Identification of Cosmic Filaments using the `Simba-C` simulation and `DisPerSE` Filament Finder

Jaydon K. Durow[1], **Renier T. Hough**[1], **Amare Abebe**[1,2], **Yin-Zhe Ma**[3]

[1]Centre for Space Research, North-West University, Potchefstroom 2520, South Africa
[2]National Institute for Theoretical and Computational Sciences (NITheCS), Potchefstroom 2520, South Africa
[3]University of Stellenbosch, Stellenbosch 7602, South Africa

E-mail: jaydondurow@gmail.com

Abstract. We used the `DisPerSE` software to identify higher dense regions of dark matter, gas, stars, galaxies and clusters called filaments. We did this for a $0.5\,\mathrm{Mpc}$ thick slice of a `Simba-C` cosmological simulation at the z=0 redshift. The filaments in the simulation that were identified correlated well with the higher density regions of gas particles in the slice. Some lower gas density regions were identified due to our choice of persistence threshold value of $4.2 \times 10^{21}\,\frac{\mathrm{g}}{\mathrm{cm}^3}$ for our noise in `DisPerSE`. We use the filaments we obtained from `DisPerSE` to identify where the gas particles that are under the identified filament's gravity reside. The amount of insignificant structures, usually attributed to noise, identified by `DisPerSE` depends strongly on the persistence threshold value you choose.

1 Introduction

The Universe at the largest scales is isotropic and homogeneous. It evolves according to the Friedmann equations, derived from general relativity [1]. But at smaller scales, which are observed to be inhomogeneous, we use gravitational perturbation theory to describe how these small deviations from homogeneity results in large changes to the structures in the Universe [1].

These complex structures at some of the largest scales of the Universe have been observed through observational catalogues of galaxies and simulations of the Universe [2]. These structures are believed to come from small anisotropic matter distributions as a result of weak ripples and quantum fluctuations in the otherwise uniform and rapidly expanding early universe [3]. These ripples have expanded and evolved due to gravitational forces into the Large-Scale Structure (LSS) or 'Cosmic-Web' that we observe today [3].

The LSS consists of filaments of high concentrations of galaxies and gas clusters that are surrounded by large voids with extremely low dark and baryonic matter density [2]. These filaments were first seen in simulations of the Universe [3]. Filaments are shaped by concentrations of dark matter that attract baryonic matter and more dark matter into the clusters of galaxies through gravity making the clusters larger [3]. As the gas and dark matter move toward the clusters they create filaments [3]. The concentration of the gas results in the density of cluster exceeding Jeans density making it dense enough for stars to form [3]. A balance between the mass and temperature of the gas in this dense environment allows the formation of stars [3]. These stars group together to form clusters of stars that then merge with each other to create galaxies and the galaxies come together to form galaxy clusters [3]. Because the Universe started out homogeneous on the scales of the LSS and evolved to the complex structure we see in the local Universe today, observations should find that the Universe becomes more homogeneous the further they observe from the local Universe. The furthest you can observe is the cosmic microwave background radiation

(CMB).

Observations of the small inhomogeneities in the CMB, corresponding to small anisotropic matter distributions in the early Universe, has given a clear set of initial conditions for the evolution of the LSS [3]. Other observations of galaxies like the 2-degree Field Galaxy Redshift Survey, or 2dFGRS, and the Sloan Digital Sky Survey, or SDSS2, were the first galaxies surveys to have mapped the distribution of the local Universe [3]. They achieved this by taking spectra of more than a million and approximately 250,000 galaxies respectively with the wide field of view, ground based telescopes at Apache Point Observatory in New Mexico and the Anglo-Australian Telescope respectively (https://www.sdss.org/ and http://www.2dfgrs.net/). The link between the early near-uniform Universe and the complex structure seen today have been showed through numerical simulation, observations and theory [3]. The DESI survey is the latest galaxy survey and has mapped millions of galaxies from $z = 0$ to $z \approx 4$ [4]. It will shed light on the LSS of the Universe like never before by giving us a more detailed and more complete look at our sky.

