giant Arc, see Figure 4.9) and if this offset matches the offset we see in the BR and quasar 'ring' identified from FilFinder. At present, we can continue to use the contour maps from the quasars, and now also the DESI clusters (Zou et al., 2021), looking at a larger FOV, and investigate the visual association of the independent datasets with the MgII absorbers (see Figure 5.17). Note, statements of association between the Mg II and other independent data are based on visual interpretations (i.e., the assessment is qualitative) from the superimposed density-contour images; we leave the quantitative assessment for future work.

The visual investigation of quasars and DESI clusters with Mg II absorbers indicates that generally there is a plausible association of both independent datasets with the absorbers. The association can be seen most clearly when comparing the 'voids' in the data; where there are no absorbers, we find fewer quasars and DESI clusters.

A striking association is seen in the bottom-middle panel of Figure 5.17 where the green contours of the DESI clusters (with $R \geq 22.5$) seem to follow the same filamentary trajectory as the Mg II absorbers in many cases. Interestingly in the bottom-left panel of Figure 5.17, a similar trend is also seen, but here the DESI clusters with $R \leq 22.5$ appear to follow the Mg II absorbers that were 'missed' by the DESI clusters with $R \geq 22.5$. The two aforementioned DESI cluster figures (bottom-left and bottom-middle panels of Figure 5.17) have richness limits applied of $R \leq 22.5$ and $R \geq 22.5$, respectively, so the crossover between the two figures is

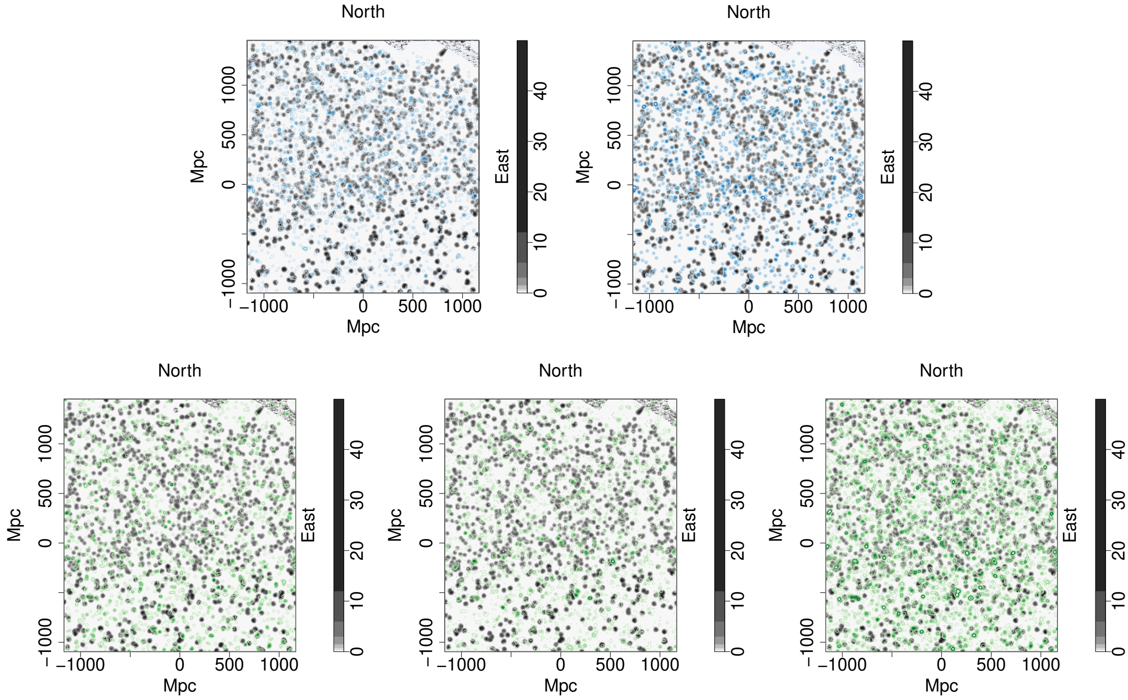


Figure 5.17: Density distribution of the Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ in a large field-of-view represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flat-fielded with respect to the background quasars. The additional blue and green contours (also smoothed with a Gaussian kernel of $\sigma = 11$ Mpc) in each of the five panels represent the superimposed density distribution of the field quasars (blue) and DESI clusters (green) that have been restricted according to their i magnitude and richness, respectively. Top panels: Note, the larger field-of-view of the figures crosses two of the SDSS borders where the quasar coverage drops sharply (see Figure 5.5). (Left) Field quasars restricted to $i \leq 20.0$ The blue contours appear to have a plausible association with the grey contours. The i magnitude limit is not very restrictive so the blue contours are visually noisy compared with the Mg II absorbers. Note, these quasars were found to have a similar-sized, ring-type FilFinder filament to the BR FilFinder filament that appeared slightly offset from the BR, so clearly it is the visual impression that is difficult to determine due to the noise. (Right) Field quasars restricted to $i \leq 19.5$. With the reduced quasars (compared with the left panel) the increased density contrasts in the blue contours help to highlight the trends with the Mg II absorbers. For example, the left, 'northern spur' coming off the BR has both a strong filament of Mg II absorbers and quasars. Generally, the dense clumps of Mg II absorbers appear to have associated quasars rather than the thin Mg II filaments.

Bottom panels: (Left) DESI clusters restricted to $R \leq 22.5$. The green contours generally (but sparsely) follow the grey Mg II contours, although the richness limit reduces the number of DESI clusters in the field. However, if one focusses on the voids in the Mg II absorbers, one can confirm that there appear to be fewer green DESI contours at these locations, e.g., the large void to the south-west of the BR. (Middle) DESI clusters restricted to $R \geq 22.5$. Similar to the left panel, it can be seen that the green contours generally follow the grey Mg II contours. In particular, the lower half of the BR has a strong filament of DESI clusters following the filamentary shape of the Mg II absorbers. (Right) DESI clusters of all richnesses ($0 < R \leq 300$). The green contours generally follow the grey Mg II contours quite well.

minimal. Here, by showing the association of low and high richness clusters with the Mg II absorbers, we might have found a way to investigate the Mg II absorbers and their physical environment.

The bottom-right panel of Figure 5.17 shows the DESI clusters of all richnesses plotted in green contours over the BR Mg II absorbers. Comparing all three DESI cluster figures (the bottom panels of Figure 5.17), it would appear that the DESI clusters of all richnesses have the strongest association with the Mg II absorbers. (Note that, with the GA previously, we did not find a strong association of the DESI clusters with the Mg II absorbers, which might worth investigating further.)

Finally, we note that the DESI clusters will have much larger redshift errors ($\Delta z \sim 0.04$) than the Mg II absorbers and quasars, so the structures in the DESI clusters will be blurred along the line of sight. Given that the visual association seems quite striking, then smaller redshift errors could lead to an even clearer association.

5.4.2 Viewing the BR from other angles

The BR and GA discoveries were made unexpectedly by eye with the method of intervening Mg II absorbers, and so are subject to observational bias — we have first detected the signal of a LSS by observing a curious shape and/or visual overdensity in the initial Mg II images that we later assess statistically. Specifically, we are only observing these LSS candidates from one line of sight (LOS), i.e., a 2D projection on the sky of the 3D matter distribution. Thus, from a different viewing angle, or LOS, the LSS candidates could look entirely different, or furthermore, there could be LSS candidates that have a seemingly average, or ‘uninteresting’ distribution of absorber members from our LOS that may be arcs, rings or interesting filamentary shapes from a different viewing angle. We investigate how different viewing angles change the perception of the BR, which first requires redefining the coordinate system on which the BR-field absorber members are projected, described here.

An initial, orthogonal, 3-vector coordinate system is defined such that u_0, v_0, w_0 are closely linked to $x_{prop}, y_{prop}, z_{prop}$. The initial normal vector (w_0) is defined as the proper coordinate to the mean $x_{prop}, y_{prop}, z_{prop}$ of the absorber members (which can be thought of as the original LOS). Then, the plane perpendicular to the normal is rotated such that u_0 points towards the most easterly absorber. Finally, all the absorber members are projected onto the new plane (see Figure 5.18). Hence, we call this method the 'project-plane method'. With the initial coordinate system defined, we can define any new normal vector (w) as some combination of the initial u_0, v_0, w_0.

In Figure 5.18, and other similar figures following, the colours of the absorbers indicate the redshift of each absorber, with the high-z absorbers represented by a darker shade, and the low-z absorbers represented by a lighter shade (see the key on the top left of the figures). The numbers associated with each absorber are a simple ID system, labelled 1 to 51 in ascending order from the most easterly absorber to the most westerly (i.e., in descending order of RA). The ID system remains the same throughout all subsequent rotations of the plane projections so that the numbers can be used to orientate oneself. In this manner we are viewing the BR from different viewing angles, similar to taking snapshots of the 3D structure from different LOSs. Note, the observational analysis presented on the BR here is based on the visually-identified BR absorber members, which are the blue points in Figure 5.2.

We then redefine the normal with combinations of the original, orthogonal, 3-vector system to give different projected planes. The newly-defined vectors u, v, w are similarly orthogonal. Figures 5.19 — 5.21 each have a different viewing angles of the BR, and the observational analysis of each plane projection follows.

In Figure 5.19 the new normal is set to u_0, which can be thought of as the side-on view of the BR (where the x-axis resembles the redshift axis). To emphasise this

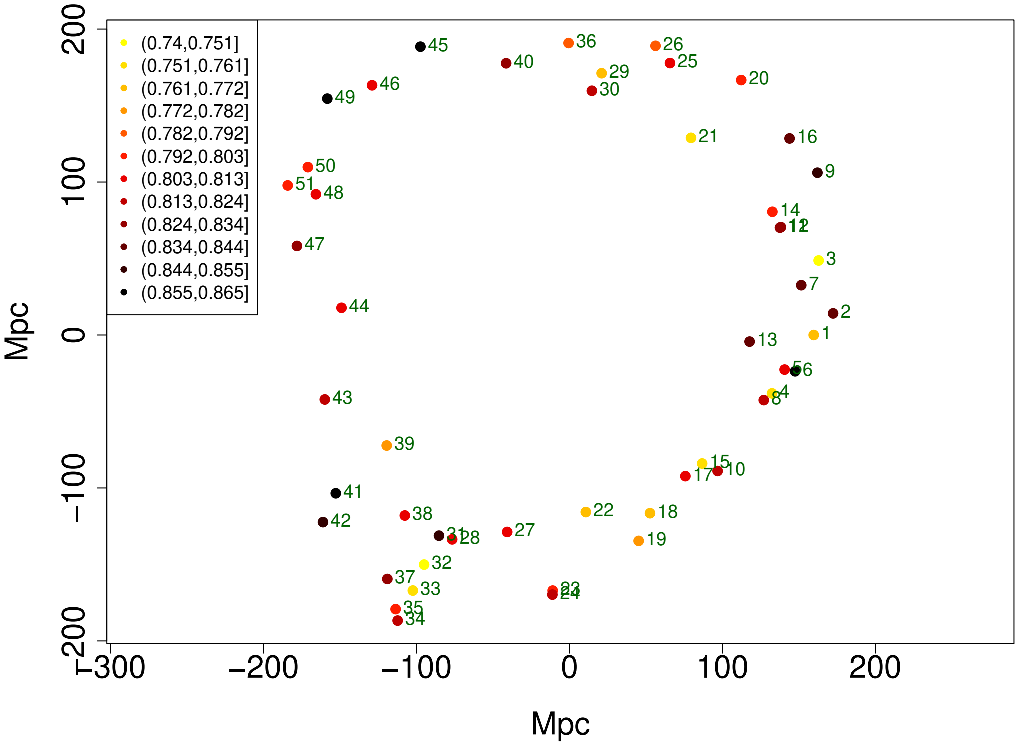


Figure 5.18: The visually-identified BR absorber members projected onto the plane perpendicular to the initial normal vector, w_0. The x-axis points towards the u_0 direction and the y-axis points towards the v_0 direction. The key in the top LHS of the figure indicates the redshifts of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number. By comparing this figure here with Figure 5.2 one will notice the slight rotation of the BR projection, indicating the small misalignment of the tangent-plane coordinate system with the new u_0, v_0, w_0 system.

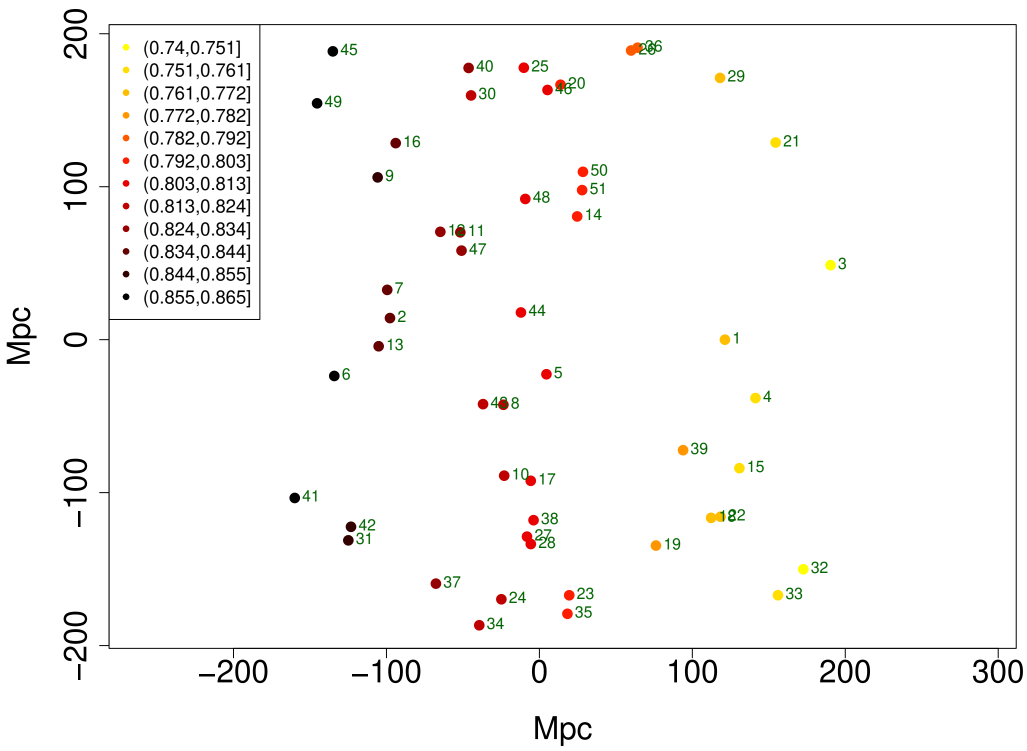


Figure 5.19: The BR absorber members projected onto the plane perpendicular to u_0. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top LHS of the figure) and the small numbers paired with each data point indicate their unique ID number. This plane-projection can be thought of as looking at the BR from the ‘side-on’ perspective (i.e., this x-axis resembles the redshift axis). The absorbers appear to be concentrated in three redshift bands. The lowest and highest redshift bands appear to form a curious backward ‘S’ shape.

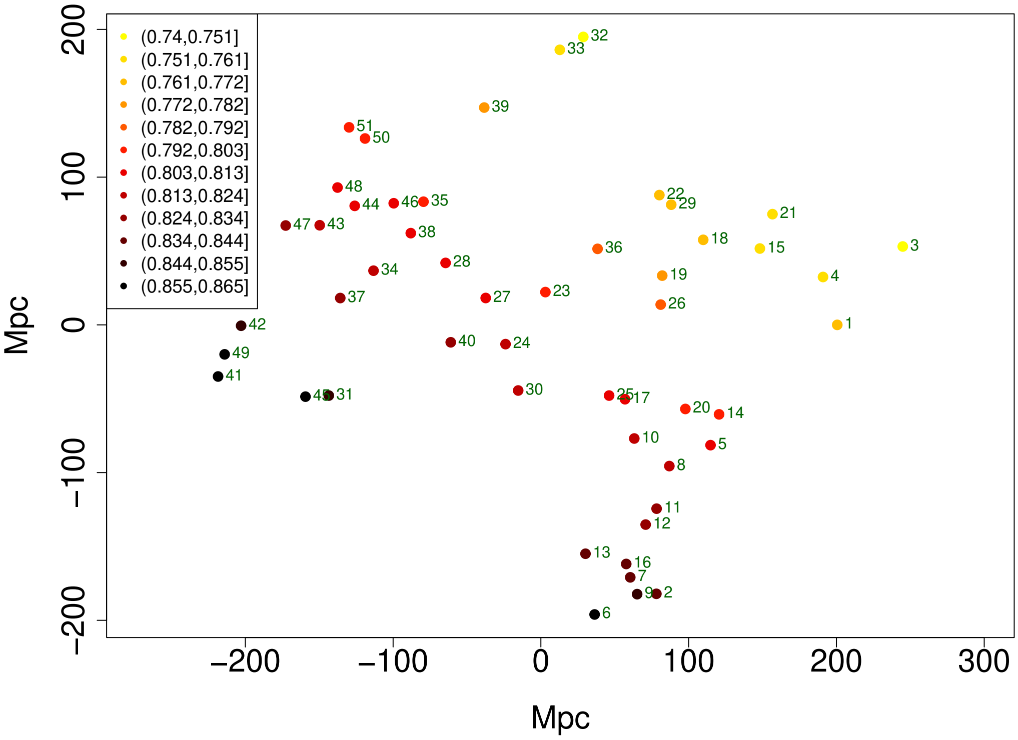


Figure 5.20: The BR absorber members projected onto the plane perpendicular to v_0. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top LHS of the figure) and the small numbers paired with each data point indicate their unique ID number. From this angle, the BR resembles an arc.

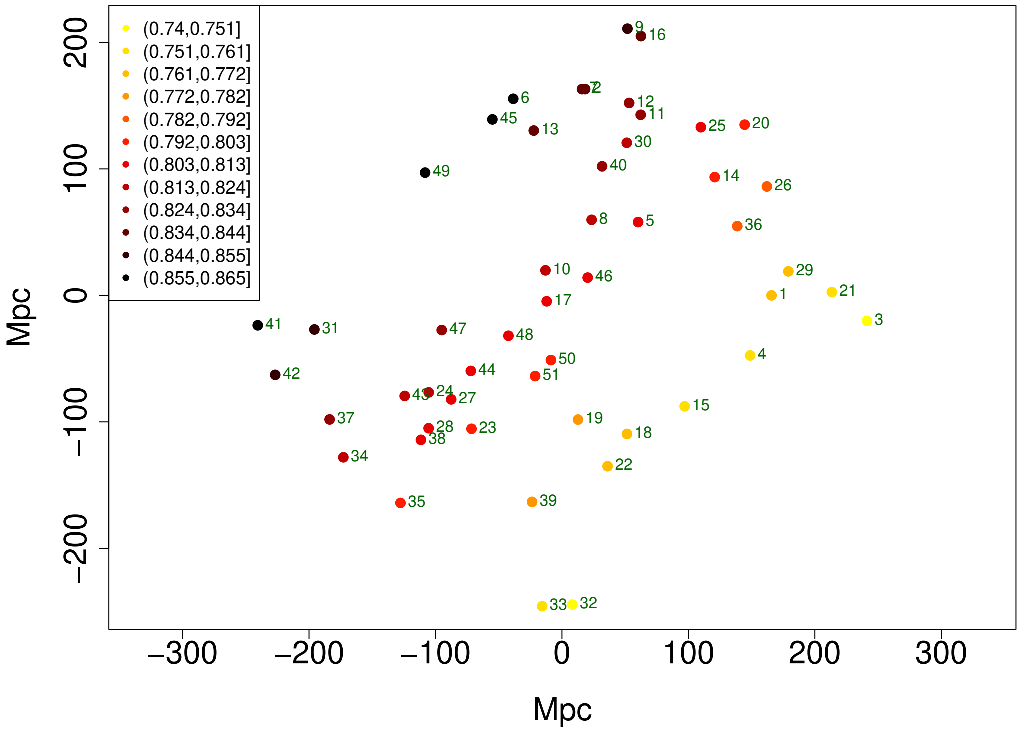


Figure 5.21: The BR absorber members projected onto the plane perpendicular to $u_0 - v_0$. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top LHS of the figure) and the small numbers paired with each data point indicate their unique ID number. From this angle, the BR appears to have a distinct ‘cork-screw’ shape.

point, the redshift colours are now banded vertically, indicating that the u-axis has essentially become the redshift axis. Of interest in this figure is the visual impression that the BR has three distinct redshift bands, and the absorbers in each band create a noticeable backwards 'S' shape (but more so in the nearest and farthest redshift bins which are the lightest and darkest coloured absorbers, respectively). It is not entirely clear what the backwards 'S' shape might indicate, or if it warrants further investigation. Initially, it was thought that the absorbers belonging to the three distinct redshift bands may also be associated particularly with some of the SLHC identified groups from Figure 5.7, but this is not the case. However, instead we find that almost all (one exception, starred in the list below) of the nearest redshift absorbers (those forming the backwards 'S' shape with the light-coloured absorbers) belong to one half of the BR closest to the RHS. This can be confirmed by cross-correlating the unique ID numbers in Figure 5.18 with Figure 5.19. The IDs of absorbers contained in the near-redshift band, i.e., the light-coloured backwards 'S' shape, are: $36, 26, 29, 21, 3, 1, 4, 39*, 15, 22, 18, 19, 32, 33$. ID 39 has been starred as the one absorber belonging in the near-redshift band that does not sit with the rest of the absorbers on one half of the BR. So the near-redshift absorbers form an arc of a circle closest to the RHS of the BR. Later we will see that this further implies a somewhat spiral or 'cork-screw' interpretation of the BR, that is aligned face-on with our LOS.

In Figure 5.20 the new normal is set to v_0, which can be thought of as viewing the BR from the bottom (south), looking up (north). Now we can see a thin, central arc which is formed from the central redshift absorbers, with clumps of absorbers at lower and higher redshifts either side of this arc. If the BR had been discovered from this particular angle, had we been located at a different position in the Universe, then we may have named this structure the 'Big Arc'. This highlights the observational bias mentioned earlier: our viewing angle only allows one LOS in which to discover interesting structures and filaments when using the Mg II image method. However, it also presents the opportunity to re-examine LSS candidates from multiple viewing angles.

In Figure 5.21 the new normal is set to $u_0 - v_0$, which can be thought of as looking at 45° through the BR towards the south-east direction. The BR now very clearly resembles that of a coil, and the central-redshift absorbers (red points) in this figure appear very thin and filamentary. The high-z and low-z absorbers form ovoid shapes connected by the central region, although the high-z absorbers appear more broken and sparse compared with the rest of the coil.

All three viewing angles, plus the original face-on view of the BR, indicate that the BR structure is a coil shape with a thin, flat, dense, central component. The flatness of the central component is particularly interesting given the recent work of Peebles (2023) on finding flat patterns in cosmic structure.

It is not yet clear how a structure like the BR (or the GA) will have formed; however, one possibility could be cosmic strings, which have become of topical interest in recent works (Ahmed et al., 2024; Cyr et al., 2024; Gouttenoire & Vitagliano, 2024; Ellis et al., 2023; Jiao et al., 2023; Peebles, 2023; Wang et al., 2023; Sanyal, 2022). Alternatively, we noted previously that the diameter of the BR was similar to the expected size of a BAO. However, the coil shape seems likely to be inconsistent with an origin in BAOs.

Finally, an ellipse can be fitted to the BR (Figure 5.22). The ellipse fitting calculates the semi-major axis (184 Mpc) and the semi-minor axis (162 Mpc) of the BR. Given that the signature of an individual BAO has a characteristic size of 150 Mpc, we again argue against the possibility of the BR occurring from a BAO. Given the ellipse calculations of the semi-major and semi-minor axes we can use Ramanujan's approximative perimeter to approximate the circumference of the

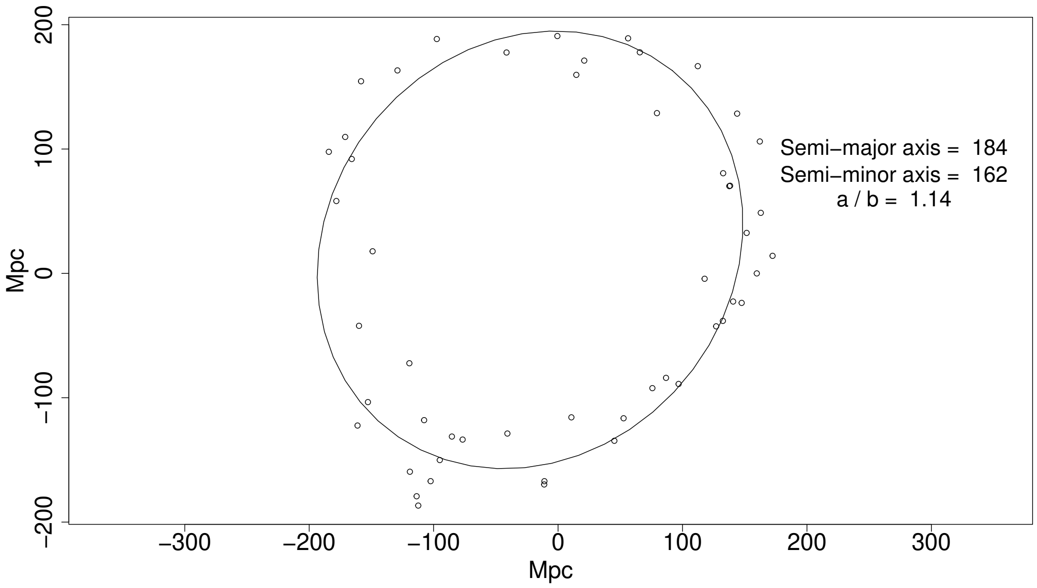


Figure 5.22: The visually-identified BR absorber members projected onto the plane perpendicular to w_0 with an added fitted ellipse.

ellipse,

$$p_R = \pi \left\{ (a+b) + \frac{3(a-b)^2}{10(a+b) + \sqrt{a^2 + 14ab + b^2}} \right\}$$

where a and b are the semi-major and semi-minor axes. The circumference of the BR ellipse is then $\sim$ 1.1 Gpc (comparable to the extent of the GA).

5.4.3 Equivalent widths

In Figure 5.18 we showed the visually-identified BR absorber members projected onto a plane normal to the vector w_0 (where w_0 resembles the LOS), with the points colour-coded according to their redshift. In Figure 5.23 we show a corresponding figure, but with the points now colour-coded according to the rest-frame equivalent widths (EWs, in Å) of the λ2796 component of the Mg II doublet. (The equivalent widths and their uncertainties, both taken from the Mg II database of Anand et al., 2021, are tabulated in Table 5.5.)

In Figure 5.23 there is no obvious (visual) pattern in the distribution of lower

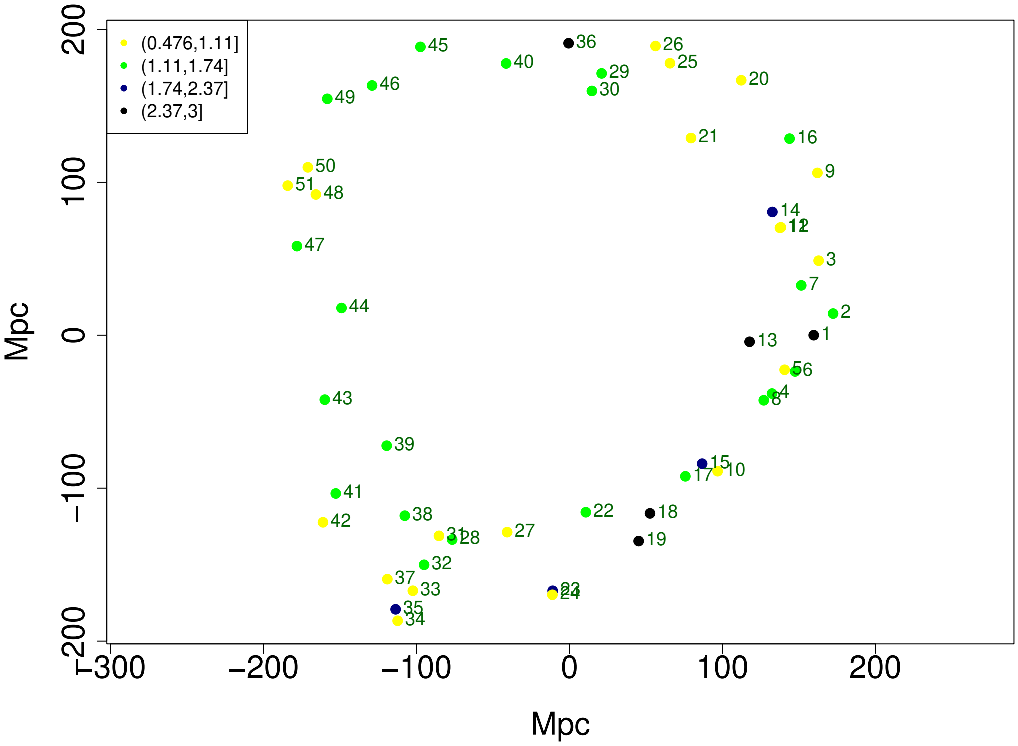


Figure 5.23: The visually-identified BR absorber members projected onto the plane perpendicular to the initial normal vector, w_0. The x-axis points towards the u_0 direction and the y-axis points towards the v_0 direction. The key in the top LHS of the figure indicates the rest-frame equivalent widths (EWs, in Å) of the $\lambda2796$ component of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number. The equivalent widths are taken from the Mg II database of Anand et al. (2021). The higher EWs seem to be concentrated on the RHS of the plot.

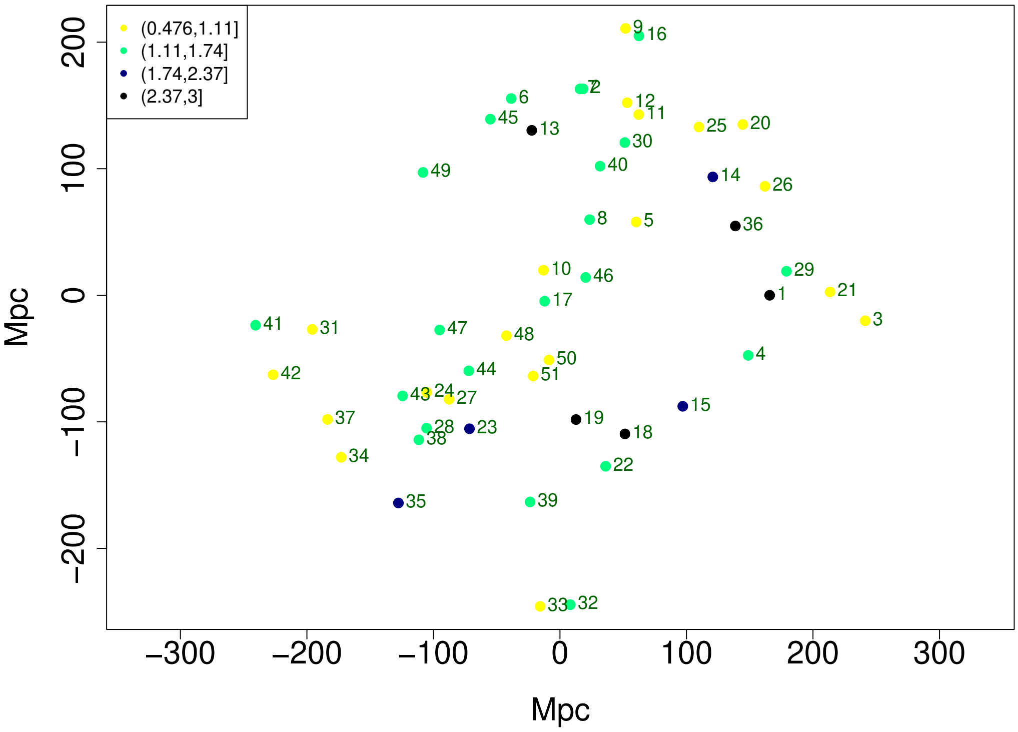


Figure 5.24: The BR absorber members projected onto the plane perpendicular to $u_0 - v_0$. The new u direction points towards the most easterly absorber, as usual. The key in the top LHS of the figure indicates the rest-frame equivalent widths (EWs, in Å) of the $\lambda2796$ component of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number. The equivalent widths are taken from the Mg II database of Anand et al. (2021). The higher EWs seem concentrated in the low-z ovoid shape.

EWs, but we note that the higher EWs tend to be concentrated on the RHS of the plot, especially the points that are numbered $14, 1, 13, 15, 18, 19, 23$ (but not $36, 35$).

In Figure 5.21 we showed the visually-identified BR absorber members projected onto a plane perpendicular to $u_0 - v_0$, which drew attention to the ovoid shapes at high-z and low-z, or otherwise the coiling pattern in the BR. In Figure 5.24 we show a corresponding figure, but now with colour-coding according to the equivalent widths (as above). It shows that the points $14, 36, 1, 15, 18, 19$ (and possibly $23, 35$

but not 13) are actually concentrated in the low-z ovoid shape. Possibly a preferential alignment with respect to the LOS of the host galaxies in this ovoid shape is enhancing the column densities.

5.5 Discussion and conclusions

In this chapter I have presented the discovery of the Big Ring; the work pertaining to this discovery was carried out (approximately) during the final two years of my (part-time) PhD. The work was first presented at the # 243rd American Astronomical Meeting (AAS) in January 2024 in New Orleans (USA). Shortly thereafter, the work was written-up, peer-reviewed, and published in the Journal of Cosmology and Astroparticle Physics (Lopez et al., 2024), which can be found appended to the back of this thesis.

The Big Ring (BR) is the second ultra-large large-scale structure (uLSS) detected in the Mg II catalogues. The BR is detected most prominently in the redshift slice $z = 0.802 \pm 0.060$, which is the exact same redshift slice containing the previously detected GA. Both structures on the sky are separated by only $\sim 12^{\circ}$, meaning that these two intriguing structures are in the same cosmological neighbourhood. We applied several inspection tests and statistical tests to the BR to support it as an ultra-large LSS. A summary of the results is as follows.

(1) Each of the visually-identified BR+filament absorbers was confirmed visually in the corresponding spectra, establishing that 100% of the absorbers are real detections, and not false-positives.

(2) The SLHC algorithm identified 46 out of 59 of the visually-identified BR and central filament) absorber members across 5 individual, overlapping or adjacent, groups. The SLHC algorithm also detected a statistically-significant arc in the redshift slices centred at $z = 0.862 \pm 0.060$ and $z = 0.922 \pm 0.060$. The arc appears to be an extension (along the line of sight) of the bottom portion of the BR, given

ID	EW (Å)	σ_{EW} (Å)	**ID**	EW (Å)	σ_{EW} (Å)
1	2.99	0.17	**27**	0.96	0.12
2	1.46	0.20	**28**	1.15	0.27
3	0.94	0.12	**29**	1.63	0.19
4	1.71	0.43	**30**	1.35	0.30
5	1.05	0.23	**31**	0.76	0.09
6	1.29	0.26	**32**	1.22	0.30
7	1.13	0.28	**33**	0.93	0.12
8	1.31	0.22	**34**	0.99	0.18
9	0.51	0.13	**35**	2.07	0.43
10	0.72	0.15	**36**	2.49	0.35
11	1.08	0.06	**37**	0.51	0.05
12	0.82	0.05	**38**	1.33	0.16
13	2.70	0.46	**39**	1.32	0.29
14	2.29	0.32	**40**	1.20	0.22
15	1.75	0.25	**41**	1.73	0.33
16	1.45	0.32	**42**	0.76	0.05
17	1.55	0.36	**43**	1.28	0.25
18	2.62	0.08	**44**	1.50	0.17
19	2.58	0.21	**45**	1.35	0.12
20	1.07	0.05	**46**	1.55	0.25
21	0.86	0.05	**47**	1.50	0.36
22	1.66	0.12	**48**	0.91	0.21
23	2.08	0.20	**49**	1.56	0.29
24	1.05	0.25	**50**	0.48	0.11
25	0.79	0.09	**51**	1.08	0.11
26	0.58	0.09			

Table 5.5: The table is for the 51 visually-identified BR absorber members. The identification numbers (ID) correspond to those in Figure 5.18 and the other similar figures that follow it. The rest-frame equivalent widths (EW) and their associated uncertainties (σ_{EW}, taken from the Mg II database of Anand et al. (2021), are for the $\lambda2796$ component of the Mg II doublet; the units are Å. Note that the EW for point 12 is likely to be enhanced by the superimposed $\lambda2803$ component from point 11.

the agreement in the on-sky position and that the redshift slices overlap by 50%.

(3) We calculated the CHMS and MST significance of the SLHC-identified absorbers, the visually-identified BR absorbers (two versions), and the FilFinder-identified absorbers. The CHMS method was found to have a much larger variation in its reported significances due to the nature of the method: the CHMS calculates the volume of the unique structure containing all the absorbers, so in the example of a ring-like structure, there is a large volume in the centre of the ring with far fewer absorbers than the annulus. The total (across all four absorber-member estimates of the BR) CHMS significance was $(3.65 \pm 1.13)\sigma$. With just the visually-identified BR-all absorbers (the BR annulus plus the BR inner absorbers) the CHMS algorithm yielded a (likely upper limit) significance of 5.2σ. The MST significance on the other hand was more consistent, as this method relies only on the mean MST edge-length; using this method the total (across all four absorber-member estimates of the BR) MST significance was $(4.10 \pm 0.45)\sigma$.

(4) The FilFinder algorithm was applied to the Mg II image of the BR field to identify filaments objectively. By incrementally increasing the size-threshold, the algorithm left only one identified filament — a ring — and that ring corresponded to the visually-identified BR. In this way, the BR was established as: (i) a ring, independently of visual perception; and (ii) the most-connected and largest filament in the image. We also applied the FilFinder algorithm to the field quasars (not to be confused with the background probes) and found that there was a large, connected, ring-like filament that mostly aligns with the BR, which provides further support of the BR structure with independent corroborative data.

(5) We applied the CE test to the field containing the BR and then 'zoomed' in on the BR to assess the contribution of the BR to the spatial clustering in the field. Tentative significant clustering in the field was detected in the second zoom of the BR, at a p-value $p = 0.022$ (corresponding to a significance of 2.0σ). We compared

this with four other unrelated fields at the same redshift as the BR and found no significant spatial clustering. However, the CE test results for the BR field are still inconclusive at the $\sigma > 3.0$ significance level, suggesting that the clustering in the BR field is not statistically significant, unlike what was seen with the GA field. Note, this test assesses the clustering in the *field* and not the statistical significance of any individual candidate structure.

(6) By superimposing contour maps of the field quasars and DESI clusters on the Mg II absorber images, we were able to show a plausible association between the datasets, thus providing independent corroboration. We found in particular that the DESI clusters, when subject to richness limits ($R \leq 22.5$ and $R \geq 22.5$), were mapping different Mg II absorbers, suggesting that there could be scope for using DESI clusters as a way to investigate the physical environment around the Mg II absorbers.

(7) Using a project-plane method we investigated the 3D distribution of the BR. Viewing the BR side-on (with the project-plane perpendicular to u_0 so that the x-axis resembles the redshift axis) revealed three, vertical redshift bands. The near and far redshift bands had a distinct, curious backwards 'S' shape, which could also been seen (less distinctly) in the central redshift band. With other viewing angles of the BR we found that the BR appears to be more of a coil shape. The central redshift band contains the majority of the absorbers in a ring shape in a thin, flat region. The near-redshift band is contained almost entirely on the RHS (east) of the BR and even appears to angle backwards into the central redshift band of the BR creating a spiral-like shape. The farthest-redshift band had a similar coiling effect, but appeared to have a patchy distribution (it made a broken, ring shape).

The data and analysis show that the BR is of particular interest for LSS studies in cosmology. In cosmology, we assume statistical homogeneity in the Cosmological Principle (CP) as the foundations of the ΛCDM model. We have shown that the BR is real and statistically significant, adding it to a growing list of LSS candidates that are in tension with the CP. The growing list of LSSs also indicates that the chance of all of these structures being a matter of statistical noise is decreasing. In addition, the list of LSS in tension with the CP is also generally difficult to understand as there are gaps in the knowledge for explaining the formation of these ultra-large structures. Perhaps we could speculate that the existence of these structures requires an extension to the standard model, perhaps in the form of cosmic strings (see Chapter 6.2).

The BR and GA together, given their sizes and morphologies, are presumably telling us something intriguing, and quite possibly important, about the Universe, but at the moment we can only speculate what that might be.

Chapter 6

Discussion and Conclusions

The Giant Arc (GA) and Big Ring (BR) ultra-large large-scale structure (uLSS) discoveries have been individually presented in this thesis in chapters 4 and 5, respectively. Here I discuss them together, and suggest the possible implications they may have for Cosmology (Section 6.1). The review in this chapter is based on work that has been published as a contribution piece of a Philosophical Transactions A issue (Lopez et al., 2025), following the Royal Society meeting held in London, in April, 2024 on *"Challenging the Standard Cosmological Model"*. The published paper can be found appended to the back of this thesis. Then, I discuss my priorities now and some potential future work (Section 6.2). Finally, I conclude the work that has been presented in this thesis (Section 6.3).

6.1 Discussion: Implications for Cosmology

The discovery of the BR is particularly interesting as it was made directly after the discovery of the GA but using the newer Anand21 Mg II catalogues instead of the older Z&M catalogues. The BR was first apparent when looking at the GA field with the Anand21 database; just to the north of the GA was an interesting ring shape, which we then investigated further.

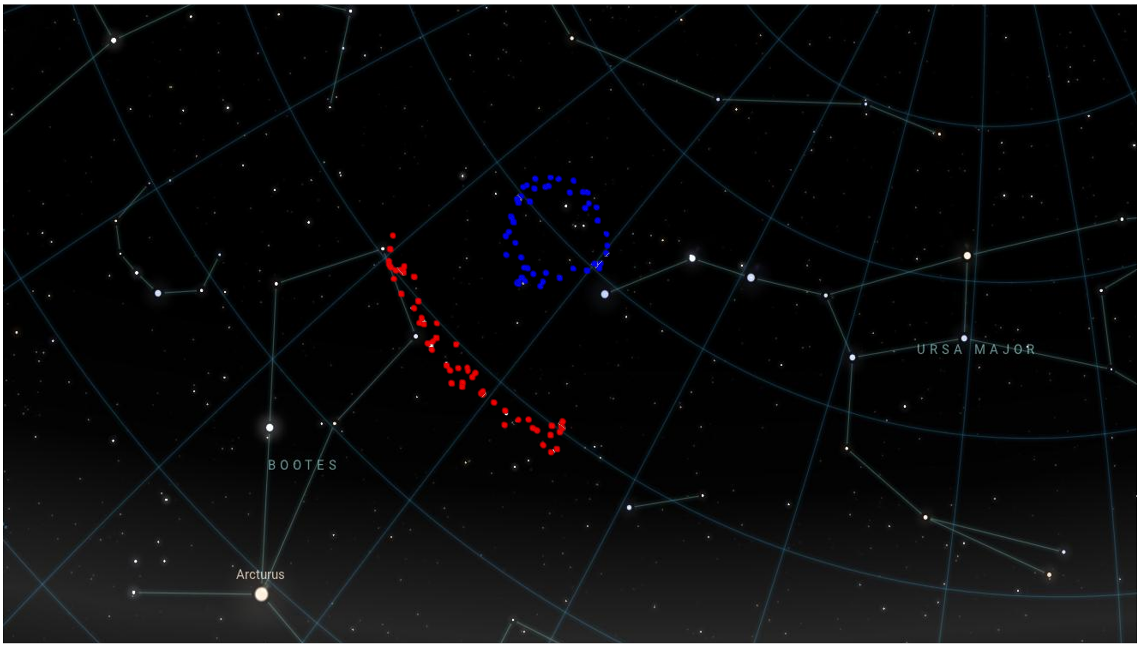


Figure 6.1: An approximate projection of the GA and BR visually-identified absorber members superimposed onto an image of the night sky taken from Stellarium. This figure gives an impression of the scale of the two uLSSs, their position on the sky, and their proximity.

On the sky, the GA and the BR are separated by only about 12°, and they both lie in exactly the same redshift slice $z = 0.802 \pm 0.060$. Figure 6.1 is an approximate projection of the BR and GA absorber members imprinted onto a night sky image taken from Stellarium. In this figure we gain an impression of the scale of the two uLSSs and how they would look in the night sky if we could see them. Their angular sizes span several degrees across the sky. Their physical sizes are of the largest structures observed, and yet we are seeing them as they were when the Universe was only half its present age (the structures are $\sim$ 9.2 billion light-years away). Naturally, several important questions arise as a result of these two intriguing discoveries, which I will discuss here.

The first question might be: how did these structures form and how will they evolve? Elucidating the answers to this question requires a theory (and a model) which makes predictions that we can then directly compare with observations. In Cosmology we have the ΛCDM model, so it follows, with the assumptions made, and with the chosen values of the ΛCDM parameters, do cosmological simulations predict huge, circular-type structures such as the GA and BR, and even the GA+BR system (two uLSSs in such close proximity)? If not, then what could that be telling us about our current standard model? (Or, are the observations just some statistical fluke that can be ignored?) Note that, this question is addressing the challenges of uLSSs towards the ΛCDM model *independently* from the challenge uLSSs present to the assumed homogeneity in the Cosmological Principle (CP). This is because, in most cosmological models, cosmic growth is assumed to follow linear theory, which does not yet appear to explain the uLSSs that have been discovered in our own Universe. Taking the intriguing discoveries to be hints about the shortfalls of our understanding of the Universe might lead us in a direction of making incremental steps towards a more complete version of our standard model. In this light, consider the following quote in a *New Scientist* article from Nobel laureate and founder of modern cosmology, Prof. Jim Peebles: *"But maybe the issue [Hubble tension] is a hint towards something we need to improve in our theory. There are other predictions in that same nature. The big hope, to my mind, is that other anomalies will appear."*. It would seem that the discovery of the GA and BR resonates with the hopes voiced by Prof. Jim Peebles, further demonstrating the importance of investigating intriguing results.

