

NEWCASTLE UNIVERSITY

SCHOOL OF MATHS AND STATS

---

# Analysis of Random Fields with Graph Theory

---

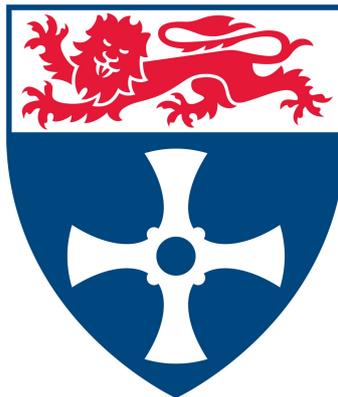
VICTORIA BUCHANAN

*Supervisors:*

PROF. ANVAR SHUKUROV

DR. ALINA VDOVINA

*May 2015*



## **Abstract**

The interstellar medium has a complicated structure with physical variables represented as random fields. There are already several ways to analyse these fields, but those of the interstellar medium are too complicated for these methods. We can consider a new approach, through graph theory, to investigate the gas density distribution obtained from a comprehensive simulation of the interstellar medium in the Milky Way. By implementing constraints on the gas density, we can isolate structures of similar value from the random field. We can then go on to create a graph of vertices and add edges by considering the distance between them. Through the use of the adjacency matrix, eigenvalues and, after considering its stability, the spectral gap, we can analyse the graphs produced from the random field.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Random Fields in the ISM</b>	<b>3</b>
2.1	Density in the ISM . . . . .	3
2.2	Logarithmic density in the ISM . . . . .	5
<b>3</b>	<b>Graph Theory</b>	<b>7</b>
3.1	Geometric centre . . . . .	8
3.2	Centre of mass . . . . .	10
3.3	Comparing centres . . . . .	11
3.4	Isomorphism . . . . .	13
<b>4</b>	<b>Spectral Gap</b>	<b>15</b>
4.1	Moving into 3 dimensions . . . . .	15
4.2	Graphs in 3 dimensions . . . . .	17
4.3	The spectral gap . . . . .	18
<b>5</b>	<b>Stability of the Spectral Gap</b>	<b>21</b>
5.1	Bootstrap statistics . . . . .	21
5.2	Varying the starting vertex . . . . .	23
<b>6</b>	<b>Results and Conclusions</b>	<b>26</b>
6.1	Graphs in 2D . . . . .	26
6.2	Graphs in 3D . . . . .	28
6.3	Conclusion . . . . .	30

# Chapter 1

## Introduction

Random fields occur in many different applications, perhaps most often within science. A field is a set of values in a Euclidean space that is defined over a parameter space of at least one dimension. It takes a data set in terms of a minimum number of parameters needed to model the essential features of a complex pattern to form simpler patterns. Most of the models produce simple random fields, which are a lot easier to work with due to some of the characteristics they have. For example, simple random fields are usually both continuous and statistically homogeneous. This means that the field is similar over the region of the model, making it easier to analyse. Another useful feature of random fields, is that the methods of analysis are applicable on many different scales, for example, time or spatial scales. They can also be used on different ranges within these scales, such as an astronomical or logarithmic scale.

The most common random field for continuous models is the Gaussian field as it follows a normal distribution. They are often multivariate fields determined by Gaussian probability density functions. Two traditional methods, which are most frequently used, for analysing these fields, are through power spectra and statistical moments. The power spectra is based on Fourier, depending on the field either the Fourier series or the fast Fourier transform is used. A Fourier series expresses a continuous random field as a sum of sinusoidal functions of the form

$$F(x) = a_0 + \sum \left[ a_n \cos \left( \frac{n\pi x}{L} \right) + b_n \sin \left( \frac{n\pi x}{L} \right) \right],$$

where  $a_0, a_n$  and  $b_n$  are the constants to be found and  $L$  is the periodicity scale. This method is a great way to find relevant information but is restricted to functions with a repeating pattern and is normally used on very simple fields. The fast Fourier transform can be applied to a field to obtain a continuous sequence of the Fourier coefficients rather than a discrete set. It can be used on more complicated fields than the Fourier series but is also restricted to a repeating pattern in the data.

With the statistical moments method, the random field would usually be defined as a set of finite dimensional distributions and the corresponding probability densities. One benefit of using a Gaussian field ( $f(t)$ ) in this method, is the distribution can be specified by the mean and covariance functions, defined by

$$m(t) = E\{f(t)\}$$

and

$$C(s, t) = E \{(f(s) - m(s)) (f(t) - m(t))\}$$

respectively.

These functions can then be used in a number of ways to find relevant information and estimates for the parameters of the data. The parameters can be used in other functions, like the correlation function, which describes a probability of obtaining similar results in two measurements of a random function. For example, the correlation function is often used to quantify the distance between galaxies, as it describes the probability that another galaxy will be found within a particular distance. Statistical moments have an advantage over the Fourier spectra, as it can be calculated for non-periodic fields. On the other hand, this method does have its disadvantages since the estimates do not always have sufficient statistics.

These methods have many advantages when being used on random fields and are quite often used together for a full analysis of the data. However, they also have many disadvantages. They are both useful for analysing simple and stationary fields but not the more complicated fields. We will be looking at complicated random fields from the interstellar medium. These can not be broken down easily into simple enough fields for methods like those mentioned above, hence, we will consider a new technique of analysis for these fields, through the use of graph theory.

# Chapter 2

## Random Fields in the ISM

The space between stars, in any galaxy, is known as the interstellar medium (ISM), these regions are thought to have very low densities and consist mainly of gas and traces of dust. The ISM can be split into three distinct regions depending on the temperature, these are known as the phases. The cold phase has a typical temperature of 10K, the warm phase is usually of the order between 100 - 1000's K and the hot phase is typically in the order of millions K. The different phases in the ISM all affect its properties and behaviours in different ways. When cold gas is perturbed, it is quick to return to its equilibrium due to short cooling times, whereas in the warm phase, the gas has a longer cooling time. This means the gas has a transient state which is out of the thermal pressure balance. The density in the ISM is also split over these phases, with very high density in the cold phase and very low density in the hot.

### 2.1 Density in the ISM

By taking snapshots of the gas density, from numerical simulations of Gent *et al.*[6], we can obtain a small section and analyse them by placing them on a Cartesian grid of three dimensions. This particular model shows a region which has boundary conditions set at  $-0.5 < x < 0.5$  kpc,  $-0.5 < y < 0.5$  kpc and  $-1 < z < 1$  kpc, about the galactic mid-plane, so that statistical properties of the ISM should be reliably captured and a large amount of density data is available. When we place the model on this Cartesian grid, the mid-plane will be at  $z = 0$  kpc and there is a resolution of about 4 parsecs between each of the points.

Figure 2.1 shows the density in the  $x$ - $y$  plane at the galactic mid-plane. It shows the two most dominant phases with the yellow colour to show the warm phase and the red showing the hot.

From Figure 2.1, we can see the majority of our data is present in the warm phase. We can implement limits in order to restrict the temperature and density to this phase. For the temperature limits, a reasonable identification for the phase can be taken from [6] restricting it between  $500 < T < 5 \times 10^5$  K. Similarly the limits for the density,  $n$ , can be found using the Figure 2.2 [6].

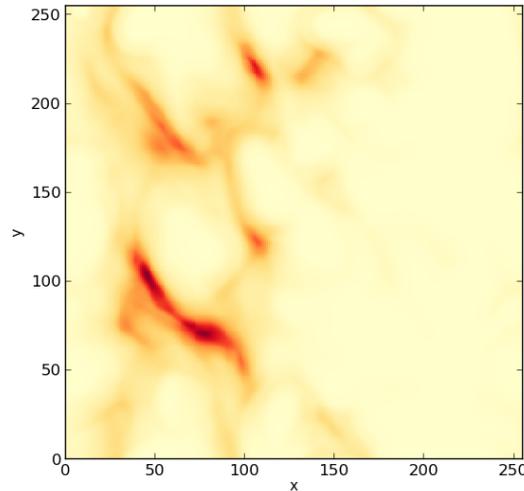


Figure 2.1: Model of the  $x$ - $y$  plane at the Galactic mid-plane ( $z = 0$  kpc).