2 Cosmological Simulations

Cosmological simulations are given a set of initial conditions, usually the CMB, to make the simulations as similar and realistic to our Universe as possible, but other random distributions are also used for statistical accuracy. These simulations then use gravity solvers and other physical theories, like AGN feedback mechanisms, to simulate cosmic events and objects. There are four types of cosmological simulations, as outlined by [5]:

- Zoom simulations with dark matter (eg. `Aquarius`).
- Zoom simulations with dark matter and baryonic matter (eg. `FIRE-2`).
- Large volume simulations with dark matter (eg. `Millennium`).
- Large volume simulations with dark matter and baryonic matter (eg. `Simba`).

Zoom simulations are primarily used to simulate smaller volumes with more detail, like simulating a galaxy [5]. Because of the greater level of detail, the computational requirements increases, this restricts the size of the zoom simulations. Large volume simulations simulate a much bigger volume than zoom simulations and are used to study large structures in the Universe [5]. Therefore, we would only find LSS filaments in large volume simulations. Dark matter only simulations are N-body simulations which are easier to simulate [5]. Because they are easier to simulate we can make larger simulations with the same computational cost, which allows us to have a larger sample size thus better statistics. Dark matter and baryonic matter simulations include the hydrodynamical equations that describe the baryonic matter and are thus more complicated to simulate compared to the N-body simulations [5]. Because of the added computational cost of the hydrodynamical simulation we can only run smaller simulations. Therefore, if we want to do statistical analysis we would use dark matter only simulations. If we want to see the effect baryons have on the simulation, we would include hydrodynamical simulation as well. Therefore, we use a large volume dark matter and baryonic matter simulation.

2.1 The `Simba-C` Simulation

The `Simba-C` simulation is based on the `Simba` large-scale cosmological simulation suite. `Simba` from [6] uses `GIZMO`, a hydrodynamics and gravity solver, to make a large-volume cosmic simulation that includes both dark matter and baryonic matter. `Simba-C` includes the `Chem5` chemical enrichment model from [7] and [8], that introduces more elements to the `Simba` simulation. The `Simba` simulation also includes black hole and star formation, as well as an improved Active Galactic Nuclei feedback model [9]. But the `Simba` simulation wasn't explicitly tuned to z = 0 stellar masses and one discrepancy is the low-mass end of the galaxy stellar mass function [6]. Some other large volume dark and baryonic matter simulations we could have used was the `EAGLE` simulation [10]. `EAGLE` aimed to create a faithful population of galaxies by calibrating feedback to match key observations (https://icc.dur.ac.uk/Eagle/index.php) [10]. Balanced calibration means many observables come out right [10]. But some fine-tuning means it's not an entirely independent "prediction" of galaxy formation but rather a model constrained by present-day data [10]. We could also have use the `IllustrisTNG` simulations [11]. `IllustrisTNG` is the next-generation `Illustris` simulation with large volumes including improved galaxy formation physics. Some strengths are its broad range of scales (from dwarf galaxies to clusters) and it is ideal for statistical studies of galaxy evolution [11]. Some limitations are moderate resolution in the largest volume (TNG300) means dwarf galaxies and internal structures are not fully resolved, and some satellite galaxy properties (sizes, quenched fractions of the smallest galaxies) are less reliable [11].

We chose to use the `Simba-C` simulation from [9], because it offers a good balance between size and resolution; so that the filaments we are studying can be resolved and we can have a good sample size. A slice of the specific simulation we used can be seen in Figure 1.