In April 2024, upon prior invitation from Prof. Subir Sarkar, I presented my findings of the GA and BR at a Royal Society Meeting in London "Challenging the Standard Cosmological Model". The meeting progressed over two days with healthy and lively discussions on the many tensions that cosmology is facing, for example: the Hubble tension (although, recent results suggest an agreement between the Kilo-Degree Survey and the Planck Legacy cosmic microwave background experiment, Wright et al., 2025); the S_8 (or σ_8) tension; the dipole anisotropy in the CMB; and cosmic bulk-flow estimates — e.g., Efstathiou (2025); Secrest (2025); Watkins (2025). However, at the end of the meeting, no particular clarity had emerged on the question of whether systematic uncertainties or ΛCDM shortfalls are the root cause of these anomalies and tensions. Additionally, in the example of the Cosmological Principle, varying interpretations amongst individuals was clearly a contributing factor to some of the disagreements over what the observational data are showing us.

The main topic I have discussed throughout this thesis is that of the potential challenges that LSSs and uLSSs present to the ΛCDM model and the CP. However, it is worth noting that there are several tensions felt by the ΛCDM model from other observational anomalies, and while I have not presented those in this thesis, they do add together with the work here and together the anomalies indicate that there might be room for improvement, or even scope for different perspectives from new avenues, in our understanding of the Universe.

I emphasise here that the discovery of intriguing LSSs and uLSSs, such as the GA and BR, are important for cosmology for much more than simply their statistical significance and I encourage the reader to wonder about their unique morphologies, huge sizes, proximities, independent corroborative data etc. What is often found in LSS studies is an overemphasis on the clustering signal versus what can be found in the noise — many authors have shown that large, overdense, highly-clustered structures can be reproduced in mock catalogues, simulations and even random Poisson-point distributions (Park et al., 2012; Nadathur, 2013). When assessing individual LSS candidates, we can simply ask: (i) do the structures exceed the 'homogeneity scale'? and (ii) how frequent is a similar structure using extreme-value analysis. Both are good, useful questions to ask, and they give us an idea of how special an individual candidate structure is in the context of their clustering in the whole field (or Universe). However, the dangers of focussing solely on these basic questions are firstly, there is not yet one agreed definition of the CP, so relying on the homogeneity scale (if such a thing really exists) to provide a consensus for what is deemed interesting can be misleading. Secondly, extreme-value statistical analysis does not have the power to highlight curious anomalies that may be hinting at the flaws in our standard model or even new physics (instead they become washed out and grouped in with 'noise'). Currently we do not know what the GA, BR and GA+BR system are or represent, but understanding them will be important for cosmology.

Another question that arises is: are the two uLSSs linked in some way, perhaps in the initial conditions or in their origin? In the beginning of this thesis I reviewed large-scale structures generally; one reason they are interesting for cosmology is they can retain the information of the initial conditions, since inflation is thought to 'freeze' the initial density perturbations which then scale up to the size of the cosmic web (Einasto et al., 2016). However, the GA and BR are extremely (significantly) large and overdense, and would presumably require an equally significant inhomogeneity in the initial conditions to explain them. Alternatively, if there is some cause for their existence outside of ΛCDM, would the two uLSS discoveries originate from this same disturbance (e.g., cosmic strings)? The two uLSSs are intriguingly close, and they are both circular-type structures (from our LOS). A circular shape is a very interesting shape as it implies something unique and specific within the formation or evolution.

If the trajectory of the GA was to continue and form a full-ring shape, then this would envelop the BR. In the BR field of the Anand21 databases there was an interesting thin filament which appears to be a potential continuation of the GA. In Figure 6.2 we have added the GA points from the Z&M catalogues on the same plot as the BR points from the Anand21 catalogues. In addition we have added the points of the interesting thin filament that appeared in the Anand21 catalogues. With fitted ellipses it can be seen that if (with more data) this interesting filament was a continuation of the GA, then the extended GA, or Giant Ring (GR) perhaps, would then encompass the BR (Figure 6.2). Does it follow then that the BR and the (potential) GR are almost concentric circles? Could it be that, if they were close to concentric, they formed from the same initial disturbance, such as a cosmic string? Or perhaps the Penrose Conformal Cyclic Cosmology (Penrose, 2006; Meissner & Penrose, 2025) lends an explanation to these huge, circular-type structures.

In the case on cosmic strings, it has been noted earlier in this thesis that they could provide an extension to the ΛCDM model that might help explain large-scale fluctuations (Hindmarsh & Kibble, 1995; Vilenkin, 1997; Shlaer et al., 2012; Duplessis & Brandenberger, 2013). Cosmic strings are topological defects that form as a result of symmetry breaking during early-Universe phase transitions. They are expected to be very large (on the order of the scale of the observable Universe) and very rare. This then makes cosmic strings a tantalising option to explain uLSSs that are also, presumably, quite rare. However, a complete search through the MgII catalogues (future work) could conceivably reveal many more ring-type, or otherwise, uLSSs all over the sky.

Cosmic strings have become of topical interest in recent years with Pulsar Timing Array (PTA) data (Wang et al., 2023; Srivastava, 2024) and with Peebles' work on flat, sheet-like LSS (Peebles, 2023), but also more generally over the last decade. In early works, cosmic-string scenarios required cosmic strings with energy scales $G\mu \sim 10^{-6}$, where $G\mu$ is a dimensionless parameter; μ is the string tension and G is the gravitational constant. These cosmic strings were later conclusively ruled out by

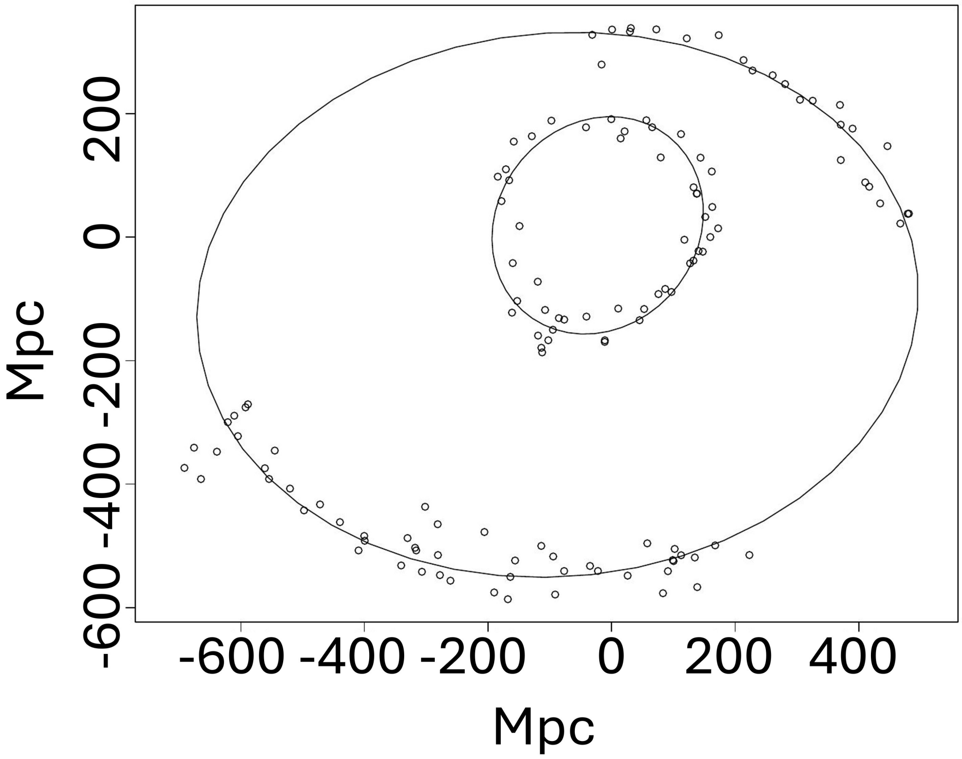


Figure 6.2: The GA points from Z&M (along the bottom) plotted with the BR points from Anand21 (in the middle). The thin filament to the north was also detected in the newer Anand21 catalogues. With the fitted ellipses it appears that the interesting thin filament could be a continuation of the GA, and if so, then this Giant Ring would encompass the BR.

CMB observations (Shlaer et al., 2012, and references therein). Additionally, since CMB fluctuations were already found to agree so well with ΛCDM power spectrum predictions, the contributions from comic strings could not account for more than $\sim 2\%$ of the spectral power (Shlaer et al., 2012). One study used WMAP and SDSS data to place bounds on cosmic strings. These authors estimated string energy scales down to $G\mu \sim 10^{-7}$ and a total power contribution to the CMB anisotropy of $\sim 10\%$ (Wyman et al., 2005; Pogosian et al., 2006).

Cosmic string scenarios require an equation of state parameter $\omega_0 = -1/3$, which is within the boundaries of the most recent constraints from DESI BAO DR2 DESI Collaboration et al. (2025). However, these most recent DESI results were found to be in the best agreement with other data (supernovae and CMB) after invoking a time-evolving equation of state defined as $\omega(a) = \omega_0 + \omega_a(1 - a_0)$, where ω_a corresponds to the scale factor $a = (1 + z)^{-1}$. It remains to be seen if cosmic string scenarios can hold in a Dynamical Dark Energy model.

A particularly interesting study from Sazhina et al. (2014) carried out a search for cosmic string candidates from radio surveys. Their prime cosmic string candidate CSc-1 (Sazhina et al., 2019) was supported by an observation of an excess of lensed galaxies in the string field. Curiously, the location of CSc-1 is roughly at the same location as the Huge-LQG.

Cosmic strings were previously considered to be observationally unsupported, but now they appear to be being considered once again, in certain contexts. They are a natural prediction in most grand unified theories (GUT), and perhaps we will see tighter constraints on cosmic strings in future work. Also, perhaps the theory of late-Universe LSS seeded by cosmic strings will be revisited.

Finally, via private communication from Dr Ely Kovetz, I was made aware of some older, but intriguing work that might relate to the question of giant, ring-type structures in the Universe. The work pertains to a string-inspired model whereby a cosmic defect seeded by a pre-inflationary particle (PIP) would leave a cosmological imprint. The theoretical groundwork appears to be originally motivated by the presence of large-scale anomalies (Fialkov et al., 2010). Similar to the work carried out by groups working on the CCC model (see below), this group have studied the CMB and found evidence for anomalous giant rings in the CMB sky (Kovetz et al., 2010) which they reason could be a signature of PIPs. Perhaps such an idea will be revisited, with a particular emphasis on what predictions can be made on the formation and evolution of cosmological LSS within this string-inspired model.

In the case on Conformal Cyclic Cosmology (CCC), the intriguing link between this alternative model that predicts huge, concentric rings in the CMB and the discovery of two circular-type uLSSs, that also may even be close to concentric, is considered. In the CCC model, collisions of supermassive blackholes (SMBHs) in a ‘previous aeon’ are predicted to leave distinct temperature rings in CMB. Could it follow that the imprinted rings in the CMB sky leave a non-Gaussian disturbance in the initial conditions of the present aeon that could seed ultra-large, circular-type structures? In private communication with Sir Roger Penrose, we discussed at some length the requirements of investigating the CCC model with late-Universe LSS signatures. Currently, late-Universe structure-formation predictions from the CCC model have not been investigated; evidence for the CCC model is primarily attached to observations of rings in the CMB (An et al., 2017; Gurzadyan, V. G. & Penrose, R., 2013), i.e., early-Universe observations, but perhaps there could be some later effect on LSS formation. Clearly, a refined understanding in the theory of both the CCC model and galaxy formation from initial primordial conditions, would be necessary. An initial suggestion from Sir Roger Penrose was that, if the huge, circular-type structures are somehow a prediction of the CCC model, then it would follow that we should find rings everywhere on the sky. Furthermore, the CCC model predicts concentric rings in the CMB from the occurrence of multiple SMBH collisions occurring at the same location from our vantage point in the previous aeon. If the GA does indeed become the Giant Ring that encompasses the BR, could the concentric rings in the LSS be further evidence for the CCC model? (Note, this idea is very speculative, but a curiosity nonetheless.) Clearly, investigating this avenue would be complex, but it might be interesting to see where it could lead.

Briefly, and for completeness, it was previously mentioned that we had thought the BR may originate from an individual Baryon Acoustic Oscillation (BAO). BAOs are a result of the sound waves that propagated through the plasma of the early Universe freezing at the stage of cosmic recombination, when photons and baryonic matter discontinue their interactions. BAOs are expected to leave imprints of low-density contrast rings in the galaxy/baryonic matter distribution. The BAOs are predicted to have a shell structure with a characteristic radius $r \sim 150$ Mpc (scaled to the present epoch; Tully et al., 2023; Einasto et al., 2016; Planck Collaboration et al., 2016b; Anderson et al., 2014; Eisenstein et al., 2005). Thus, when we first detected the BR — a ring shaped structure with roughly 300 Mpc diameter — we suggested that BAOs might be a reasonable explanation. However, for a few reasons, we have since decided that an origin of the BR in BAOs is unlikely. For example, the BR is in fact elliptical, and the ring shape is only apparent from our perspective on the sky (resembling a cork-screw from the side-on perspective).

Since we made a direct discovery of a second uLSS, the BR, in the same cosmological neighbourhood as the first discovery, the GA, this leads to the final question of whether we would expect to find many more interesting, huge structures with circular or other conic-section morphologies. Perhaps, a more important question would be to determine *a priori* (i) whether ΛCDM can explain the occurrence of such huge structures, and (ii) whether, with the CP assumed (and a particular version of the CP — see Section 1.2), we should expect to find no other, or many more, of these types of structures.

6.2 Future Work

With the above discussions and questions in mind there still remains plenty of work to understand the implications of the GA and BR discoveries. Here I would like to discuss some of my priorities now for future research.

Firstly, we have thus far looked at only the small patch of sky containing the GA and BR, with the exception of neighbouring fields or redshift slices for statistical analysis, so there remains the full SDSS footprint to investigate. We intend to search through the available catalogues and document what is found. This work will probably be carried out in the manner of applying the Single-Linkage Hierarchical Clustering (SLHC) algorithm to the entirety of the Mg II catalogues as a first-check for potential LSS candidates. Although, previously we have discussed the shortfalls of applying the SLHC algorithm to an essentially incomplete dataset (where there are no quasars, we can not detect the intervening matter there). In this case, it may be worthwhile searching for a statistical test that will be more suitable for the data. The Cuzick and Edwards (CE) test was an example of this, but the CE test was created to assess only the general clustering in the field and not to identify individual candidate structures. Ultimately, the SLHC algorithm will likely be the best choice (for now) for the purposes of searching for LSS and uLSS candidates in the Mg II catalogues, but we will have to be mindful of the effects of an inhomogeneous set of background probes on our ability to detect intervening matter.

Secondly, in Section 3.2 I discussed the two catalogues of Zhu & Ménard (2013) and Anand et al. (2021). We had found that there was a substantial loss of absorbers from the old catalogue to the new catalogue, despite the newer catalogue having access to all the previously observed quasar spectra. We reason that this disagreement could be a result of (a) spurious or unreliable absorbers contaminating the Z&M catalogue that were filtered out by Anand21 or (b) Z&M allowing detections to smaller EWs or perhaps complex line profiles — likely some combination of both (a) and (b) are true. Equally we found that the Z&M catalogue was missing absorbers that were detected by Anand21 for quasar spectra that was available to both authors. Accessing the Mg II-finder code of each author will possibly elucidate the disagreements between the two catalogues. Alternatively, through visual inspection of the quasar spectra hosting Mg II absorbers that are found by one author and not the other (an onerous job in reality), we might spot trends that would explain the disagreement between the two catalogues.

We intend to produce our own Mg II catalogue using our own, unique Mg II-finder algorithm. We will then have the ability to apply our pipeline on new quasar catalogues as they appear, such as the anticipated DESI quasar catalogue (not yet publicly available; DESI Collaboration et al., 2024). In addition, we can apply our pipeline to the previous SDSS quasar catalogues, providing us with a third, independent catalogue — we could run further cross-checks between ours, the Anand21, and Z&M Mg II catalogues. Finally, we would have the freedom to adapt the code as needed, as well as knowing exactly what cuts (i magnitude, S/N cuts) have been made.

At the beginning of this thesis, and throughout, I have reiterated that the definition of 'large scales' in the Cosmological Principle is ambiguous and vague. As a result, there appear to be multiple interpretations of the CP, some of which are possibly even contradictory, which makes it difficult to assess the data in a consistent and precise way. For LSS especially, we are relying on comparing results with the assumption of homogeneity. However, we do not yet know, or agree, to what extent we should observe homogeneity, or even how often we should statistically expect to find deviations from homogeneity. In this thesis I have presented two intriguing uLSS discoveries that individually and together potentially challenge the CP; but the accumulation of all of the uLSS discoveries to date pose a greater potential challenge to the CP (unless, of course, we increase or even abandon the idea of a homogeneity scale). Therefore, it is imperative for the critical assessment of our current standard model to establish an agreed, precise definition of the CP definitively. Although I concur on this matter, it is not yet clear to me how this problem in cosmology will be addressed.

In Chapter 2 of this thesis I reviewed the physics and astrophysics of the Mg II absorbers. The Mg II gas is known to trace galaxies, galaxy clusters, and star-formation regions generally, but the precise nature of the complex, patchy gas distribution is not very-well understood. Several studies have been conducted to understand the relationship between observed Mg II-line profile properties and their physical environments, such as mass, morphology, inclination, impact parameter, number density etc. Presumably, as the research develops in this field, we can utilise new results for our specific application of analysing cosmological LSS using the Mg II method. Perhaps, too, we might be able to learn about the environment of LSS candidates identified with the Mg II method.

I showed in the previous section a ‘northern arc’ that appears in the Anand21 catalogue which might be extending the GA into the Giant Ring (GR) that encompasses the BR. If this is the case, the BR+GR system would have quite strong implications for cosmology (overturn cosmology as we know it?). Therefore, we will investigate the possibility of the GA extending into the GR. This work will be carried out using the Anand21 data, and, upon completion, our own Mg II catalogue. Standard statistical techniques will at least be initially utilised, but we envision a need for stronger, more applicable, statistical techniques.

Finally, there appear to be at least a few different avenues in cosmology that could be explored for explaining these uLSS anomalies (and perhaps other cosmological anomalies too). I think it is sensible to consider the anomalies as hints to the shortfalls of our standard model. We will continue to monitor new ideas in cosmology, as well as how the standard model addresses the tensions challenging it.

The hope is a consensus will begin to emerge on whether systematics will ultimately mend the tensions in cosmology, or whether new physics will be required.

After the PhD, in January 2025, I began my first postdoctoral research assistant position at the University of Lancashire (formerly Central Lancashire). In the first third of my 1-year postdoctoral position, we addressed the Large Scale Structure (LSS) statistical analysis (using Minimal Spanning Tree methods) of the FLAMINGO-10K simulation data for cold dark matter particles (Lopez & Clowes, 2025). In summary, we find that there are no gigaparsec structures to be seen in Λ*CDM, suggesting that either (a) the simulations have not been able to adequately represent the real LSS of the Universe, or (b) the accumulation of ultra-large LSS (uLSS) discoveries present a more direct challenge to the standard cosmological model. For transparency, we believe it is probably a mixture of both.*

6.3 Conclusions

In this thesis I have presented two uLSSs: the Giant Arc and the Big Ring (GA and BR). Both were discovered visually as prominent overdensities in Mg II images, and subsequently supported by a range of statistical assessments and possibly by independent corroboration from DR16Q quasars and DESI clusters. Both exceed the Yadav estimated upper-limit to the scale of homogeneity and add to an accumulating list of uLSSs that challenge the Cosmological Principle (CP) in this respect. (Note, the interpretations of the CP vary, see Section 1.2.)

The GA and BR have intriguing morphologies, as indicated by their names. There is even a hint that the GA could extend into a Giant Ring that envelops the BR. While the BR appears entirely as a ring in projection on the sky, in 3D visualisations it appears to coil into and out of a central flat ring (which contains most of the member absorbers).

Extreme-value analysis and cosmological simulations have shown that some individual occurrences of uLSSs may be consistent with the CP and with ΛCDM, but it is not obvious that the entire accumulating set of uLSSs will be consistent. Additionally, our most recent work suggests that gigaparsec *structures* cannot be reproduced in ΛCDM using the FLAMINGO-10K simulation data (Lopez & Clowes, 2025).

Given their morphologies, perhaps the GA and BR in particular require an explanation outside ΛCDM. One possibility might be cosmic strings, which have become of topical interest in recent work (Ahmed et al., 2024; Cyr et al., 2024; Ellis et al., 2023; Gouttenoire & Vitagliano, 2024; Jiao et al., 2023; Peebles, 2023; Sanyal, 2022; Wang et al., 2023). For the BR specifically, we noted its similarity in projection and radius to an individual baryonic acoustic oscillation (BAO) (Tully et al., 2023; Einasto et al., 2016; Planck Collaboration et al., 2016b; Anderson et al., 2014; Eisenstein et al., 2005). However, given the non-spherical, coiling nature of the BR, and its actually being larger than a BAO ($r \sim 200$ Mpc compared with $r \sim 150$ Mpc for a BAO), the BR having an origin in BAOs is probably ruled out.

Unexpected and apparently anomalous discoveries in cosmology, such as uLSSs in general and the GA and BR in particular, may be indicating a route to further understanding and to refinements of the standard model. It may be productive to see where they lead.

Bibliography

Ahmed, W., Chowdhury, T. A., Nasri, S., & Saad, S. 2024, *Phys. Rev. D*, 109, 015008

Aluri, P. k., Cea, P., Chingangbam, P., et al. 2023, *Class. Quantum Gravity*, 40, 094001

An, D., Meissner, K. A., & Nurowski, P. 2017, *Mon. Not. Roy. Astron. Soc.*, 473, 3251

Anand, A., Nelson, D., & Kauffmann, G. 2021, *Mon. Not. Roy. Astron. Soc.*, 504, 65

Anderson, L., Aubourg, É., Bailey, S., et al. 2014, *Mon. Not. Roy. Astron. Soc.*, 441, 24

Andrade, U., Gonçalves, R. S., Carvalho, G. C., et al. 2022, *J. Cosmo. Astroparticle Phys.*, 2022, 088

Bagchi, J., Sankhyayan, S., Sarkar, P., et al. 2017, *Astrophys. J.*, 844, 25

Balázs, L. G., Bagoly, Z., Hakkila, J. E., et al. 2015, *Mon. Not. Roy. Astron. Soc.*, 452, 2236

Barnes, L. A., Garel, T., & Kacprzak, G. G. 2014, *Pub. Astron. Soc. Pac.*, 126, 969

Bergeron, J. 1988, in Large Scale Structures of the Universe, ed. J. Audouze, M. C. Pelletan, A. Szalay, Y. B. Zel'dovich, & P. J. E. Peebles, Vol. 130, 343

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, 380, 603

Burenin, R. A., Bikmaev, I. F., Khamitov, I. M., et al. 2018, *Astron. Letters* , 44, 297

Cherrey, M., Bouché, N. F., Zabl, J., et al. 2024, *Mon. Not. Roy. Astron. Soc.*, 528, 481

Chow-Martínez, M., Andernach, H., Caretta, C. A., & Trejo-Alonso, J. J. 2014, *Mon. Not. Roy. Astron. Soc.*, 445, 4073

Christian, S. 2020, *Mon. Not. Roy. Astron. Soc.*, 495, 4291

Churchill, C. W., Kacprzak, G. G., & Steidel, C. C. 2005, in IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines, ed. P. Williams, C.-G. Shu, & B. Menard, 24–41

Churchill, C. W., Mellon, R. R., Charlton, J. C., et al. 2000a, *Astrophys. J. Suppl.*, 130, 91

Churchill, C. W., Mellon, R. R., Charlton, J. C., et al. 2000b, *Astrophys. J.*, 543, 577

Churchill, C. W., Nielsen, N. M., Kacprzak, G. G., & Trujillo-Gomez, S. 2013, *Astrophys. J. Letters*, 763, L42

Churchill, C. W., Rigby, J. R., Charlton, J. C., & Vogt, S. S. 1999, *Astrophys. J. Suppl.*, 120, 51

Clowes, R. G. 1986, *Mon. Not. Roy. Astron. Soc.*, 218, 139

Clowes, R. G. & Campusano, L. E. 1991, *Mon. Not. Roy. Astron. Soc.*, 249, 218

Clowes, R. G., Campusano, L. E., Graham, M. J., & Söchting, I. K. 2012, *Mon. Not. Roy. Astron. Soc.*, 419, 556

Clowes, R. G., Harris, K. A., Raghunathan, S., et al. 2013, *Mon. Not. Roy. Astron. Soc.*, 429, 2910

Colless, M. 2004, in Measuring and Modeling the Universe, ed. W. L. Freedman, 196

Cuzick, J. & Edwards, R. 1990, *J. Roy. Stat. Soc. Series B (Methodological)*, 52, 73

Cyr, B., Chluba, J., & Acharya, S. K. 2024, *Phys. Rev. D*, 109, L121301

De Marzo, G., Sylos Labini, F., & Pietronero, L. 2021, *Astron. Astrophys.*, 651, A114

DESI Collaboration, Abdul-Karim, M., Aguilar, J., et al. 2025, arXiv e-prints, arXiv:2503.14738

DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, *Astron. J.*, 167, 62

Duplessis, F. & Brandenberger, R. 2013, *J. Cosmo. Astroparticle Phys.*, 2013, 045

Dutta, R., Srianand, R., Gupta, N., & Joshi, R. 2017, *Mon. Not. Roy. Astron. Soc.*, 468, 1029

Efstathiou, G. 2025, Philosophical Transactions of the Royal Society of London Series A, 383, 20240022

Einasto, M., Heinämäki, P., Liivamägi, L. J., et al. 2016, *Astron. Astrophys.*, 587, A116

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *Astrophys. J.*, 633, 560

Ellis, J., Lewicki, M., Lin, C., & Vaskonen, V. 2023, *Phys. Rev. D*, 108, 103511

Evans, J. L., Churchill, C. W., Murphy, M. T., Nielsen, N. M., & Klimek, E. S. 2013, *Astrophys. J.*, 768, 3

Fialkov, A., Itzhaki, N., & Kovetz, E. D. 2010, *J. Cosmo. Astroparticle Phys.*, 2010, 004

French, J. 2020, Statistical Methods for the Analysis of Case-Control Point Data

Friday, T., Clowes, R. G., & Williger, G. M. 2022, *Mon. Not. Roy. Astron. Soc.*, 511, 4159

Geller, M. J. & Huchra, J. P. 1989, *Science*, 246, 897

Gonçalves, R. S., Carvalho, G. C., Andrade, U., et al. 2021, *J. Cosmo. Astroparticle Phys.*, 2021, 029

Gonçalves, R. S., Carvalho, G. C., Bengaly, C. A. P., J., et al. 2018, *Mon. Not. Roy. Astron. Soc.*, 475, L20

Gott, R. J., Jurić, M., Schlegel, D., et al. 2005, *Astrophys. J.*, 624, 463

Gouttenoire, Y. & Vitagliano, E. 2024, *Phys. Rev. D*, 110, L061306

Guha, L. K., Srianand, R., & Petitjean, P. 2024, *Mon. Not. Roy. Astron. Soc.*, 527, 5075

Gurzadyan, V. G. & Penrose, R. 2013, *Eur. Phys. J. Plus*, 128, 22

Hindmarsh, M. B. & Kibble, T. W. B. 1995, *Rep. Prog. Phys.*, 58, 477

Hinrichsen, V. L., Klassen, A. C., Song, C., & Kulldorff, M. 2009, *Int. J. Health Geogr.*, 8

Horváth, I., Hakkila, J., & Bagoly, Z. 2014, *Astron. Astrophys.*, 561, L12

Horvath, I., Szécsi, D., Hakkila, J., et al. 2020, *Mon. Not. Roy. Astron. Soc.*, 498, 2544

Huang, Y.-H., Chen, H.-W., Shectman, S. A., et al. 2021a, *Mon. Not. Roy. Astron. Soc.*, 502, 4743

Huang, Y.-H., Chen, H.-W., Shectman, S. A., et al. 2021b, *Mon. Not. Roy. Astron. Soc.*, 502, 4743

Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *Astrophys. J. Suppl.*, 199, 26

Huterer, D. 2023, *Astron. Astrophys. Rev.*, 31, 2

Hutsemékers, D. 1998, *Astron. Astrophys.*, 332, 410

Hutsemékers, D., Braibant, L., Pelgrims, V., & Sluse, D. 2014, *Astron. Astrophys.*, 572, A18

Hutsemékers, D., Cabanac, R., Lamy, H., & Sluse, D. 2005, *Astron. Astrophys.*, 441, 915

Hutsemékers, D. & Lamy, H. 2001, *Astron. Astrophys.*, 367, 381

Jiao, H., Brandenberger, R., & Refregier, A. 2023, *Phys. Rev. D*, 108, 043510

Kacprzak, G. G., Churchill, C. W., & Steidel, C. C. 2005, in IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines, ed. P. Williams, C.-G. Shu, & B. Menard, 80–85

Kacprzak, G. G., Churchill, C. W., Steidel, C. C., & Murphy, M. T. 2008, *Astron. J.*, 135, 922

Keenan, R. C., Barger, A. J., & Cowie, L. L. 2013, *Astrophys. J.*, 775, 62

Koch, E. W. & Rosolowsky, E. W. 2015, *Mon. Not. Roy. Astron. Soc.*, 452, 3435

Kovetz, E. D., Ben-David, A., & Itzhaki, N. 2010, *Astrophys. J.*, 724, 374

Lanzetta, K. M. & Bowen, D. 1990, *Astrophys. J.*, 357, 321

Lee, J. C., Hwang, H. S., & Song, H. 2021, *Mon. Not. Roy. Astron. Soc.*, 503, 4309

Lietzen, H., Tempel, E., Liivamägi, L. J., et al. 2016, *Astron. Astrophys.*, 588, L4

Lopez, A. M. 2019, Assessing the Potential of Intervening MgII Absorbers in Cosmology (University of Central Lancashire)

Lopez, A. M., Clowes, R., & Williger, G. 2025, Philosophical Transactions of the Royal Society of London Series A, 383, 20240029

Lopez, A. M. & Clowes, R. G. 2025, arXiv e-prints, arXiv:2504.14940

Lopez, A. M., Clowes, R. G., & Williger, G. M. 2022, *Mon. Not. Roy. Astron. Soc.*, 516, 1557

Lopez, A. M., Clowes, R. G., & Williger, G. M. 2024, *J. Cosmo. Astroparticle Phys.*, 2024, 055

Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, *Astrophys. J. Suppl.*, 250, 8

Marchã, M. J. M. & Browne, I. W. A. 2021, *Mon. Not. Roy. Astron. Soc.*, 507, 1361

Marinello, G. E., Clowes, R. G., Campusano, L. E., et al. 2016, *Mon. Not. Roy. Astron. Soc.*, 461, 2267

Meidt, S. E., Rosolowsky, E., Sun, J., et al. 2023, *Astrophys. J. Letters*, 944, L18

Meissner, K. A. & Penrose, R. 2025, arXiv e-prints, arXiv:2503.24263

Meliker, J. R., Jacquez, G. M., Goovaerts, P., & Yassine, M. 2009, *Cancer Causes Control*, 20, 1061

Mookerjea, B., Veena, V. S., Güsten, R., Wyrowski, F., & Lasrado, A. 2023, *Mon. Not. Roy. Astron. Soc.*, 520, 2517

Nadathur, S. 2013, *Mon. Not. Roy. Astron. Soc.*, 434, 398

Napolitano, L., Pandey, A., Myers, A. D., et al. 2023, *Astron. J.*, 166, 99

Oort, J. H. 1983, *Ann. Rev. Astron, Astrophys.*, 21, 373

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, *Astron. Astrophys.*, 597, A79

Park, C., Choi, Y.-Y., Kim, J., et al. 2012, *Astrophys. J. Letters*, 759, L7

Peebles, P. J. E. 1993, Principles of Physical Cosmology, Vol. 27 (Princeton university press)

Peebles, P. J. E. 2023, *Mon. Not. Roy. Astron. Soc.*, 526, 4490

Penrose, R. 2006, in Proceedings of EPAC 2006, 2759–2762

Pilipenko, S. V. 2007, *Astro. Rep.*, 51, 820

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016a, *Astron. Astrophys.*, 594, A27

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, *Astron. Astrophys.*, 594, A13

Pogosian, L., Wasserman, I., & Wyman, M. 2006, arXiv e-prints, astro

Pomarède, D., Tully, R. B., Graziani, R., et al. 2020, *Astrophys. J.*, 897, 133

Sanyal, S. 2022, *Eur. Phys. J. Spec. Top.*, 231, 83

Sazhina, O. S., Scognamiglio, D., Sazhin, M. V., & Capaccioli, M. 2019, *Mon. Not. Roy. Astron. Soc.*, 485, 1876

Sazhina, O. S., Sementsov, V. N., & Ashimbaeva, N. T. 2014, *Astron. Rep.*, 58, 16

Schaye, J., Kugel, R., Schaller, M., et al. 2023, *Mon. Not. Roy. Astron. Soc.*, 526, 4978

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, *Astron. J.*, 139, 2360

Schwarz, D. J. 2010, in Fundamental Interactions: A Memorial Volume for Wolfgang Kummer, ed. D. Grumiller & et al. (World Scientific Publishing Co. Pte. Ltd), 267–276

Secrest, N. J. 2025, Philosophical Transactions of the Royal Society of London Series A, 383, 20240027

Shimakawa, R., Okabe, N., Shirasaki, M., & Tanaka, M. 2023, *Mon. Not. Roy. Astron. Soc.*, 519, L45

Shlaer, B., Vilenkin, A., & Loeb, A. 2012, *J. Cosmo. Astroparticle Phys.*, 2012, 026

Söchting, I. K., Clowes, R. G., & Campusano, L. E. 2002, *Mon. Not. Roy. Astron. Soc.*, 331, 569

Söchting, I. K., Clowes, R. G., & Campusano, L. E. 2004, *Mon. Not. Roy. Astron. Soc.*, 347, 1241

Song, C. & Kulldorff, M. 2003, *Int. J. Health Geogr.*, 2

Srivastava, A. M. 2024, *Phys. Lett. B*, 853, 138683

Steidel, C. C. 1995, in QSO Absorption Lines, ed. G. Meylan (Berlin, Heidelberg: Springer Berlin Heidelberg), 139–152

Steidel, C. C. & Sargent, W. L. W. 1992, *Astrophys. J. Suppl.*, 80, 1

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

Tully, R. B., Howlett, C., & Pomarède, D. 2023, *Astrophys. J.*, 954, 169

van de Weygaert, R., Shandarin, S., Saar, E., & Einasto, J., eds. 2016, IAU Symposium, Vol. 308, The Zeldovich Universe: Genesis and Growth of the Cosmic Web

Vilenkin, A. 1997, in Critical Dialogues in Cosmology, ed. N. Turok, 415

Wang, Z., Lei, L., Jiao, H., Feng, L., & Fan, Y.-Z. 2023, *Sci. China. Phys. Mech. Astron.*, 66, 120403

Watkins, R. 2025, Philosophical Transactions of the Royal Society of London Series A, 383, 20240031

Webster, A. 1976a, *Mon. Not. Roy. Astron. Soc.*, 175, 61

Webster, A. 1976b, *Mon. Not. Roy. Astron. Soc.*, 175, 71

Webster, A. 1982, *Mon. Not. Roy. Astron. Soc.*, 199, 683

Whitbourn, J. R. & Shanks, T. 2016, *Mon. Not. Roy. Astron. Soc.*, 459, 496

Williger, G. M., Campusano, L. E., Clowes, R. G., & Graham, M. J. 2002, *Astrophys. J.*, 578, 708

Wright, A. H., Stölzner, B., Asgari, M., et al. 2025, arXiv e-prints, arXiv:2503.19441

Wyman, M., Pogosian, L., & Wasserman, I. 2005, *Phys. Rev. D*, 72, 023513

Yadav, J. K., Bagla, J. S., & Khandai, N. 2010, *Mon. Not. Roy. Astron. Soc.*, 405, 2009

Zhang, S., Wang, K., Liu, T., et al. 2023, *Mon. Not. Roy. Astron. Soc.*, 520, 322

Zhu, G. & Ménard, B. 2013, *Astrophys. J.*, 770, 130

Zou, H., Gao, J., Xu, X., et al. 2021, *Astrophys. J. Suppl.*, 253, 56

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **516,** 1557–1572 (2022) https://doi.org/10.1093/mnras/stac2204
Advance Access publication 2022 August 6

A Giant Arc on the Sky

Alexia M. Lopez ,[1]★ Roger G. Clowes [1] and Gerard M. Williger[2]
[1] *Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK*
[2] *Department of Physics and Astronomy, University of Louisville, Louisville, KY 40292, USA*

Accepted 2022 August 1. Received 2022 August 1; in original form 2021 August 30

ABSTRACT

We present the serendipitous discovery of a 'Giant Arc on the Sky' at $z \sim 0.8$. The Giant Arc (GA) spans ∼1 Gpc (proper size, present epoch) and appears to be almost symmetrical on the sky. It was discovered via intervening Mg II absorbers in the spectra of background quasars, using the catalogues of Zhu & Ménard. The use of Mg II absorbers represents a new approach to the investigation of large-scale structures (LSSs) at redshifts $0.45 \lesssim z \lesssim 2.25$. We present the observational properties of the GA, and we assess it statistically using methods based on (i) single-linkage hierarchical clustering ($\sim 4.5\sigma$); (ii) the Cuzick-Edwards test ($\sim 3.0\sigma$); and (iii) power-spectrum analysis ($\sim 4.8\sigma$). Each of these methods has distinctive attributes and powers, and we advise considering the evidence from the ensemble. We discuss our approaches to mitigating any post hoc aspects of analysing significance after discovery. The overdensity of the GA is $\delta\rho/\rho \sim 1.3 \pm 0.3$. The GA is the newest and one of the largest of a steadily accumulating set of very large LSSs that may (cautiously) challenge the Cosmological Principle, upon which the 'standard model' of cosmology is founded. Conceivably, the GA is the precursor of a structure like the Sloan Great Wall (but the GA is about twice the size), seen when the Universe was about half its present age.

Key words: galaxies: clusters: general – quasars: absorption lines – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

We are using intervening Mg II absorbers to investigate cosmological structure at redshifts $0.45 \lesssim z \lesssim 2.25$. Previous work at such redshifts has generally depended on either (i) quasars, for which there are accurate spectroscopic redshifts, or (ii) galaxies (and clusters), for which there are typically only somewhat less accurate photometric redshifts. However, with Mg II absorbers, we can obtain accurate spectroscopic redshifts for faint galaxies; subject, of course, to the limitations that the detected galaxies will be relatively sparse and that they can be only at the sky coordinates of the background quasars that are used as probes. These limitations need not, however, be a restriction on investigating large-scale structure (LSS).

In this way, we have serendipitously discovered a 'Giant Arc on the Sky' at $z \sim 0.8$. The Giant Arc (GA) spans ∼1 Gpc (proper size, present epoch) and appears to be intriguingly symmetrical on the sky. It is one of the largest of a steadily accumulating set of very large LSSs that may (cautiously) challenge the Cosmological Principle (CP), upon which the 'standard model' of cosmology is found. (Note that the word 'challenge' is not synonymous with 'contradict', but it does imply something to be investigated further.)

1.1 The largest large-scale structures

Yadav, Bagla & Khandai (2010) gave ∼370 Mpc as an ideal or upper limit to the scale of homogeneity in the concordance cosmology, beyond which departures from homogeneity should not be evident. As Yadav et al. (2010) state, above this scale it should not be possible to distinguish a given point distribution from a homogeneous distribution. We can therefore take ∼370 Mpc as an indication of the size (and, incidentally, separation also) beyond which LSSs become cosmologically interesting. We list in Table 1 some of the very large LSSs that have been reported in the literature, noting mainly those of (present-epoch) size $\gtrsim$ 370 Mpc. Many of the sizes quoted in this table will be somewhat uncertain, usually because of uncertainty in the boundaries, but sometimes because of uncertainty in what is being quoted in the papers (e.g. cosmological model and parameters, cosmological epoch). Also, in the lower part of Table 1, we list some results that, while not strictly being instances of LSS, are certainly of interest for considering the validity of the CP.

1.2 The Mg II approach

The presence of metal-rich intervening absorption lines in the spectra of quasars reveals foreground gas associated with galaxies. Specifically, the prominent and distinctive Mg II doublet feature, which can be seen over a broad range of redshifts, $0.45 \lesssim z \lesssim 2.25$, is strongly associated with the low-ionized gas around galaxy haloes, and is therefore an easily identifiable tracer of galaxies. In particular, Mg II is known to trace the H I regions indicative of star formation regions. Mg II absorbers can be expected to be useful tracers of large-scale structure, as they trace metal-enriched gas associated with galaxies and clusters.

It is generally believed that the strength of the absorption doublet (rest equivalent widths $W_{r,\,2796}$, $W_{r,\,2803}$) arising in the haloes corresponds to the properties of the galaxy and galaxy clusters,

---

★ E-mail: missalexia.lopez@gmail.com

Published by Oxford University Press on behalf of Royal Astronomical Society

Table 1. Some of the very large LSSs reported in the literature. The columns are the name of the LSS; the mean redshift; the reported size in Mpc (present epoch); and references. The lower part of the table lists some results, which, while not strictly instances of LSS, are of interest for the CP.

Name	Mean z	Size (Mpc)	References
HCB Great Wall	$\sim$2	2000–3000	Horváth, Hakkila & Bagoly (2014), Horvath et al. (2020)[a]
Giant GRB Ring	0.82	1720	Balázs et al. (2015)
Correlated LQG orientations	1.0–1.8	1600	Friday, Clowes & Williger (2022)
U1.27, Huge-LQG	1.27	1240	Clowes et al. (2013)[b]
Coherent quasar polarisation[c]	1–2	1000	Hutsemékers (1998), Hutsemékers & Lamy (2001), Hutsemékers et al. (2005)
U1.11	1.11	780	Clowes et al. (2012)
U1.28, CCLQG	1.28	630	Clowes & Campusano (1991), Clowes et al. (2012)
Sloan Great Wall	0.073	450	Gott et al. (2005)
South Pole Wall	0.04	420	Pomarède et al. (2020)
Blazar LSS	$\sim$0.35	350	Marchã & Browne (2021)
Local void	$<$0.07	300	Keenan, Barger & Cowie (2013), Whitbourn & Shanks (2016)
BOSS Great Wall (BGW)	0.47	250	Lietzen et al. (2016)
Great Wall	0.029	240	Geller & Huchra (1989)
Saraswati supercluster	0.28	200	Bagchi et al. (2017)
CMB anomalies			Schwarz et al. (2016)
Acceleration anisotropy			Colin et al. (2019)
Cluster anisotropy			Migkas et al. (2020)
Quasar dipole	0.5–2.0		Secrest et al. (2021)
Statistical power, LRGs	$\sim$0.6	1000	Thomas, Abdalla & Lahav (2011)

Notes. Further references.

[a] See also Christian (2020).

[b] See also Nadathur (2013), Marinello et al. (2016), and Hutsemékers et al. (2014).

[c] See also Marchã & Browne (2021).

although there is still uncertainty about the relative importance of morphology, luminosity, impact parameter, galaxy inclination, etc., and combinations of these that lead to the different classes (weak, strong) of absorbers (see e.g. Lanzetta & Bowen 1990; Churchill et al. 2000; Steidel et al. 2002; Churchill, Kacprzak & Steidel 2005; Chen et al. 2010; and Bordoloi et al. 2011).

The roughly spherical haloes hosting strong Mg II systems are considered to extend to radii $43 \leq r \leq 88$ kpc (Kacprzak et al. 2008). It has been suggested by Steidel (1995) and Churchill et al. (2005) that Mg II absorption with rest-frame equivalent width $W_{r,\,2796} < 0.3$ Å occurs predominantly in the outer regions of a halo, whereas stronger absorption with $W_{r,\,2796} > 0.3$ Å occurs predominantly in the inner regions. Such an association of equivalent width with radius is seen also in C IV and Lyman-limit absorption systems. Of course, there is probably patchy structure so generalizations may be misleading.

Lee, Hwang & Song (2021) investigate the rate of Mg II absorption in and around clusters of galaxies. They find that although the detection rate per quasar is higher inside the clusters, the rate is in fact quite low when considering the number of galaxies in the clusters – that is, the galaxy-to-absorber ratio is lower inside clusters, presumably because the environment within clusters modifies the galaxy haloes.

1.2.1 Data sources

We have constructed our Mg II absorber database from the Zhu & Ménard (Z&M) catalogues that are publicly available at the website https://www.guangtunbenzhu.com/jhu-sdss-metal-absorber-catalog. We have used their DR7 and DR12 ‘Trimmed’ catalogues (not the ‘Expanded’ versions). DR7 and DR12 indicate that the sources of the quasars that have been used as background probes are data releases 7 and 12 of the Sloan Digital Sky Survey (SDSS). The detection of the absorbers and the construction of the catalogues is described in Zhu & Ménard (2013) and the above website.

We paired the Z&M absorber catalogues on RA, Dec. to the ‘cleaned’ quasar data bases DR7QSO (Schneider et al. 2010) and DR12Q (Pâris et al. 2017). Thus, the absorbers can all be associated subsequently with either DR7QSO or DR12Q.

We removed entries for repeat spectra (see the above website) within the Z&M DR12(Q) absorber catalogue, thus avoiding duplication of absorbers. There were no entries for repeat spectra within the Z&M DR7(QSO) absorber catalogue.

When a particular absorber (RA, Dec., z) appeared in both the Z&M DR7(QSO) and DR12(Q) catalogues, we removed the DR7(QSO) entry, thus giving preference to DR12(Q) parameters. The final data base has 63876 Mg II absorbers.

We also produced a corresponding data base of probes from the Z&M ‘Quasars searched’ catalogues, similarly restricting them to those that appear in either DR7QSO or DR12Q. This data base has 123 351 member quasars.