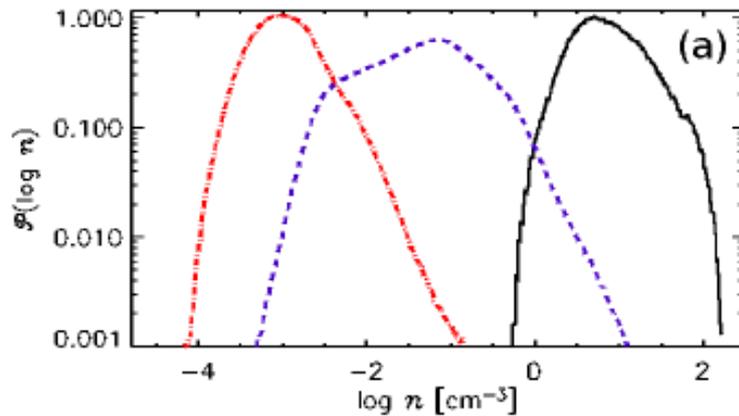


Figure 2.2: Probability distribution of log density, [6].

As shown, in Figure 2.2, for the density probability distribution functions in the three phases, the density values in the warm phase are between the limits  $0.01 < n < 10 \text{ cm}^{-3}$ . The spread of the data in this phase can be seen in Figure 2.3.

The histogram in Figure 2.3(b) shows that most of the density values are in the range  $0 < n < 1 \text{ cm}^{-3}$ , therefore most of the summary statistics, for example the mean or standard deviation, would all lie within this first large bar also. This is reinforced through the Figure 2.3(a), which again shows the density on the  $x$ - $y$  plane at the galactic mid-plane. The black areas of the data are the points which are to be ignored, such as the points in a different phase. From Figure 2.3(a), we can easily see there is no significant change in the colour, which represents the data, across the region. Therefore, it is also describing the data to be within a relatively small range. As a result of this, it is difficult to interpret any results from the figure.

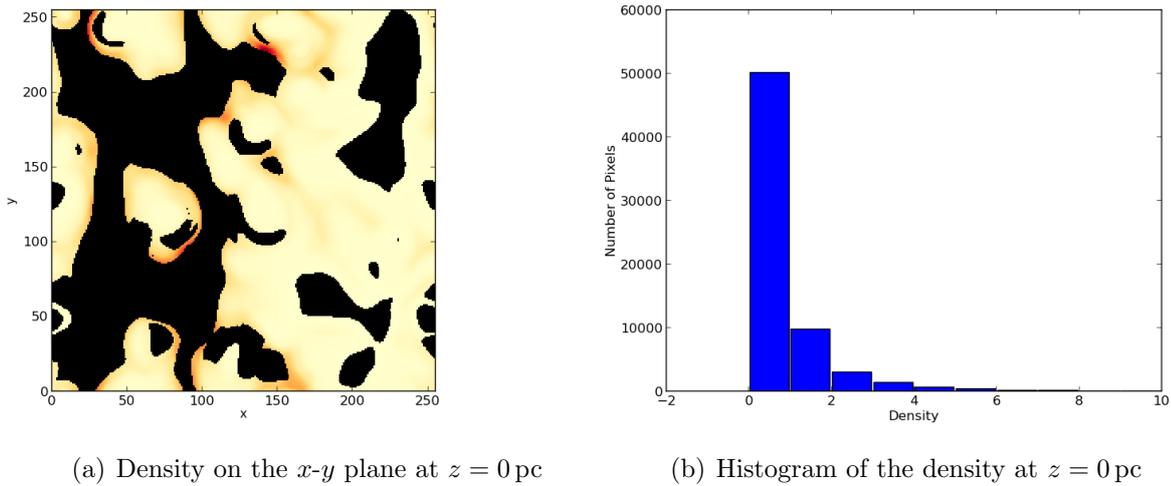


Figure 2.3: Density data in the warm phase.

## 2.2 Logarithmic density in the ISM

By taking the logarithm of the density, our data is spread better over the region and easier to interpret. The histogram in Figure 2.4(b) now has a much improved representation of the data, with a more even spread across the range. This again is reflected in Figure 2.4(a) which shows the contrast in the data is a lot clearer and it is easier to interpret where the high and low density values fall. The image shows a colour spectrum from yellow to red, where the yellow shades are the low densities and the red represents the higher densities. Therefore taking the log density, as the other paper has done, does give a better representation of our data and so we will also use it to analyse the data.

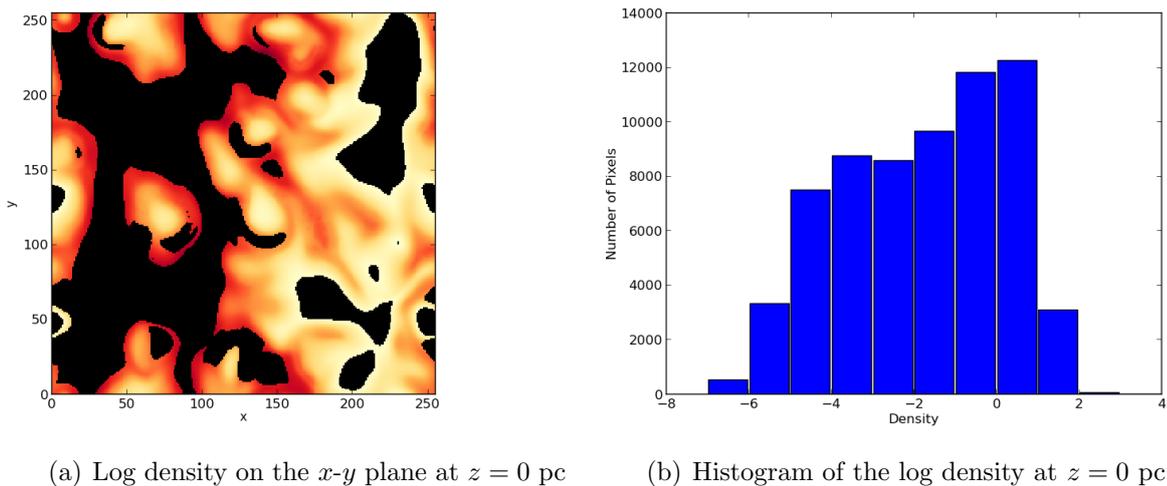


Figure 2.4: Log density data in the warm phase.

From the log density image in Figure 2.4(a), it is clear to see that the gas density exhibits several structures and we want to investigate if there are any connections between them. The connections of interest would be between structures with similar density values,

such as between high density or the values about zero. To only consider similar density values it is necessary to isolate the small range needed and to ignore the rest of the data. Consequently this would make any values not in this range appear black in an image similar to that in Figure 2.4(a).

To isolate this range we consider the equation

$$\log(n) = \langle \log n \rangle + \nu \sigma_{\log n}, \quad (2.1)$$

where  $n$  represents the density,  $\langle \log n \rangle$  is the mean average and  $\sigma_{\log n}$  is the standard deviation of the log density. The constant  $\nu$  can take any real value to give a distinct value of log density, which can then be plotted on an image, like the one in Figure 2.4(a), to show where this value occurs. Figure 2.5 shows the data that are only in the warm phase and within the range  $0 \leq \nu \leq 1$ .

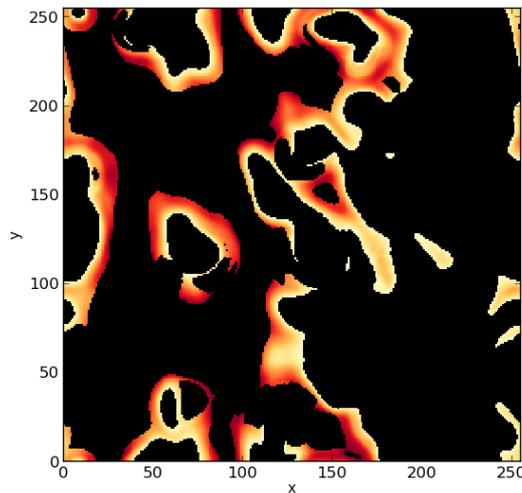


Figure 2.5: Log density in the range  $0 \leq \nu \leq 1$ .

The structures left in the figure can be defined as clusters of similar density. A cluster is defined by two or more values which lie next to each other, including diagonally, i.e it is still one cluster if the value is in one of the eight grid points around the first value (in 2D). However, we will ignore any cluster that is the size of two grid points or smaller, as it is too small to make any significant difference. Two clusters will then be connected if they lie within an appropriate distance to one another.

Most of the clusters are filamentary in shape, hence the connection between two clusters could change depending on where the connection is being made from. An example is if it is being taken from one end or the centre of the shape. Another shape we found in Figure 2.5 is a ring type shape, this means that the middle of the structure is not in the required range. This type of shape could be misleading when finding the connections, depending on where on the ring you are choosing to connect from. A solution to this can be found by using a graph theory approach. In graph theory, each of the clusters could be represented through singular points making it easier to find where they would be connected.