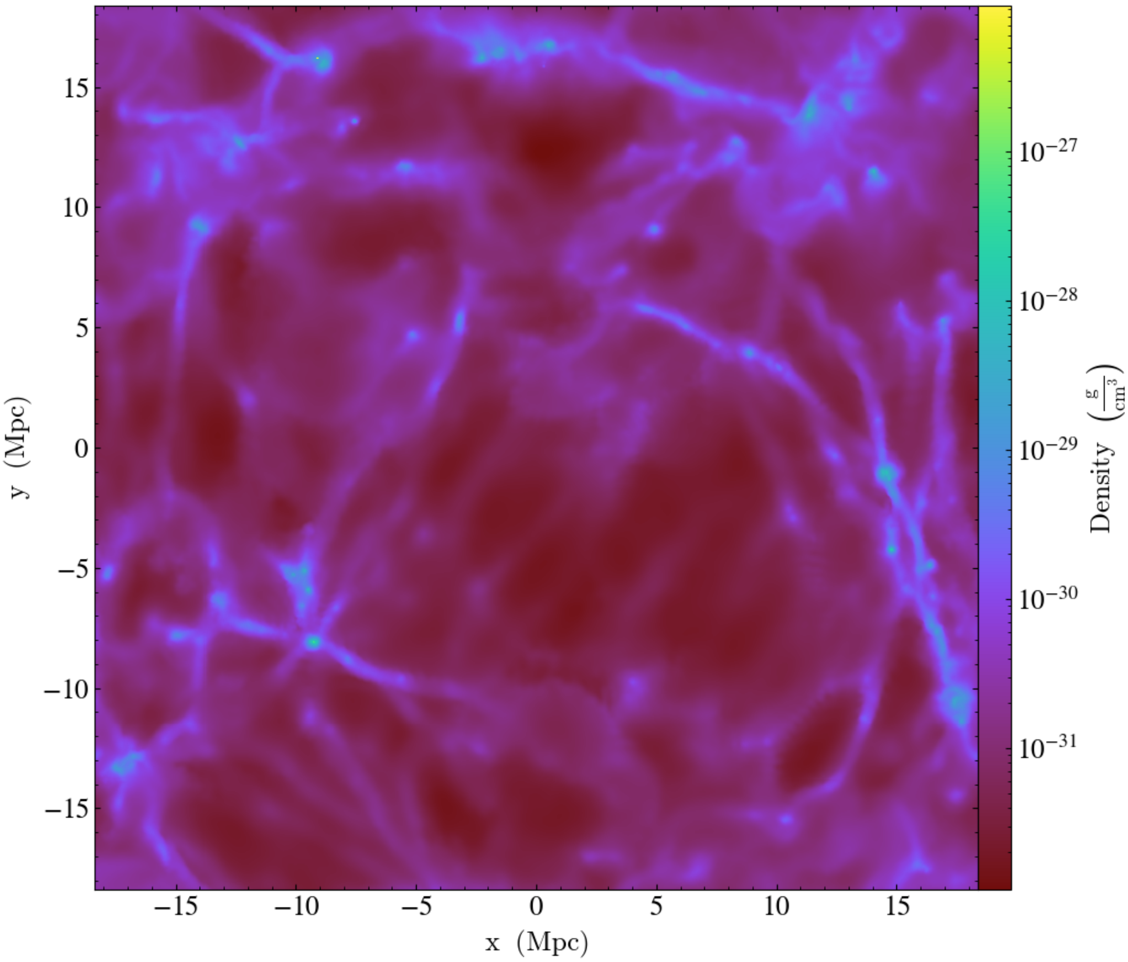

Figure 1: A $0.5\,\mathrm{Mpc}$ slice from the `Simba-C` simulation at the z=0 coordinate at redshift 0. Higher dense regions of gas are light blue indicating a density of about $10^{-29}\,\frac{\mathrm{g}}{\mathrm{cm}^3}$ which we can see form filamentary structures.

3 DisPerSE

`DisPerSE` uses Morse theory and Persistence theory to identify astrophysical structures. Morse theory is a mathematical theory that describes how points in an arbitrary space can be grouped together depending on which maxima or minima they go to [12]. This can be seen in Figure 2. The set of all these groups of points is called a Morse complex and can be used to describe the space [12]. Persistence theory uses the differences of the values between different points to determine which structures are more or less significant [12]. `DisPerSE` breaks up any given space into different sections and uses the sample of data as a density function and assigns a density to each section as shown in Figure 3 [12]. The difference between the values of the sections indicate the gradient of the space [12]. Using these gradients to group sections that go to the same maxima or minima together you make a Morse complex [12]. Which you can see in the bottom right panel in Figure 2. Different types of complexes represent filaments, walls and voids.

The algorithm applies these theories by going over all the give data points and identifying maxima and minima, as well as the points that go towards or away from those critical points, by comparing the values of the data points [12]. All the points that go away from or towards a critical point are grouped together [12]. All the groups of points are then group together depending on the type of critical point and are called manifolds [12]. These manifolds describe the topology of the space, thus identifies the filaments [12].