1.3 Cosmological model

The concordance model is adopted for cosmological calculations, with $\Omega_{T0} = 1$, $\Omega_{M0} = 0.27$, $\Omega_{\Lambda 0} = 0.73$, and $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. All sizes given are proper sizes at the present epoch. (For consistency, we are using the same values for the cosmological parameters that were used by Clowes et al. 2013.)

2 THE GA

The GA is a large, filamentary, crescent-shaped structure that was discovered serendipitously in Mg II catalogues (Zhu & Ménard 2013). The GA extends $\sim$1 Gpc (proper size, present epoch) in its longest dimension, at a redshift of $z \sim 0.8$. We now discuss: the discovery of the GA; observational properties; connectivity and statistical properties; overdensity and comparisons with other data.

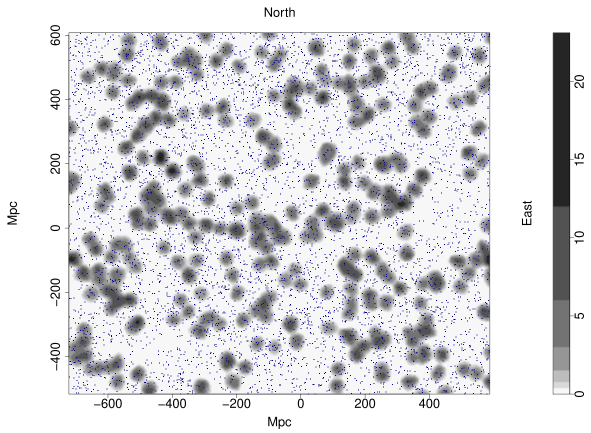


Figure 1. The tangent-plane distribution of Mg II absorbers centred in the redshift interval $z = 0.802 \pm 0.060$. The grey contours, increasing by a factor of 2, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes. The dark blue dots represent the background probes (quasars). The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top. The GA runs west-east in the centre of the figure, spanning ∼1 Gpc.

2.1 Discovery of the GA

The discovery and the preliminary analysis of the GA were first discussed in Lopez (2019) and are briefly summarized here. The GA was discovered when testing the Mg II approach on six small fields containing published structures (e.g. clusters and superclusters) at $z \gtrsim 0.7$. The Sunyaev–Zeldovich (SZ) cluster candidate PSZ2 G069.39 + 68.05 (Burenin et al. 2018, subsequently B18), at $z = 0.763$, was one such field: it indicated a very large LSS extending on the sky, as a dense, long, thin band of Mg II absorbers, roughly symmetrically to both sides of the cluster. In Lopez (2019), the central redshift and the redshift interval of the GA were estimated by stepping through thin redshift slices and visually inspecting the density and connectivity of the Mg II absorbers. We have since refined a little the estimates of the central redshift and the redshift interval (see Section 2.3.1 for the details). In Fig. 1, the GA can be seen stretching ∼1 Gpc (proper distance, present epoch) horizontally across the centre of the field. Visually, the GA appears densely concentrated and with the distinctive shape of a GA.

The Mg II density images, such as Fig. 1 here, and others throughout, are intended to give a useful impression of the connectivity of the absorbers. The images are constructed by smoothing the 2D distributions of the Mg II absorbers and the background probes (quasars) with a Gaussian kernel, with the same smoothing scale for both. In a process of ‘flat-fielding’ the absorber image is then divided by the normalized probe image to correct for non-uniformities in the distribution of the probes on that smoothing scale. The grey contours in the Mg II density images increase by a factor of 2. We use tangent-plane coordinates, scaled, using the central redshift, to present-epoch proper coordinates in Mpc.

2.2 Observational properties of the GA

We investigate the observational properties of the GA, in a visual manner, including equivalent width (EW) distribution ($W_{r,\,2796}$); signal-to-noise ratio (S/N) of the λ2796 Mg II line; S/N of the continuum of the spectra; the *i* magnitude (*i*) of the probes (background quasars); and the redshift distribution (z_{2796}). Previously, we noted that EW distribution could be related to the galaxy properties (morphology, luminosity, impact parameter, galaxy inclination, etc.), but these aspects are still not fully understood. While it may not yet be clear what the EW distribution within the GA indicates, future studies of EW in Mg II data should ultimately lead to more understanding of the origins of the GA and its environment.

The values of the EW are often classed as ‘strong’ or ‘weak’, although there seems to be no agreement on what defines ‘strong’ and ‘weak’. For example, in the literature one might find strong EW variously defined as $W_{r,\,2796} \geq 0.3$ Å, $W_{r,\,2796} \geq 0.6$ Å, and $W_{r,\,2796} \geq 1.0$ Å (see e.g. Williger et al. 2002; Churchill et al. 2005; Evans et al. 2013; Dutta et al. 2017). We shall follow Zhu & Ménard (2013) and use the definitions of strong and weak EWs as $W_{r,\,2796} \geq 0.6$ Å and $W_{r,\,2796} < 0.6$ Å, respectively.

We divide the $W_{r,\,2796}$ EWs into four bins with boundaries at 0.0, 0.3, 0.6, 1.0, 10.0 Å. (The boundaries were chosen to reflect the above diversity of what corresponds to ‘strong’ in the literature.) The on-sky spatial coordinates of the absorbers in the GA and its immediate field are then plotted, with colour coding according to the four EW bins (see Fig. 2a). The shade of the blue dots in the figure represents the EW bin, with the lightest shade representing the first bin $0.0 < W_{r,\,2796} \leq 0.3$ Å, and the darkest shade representing the last bin $1.0 < W_{r,\,2796} \leq 10.0$ Å.

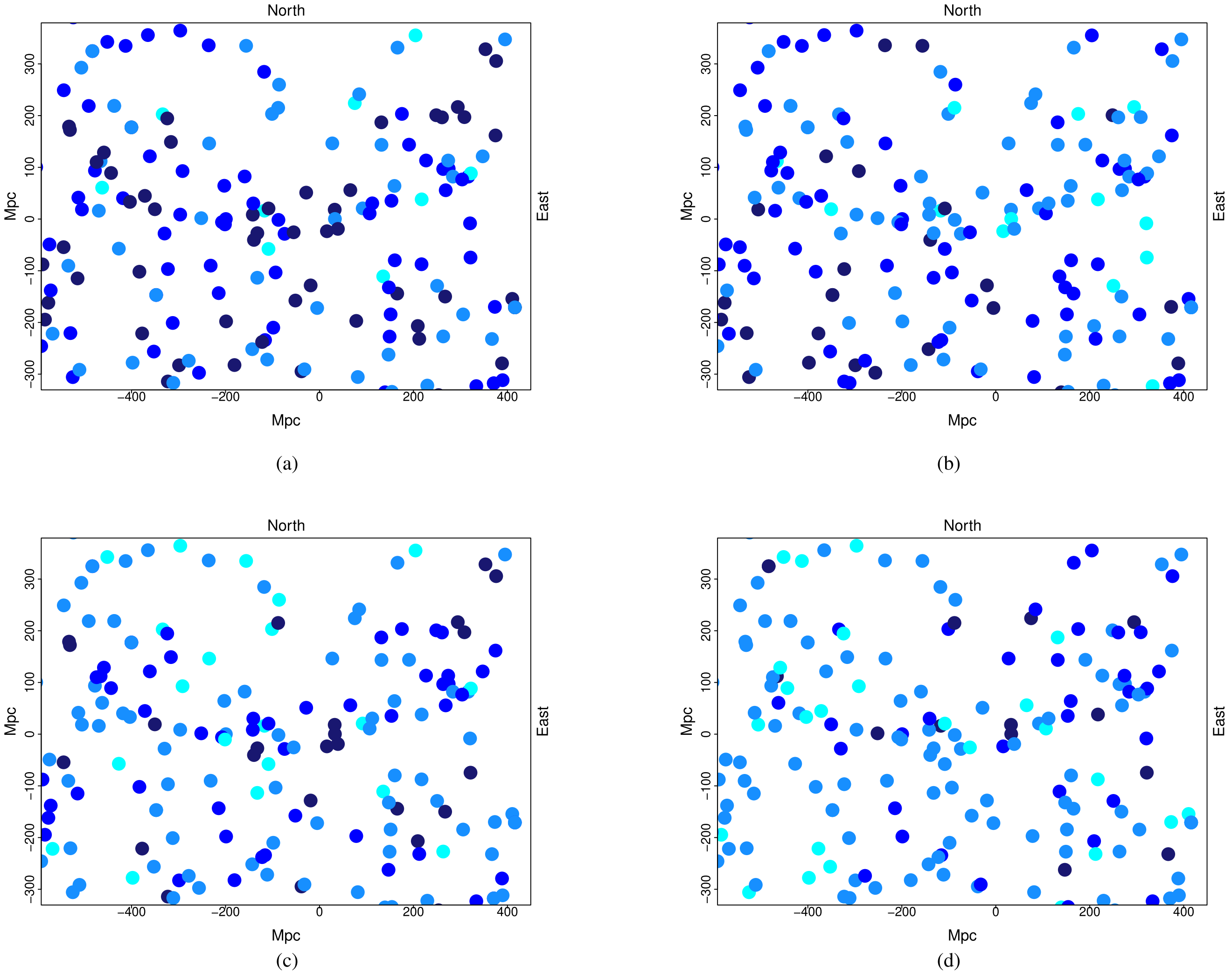
Similarly, with the same set of four blue shades, we show the S/N of the λ2796 Mg II line, the S/N of the continuum, and the *i* magnitude of the probes (see Figs 2b–d). The boundaries of the bins are as follows: (i) S/N of the λ2796 Mg II line – 0, 3, 6, 12, 37; (ii) S/N of the continuum – 0, 8, 16, 24, 47; and (iii) *i* magnitude of the background quasars – 16.0, 17.8, 18.7, 19.6, 21.0. The colour coding again represents the smallest values by the lightest shade of blue, and the largest values by the darkest shade. Note, that for *i*, the lightest shade thus represents the brightest probes.

The EW of the λ2796 Mg II line should correlate with the S/N of the λ2796 Mg II line; Figs 2(a) and (c) show that this is indeed the case. The brightness (*i*) of the background quasar should correlate with the S/N of the quasar continuum; Figs 2(b) and (d) show that this is indeed the case. Note that brighter quasars, having a higher continuum S/N, can detect absorbers to a lower threshold EW.

An asymmetry is apparent in the distribution of Mg II EWs within the GA: the EWs tend to be stronger on the LHS (lower RA), and in the centre of the GA, than on the RHS (higher RA). Conversely, there is a tendency for the RHS to have brighter probes and higher continuum S/N, so, on the RHS, the threshold EW for detection should tend to be lower and certainly able to detect stronger absorbers, should they be present. Thus, the collected observations are consistent with the reality of the observed asymmetry of stronger absorbers on the LHS.

As discussed earlier, in Section 1.2, there have been many attempts to understand the relationship between EW and galaxy properties (morphology, luminosity, impact parameter, galaxy inclination, etc.), but so far without a clear understanding of the connections between them. Conceivably, the asymmetry in the EW distribution could arise from the details of the geometry of the GA and the orientations of the galaxies within it. Future sky surveys and targeted observations seem likely to be necessary for progress on these details.

We note that there appears to be a preference for the strongest ($W_{r,\,2796}$) Mg II absorbers in the GA to clump together into groups of a few. See the dark blue points in Fig. 2(a), and note in particular those on the LHS of the GA (lower RA), the centre of the GA, and the group just above the tip of the RHS (higher RA) of the GA. As the GA is denser than the rest of the field, we can speculate that the



(a)

(b)

(c)

(d)

Figure 2. The on-sky spatial coordinates of the absorbers in the GA and its immediate field, with colour coding according to the distribution of equivalent width (EW) of the Mg II λ2796 line ($W_{r,\,2796}$); quasar brightness (i); S/N of the λ2796 Mg II line; S/N of the quasar continuum. The lightest shades of blue represent the smallest values (for i, this means that the lightest shades represent the brightest quasars). (a) The on-sky spatial coordinates of the absorbers in the GA and its immediate field, with colour coding according to the four equivalent width (EW) bins of the λ2796 Mg II line ($W_{r,\,2796}$). The four EW bins are divided: 0.0, 0.3, 0.6, 1.0, 10.0 Å, with the lightest shades of blue representing the smallest values. There are more strong Mg II absorbers on the LHS (lower RA) of the GA than the RHS. Also there appears to be a tendency for the strong Mg II absorbers in the GA to clump into groups of a few. (b) The on-sky spatial coordinates of the absorbers in the GA and its immediate field, with colour coding according to the four i magnitude bins of the background quasars (probes). The i bins are divided: 16.0, 17.8, 18.7, 19.6, 21.0, with the lightest shades of blue representing the smallest values (this means that the lightest shade thus represents the brightest quasars). There appear to be more bright quasars on the RHS (higher RA) of the GA than the LHS. (c) The on-sky spatial coordinates of the absorbers in the GA and its immediate field, with colour coding according to the four S/N bins of the λ2796 Mg II line. The four S/N bins are divided: 0, 3, 6, 12, 37, with the lightest shades of blue representing the smallest values. By comparing Fig. 2(a) with the figure here, one will notice that the two properties are correlated. (d) The on-sky spatial coordinates of the absorbers in the GA and its immediate field, with colour coding according to the four S/N bins of the quasar continuum. The four S/N bins are divided: 0, 8, 16, 24, 47, with the lightest shades of blue representing the smallest values. By comparing Fig. 2(b) with the figure here, one will notice that the two properties are correlated.

occurrence of the strongest EWs in proximity is not accidental but is connected with the origin and environment of the GA.

Recall that the SZ cluster B18, $z = 0.763$ (Burenin et al. 2018), is at the centre of the GA. (It is what led to the discovery of the GA.) At the centre of the GA is a small, circular 'hole', and surrounding this hole is a group of the stronger absorbers. SZ clusters create a highly ionized environment, but Mg II absorption occurs in low-ionised regions. Possibly, a region of high ionisation can account for the hole, but an origin, in environment, of the enveloping group of stronger absorbers is not then obvious.

The investigation of small ($\Delta z = 0.030$), overlapping (by 50 per cent) redshift slices reveals a noticeable difference between the left-hand and right-hand sides of the GA. For example, the LHS (lower RA) of the GA appears concentrated in the small redshift slice located farthest away ($z = 0.832 \pm 0.030$), whereas the RHS (higher RA) of the GA appears spread diffusely through the larger redshift slice. Interestingly, the LHS of the GA has both a narrower redshift distribution and a preference for stronger Mg II EWs.

Finally, the investigation of the redshift distribution suggests that, if the GA is represented as a segment of a cylindrical shell, then the

LHS would be tilted away along the line of sight. That is, if the GA can indeed be represented as a segment of a cylindrical shell, then it is not precisely orthogonal to the line-of-sight but is rotated with respect to a north–south axis.

2.3 Connectivity and statistical properties

The GA was discovered visually, from a Mg II density image (e.g. Fig. 1). Albeit after the event, we now discuss its connectivity and statistical properties. The Mg II absorbers can, of course, be found only where there are background quasars to act as probes, and those probes may themselves be subject to spatial variations arising from large-scale structure and, in particular, from artefacts in the surveys.

We apply three different statistical methods for assessing the GA, as follows.

(i) SLHC/CHMS (see Clowes et al. 2012). This method depends first on constructing the 3D minimal spanning tree (MST), and then separating it at some specified linkage scale. At this stage, it is equivalent to single-linkage hierarchical clustering (SLHC). The statistical significance of a candidate structure is then assessed using its volume obtained as the volume of the CHMS. Note the important feature that this method assesses the significance of *individual* candidate structures.

(ii) The Cuzick–Edwards (CE) test (see Cuzick & Edwards 1990). It is a 2D 'case-control' method that is designed to correct the incidence of cases for spatial variations in the controls (the underlying population). It depends on the number of cases that occur within the k nearest neighbours. The CE test can detect the presence of clustering in the field, while correcting for variations in the background, and can assess its statistical significance. It cannot, however, assess the physical scale of the clustering.

(iii) 2D Power Spectrum Analysis (2D PSA) (see Webster 1976a). It is a powerful Fourier method for detecting clustering in the field. It can be effective even for detecting weak clustering. The 2D PSA can detect the scale of clustering and assess the statistical significance of the clustering at that scale.

Each of these tests has distinctive attributes, and the reader should judge the evidence provided by the ensemble. Only the SLHC/CHMS method assesses the significance of individual candidate structures, whereas the CE test and the 2D PSA address clustering in the field. We shall describe below the 'polygon approach', in which we assess the contribution that the GA makes to the results from the CE test and the 2D PSA for the field. Only the CE test can correct for spatial variations in the underlying population. However, we shall describe below, again using the polygon approach, that the 2D PSA has more power to discriminate than the CE test.

Finally, we emphasize again, that given the nature of the discovery the statistical analysis is necessarily performed post hoc. The reader will find that we have used techniques to compare the field containing the GA with other, unrelated fields (within the same Mg II data set). This of course has its limitations due to the non-uniformity of the background quasars (probes) and potential survey artefacts. We have also compared with randomized simulations, in which we attempt to preserve these subtleties of the Mg II data.

2.3.1 SLHC/CHMS (minimal spanning tree)

The minimal spanning tree (MST) is a widely used algorithm for assessing large-scale structure in astronomy and cosmology. When the MST is separated at some specified linkage scale it is equivalent to the algorithm for SLHC. An approach to assessing the statistical significance of the agglomerations found in this way was introduced by Clowes et al. (2012): the Convex Hull of Member Spheres (CHMS) method. It was further used by Clowes et al. (2013) in the analysis of the Huge-LQG, the Huge Large Quasar Group that they discovered.

Here, we apply the sequence of SLHC and CHMS to the Mg II absorbers in the GA field. By specifying a linkage scale and a minimum membership, the SLHC identifies the 3D agglomerations or groups within the coordinates of the absorbers. Within each identified group the CHMS constructs a sphere around each member point with a radius of half the mean linkage separation for that group. A volume for the group is then computed as the volume of the convex hull of its member spheres (and note that the convex hull is a unique construction). An expected density of absorbers is determined from a control field and the observed redshift interval of a group. (Here, the control field is specified as the same field that is being assessed.) The observed number of member points within a group are then scattered randomly within a cube at the expected density, and their CHMS volume is calculated; this is done 1000 times. The significance of the group is calculated by the rate of occurrence of randomly generated CHMS volumes that are smaller than the observed volume. See Clowes et al. (2012) for full details of the CHMS method.

In principle, this SLHC/MST approach should be applied only to surveys that have no intrinsic spatial variations. The background quasars – the probes of the Mg II absorbers – are drawn from a merger of the SDSS DR7QSO and DR12Q data bases. While a reasonably spatially uniform subset can be extracted from DR7QSO, DR12Q is much more strongly affected by spatial artefacts arising from deeper areas. Thus, the distribution of the background quasars can conceivably affect the distribution of the Mg II absorbers in some, possibly complicated, way. However, if the distribution of the background quasars appears to be reasonably homogeneous in the area of interest, then we can assume that the distribution of Mg II absorbers is predominantly a product of the LSS and not the availability of background quasars. Of course, the distribution of background quasars can still have some effects – such as occasional gaps in connectivity – on the Mg II absorbers even in such reasonably homogeneous regions.

Fig. 3 shows the kernel-smoothed distribution ($\sigma = 11$ Mpc, present epoch) of the background probes (quasars) in the area of the GA for $z > 0.862$. It is clear that there are denser areas, less dense areas, and even empty patches, across the whole image, indicating the spatial non-uniformity of the background probes. There is a particularly dense band in approximately the northern third, which arises from a deeper area of the DR12Q survey. However, there are evidently no artefacts that correspond to the dimensions and orientation of the GA.

We are now taking the GA to be predominantly concentrated in the redshift interval 0.802 ± 0.060, so $0.742 \rightarrow 0.862$. Its (present-epoch) depth is then $\sim$340 Mpc.

This redshift interval appears to be the optimum, following a heuristic process of stepping through a range of redshift intervals and determining the membership and significance of the GA through the SLHC/CHMS method. Redshift intervals of thickness $\Delta z = \pm 0.050$ between $0.760 < z < 0.810$ were tested, using various linkage scales, and a clear peak signal for the GA was seen around the redshift $z = 0.810$ using a linkage scale of 95 Mpc (see Section 2.3.2 for the details of choosing the optimum linkage scale). Further, finer-scale, testing revealed the greatest number of connected GA members to be more precisely located at $z = 0.802$, again for the linkage scale of 95 Mpc. The significance for $z = 0.802 \pm 0.050$ is 4.15σ. Widening

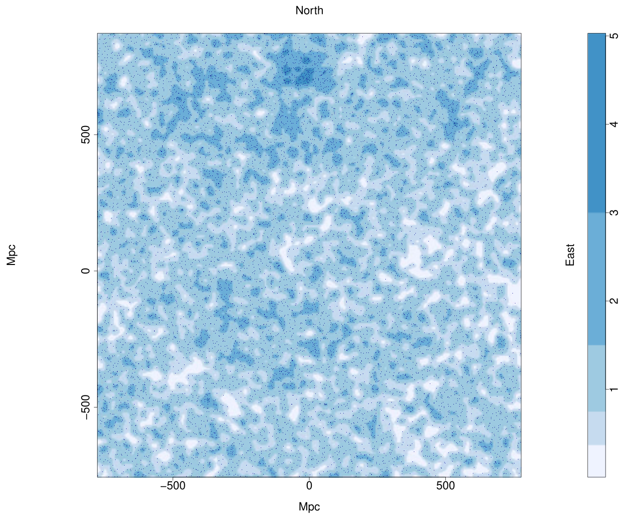


Figure 3. The tangent-plane distribution of background probes (quasars) in the GA field with redshifts $z > 0.862$, represented by the dark blue points. The blue contours, increasing by a factor of 2, represent the density distribution of the quasars using a Gaussian kernel with a smoothing scale of $\sigma = 11$ Mpc. The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top. It is clear that there are denser areas, less dense areas, and even empty patches, across the whole image, indicating the spatial non-uniformity of the background probes. There is a particularly dense band in approximately the northern third. However, there are evidently no artefacts that correspond to the dimensions and orientation of the GA.

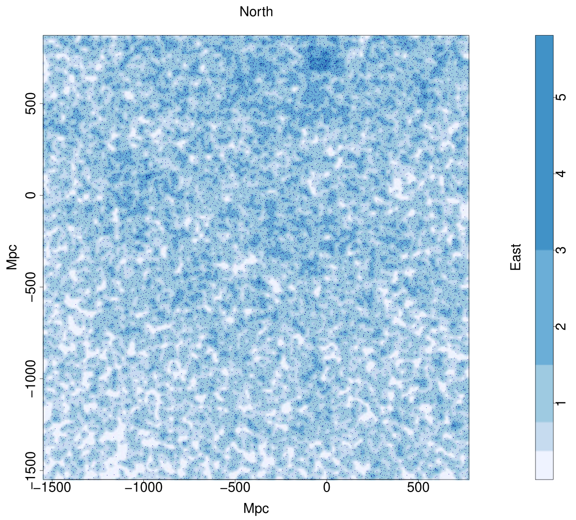


Figure 4. The tangent-plane distribution of background probes (quasars), in the same region of sky as the GA for redshifts $z > 0.862$, represented by the dark blue points. The field-of-view has here been increased by extending the western and southern boundaries. The blue contours, increasing by a factor of 2, represent the density distribution of the quasars using a Gaussian kernel with a smoothing scale of $\sigma = 11$ Mpc. The axes are labelled in Mpc, scaled to the present epoch. East is towards the right and north is towards the top.

the redshift interval to $z = 0.802 \pm 0.060$ gives a slightly higher significance of 4.30σ, and the number of member Mg II absorbers is then increased from 42 to 44.

A second, smaller, agglomeration made up of 10 and 11 absorber members at both $z = 0.802 \pm 0.050$ and $z = 0.802 \pm 0.060$ respectively, although not formally significant (1.75σ and 2.04σ, respectively), is clearly also part of what we identified visually as the GA. Conceivably, just one further background probe would be sufficient to yield one further absorber that would then connect both agglomerations as one significant unit. We have emphasized the limitations of the SLHC/CHMS method for this data set, and we might here be seeing their consequences.

As noted above, the estimation of the CHMS significance requires a control field from which the expected average density is calculated. For the CHMS significances given above, we used the same field as that containing the GA. This was a deliberate choice, given the spatial variations of the wider survey. Clearly, the GA must then represent a small fraction of the total area and number of absorbers (∼ 7 per cent of the absorbers are from the GA).

Even so, the distribution of background probes (quasars) in the field of the GA is not uniform – notably the denser band in the northern third. This non-uniformity could affect the CHMS calculations of significance, either by overestimating or underestimating, depending on whether the probes are generally underpopulated or overpopulated in the control field. A second estimate of the significance can be calculated from the CHMS method by increasing the field-of-view (FOV) containing the GA, and using it as a new control field. Fig. 4 shows the background probes in the field containing the GA with the western and southern boundaries extended. Note that the eastern and northern boundaries were not extended because of proximity to the edge of the survey area.

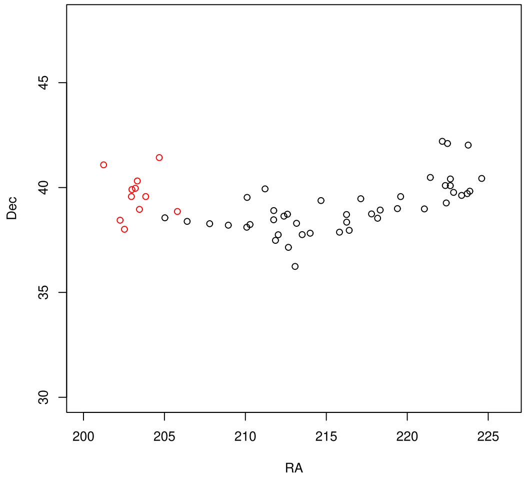


Figure 5. The GA connected via the SLHC algorithm. There are 44 black points which indicate the principal agglomeration of the GA – the largest, and most significant agglomeration in the field –, which comprises the majority of the GA. There are 11 red points which indicate the secondary agglomeration of the GA, which, although by itself not statistically significant, is clearly part of the GA. The axes are labelled RA and Dec where east is towards the right and north is towards the top.

Using the *larger* FOV in the redshift interval $z = 0.802 \pm 0.060$, the CHMS method calculates a significance of 4.53σ for the principal agglomeration of the GA. As noted previously, the GA is split into two agglomerations by the SLHC algorithm, shown in Fig. 5. For this entire, larger, FOV, there are 35 agglomerations in total, with the principal agglomeration of the GA being the largest and most significant, and the only agglomeration with a significance $>3.5\sigma$. Fig. 5 shows the GA as located by the SLHC/CHMS method, with the principal agglomeration represented by the black points; the red

Table 2. Results from applying the SLHC/CHMS algorithm to three lower redshift slices in the same (larger) FOV as the GA (see Fig. 4). The columns are: the redshift interval of the Mg II absorbers; the redshift of the background probes (quasars); the number of structures found by the SLHC algorithm; the number of structures exceeding 3σ significance; the number of structures exceeding 4σ significance; the number of structures exceeding 3σ significance after removing those that reside in artefacts of the probes. Note that the three lower redshift slices have two sets of probes with different redshifts: one for the probes corresponding to those in the GA field, and one for the probes corresponding to the redshift slice.

Redshift slice	Probes	N total	$N \geq 3.0\sigma$	$N \geq 4.0\sigma$	N not rejected as artefacts
$z = 0.802 \pm 0.060$	$z > 0.862$	35	4	1 (GA)	2 (incl. GA)
$z = 0.682 \pm 0.060$	$z > 0.862$	31	3	1	3
	$z > 0.742$	34	6	2	4
$z = 0.562 \pm 0.060$	$z > 0.862$	17	6	3	2
	$z > 0.622$	19	7	4	3
$z = 0.442 \pm 0.060$	$z > 0.862$	9	1	0	0
	$z > 0.502$	12	2	0	0

Notes. (1) For each redshift slice the overall density of absorbers in the field varies, so we modify the linkage scale according to the relation $s = (\rho_0/\rho)^{1/3} \times 95$ Mpc, where s is the linkage scale, ρ_0 is the density for the GA field and ρ is the density for the new field concerned.

(2) The number of structures reduces in each successively lower redshift slice, because the number of Mg II absorbers reduces as the lower-wavelength limit of detectability is approached.

points indicate the smaller, separate agglomeration of much lower significance, but visually it can clearly be seen as part of the GA.

The SLHC/CHMS algorithm is next applied to three lower redshift slices – $z = 0.682 \pm 0.060$, $z = 0.562 \pm 0.060$, and $z = 0.442 \pm 0.060$ – in the same (larger) FOV as the GA (see Fig. 4). Using lower redshift slices and the same FOV means that we can conveniently compare the Mg II absorbers arising from the same probes as those in the GA field by restricting the probes to $z > 0.862$. We can also apply the SLHC/CHMS method to the same three lower redshift slices *without* this redshift restriction on the probes. Recall, however, that the probes may show density artefacts with a large FOV. As mentioned previously, the SLHC/CHMS method can be problematic for the Mg II analysis because of these artefacts. Therefore, superficially significant structures that correspond to particularly dense patches of probes are likely to be discarded. The results are summarized in Table 2.

As can be seen in Table 2 there are significant ($\sim 3\sigma$) structures. At a more cautious limit of 4σ, however, there were only two candidate structures which did not reside in an artefact (in redshift slices $z = 0.682$ and $z = 0.562$). We shall in due course investigate them further, starting with optimisation of the redshift intervals.

Finally, we introduce a random-simulation aspect to the SLHC/CHMS analysis. We have carried out 1000 random simulations as follows. (i) We consider the large, extended area that corresponds to (Fig. 4). (ii) We consider only the probes at higher redshift than the redshift slice of the GA – that is, we continue (as with the slices of redshift lower than that of the GA) to use only the probes appropriate to the GA, so that density artefacts in the probes remain identical. (iii) We reassign at random Mg II absorbers of any redshift to the probes, while not splitting occurrences of multiple absorbers per line of sight. (Note that splitting absorbers would have the undesirable effect of changing the total number of probes with absorbers.) (iv) We then analyse the random-simulated data as for the actual GA slice, selecting absorber redshifts for the redshift slice.

Within the simulations, we looked for 'structures' that had properties comparable to, or more extreme than, the observed properties of the GA (precisely, of GA-main – see below). The properties considered were the set of: number of members; SLHC/CHMS significance; and overdensity. In all cases (roughly one occurrence per two simulations), we found that these 'comparable structures' were in the regions of the visually-obvious density artefacts, and never in the region occupied by the real GA. The occurrence of the 'comparable structures' in the density artefacts is as expected: for those artefacts, the linkage scale and the control density would clearly not be appropriate. We can infer that the probability of the real GA (precisely, GA-main) occurring as a random event is <0.001.

The SLHC algorithm easily identifies the GA, with 44 connected Mg II absorbers, and the CHMS method estimates a significance of $\sim 4.5\sigma$ using the central redshift $z = 0.802$. In every redshift interval investigated, the GA appears in two parts: (i) the principal agglomeration, which is large in both physical size and membership, and statistically very significant; and (ii) the secondary agglomeration, small in size and membership, and by itself statistically not significant. As mentioned earlier, the Mg II absorbers depend on the availability of background probes (quasars), and without those, Mg II would not be detected. Thus an artefact in the distribution of probes – i.e. a gap, perhaps of just one missing probe – could lead to an artefact of apparent splitting into two agglomerations.

We have seen previously, in Section 2.2, that there is a noticeable difference between the LHS and RHS of the GA with regards to redshift distribution. Investigating small ($\Delta z = 0.030$), overlapping (by 50 per cent) redshift slices has highlighted the sub-structure of the GA along the redshift axis. We find that the larger agglomeration of the GA is distributed more evenly and widely along the redshift axis, while the smaller agglomeration is concentrated in a narrower redshift slice. It becomes clear from the central redshift slice and below ($z < 0.802$) that there are no Mg II absorbers available that can connect the small agglomeration to the large agglomeration.

More data, such as the new SDSS DR16Q quasar database (Lyke et al. 2020), could provide additional information to investigate the GA further. This includes, conceivably, the possibility of connecting the small agglomeration to the large agglomeration of the GA. However, this would require construction of a new Mg II catalogue corresponding to the DR16Q quasars. In the future, we plan to create our own Mg II catalogues from previous and future quasar data releases.

2.3.2 Selecting a Linkage Scale

The linkage scale that is set in the SLHC/CHMS method determines both the number of agglomerations and their memberships. It was set at 95 Mpc for the GA. This setting was partly guided by the linkage scale that was known to be effective for the CCLQG, and which, when used subsequently, led to the discovery of the Huge-LQG and many other LQGs (see Clowes et al. (2012) and Clowes et al. (2013)). Clearly, the linkage scale must be adjusted for field density; in this case, starting from the linkage scale that was effective for LQGs (i.e. for LSS in quasars) we calculate a linkage scale of 85 Mpc for Mg II absorbers in the GA field. From here we followed the heuristic process described in Section 2.3.1 to identify an optimum linkage scale of 95 Mpc for the Mg II absorbers.

One must remember that Mg II absorbers are distinctly different from quasars and therefore cannot be treated in quite the same way. For example, the linkage scale that works for quasars will not necessarily work for the Mg II absorbers since the latter is a case of inhomogeneities (the absorbers) superimposed on inhomogeneities (the quasars and survey artefacts). Future work will address the development of a clustering analysis that is specifically addressed to the requirements of Mg II absorbers.

It is in the nature of discoveries that there will be a post hoc aspect to the analysis. What turned out to be effective for the CCLQG led to the discovery of the Huge-LQG: an initial discovery, followed by a heuristic process, followed by an entirely objective a priori new discovery. It is in this spirit that we present the discovery of the GA: one for which the techniques and parameters used to assess and characterise it can subsequently be applied to the whole Mg II database.

For completeness, we can briefly mention what results from instead setting the linkage scale to 85 Mpc, 90 Mpc, 100 Mpc, and 105 Mpc for the adopted redshift slice. In five runs, using linkage scales of 85 Mpc, 90 Mpc, 95 Mpc, 100 Mpc, and 105 Mpc, there are totals of 4, 25, 35, 43, and 3 agglomerations found respectively, with the GA always being the most significant in all except the 105 Mpc run. The middle three runs split the GA into two parts – one large and significant, and one smaller and less significant. For the following comments, we concentrate only on the large, significant part of the GA located in the middle three runs (90, 95 and 100 Mpc), which makes up the majority of what we visually identified as the GA. (1) It has a significance greater than 3.8σ in all of the three runs. (2) It is the only agglomeration that has a significance greater than 3.5σ, with only two or three agglomerations above 3σ (all others being below a 3σ threshold). (3) In both the 90 Mpc and 95 Mpc runs, it is the largest agglomeration by membership, and is the second largest by membership in the 100 Mpc run. Lastly, we mention what arises from setting the linkage scale to 85 Mpc and 105 Mpc. Using these linkage scales one can see that the SLHC / MST method has reached its maximum and minimum limit with the linkage scales, as there are only 4 and 3 structures found, respectively. The corresponding memberships are 20 and 133 with significances of 3.6σ and 0.8σ, respectively. As is demonstrated, going any lower or higher with the linkage scale would yield nothing of consequence: either no structures, or one large structure containing almost everything. It is worth noting that, although the 85 Mpc linkage scale is the minimum at which structures can be found in the Mg II data at this redshift/density, the GA is still mostly detected, still the largest agglomeration in the field, and the only structure detected with a significance over 3σ (see Fig. 6).

Note that for future work, involving the remainder of the Mg II database, we can adopt this scale of 95 Mpc as a standard, with scaling according to the density of absorbers (i.e. $s = (\rho_0/\rho)^{1/3} s_0$) in the volumes of interest.

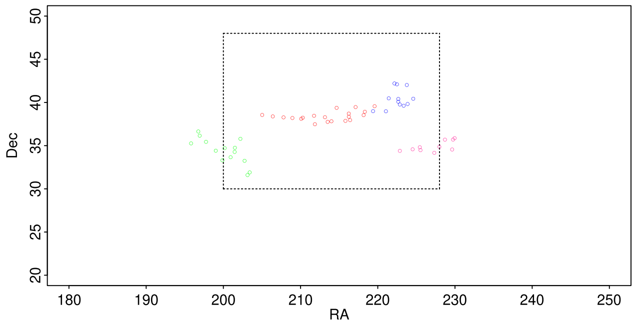


Figure 6. The four structures found via the SLHC/CHMS method using a reduced linkage scale of 85 Mpc. The black dashed rectangle corresponds to Fig. 5 for comparison. The GA is still mostly detected in what appears to be two parts. The largest agglomeration, containing 20 Mg II absorber members represented by the red points in the figure, makes up the majority of what we visually identify as the GA. The second largest agglomeration, containing 14 Mg II absorber members represented by the blue points in the figure, is also part of what we visually identified as the GA and can be seen as an extension of the red points. The axes are labelled RA and Dec where east is towards the right and north is towards the top.

2.3.3 Cuzick-Edwards test

We mentioned above that, strictly, the SLHC / MST approach should be applied only to surveys that have no intrinsic spatial variations. However, a statistical test for (two-dimensional) clustering exists that is designed to manage spatial variations in the source data: the Cuzick-Edwards test (Cuzick & Edwards 1990). We apply it here.

The Cuzick-Edwards test (hereafter CE test) has been used mainly in medical research, such as the clustering patterns of diseases within unevenly populated geographical regions. (The essential character of our problem is the same.) It adopts a 'case-control' approach to a k nearest-neighbour (NN) analysis. Several papers have compared the properties of the CE test amongst various spatial clustering analyses and assert that the CE test is powerful and sensitive in estimating clustering significance within a point data set (see e.g. Song & Kulldorff 2003; Hinrichsen et al. 2009; Meliker et al. 2009). In Song & Kulldorff (2003), the authors note that the CE test is used more appropriately if the level of clustering is known beforehand.

Inevitably, for our problem, the statistical properties of the GA are tested after the event of discovery (i.e. the level of clustering is known).

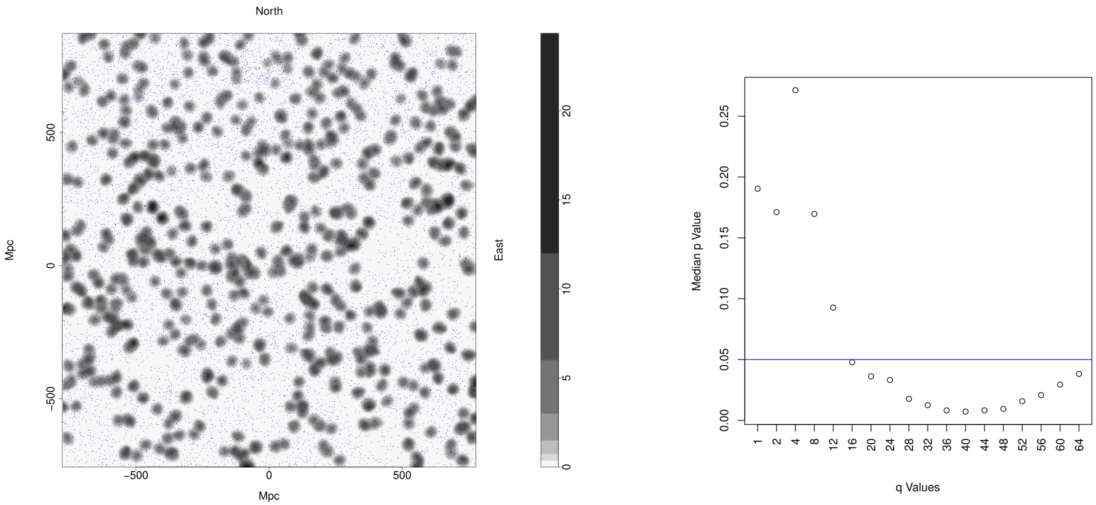
We used the CE test that is coded in the application *qnn.test* in the R package SMACPOD (*Statistical Methods for the Analysis of Case-Control Point Data*; French 2020). The probes (i.e. the background quasars) are labelled as 'controls' and the Mg II absorbers in the redshift interval are labelled as 'cases'. The *qnn.test* then uses a NN algorithm to find the q (or k) NNs of any case to another case.

The test statistic is then calculated as

$$T_k = \sum_{i=1}^{n} \delta_i d_i^k$$

where

$\delta_i = 1$ if the data point is a case or 0 if it is a control;



(a)

(b)

Figure 7. (a) Density distribution of the flat-fielded Mg II absorbers represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). (b) The median p-value over 100 runs of 2000 simulations as a function of chosen q (k) value in the GA field; see the adjacent image. The blue, horizontal line is set to $p = 0.05$. The p-value is at a minimum of 0.0072 when q is 40.

$d_i^k = 1$ if the NN is a case and 0 if it is a control.

The p-value from *qnn.test* is calculated from simulations under the random-labelling hypothesis (French 2020) for nsim = 2000 simulations.

The choice of maximum q (k) value that is adopted for the test will depend on the control-case ratio, as can be seen from the test statistic calculation. There are ∼20 times as many probes (controls) as Mg II absorbers (cases) in the redshift interval of the GA. Cuzick & Edwards (1990) examine the power of the CE test with varying control-case ratios and find that a control-case ratio of between 4 and 6 is optimum (see their fig. 5).

Therefore, we choose to use a control-case ratio of ∼ 5:1. To achieve this we randomly select 25 per cent of the probes, for each of 100 runs of *qnn.test*. (Randomly-selected controls that duplicate the coordinates of the cases in a given run are removed.) The 100 runs also allow us to assess how robust are the estimates of significance for the Mg II absorbers.

We use a set of q (k) values: 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, and 64. We start by applying the *qnn.test* to the basic GA field.

Then, to assess the (presumed) dominance of the GA itself we apply the test to a succession of smaller fields (smaller in the north–south direction), all centred on the GA. In Figs 7–9, the median p-value over 100 runs of 2000 random simulations is shown plotted against the chosen q values, with the corresponding flat-fielded Mg II image shown alongside.

The process of zooming into the GA allows the GA to become the dominating feature in the field, which, as a result, increases the significance of clustering (i.e. smaller p-value). In the first Mg II field, Fig. 7(a), the minimum p-value is 0.0072 at a q-value of 40, which (assuming a normal distribution) is equivalent to a significance of $\sim 2.68\sigma$, Fig. 7(b). Whereas in the third Mg II field, Fig. 9(a), the minimum p-value drops to 0.0027 at a q-value of 40, which is equivalent to a significance of 3.00σ, Fig. 9(b). In this way we can judge that the GA is the dominant, contributing factor to the significant level of clustering in the field.

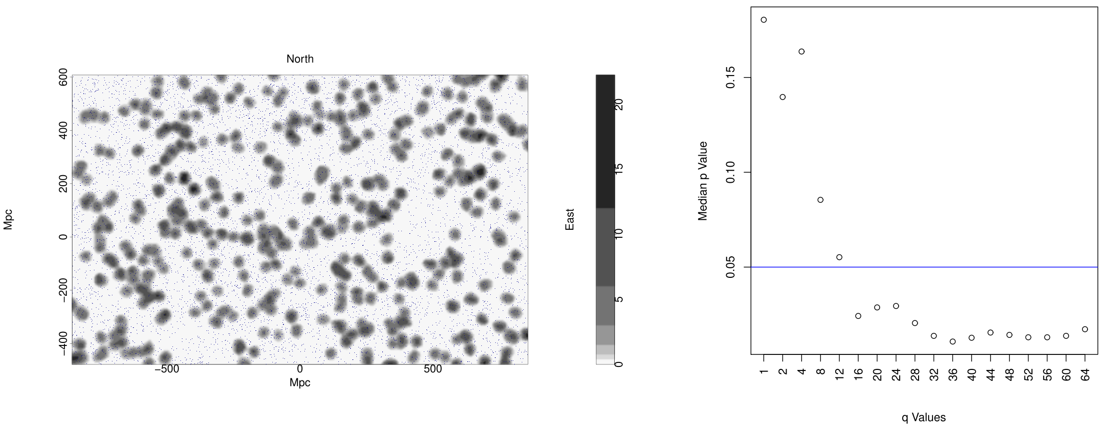
The heuristic process of 'zooming' into the GA was next applied to three other fields at lower redshift slices (z: 0.682, 0.562, 0.442) centred on the sky coordinates of the GA. The background probes are kept the same in the three new fields as those in the GA field, allowing a direct comparison of clustering in just the Mg II absorbers (as in Section 2.3.1). Figs 10–12 show the results of the CE test for the three lower redshift fields using the smallest field size (i.e. the second 'zoom'). The p-value profiles as a function of q-value in each of the lower redshift fields appear more scattered and varied compared with the GA results.

The major difference between the p-value profiles for the GA field and the p-value profiles for three lower redshift fields is that there is no sign of any significant results (p-value <0.05) in any of the three lower redshift fields for any of the chosen q values. The background probes were the same in all four fields (GA field and the three lower redshift fields), indicating that the Mg II absorbers are responsible for the different p-value profiles. From this we can assert that the GA field is markedly distinct, with significant clustering attributable to the GA.

As a further test of the dominance of the GA in the CE statistics, we have applied our 'polygon-approach'. Visually, we draw a polygon around what we identify visually as the member absorbers of the GA. We leave the absorbers in the polygon untouched but reassign (i.e. shuffle) at random the y-coordinates of absorbers outside the polygon, while avoiding the area within the polygon. We apply this process to the data of Fig. 9(a). In this way, we can compare the CE statistics arising from the original data with those in which the GA points inside the polygon are unchanged but those outside the GA polygon are randomized. We find that the range of p-values for the original data (p-values ∼ 0.002–0.003) is very similar to that of the GA + randomized data (p-values ∼ 0.001–0.003), suggesting that the GA is indeed the dominant source of the clustering signal.