# Chapter 3

## Graph Theory

A graph  $G = (V, E)$  consists of the finite non-empty set of vertices,  $V$ , and the set of edges,  $E$ . An edge  $e \in E$  is formed from an unordered pair of vertices, defined by  $e = (u, v)$  where  $u, v \in V$ . For any graph  $G$ , the following properties always apply:

1. If the set of vertices is empty or only has one component, the graph is *trivial*.
2. A graph with no edges, i.e if the set  $E$  is empty, is known as the empty graph.
3. Any two edges that share one end vertex are *adjacent*.
4. Two vertices  $u$  and  $v$  are *adjacent* if they are connected by an edge.
5. The degree of a vertex  $v$  is the number of edges with  $v$  as an end vertex.
6. A vertex whose degree is 0 is an *isolated* vertex.

Within a graph, we say two vertices are connected if it is possible to go from one to the other by passing only over edges. If this is possible for all vertices then the graph is also called connected. The adjacent vertices and corresponding edges can be expressed as a matrix known as the adjacency matrix. The adjacency matrix for a graph  $G$  with  $p$  vertices is of size  $p \times p$  represented as  $A = (a_{ij})_{p \times p}$  where

$$a_{ij} = \begin{cases} 1, & \text{if vertex } v_i \text{ is adjacent to vertex } v_j, \\ 0, & \text{otherwise.} \end{cases} \quad (3.1)$$

for vertices  $v_i$  and  $v_j$  in the graph  $G$ .

To analyse the clusters for the density data in this way, we need to represent each of the clusters by a vertex and an edge would then connect two different clusters which lie within a chosen distance. As each cluster will be represented as a single vertex, there is a worry that some of the information will be lost. However, the simplified model should still be sufficient as each of the vertices will be placed by the information in the clusters. We will consider two different ways of transforming the clusters into vertices. By using the geometric centre or the centre of mass; we can find which will represent the information in each cluster the best.

### 3.1 Geometric centre

The geometric centre of each cluster can easily be found by taking the minimum and maximum point on each of the axes and finding the midpoint of this range. The centre of a particular cluster on the  $x$ -axis is found by

$$x_{\text{centre}} = x_{\text{min}} + \frac{1}{2}(x_{\text{max}} - x_{\text{min}})$$

and similarly for the  $y$  and  $z$  axes. The points can then be placed onto the clusters to see how well they represent each one. Using the same range as before to view the clusters,  $0 \leq \nu \leq 1$  on the galactic mid-plane, Figure 3.1 shows us how well the geometric centres represent each of the clusters. The geometric centres are represented by the green points.

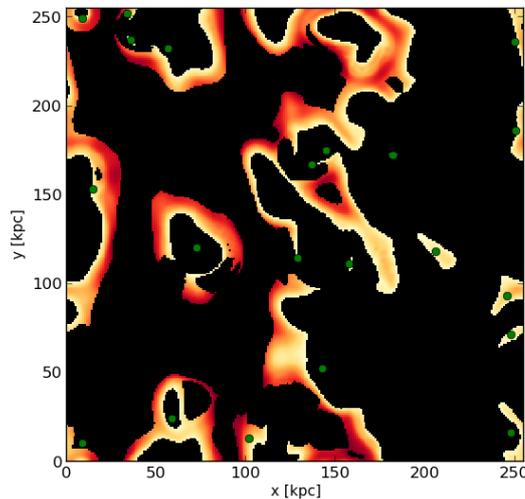


Figure 3.1: The geometric centres (green points) placed on the clusters in the range  $0 \leq \nu \leq 1$ .

The centres, shown in Figure 3.1, vary in how well they appear to represent the clusters. We can take a closer look at several of the clusters in Figure 3.1. The cluster displayed in Figure 3.2(a) has a filamentary shape; these are the most common type of clusters found within the region. We can see that the geometric centre point in Figure 3.2(a) is not actually placed on the structure but just inwards of the left side. The centre is leaning towards the part of the cluster where most of the points occur. This could be difficult to see at first glance due to the nature of the structure, but through this point we can now easily find where the main values in the cluster lie.

Similarly, the geometric centre for the cluster in Figure 3.2(b) is shown to not be on the structure itself. This cluster is a ring-type shape which, as we mentioned previously, could be problematic in finding the connections to other clusters. Using the geometric centre solves this problem because it gives us a point to make the connection from within the ring, whilst also taking into consideration where most of the points are clustered on the structure. In Figure 3.2(b), most of the points are based near the right of the ring,

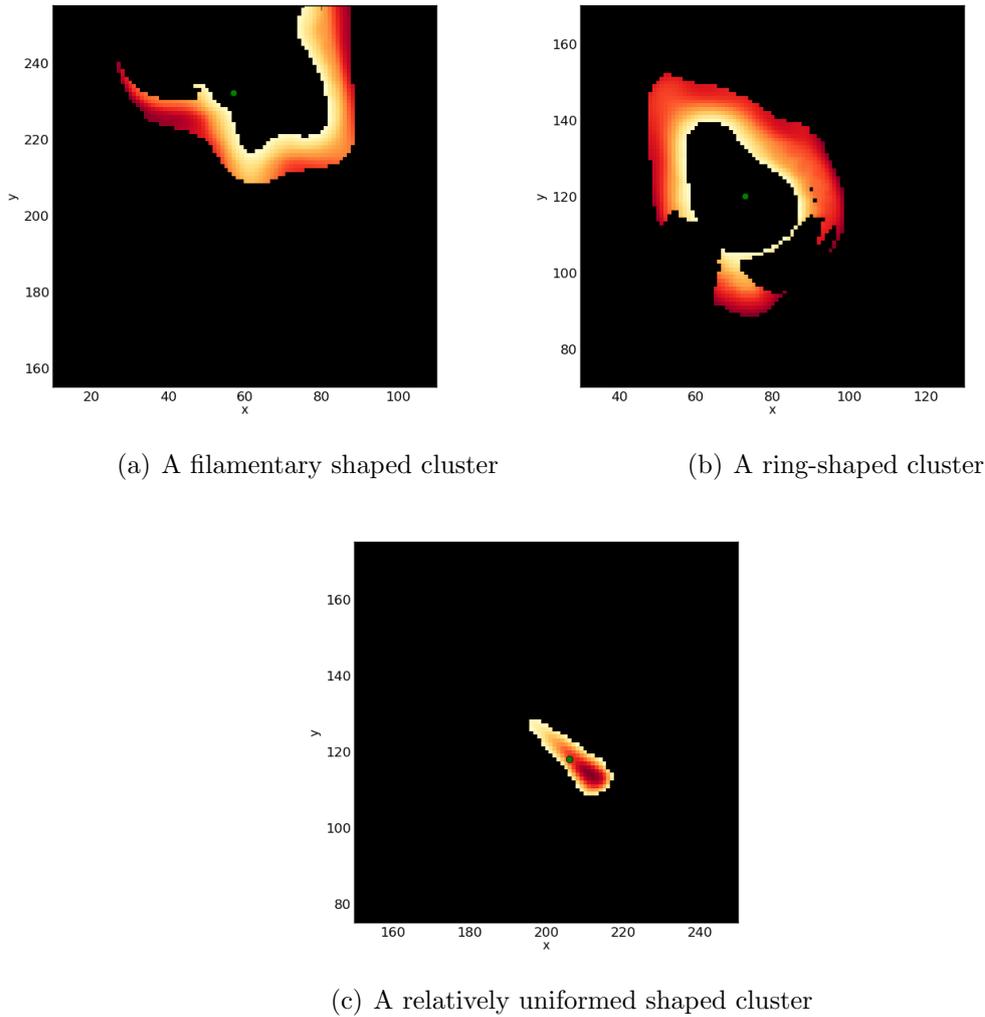


Figure 3.2: Individual clusters from the range  $0 \leq \nu \leq 1$  showing the geometric centre (green point) of each.

therefore the centre lies closer to that side. However, it is clear to see that the other points have been taken into consideration as it does fall near the centre of the ring itself.

Finally Figure 3.2(c) shows a cluster of a relatively more uniform shape compared to the others. This type of shape is a lot less common than the other two, but does occur in the region. One benefit of this type of shape is that we can easily find where a connection would be made from, therefore the geometric centre will be directly on and in the centre of each of these types of structures.