We converted the simulation file to a fits file and made a mask. With the mse command we used the upSkl flag to identify the filaments and the cut flag with a value of $4.2 \times 10^{21} \frac{\mathrm{g}}{\mathrm{cm}^3}$ for the persistence threshold value. We then used the breakdown and smooth flags as recommended with the skelconv command to combine filaments that overlap and to get a smoother profile of the identified filaments respectively. We then used the toFITS flag with the skelconv to get a fits file of the identified filaments. [1]

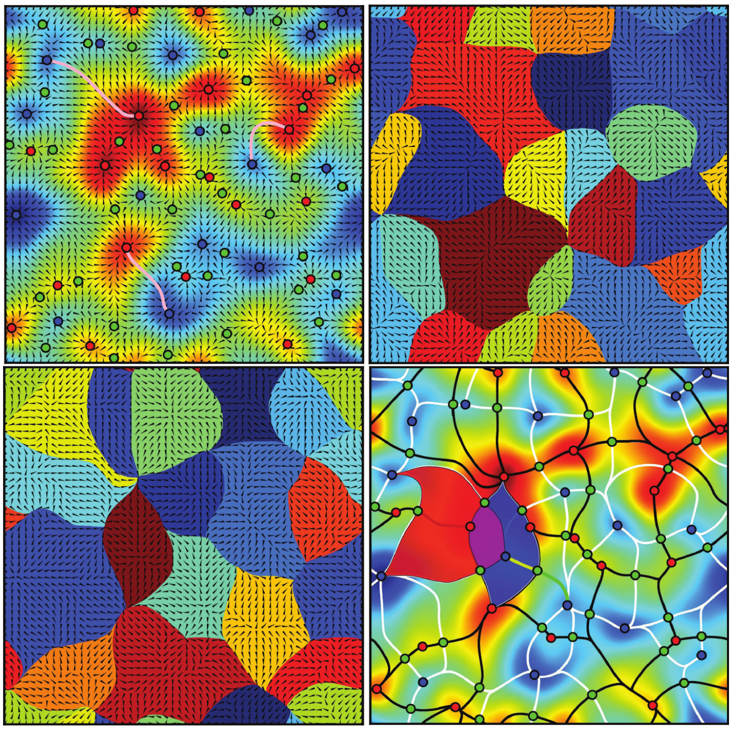

Figure 2: Top left: Gradient field with maxima (blue), minima (red) and saddle (green) critical points. Top right: Collection of sets of points which gradients point away from a particular maxima point. Bottom left: Collection of sets of points which gradients point towards a particular minima. Bottom right: Set of manifolds of the different types of critical points called a Morse-Smale complex. Figure was taken from [12].

4 Results

In the right panel of Figure 4, we see the identified filaments in red that we obtained from `DisPerSE` over layed with the slice from the `Simba-C` simulation. We note that `DisPerSE` identified the filaments well except for the filaments identified in the void in the center of the right panel as well as the small of shoots from the main identified filaments. This could be because of the persistence threshold value we chose for the noise. We tried different threshold values and chose the one that fit the best. In the left panel we used the identified filaments to map which gas particles in the simulation were within a 3D tube with a radius of $0.5\,\mathrm{Mpc}$ around the filament spline. We chose a $0.5\,\mathrm{Mpc}$ radius because it alligned better with the filaments. From the panels we can see that we can now use `DisPerSE` to identify filaments well, as well as identify where gas particles are in relation to these identified filaments. We can use this too observe how the proximity of galaxies to the filaments effects their H1 mass, growth and accretion by comparing their values at different distances from the filaments. In addition, we can look at the filaments effects on the halo mass function, σ_8 value and matter power spectrum at late times by calculating their values at different environments, i.e. in voids, filaments and clusters.

[1] For a detailed guide follow the tutorial at https://www2.iap.fr/users/sousbie/web/html/indexd41d.html?.