2.3.4 *Power spectrum analysis*

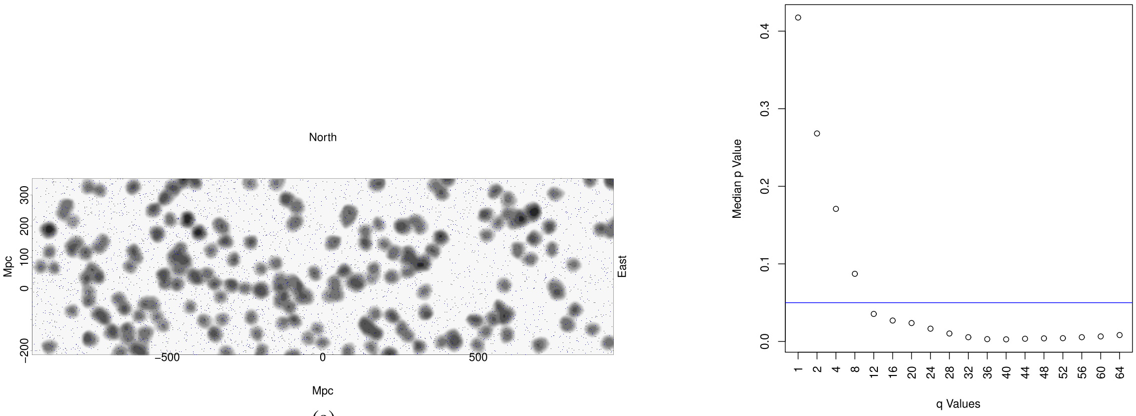
Power spectrum analysis (PSA) – see mainly Webster (1976a), but also Webster (1976b) and Webster (1982) – is a powerful Fourier method for assessing the presence and significance of clustering in



(a)

(b)

Figure 8. (a) Density distribution of the flat-fielded Mg II absorbers represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). This is the first 'zoom' of the GA, where the GA field has been reduced in the north and south boundaries. (b) The median *p*-value over 100 runs of 2000 simulations as a function of chosen *q* (*k*) value in the first 'zoomed' field containing the GA; see the adjacent image. The blue, horizontal line is set to $p = 0.05$. The *p*-value is at a minimum of 0.0107 when *q* is 36.



(a)

(b)

Figure 9. (a) Density distribution of the flat-fielded Mg II absorbers represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). This is the second 'zoom' of the GA, where the GA field has been further reduced in the north and south boundaries. (b) The median *p*-value over 100 runs of 2000 simulations as a function of chosen *q* (*k*) value in the second 'zoomed' field containing the GA; see the adjacent image. The blue, horizontal line is set to $p = 0.05$. The *p*-value is at a minimum of 0.0027 when *q* is 40.

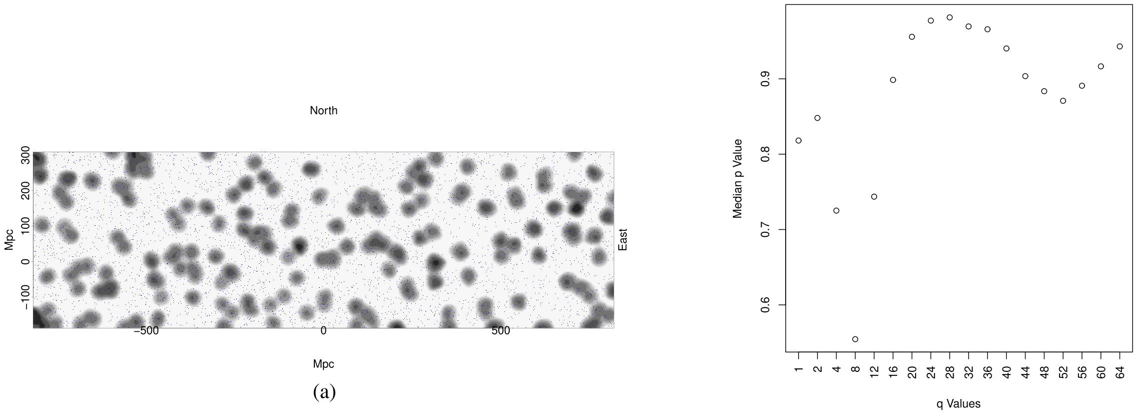
rectangular (2D PSA) or cuboidal fields (3D PSA). PSA was designed to be effective for the detection of clustering that may be weak and escape detection by other methods; it is, however, not a case-control method. A brief summary of the theory of PSA may be found in section 5 of Clowes (1986).

We apply 2D PSA to the same rectangular field, illustrated in Fig. 9(a), that was used for the CE analysis above. Fig. 13 shows the plot of the intermediate PSA statistic $Q^{'}$ against $1/\lambda$. The (six) high points towards the left of the plot allow a clustering scale of $\lambda_c \sim 270$ Mpc to be identified. The final PSA statistic *Q* for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ.

We have applied the polygon-approach here also. As in the discussion above for the CE method, we leave the GA absorbers in the polygon untouched but reassign (i.e. shuffle) at random the *y*-coordinates of absorbers outside the polygon, while avoiding the area within the polygon. In this way we can establish that the significance from the 2D PSA of the GA absorbers *alone* – i.e. with other absorbers randomized – has a mean value $\sim 3.5\sigma$ (with a range 3.0–4.4σ) for the original scale $\lambda_c \sim 270$ Mpc. In fact, the value of λ_c for the polygon-approach varies too, with the mean significance at the actual values of λ_c being $\sim 3.8\sigma$.

From this polygon approach, it appears likely that, while the GA is the dominant contributor to the PSA result, a smaller contribution from other absorbers in the field is detected too. This outcome might be expected, given the power of the PSA method. The failure to detect a contribution from the other absorbers with the CE method could be because the CE method is intrinsically less sensitive, or because its case-control correction has successfully eliminated artefacts from the background probes (the controls).

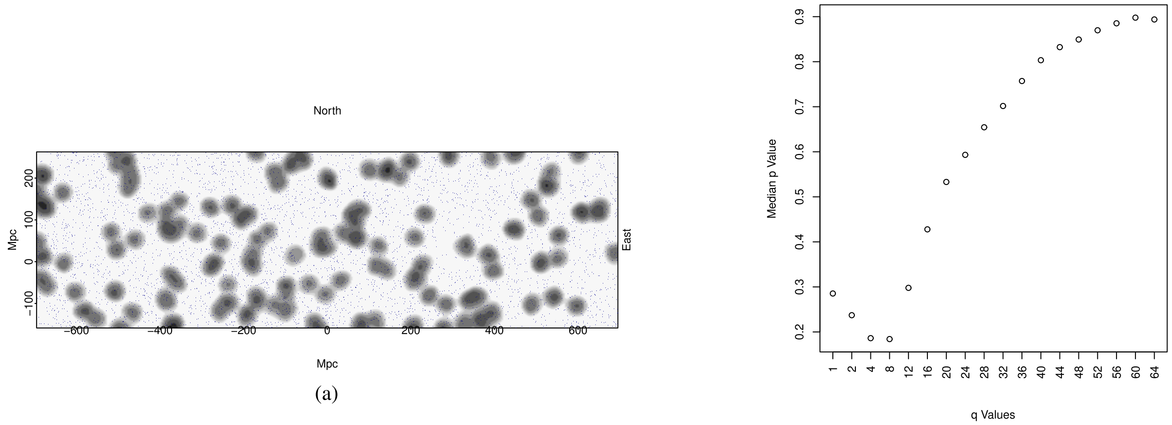
The polygon-approach can also be used to assess the relative power of the 2D PSA and the CE test. For example, we reduced the number of GA absorbers in the polygon from 52 to 42 by random selection (with the points outside the polygon being randomized as usual but unchanged in total number). In that case, the GA is generally not detected by the CE test, at a significance level of 0.01 (2.3σ), but is generally still detected by the PSA, at $>2.7\sigma$. The power



(a)

(b)

Figure 10. (a) Density distribution of the flat-fielded Mg II absorbers in the redshift interval $z = 0.682 \pm 0.060$, on the same sky coordinates of the 'zoomed' GA field, represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). (b) The median p-value over 100 runs of 2000 simulations as a function of chosen q (k) value in the $z = 0.682 \pm 0.060$ redshift interval on the same sky coordinates as the 'zoomed' GA field; see the adjacent image.



(a)

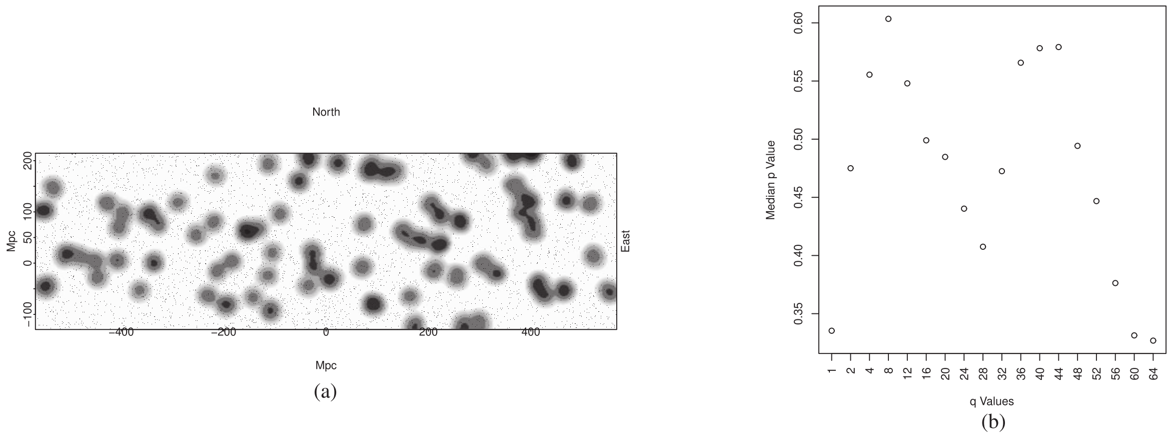
(b)

Figure 11. (a) Density distribution of the flat-fielded Mg II absorbers in the redshift interval $z = 0.562 \pm 0.060$, on the same sky coordinates of the 'zoomed' GA field, represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). (b) The median p-value over 100 runs of 2000 simulations as a function of chosen q (k) value in the $z = 0.562 \pm 0.060$ redshift interval on the same sky coordinates as the 'zoomed' GA field; see the adjacent image.

of a statistical test to discriminate is an important factor, and an uninteresting p-value does not necessarily mean nothing interesting in the data. Webster (1976a) demonstrates that the PSA has more power to detect clustering than a simple nearest-neighbour test. The CE test uses multiple neighbours, and so can be expected to have more power than a nearest-neighbour test, but, as our tests with the polygon-approach suggest, still has less power than the PSA. Of course, the CE test has the useful feature of case-control comparisons, whereas the PSA does not.

2.4 Overdensity

As seen previously, the SLHC/CHMS method splits the GA into two agglomerations – one large, statistically significant portion which makes up the majority of what we visually identified as the GA, and one smaller, statistically not significant portion that makes up the remainder of what we visually identified as part of the GA. We will refer to these agglomerations as GA-main and GA-sub, respectively, for simplicity. The overdensity of the GA can be calculated using the CHMS approach as described earlier in the paper (Section 2.3.1). However, in the case of a strongly curved structure such as the GA (and GA-main), the MST-based method of Pilipenko (2007) can have some advantages. We shall refer to these two methods as CHMS-overdensity and MST-overdensity, respectively. The MST-overdensity does not consider the physical volume of the structure being assessed. Instead, it calculates the overdensity based on the MST edge-lengths: $\delta = \langle l_0^3 \rangle / \langle l^3 \rangle - 1$ where l is MST edge-length for the structure and l_0 is that for a control field. Given the curvature of GA-main, the CHMS volume and CHMS-overdensity refer to a volume that encloses both GA-main and some space above it (where there are rather fewer absorbers, and those are not related to the GA). Therefore, the CHMS method is likely to overestimate the volume and underestimate the overdensity. In contrast, the MST-overdensity, which is an internal measure that considers only the points belonging to the group, and no additional space arising from curvature, is likely to be a better estimate of the overdensity. Conversely, GA-sub is a globular shape, so it is possible to construct a unique volume enclosing only the absorbers attached to GA-sub and not additional



(a)

(b)

Figure 12. (a) Density distribution of the flat-fielded Mg II absorbers in the redshift interval $z = 0.446 \pm 0.060$, on the same sky coordinates of the 'zoomed' GA field, represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue dots represent the background probes (quasars). (b) The median p-value over 100 runs of 2000 simulations as a function of chosen q (k) value in the $z = 0.446 \pm 0.060$ redshift interval on the same sky coordinates as the 'zoomed' GA field; see the adjacent image.

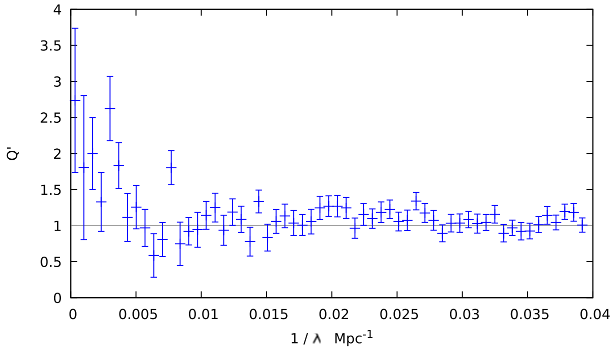


Figure 13. The plot shows the PSA statistic $Q^{'}$ plotted against 1/λ, with λ in Mpc, for the 2D PSA. The bin size is 6.7×10^{-4} Mpc^{-1} and the error bars are $\pm\sigma$. The horizontal line $Q^{'} = 1$ indicates the expectation value in the case of no clustering. The (six) high points towards the left of the plot allow a clustering scale of $\lambda_c \sim 270$ Mpc to be identified. The final PSA statistic Q for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ.

ones at lower density. Therefore, the CHMS-overdensity calculation for GA-sub is likely to be a fair estimate.

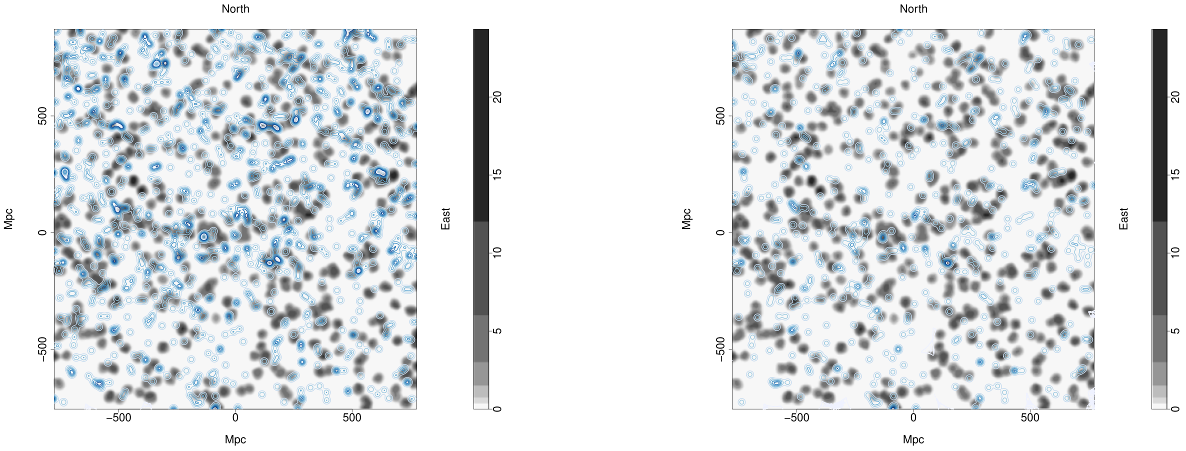
GA-main, as mentioned earlier, has a significance of 4.5σ, while GA-sub has a much smaller significance of 2.3σ. By splitting the usual control field (using the *larger* field-of-view, see Fig. 4) into eight portions – four quarter segments and four half segments – we repeatedly calculate the significance and the overdensities of GA-main and GA-sub using different control fields. An uncertainty can then be estimated for the significance and overdensities for both portions of the GA. Our results are as follows: (1) GA-main, containing 44 Mg II absorbers, has a significance of $(4.5 \pm 0.6)\sigma$; a CHMS-overdensity of $\delta\rho_{\rm CHMS}/\rho_{\rm CHMS} = 0.9 \pm 0.6$; and an MST-overdensity of $\delta\rho_{\rm MST}/\rho_{\rm MST} = 1.3 \pm 0.3$; (2) GA-sub, containing 11 Mg II absorbers, has a significance of $(2.1 \pm 0.9)\sigma$; a CHMS-overdensity of $\delta\rho_{\rm CHMS}/\rho_{\rm CHMS} = 1.5 \pm 0.3$; and an MST-overdensity of $\delta\rho/\rho = 1.3 \pm 0.3$. As expected, the CHMS-overdensity is lower than the MST-overdensity for GA-main, indicating that the CHMS unique volume encapsulating GA-main is likely to be an overestimate because of the curvature of the arc. In contrast, for GA-sub, which has a globular-shape, the CHMS-overdensity and the MST-overdensity have similar values, as expected when there is no marked curvature. In addition, the CHMS-overdensity has a much larger error than the MST-overdensity that suggests giving preference to the latter. Notice here that both GA-main and GA-sub have the same MST-overdensity, which supports their belonging to the same structure.

A final method of calculating the number overdensity is to simply draw a rectangle around the visually selected Mg II absorbers in the GA and compare the number of absorbers per unit area in the rectangle to the number of absorbers in the whole field. The method will underestimate the GA overdensity for three reasons: (i) the GA contributes to the density of the whole field, although only by a small fraction; (ii) the rectangular shape around the GA overestimates the area encompassing the GA as the GA is curved, therefore having a large portion of 'empty' space (with non-GA absorber members); (iii) this method encompasses the *whole* GA from visual inspection, rather than separating it into two agglomerations like the CHMS method, thus reducing the overall density in the GA rectangle. Using this method we calculate an overdensity of $\delta\rho/\rho \sim 0.93$.

In addition to the number overdensity, we can estimate the mass excess by assuming $\delta_n = \delta_m$, where δ_n is the MST-overdensity and δ_m is the mass overdensity. We use here the MST-overdensity, rather than the CHMS-overdensity. We are here taking the critical density of the Universe to be $9.2 \times 10^{-27}\,{\rm kg\,m^{-3}}$, as calculated using the cosmological parameters used throughout this paper, and the matter–energy density parameter to be $\Omega_{M0} = 0.27$. The mass excesses for GA-main and GA-sub are then $1.8 \times 10^{18}{\rm M}_{\odot}$ and $3.4 \times 10^{17}{\rm M}_{\odot}$, respectively. Note that the mass excess of GA-main + GA-sub is comparable to that of the Huge-LQG (Clowes et al. 2013).

2.5 Comparisons with other data

Independent corroboration of a very large LSS by an independent tracer can provide compelling support. In the case of the Huge-LQG (Clowes et al. 2013), a ∼ Gpc structure of quasars, independent corroboration was provided by Mg II absorbers. Here, we can invert this approach and look for corroboration of the GA, a ∼ Gpc structure of Mg II absorbers, in quasars. We use the SDSS DR16Q data base



(a) (b)

Figure 14. (a) Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours that have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue contours represent the DR16Q quasars with $i \le 20.0$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. The GA can be seen stretching across ∼1 Gpc in the centre of the figure (at tangent-plane y-coordinate ∼0 Mpc). Visually, the blue contours can be seen to follow the same general trajectory as the grey contours, indicating an association between the Mg II absorbers and the DR16Q quasars. (b) Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue contours represent the DR16Q quasars, with $i \le 19.5$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. The GA can be seen stretching across ∼1 Gpc in the centre of the figure (at tangent-plane y-coordinate ∼0 Mpc). Visually, the blue contours can be seen to follow the same general trajectory as the grey contours, indicating an association between the Mg II absorbers and the DR16Q quasars.

(Lyke et al. 2020). In addition, we look at the data bases of DESI galaxy clusters from Zou et al. (2021).

We are concerned at this stage with simple visual inspection, and will leave the subtleties of correcting for possible artefacts in the DR16Q quasars and the DESI clusters to future work. Our approach here will be simply to superimpose contours for the spatial distribution of the quasars (in blue) and the clusters (in green) on to the Mg II density images (grey, as previously).

We begin with the quasars, selected for the same redshift interval as the GA (Fig. 14). We show two cases, one for quasars with $i \le 20.0$ (Fig. 14a) and one for $i \le 19.5$ (Fig. 14b). We anticipate that we should then be restricting to ‘traditional’ high-luminosity quasars. In both cases, it is immediately clear that the quasars follow the same general trajectory as the GA. The quasars are entirely unrelated to the probes of the GA, and so we have in these plots quite striking independent corroboration of the GA. Furthermore, the tendency of the Mg II absorbers in general and the quasars to share common paths and voids is apparent, especially so in Fig. 14(b).

Note that there is a density boundary in the distribution of the DR16Q quasars: in roughly the lower third of the plots the density of the quasars is lower than above. This artefact, however, is well separated from the GA and does not affect our visual assessment.

We continue with the DESI clusters, again selected for the same redshift interval as the GA (Fig. 15). Note that the redshifts for the DESI clusters are photometric, with redshift errors ∼0.024 at $z \sim 0.9$ (Zou et al. 2021). (In contrast, we might expect the redshift errors for the quasars to be ∼0.003.)

There is no compelling association of the DESI clusters and the GA, although there is perhaps a hint on the RHS. Possibly the substantial errors in the photometric redshifts are a factor in diluting any correspondence that might exist. An interesting feature in Fig. 15 is, however, the ‘cluster of clusters’ in the centre of the GA, largely coinciding with the central small gap in the Mg II absorbers of the GA. It could be a large supercluster, with the SZ cluster B18, mentioned previously, as one of its member clusters. We previously mentioned, in Section 2.2, that there appears to be a set of strong Mg II absorbers enveloping a circular hole in the centre of the GA. It seems likely that these enveloping strong absorbers and the central hole are related to this putative supercluster.

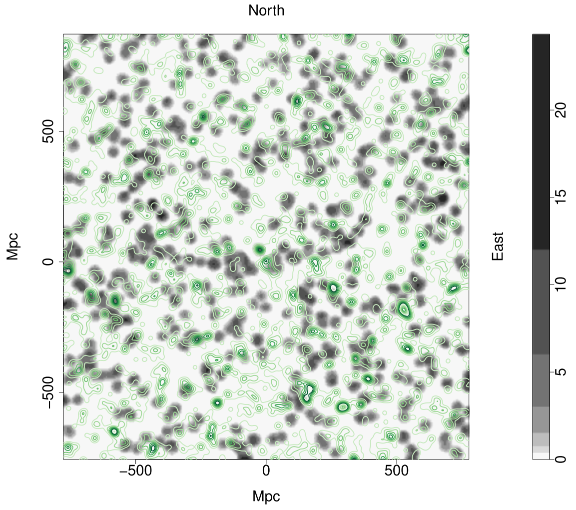


Figure 15. Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Green contours represent the DESI clusters, of all richnesses, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. The GA can be seen stretching across ∼1 Gpc in the centre of the figure (at tangent-plane y-coordinate ∼0 Mpc). There are no compelling connections between the DESI clusters and the Mg II absorbers.

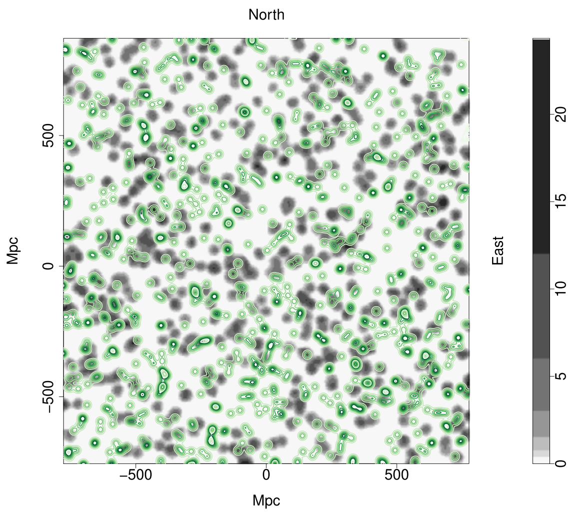


Figure 16. Density distribution of the flat-fielded Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Green contours represent the DESI clusters, with the richness limit $R \leq 22.5$, in the same redshift slice as the Mg II absorbers, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. The GA can be seen stretching across ∼1 Gpc in the centre of the figure (at tangent-plane y-coordinate ∼0 Mpc). There are a few occurrences of the green contours following the grey contours, indicating that there might be some association of low richness clusters with Mg II absorbers.

The mean richness limit for the DESI clusters is 22.5 (Zou et al. 2021). Fig. 16 shows the relationship between the Mg II absorbers and DESI clusters with richness $R \leq 22.5$. It suggests that there could be some association of the low-richness clusters with the Mg II absorbers, both for the GA and in general.

Finally, we compare the the DESI clusters with the DR16Q quasars (Fig. 17). As with the Mg II absorbers and DESI clusters, there appears to be no compelling association. However, again, the lowest richness clusters suggest some association (Fig. 18).

From the independent corroboration above, we suggest that the GA, and the Mg II absorbers in general, are associated with luminous quasars but not strongly with DESI clusters. However, there is potentially an association of the Mg II absorbers and the quasars with the low richness clusters. More statistical details of the relationship between Mg II absorbers, quasars and clusters will be investigated in our future work.

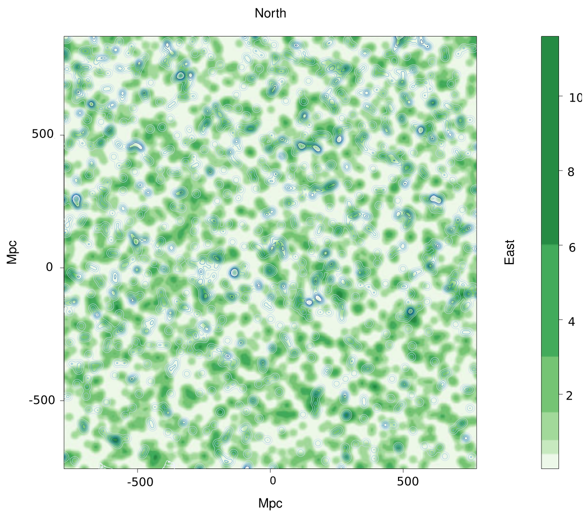


Figure 17. Density distribution of the DESI clusters in the redshift slice $z = 0.802 \pm 0.060$ represented by the green contours that have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increase by a factor of 2. Blue contours represent the DR16Q quasars, with the magnitude limit $i \leq 20.0$, in the same redshift slice as the DESI clusters, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. There are no compelling connections between the DR16Q quasars and the DESI clusters.

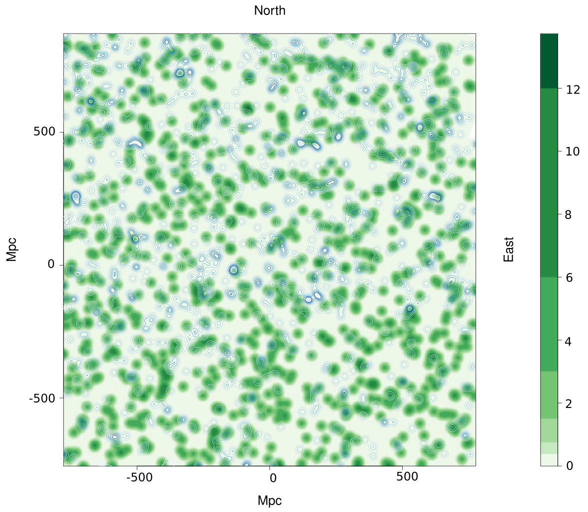


Figure 18. Density distribution of the DESI clusters in the redshift slice $z = 0.802 \pm 0.060$ represented by the green contours, increasing by a factor of 2, which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and limited to show only those clusters with a richness $R \leq 22.5$. Blue contours represent the DR16Q quasars, with the magnitude limit $i \leq 20$, in the same redshift slice as the DESI clusters, smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and increasing by a factor of 2. There are a few occurrences of the blue contours following the green contours, indicating that there might be some association of DR16Q quasars with low richness clusters.

3 DISCUSSIONS AND CONCLUSIONS

In this paper, we have presented the discovery of the Giant Arc (GA): a ∼ Gpc LSS at $z \sim 0.8$, mapped by Mg II absorption systems in the spectra of background quasars. The GA forms a large crescent shape on the sky that appears almost symmetrical. However, deeper analysis reveals some asymmetries in the GA, in the redshift and equivalent width (EW) distributions. The GA spans ∼1 Gpc on the sky and has a redshift depth of ∼340 Mpc (both proper sizes, present epoch). Visually, we determine the GA as a single unit, but using a Minimal Spanning Tree (MST) type algorithm (see Section 2.3.1) it splits into two portions: a large portion (GA-main) and a small portion (GA-sub). We proposed in Section 2.3.1 that the two portions of the GA could in fact be connected in reality, since potentially one more background probe could lead to one more Mg II absorber that would connect the two portions. On its own, GA-main is a statistically significant clustering of Mg II absorbers, with a membership of 44 Mg II absorbers, an MST-overdensity of 1.3 ± 0.3, and a mass excess of $1.8 \times 10^{18} \mathrm{M}_{\odot}$. In these respects, the GA is comparable to the Huge-LQG (Clowes et al. 2013).

Three different statistical tests were applied to the GA to assess the significance of connectivity and clustering. The results of each are summarized here. (i) The SLHC/CHMS method calculates the significance of clustering between points of close proximity by comparing the volumes of the CHMS for each structure to the CHMS of structures in randomly distributed points in a cube. GA-main, containing 44 Mg II absorbers, has a significance of $(4.5 \pm 0.6)\sigma$. GA-sub, containing 11 Mg II absorbers, has a significance of $(2.1 \pm 0.9)\sigma$. Both GA-main and GA-sub have the same MST-overdensity of $\delta\rho/\rho = 1.3 \pm 0.3$. This fact could indicate, as we suspect, that both agglomerations are connected in reality. (ii) The CE test is a case-control k nearest-neighbour algorithm that assesses the p-value of clustering in the field within an unevenly distributed population. A process of zooming into the GA field allows the GA to become increasingly dominant. In this way, we detect a p-value of 0.0027 from the field seen in Fig. 9(a), equivalent to a significance of 3.0σ. Applying this process of zooming to lower redshift fields at the same sky coordinates of the GA, we do not detect any significant clustering. We then use our polygon approach that randomizes points outside the GA, while keeping the visually selected absorbers contained within the GA the same. The CE test detects similar p-value of clustering with the polygon approach, indicating that the GA is the true, dominating feature causing significant clustering. (iii) The PSA is a Fourier method of detecting clustering in the field on a physical scale. We apply the 2D PSA to the 'zoomed' GA field, Fig. 9(a), and find significant clustering at $\lambda_c \sim 270$ Mpc with a significance of 4.8σ. As with the CE test, we use our polygon approach and detect similar significant clustering scales. However, a small contribution from other absorbers in the field is also detected. We do expect this given the power of the PSA test, and it is clear that the GA is still the dominant contributor to the PSA result.

Clearly, the analysis of the GA is after the event of its discovery, as is unavoidable with unexpected discoveries in astrophysics and cosmology. We have applied several different approaches to mitigating any post hoc aspects of analysing the statistical significance of the GA after discovery. We have performed techniques that aim to assess the GA unbiassedly, such as the polygon approach, varying redshift slices, zooming into the GA field, and randomized simulations. In the future, we can apply the same techniques used for the GA field to the whole of the Mg II data set. In addition, the Mg II data set is quite complex, with features that need careful attention: for example, the inhomogeneities of the Mg II absorbers are superimposed on the inhomogeneities of the quasars (background probes) and of the survey. Finally, there are different Mg II data bases available from different authors, each using different detection processes. We intend eventually to produce our own data bases of Mg II detections that can be used consistently with past and future quasar-survey data releases.

The GA is now amongst several other very large LSS discoveries with sizes that exceed the theoretical upper-limit scale of homogeneity of Yadav et al. (2010). Potentially, there are other such significant structures in the rest of the Mg II data base. We discuss that there are challenges in fairly characterizing the population of structures due to the inhomogeneities in the background probes (quasars). However, the challenges can be managed with suitable care, allowing for the Mg II method of studying LSS to be fully exploited.

In Table 1, we listed some of the very large LSSs, and also some of the reported CMB anomalies. In standard cosmology, we expect to find evidence for a homogeneous and isotropic Universe. However, the accumulated set of LSS and CMB anomalies now seems sufficient to constitute a prima facie challenge to the assumption of the Cosmological Principle (CP). A single anomaly, such as the GA on its own, could be expected in the standard cosmological model. For example, Marinello et al. (2016) find that the Huge-LQG (Clowes et al. 2013), a structure comparable in size to the GA, is, by itself (there are others), compatible with the standard cosmological model. However, Marinello et al. (2016) state that this is on the condition that only one structure as large as the Huge-LQG is found in a field ∼5 times the sample survey, in this case, the DR7QSO quasar database for $1.2 \leq z \leq 1.6$. Note that the GA is found in the combined footprint from DR7QSO and DR12Q (the combined footprint being almost the same area as the individual footprints), in a narrow redshift interval, so its challenge to the CP seems likely to be exacerbated. Of course, the GA is now the fourth largest LSS, so there are, at minimum, four LSSs comparable to the size of the Huge-LQG, plus several other LSSs exceeding the scale of homogeneity. We suggest that there is a need to explore other avenues within cosmology that could explain multiple, very large LSSs.

We bring attention to the Sloan Great Wall (SGW; Gott et al. 2005), which is a large, wall-like filament in the relatively local Universe. The SGW is ∼450 Mpc in its longest dimension, which is ∼0.5 times the length of the GA. One can note some of the similarities between the SGW and the GA, such as the general shape and comoving size – they are both long, filamentary and curved walls made up of galaxies and galaxy clusters –, and so perhaps also envision a LSS such as the GA as a precursor to the SGW. The GA is at a redshift of ∼0.8 which means we are seeing it when the Universe was only half its present age. Perhaps the SGW, at an earlier epoch, initially looked more like the GA. At this point, these ideas are speculative only, but experimenting with simulations (possibly even with alternative cosmological models) could conceivably elucidate such hypothetical connections between structures like the GA and the SGW.

ACKNOWLEDGEMENTS

We thank Srinivasan Raghunathan for many helpful discussions, and we thank Ilona Söchting for suggesting use of the Cuzick-Edwards test.

This paper has depended on SDSS data and on the R software.

We thank the replacement referee for careful reading and thoughtful comments.

DATA AVAILABILITY STATEMENT

The data sets were derived from sources in the public domain: https://www.guangtunbenzhu.com/jhu-sdss-metal-absorber-catalog.

REFERENCES

Bagchi J., Sankhyayan S., Sarkar P., Raychaudhury S., Jacob J., Dabhade P., 2017, ApJ, 844, 25

Balázs L. G., Bagoly Z., Hakkila J. E., Horváth I., Kóbori J., Rácz I. I., Tóth L. V., 2015, MNRAS, 452, 2236

Bordoloi R., Lilly S. J., Kacprzak G. G., Churchill C. W., 2011, ApJ, 784, 108

Burenin R. A. et al., 2018, Astron. Lett., 44, 297

Chen H.-W., Wild V., Tinker J. L., Gauthier J.-R., Helsby J. E., Shectman S. A., Thompson I. B., 2010, ApJ, 724, L176

Christian S., 2020, MNRAS, 495, 4291

Churchill C. W. et al., 2000, ApJ, 130, 91

Churchill C. W., Kacprzak G. G., Steidel C. C., 2005, in Williams P. R., Shu C.-G., Menard B., eds, Proc. IAU Colloq. 199, Probing Galaxies through Quasar Absorption Lines. Cambridge Univ. Press, Cambridge, p. 24

Clowes R. G., 1986, MNRAS, 218, 139

Clowes R. G., Campusano L. E., 1991, MNRAS, 249, 218

Clowes R. G., Campusano L. E., Graham M. J., Söchting I. K., 2012, MNRAS, 419, 556

Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Söchting I. K., Graham M. J., 2013, MNRAS, 429, 2910

Colin J., Mohayaee R., Rameez M., Sarkar S., 2019, A&A, 631, L13

Cuzick J., Edwards R., 1990, J. R. Statist. Soc. B, 52, 73

Dutta R., Srianand R., Gupta N., Joshi R., 2017, MNRAS, 468, 1029

Evans J. L., Churchill C. W., Murphy M. T., Nielsen N. M., Klimek E. S., 2013, ApJ, 768, 3

French J., 2020, available at: https://cran.r-project.org/web/packages/smacpod

Friday T., Clowes R. G., Williger G. M., 2022, MNRAS, 511, 4159

Geller M. J., Huchra J. P., 1989, Science, 246, 897

Gott J. R. et al., 2005, ApJ, 624, 463

Hinrichsen V. L., Klassen A. C., Song C., Kulldorff M., 2009, Int. J. Health Geogr., 8, 41

Horváth I., Hakkila J., Bagoly Z., 2014, A&A, 561, L12

Horvath I., Szécsi D., Hakkila J., Szabó Á., Racz I. I., Tóth L. V., Pinter S., Bagoly Z., 2020, MNRAS, 498, 2544

Hutsemékers D., 1998, A&A, 332, 410

Hutsemékers D., Lamy H., 2001, A&A, 367, 381

Hutsemékers D., Cabanac R., Lamy H., Sluse D., 2005, A&A, 441, 915

Hutsemékers D., Braibant L., Pelgrims V., Sluse D., 2014, A&A, 572, A18

Kacprzak G. G., Churchill C. W., Steidel C. C., Murphy M. T., 2008, ApJ, 135, 922

Keenan R. C., Barger A. J., Cowie L. L., 2013, ApJ, 775, 62

Lanzetta K. M., Bowen D., 1990, ApJ, 357, 321

Lee J. C., Hwang H. S., Song H., 2021, MNRAS, 503, 4309

Lietzen H. et al., 2016, A&A, 588, L4

Lopez A. M., 2019, MSc thesis, Univ. of Central Lancashire

Lyke B. W. et al., 2020, ApJS, 250, 8

Marchã M. J. M., Browne I. W. A., 2021, MNRAS, 507, 1361

Marinello G. E., Clowes R. G., Campusano L. E., Williger G. M., Söchting I. K., Graham M. J., 2016, MNRAS, 461, 2267

Meliker J. R., Jacquez G. M., Goovaerts P., Copeland G., Yassine M., 2009, Cancer Causes Control, 20, 1061

Migkas K., Schellenberger G., Reiprich T. H., Pacaud F., Ramos-Ceja M. E., Lovisari L., 2020, A&A, 636, A15

Nadathur S., 2013, MNRAS, 434, 398

Pâris I. et al., 2017, A&A, 597, A79

Pilipenko S. V., 2007, Astron. Rep., 51, 820

Pomarède D., Tully R. B., Graziani R., Courtois H. M., Hoffman Y., Lezmy J., 2020, ApJ, 897, 133

Schneider D. P. et al., 2010, AJ, 139, 2360

Schwarz D. J., Copi C. J., Huterer D., Starkman G. D., 2016, Class. Quantum Grav., 33, 184001

Secrest N. J., von Hausegger S., Rameez M., Mohayaee R., Sarkar S., Colin J., 2021, ApJL, 908, L51

Song C., Kulldorff M., 2003, Int. J. Health Geogr., 2, 9

Steidel C. C., 1995, in Meylan G., ed., ESO Workshop on Quasar Absorption Lines. Springer-Verlag, Berlin, Heidelberg, p. 139

Steidel C. C. et al., 2002, ApJ, 570, 526

Thomas S. A., Abdalla F. B., Lahav O., 2011, PRL, 106, 241301

Webster A. S., 1976a, MNRAS, 175, 61

Webster A. S., 1976b, MNRAS, 175, 71

Webster A. S., 1982, MNRAS, 199, 683

Whitbourn J. R., Shanks T., 2016, MNRAS, 459, 496

Williger G. M., Campusano L. E., Clowes R. G., Graham M. J., 2002, ApJ, 578, 708

Yadav J. K., Bagla J. S., Khandai N., 2010, MNRAS, 405, 2009

Zhu G., Ménard B., 2013, ApJ, 770, 130

Zou H. et al., 2021, ApJS, 253, 56

This paper has been typeset from a TEX/LATEX file prepared by the author.

Journal of Cosmology
and Astroparticle
Physics

PAPER • **OPEN ACCESS**

A Big Ring on the sky

To cite this article: A.M. Lopez *et al* JCAP07(2024)055

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *February 13, 2024*
Revised: *June 13, 2024*
Accepted: *July 8, 2024*
Published: *July 24, 2024*

A Big Ring on the sky

A.M. Lopez,[a] **R.G. Clowes**[a] and **G.M. Williger**[b]

[a]*Jeremiah Horrocks Institute, University of Central Lancashire,*
Preston, PR1 2HE, United Kingdom
[b]*Department of Physics and Astronomy, University of Louisville,*
Louisville, KY 40292, U.S.A.

E-mail: amlopez@uclan.ac.uk, rgclowes@uclan.ac.uk,
gmwill06@louisville.edu

Abstract: We present the discovery of 'A Big Ring on the Sky' (BR), the second ultra-large large-scale structure (uLSS) found in Mg II-absorber catalogues, following the previously reported Giant Arc (GA). In cosmological terms the BR is close to the GA — at the same redshift $z \sim 0.8$ and with a separation on the sky of only $\sim 12^\circ$. Two extraordinary uLSSs in such close configuration raises the possibility that together they form an even more extraordinary cosmological system. The BR is a striking circular, annulus-like, structure of diameter ~ 400 Mpc (proper size, present epoch). The method of discovery is as described in the GA paper, but here using the new Mg II-absorber catalogues restricted to DR16Q quasars. Using the Convex Hull of Member Spheres (CHMS) algorithm, we estimate that the annulus and inner absorbers of the BR have departures from random expectations, at the density of the control field, of up to 5.2σ. We present the discovery of the BR, assess its significance using the CHMS, Minimal Spanning Tree (MST), FilFinder and Cuzick & Edwards (CE) methods, discuss it in the context of the GA+BR system, and suggest some implications for the origins of uLSS and for our understanding of cosmology. For example, it may be that unusual geometric patterns, such as these uLSSs, have an origin in cosmic strings.

Keywords: cosmic web, superclusters, baryon acoustic oscillations, Cosmic strings, domain walls, monopoles

ArXiv ePrint: 2402.07591

https://doi.org/10.1088/1475-7516/2024/07/055

Contents

1 Introduction **1**
1.1 Cosmological model 3

2 The Big Ring **3**
2.1 Data sources 4
2.2 Initial checks of the data 5

3 Statistical analysis **8**
3.1 Single-Linkage Hierarchical Clustering algorithm 9
3.2 Significance: CHMS and MST 17
3.3 Filament identification algorithm 20
3.4 Cuzick and Edwards test 24

4 Observational properties **26**
4.1 Corroboration with independent data 26
4.2 Viewing the BR from other angles 27
4.3 Equivalent widths 33

5 Discussion and conclusions **34**

1 Introduction

We continue to make use of the method of intervening Mg II absorbers in the spectra of quasars to trace faint matter at intermediate redshifts [1, 2]. The Mg II method relies on both the spectroscopic measurement of luminous, high redshift quasars from the Sloan Digital Sky Survey (SDSS), and the highly accurate, spectroscopic redshifts of the intervening Mg II absorption doublets present in the quasar spectra, documented by independent authors [3, 4]. The intervening Mg II absorption doublet indicates the presence of galaxies and galaxy clusters [5–7]. Together, with the quasars and Mg II absorbers, we have the information on the on-sky position of intervening matter and the redshift of the intervening matter, so in mapping the 3D distribution of the Mg II absorption features in the spectra of quasars, we can infer the LSS of intermediate-to-high redshift, faint matter.

The multiple discoveries of LSSs made throughout the past few decades are well known to challenge our understanding of the Standard Cosmological Model (ΛCDM) [2, 8–12], in particular due to a possible violation of a fundamental assumption, the Cosmological Principle (CP), which states that our Universe is both homogeneous and isotropic on large scales. In addition, there are numerous results in cosmology posing similar challenges and tensions for our current standard model (see [13] for a recent review). Large structures in the Universe are interesting for several reasons, such as: how did the structures form so early on in the evolution of the Universe, given the current understanding of CDM structure formation; how might the structures evolve to the present day; do the seeds of such large LSSs lie in the density perturbations that are amplified by inflation [14, 15]. Answers to such questions may lie outside concordance cosmology, with either the inclusion of extensions to the standard model, e.g., cosmic strings (CS) [9, 16–18], or alternative theories to the ΛCDM model, e.g., Conformal Cyclic Cosmology (CCC) [19] and modified gravity theories [20–22].

During the discovery of the Giant Arc (GA) [2] (hereafter Lopez22), the SDSS Data Release 16 quasar database (DR16Q) became available [23]. Then, independent authors created the most up-to-date Mg II database from the DR16Q quasars [3] (hereafter Anand21). Following the discovery of the GA, made using the older Mg II database from [4] (hereafter Z&M, also the corresponding DR7QSO and DR12Q quasar catalogues from [24, 25]), we are now in a position to continue LSS investigations with the new DR16Q database and corresponding DR16Q Mg II database. We have found an interesting ring shape in the Mg II absorbers, indicating a LSS of galaxies and galaxy clusters, that spans a diameter of $\sim 400\,\mathrm{Mpc}$ scaled to the present epoch. Incidentally, the estimated size of the BR is close to that which could be expected in a detection of an individual Baryon Acoustic Oscillation (BAO), $r \sim 150\,\mathrm{Mpc}$ [26–30], but we later suggest that the BR is unlikely to have its origins in BAOs. The BR shape and size are both hard to understand in our current theoretical framework. Additionally, the BR is in the same redshift slice as the GA and to the north of the GA by $\sim 12°$ which raises further questions about their origin both together and independently.