When using the geometric centres to represent the clusters in this way, we need to consider that the density values in each of the structures are being ignored. The connections being made between the structures are for similar values of density, so maybe these values should be taken into consideration when we transform each cluster into a vertex. On the other hand, the structures that are connected, like those in Figure 3.1, are already of a similar density by construction. As the density values were used to place these constraints, the geometric centre could be a sufficient representation of the clusters.

## 3.2 Centre of mass

Using the centre of mass to transform clusters into vertices means the density values in each of the structures are now being taken into consideration. The centre of mass is the point where the density of a cluster is concentrated. We can find this centre, for a single cluster  $i$ , on the  $x$ -axis by

$$x_{\text{centre}} = \frac{\sum n_i x_i}{\sum n_i},$$

with similar equations for  $y$  and  $z$  are also used.

Again using the same range from  $0 \leq \nu \leq 1$ , Figure 3.3 shows how the centres of mass lie on the clusters. The centres of mass are represented by the light blue points on the clusters.

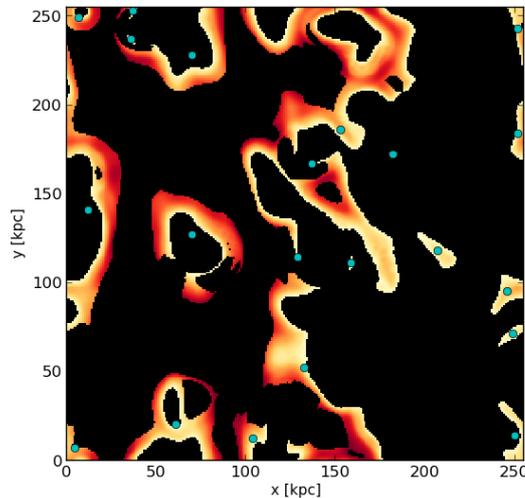


Figure 3.3: The centres of mass (blue points) placed on the clusters in the range  $0 \leq \nu \leq 1$ .

To see how well these centres represent the clusters, we can isolate several of them to take a closer look in Figure 3.4. By taking the same clusters as before, the two different centres can then be compared.

Firstly, Figure 3.4(a), shows a filamentary cluster, as this is the most common type found. The cluster shows the small range of density within it through a colour spectrum from yellow to red, with red being the higher density. Similar to the geometric centre, the centre of mass, for the Figure 3.4(a), does not lie on the structure itself, but is near the centre with the cluster curling around it.

A ring-type cluster is displayed in Figure 3.4(b) and, much like the geometric centre and the filamentary shaped cluster, the centre lies in the middle of the structure and not on it. Just like the geometric centre, it also solves the issue of where we would make the connections from for this type of shape. The centre of mass lies near to the top of the ring, as we can see from the colour spectrum that the higher density values lie here, but has also taken into consideration those that rest at the bottom to find the point which is shown.

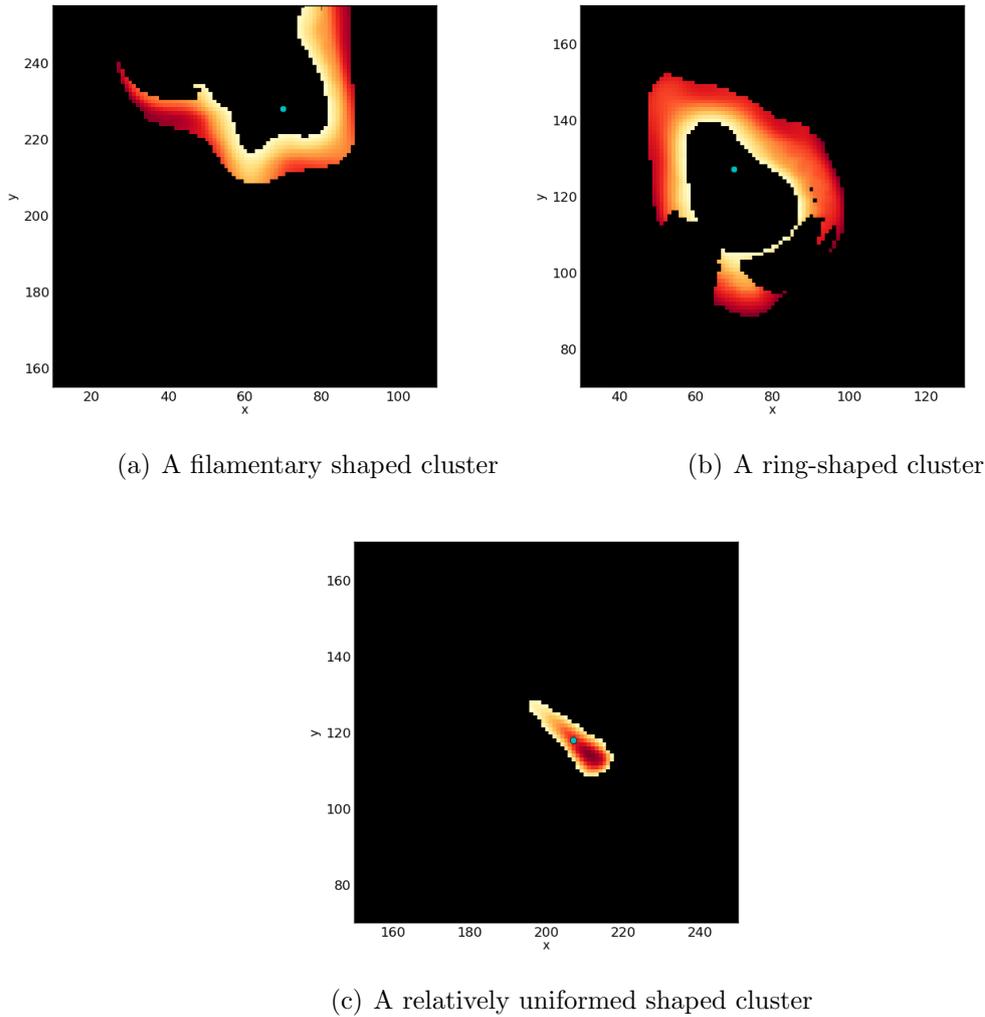


Figure 3.4: Individual clusters from the range  $0 \leq \nu \leq 1$  showing the centre of mass (blue point) of each.

In Figure 3.4(c), the uniformly shaped cluster from before shows the centre of mass on the cluster and close to the centre of it. As the shape in Figure 3.4(c) is relatively uniform compared to the others, the centre of mass will at least lie really close to the geometric centre, if not on top of it.

### 3.3 Comparing centres

Both, the geometric centres and the centres of mass, have advantages and disadvantages to their uses and both represent the data a unique way. By placing both sets of points on top of the clusters, we can compare the different centres to see what differences they have. Figure 3.5 shows the geometric centres in green and the centres of mass in blue on top of the clusters that are in the mid-plane between the range  $0 \leq \nu \leq 1$ .

For most of the clusters in Figure 3.5, the geometric and the centre of mass points can both be seen. However, for those few clusters that are small enough in size or are uniform

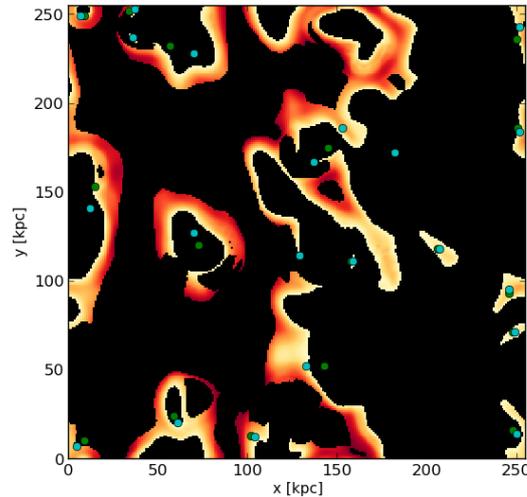
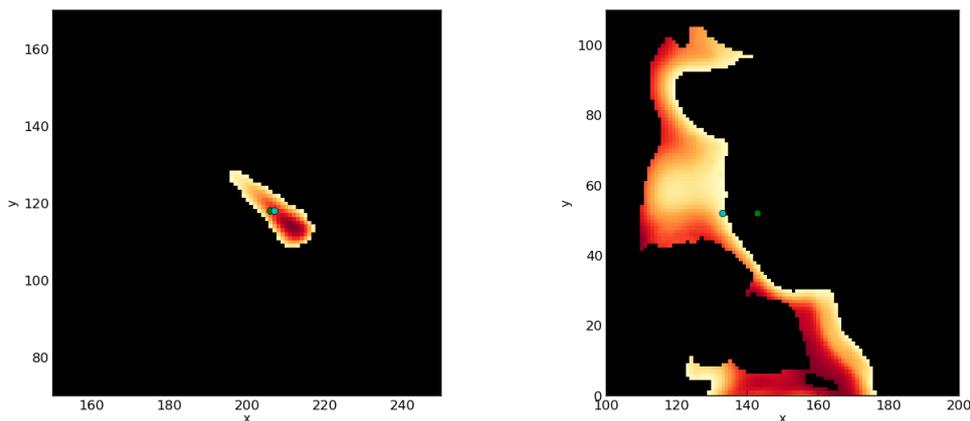


Figure 3.5: The centres of mass (blue points) and geometric centres (green points) placed on the clusters in the range  $0 \leq \nu \leq 1$  at the  $z$  mid-plane.

in shape, we can see only part or none of the geometric centre, as the centre of mass point lies on top of it. One of the problems looking at this figure, is that sometimes finding the corresponding points for each can be difficult. Looking at individual clusters would make it easier for us to see what differences there are between the two centres.