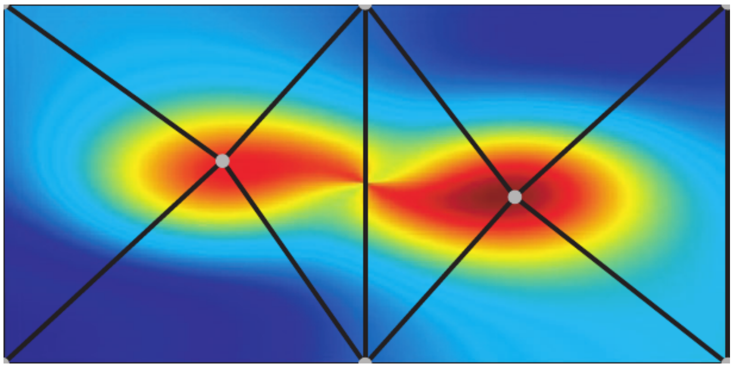

Figure 3: Illustrates a field of values, this can be a smooth function or a set of measured data, e.g. a piece of the sky. This field is broken up into triangles, the most simple 2D shape. Each point, line and area is then assigned the value underneath them or the average of values underneath them. Figure was taken from [12].

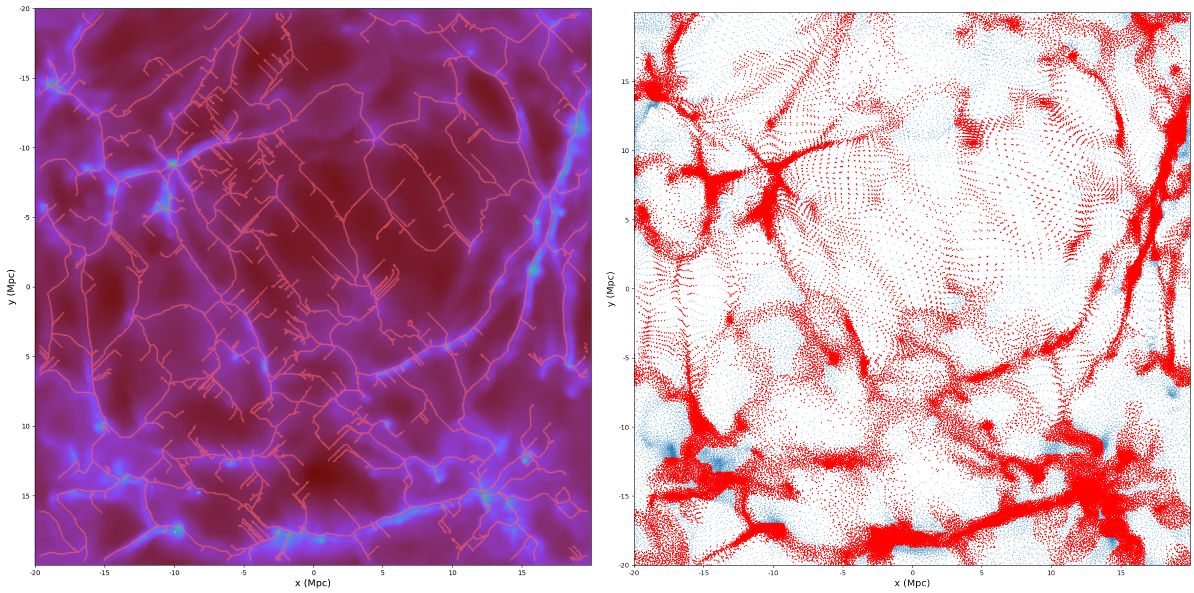

Figure 4: (Right) Filaments (red) overlayed with a slice from the `Simba-C` simulation. (Left) Gas particles (red) in a 3D tube with a radius $0.5\,\mathrm{Mpc}$ around the identified filament spline and gas particles (blue) outside the tube around the filaments.

5 Conclusions

At the largest scales in our Universe we observe higher and lower dense regions called clusters, filaments, walls and voids. The identification of these filaments is important if we want to study how galaxies change as we move from voids to clusters. These structures are also in cosmological simulations like, `Simba-C`. `Simba-C` is a large volume hydrodynamical simulation. We can use a cosmological filament finder called `DisPerSE` to help in identifying these filaments. This software uses mathematical theories to order and describe the data given to it, and outputs the identified structure. We can now identify the filaments in any slice of a `Simba-C` simulation, as well as identify where the gas particles are in relation to the identified filaments seen in Figure 4. The filaments in the simulation that were identified correlated well with the higher density regions of gas particles in the slice. Some lower gas density regions were identified due to our choice of persistence threshold value for our noise. The smaller values for the threshold had more identified lower density regions and the larger values had fewer. We chose the threshold that looked the best.