For the work that led to the original discovery of the GA, using the Z&M database, we had looked at only a few small areas of sky and redshift slices [1], essentially to test the viability of the Mg II approach itself. Following the discovery of the GA, and now using the Anand21 database, we have so far concentrated on the GA field and redshift slice because we immediately made the further discovery of the BR there; the only exceptions to this statement are (i) the use of adjacent redshift slices to test that the BR was not arising from artefacts and to test for extensions of the BR into adjacent redshift slices, and (ii) the use of neighbouring fields at the same redshift as the GA/BR field for comparing the spatial clustering results of the Cuzick and Edwards test (see section 3.4). Consequently, a 'look-elsewhere' effect on the statistical assessments should not be a factor. In future, of course, we intend to explore both databases in their entireties.

The Mg II data we use here are complicated and quite difficult to manage. The advantage, of course, is the precise redshifts, and the concomitant possibility of discovering intriguing structures such as the Big Ring and the Giant Arc. In future, the Mg II approach to LSS should be enhanced by DESI spectra [31], taken with the KPNO 4m telescope, allowing detection of Mg II to lower equivalent widths, and hence allowing the exploration of finer detail in the morphology of structures.

In this paper we assess the reality of the BR and its statistical significance with respect to the assumed, homogeneous large-scale distribution of matter.

Note that we have introduced the term 'ultra-large LSS' (uLSS) to distinguish those structures that exceed the estimated $\sim 370\,\mathrm{Mpc}$ upper limit to the scale of homogeneity [32]. This limit is often adopted in discussions of homogeneity and the CP.

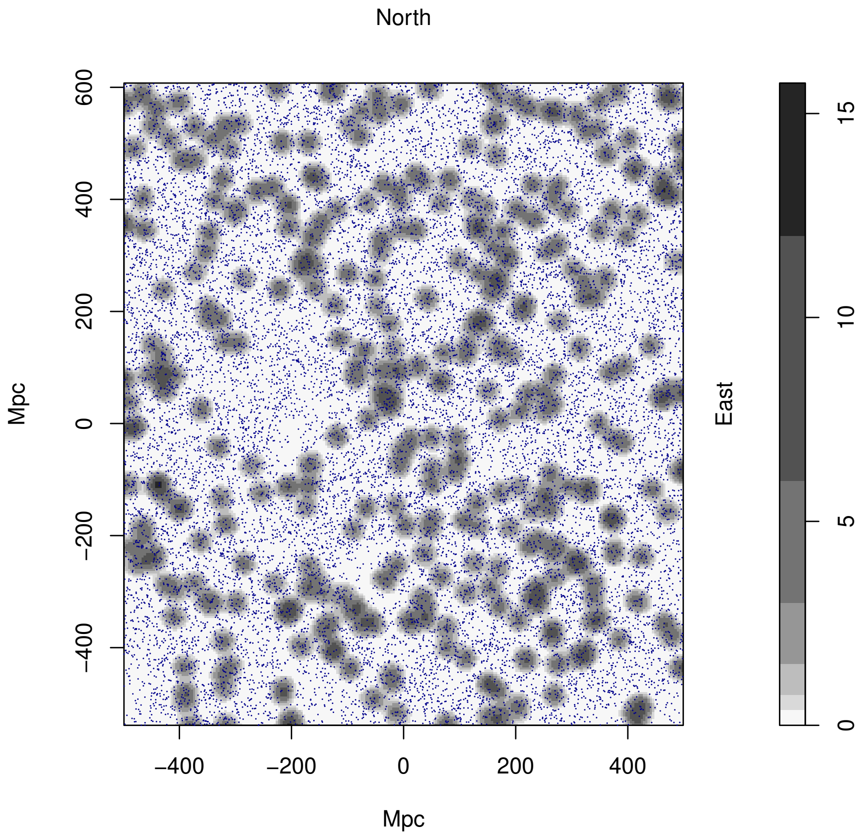


Figure 1. The tangent-plane distribution of Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$. The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$, and flat-fielded with respect to the distribution of background probes (quasars). The dark blue dots represent the background probes. S/N limits of: $4, 2$ and 4 were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively (details of S/N cuts are discussed in section 3.1). The BR can be seen to the north of the centre point spanning $\sim 400\,\mathrm{Mpc}$ in diameter. The field-of-view corresponds to the small, pink area seen in figure 5.

1.1 Cosmological model

The concordance model is adopted for cosmological calculations, with $\Omega_{T0} = 1$, $\Omega_{M0} = 0.27$, $\Omega_{\Lambda 0} = 0.73$, and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All sizes given are proper sizes at the present epoch. (For consistency, we are using the same values for the cosmological parameters that were used in Lopez22.)

2 The Big Ring

The existence of a large, circular structure of Mg II absorbers, the Big Ring (BR), became apparent when investigating the new SDSS DR16Q catalogues, and corresponding Mg II databases, at the same redshift and position as the previously documented Giant Arc (GA). In figure 1, we are seeing the BR, which is the visually overdense ring shape of Mg II absorbers centred at approximately $x = 0\,\mathrm{Mpc}$ and $y = 240\,\mathrm{Mpc}$. (The large 'void' to the south-west of the BR is also particularly striking.) In this figure, and others, the grey contours,

JCAP07(2024)055

increasing by a factor of two, represent the density distribution of Mg II absorbers in the specified redshift slice and field-of-view (FOV). The Mg II contours have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$, and flat-fielded with respect to the distribution of background probes (quasars). The smoothing gives a useful impression of the connectivity. The background probes (quasars) are represented by the small, blue points. The axes are labelled in Mpc, scaled to the present epoch; for details on obtaining the Mpc scaling, refer to the GA paper (Lopez22). East is towards the right and north is towards the top. From figure 1 we can estimate the BR diameter is $\sim$ 300–400 Mpc, which would make its circumference comparable to the extent of the GA.

2.1 Data sources

The use of Mg II absorbers for analysing LSS has the particular advantage of providing very precise redshifts. A disadvantage of course, which can require very careful handling, is that one must take the background probes — the quasars — where they are given by the catalogues. The catalogues now have suitably dense coverage on the sky, but generally they are affected by variations in selection criteria.

At $z_{\mathrm{abs}} \sim 0.8$, the parameter in the Anand21 catalogues for the redshift error (Z_ABS_ERR) indicates a median error of $\sigma_{z_{\mathrm{abs}}} \approx 4.2 \times 10^{-5}$. (The emission-redshift error for a quasar catalogue might be about two orders of magnitude larger at ~ 0.004.) This $\sigma_{z_{\mathrm{abs}}}$ corresponds to a velocity difference of $\sim 7\,\mathrm{km\,s^{-1}}$. A comparison of repeated observations in the basic Anand21 database suggests that a practical estimate of the redshift error at $z_{\mathrm{abs}} \sim 0.8$ is a little larger at $\sigma_{z_{\mathrm{abs}}} \approx 1.7 \times 10^{-4}$, corresponding to a velocity difference of $\sim 28\,\mathrm{km\,s^{-1}}$.

When considering the finer details of the morphology of individual LSSs, any blurring will then be due to peculiar velocities, for which plausible values might be $\sim 400\,\mathrm{km\,s^{-1}}$, corresponding to $\sigma_{z_{\mathrm{pec}}} \sim 0.0024$ at $z \sim 0.8$, or $\sim 7\,\mathrm{Mpc}$ in proper distance for the present epoch. We therefore expect that any blurring effects should be minor.

The BR is detected in the new Anand21 Mg II database, whereas the previously documented GA was detected in the older Z&M Mg II database, so we investigate the differences in the databases (e.g. the Mg II absorbers detected from probes that were in common to both databases) in the following manner. First, we choose the standard BR FOV at $z = 0.802 \pm 0.060$ as in figure 1. Then, we select the absorbers and probes arising in the chosen field from both Anand21 and Z&M (making no additional cuts to the S/N or i-magnitude). Using TOPCAT[1] [33], the probe (quasar) and Mg II files are paired in various ways for comparison.

A summary of the results is as follows. (1) There are over three times as many background probes in the field for Anand21 than Z&M. (These are the probes that were initially searched by the authors for Mg II absorbers.) Almost all of the additional probes searched by Anand21 were new observations between SDSS DR12 to DR16. However, a small fraction (just 37 out of 7257 probes) were not included in the Mg II search by Anand21. (2) There are just over twice as many Mg II absorbers detected by Anand21 than by Z&M: 852 Mg II absorbers in Anand21 and 359 Mg II absorbers in Z&M. Of those absorbers, 597 were unique to Anand21 and 104 were unique to Z&M. (3) Of the probes that both authors had in common and

---

[1] https://www.star.bris.ac.uk/mbt/topcat/.

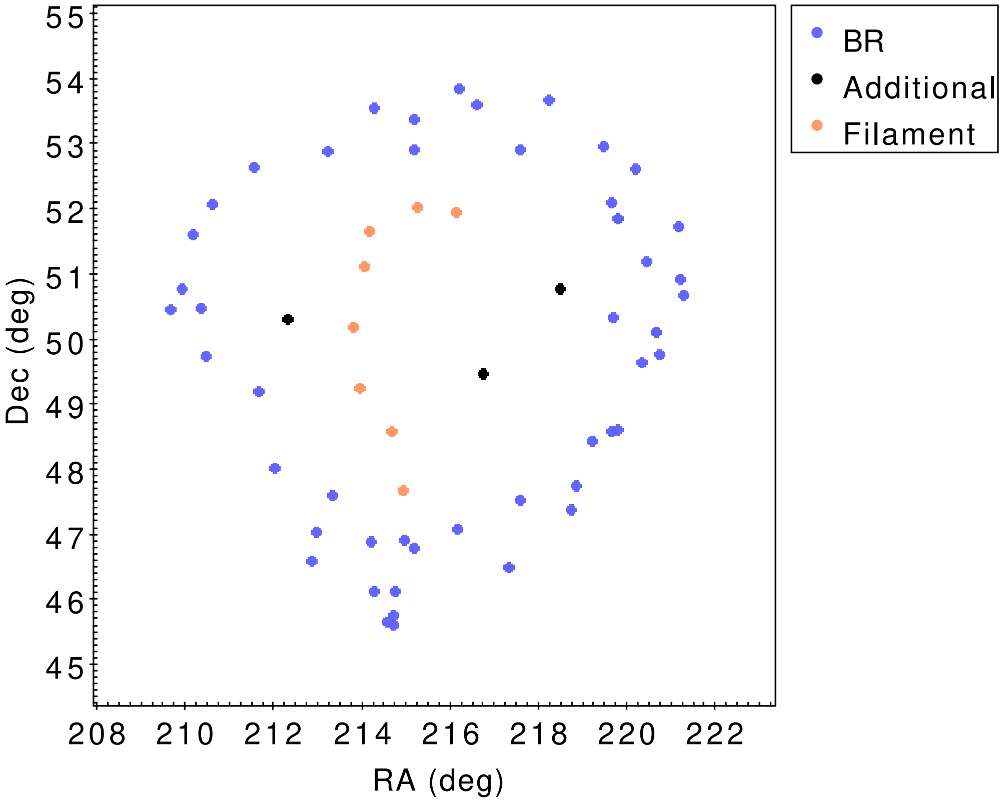


Figure 2. The Mg II absorbers in the visually-identified BR (blue), the inner filament (orange), and the additional absorbers inside the BR (black) from the on-sky perspective. Three of the blue points (BR) correspond to two absorbers occurring in one spectrum (see figure 8).

searched, Anand21 found 67 additional absorbers that Z&M missed, and Z&M found 103 additional absorbers that Anand21 missed. In total, there were 255 Mg II absorbers in common to both authors ($\sim 60\%$ total agreement of the probes in common to both authors). We found that, despite the greater number of absorbers in Anand21, they missed around 1.5 times as many absorbers found by Z&M, compared with the number of absorbers missed by Z&M that were found by Anand21, for shared probes. Without access to their software it will probably remain unclear why Anand21 and Z&M have this disagreement. One possibility might be that Anand21 have more refined search criteria, given that their percentage Mg II detection rate per quasar is much lower than for Z&M. Or, perhaps Z&M are detecting to lower thresholds, without increasing spurious detections.

2.2 Initial checks of the data

First, we check that the Mg II absorbers belonging to the visually-identified BR are real (not false positive detections). We visually inspected spectra of 56 DR16Q quasars that are the probes that correspond to the visually-identified BR and inner filament absorbers (see figure 2). Since 6 of the Mg II absorbers are multiples per probe occurring in three quasar spectra, we were checking for a total of 59 Mg II doublet systems. Each Mg II absorption doublet that we searched for was visually confirmed and in agreement with the documented redshifts from

Anand21, so we can confirm that 100% of the Mg II absorber members of the BR and inner filament are real, physical Mg II absorbers indicating the presence of intervening matter.

There are two unusual Mg II systems, occurring in the same spectrum, for which the Anand21-documented redshifts suggest that the λ_{2796} of the lower-z absorption doublet appears at the same wavelength as the λ_{2803} of the higher-z absorption doublet. This is indeed the case: the two Mg II absorption doublets appear as 3 absorption lines in the spectrum, which is a rare oddity. Although Anand21 recognise the 3 absorption lines as two systems, they appear not to have disentangled the EWs of the centre absorption line of the triplet (the higher-z λ_{2796} and the lower-z λ_{2803} EWs being the same).

Secondly, we investigate if the visually obvious BR is an artefact of the probes. This can be done in two ways: simply checking the density distribution of background probes and checking for obvious artefacts; and looking at the next redshift slice down from the BR field (on the near side) and checking for repeating Mg II features that correspond to any obvious artefacts in the probes. For the former, see figure 3.

Many overdensities and underdensities can be seen clearly in figure 3. In particular, there are a few overdense regions (small, dark clumps) centred at 0 Mpc on the x-axis and between roughly 100 Mpc to 400 Mpc on the y-axis, a few of which appear to coincide with the inner filament of BR. This could imply that the inner filament of the BR is suspect, so we will need to be sure that the filament is not an artefact of the probes. We can visually check the rate of occurrence of Mg II absorbers at the position of the overdense artefacts (dense blobs) by blinking the image of the probes (figure 3) with the image of the absorbers (figure 1). Doing this shows that 8 out of 20 randomly selected overdense artefacts had Mg II absorbers present, so less than half of the artefacts. Incidentally, we could also confirm that most of the artefacts that appeared to coincide with parts of the BR (including the inner filament) were in fact offset, so not responsible for the Mg II absorbers arising there. Checking the artefacts shows that there is no particular association of the artefacts with the Mg II absorbers present in the BR field. There is also a much larger region of underdense probes spanning -100 Mpc to 400 Mpc in the y-axis and centred at -200 Mpc in the x-axis which partially coincides with the l.h.s. of the BR. The fact that part of the BR is located in an underdense region of probes is noteworthy.

For the latter way to test if the MgII absorbers are artefacts of the probes, we can check the absorbers arising from the same set of probes corresponding to the BR field in the next, non-overlapping redshift slice on the near side of the BR. To do this we keep the probes of the BR field the same (having $z > 0.862$ — i.e., a redshift greater than the far edge of the BR Mg II redshift slice) and map the Mg II absorption in the nearest, non-overlapping redshift slice (i.e., $z = 0.682 \pm 0.060$). In this way we are able to search for any obvious artefacts of the probes that could be responsible for the specific distribution of Mg II absorbers in the BR field by comparing the Mg II image in the usual BR field with the neighbouring redshift slice. We apply the SLHC/CHMS and FilFinder algorithms to the field centred at $z = 0.682 \pm 0.060$ and corresponding to the usual BR FOV (see sections 3.1 and 3.3 for details on the SLHC/CHMS and FilFinder methods). In the redshift slice $z = 0.682 \pm 0.060$ we find with the SLHC/CHMS method that there are no structures detected by the SLHC algorithm corresponding to the BR, and in addition, no structures at all that are statistically significant.

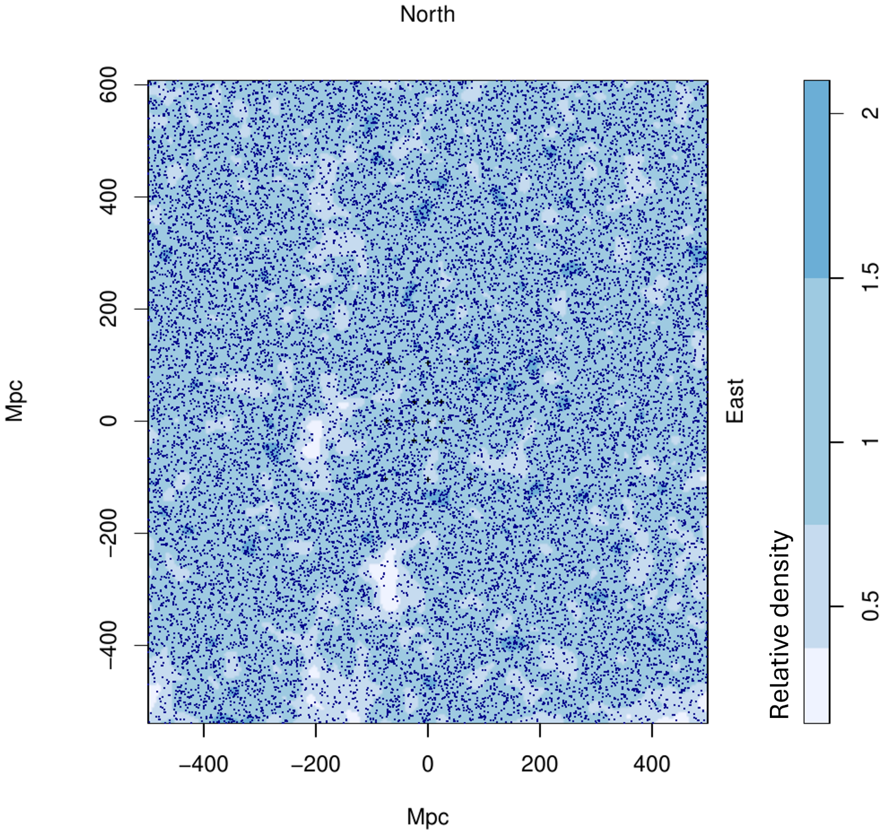


Figure 3. The tangent-plane distribution of probes (background quasars) in the BR FOV with the redshift condition $z > 0.862$ — i.e., the probes that are responsible for the Mg II absorbers arising in the BR field. The blue contours, increasing by a factor of two, represent the density distribution of the probes which have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$. S/N limits were applied to the quasar continuum such that $\mathrm{S/N_{con}} > 4$ (details of S/N are discussed in section 3.1). The field-of-view corresponds to the small, pink area seen in figure 5. The figure shows many areas of overdensities and underdensities. In particular, there are a few overdense regions (small, dark clumps), centred at 0 Mpc on the x-axis and between roughly 100 Mpc to 400 Mpc on the y-axis, a few of which appear to coincide with the inner filament of BR. There is also a much larger region of underdense probes spanning −100 Mpc to 400 Mpc in the y-axis and centred at −200 Mpc in the x-axis which coincides with the l.h.s. of the BR.

We also find with the FilFinder method, no filaments detected in the field that correspond to the BR (figure 4). Therefore, we conclude that the BR is not an artefact of the probes.

It is worth highlighting that the BR appears in the same redshift slice and FOV as the previously documented GA, but we are now using the new Anand21 databases and not the previously used Z&M databases. As mentioned earlier, the overall agreement between the two datasets is $\sim 60\%$ for the probes in common in the BR field, so the GA appears somewhat different in this new dataset. The GA is still the most significant, most numerous and overdense structure detected in the field despite the slight change in appearance, so quantitatively, there is very little difference in the GA in the new dataset. However, qualitatively, there are two main reasons the GA appears different: (1) the field overall is much more dense with the new dataset due to many more quasar observations; (2) Anand21 miss several GA absorbers. We

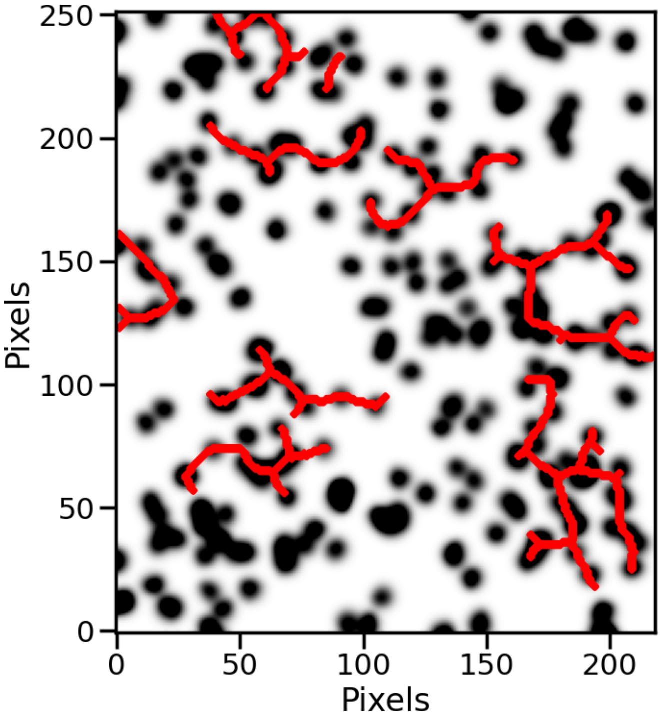


Figure 4. The FilFinder algorithm applied to the tangent-plane distribution of Mg II absorbers in the closest, non-overlapping redshift slice to the BR field — i.e., the field centred at $z = 0.682 \pm 0.060$ and corresponding to the usual BR field-of-view. Axes are labelled in pixels, where 1 pixel = $4^2\,\mathrm{Mpc}^2$. S/N limits of: 4, 2 and 4 were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively (details of S/N cuts are discussed in section 3.1). The probes (quasars) responsible for the Mg II arising here have redshifts $z > 0.862$ so that they are the same probes responsible for the Mg II absorbers in the BR field centred on the usual redshift slice ($z = 0.802 \pm 0.060$). The figure shows that there are no filaments correlating to the BR indicating that the BR is not a result of artefacts in the probes.

investigate the second point by manually checking each of the quasar spectra that are probes to the Mg II absorbers in the GA that Anand21 missed. There were 16 from 51 absorbers that Anand21 missed, and none was due to the small fraction of removed quasars in DR16Q that was mentioned earlier. In each of the quasar spectra (corresponding to the GA) the Mg II doublets were visually confirmed, but the 16 absorbers missed by Anand21 had profiles that were generally complex, broad or weak. This again suggests that the Anand21 Mg II detection algorithm has a narrower detection window at the cost of losing some real absorbers. Conversely, the Z&M Mg II detection algorithm could conceivably contain more spurious absorbers while likely managing to detect a higher percentage of the real absorbers.

3 Statistical analysis

The discovery of the BR was made serendipitously when looking at the previously documented GA field with the Anand21 data. We now present the statistical analysis of the BR, taking a similar approach to the previous work in Lopez22. Given the nature of the discovery, the analysis is necessarily post-hoc. However, given the previous work presented in Lopez22 we

JCAP07(2024)055

are able to follow the guidelines set there, consequently alleviating many of the problems associated with post-hoc analysis.

Simulations are often advocated in contemporary astrophysics and cosmology, but we do not consider them likely to be effective or efficient here. Their complexity would be too great and would have too many unknowns and uncertainties. Consider, for example, that the simulations would have to incorporate: simulating the universe in general; the occurrence of quasars in that simulated universe; the observational parameters of the imaging and spectroscopic surveys and their on-sky variations; and the detection of the Mg II by software. Instead, we have taken the more practical approach of (i) using the data to correct the data, and (ii) seeking independent corroboration of features using independent tracers.

This section is divided in the following manner. (3.1) We assess the BR from a 'first look' perspective by using a heuristic process of stepping through redshift slices and determining the optimum redshift for the BR. (3.2) We use the Convex Hull of Member Spheres (CHMS) and the Minimal Spanning Tree (MST) significance calculations for assessing the significance of the BR. We apply these two methods of significance calculations to four sets of BR absorber-member estimates: the SLHC groups; the visually-identified BR absorbers (both including and excluding the inner absorbers); and the FilFinder-identified absorbers. (3.3) The 2D FilFinder algorithm is applied to the pixel image containing the BR to objectively identify filaments in the field. (3.4) Finally, we apply the 2D Cuzick and Edwards test to the BR field to determine the significance of clustering in the field (not the candidate structure itself).

3.1 Single-Linkage Hierarchical Clustering algorithm

The Single-Linkage Hierarchical Clustering (SLHC) algorithm is equivalent to a Minimal Spanning Tree (MST in a generic sense, not to be confused with the MST *significance* calculation in section 3.2) when separated at a specified linkage scale. Our particular application of the algorithm was first described in [12], in combination with the Convex Hull of Member Spheres (CHMS) algorithm which assesses the significance of a specified structure. The SLHC/CHMS method has been used to locate and assess LSS in both quasars and Mg II absorbers [2, 11, 12].

Previously, the inherent difficulties of the Mg II method have been discussed, especially in relation to applying statistics on an intrinsically inhomogeneous dataset. The quasars act as probes of the intervening matter, so the varying availability of background quasars leads to an incomplete image or map of the intervening matter, in this case, the Mg II absorbers. The inhomogeneity of the quasars, both intrinsic and of the survey, contribute in a complicated way to the inhomogeneity of the Mg II absorbers. Of course, with large data and large survey areas, approximations can be made. Before applying the SLHC algorithm we define the field-of-view (FOV) for the analysis presented here.

First, the BR appears in the north of the usual GA field, so we re-centre the Mg II images accordingly. Second, the field containing the BR is close to the northern SDSS footprint border as well as a southern border arising from much lower quasar coverage. We choose to shrink slightly the typical size of a Mg II image to avoid these areas. Third, given the generally patchy DR16Q quasars, and thus the corresponding Mg II absorbers in the Anand21 databases, we apply signal-to-noise (S/N) limits to the Mg II absorption lines and the continuum. We calculate the S/N of the Mg II lines by:

$$S/N = W_r/W_{\text{err}}$$

where Wr is the equivalent width of the line, and W_{err} is the corresponding error in the Mg II line. For the continuum, Anand21 provide the median quasar S/N. Applying S/N limits has the effect of removing spurious and potentially false positive absorbers as well as generally reducing the patchiness in the data. Setting additional[2] limits to the magnitude of the quasars would also have a similar, desirable effect. However, not all quasars will have the same integration time, so faint quasars could have good S/N due to long exposures. Following the example by Z&M, we apply a S/N limit of 4 and 2 for the λ_{2796} and λ_{2803} lines respectively. Since the S/N of the quasar continuum necessarily has equal or higher S/N than the λ_{2796} line we apply a S/N limit to the quasar continuum of 6. Applying the condition of S/N $\geq$ 6 to the quasar continuum could be too restrictive, and later we will apply a less conservative condition to the continuum for comparison.

The CHMS significance is calculated by the rate of occurrence of volumes smaller than the CHMS volume of the structure by randomly distributing the absorbers belonging to the structure at a density equal to the control field density — the simulations are repeated 1000 times. Previously, the control field was chosen to be that of the field being assessed. However, this would mean that a percentage of the field absorbers are those belonging to the structures of interest, e.g., the GA and BR in the GA/BR FOV. In addition, the FOV containing the BR is small, so small-scale inhomogeneities have a much larger effect on the average density. Conversely, choosing a control field that is too large will lead to problems involving the large-scale inhomogeneities of the SDSS survey, explained above. Subsequently, we have designed two versions of control fields accounting for the northern portion of the SDSS footprint (overdense region) and the southern portion of the SDSS footprint (underdense). Each version can be chosen depending on the location of the field of interest. So here, we will be using version-1 of the control field — figure 5.

The SLHC is equivalent to an MST when separated at a specified linkage scale: thus the choice of linkage scale will determine the maximum distance between points that would be considered 'joined', or a candidate structure. The term 'structure'[3] here is not be confused with a gravitationally-bound system, but is instead referring to a grouping of more than 10 (a specified minimum) members (in this case, Mg II absorbers) that have an MST with distances smaller than the linkage scale. LSSs are not expected to be gravitationally bound, as, indeed, superclusters and great walls are not expected to be gravitationally bound. Similar usages and definitions of 'structure' are common in LSS studies [9, 10, 34, 35]. The candidate structures are then assessed by the CHMS to determine whether their volumes are statistically

---

[2] A base-line magnitude limit of $i <= 20.5$ is applied to the DR16Q quasars as part of the read-in process since the completeness declines steeply for fainter quasars.

[3] For a working definition, in investigating LSS, we often consider a candidate structure to be a set of N connected tracers, the containing volume of which is a $n\sigma$ departure from the containing volume expected for a uniform, random distribution. We might choose to consider further only those candidates for which the amplitude $n\sigma$ exceeds some threshold. 'Connected' and 'containing' volume will often be determined *algorithmically*; for both, there is an implicit assumption of a uniform host survey, which might be approximately true only in restricted areas.

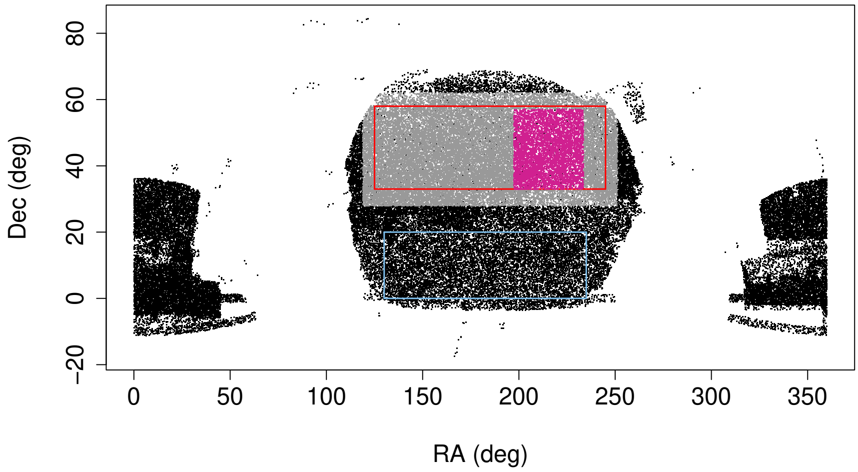


Figure 5. The SDSS DR16Q footprint. The grey points are the input quasars for the control field and the pink points correspond to the Mg II absorbers in the field of interest. The control fields are outlined by lines in red (overdense region, version-1) and blue (underdense region, version-2).

significant. In Lopez22 we discuss the effects of varying the linkage scale, so here we choose to scale the linkage scale according to the Mg II number density of the control field and the number density of the GA field — the GA field being the candidate field from which to scale all other fields, i.e.,

$$s = (\rho_0/\rho)^{1/3} s_0$$

where s and s_0 are the linkage scale for the control field and the GA field respectively, and similarly ρ and ρ_0 are the densities of the corresponding fields. The linkage scale is to be taken as a guideline; if the linkage scale is too small then potentially interesting candidate structures will be missed, and if the linkage scale is too high then too many points will be grouped as one seemingly coherent structure, but that would of course reflect in the CHMS significance. We find that the linkage scale set for the GA field (the field which we now use as the base-line) was an appropriate choice for the specific field density. However, we saw that even with the chosen linkage scale the SLHC algorithm identified the GA as two, individual, overlapping candidate structures. We reason that, when using a smaller linkage scale, candidate structures that are overlapping or adjacent could reasonably belong to the same structure. It is important to recognise that multiple candidate structures overlapping and adjacent to each other will still need to be assessed with the CHMS or MST significance, which would then objectively determine whether their agglomeration is statistically significant (remember, if the whole field was joined as one structure then this would of course not be statistically significant).

As with the GA analysis, we step through overlapping redshift slices and use the SLHC/CHMS algorithms to determine the redshift of the peak signal of the BR. Given the much larger Mg II database from Anand21, even after S/N limits are applied, it is expected that the field density containing the BR is much higher than the GA field in the Z&M databases, and therefore the chosen linkage scale will be correspondingly lower. The five redshift slices assessed are centred at: 0.682, 0.742, 0.802, 0.862 and 0.922, each with a redshift thickness of $\Delta z = 0.060$; the results are shown in table 1.

Central redshift	**Linkage scale (Mpc)**	**No. candidate structures**	**No. candidate structures with $\sigma_{\rm CHMS} \geq 3.5\sigma$**	**Maximum $\sigma_{\rm CHMS}$ (σ)**	**Maximum Mg II absorber membership**
0.682	79.3	4	0	3.3	17
0.742	79.3	5	1	3.5	16
0.802	79.5	8	1	4.5	28
0.862	79.6	6	2	4.7	42
0.922	81.1	4	1	4.1	26

Table 1. Results from the SLHC/CHMS on five, overlapping redshift slices to determine the optimum redshift slice for the BR signal. The FOV of each redshift slice corresponds to the small, pink area seen in figure 5. S/N limits of 4, 2 and 6 were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively. All redshift slices have a thickness of $\Delta z = \pm 0.060$. The columns from left to right are: the central redshift of the field being assessed; the linkage scale used for the field being assessed, calculated as $s = (\rho_0/\rho)^{1/3} s_0$ (see the main text); the number of candidate structures identified in the field; the number of candidate structures identified in the field with a CHMS significance equal to or exceeding 3.5σ; the maximum CHMS significance calculated from the candidate structures; the maximum Mg II absorber membership identified from the candidate structures.

The BR that was originally identified visually appears almost fully (in a partly-open ring) in only the central redshift slice $z = 0.802$ indicating that this is the optimum redshift slice for the BR, as it was for the GA. (Note, the GA is also identified, and is statistically significant in the central redshift slice). The four structures contributing to the visually-identified BR are adjacent or overlapping on the sky, indicating that the separate structures plausibly belong to the same structure. The apparent splitting of a seemingly coherent structure was also seen with the GA, which was made up of two overlapping SLHC groups; the splitting of structures is an example of the limitations of applying the SLHC algorithm to an essentially incomplete dataset. Interestingly, the SLHC group corresponding to the bottom portion of the BR appears also to extend into the two higher redshift slices, $z = 0.862$ and $z = 0.922$, since there are similarly-shaped arcs (corresponding to the bottom portion of the BR) appearing at the same on-sky position in all three redshift slices ($z = 0.802, 0.862, 0.922$). (Note, the probes are not here restricted to be identical.)

As mentioned earlier, the condition of S/N ≥ 6 for the quasar continuum could be too restrictive. Anand21 calculate the S/N over the whole quasar continuum, rather than the local continuum at the point of an absorber. Therefore, the whole quasar continuum could have lower S/N overall compared with local S/N at the position of an absorber. Accordingly, we slightly relax the S/N conditions to 4, 2, 4 for the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively and repeat the above described analysis (see table 2).

We see again that the full BR is detected — this time a full, closed ring, and with the inclusion of the inner filament — in the central redshift slice $z = 0.802 \pm 0.060$, figure 6. The BR is located north of the centre point, spanning ~ 10 degrees in the RA and Dec axes (x and y axes respectively), and is made up of a collection of five structures identified by the SLHC algorithm.

Central redshift	**Linkage scale (Mpc)**	**No. candidate structures**	**No. candidate structures with $\sigma_{\text{CHMS}} \geq 3.5\sigma$**	**Maximum σ_{CHMS} (σ)**	**Maximum Mg II absorber membership**
0.682	76.7	6	0	3.2	16
0.742	76.2	7	0	3.3	19
0.802	75.8	10	1	4.1	30
0.862	75.7	7	1	3.7	27
0.922	77.1	6	0	3.4	15

Table 2. As with table 1. Slightly relaxed S/N limits of $4, 2, 4$ were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively. All redshift slices have a thickness of $\Delta z = \pm 0.060$. The columns from left to right are: the central redshift of the field being assessed; the linkage scale used for the field being assessed, calculated as $s = (\rho_0/\rho)^{1/3} s_0$ (see the main text); the number of candidate structures identified in the field; the number of candidate structures identified in the field with a CHMS significance equal to or exceeding 3.5σ; the maximum CHMS significance calculated from the candidate structures; the maximum Mg II absorber membership identified from the candidate structures.

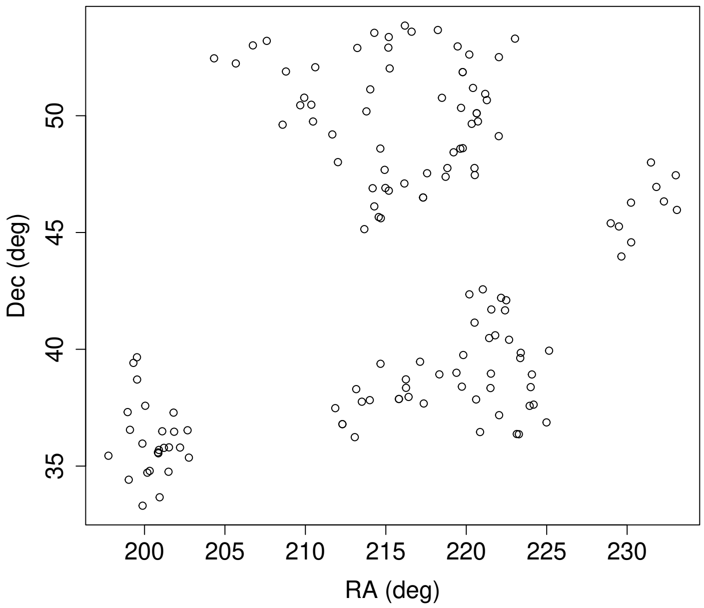


Figure 6. All the Mg II absorbers belonging to a candidate structure identified by the SLHC algorithm from results-2 in the redshift slice centred at $z = 0.802 \pm 0.060$. The field-of-view seen here corresponds to the pink points in figure 5. S/N limits of: $4, 2, 4$ were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively. The BR can been seen to the north of the centre point. The visually-identified BR is seen here with an additional extension heading towards to the north-west direction, and the visually-identified inner filament is the central line cutting through the BR here. The BR spans ~ 10 degrees in RA and Dec coordinates. The large structure south of the BR belongs to the previously identified GA.

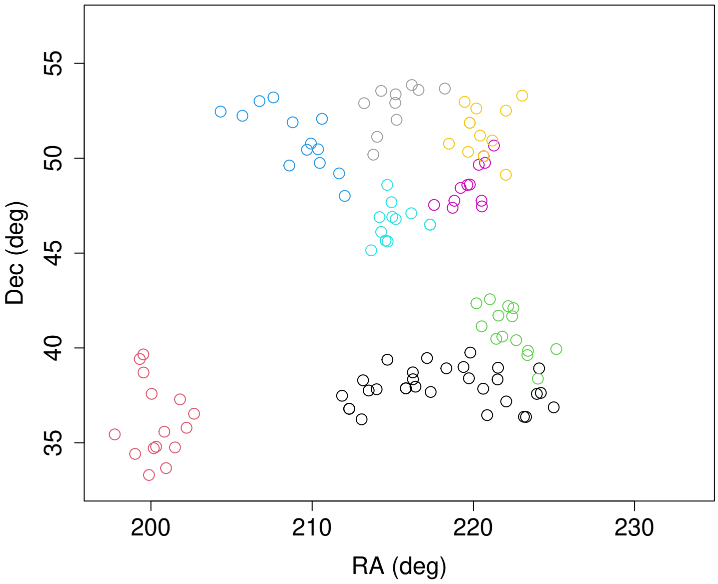


Figure 7. Eight of the 10 highest membership candidate structures identified by the SLHC/CHMS algorithms from results-2 in the redshift slice centred at $z = 0.802 \pm 0.060$. The colours represent the memberships which are ordered from high to low in the following way: black, red, green, blue, turquoise, pink, yellow, grey. The field-of-view here corresponds to the pink points in figure 5. The BR and inner filament are detected, but separated into five structures, that can visually be seen adjacent to each other or overlapping. In this figure, only the black points, representing absorbers belonging to the GA, are statistically significant.

Note the very high similarity of the two tables of SLHC/CHMS results (tables 1 and 2), where the only change is reducing the S/N limit of the quasar continuum from 6 to 4. For convenience, the first set of results with the more restrictive S/N limits will be referred to by MST results-1, and the second set of results with the less restrictive S/N limits will be referred to by MST results-2. We deduce that setting restrictive S/N limits (> 4) to the quasar continuum, after already applying S/N limits to the Mg II lines, is not absolutely essential, and possibly adds to the incompleteness of data in a detrimental way. An overview of the results is as follows. (1) In the two lowest redshift slices, in MST results-1 and results-2, there is only one statistically-significant ($> 3.5\sigma$) structure detected in total. The statistically-significant structure belongs to the redshift slice centred at $z = 0.742 \pm 0.060$ from results-1; it is a small group of absorbers located at the lower l.h.s. of the Mg II image, and of no relevance to the BR. In addition, other than a possibility of a thin filament forming in the Mg II image in the redshift slice centred at $z = 0.742$, there is no strong indication or detection of the BR in the two lowest redshift slices. (2) In the central redshift slice, $z = 0.802$, both results-1 and results-2 find one significant structure corresponding to the GA. The BR is separated into four and five (respective to results-1 and results-2) individual, adjacent or overlapping, structures, that are statistically insignificant on their own (see figure 7).

We could again be seeing here limitations of the Mg II method, as was seen with the GA, since the individual structures are overlapping or adjacent. In results-2, the visually-identified BR and inner filament absorbers are mostly detected (46 out of 59 absorbers, 78%) and form

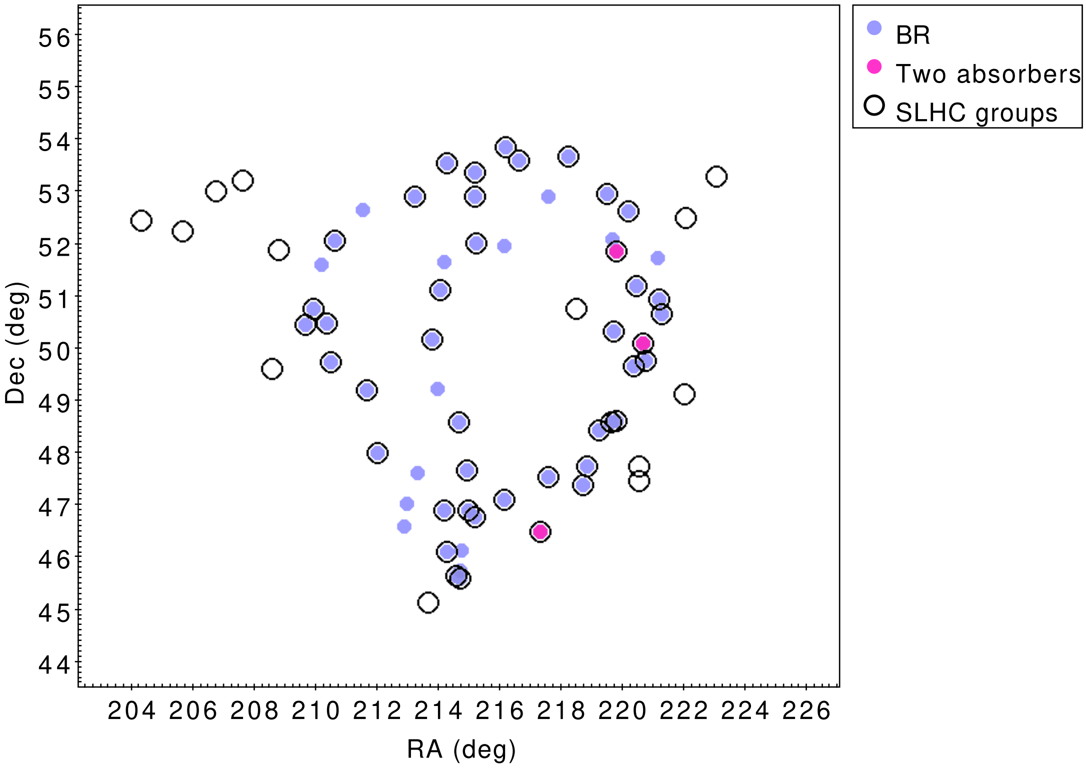


Figure 8. The visually-identified BR absorbers (lilac) and the SLHC-identified absorbers belonging to the 5 candidate structures from results-2 at $z = 0.802 \pm 0.060$ (black circles) that correspond to the visually-identified BR. The pink points indicate the positions where there are two absorbers occurring in a single spectrum. There are 46 out of 59 absorbers in common to the SLHC-identified and visually-identified BR and inner filament absorbers. Of the 13 absorbers in the visually-identified BR and inner filament absorbers that were not connected by the SLHC algorithm, 9 of these occur at the most extreme edges of the redshift range, possibly explaining their exclusion.

the full BR shape (see figure 8). (3) In the two highest redshift slices there are totals of 3 and 1 significant structures from MST results-1 and results-2 respectively. For results-1 the most significant structure in both redshift slices is an arc corresponding to the lower portion of the BR (see figure 9).

However, in results-2, we find the same arc that was detected in results-1, but over multiple structures, highlighting again the nuances of applying the SLHC algorithm to an essentially incomplete dataset. To clarify this, reducing the S/N limits in the quasar continuum from 6 to 4 initially appears inconsequential — i.e., their Mg II images appear on the whole unchanged and there is only a 15% increase of absorbers in the whole field from results-1 to results-2. But, when the SLHC algorithm is applied, then this small increase in absorbers increases the density-scaled linkage scale thus creating more broken structures. We scale the linkage scale of the SLHC algorithm to the density for a general approach to the wide-varying densities in the dataset (mostly due to survey bias). We can see that the absorbers identified as candidate structures corresponding to the arc in the redshift slice $z = 0.862 \pm 0.060$ in results-2 are very similar to the absorbers identified as one candidate structure in results-1. To further clarify this point, compare figure 9 with figure 10, and

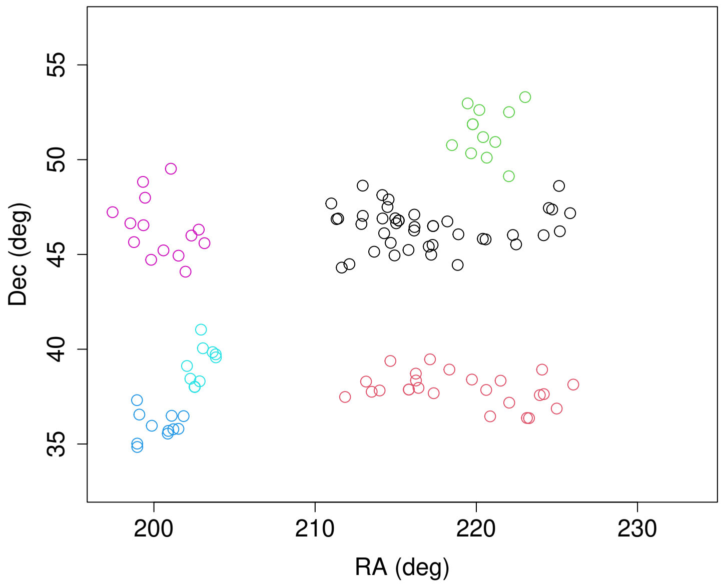


Figure 9. The 6 candidate structures identified by the SLHC algorithm from results-1 in the redshift slice centred at $z = 0.862 \pm 0.060$. The different colours represent the significances, and are ordered from high to low in the following way: black, red, green, blue, turquoise, pink. The two most significant structures belong to the BR and GA respectively, both which are also statistically significant.