Figure 3.6(a) shows an example of a uniform cluster, where the geometric centre and the centre of mass lie on top of one another, therefore only one can be seen. If they lie really close or on top of one another, such as this example, it would not matter which type of centre we use to form the vertices, as the same points will be formed.



(a) A uniformly shaped cluster

(b) A filamentary shaped cluster

Figure 3.6: Individual clusters from the range  $0 \leq \nu \leq 1$  showing the centre of mass (blue point) and geometric centre (green point) of each.

Most of the centres on the clusters will not lie close to one another, especially as most of the clusters are a non-uniform shape. The cluster in Figure 3.6(b) has a more complicated

structure. It appears to be filamentary in shape, but with a larger number of points occurring in certain areas, like across the bottom. The two different centres are at different positions on the cluster. The geometric centre is to the right of the cluster as there is a large number of points close on the left of it and at the bottom, while the centre of mass lies on the cluster near the middle. From the yellow to red spectrum, we see the higher density is gathered at the bottom and along the filamentary part of the structure. They are both affected by the large cluster of points as these areas also have a higher concentration of density within them.

The difference between these two centres will have an effect on the graphs that each set of vertices will draw. When connecting up the clusters with a distinct radial distance, the centre of mass could connect with a cluster on the left of that in Figure 3.6(b), while the geometric centre may not connect with the same cluster, as the distance between those points could be larger. By choosing a sensible radial distance, the graphs can be compared to see if there are any distinct differences between them.

### 3.4 Isomorphism

Two graphs are said to be isomorphic if, under a transformation, the edges and adjacent vertices are preserved. So even though the shape of the graph has been changed, the vertices and corresponding edges can be mapped onto one another. For example, there is a simple graph in Figure 3.7(a) of four vertices and four edges.

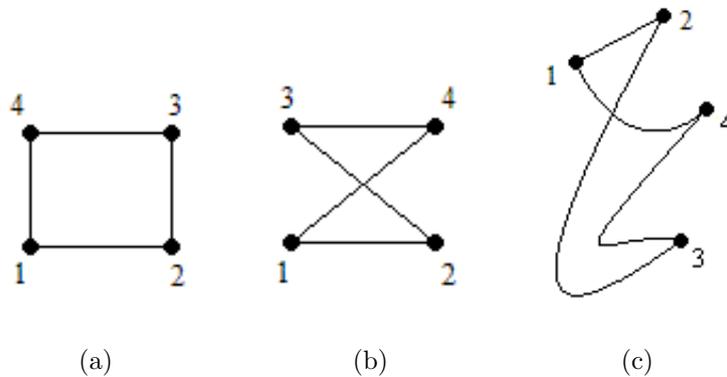


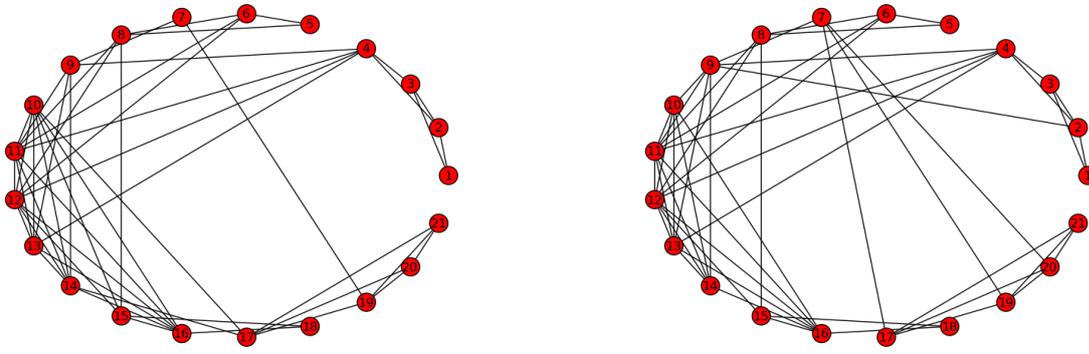
Figure 3.7: Example of isomorphic graphs,[11].

This graph is then transformed to create the graph in Figure 3.7(b) by switching around the two top vertices. This is a simple transformation, so we can see that the second graph preserves the edges between vertices as all the edges still have the same end vertices. Therefore the graphs in Figures 3.7(a) and 3.7(b) are isomorphic to one another. Going back to the simple graph in Figure 3.7(a), this graph can be transformed in a different way to create the graph in Figure 3.7(c). Unlike before, it is harder for us to see how the graph has been changed, but by checking each of the vertices, it can still be shown that the edges and vertices again remain unchanged and can be mapped onto one another. This means that all three graphs in Figure 3.7 are isomorphic to each other.

There is no known algorithm for checking if two graphs are isomorphic to one another,

but comparing each of the vertices can take a large amount of time, especially for larger graphs. A quicker way of finding if two graphs are isomorphic is to compare their adjacency matrices. If they are identical then the two graphs are isomorphic to one another. The two graphs are also isomorphic if the adjacency matrices have the same element values over the shape of the matrix. This means that if the rows of the matrix are permuted, by renumbering the vertices, the adjacency matrix of the graph can be changed into the matrix of the other.

In our case, the graphs of the centres of mass and the geometric centres were formed by connecting the centres by edges that are not longer than 400 parsecs. A few different values were tried, but, for 400 parsecs, the graphs created were not over crowded with edges, nor were there many isolated vertices. Using the clusters in the range  $0 \leq \nu \leq 1$  and the vertices found from the different centres, the graphs in Figure 3.8 were formed.



(a) Graph formed from the centres of mass as vertices.

(b) Graph formed from the geometric centres as vertices

Figure 3.8: Graphs for the range  $0 \leq \nu \leq 1$ , vertices are in an anticlockwise direction, starting with vertex 1 at the clock position of quarter past three.

The graphs have a few properties in common. Neither of them show any isolated vertices and they both show grouped clusters. For example, in both graphs, the vertices from 1 to 4 have edges between them, linking them up in a square-like formation. Also, both graphs are connected, as it is possible to travel to any vertex no matter which one is the beginning vertex. However, there are some differences in these graphs. By comparing the vertex 7, we can see that in the graph 3.8(b) it has a degree of 4, whereas in the graph 3.8(a), this vertex only has a degree of 2. We can also clearly see that on the graph formed from the geometric centres, there is an edge between the vertices 2 and 9, but on the centre of mass graph, this edge is not there as now the centres are too far apart. Therefore these graphs are not isomorphic to one another, though they are very similar.

Comparing the use of the two different centres as vertices has shown that they both represent the clusters in a unique way and both form useful graphs which can be analysed. But from these graphs, it can not be said that the use of one centre is any better than the other, so a different approach to compare them will be discussed in the next Chapter.

# Chapter 4

## Spectral Gap

For any adjacency matrix,  $A$ , of a graph, the *eigenvalues*,  $\lambda$ , of the graph can be found by

$$Ax = \lambda x.$$

As the adjacency matrix is symmetric along its diagonal, by definition, the eigenvalues are real. The set of eigenvalues of the adjacency matrix is known as the *spectrum* of the graph,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . For any given adjacency matrix, it has a unique set of eigenvalues, but in general there is no unique adjacency matrix for a given spectrum. The spectrum of a graph is also known to be *invariant* as the eigenvalues do not depend on the labeling of vertices.