In the future we could refine the persistence threshold value to improve the identification of the filaments. We could also expand this to the 3D case, as well as identify filaments at different times. Using the identified filaments we could look at the effect proximity of galaxies to the filaments has on their H1 mass, accretion and growth. As well as see the effect on halo mass function, σ_8 value and matter power spectrum at late times. We can change some physical process, like modified gravity and changes to the initial power spectrum, in the simulation to see the effect it has on the filaments. Lastly, we can compare this method for identifying filaments to other methods, like friend-of-friend (fof) finder.

References

[1] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby, *General relativity: an introduction for physicists*. Cambridge University Press, 2006. [Online]. Available: https://doi.org/10.1017/cbo9780511790904

[2] A. L. Coil, *The Large-Scale Structure of the Universe*. Dordrecht: Springer Netherlands, 2013, pp. 387–421. [Online]. Available: https://doi.org/10.1007/978-94-007-5609-0_8

[3] V. Springel, C. S. Frenk, and S. D. White, "The large-scale structure of the universe," *nature*, vol. 440, no. 7088, pp. 1137–1144, 2006. [Online]. Available: https://doi.org/10.1038/nature04805

[4] A. Adame, J. Aguilar, S. Ahlen, S. Alam, D. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud *et al.*, "Desi 2024 vi: cosmological constraints from the measurements of baryon acoustic oscillations," *Journal of Cosmology and Astroparticle Physics*, vol. 2025, no. 02, p. 021, 2025. [Online]. Available: https://doi.org/10.1088/1475-7516/2025/02/021

[5] M. Vogelsberger, F. Marinacci, P. Torrey, and E. Puchwein, "Cosmological simulations of galaxy formation," *Nature Reviews Physics*, vol. 2, no. 1, pp. 42–66, 2020. [Online]. Available: https://doi.org/10.1038/s42254-019-0127-2

[6] R. Davé, D. Anglés-Alcázar, D. Narayanan, Q. Li, M. H. Rafieferantsoa, and S. Appleby, "Simba: Cosmological simulations with black hole growth and feedback," *Monthly Notices of the Royal Astronomical Society*, vol. 486, no. 2, pp. 2827–2849, 2019. [Online]. Available: https://doi.org/10.1093/mnras/stz937

[7] C. Kobayashi, A. I. Karakas, and M. Lugaro, "The origin of elements from carbon to uranium," *The Astrophysical Journal*, vol. 900, no. 2, p. 179, 2020. [Online]. Available: https://doi.org/10.3847/1538-4357/abae65

[8] C. Kobayashi, S.-C. Leung, and K. Nomoto, "New type ia supernova yields and the manganese and nickel problems in the milky way and dwarf spheroidal galaxies," *The Astrophysical Journal*, vol. 895, no. 2, p. 138, 2020. [Online]. Available: https://doi.org/10.3847/1538-4357/ab8e44

[9] R. T. Hough, D. Rennehan, C. Kobayashi, S. I. Loubser, R. Davé, A. Babul, and W. Cui, "Simba-C: an updated chemical enrichment model for galactic chemical evolution in the simba simulation," *Monthly Notices of the Royal Astronomical Society*, vol. 525, no. 1, pp. 1061–1076, 2023. [Online]. Available: https://doi.org/10.1093/mnras/stad2394

[10] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. McCarthy, J. C. Helly *et al.*, "The eagle project: simulating the evolution and assembly of galaxies and their environments," *Monthly Notices of the Royal Astronomical Society*, vol. 446, no. 1, pp. 521–554, 2015. [Online]. Available: https://doi.org/10.1093/mnras/stu2058

[11] A. Pillepich, D. Nelson, L. Hernquist, V. Springel, R. Pakmor, P. Torrey, R. Weinberger, S. Genel, J. P. Naiman, F. Marinacci *et al.*, "First results from the illustristng simulations: the stellar mass content of groups and clusters of galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 475, no. 1, pp. 648–675, 2018. [Online]. Available: https://doi.org/10.1093/mnras/stx3112

[12] T. Sousbie, "The persistent cosmic web and its filamentary structure–i. theory and implementation," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 1, pp. 350–383, 2011. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2011.18394.x