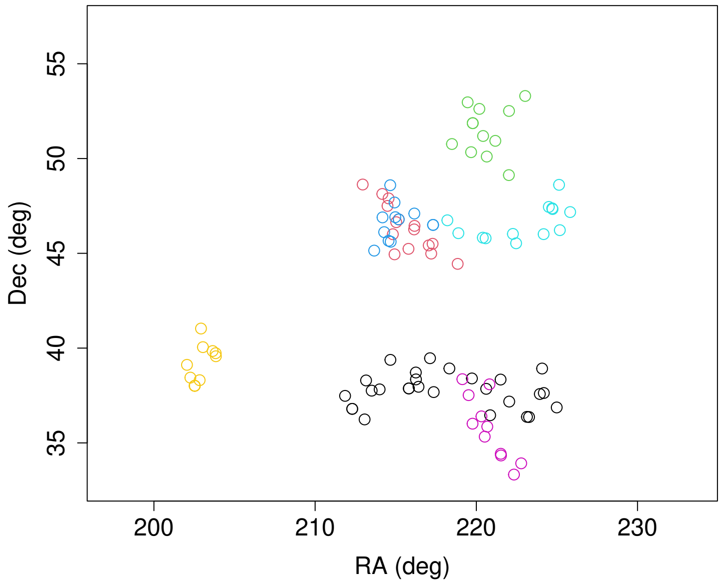


Figure 10. The 7 candidate structures identified by the SLHC algorithm from results-2 in the redshift slice centred at $z = 0.862 \pm 0.060$. The different colours represent the significances, and are ordered from high to low in the following way: black, red, green, blue, turquoise, pink, yellow. The absorbers belonging to the GA are statistically significant, shown in black. By comparing this figure with figure 9 we can see that the absorbers identified by the 3 individual, overlapping or adjacent structures, coloured red, blue and turquoise, are clearly the same absorbers identified by *one* full, statistically-significant structure that was found in results-1, highlighting the complications of applying the SLHC algorithm to an essentially incomplete dataset.

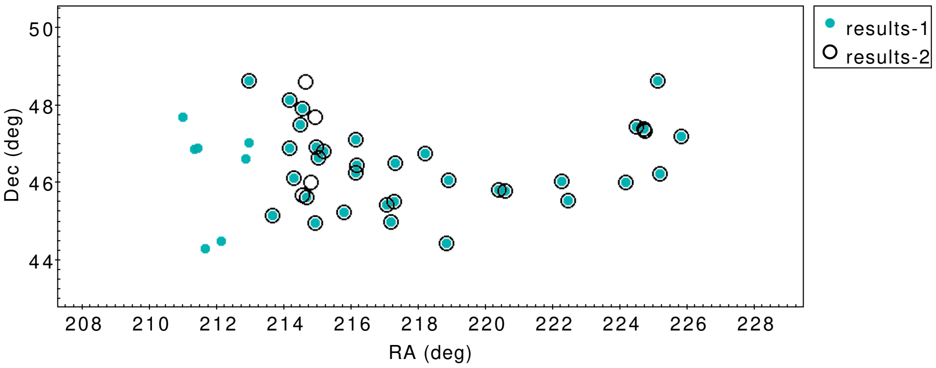


Figure 11. Mg II absorbers in the redshift slice $z = 0.862 \pm 0.060$ belonging to the arc identified by the SLHC algorithm as one, full, candidate structure from results-1 (turquoise) and as three, overlapping or adjacent, candidate structures from results-2 (black circles). There are 42 absorbers belonging to results-1, four of which appear as multiples per probe in two quasars, and 40 absorbers belonging to results-2, the same four of which appear as multiples per probe in two quasars. Both results connect many of the same absorbers, but in results-1, where the linkage scale is higher, there are additional absorbers connected by the SLHC algorithm all appearing on the l.h.s. of the arc. The additional absorbers in results-2 occur in the middle of the arc, and there are much fewer additional absorbers than for results-1, suggesting that these absorbers were identified due to the lowered restriction on the S/N on the continuum.

see also figure 11. There are 42 absorbers in the SLHC-identified arc in the redshift slice $z = 0.862 \pm 0.060$ from results-1 represented by the black points in figure 9, and a combined total of 40 absorbers in the three SLHC-identified candidate structures corresponding to the arc in the redshift slice $z = 0.862 \pm 0.060$ from results-2 represented by the red, blue and turquoise points in figure 10. 35 of these absorbers from results-1 and results-2 are in common to both results, confirming that the SLHC identified a high fraction (67% overlap) of the same absorbers between the two results. In fact, all 7 of the additional absorbers connected by results-1 all occur on the l.h.s. of the arc, indicating that the increased linkage scale in results-1 (due to lower field density) is only *extending* the arc, and not responsible for the central components of the arc. Due to the above, we reason that individual, SLHC-identified candidate structures that are overlapping or adjacent could reasonably be connected as one structure if given a more complete dataset.

In conclusion, the BR appears most obvious in the original, central redshift slice, $z = 0.802 \pm 0.060$, both visually and by the presence of individual SLHC structures overlapping or adjacent on the sky that comprise the visually-identified BR. The strongly, detected arc corresponding to the lower portion of the BR detected in a slightly higher redshift slice ($z = 0.862 \pm 0.060$) is also of particular interest in relation to the BR+GA system.

3.2 Significance: CHMS and MST

The CHMS and MST significance calculations are applied to: the SLHC-identified absorbers; the visually-identified absorbers; and the FilFinder-identified absorbers. In addition, the Alpha Hull algorithm is applied to the visually-identified absorbers for an estimate of the volume, overdensity and significance using simple Poisson statistics.

	No. Mg II absorber members	CHMS signif.	MST signif.
BR only	51	3.3	4.0
BR all	62	5.2	4.1

Table 3. The CHMS and MST significances for the BR-all and BR-only absorbers. The CHMS significance is dependent on the volume and number of absorbers so, clearly, removing the inner absorber members and keeping the volume the same will reduce the CHMS significance. However, for the MST significance, the volume is not directly related, but instead it is related to the MST mean edge lengths. This is clearly shown in the results as the MST significance stays mostly the same, at $\sim 4\sigma$, and the CHMS significance decreases from 5.2σ to 3.3σ by excluding the absorbers enveloped by the BR.

3.2.1 SLHC-identified Mg II absorbers

To determine the significance of the BR in its entirety, we take the SLHC-identified absorbers from results-2 that make up the BR (i.e., the BR in figure 6), and apply the CHMS algorithm. Remember that the CHMS has the ability to assess the significance of a structure by comparing the observed convex-hull volume with the volumes that would be expected for a set of random distributions of those same absorbers at the control density of absorbers for the same redshift interval. Based on the definition of the CHMS volume and significance calculation, the algorithm is optimal when applied to clumpy structures, with no obvious gaps, holes or curvature that would lead to an overestimation of the volume. Clearly, this is not the case with the BR, having a large volume mostly unoccupied by absorbers in its centre. Instead, the MST-significance calculation introduced by [36] could be more appropriate, which uses the mean MST edge-length between neighbouring data points. Using both methods we then find that the SLHC-identified BR in its entirety has a CHMS significance of 3.6σ, and an MST significance of 4.7σ. Both tests indicate statistical significance, but we can see that CHMS has likely overestimated the BR volume leading to a much lower significance compared with the MST-significance test.

3.2.2 Visually-identified Mg II absorbers

We then take the visually selected absorbers of the BR, and everything within the BR, for which there are a total of 62 absorbers (see figure 2), and apply the CHMS to these absorbers only. The CHMS algorithm then calculates a significance of 5.2σ. The significance calculated here is likely the upper limit estimate for the BR, as the algorithm was applied to those absorbers that were visually-selected. We can similarly repeat this work for the BR-only absorbers, as well as compare the MST significance with the CHMS significance. Clearly, removing the BR innards will reduce the significance calculations for the CHMS, since the volume will remain the same but the number of absorbers will be reduced. In contrast, for the MST significance, the mean MST edge length may not be greatly affected by removing the BR innards, as seen in table 3. We find that on both occasions, using the BR-all and BR-only absorbers, the MST significance is roughly the same, at $\sim 4\sigma$. In contrast, for the CHMS calculation, the significance drops from 5.2σ to 3.3σ after removing the absorbers contained within the BR. However, even after removing all of the BR inner absorbers the CHMS significance is still greater than 3.0σ, but not quite reaching 3.5σ which is the usual standard we apply for comparing structures. Of course, the CHMS calculation on the BR-*only* absorbers gives a drastic under-representation of the true significance since we have forcefully removed absorbers contained within the BR while keeping the unique volume the same.

For comparison, we can estimate the volume of the BR using the 2D Alpha Hull algorithm. First, the Alpha Hull area of the BR is repeatedly calculated 500 times from the BR by drawing a cloud of points in a circle with radius equal to half the mean MST edge length around each of the Mg II absorbers in the BR. Then, the area is multiplied by the physical size of the redshift range of the BR absorber members. In this manner, we are calculating the volume of a somewhat cylindrical, tube shape. The benefits of this method versus the CHMS is that we can eliminate the central region of the BR where there are very few absorbers. The downside of this method is that, although the absorbers on the sky make up a ring shape, we later see that the 3D distribution of the absorbers is more of a coil shape, so we are again overestimating the volume of the BR (and underestimating the overdensity and significance of the structure). Nevertheless, using this method we obtain a volume of $21.8 \times 10^6\,\mathrm{Mpc}^3$, an overdensity of 0.75 and a significance of 4.0σ for the number of absorbers in this volume based on Poisson statistics. The significance calculated from the Alpha Hull here, although simple, agrees with the MST-significances of the SLHC-identified absorbers and the visually-identified absorbers, which adds further confidence to the statistical assessment of this structure.

3.2.3 FilFinder-identified Mg II absorbers

Finally, we have seen the CHMS algorithm as well as the MST significance test applied to the SLHC-identified and visually-identified absorbers, so now we will apply both algorithms to the FilFinder-identified absorbers. The FilFinder work is discussed next, in section 3.3, and we are referencing figure 13(d) for the work in this section. The FilFinder algorithm is applied to 2D pixel images, so the physical absorber points are irrelevant to the algorithm, but we can estimate the absorbers identified by the filament by including those that are attached to the filament in the figure. Our estimation of the absorbers connected by the FilFinder algorithm is shown in figure 12. We can intuitively see that a volume containing all of the FilFinder-identified absorbers will include many large, empty sub-volumes — e.g., the centre of the BR and the volume between the ‘extended filaments’ connected to the BR. The BR volume was also overestimated with all previous identifications of the BR, which can be confirmed with the comparison of, in particular, the BR-all with the BR-only absorbers, and the CHMS with the MST significances. So, when the CHMS algorithm is applied to the FilFinder-identified absorbers, it calculates a 2.5σ significance. In contrast, the MST-significance test estimates that the FilFinder-identified absorbers have a significance of 3.6σ, which is generally in agreement with all previous estimates of the BR significance using the MST calculation and the Alpha Hull Poisson statistics. So, again, the BR volume is considerably overestimated, this time using the FilFinder-identified absorbers, and the CHMS significance is therefore underestimated as a consequence, which can be confirmed with the use of the MST-significance test.

JCAP07(2024)055

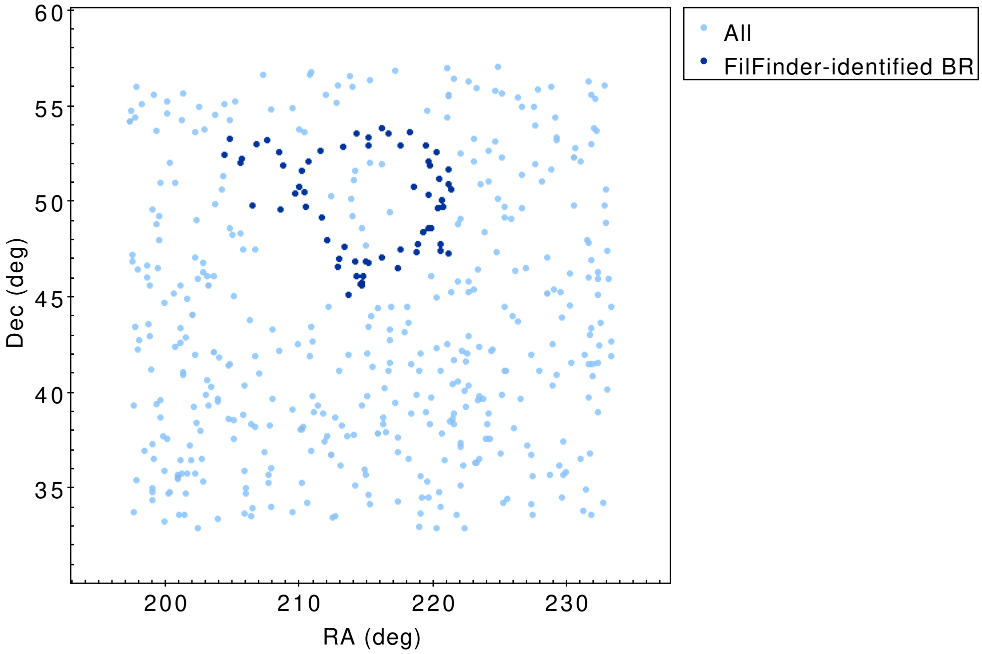


Figure 12. The Mg II absorbers belonging to the whole BR field, represented by light blue points, and the FilFinder-identified BR absorbers represented by dark blue points.

	SLHC	Visual BR-all	Visual BR-only	FilFinder
CHMS	3.6	5.2	3.3	2.5
MST	4.7	4.1	4.0	3.6

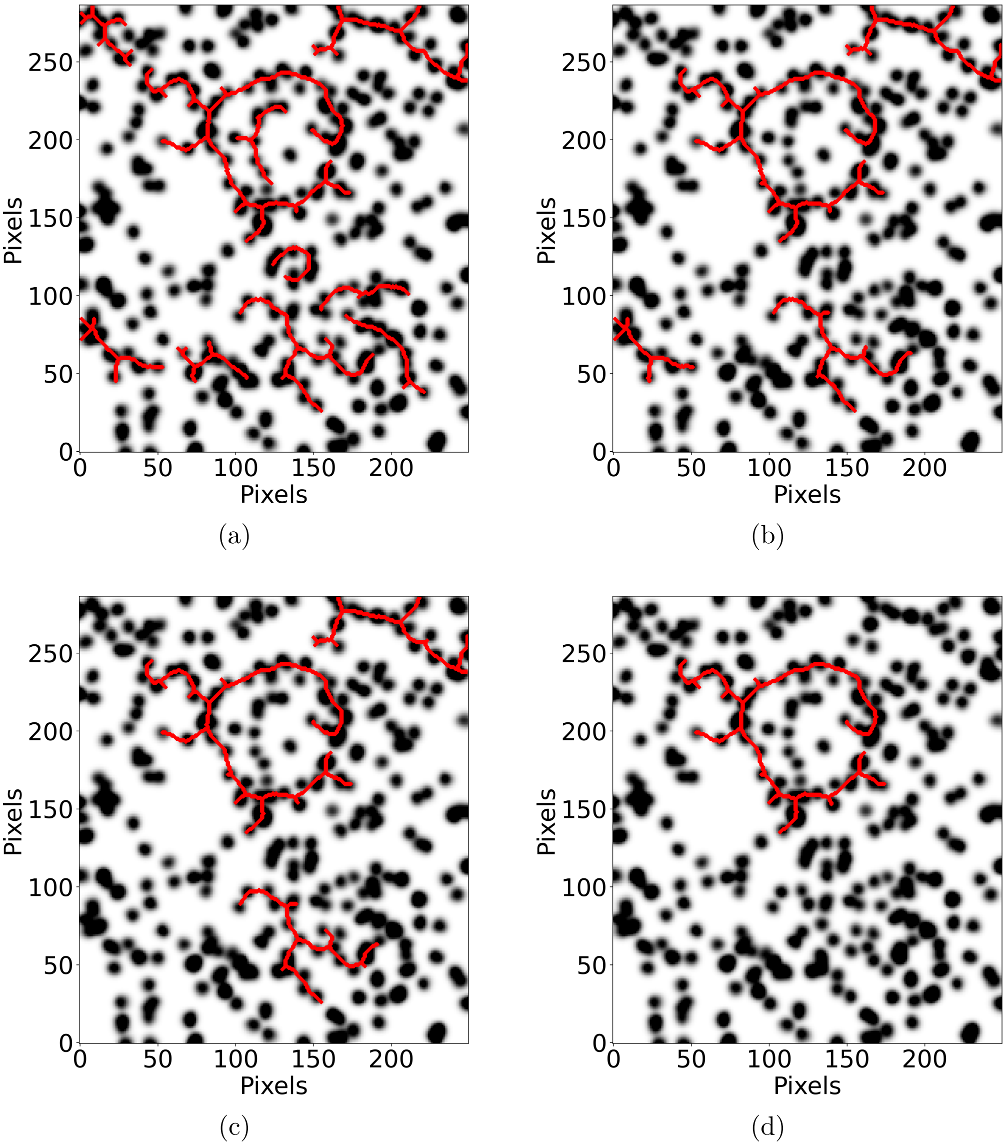
Table 4. The CHMS and MST significances of the BR calculated for each set of identified absorbers (SLHC, visual BR-all, visual BR-only and FilFinder).

3.2.4 Summary of significances

We have shown our application of the CHMS and MST significance calculations to four sets of uniquely-identified BR absorber members. Table 4 summarises all of the above significance calculations. The results are: the total mean significance of the BR (all results from table 4 considered) is $(3.88 \pm 0.83)\sigma$; the mean CHMS-significance is $(3.65 \pm 1.13)\sigma$; and the mean MST-significance is $(4.10 \pm 0.45)\sigma$. From these results we can see that the variation of the CHMS-significance is a considerable fraction of the average, indicating its results are to be taken with caution. On the other hand, the MST-significance has much lower spread, indicating that for the purpose of analysing the BR, the MST-significance test may be more appropriate, given the difficulty of defining a volume around a ring-like structure without incorporating overestimations of the volume (as is the case with the CHMS algorithm).

3.3 Filament identification algorithm

FilFinder is a filament identification algorithm created by [37]. It uses mathematical morphology to identify filaments ranging in size, shape and brightness on a 2D pixel image. The



(a)

(b)

(c)

(d)

Figure 13. Applying the FilFinder algorithm to the standard BR field and increasing the size threshold incrementally to show the elimination process of small filaments. All axes are labelled in pixels, where 1 pixel = $4^2\,\mathrm{Mpc}^2$. (a) The standard parameters, for which adaptive threshold is 18 pixels, smooth size is 12 pixels, and size threshold is 576 pixels. For the remaining figures we incrementally increase the size threshold. (b) Size threshold is 800 pixels. (c) Size threshold is 1000 pixels. (d) Size threshold is 2000 pixels.

algorithm was intended for use in small (10^1–10^3 pc), gaseous areas, such as star formation regions and the interstellar medium [38–40]. So, applying it to cosmological LSS is new, and we have had to make adaptations in the parameter settings to suit the data. We use the following input parameter settings. (1) Adaptive threshold — the expected width of a typical filament. We choose a value (in number of pixels) equivalent to a filament that is only one absorber wide. Since the Mg II absorbers in the Mg II images span a diameter of approximately 12 pixels, the adaptive threshold is set to 12. (2) Smooth size — scale on which to smooth the data. Our data is already smoothed and flat-fielded, so it is possibly unnecessary to smooth beyond the size of a Mg II absorber: i.e., we shall implement a smoothing size that does not affect the already smoothed Mg II absorbers. Therefore, we set smooth size to a value of 12 pixels. (3) Size threshold — the smallest area to be considered a filament. Using again the estimated value for the number of pixels across a single Mg II absorber, we set this value at $4 \times 12^2 = 576$ pixels. That is, the minimum area to be considered a filament is made up of 4 Mg II absorbers (with the absorbers imagined as squares rather than as circles, for simplicity).

We use the FilFinder package available in Python to objectively identify the filaments present in the BR field. For the above parameters we find that there are no visually-obvious consequences for relatively small changes (e.g., within a few percent of the chosen value). However, much larger changes in the parameter choices render noticeably different, and consequential, results. For this reason we experiment with changing the standard parameter values above to gain a better understanding of the FilFinder algorithm when applied to cosmological data. The FilFinder algorithm is applied multiple times, with different parameter settings, on the Mg II image containing the BR (figure 1), and the results are discussed below.

First, the adaptive threshold and smoothing size are related in a way that increasing one has an almost indistinguishable effect from decreasing the other, for reasonable values (e.g., < 40 pixels, since larger values flag a warning within the FilFinder algorithm). More specifically, the most meaningful and interesting effects can be found when one focuses on the ratio between the values. This can be understood intuitively since smoothing size is the scale on which to smooth the filaments and adaptive threshold is the typical width of those filaments, the size of which is affected by the smoothing size. When the adaptive threshold and smoothing size ratio is $1 : 2$ there is an undesirable effect of creating 'blurred' masking borders over the filaments; this is when the borders of the mask around individual filaments overlap. However, the opposite is true when the adaptive threshold and smoothing size ratio is $2 : 1$, which has the effect of creating more concise borders around the filaments, but it also drastically reduces the number of filaments that can be detected. With this in mind, we choose to slightly increase the adaptive threshold from 12, a single Mg II absorber, to 18, 1.5 times a Mg II absorber, while keeping the smoothing size at 12 as usual, thus creating a ratio of $3 : 2$ for the adaptive threshold and smooth size.

Secondly, the size threshold simply determines the minimum size to be considered a filament. Increasing the size threshold incrementally and comparing results demonstrates the elimination process of small filaments. So now using the updated parameter values, where the adaptive threshold is set at 18 pixels and the smoothing size is kept at 12 pixels, we incrementally increase the size threshold and show the filaments that survive the elimination process (see figure 13). Immediately it becomes clear that the BR is the largest, and most dense, filament in the field. The BR filament is the only filament left with the largest size threshold limit. In fact, the BR is not eliminated until the size threshold exceeds 4200 pixels. Using the elimination process with the FilFinder algorithm has given an impressive indication for the size and uniqueness of the BR compared with the rest of the field.

Next, we apply the same method of analysis to the SDSS DR16Q quasars in the same field and the same redshift slice as the Mg II absorbers containing the BR for comparison

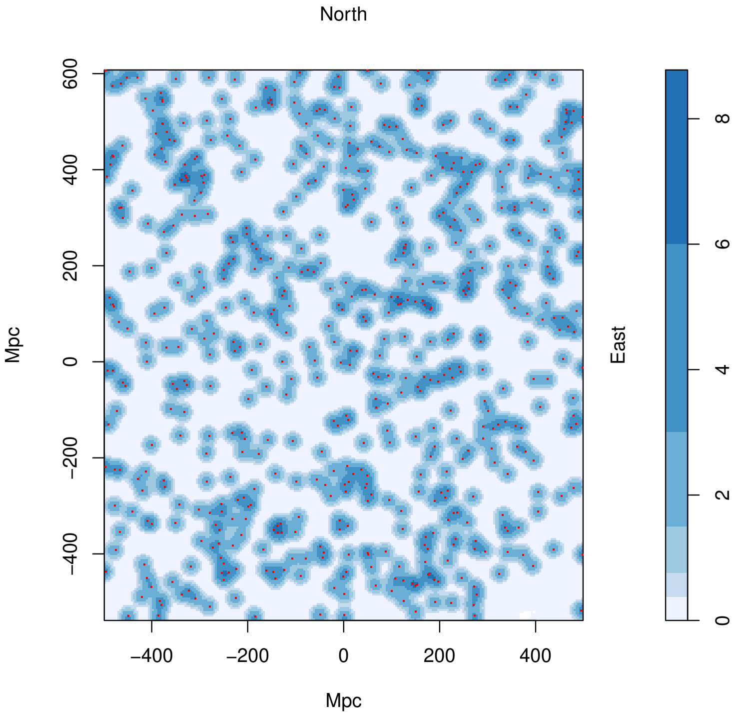


Figure 14. The tangent-plane distribution of quasars in the redshift slice $z = 0.802 \pm 0.060$. The blue contours, increasing by a factor of two, represent the density distribution of the field quasars which have been smoothed using a Gaussian kernel of $\sigma = 11\,\text{Mpc}$. Magnitude limits have been applied to the quasars such that $i \leq 20.0$. The field-of-view corresponds to the small, pink area seen in figure 5. There is no strikingly obvious structure by eye, but when the FilFinder algorithm is applied to this field it can detect a filamentary ring-like shape that coincides with the BR.

with an independent data source. We want to reduce the noise of the high number density of quasars so we apply an i-magnitude (i) limit of $i \leq 20.0$, such that only the intrinsically very bright quasars are included. The field size and redshift interval for producing the quasar image is the same as that for the Mg II absorbers from figure 1: that is, the quasars are chosen to be those that lie in the same field as Mg II absorbers, not to be confused with the background quasars responsible for the Mg II absorbers. The quasar image is seen in figure 14. The FilFinder algorithm is applied to the quasar image in the same manner as described above with the Mg II absorbers, with the following parameter settings: adaptive threshold = 18, smoothing size = 12 and size threshold = 576 (see figure 15(a)). Given the much higher density of the field quasars compared with the Mg II absorbers, even after i magnitude cuts have been made, we can see that there are generally more filaments identified (figure 15(a)). Incrementally increasing the size threshold will remove the small filaments and leave only those able to survive the cuts — so just the large filaments of interest. Figures 15(a)–15(d) show the results from FilFinder applied to the field quasars with increasing size threshold.

When the FilFinder algorithm is applied to the field quasars, and the size threshold is incrementally increased, we are left with only a ring-like filament that coincides with the BR. This is particularly interesting as it demonstrates from independent corroboration that the quasars also follow a similar shape to the Mg II absorbers. We can also visualise

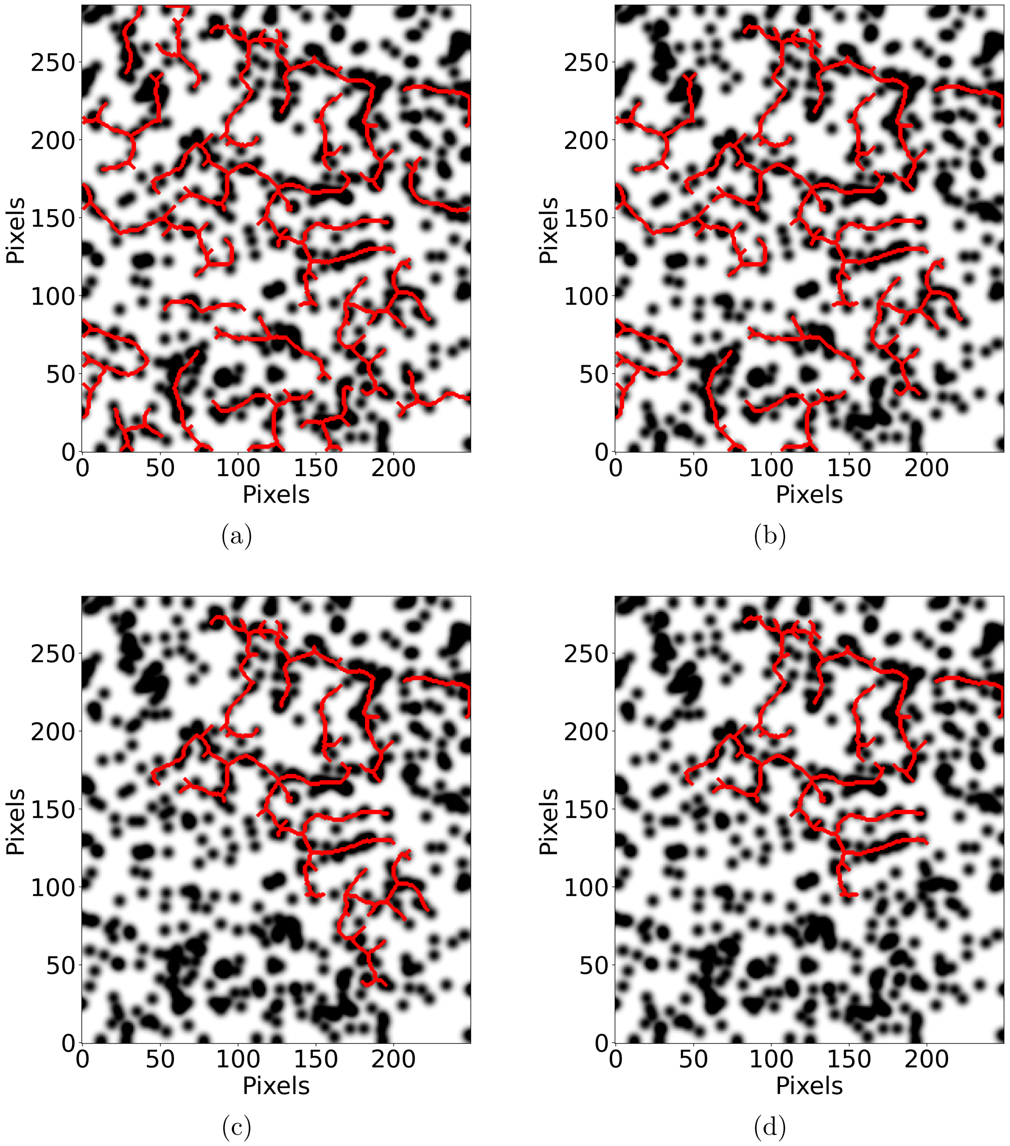


Figure 15. Applying the FilFinder algorithm to the field quasars corresponding to the standard BR field and incrementally increasing the size threshold to show the elimination process of small filaments. All axes are labelled in pixels, where 1 pixel = 4^2 Mpc^2. (a) Using the standard FilFinder parameters where adaptive threshold is 18 pixels, smooth size is 12 pixels, and size threshold is 576 pixels. For the remaining figures we incrementally increase the size threshold. (b) Size threshold is 1000 pixels. (c) Size threshold is 2000 pixels. (d) Size threshold is 4000 pixels.

the relationship of the Mg II absorbers with the bright field quasars by superimposing the quasar image onto the Mg II image (see figure 16). In section 4.1 we further investigate the observational properties of both field quasars and DESI clusters in the BR field.

3.4 Cuzick and Edwards test

The Cuzick-Edwards (CE) test [47] was created to assess the clustering of cases in an inhomogeneous population. It is a 2D, case-control statistical method that deals with variations in spatial populations. This method has shown to be useful for the particular data

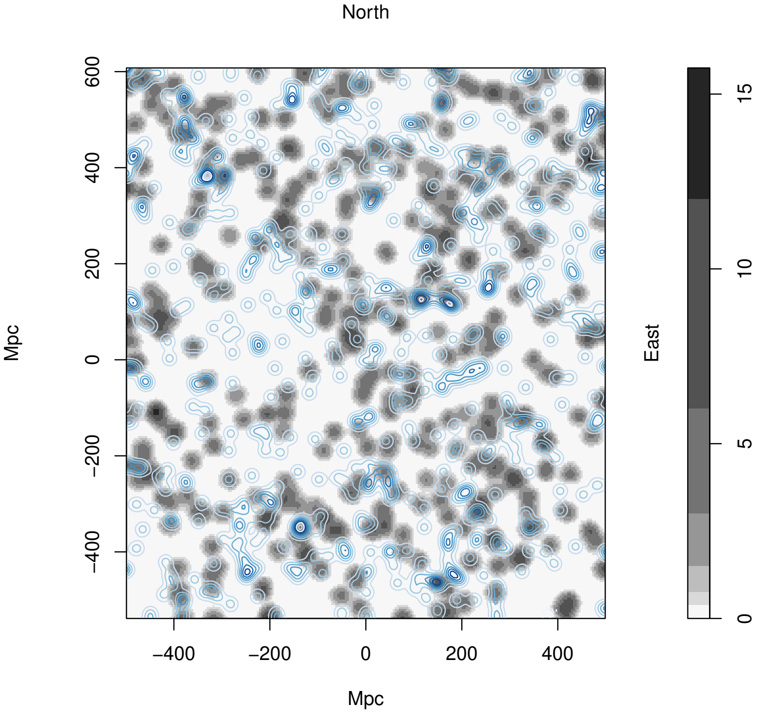


Figure 16. The tangent-plane distribution of Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ superimposed with the tangent-plane distribution of quasar positions in the same redshift slice (not to be confused with the background probes). The grey contours, increasing by a factor of two, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$, and flat-fielded with respect to the distribution of background probes (quasars). The blue contours, increasing by a factor of two, represent the density distribution of the field quasars which have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$. In the Mg II image, S/N limits of 4, 2 and 4 were applied to the $\lambda_{2796}, \lambda_{2803}$ Mg II lines and quasar continuum, respectively. In the quasar image, magnitude limits were applied such that $i \leq 20.0$. In general, we find that the blue contours follow the grey contours.

that we are working with; the Mg II are the cases to be assessed, which are affected by the inhomogeneous spatial distribution of the background probes (quasars). This method assesses the statistical significance of clustering by assessing the occurrence of cases within the k nearest neighbours, but it cannot, however, assess the significance of individual candidate structures. Therefore, it is a useful additional method to analyse the field containing the BR.

The CE test was first applied to cosmological (and to our knowledge, astrophysical) data when analysing the GA (section 2.3.3 in Lopez22); before this, the method had mainly been used in the context of medical research (e.g. for patterns in diseases across varying population sizes). We can now follow the steps and process presented in Lopez22 for the BR. (We are using the application *qnn.test* in the R package SMACPOD by [48].)

As mentioned previously in Lopez22, the power of the CE test is affected by the control-case ratio, and a ratio between 4 and 6 is found to be optimal [47]. For this reason, we continue to use a control-case (i.e., probes to absorbers) ratio of 5 : 1, which, for the BR field, requires removing ∼ 90% of the probes. The ∼ 10% remaining probes are randomly selected for each of the 200 runs of the *qnn.test*, which itself is computed for 499 simulations (the default setting).

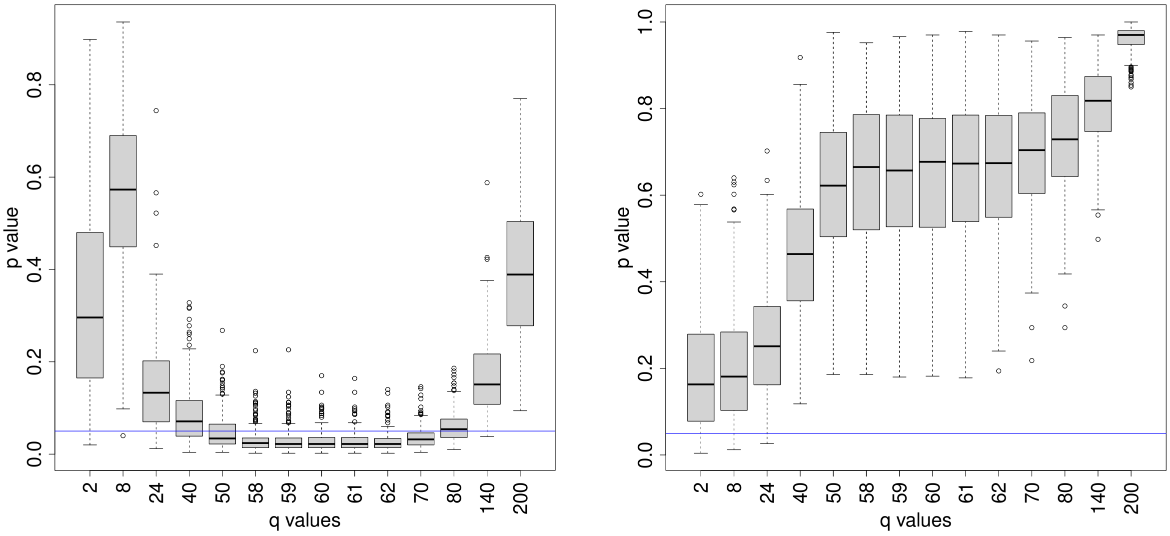
We begin by assessing the BR field, and then we successively shrink the FOV to 'zoom in' on the BR, which is presumably the main feature contributing to any clustering patterns detected in the field. Unlike the GA, the BR is a full ring shape, so the successive zooming is limited in both axes. Therefore, we apply only two successive zooms to the BR, beyond which the BR would extend beyond the FOV. We then apply the CE test on four other fields that are unrelated and detached from the BR field, for comparison with the BR field. The four unrelated fields are chosen to be at the same redshift as the BR field ($z = 0.802 \pm 0.060$), the same field size (equivalent to the second 'zoom' of the BR field) and the same declination ($\delta = 50°$), so that the only change from the BR field to the unrelated fields is to move the right ascension (RA) centre point — the unrelated fields are then outside the BR field but within version-1 of the control area (see figure 5). For the BR field, we repeat the CE test (with 200 runs and 499 simulations) 10 times altogether to investigate how the random selection of probes (controls) could affect the CE statistics. The distribution of p-values calculated from applying the CE test to the BR field indicates tentative significant clustering ($p \leq 0.05$) for a range of chosen q (k) values. The median p-value dropped to a minimum of $p = 0.0238 \pm 0.0015$ over the 10 repetitions of the CE test at $q = 61$, corresponding to a significance of 2.0σ. However, in the four other unrelated fields there was no (with the exception of exactly two data points) significant (significant is here defined as $p \leq 0.05$) clustering detected, suggesting that the clustering seen in the BR field is special (see figure 17) — as was similarly found with the GA previously. In one of the unrelated fields there were two points, at $q = 2$ and $q = 8$, where the median p-value dropped below the $p = 0.05$ threshold, which is here considered tentative significant clustering. However, since it was at a much lower q value, the CE test was likely here detecting very small-scale clustering, and not related to the same type of clustering seen in the BR (and GA) field. Overall, the clustering seen in the BR field mimicked the clustering that was also seen in the GA field, but here at a tentative ($p \leq 0.05$) significance level; the clustering in the BR field was inconclusive at the $\sigma > 3.0$ significance level.

4 Observational properties

4.1 Corroboration with independent data

Earlier, in section 3.3, we saw that the bright field quasars have identified filaments similar to those of the Mg II absorbers. It was also shown that there is a plausible visual association of the field quasars and Mg II absorbers from superimposing the contour maps of each dataset (figure 16). We can continue to use the contour maps from the quasars, and now also the DESI clusters [49], looking at a larger FOV, and investigate the visual association of the independent datasets with the MgII absorbers on a larger scale (see figure 18).

The visual investigation of quasars and DESI clusters with Mg II absorbers indicates that generally there is plausible association of both independent datasets with the absorbers. The association can be seen clearly when comparing the 'voids' in the data; where there are no absorbers there tend also not to be very many quasars or DESI clusters. The visually most striking association is seen in figure 18(d) where the green contours of the DESI clusters seem to follow the same filamentary trajectory as the Mg II absorbers in many cases. Interestingly,



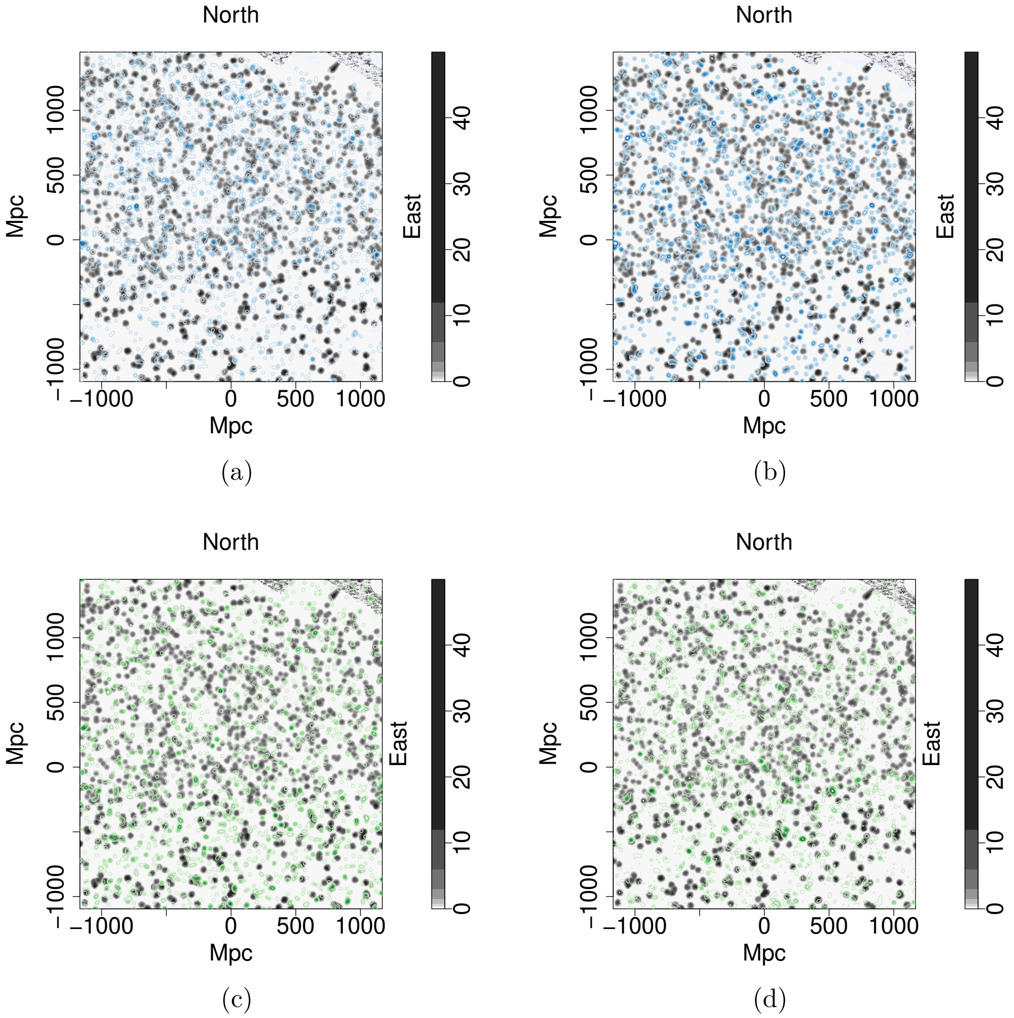
(a) (b)

Figure 17. The Cuzick and Edwards (CE) test applied to the second 'zoom' of the BR field and an unrelated field (a and b respectively) for comparison. The figures show a box-plot of the distribution of p-values over 200 runs of 499 simulations as a function of chosen q (k) value. (a) The BR field shows tentative significant clustering ($p \leq 0.05$) between $q = 50$ and $q = 70$, reaching a minimum of $p = 0.022$ between 59 and $q = 62$ corresponding to a significance of 2.0σ. Note, this is *one* of the figures produced from the 10 repeated tests. (b) The unrelated field shows no significant clustering on any scale.

however, in figure 18(c), similar trajectory is also seen, but on other Mg II absorbers that were missed in figure 18(d). The two DESI cluster figures 18(c) and 18(d) have richness limits applied of $R \leq 22.5$ and $R \geq 22.5$, respectively, so the crossover of the two figures is minimal. (The median richness limit for the DESI clusters is 22.5 [49].) We might thus have a potentially useful technique of investigating the Mg II absorbers and their association with low and high richness clusters, from which we might learn more about the physical origin of the Mg II absorbers. We can also see in figure 19 the DESI clusters of all richnesses plotted in green contours over the BR Mg II absorbers. Comparison with figures 18(c) and 18(d), which are for $R \leq 22.5$ and $R \geq 22.5$, indicates that figure 19 shows the strongest association of the DESI clusters with the Mg II absorbers. Of course, the DESI clusters will have much larger redshift errors ($\Delta z \sim 0.04$) than the Mg II absorbers and quasars, so the structures in the DESI clusters will be blurred. Given that the association seems quite clear, then presumably smaller redshift errors would lead to an even clearer association.

4.2 Viewing the BR from other angles

The BR and GA discoveries were made unexpectedly with the method of intervening Mg II absorbers, and so are subject to observational bias — we have first detected the signal of a LSS by observing a curious shape and/or visual overdensity in the initial Mg II images that we later assess statistically. Specifically, we are only observing these LSS candidates from one line of sight (LOS), i.e., a 2D projection on the sky of the 3D matter distribution. Thus, from a different viewing angle, or LOS, the LSS candidates could look entirely different, or



(a) (b) (c) (d)

Figure 18. Density distribution of the Mg II absorbers in the redshift slice $z = 0.802 \pm 0.060$ in a large field-of-view represented by the grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11\,\mathrm{Mpc}$ and flat-fielded with respect to the background quasars. The additional contours (in blue and green) represent the superimposed density distribution of the field quasars (a and b) and the DESI clusters (c and d), respectively. (a) The field quasars, represented by the blue contours, are restricted to $i \leq 20.0$. The blue contours appear to have a plausible association with the grey contours. The i magnitude limit is not very restrictive so the blue contours are visually noisy compared with the Mg II absorbers. Note, these quasars were found to have a FilFinder filament that linked closely to the shape of the BR, so clearly it is the visual impression that is difficult to determine due to the noise. (b) The field quasars, represented by the blue contours, are restricted to $i \leq 19.5$. With the reduced quasars (compared with a) a clearer trend of quasar and Mg II absorber association can be seen. Generally, the dense clumps of Mg II absorbers appear to have associated quasars rather than the thin Mg II filaments. (c) The DESI clusters, represented by the green contours, are restricted to $R \leq 22.5$. The green contours generally follow the grey contours, although the richness limit reduces the number of DESI clusters in the field. (d) The DESI clusters, represented by the green contours, are restricted to $R \geq 22.5$. Again, it can be seen that the green contours generally follow the grey contours. In particular, the lower half of the BR has a strong filament of DESI clusters following the filamentary shape of the Mg II absorbers. Note, the larger field-of-view of the above figures crosses two of the SDSS borders where the quasar coverage drops sharply (see figure 5).