If the number of edges, in a graph, is close to the maximum, when every vertex is connected to every other vertex, the graph is said to be *dense*. A graph in which the number of edges is close to the minimal is called *sparse*. A sparse graph does not have to be connected, but if it is an empty graph, it is trivial. Using the two largest eigenvalues of the spectrum, the *spectral gap*,  $\Upsilon$ , can be defined as

$$\Upsilon = \lambda_1 - \lambda_2. \tag{4.1}$$

The spectral gap is more useful for analysing large graphs with lots of vertices. We would apply it to graphs from the 3D range, rather than the the 2D density data used above.

### 4.1 Moving into 3 dimensions

Figure 4.1 shows the warm phase of the data, taken previously, on the 3D range. A colour code is on the right of the figure to show the spread of density values. The warm phase normally contains about 95% of the data. The mid-plane of the snapshot lies where  $z = 0$  Kpc, this is the 2D slice we originally took and only showed a small fraction of what was happening within the region. By now looking at the larger range, it is clear that the clusters are a lot bigger and have a larger range in density values than could be shown in 2D.

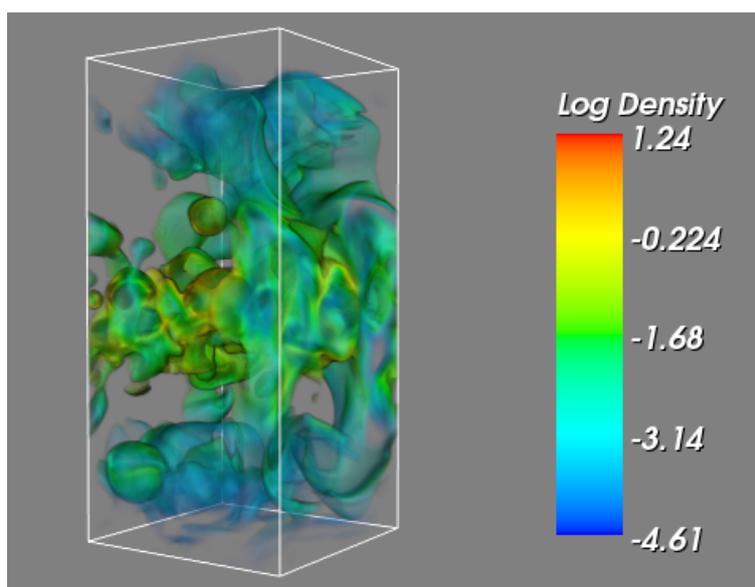


Figure 4.1: Log density of the warm phase in the 3D region.

At the bottom and top of the region there is not much fluctuation as most of these positions have low density. The middle of the region is where most of the fluctuations occur. If we took a similar region as before, the graphs formed would be too complicated. To avoid these overcomplicated graphs, a narrower range of density needs to be taken. Figure 4.2 shows the range  $0 \leq \nu \leq 0.2$  and the colour bar to the right shows how small the spread of the data is.

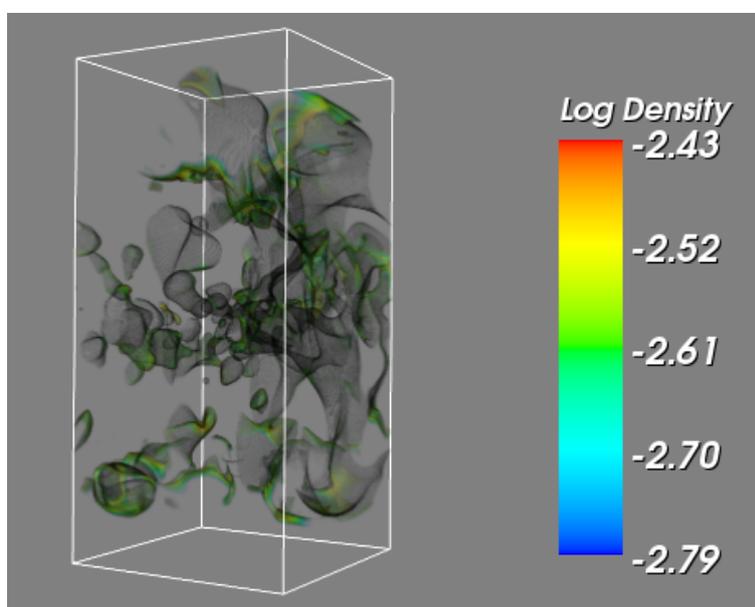
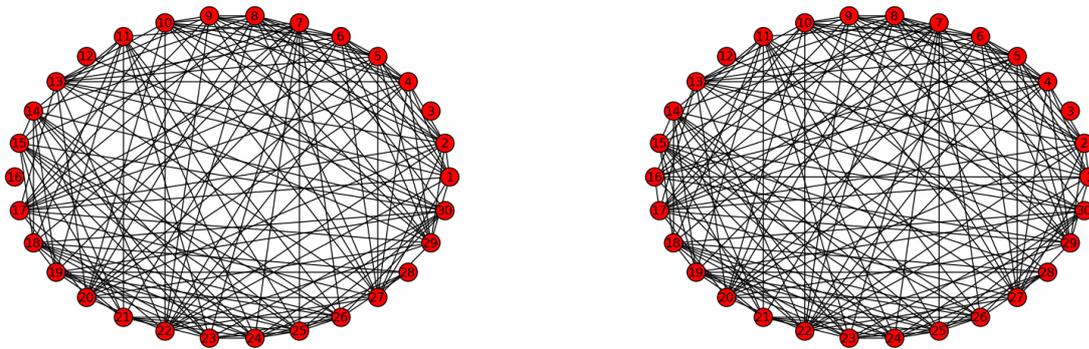


Figure 4.2: 3D image of the log density within the range  $0 \leq \nu \leq 0.2$ .

There are 343 clusters in this range of  $\nu$ . Similar to the 2D image, the geometric centres and the centres of mass can be found for each of the clusters, and then used to form the vertices of a graph.

## 4.2 Graphs in 3 dimensions

Using the same radial distance of 400 parsecs, we can form two graphs, however they will be too large to take any information from. So the graphs could still be analysed, I suggested taking smaller graphs with a limited number of vertices to form subgraphs. In Figure 4.3, the first 30 vertices of each of the graphs and the corresponding connections have been taken to form the two subgraphs. As these graphs have so many edges, any differences in the graphs are very hard to find. Hence comparing the eigenvalues and the spectral gap is very handy. Because the number of eigenvalues for a graph increases as the number of vertices increases, we will focus only on the first two eigenvalues, as these are the values needed for the spectral gap.



(a) Graph formed from the centres of mass as vertices

(b) Graph formed from the geometric centres as vertices

Figure 4.3: Subgraphs of the first 30 clusters in the 3D field for the range  $0 \leq \nu \leq 0.2$ . Vertices are arranged in an anticlockwise direction, starting with vertex 1 positioned at quarter past three on a clock face.

The graphs for the centres of mass and geometric centres in Figure 4.3 would have the first two eigenvalues,  $\lambda_1$  and  $\lambda_2$ , and the spectral gap,  $\Upsilon$  shown in Table 4.1.

Table 4.1: Eigenvalues and spectral gap for the graphs in Figure 4.3 for the range  $0 \leq \nu \leq 0.2$ .

	$\lambda_1$	$\lambda_2$	Spectral Gap, $\Upsilon$
centre of mass	12.74	11.05	1.69
geometric centre	13.19	11.34	1.85

The spectral gaps are rather similar in value. The relative difference,  $\delta$ , can be found by taking the difference in values over the mean of them, using

$$\delta = \frac{|\lambda_1 - \lambda_2|}{\bar{\lambda}} \approx 0.09,$$

where  $\bar{\lambda} = (\lambda_1 + \lambda_2)/2$ . This difference provides an estimate of the accuracy of the spectral gap for the gas density distribution in Figure 4.2. Table 4.2 considers the relative difference for the graphs formed in the interval  $-1 \leq \nu \leq 1$  over a range of 0.2 each time

Table 4.2: The first eigenvalue and spectral gap for the centres of mass, geometric centres and the relative distance,  $\delta$ , over different ranges of  $\nu$ .

Range of $\nu$	Centre of Mass		Geometric centre		Delta, $\delta$
	$\lambda_1$	Spectral Gap, $\Upsilon$	$\lambda_1$	Spectral Gap, $\Upsilon$	
$-1.0 \leq \nu \leq -0.8$	8.91	1.65	8.71	1.47	0.16
$-0.8 \leq \nu \leq -0.6$	8.27	1.21	8.32	1.53	0.23
$-0.6 \leq \nu \leq -0.4$	18.5	13.21	19.05	13.78	0.04
$-0.4 \leq \nu \leq -0.2$	17.46	10.08	17.46	10.3	0.02
$-0.2 \leq \nu \leq 0$	15.05	3.77	15.04	3.77	0
$0 \leq \nu \leq 0.2$	12.74	1.69	13.19	1.85	0.09
$0.2 \leq \nu \leq 0.4$	11.35	6.3	11.37	6.24	0.01
$0.4 \leq \nu \leq 0.6$	10.37	2.42	10.52	2.89	0.18
$0.6 \leq \nu \leq 0.8$	12.23	5.92	12.32	6.15	0.04
$0.8 \leq \nu \leq 1.0$	8.24	2.13	8.02	0.01	1.98

For almost all the ranges in Table 4.2, the relative difference is very small and therefore our choice of centre does not matter much when analysing the graphs. Moreover, the relative difference is larger for  $|\nu| \geq 0.4-0.6$ , where the number of vertices begins to decrease again. We will use the centres of mass to form our graphs from now on.

### 4.3 The spectral gap

By finding  $\Upsilon$  for an increasing number of vertices, it is easy to tell if a graph is sparse and if it is highly connected. The best way to find this is to create a plot of spectral gap by adding in vertices of a graph one at a time. For under about 100-150 vertices the graph will be very unclear, as there are too few vertices. However above this number,  $\Upsilon$  should fluctuate over an asymptotic value.

If the graph does not have a clear lower bound but continues to rise gradually, the graph is well connected, however it is far too dense, so there are too many connections. Whereas if the spectral gap largely rises and falls when the number of vertices increases, but there is still a lower bound, the graph is sparse, but it is not well connected. This means there will be too few edges and hardly any connections between the different clusters.

The spectral gap can be used to find an appropriate connection radius in order to create a graph that will be both sparse and highly connected. Looking at the range  $0 \leq \nu \leq 0.2$  and a radial distance of 400 parsecs, Figure 4.4 shows the spectral gap over this range for an increasing number of vertices.

As previously stated, the spectral gap for under 100-150 vertices can be unpredictable and does not give much information. From Figure 4.4, we can see that for under 100 vertices the graph is not much use, but then it rises suddenly and levels off to form a decent plot

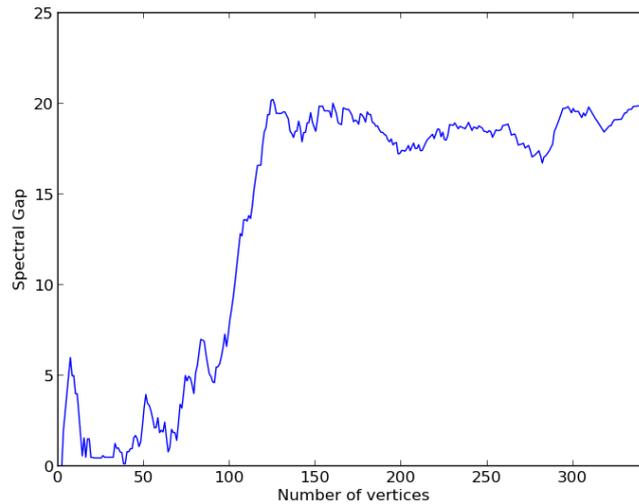
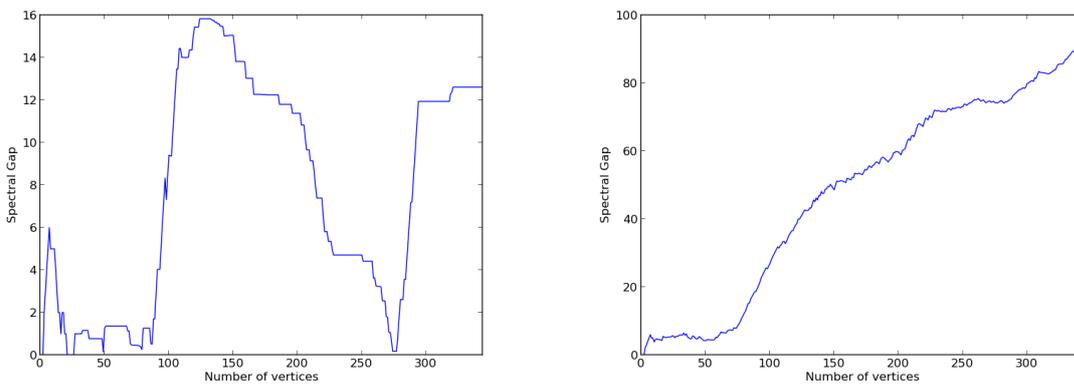


Figure 4.4: Spectral gap versus the number of vertices for the graph with a connection radius of 400 parsecs and  $0 \leq \nu \leq 0.2$ .

of a spectral gap. From about 150 vertices onwards, the graph has mainly levelled out with just a few fluctuations. Ignoring the part of the graph under 150 vertices, we can see that the graph, after it has levelled out, does show that  $\Upsilon$  is bounded from below. Therefore using a connection radius of 400 parsecs was a good choice.

In the Figure 4.5, the spectral gap is shown for two different connection distances to demonstrate that 400 parsecs is a good value for the connection radius, at least for this range.



(a) Using a connection radius or 200 parsecs      (b) Using a connection radius of 600 parsecs

Figure 4.5: Spectral gap over the number of vertices for the whole graph, in the range  $0 \leq \nu \leq 0.2$ .

Figure 4.5(a) shows the spectral gap for a connection radius of 200 parsecs. A large rise, similar to the one in Figure 4.4, is followed by a decline. It does not contain small fluctuations about a value, however it does show some form of a lower bound. The spectral gap shows that the graph is sparse, but has very few connections. There will be many isolated vertices in the graph and therefore many isolated structures. By not

having many connections, the graph will not be useful and consequently the connection radius is too small.

The spectral gap for a connection radius of 600 parsecs is shown in Figure 4.5(b), again we will ignore  $\Upsilon$  for the first 100 vertices. The spectral gap has a gradual incline, but never seems to level off or give an indication of a lower bound. Maybe, over more vertices, it would eventually level off but for this range of  $\nu$ , we can only see that it rises. Therefore, there is no clear lower bound for the spectral gap. This means that even though the graph is highly connected, it is actually too connected and not much information can be gathered from it. This is a result of almost every vertex being connected to every other, so we can conclude that a connection radius of 600 parsecs is too large.

From the spectral gap, we can see that choosing the right connection radius can have a strong effect on the graph. By choosing a radius too big or small, the graph will be of no use. Therefore, 400 parsecs is the best choice to use as the connection radius as it gives a graph that is both sparse and highly connected.

# Chapter 5

## Stability of the Spectral Gap

In this Chapter, we consider the stability of the spectral gap to see how accurate and stable it is. The first approach is a statistical method called bootstrapping, which is commonly used to calculate the scatter of estimated random variables. Then we varied the order of the vertices to see if it had any effect on the spectral gap.

### 5.1 Bootstrap statistics

Bootstrapping involves random replacement sampling and is used to analyse any statistic, like the spectral gap. Although bootstrapping is a relatively simple procedure, the technique is heavily dependent on computer calculations. To use the bootstrap procedure, we randomly choose a certain percentage of the data and allow for replacement to see how the spectral gap is affected. This is done many times, taking a different random sample each time. Some points may be used numerous times and some not at all, due to the randomness of the points being taken.

Because of the simplicity of the bootstrapping technique, it is straight forward to estimate the error for a complex parameter, like the spectral gap, and the stability of the results can be easily checked. But the simplicity is also a disadvantage to this statistic because it hides any assumptions taken when doing the analysis, such as the independence of samples. Bootstrapping a sample is asymptotically more accurate than most other statistical methods for finding the error, so for the spectral gap it will be a good estimator of this error.

To form a sample we chose to replace a random 30% of the vertices. The percentage picked is small enough that there are still enough vertices to find a meaningful spectral gap, but it is also large enough to expect some effect on the spectral gap. Applying this type of statistic once gave the plot in Figure 5.1.

The missing data has had some effect on the spectral gap, but not enough to change it in a big way. The data still has a big rise before levelling off with some small fluctuations. After levelling off, the spectral gap still has a slight rise, but this may be caused by the reduced number of vertices. Comparing this to Figure 4.4, we note that the fluctuations are larger.