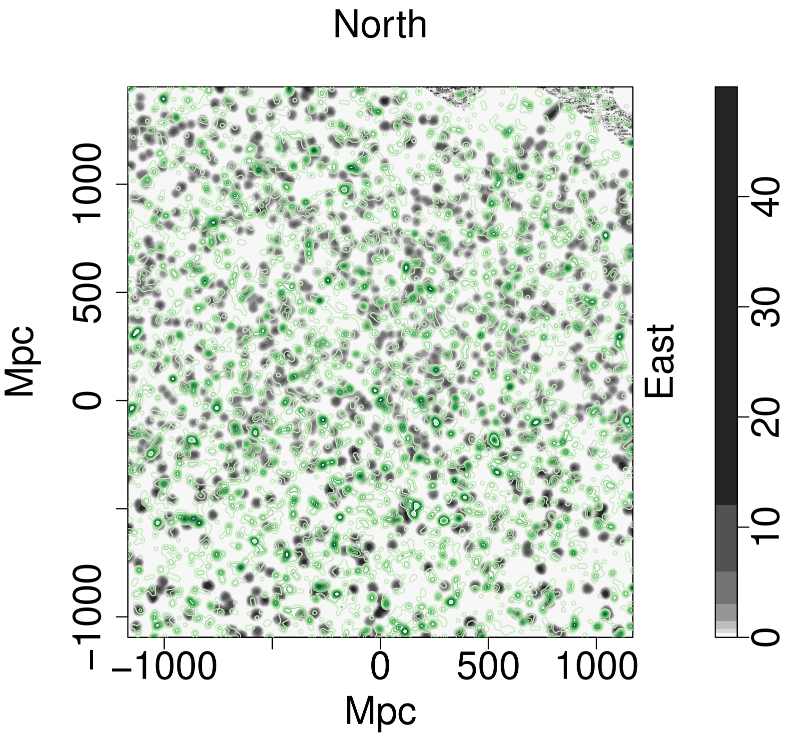


Figure 19. All of the DESI clusters with richnesses $0 < R \leq 300$, represented by the green contours. Here it can be seen that the green contours generally follow the grey contours quite well. Compare with figures 18(c) and 18(d), which are for $R \leq 22.5$ and $R \geq 22.5$. Note, the larger field-of-view of the figure crosses two of the SDSS borders where the quasar coverage drops sharply (see figure 5).

furthermore, there could be LSS candidates that have a seemingly average, or 'uninteresting' distribution of absorber members from our LOS that may be arcs, rings or interesting filamentary shapes from a different viewing angle.

We investigate how different viewing angles change the perception of the BR, which first requires redefining the coordinate system on which the BR-field absorber members are projected, described here. An initial, orthogonal, 3-vector coordinate system is defined such that u_0, v_0, w_0 are closely linked to $x_{\rm prop}, y_{\rm prop}, z_{\rm prop}$. The initial normal vector (w_0) is defined as the proper coordinate to the mean $x_{\rm prop}, y_{\rm prop}, z_{\rm prop}$ of the absorber members (which can be thought of as the original LOS). Then, the plane perpendicular to the normal is rotated such that u_0 points towards the most easterly absorber. Finally, all the absorber members are projected onto the new plane (see figure 20). Hence, we call this method the 'project-plane method'. With the initial coordinate system defined, we can define any new normal vector (w) as some combination of the initial u_0, v_0, w_0.

In figure 20 and other similar figures following, the colours of the absorbers indicate the redshift of each absorber, with the high-z absorbers represented by a darker shade, and vice versa (see the key on the top left of the figures). The numbers associated with each absorber are a simple ID system, labelled 1 to 51 in ascending order from the most easterly absorber to the most westerly (i.e., in descending order of RA). The ID system remains the same throughout all subsequent rotations of the plane projections so that the numbers can be used

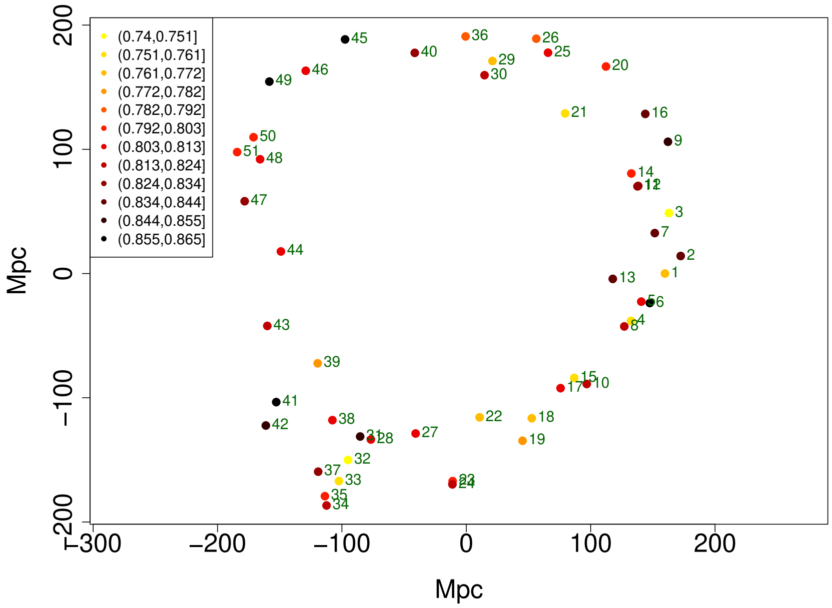


Figure 20. The visually-identified BR absorber members projected onto the plane perpendicular to the initial normal vector, w_0. The x-axis points towards the u_0 direction and the y-axis points towards the v_0 direction. By comparing the figure here with figure 2 one will notice the slight rotation of the BR projection, indicating the small misalignment of the tangent-plane coordinate system with the new u_0, v_0, w_0 system. The key in the top l.h.s. of the figure indicates the redshifts of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number.

to orientate oneself. In this manner we are viewing the BR from different viewing angles, similar to taking snapshots of the 3D structure from different LOSs. Note, the observational analysis presented on the BR here is based on the visually selected BR absorber members, which are the blue points in figure 2.

We then redefine the normal with combinations of the original, orthogonal, 3-vector system to give different projected planes. The newly-defined vectors u, v, w are similarly orthogonal. Figures 21–23 have different viewing angles of the BR, and the observational analysis of each plane projection follows.

In figure 21 the new normal is set to u_0, which can be thought of as the side-on view of the BR. To emphasise this point, the redshift colours are banded vertically indicating that the u-axis has essentially become the redshift z-axis. Of interest in this figure is the curious backwards 'S' shape which is in both the dark-coloured high-z absorbers and the light-coloured low-z absorbers. The impression is that the BR has three distinct redshift bands, and the absorbers in each band create a noticeable backwards 'S' shape, but more so in the nearest and farthest redshift bins (the lightest and darkest coloured absorbers). It is not entirely clear what the backwards 'S' shape might indicate, or if it warrants further investigation. Initially, it could be thought that the absorbers belonging to the three distinct redshift bands may also be associated particularly with some of the SLHC identified groups from figure 7, but this is not the case. However, instead we find that almost all (one exception, starred in the list below) of the nearest redshift absorbers (those forming the backwards

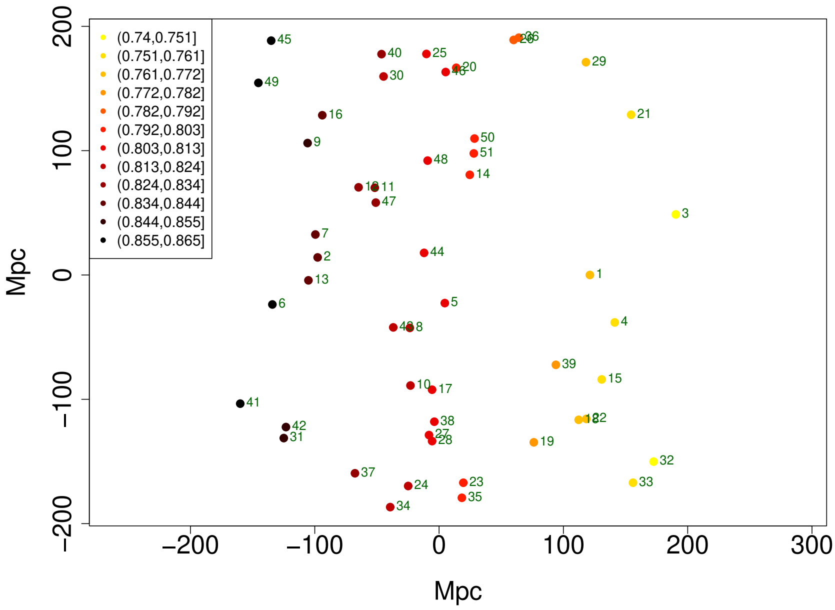


Figure 21. The BR absorber members projected onto the plane perpendicular to u_0. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top l.h.s. of the figure) and the small numbers paired with each data point indicate their unique ID number.

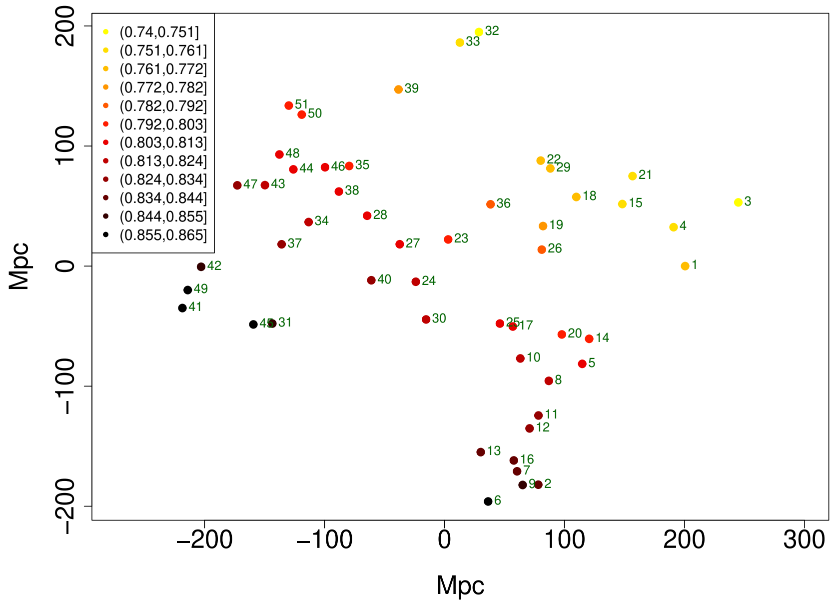


Figure 22. The BR absorber members projected onto the plane perpendicular to v_0. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top l.h.s. of the figure) and the small numbers paired with each data point indicate their unique ID number.

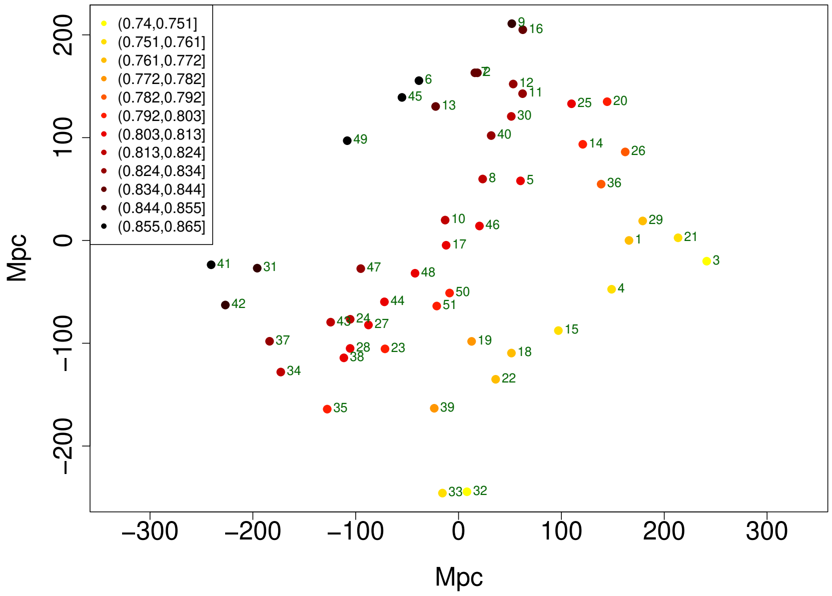


Figure 23. The BR absorber members projected onto the plane perpendicular to $u_0 - v_0$. The new u direction points towards the most easterly absorber, as usual. The colours represent the redshifts of the absorbers (see the key in the top l.h.s. of the figure) and the small numbers paired with each data point indicate their unique ID number.

'S' shape with the light-coloured absorbers) belong to one half of the BR closest to the r.h.s. This can be confirmed by cross-correlating the number IDs in figure 20 and figure 21. (IDs of absorbers contained in the near-redshift band, i.e., the light-coloured backwards 'S' shape: $36, 26, 29, 21, 3, 1, 4, 39*, 15, 22, 18, 19, 32, 33$. 39 has been starred as the one absorber belonging in the near-redshift band that does not sit with the rest of the absorbers on one half of the BR.) So the near-redshift absorbers form an arc of a circle closest to the r.h.s. of the BR. Later we will see that this further implies a somewhat spiral or 'cork-screw' interpretation of the BR, that is aligned face-on with our LOS.

In figure 22 the new normal is set to v_0, which can be thought of as viewing the BR from the bottom, looking up. Now we can see a thin, central arc which is formed from the most central redshift absorbers, with clumps of absorbers at lower and higher redshift either side of this arc. If the BR had been discovered from this particular angle, had we been located at a different position in the Universe, then we may have named this structure the 'Big Arc'. This highlights the observational bias mentioned earlier: our viewing angle only allows one LOS in which to discover interesting structures and filaments when using the Mg II image method. However, it also presents the opportunity to re-examine LSS candidates from multiple viewing angles.

In figure 23 the new normal is set to $u_0 - v_0$, which can be thought of as looking at 45° through the BR towards the south-east direction. The BR now very clearly resembles that of a coil, and the inner filament is very thin and filamentary. The high-z and low-z absorbers form ovoid shapes connected by the inner filament, although the high-z absorbers appear more broken and sparse compared with the rest of the coil.

All three viewing angles, plus the original face-on view of the BR, indicate that the BR structure is a coil shape with a thin, flat, dense, central component. The flatness of the

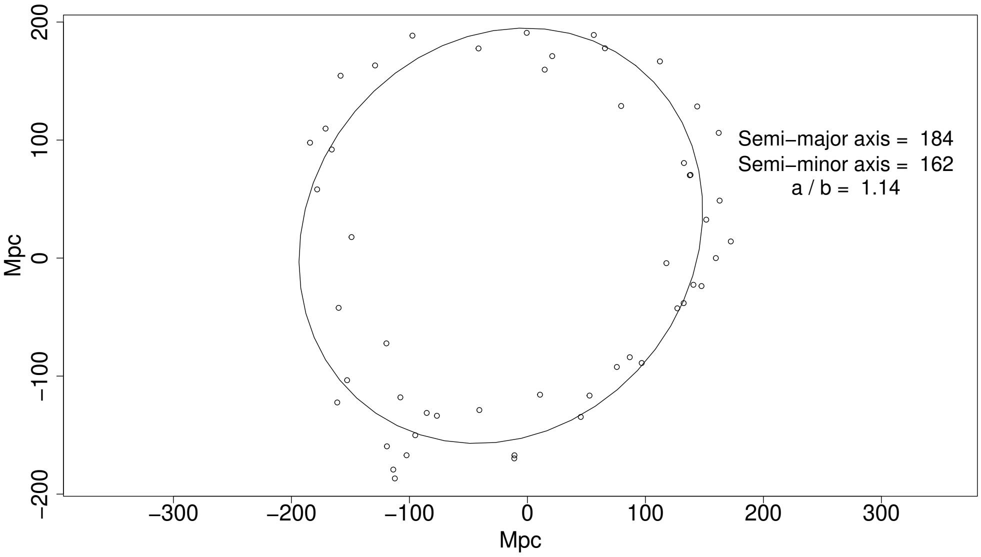


Figure 24. The visually-identified BR absorber members projected onto the plane perpendicular to w_0 with an added fitted ellipse.

central component is particularly interesting given the recent work of finding flat patterns in cosmic structure presented in [17]. It is not yet clear how a structure like this will have formed; however, one possibility could be cosmic strings, which have been suggested to explain other recent discoveries and data [16, 17, 41–46].

Previously, the diameter of the BR was noted as similar to the expected size of a BAO. However, the coil shape seems likely to be inconsistent with an origin in BAOs.

Finally, an ellipse can be fitted to the BR (figure 24). The ellipse calculates the semi-major and semi-minor axes as 184 Mpc and 162 Mpc, respectively. Given that the signature of an individual BAO has a characteristic size of 150 Mpc, we again argue against the possibility of the BR occurring from a BAO. Given the ellipse calculations of the semi-major and semi-minor axes we can use Ramanujan's approximative perimeter to approximate the circumference of the ellipse,

$$p_R = \pi \left\{ (a+b) + \frac{3(a-b)^2}{10(a+b) + \sqrt{a^2 + 14ab + b^2}} \right\}$$

where a and b are the semi-major and semi-minor axes. The circumference of the ellipse is then $\sim$ 1.1 Gpc.

4.3 Equivalent widths

In figure 20 we showed the visually-identified BR absorber members projected onto a plane normal to the vector w_0 (which corresponds to the LOS), with the points colour-coded according to redshift. Here, in figure 25, we show a corresponding figure but with the points now colour-coded according to the rest-frame equivalent widths (in Å) of the λ_{2796} component of the Mg II doublet. The equivalent widths and their uncertainties, both taken from the Mg II database of [3], are tabulated in table 5.

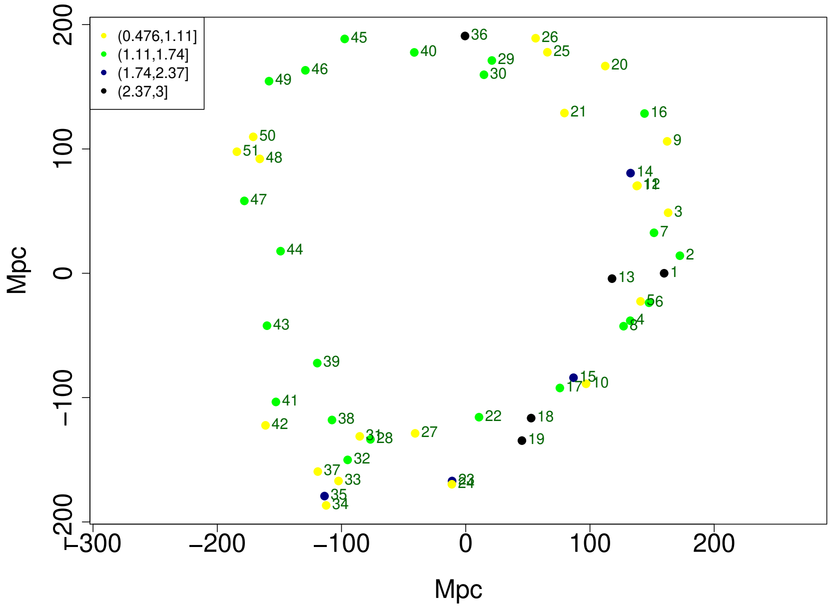


Figure 25. The visually-identified BR absorber members projected onto the plane perpendicular to the initial normal vector, w_0. The x-axis points towards the u_0 direction and the y-axis points towards the v_0 direction. The key in the top l.h.s. of the figure indicates the rest-frame equivalent widths (in Å) of the λ_{2796} component of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number. The equivalent widths are taken from the Mg II database of [3].

In figure 25 there is no obvious pattern for the lower equivalent widths, but we note that the higher equivalent widths tend to be concentrated on the r.h.s. of the plot, especially the points that are numbered $14, 1, 13, 15, 18, 19, 23$ (but not $36, 35$).

The earlier figure 23, for the plane perpendicular to $u_0 - v_0$, drew attention to ovoid shapes at high-z and low-z. Here, figure 26 corresponds to that figure, but again with colour-coding according to the equivalent widths. It shows that the points $14, 36, 1, 15, 18, 19$ (and possibly $23, 35$ but not 13) are actually concentrated in the low-z ovoid shape. Possibly a preferential alignment with respect to the LOS of the host galaxies in this ovoid shape is enhancing the column densities.

5 Discussion and conclusions

In this paper we have presented a 'Big Ring on the Sky', which is the second ultra-large large-scale structure (uLSS) detected in the Mg II catalogues The BR is detected most prominently in the redshift slice $z = 0.802 \pm 0.060$. The redshift slice containing the BR is the exact same redshift slice containing the previously detected GA, and both structures on the sky are separated by only $\sim 12°$, meaning that these two intriguing structures are in the same cosmological neighbourhood.

We applied several inspection tests and statistical tests to the BR to support it as an ultra-large LSS. A summary of the results is as follows. (1) Each of the visually-identified BR+filament absorbers was confirmed visually in the corresponding spectra, establishing that

ID	EW (Å)	σ_{EW} (Å)	**ID**	EW (Å)	σ_{EW} (Å)
1	2.99	0.17	**27**	0.96	0.12
2	1.46	0.20	**28**	1.15	0.27
3	0.94	0.12	**29**	1.63	0.19
4	1.71	0.43	**30**	1.35	0.30
5	1.05	0.23	**31**	0.76	0.09
6	1.29	0.26	**32**	1.22	0.30
7	1.13	0.28	**33**	0.93	0.12
8	1.31	0.22	**34**	0.99	0.18
9	0.51	0.13	**35**	2.07	0.43
10	0.72	0.15	**36**	2.49	0.35
11	1.08	0.06	**37**	0.51	0.05
12	0.82	0.05	**38**	1.33	0.16
13	2.70	0.46	**39**	1.32	0.29
14	2.29	0.32	**40**	1.20	0.22
15	1.75	0.25	**41**	1.73	0.33
16	1.45	0.32	**42**	0.76	0.05
17	1.55	0.36	**43**	1.28	0.25
18	2.62	0.08	**44**	1.50	0.17
19	2.58	0.21	**45**	1.35	0.12
20	1.07	0.05	**46**	1.55	0.25
21	0.86	0.05	**47**	1.50	0.36
22	1.66	0.12	**48**	0.91	0.21
23	2.08	0.20	**49**	1.56	0.29
24	1.05	0.25	**50**	0.48	0.11
25	0.79	0.09	**51**	1.08	0.11
26	0.58	0.09			

Table 5. The table is for the 51 visually-identified BR absorber members. The identification numbers (ID) correspond to those in figure 20 and the other similar figures that follow it. The rest-frame equivalent widths (EW) and their associated uncertainties (σ_{EW}, taken from the Mg II database of [3], are for the λ_{2796} component of the Mg II doublet; the units are Å. Note that the EW for point 12 is likely to be enhanced by the superimposed λ_{2803} component from point 11.

100% of the absorbers are real detections, and not false-positives. (2) The SLHC algorithm identified 46 out of 59 of the visually-identified BR+filament absorbers across 5 individual, overlapping or adjacent, groups. The SLHC algorithm also detected a statistically-significant arc in the redshift slices centred at $z = 0.862 \pm 0.060$ and $z = 0.922 \pm 0.060$. The arc appears to be an extension of the bottom arc of the BR, given the agreement in the on-sky position and that the redshift slices overlap by 50%. (3) We calculated the CHMS and MST significance of the SLHC-identified absorbers, the visually-identified BR absorbers (two versions), and the FilFinder-identified absorbers. The CHMS method was found to have a much larger variation in its reported significances due to the nature of the method: the CHMS calculates the volume

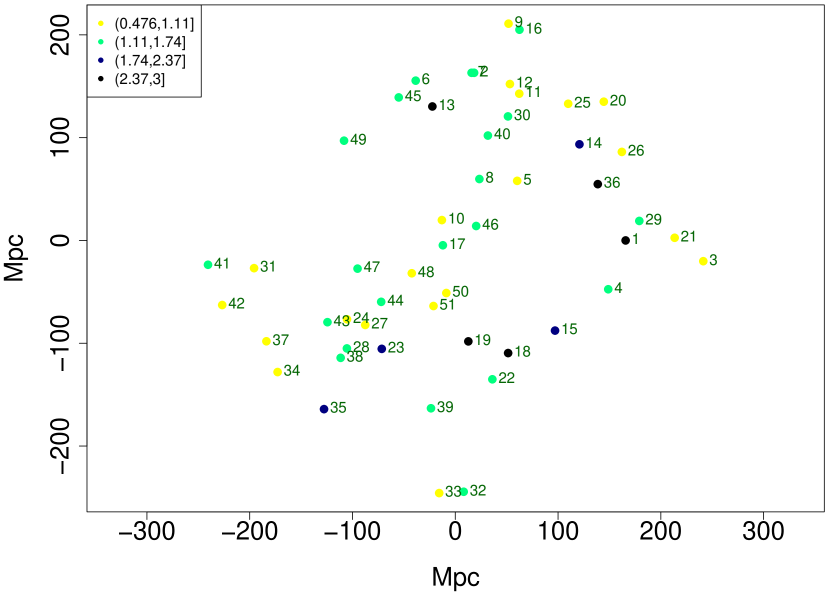


Figure 26. The BR absorber members projected onto the plane perpendicular to $u_0 - v_0$. The new u direction points towards the most easterly absorber, as usual. The key in the top l.h.s. of the figure indicates the rest-frame equivalent widths (in Å) of the λ_{2796} component of the absorbers associated by the colours, and the small numbers paired with each data point indicate their unique ID number. The equivalent widths are taken from the Mg II database of [3].

of the unique structure containing all the absorbers, so in the example of a ring-like structure, there is a large volume in the centre of the ring with far fewer absorbers than the annulus. The CHMS calculates a $(3.65 \pm 1.13)\sigma$ significance. By using the visually-identified absorbers, and keeping the BR innards, the CHMS calculates a (likely) upper limit of 5.2σ significance. The MST significance on the other hand was more consistent, as this method relies only on the mean MST edge-length, and this method calculates a significance of $(4.10 \pm 0.45)\sigma$. (4) The FilFinder algorithm was applied to the Mg II image of the BR field to identify filaments objectively. By incrementally increasing the size-threshold, the algorithm left only one identified filament — a ring — and that ring corresponded to the visually-identified BR. In this way, the BR was established as: (i) a ring, independently of visual perception; and (ii) the most-connected and largest filament in the image. We also applied the FilFinder algorithm to the field quasars (not to be confused with the background probes) and this found that there was a large, connected, ring-like filament that mostly aligns with the BR. (5) We applied the CE test to the field containing the BR and then 'zoomed' in on the BR to assess the contribution of the BR to the spatial clustering in the field. Tentative significant clustering in the field was detected in the second zoom of the BR, at a p-value $p = 0.022$ (corresponding to a significance of 2.0σ). We compared this with four other unrelated fields at the same redshift as the BR and found no significant spatial clustering. However, the CE test results for the BR field are still inconclusive at the $\sigma > 3.0$ significance level, suggesting that the clustering in the BR field is not statistically significant, unlike what was seen with the GA field. Note, this test assesses the clustering in the *field* and not the statistical significance of any individual candidate structure. (6) By superimposing contour maps of the field quasars and DESI clusters on the Mg II absorber images we were able to show plausible association between the datasets, thus providing independent corroboration. We found in particular that the DESI clusters that were subject to richness limits were mapping different Mg II absorbers, suggesting that there could be scope for using DESI clusters as a way to investigate the physical environment around the Mg II absorbers. (7) Using a project-plane method we could investigate the 3D distribution of the BR. Viewing the BR on different angles showed that there are 3 distinct redshift bands, where the central redshift band contains the majority of the absorbers in a ring shape in a thin, flat region. The near-redshift band is contained almost entirely on the r.h.s. of the BR and even appears to angle backwards into the central redshift bin of the BR creating a spiral-like shape. The farthest-redshift band had a similar coiling effect, but appeared to be distributed randomly throughout the BR (it made a broken, ring shape). We also found, with the project-plane method, curious backwards 'S' shapes when viewing the BR side on. The S shapes are most apparent in the near and far redshift bands but can also be seen in the central redshift band.

The data and analysis show that the BR is of particular interest for LSS studies in cosmology. In cosmology, we assume statistical homogeneity in the Cosmological Principle (CP) as the foundations of the ΛCDM model. We have shown that the BR is real and statistically significant, adding it to a growing list of LSS candidates that are in tension with the CP. The growing list of LSSs also indicates that the chance of all of these structures being a matter of statistical noise is decreasing. In addition, the list of LSSs in tension with the CP is also generally difficult to understand as there are gaps in the knowledge for explaining the formation of these ultra-large structures. Perhaps we could speculate that the existence of these structures requires an extension to the standard model, perhaps in the form of cosmic strings.

The BR and GA together, given their sizes and morphologies, are presumably telling us something intriguing, and quite possibly important, about the Universe, but at the moment we can only speculate what that might be.

Acknowledgments

We acknowledge the use of the public R software (v4.1.2).[4] Our data has depended on the publicly-available Sloan Digital Sky Survey quasar catalogue and the corresponding Mg II catalogues.[5] AML was supported by a UCLan/JHI PhD studentship.

We thank the referee for thoughtful comments and useful suggestions.

---

[4] https://www.R-project.org/.

[5] https://wwwmpa.mpa-garching.mpg.de/SDSS/MgII/.

References

[1] A.M. Lopez, *Assessing the potential of intervening Mg* II *absorbers for cosmology*, M.Sc. Thesis, University of Central Lancashire (2019).

[2] A.M. Lopez, R.G. Clowes and G.M. Williger, *A Giant Arc on the Sky*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 1557 [`arXiv:2201.06875`] [INSPIRE].

[3] A. Anand, D. Nelson and G. Kauffmann, *Characterizing the abundance, properties, and kinematics of the cool circumgalactic medium of galaxies in absorption with SDSS DR16*, *Mon. Not. Roy. Astron. Soc.* **504** (2021) 65 [`arXiv:2103.15842`] [INSPIRE].

[4] G. Zhu and B. Ménard, *The JHU-SDSS metal absorption line catalog: redshift evolution and properties of Mg II absorbers*, *Astrophys. J.* **770** (2013) 130 [`arXiv:1211.6215`] [INSPIRE].

[5] J. Bergeron, *Galaxies Giving Rise to MgII Absorption Systems in Quasar Spectra*, in the proceedings of the *130th Symposium of the IAU on Large scale structures of the Universe*, Balatonfured, Hungary, 15–20 June 1987, `DOI:10.1017/s007418090013623x`.

[6] C.W. Churchill, G.G. Kacprzak and C.C. Steidel, *MgII absorption through intermediate redshift galaxies*, in the proceedings of the *IAU Colloquium 199: Probing Galaxies throught Quasar Absporption Lines*, Shanghai, People's Republic of China, 14–18 March 2005 [`DOI:10.1017/S1743921305002401`] [`astro-ph/0504392`] [INSPIRE].

[7] C.C. Steidel, *The nature and evolution of absorption: Selected galaxies*, in the proceedings of the *ESO Workshop on QSO absorption lines*, Garching, Germany, November 1994 [`astro-ph/9509098`] [INSPIRE].

[8] J. Bagchi et al., *Saraswati: An Extremely Massive* ~ 200 *Megaparsec Scale Supercluster*, *Astrophys. J.* **844** (2017) 25 [`arXiv:1707.03082`] [INSPIRE].

[9] L.G. Balázs et al., *A giant ring-like structure at* $0.78 < z < 0.86$ *displayed by GRBs*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 2236 [`arXiv:1507.00675`] [INSPIRE].

[10] I. Horváth, J. Hakkila and Z. Bagoly, *Possible structure in the GRB sky distribution at redshift two*, *Astron. Astrophys.* **561** (2014) L12 [`arXiv:1401.0533`] [INSPIRE].

[11] R.G. Clowes et al., *A structure in the early universe at* $z \sim 1.3$ *that exceeds the homogeneity scale of the R-W concordance cosmology*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 2910 [`arXiv:1211.6256`] [INSPIRE].

[12] R.G. Clowes, L.E. Campusano, M.J. Graham and I.K. Soechting, *Two close Large Quasar Groups of size* ~ 350 *Mpc at* $z \sim 1.2$, *Mon. Not. Roy. Astron. Soc.* **419** (2012) 556 [`arXiv:1108.6221`] [INSPIRE].

[13] P.K. Aluri et al., *Is the observable Universe consistent with the cosmological principle?*, *Class. Quant. Grav.* **40** (2023) 094001 [`arXiv:2207.05765`] [INSPIRE].

[14] D. Huterer, *Growth of cosmic structure*, *Astron. Astrophys. Rev.* **31** (2023) 2 [`arXiv:2212.05003`] [INSPIRE].

[15] P. Bari et al., *Signatures of Primordial Gravitational Waves on the Large-Scale Structure of the Universe*, *Phys. Rev. Lett.* **129** (2022) 091301 [`arXiv:2111.06884`] [INSPIRE].

[16] W. Ahmed, T.A. Chowdhury, S. Nasri and S. Saad, *Gravitational waves from metastable cosmic strings in the Pati-Salam model in light of new pulsar timing array data*, *Phys. Rev. D* **109** (2024) 015008 [`arXiv:2308.13248`] [INSPIRE].

[17] P.J.E. Peebles, *Flat patterns in cosmic structure*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 4490 [`arXiv:2308.04245`] [INSPIRE].

[18] B. Shlaer, A. Vilenkin and A. Loeb, *Early structure formation from cosmic string loops*, *JCAP* **05** (2012) 026 [arXiv:1202.1346] [INSPIRE].

[19] R. Penrose, *On the Gravitization of Quantum Mechanics 2: Conformal Cyclic Cosmology*, *Found. Phys.* **44** (2014) 873 [INSPIRE].

[20] S. Mazurenko, I. Banik, P. Kroupa and M. Haslbauer, *A simultaneous solution to the Hubble tension and observed bulk flow within 250 h−1 Mpc*, *Mon. Not. Roy. Astron. Soc.* **527** (2024) 4388 [arXiv:2311.17988] [INSPIRE].

[21] C. Llinares, A. Knebe and H.S. Zhao, *Cosmological Structure Formation under MOND: a new numerical solver for Poisson's equation*, *Mon. Not. Roy. Astron. Soc.* **391** (2008) 1778 [arXiv:0809.2899] [INSPIRE].

[22] M. Milgrom, *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis*, *Astrophys. J.* **270** (1983) 365 [INSPIRE].

[23] eBOSS collaboration, *The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release*, *Astrophys. J. Suppl.* **250** (2020) 8 [arXiv:2007.09001] [INSPIRE].

[24] SDSS collaboration, *The Sloan Digital Sky Survey Quasar Catalog V. Seventh Data Release*, *Astron. J.* **139** (2010) 2360 [arXiv:1004.1167] [INSPIRE].

[25] I. Pâris et al., *The Sloan Digital Sky Survey Quasar Catalog: twelfth data release*, *Astron. Astrophys.* **597** (2017) A79 [arXiv:1608.06483] [INSPIRE].

[26] R.B. Tully, C. Howlett and D. Pomarède, *Ho'oleilana: An Individual Baryon Acoustic Oscillation?*, *Astrophys. J.* **954** (2023) 169 [arXiv:2309.00677] [INSPIRE].

[27] M. Einasto et al., *Shell-like structures in our cosmic neighbourhood*, *Astron. Astrophys.* **587** (2016) A116 [arXiv:1506.05295] [INSPIRE].

[28] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [arXiv:1502.01589] [INSPIRE].

[29] BOSS collaboration, *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 24 [arXiv:1312.4877] [INSPIRE].

[30] SDSS collaboration, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560 [astro-ph/0501171] [INSPIRE].

[31] L. Napolitano et al., *Detecting and Characterizing Mg ii Absorption in DESI Survey Validation Quasar Spectra*, *Astron. J.* **166** (2023) 99 [arXiv:2305.20016] [INSPIRE].

[32] J.K. Yadav, J.S. Bagla and N. Khandai, *Fractal Dimension as a measure of the scale of Homogeneity*, *Mon. Not. Roy. Astron. Soc.* **405** (2010) 2009 [arXiv:1001.0617] [INSPIRE].

[33] M.B. Taylor, *TOPCAT & STIL: Starlink Table/VOTable Processing Software* in the proceedings of the *Astronomical Society of the Pacific Conference Series on Astronomical Data Analysis Software and Systems XIV*, Pasadena, California, U.S.A., 24–27 October 2004.

[34] D. Pomarède et al., *Cosmicflows-3: The South Pole Wall*, *Astrophys. J.* **897** (2020) 133 [arXiv:2007.04414] [INSPIRE].

[35] I.I.I.J.R. Gott et al., *A map of the universe*, *Astrophys. J.* **624** (2005) 463 [astro-ph/0310571] [INSPIRE].

[36] S.V. Pilipenko, *The space distribution of quasars*, *Astron. Rep.* **51** (2007) 820.

[37] E.W. Koch and E.W. Rosolowsky, *Filament Identification through Mathematical Morphology*, arXiv:1507.02289.

[38] B. Mookerjea et al., *Spiral Structure and Massive Star formation in the Hub-Filament-System G326.27-0.49*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 2517 [arXiv:2301.09775].

[39] S. Zhang et al., *ATOMS: ALMA Three-millimeter Observations of Massive Star-forming regions — XIII. Ongoing triggered star formation within clump-fed scenario found in the massive* ($\sim 1500\,M_\odot$) *clump*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 322 [arXiv:2301.01937].

[40] S.E. Meidt et al., *PHANGS-JWST First Results: ISM structure on the turbulent Jeans scale in four disk galaxies observed by JWST and ALMA*, *Astrophys. J. Lett.* **944** (2023) L18 [arXiv:2212.06434].

[41] B. Cyr, J. Chluba and S.K. Acharya, *Cosmic string solution to the radio synchrotron background*, *Phys. Rev. D* **109** (2024) L121301 [arXiv:2308.03512] [INSPIRE].

[42] J. Ellis, M. Lewicki, C. Lin and V. Vaskonen, *Cosmic superstrings revisited in light of NANOGrav 15-year data*, *Phys. Rev. D* **108** (2023) 103511 [arXiv:2306.17147] [INSPIRE].

[43] Y. Gouttenoire and E. Vitagliano, *Domain wall interpretation of the PTA signal confronting black hole overproduction*, arXiv:2306.17841 [INSPIRE].

[44] H. Jiao, R. Brandenberger and A. Refregier, *Early structure formation from cosmic string loops in light of early JWST observations*, *Phys. Rev. D* **108** (2023) 043510 [arXiv:2304.06429] [INSPIRE].

[45] S. Sanyal, *Nambu Goto cosmic strings in the early universe*, *Eur. Phys. J. ST* **231** (2022) 83 [INSPIRE].

[46] Z. Wang et al., *The nanohertz stochastic gravitational wave background from cosmic string loops and the abundant high redshift massive galaxies*, *Sci. China Phys. Mech. Astron.* **66** (2023) 120403 [arXiv:2306.17150] [INSPIRE].

[47] J. Cuzick and R. Edwards, *Spatial Clustering for Inhomogeneous Populations*, *J. Roy. Statist. Soc. B* **52** (1990) 73.

[48] J. French, *Statistical methods for the analysis of case-control point data*, CRAN, https://cran.r-project.org/web/packages/smacpod/ (2023).

[49] J. Gao et al., *Galaxy Clusters from the DESI Legacy Imaging Surveys. I. Cluster Detection*, *Astrophys. J. Suppl.* **253** (2021) 56 [arXiv:2101.12340] [INSPIRE].

PHILOSOPHICAL
TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Opinion piece

Cite this article: Lopez AM, Clowes R, Williger G. 2025 Investigating ultra-large large-scale structures: potential implications for cosmology. *Phil. Trans. R. Soc. A* **383**: 20240029.
https://doi.org/10.1098/rsta.2024.0029

Received: 28 May 2024
Accepted: 21 September 2024

One contribution of 12 to a discussion meeting issue 'Challenging the standard cosmological model'.

Subject Areas:
cosmology, astrophysics, galaxies, observational astronomy, quasars

Keywords:
large-scale structure, cosmological principle, intervening absorption

Author for correspondence:
Alexia M. Lopez
e-mail: missalexia.lopez@gmail.com

THE ROYAL SOCIETY
PUBLISHING

Investigating ultra-large large-scale structures: potential implications for cosmology

Alexia M. Lopez[1], Roger Clowes[1] and Gerard Williger[2]

[1]Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK
[2]Department of Physics and Astronomy, University of Louisville, Louisville, KY 40292, USA

AML, 0000-0003-4634-3124

Large-scale structure (LSS) studies in cosmology map and analyse matter in the Universe on the largest scales. Understanding LSSs can provide observational support for the cosmological principle (CP) and the standard cosmological model (ΛCDM). In recent years, many discoveries have been made of LSSs that are so large that they become difficult to understand within ΛCDM. Reasons for this are that they potentially challenge the CP (i.e. the scale of homogeneity) and their formation and origin are not fully understood. In this article, we review two recent LSS discoveries: the Giant Arc (GA, approx. 1 Gpc) and the Big Ring (BR, approx. 400 Mpc). Both structures are in the same cosmological neighbourhood—at the same redshift ($z \sim 0.8$) and with a separation on the sky of approximately only 12°. Both structures exceed the often-cited scale of homogeneity (Yadav+ 2010), so individually and together, these two intriguing structures raise more questions for the validity of the CP and potentially hint at new physics beyond the standard model. The GA and BR were discovered using a novel method of mapping faint matter at intermediate redshifts, interpreted from the Mg II absorption doublets seen in the spectra of background quasars.

This article is part of the discussion meeting issue 'Challenging the standard cosmological model'.

1. Introduction

Large-scale structure (LSS) studies are motivated by the need for observational data to confirm the predictions of the standard cosmological model (ΛCDM). In particular, from studying the LSS of matter on the very largest scales, one can learn about the growth of cosmic structure [1] and about the dynamical history of the Universe [2], thus allowing comparison with Λ CDM. Furthermore, and of some current interest, LSS can test the assumption of large-scale homogeneity, which is a fundamental aspect of the cosmological principle (CP) and hence of the theoretical framework in cosmology.

Unfortunately, the CP lacks a precise and agreed definition. Different interpretations can be encountered in the cosmological literature and across the history of cosmology. The details can be vague. For example, the textbook version of the CP might say that the Universe on large scales is homogeneous and isotropic. However, what those large scales might be is often not clearly specified, and indeed, the expectation of what is plausible seems to have increased by at least a factor of 10 over the years.

Consider the following three interpretations of the CP, and specifically what is meant by homogeneity: (i) There exists some large scale, known as the scale of homogeneity, on which the Universe can be smoothed and the distribution of matter would then be well represented by a stationary random process (e.g. [3–5]). (ii) The power spectrum suggests that there can always be some large scale at which statistically significant deviations might be found in the matter distribution, but such deviations on large scales should be rare (e.g. [2,6]). (iii) There should be similarity everywhere (maximally symmetric). Any observed large-scale structuring indicates that the scale of homogeneity, if it exists, must then be larger than these scales. The occurrence of a particular LSS, even the largest known, does not imply that the *probability* of finding a comparable LSS elsewhere is any different. Points (ii) and (iii) could be contradictory, given that point (ii) suggests that the largest structures should be rare, and point (iii) suggests that the largest structures need not be a problem for the CP if their probabilities are homogeneous and isotropic. (How would we know that?) A useful overview of the various interpretations of the CP can be found in [7].

The diversity in the interpretation of homogeneity in the CP has led to differing conclusions on whether the observed matter supports a homogeneous Universe. For example, there have been claims that large-survey analysis supports homogeneity in luminous red galaxies and in quasars [8–10]. However, the accumulating set of ultra-large LSS (uLSS)[1] discoveries might indicate that homogeneity is not supported. For a recent list of the largest LSSs that appear to extend beyond Yadav's [4] estimated approximately 370 Mpc upper limit to the scale of homogeneity, see table 1 in [11].

Perhaps individual uLSS discoveries can instead be explained by appealing to extreme-value statistics in some form. For example, Park *et al*. [6] and Nadathur [12] address the Sloan Great Wall (SGW) and the Huge Large Quasar Group (Huge-LQG) with mock catalogues and random catalogues, respectively, finding that mock/random structures of comparable size and overdensity may be readily reproduced. Somewhat differently for the Huge-LQG, Marinello *et al*. [13] used the Horizon Run 2 cosmological simulation and extreme-value analysis to show that the Huge-LQG is compatible with the ΛCDM model if it should happen to be the largest such structure in a volume over five times larger. Similarly, for the SGW, Park *et al*. [6] found that, while structures comparable with the size and overdensity of the SGW were reproduced in their simulations, they were always in the top six largest and richest structures detected in the 200 mock samples.

---

[1] We have introduced the new term 'ultra-large LSS' (uLSS) to denote those structures that exceed the Yadav estimated ~ 370 Mpc upper limit to the scale of homogeneity [4].

Table 1. The SLHC and MST significance results applied to the four versions of the BR from the SLHC-identified absorbers, visually identified (BR-all and BR-only) absorbers and FilFinder-identified absorbers. These results are based on data from the Anand21 database.

	SLHC (σ)	visual BR-all (σ)	visual BR-only (σ)	FilFinder (σ)	mean (σ)
CHMS	3.6	5.2	3.3	2.5	3.7 ± 1.1
MST	4.7	4.1	4.0	3.6	4.1 ± 0.5

BR, Big Ring; MST, Minimal Spanning Tree; SLHC, Single-Linkage Hierarchical Clustering.

The results from [6] and [13] suggest that, in the matter of the compatibility of uLSSs with the ΛCDM model, the *accumulated* set of uLSSs might be of more importance than any one individual structure.

In this review article, we discuss two intriguing uLSS discoveries, the Giant Arc (GA) [11] and the Big Ring (BR) [14]. See figure 1 for an artistic impression of both these structures in the sky. We summarize their method of discovery and statistical analysis, describe their observed properties and comment or speculate on their possible origins. The GA and the BR are in the same cosmological neighbourhood, at $z \sim 0.8$ and separated by approximately only 12° on the sky. Individually and together, they exceed the Yadav 370 Mpc estimated scale of homogeneity, and thus may at least challenge some interpretations of the CP. (Note that the word ‘challenge’ is not synonymous with ‘contradict’, but it does imply something to be investigated further.) Their sizes and morphologies appear to be hard to explain in the ΛCDM model. Quite possibly, new developments in cosmology will follow from the continued investigation of intriguing and unexpected anomalies such as these.

2. The Mg II method

Observations for studying LSS in the Universe can be broadly divided into two categories: (i) observations of low-luminosity objects at low to intermediate redshifts (e.g. [15–17]) and (ii) observations of high-luminosity objects at intermediate to high redshifts (e.g. [18–25]). There are benefits and challenges to both categories. In category (i), photometric redshifts, for very large numbers of objects, are commonplace, but the large redshift errors can lead to blurring of structures along the line of sight. Spectroscopic redshifts, generally for smaller numbers of objects, are more demanding of telescope time, and are more likely to feature in category (ii). The redshift errors are then smaller but may still be associated with some blurring.

A novel method for analysing LSS is to infer the low-luminosity matter at intermediate redshifts from the presence of sharp metal absorption lines in the spectra of high-redshift quasars. The Mg II doublet, specifically, arises from low-ionized metal-enriched gas which is well-known to trace star formation regions [26–30]. From many quasar observations covering a large area of sky, such as the Sloan Digital Sky Survey (SDSS) footprint or the recent Dark Energy Spectroscopic Instrument (DESI) survey footprint (at the time of writing, this is not yet publicly available), one can then map the inferred, low-luminosity, intervening matter at intermediate redshifts and learn about the LSS.

The sources of data for the Mg II method are as follows. The first-order data are the spectroscopic quasar observations from the SDSS quasar catalogues. For previous work (including the discovery and analysis of the GA), we used the ‘cleaned’ quasar catalogues DR7QSO [31] and DR12Q [32]. For recent work (including the discovery and analysis of the BR), we used the newer ‘cleaned’ DR16Q quasar catalogue [33]. The second-order data are the corresponding Mg II catalogues from independent authors. For the Mg II absorber catalogues corresponding to DR7QSO and DR12Q, we downloaded the Zhu and Ménard (Z&M) [34] data,

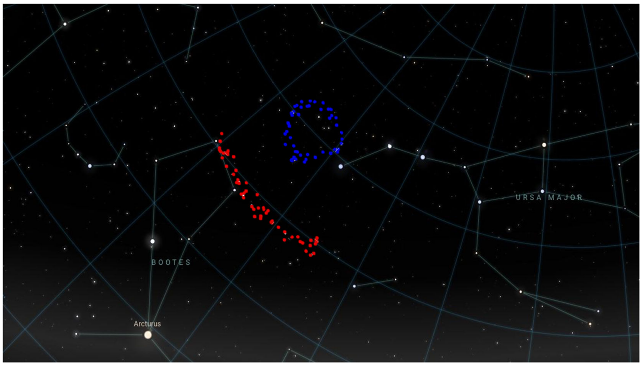


Figure 1. An approximate projection of the GA and BR visually identified absorber members superimposed on to an image of the night sky taken from Stellarium. This figure gives an impression of the scale of the two uLSSs, their position in the sky and their proximity.

and for the absorber catalogues corresponding to DR16Q, we downloaded the Anand *et al*. (Anand21) [35] data.

3. Statistical analysis

In this section, we summarize the statistical analyses performed on the GA and BR to assess them, subdivided by the different statistical tests used. Full details can be found in the respective papers.

(a) Power spectrum analysis

Two-dimensional spectrum analysis (2D PSA) [36,37] is a powerful statistical tool for detecting clustering of sources in a rectangular field using Fourier methods; it has the power to detect the scale of clustering as well as its statistical significance on that scale. We applied this test to the field containing the GA. Given that the (probable) dominant feature of clustering in the GA field was the GA itself, we reduced the typical Mg II image size, mostly along the north-south axis (figure 2), before applying 2D PSA.

In figure 3, we show the 2D PSA statistic Q' corresponding to the GA field seen in figure 2. The (six) high points towards the left of the figure allow a clustering scale of $\lambda_c \sim 270$ Mpc to be identified. The final PSA statistic Q for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ. Given the clustering scale of $\lambda_c \sim 270$ Mpc, this is probably detecting the width of the GA as perceived along its length.

(b) Cuzick and Edwards test

The Cuzick and Edwards (CE) test [38] is a case-control k (or q here) nearest-neighbours algorithm, originally intended to assess geographical, spatial clustering of medical illnesses in inhomogeneous populations. Given the complications of the data with which we work—the possible inhomogeneity of the quasars (the probes) available for detecting intervening Mg II absorbers—we applied this test to both the GA and the BR.

CE found the test to be most powerful when the ratio of controls to cases is between four and six. However, for the GA field, there were approximately 20 times as many probes (controls) as Mg II absorbers (cases), and for the BR field, there were approximately 50 times as many probes as Mg II absorbers. For the CE test calculation, in both cases of the GA and BR, a subset of probes was randomly selected to reduce the ratio of controls to cases to 5:1, and this procedure

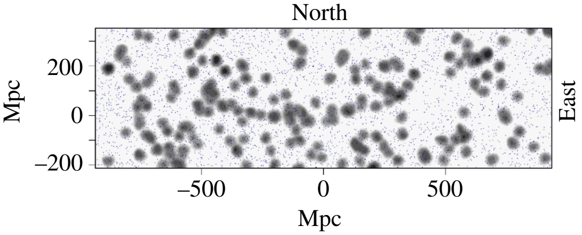


Figure 2. The tangent-plane distribution of Mg II absorbers centred on the GA in the redshift slice $z = 0.802 \pm 0.060$ using the Z&M data. The grey contours, increasing by a factor of 2, represent the density distribution of the absorbers which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc, and flat-fielded with respect to the distribution of background probes (quasars). The dark blue dots represent the background probes. This figure corresponds to fig. 9*a* of the GA paper [11].

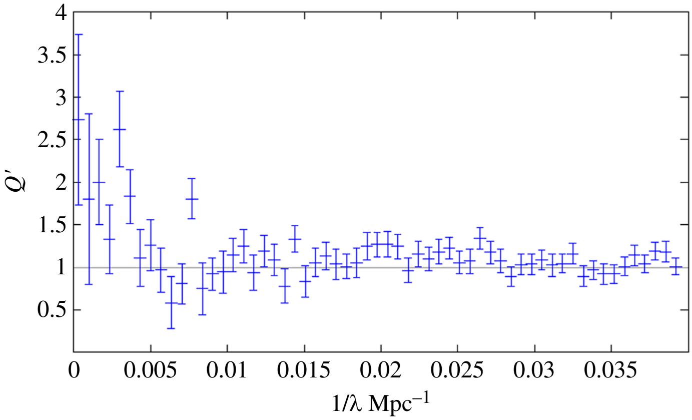


Figure 3. The 2D PSA test statistic Q' as a function of clustering scale $1/\lambda$ with λ in Mpc. The bin size is 6.7×10^{-4} Mpc−1 and the error bars are $\pm\sigma$. The horizontal line $Q' = 1$ indicates the expectation value in the case of no clustering. The (six) high points towards the left of the plot allow a clustering scale of $\lambda_c \sim 270$ Mpc to be identified. The final PSA statistic Q for this scale λ_c corresponds to a detection of clustering at a significance of 4.8σ. This figure corresponds to fig. 13 in the GA paper [11].

was then repeated 100 times. Within each run of randomly selected probes, the CE test was run with 2000 simulations.

For the GA field, seen in figure 2, the CE test found significant $p = 0.0027$ field clustering at $q = 40$ (figure 4*a*), corresponding to a significance of 3.0σ. For the (reduced) BR field seen in figure 5, the CE test did not find conclusive significant clustering. In figure 4*b*, the *p*-value drops below 0.05 (indicated by the blue, horizontal line), reaching a minimum of $p = 0.043$ (1.7σ) at $q = 58$. Note, updated statistical analysis using the CE test on the BR can be found in the BR paper. However, the differences are minor and of no consequence.

(c) FilFinder algorithm

The FilFinder algorithm [39] is a 2D filament identification tool, originally designed to trace filamentary structures in small gaseous regions, such as the ISM and star-formation regions [40–42]. We applied it to the field containing the BR to objectively trace the longest and most connected filaments in the field. Figure 6*a* shows the FilFinder algorithm applied to the BR field (for details of the parameter choices, see the original discovery paper). By incrementally increasing the size threshold of the algorithm, it removes the smaller and less connected filaments (figure 6*b*). In doing so, only the filament tracing the BR remains. Applying the FilFinder algorithm to the BR field shows that the BR is the most connected and largest filament in the field.

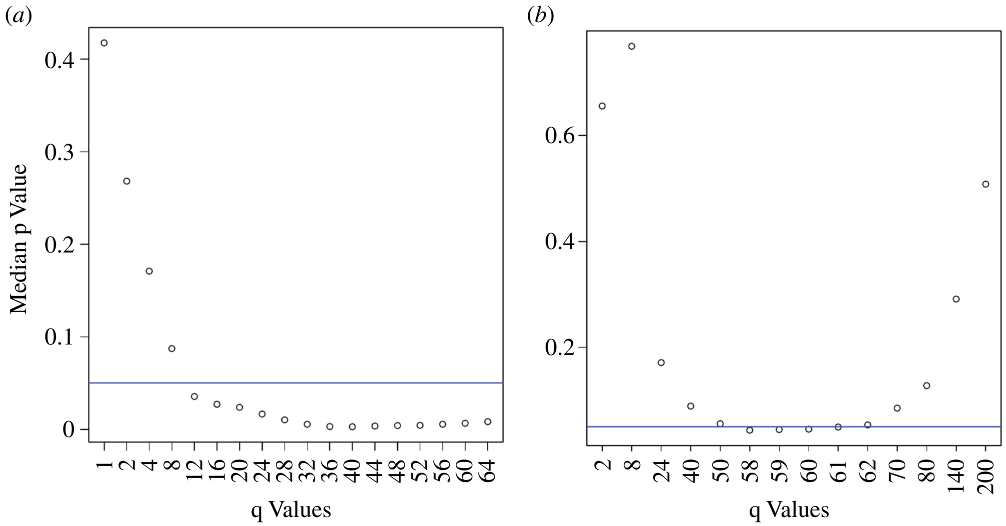


Figure 4. (*a*) The median p-value over 100 runs of 2000 simulations as a function of chosen q value. These results correspond to the GA field seen in figure 2. The p-value reaches a minimum of $p = 0.0027$ (corresponding to a 3σ detection) at $q = 40$. These results are based on data from the older Z&M database. The x-axis labels are 1, 2, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64. (*b*) The median p-value over 100 runs of 2000 simulations as a function of chosen q value. These results correspond to the BR field seen in figure 5. The p-value reaches a minimum of $p = 0.043$ at $q = 58$. These results are based on data from the Anand21 database. The x-axis labels are 2, 8, 24, 40, 50, 58, 60, 61, 62, 70, 80, 140, 200.

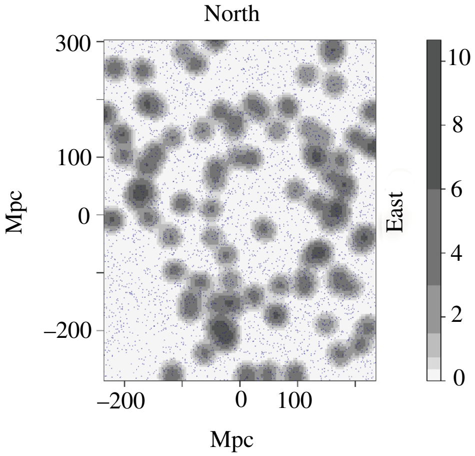


Figure 5. The tangent-plane distribution of Mg II absorbers centred on the BR in a reduced field size in the redshift slice $z = 0.802 \pm 0.060$ using the Anand21 data. The grey contours represent the density distribution of the absorbers, which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flat-fielded with respect to the distribution of background probes (quasars). The key on the right-hand side shows the relative density of the Mg II absorbers, which are increasing by a factor of 2. The dark blue dots represent the background probes.

(d) Single-linkage hierarchical clustering algorithm

The single-linkage hierarchical clustering (SLHC) algorithm is equivalent to a minimal spanning tree (MST) when the MST is separated at a specified linkage scale. When applied to three-dimensional (3D) spatial data, the SLHC algorithm determines the candidate

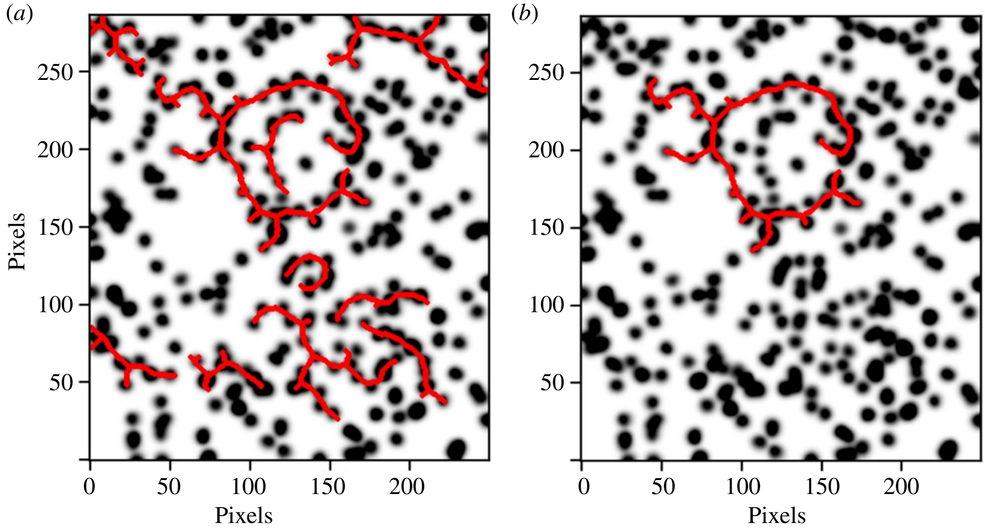


Figure 6. (*a*) Filaments, shown in red, identified by the FilFinder algorithm in the BR field. (*b*) Starting from (*a*), the FilFinder size threshold parameter was incrementally increased, leaving only the filament that traces the absorbers belonging to the BR. The Mg II data are from the Anand21 database.

(algorithmic) 'structures' within the dataset. The linkage scale will determine the set of data points included in a candidate structure. Naturally, the choice of linkage scale must be related to the field density: in a high-density region, the chosen linkage scale should be smaller than in a low-density region. We adopt the GA field as a standard field (linkage scale s_0, density ρ_0) and use the relation $s = (\rho_0/\rho)^{1/3}s_0$ to obtain the linkage scale s for any other field of density ρ.

The SLHC algorithm identified the majority of the absorbers belonging to the visually identified GA (using the older Z&M database) in two agglomerations: a large, statistically significant portion and a smaller, not statistically significant portion, which were named GA-main and GA-sub, respectively. The two GA agglomerations visually overlapped on the sky, indicating that they could belong to the same candidate structure if given a more complete survey coverage. (Recall that the background quasars are responsible for the detection of intervening Mg II systems.) The statistical significance of GA-main was then calculated using the convex hull of member spheres (CHMS; see originally [43] and the corresponding GA or BR paper for more details). The CHMS statistical significance of GA-main was 4.53σ. The overdensity of GA-main was calculated both with the CHMS method and the Pilipenko MST-overdensity method [44]. The CHMS-overdensity and MST-overdensity for GA-main were $\delta\rho_{\mathrm{CHMS}}/\rho = 0.9 \pm 0.6$ and $\delta\rho_{MST}/\rho = 1.3 \pm 0.3$, respectively. Clearly, the larger error estimates from the CHMS overdensity indicate that those results should be taken with caution. A mass excess was estimated from the MST-overdensity by assuming that $\delta_n = \delta_m$, where δ_n is the MST-overdensity and δ_m is the mass overdensity. We took the critical density of the Universe to be 9.2×10^{-27} kg m^{-3}. The estimated mass excess for GA-main was $1.8 \times 10^{18} M_\odot$.

The SLHC algorithm applied to the BR field identified the majority of the absorbers belonging to the visually identified BR (using the Anand21 database) in five, adjacent or overlapping, candidate structures, which, individually, were not statistically significant (see figure 7, also see figure 8 for our selection of the visually identified BR). Similar to what was seen with the SLHC algorithms on the GA field, the individual candidate structures in the BR appear connected on the sky, indicating that they could belong to the same structure. In addition, the results from the FilFinder add further support to the connectivity of the BR. We then applied the CHMS and MST significance tests to four versions of the BR: those identified by the FilFinder algorithm (FilFinder-identified); those identified by the SLHC in five candidate structures (SLHC-identified); the visually identified absorbers making the circumference of the

BR (BR-only); the visually identified absorbers of the BR and everything contained within the BR (BR-all). Table 1 lists the results of the CHMS and MST significances applied to the four versions of the BR.

In table 1, looking across the four versions in the CHMS significance calculations, the likely upper-limit estimate is from the visually identified BR-all absorbers, and the lower-limit estimate is from the FilFinder-identified absorbers. This variation is expected, as the CHMS draws a unique volume around the absorber members, and empty volumes existing between absorbers will also be included. So, with large, empty volumes incorporated by the CHMS algorithm, the volume of a filamentary-type structure will be overestimated, and the significance of the unique CHMS volume will be correspondingly underestimated. For example, compare the filament-identified absorbers in figure 6*b* with the visually identified absorbers in figure 8. The filament-identified absorbers envelop a large, empty region in the middle, as well as creating additional empty regions from the additional spurs to the northwest, whereas the visually identified absorbers, including all of those absorbers contained within the visually identified BR, have fewer empty regions. The differences are then reflected in the CHMS significance estimates. Ideally, the CHMS algorithm is more appropriately used for clumpy, globular-type structures and is less applicable to filamentary-type structures.

By contrast, also in table 1, the MST-significance values are less varying than the CHMS-significance values, as was also the case for the GA. The CHMS algorithm relies on the unique volume encompassing the absorber members, whereas the MST calculation relies on the mean MST edge length between absorber members, which might be better suited for filamentary structures than the CHMS.

4. Observational properties

In this section, we give examples of and review some of the observational properties of the GA and the BR; further details can be found in the respective papers.

(a) Independent corroboration

We used the SDSS DR16Q quasars and the DESI cluster catalogue from [45] to provide independent corroboration of the Mg II structures.

We used coloured contours superimposed on the Mg II images to visually inspect the relationship between the independent datasets. In figure 9*a,b*, we see two examples of independent corroboration of the Mg II with the DR16Q field quasars and the DESI clusters, respectively. Visually, the coloured contours show a tendency to follow the grey contours, and since the quasars and DESI clusters are independent sources, they both provide independent corroboration of the GA and BR.

(b) The proximity of the BR to the GA

The BR discovery is particularly interesting as it was made directly after the discovery of the GA but using the newer Anand21 Mg II catalogues instead of the older Z&M catalogues. The BR was first apparent when looking at the GA field with the Anand21 database; just to the north of the GA was an interesting ring shape, which we then investigated further.

On the sky, the GA and the BR are separated by only approximately 12°, and they both lie in exactly the same redshift slice $z = 0.802 \pm 0.060$. Figure 1 shows an approximate projection of the BR and GA absorber members imprinted on to a night sky image taken from Stellarium.

The direct discovery of a second uLSS, the BR, in the same cosmological neighbourhood as the first discovery, the GA, leads to the question of whether we would expect to find many more interesting, huge structures with circular or other conic-section morphologies. Perhaps,

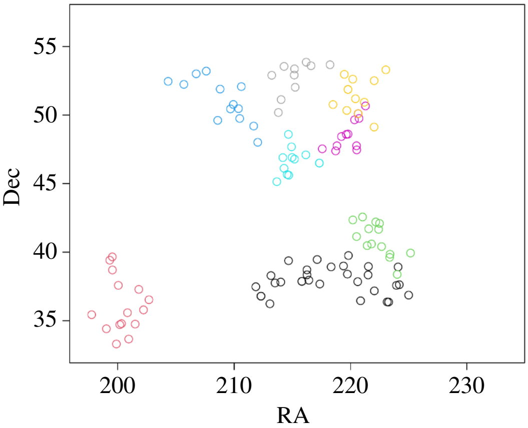


Figure 7. Eight of the ten highest membership candidate structures identified by the SLHC algorithm applied to the BR field in the redshift slice $z = 0.802 \pm 0.060$. The colours represent the memberships, which are ordered from high to low in the following way: black, red, green, blue, turquoise, pink, yellow and grey. In this figure, only the black points, representing absorbers belonging to the GA, are statistically significant. Data points are from the Anand21 database.

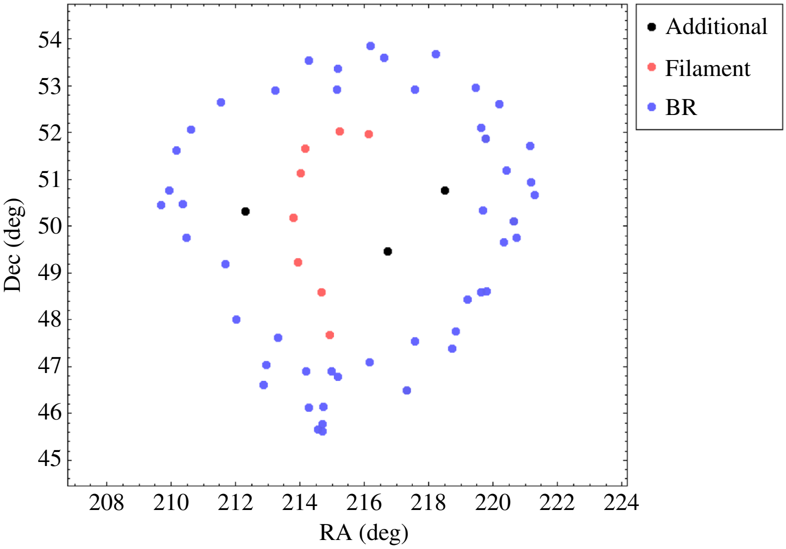


Figure 8. The visually identified BR is shown in blue points. The orange points mark an interesting filament running through the BR, and the black points are the additional absorbers encompassed by the BR. Data points are from the Anand21 database.

a more important question would be to determine *a priori*: (i) whether ΛCDM can explain the occurrence of such huge structures and (ii) whether, with the CP assumed, we should expect to find no other, or many more, of these types of structures.

(c) Hints of an extended GA

In the BR field of the Anand21 databases, there was also an interesting thin filament that appears to be a potential continuation of the GA. In figure 10, we have added the GA points from the Z&M catalogues on the same plot as the BR points from the Anand21 catalogues. In addition, we have added the points of the interesting thin filament that appeared in the

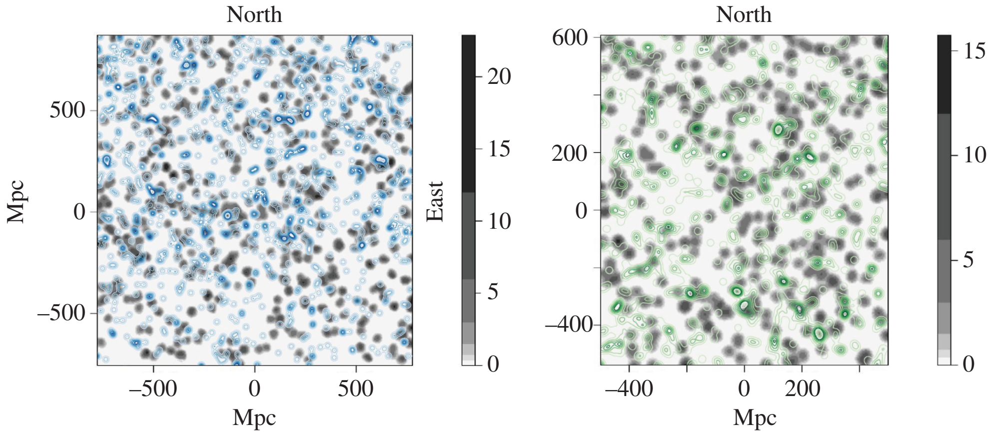


Figure 9. (*a*) Density distribution of the flat-fielded Mg II absorbers in the GA field in the redshift slice $z = 0.802 \pm 0.060$ represented by grey contours which have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flat-fielded with respect to the background probes (background quasars). The blue (line) contours represent the *field* quasars from DR16Q—i.e. the quasars that are in the same field and redshift slice as the Mg II absorbers —which have been restricted to $i \leq 20.0$ and have also been smoothed with a Gaussian kernel of $\sigma = 11$ Mpc. The GA can be seen stretching approximately 1 Gpc across the field at the tangent-plane y-coordinate ~ 0. This figure corresponds to figure 14*a* in the GA paper. (*b*) Density distribution of the flat-fielded Mg II absorbers in the BR field in the redshift slice $z = 0.802 \pm 0.060$ represented by grey contours that have been smoothed using a Gaussian kernel of $\sigma = 11$ Mpc and flat-fielded with respect to the background probes (background quasars). The green (line) contours represent the DESI clusters that are in the same field and redshift slice as the Mg II absorbers, which have been restricted to a richness limit $R \geq 22.5$, and have also been smoothed with a Gaussian kernel of $\sigma = 11$ Mpc. The BR can be seen centred approximately on tangent-plane x, y coordinates $\sim (50, 300)$. In both figures, the key on the right-hand side shows the relative density of the Mg II absorbers, which is increasing by a factor of 2.

Anand21 catalogs. With fitted ellipses, it can be seen that if (with more data) this interesting filament was a continuation of the GA, then the GA or Giant Ring (GR) would encompass the BR.

(d) The 3D distribution of the BR

For the BR, we define a 3D coordinate system so that the distribution of Mg II absorber points in BR can be projected on to different planes, allowing for the investigation of the BR at different viewing angles.

In figure 11*a*, we see the BR projected on to a plane, where the 3D coordinate system has been defined to resemble the original line-of-sight projection. The absorber points are numbered 1–51, and these values are fixed to their absorber points through all rotations of the plane. The colours represent the redshift distribution, with nearer-z absorbers being lighter in colour and farther-z absorbers being darker.

In figure 11*b*, we are next seeing the BR absorber points projected on to a plane that corresponds to viewing the BR from the southeast direction. From this viewing angle, the BR appears to resemble a coil shape. There also appear to be three distinct redshift bands, which were similarly seen from a side-on viewing angle (see the BR paper).

A 3D visualization tool was also used to move around the BR absorber points to investigate the 3D structure. Using this tool also supported the coil nature of the BR, and it gave an impression of the near-z absorbers 'looping' into the central-z absorbers. The central-z absorbers then contained the majority of the BR absorbers and created the main ring shape in a thin, flat region. The far-z absorbers showed less continuity, with the appearance resembling a broken ring shape.

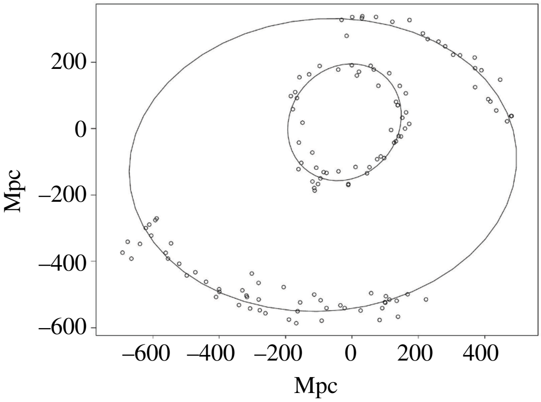


Figure 10. The GA points from Z&M (along the bottom) plotted with the BR points from Anand21 (in the middle). The thin filament to the north was also detected in the newer Anand21 catalogues. With the fitted ellipses, it appears that the interesting thin filament could be a continuation of the GA, and if so, then this GR would encompass the BR.

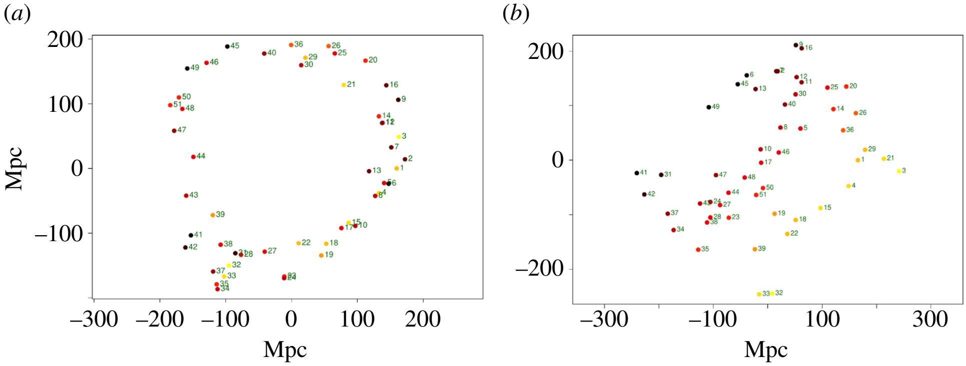


Figure 11. (*a*) The BR absorber members are projected on to a plane that resembles the original line-of-sight projection. (*b*) The BR absorber members are projected on to a plane that resembles viewing the BR from a southeast direction. In this figure, the BR appears to resemble a coil shape. In both figures, the colours represent the redshifts of the absorbers, with nearer-z absorbers being lighter in colour and farther-z absorbers being darker. The small numbers paired with each data point indicate their unique ID numbers.

5. Discussion and conclusions

We have reviewed two recently discovered uLSSs: the GA and BR. Both were discovered visually as prominent overdensities in Mg II images and subsequently supported by a range of statistical assessments and by independent corroboration from DR16Q quasars and DESI clusters. Both exceed the Yadav estimated upper limit to the scale of homogeneity and add to an accumulating list of uLSSs that challenge the CP in this respect. We note that interpretations of the CP vary, however.

The GA and BR have intriguing morphologies, as indicated by their names. There is even a hint that the GA could extend into a GR that envelops the BR. While the BR appears entirely as a ring in projection on the sky, in 3D visualizations, it appears to coil into and out of a central flat ring (which contains most of the member absorbers).

Extreme-value analysis and cosmological simulations have shown that some individual occurrences of uLSSs may be consistent with the CP and with ΛCDM, but it is not obvious that the entire accumulating set of uLSSs will be consistent.

Given their morphologies, perhaps the GA and BR, in particular, require an explanation outside ΛCDM. One possibility might be cosmic strings, which have become of topical interest in recent work [46–53]. For the BR specifically, we noted its similarity in projection and radius to an individual baryonic acoustic oscillation (BAO) [54–58]. However, given the non-spherical, coiling nature of the BR, and its actually being larger than a BAO ($r \sim 200$ Mpc compared with $r \sim 150$ Mpc for a BAO), the BR having an origin in BAOs is probably ruled out.

Unexpected and apparently anomalous discoveries in cosmology, such as uLSSs in general and the GA and BR in particular, may indicate a route to further understanding and refinements of the standard model. It may be productive to see where they lead.

Data accessibility. Our data have depended on the publicly available Sloan Digital Sky Survey quasar catalogue and the corresponding Mg II catalogues.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. A.M.L.: writing—original draft; R.C.: supervision, writing—original draft; G.W.: writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. Our data have depended on the publicly available Sloan Digital Sky Survey quasar catalogue and the corresponding Mg II catalogues.[2] A.M.L. was supported by a UCLan/JHI PhD studentship.

References

1. Huterer D. 2023 Growth of cosmic structure. *Astron. Astrophys. Rev.* **31**, 2. (doi:10.1007/s00159-023-00147-4)
2. Chow-Martínez M, Andernach H, Caretta CA, Trejo-Alonso JJ. 2014 Two new catalogues of superclusters of Abell/ACO galaxy clusters out to redshift 0.15. *Mon. Not. R. Astron. Soc.* **445**, 4073–4085. (doi:10.1093/mnras/stu1961)
3. Peebles PJE. 1993 *Principles of physical cosmology*. Princeton: Princeton University Press.
4. Yadav JK, Bagla JS, Khandai N. 2010 Fractal dimension as a measure of the scale of homogeneity. *Mon. Not. R. Astron. Soc.* **405**, 2009–2015. (doi:10.1111/j.1365-2966.2010.16612.x)
5. Kumar Aluri P *et al.* 2023 Is the observable Universe consistent with the cosmological principle? *Class. Quantum Gravity* **40**, 094001. (doi:10.1088/1361-6382/acbefc)
6. Park C, Choi YY, Kim J, Gott III JR, Kim SS, Kim KS. 2012 The challenge of the largest structures in the universe to cosmology. *Astrophys. J. Lett.* **759**, L7. (doi:10.1088/2041-8205/759/1/l7)
7. Schwarz DJ. 2009 Thoughts on the Cosmological Principle. In *Fundamental interactions: a memorial volume for wolfgang kummer*, pp. 267–276. Singapore: World Scientific Publishing Co. Pte. Ltd. (doi:10.1142/9789814277839_0015)
8. Andrade U, Gonçalves RS, Carvalho GC, Bengaly CAP, Carvalho JC, Alcaniz J. 2022 The angular scale of homogeneity with SDSS-IV DR16 luminous red galaxies. *J. Cosmol. Astropart. Phys.* **2022**, 088. (doi:10.1088/1475-7516/2022/10/088)
9. Gonçalves RS, Carvalho GC, Andrade U, Bengaly CAP, Carvalho JC, Alcaniz J. 2021 Measuring the cosmic homogeneity scale with SDSS-IV DR16 quasars. *J. Cosmol. Astropart. Phys.* **2021**, 029. (doi:10.1088/1475-7516/2021/03/029)
10. Gonçalves RS, Carvalho GC, Bengaly CAP, Carvalho JC, Alcaniz JS. 2018 Measuring the scale of cosmic homogeneity with SDSS-IV DR14 quasars. *Mon. Not. R. Astron. Soc.* **481**, 5270–5274. (doi:10.1093/mnras/sty2670)
11. Lopez AM, Clowes RG, Williger GM. 2022 A Giant Arc on the Sky. *Mon. Not. R. Astron. Soc.* **516**, 1557–1572. (doi:10.1093/mnras/stac2204)

---

[2] For the Anand21 Mg II catalogues, see https://wwwmpa.mpa-garching.mpg.de/SDSS/MgII/. For the Z&M Mg II catalogues, see https://www.guangtunbenzhu.com/jhu-sdss-metal-absorber-catalog.

12. Nadathur S. 2013 Seeing patterns in noise: gigaparsec-scale 'structures' that do not violate homogeneity. *Mon. Not. R. Astron. Soc.* **434**, 398–406. (doi:10.1093/mnras/stt1028)
13. Marinello GE, Clowes RG, Campusano LE, Williger GM, Söchting IK, Graham MJ. 2016 Compatibility of the Large Quasar Groups with the Concordance Cosmological Model. *Mon. Not. R. Astron. Soc.* **461**, 2267–2281. (doi:10.1093/mnras/stw1513)
14. Lopez AM, Clowes RG, Williger GM. 2024 A Big Ring on the sky. *J. Cosmol. Astropart. Phys.* **2024**, 055. (doi:10.1088/1475-7516/2024/07/055)
15. Geller MJ, Huchra JP. 1989 Mapping the Universe. *Science* **246**, 897–903. (doi:10.1126/science.246.4932.897)
16. Gott III JR, Jurić M, Schlegel D, Hoyle F, Vogeley M, Tegmark M, Bahcall N, Brinkmann J. 2005 A Map of the Universe. *Astrophys. J.* **624**, 463–484. (doi:10.1086/428890)
17. Pomarède D, Tully RB, Graziani R, Courtois HM, Hoffman Y, Lezmy J. 2020 Cosmicflows-3: The South Pole Wall. *Astrophys. J.* **897**, 133. (doi:10.3847/1538-4357/ab9952)
18. Webster A. 1982 The clustering of quasars from an objective-prism survey. *Mon. Not. R. Astron. Soc.* **199**, 683–705. (doi:10.1093/mnras/199.3.683)
19. Crampton D, Cowley AP, Hartwick FDA. 1987 The space distribution of faint quasars from the CFHT survey. *Astrophys. J.* **314**, 129. (doi:10.1086/165045)
20. Crampton D, Cowley AP, Hartwick FDA. 1989 Redshifts of quasars in the CFHT/MMT survey - Further evidence for isolated groups. *Astrophys. J.* **345**, 59. (doi:10.1086/167881)
21. Clowes RG, Harris KA, Raghunathan S, Campusano LE, Söchting IK, Graham MJ. 2013 A structure in the early Universe at $z \sim 1.3$ that exceeds the homogeneity scale of the R-W concordance cosmology. *Mon. Not. R. Astron. Soc.* **429**, 2910–2916. (doi:10.1093/mnras/sts497)
22. Horváth I, Hakkila J, Bagoly Z. 2014 Possible structure in the GRB sky distribution at redshift two. *Astron. Astrophys.* **561**, L12. (doi:10.1051/0004-6361/201323020)
23. Horváth I, Bagoly Z, Hakkila J, Tóth LV. 2015 New data support the existence of the Hercules-Corona *Borealis* Great Wall. *Astron. Astrophys.* **584**, A48. (doi:10.1051/0004-6361/201424829)
24. Balazs LG, Bagoly Z, Hakkila JE, Horvath I, Kobori J, Racz II, Toth LV. 2015 A giant ring-like structure at displayed by GRBs. *Mon. Not. Roy. Astron. Soc.* **452**, 2236. (doi:10.1093/mnras/stv1421)
25. Balázs LG, Rejtő L, Tusnády G. 2018 Some statistical remarks on the giant GRB ring. *Mon. Not. R. Astron. Soc.* **473**, 3169–3179. (doi:10.1093/mnras/stx2550)
26. Bergeron J. 1988 Galaxies Giving Rise to MgII Absorption Systems in Quasar Spectra. In *Large scale structures of the universe*, pp. 343–358. Netherlands: Springer. (doi:10.1007/978-94-009-2995-1_48)
27. Steidel CC, Sargent WLW. 1992 MG II absorption in the spectra of 103 QSOs - Implications for the evolution of gas in high-redshift galaxies. *Astrophys. J. Suppl. Ser.* **80**, 1–108. (doi:10.1086/191660)
28. Churchill CW, Kacprzak GG, Steidel CC. 2005 MgII absorption through intermediate redshift galaxies. *Proc. Int. Astron. Union* **1**, 24–41. (doi:10.1017/s1743921305002401)
29. Kacprzak GG, Churchill CW, Steidel CC, Murphy MT. 2008 Halo gas cross sections and covering fractions of Mg II absorption selected galaxies. *Astron. J.* **135**, 922–927. (doi:10.1088/0004-6256/135/3/922)
30. Barnes LA, Garel T, Kacprzak GG. 2014 Ly-α and Mg II as Probes of Galaxies and Their Environment. *Publ. Astron. Soc. Pac.* **126**, 969–1009. (doi:10.1086/679178)
31. Schneider DP *et al.* 2010 The sloan digital sky survey quasar catalog. V. seventh data release. *Astron. J.* **139**, 2360–2373. (doi:10.1088/0004-6256/139/6/2360)
32. Pâris I *et al.* 2017 The Sloan Digital Sky Survey Quasar Catalog: Twelfth data release. *Astron. Astrophys.* **597**, A79. (doi:10.1051/0004-6361/201527999)
33. Lyke BW *et al.* 2020 The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release. *Astrophys. J. Suppl. Ser.* **250**, 8. (doi:10.3847/1538-4365/aba623)
34. Zhu G, Ménard B. 2013 The jhu-sdss metal absorption line catalog: redshift evolution and properties of Mg II absorbers. *Astrophys. J.* **770**, 130. (doi:10.1088/0004-637x/770/2/130)
35. Anand A, Nelson D, Kauffmann G. 2021 Characterizing the abundance, properties, and kinematics of the cool circumgalactic *medium* of galaxies in absorption with SDSS DR16. *Mon. Not. R. Astron. Soc.* **504**, 65–88. (doi:10.1093/mnras/stab871)

36. Webster A. 1976 The Clustering of Radio Sources -- I the theory of power-spectrum analysis. *Mon. Not. R. Astron. Soc.* **175**, 61–70. (doi:10.1093/mnras/175.1.61)
37. Webster A. 1976 The Clustering of Radio Sources -- II the 4C, GB and MC1 surveys. *Mon. Not. R. Astron. Soc.* **175**, 71–83. (doi:10.1093/mnras/175.1.71)
38. Cuzick J, Edwards R. 1990 Spatial Clustering for Inhomogeneous Populations. *J. R. Stat. Soc. Ser. B* **52**, 73–104. (doi:10.1111/j.2517-6161.1990.tb01773.x)
39. Koch EW, Rosolowsky EW. 2015 Filament identification through mathematical morphology. *Mon. Not. R. Astron. Soc.* **452**, 3435–3450. (doi:10.1093/mnras/stv1521)
40. Mookerjea B, Veena VS, Güsten R, Wyrowski F, Lasrado A. 2023 Spiral structure and massive star formation in the hub-filament-system G326.27-0.49. *Mon. Not. R. Astron. Soc.* **520**, 2517–2533. (doi:10.1093/mnras/stad215)
41. Zhang S *et al.* 2023 ATOMS: ALMA three-millimeter observations of massive star-forming regions – XIII. Ongoing triggered star formation within clump-fed scenario found in the massive (∼1500 M⊙) clump. *Mon. Not. R. Astron. Soc.* **520**, 322–352. (doi:10.1093/mnras/stad011)
42. Meidt SE *et al.* 2023 PHANGS–JWST First Results: Interstellar Medium Structure on the Turbulent Jeans Scale in Four Disk Galaxies Observed by JWST and the Atacama Large Millimeter/submillimeter Array. *Astrophys. J. Lett.* **944**, L18. (doi:10.3847/2041-8213/acaaa8)
43. Clowes RG, Campusano LE, Graham MJ, Söchting IK. 2012 Two close large quasar groups of size ∼ 350 Mpc at. *Mon. Not. R. Astron. Soc.* **419**, 556–565. (doi:10.1111/j.1365-2966.2011.19719.x)
44. Pilipenko SV. 2007 The space distribution of quasars. *Astron. Rep.* **51**, 820–829. (doi:10.1134/s106377290710006x)
45. Zou H *et al.* 2021 Galaxy Clusters from the DESI Legacy Imaging Surveys. I. Cluster Detection. *Astrophys. J., Suppl. Ser.* **253**, 56. (doi:10.3847/1538-4365/abe5b0)
46. Ahmed W, Chowdhury TA, Nasri S, Saad S. 2024 Gravitational waves from metastable cosmic strings in the Pati-Salam model in light of new pulsar timing array data. *Phys. Rev. D* **109**, 015008. (doi:10.1103/physrevd.109.015008)
47. Cyr B, Chluba J, Acharya SK. 2024 Cosmic string solution to the radio synchrotron background. *Phys. Rev. D* **109**. (doi:10.1103/physrevd.109.l121301)
48. Ellis J, Lewicki M, Lin C, Vaskonen V. 2023 Cosmic superstrings revisited in light of NANOGrav 15-year data. *Phys. Rev. D* **108**, 103511. (doi:10.1103/physrevd.108.103511)
49. Gouttenoire Y, Vitagliano E. 2024 Domain wall interpretation of the PTA signal confronting black hole overproduction. *Phys. Rev. D* **110**. (doi:10.1103/physrevd.110.l061306)
50. Jiao H, Brandenberger R, Refregier A. 2023 Early structure formation from cosmic string loops in light of early JWST observations. *Phys. Rev. D* **108**, 043510. (doi:10.1103/physrevd.108.043510)
51. Peebles PJE. 2023 Flat patterns in cosmic structure. *Mon. Not. R. Astron. Soc.* **526**, 4490–4501. (doi:10.1093/mnras/stad3051)
52. Sanyal S. 2022 Nambu Goto cosmic strings in the early universe. *Eur. Phys. J. Spec. Top.* **231**, 83–89. (doi:10.1140/epjs/s11734-021-00359-8)
53. Wang Z, Lei L, Jiao H, Feng L, Fan YZ. 2023 The nanohertz stochastic gravitational wave background from cosmic string loops and the abundant high redshift massive galaxies. *Sci. China Phys. Mech. Astron.* **66**, 120403. (doi:10.1007/s11433-023-2262-0)
54. Tully RB, Howlett C, Pomarède D. 2023 Ho'oleilana: An Individual Baryon Acoustic Oscillation? *Astrophys. J.* **954**, 169. (doi:10.3847/1538-4357/aceaf3)
55. Einasto M, Heinämäki P, Liivamägi LJ, Martínez VJ, Hurtado-Gil L, Arnalte-Mur P, Nurmi P, Einasto J, Saar E. 2016 Shell-like structures in our cosmic neighbourhood. *Astron. Astrophys.* **587**, A116. (doi:10.1051/0004-6361/201526769)
56. Planck Collaboration. 2016 Planck 2015 results. XIII. Cosmological parameters, Astron. *Astrophys.* **594**, A13. (doi:10.1051/0004-6361/201525830)
57. Anderson L *et al.* 2014 The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples. *Mon. Not. R. Astron. Soc.* **441**, 24–62. (doi:10.1093/mnras/stu523)
58. Eisenstein DJ *et al.* 2005 Detection of the Baryon Acoustic Peak in the Large‐Scale Correlation Function of SDSS Luminous Red Galaxies. *Astrophys. J.* **633**, 560–574. (doi:10.1086/466512)