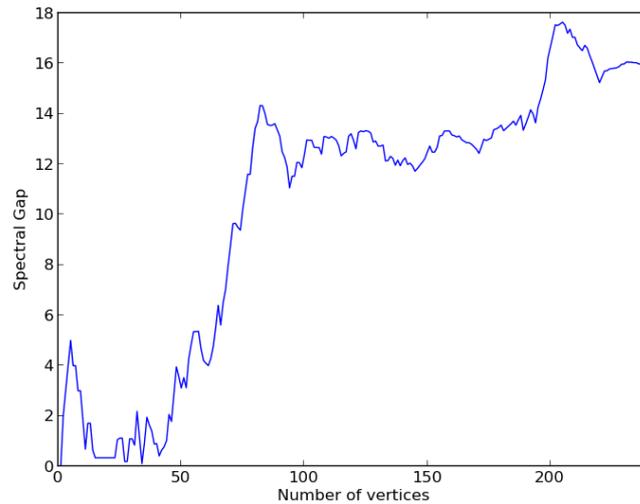


Figure 5.1: The spectral gap for the range  $0 \leq \nu \leq 0.2$ , ignoring at random 30% of the data.

Using only one iteration for the bootstrap method does not tell us much about the stability of the spectral gap. By taking several iterations, a different 30% will be replaced each time. Figure 5.2 shows ten iterations of the bootstrap method in the range  $0 \leq \nu \leq 0.2$ .

In all cases the spectral gap increases just before 100 vertices, and then levels off with only minor fluctuations. From Figure 5.2, we notice that each iteration levels off at a different value. In all cases there are unique fluctuations after they level off and there is a lower bound for the spectral gap. Using the bootstrapping method has shown us that the spectral gap is both stable and gives an accurate result each time.

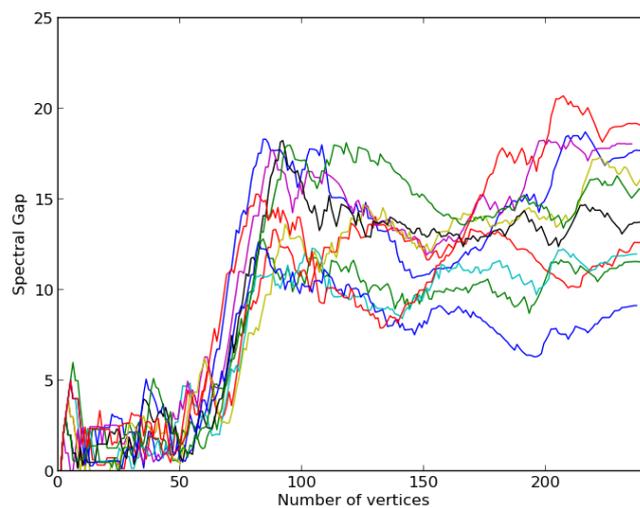


Figure 5.2: Ten iterations for the spectral gap in the range  $0 \leq \nu \leq 0.2$ , ignoring a random 30% of the data.

## 5.2 Varying the starting vertex

In the description of the spectral gap, we said that it is known to be invariant under a relabelling of the vertices. Instead, I suggested relabelling the clusters, so that when they are transformed into vertices they would be connected in a different order.

In all the work done so far, the numbering of the clusters has started at the top and worked its way down, so the starting cluster has always been quite close to the edge of a region, and has only had a few neighbours. By selecting a different cluster as vertex 1, the starting cluster will have a different number of neighbours to which it is connected. Using a cluster at the other end of the region will most likely have around the same amount of clusters neighbouring it. Whereas using a cluster near the centre of the region will have more clusters surrounding it and therefore should have more connections. Starting from near the centre of the domain, where clusters will have more connections, could have an effect on the spectral gap.

The spectral gaps, starting at two different vertices, is shown in Figure 5.3. These can be compared to Figure 4.4, which started at the top of the region.

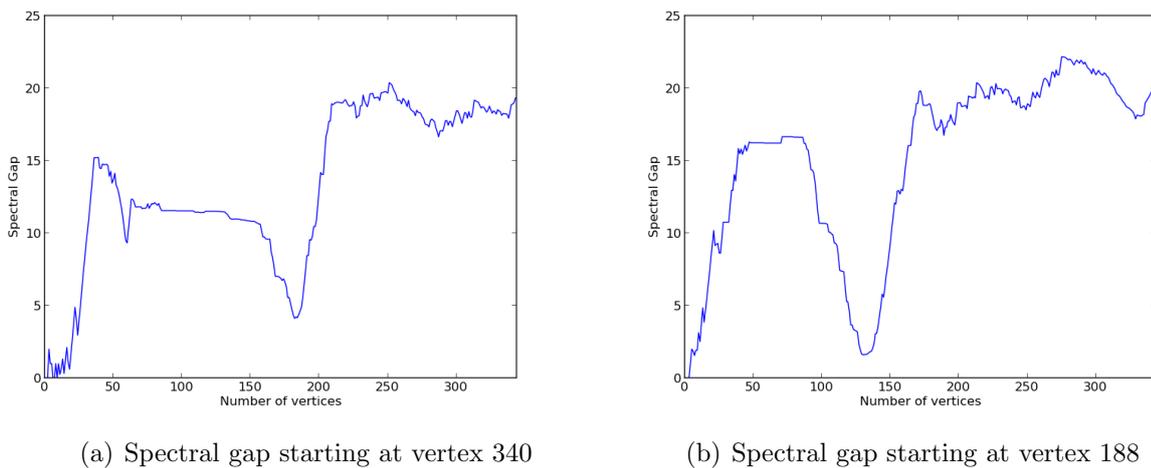


Figure 5.3: Spectral gap, starting at different vertices for the graph in the range  $0 \leq \nu \leq 0.2$ .

The spectral gap in Figure 5.3(a) starts from a vertex near the bottom of the region (vertex 340), and Figure 5.3(b) shows the spectral gap starting from a vertex near the centre of the region (vertex 188).

There are significant differences in the spectral gaps, but they both tend to about the same asymptotic value with only minor fluctuations. The main differences in the graphs happen below 200 vertices, with any major differences happening under 100 vertices. For example, for both end vertices at vertices 0 and 340, there are a few minor perturbations before the initial rise. Whereas for the starting vertex 188, it rises straight away with not too many perturbations. This is a consequence of starting at a vertex with many more connections. It quickly makes the spectral gap rise, but then falls appear when the less connected vertices are included. As all these differences are earlier on in the spectral gap, they can possibly be ignored due to the nature of the spectral gap. As the number of vertices rises, the graphs look like they tend to a similar value and both show it is

bounded from below. It is in line with what we predicted to happen, as now most of the clusters are included in finding the spectral gap.

By taking several versions of the spectral gap and placing them on the same plot, we can clearly see how similar or different these plots are. Figure 5.4 shows the spectral gap starting at 10 different clusters.

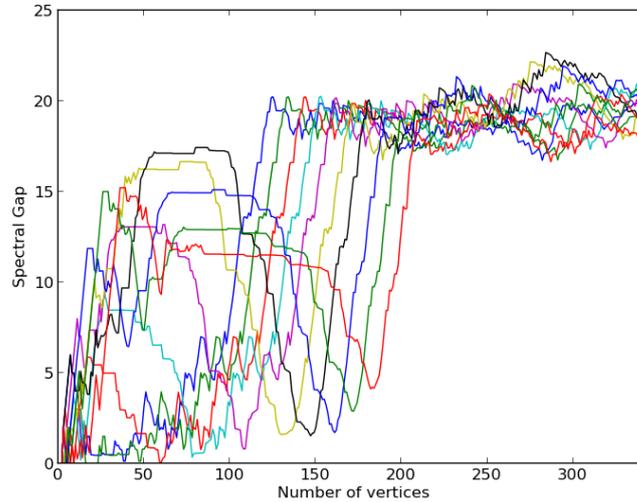


Figure 5.4: Spectral gap for the range  $0 \leq \nu \leq 0.2$  starting at vertices 1, 36, 76, 112, 148, 188, 224, 264, 300 & 340.

From Figure 5.4, we can see that below 100 vertices, there are many differences between the spectral gaps. A few of them rise first before dropping and rising again, whereas others perturbate over a small region before having the main rise. But all of the spectral gaps have the main rise and then level off at about the same value. This happens at a different amount of vertices for each of the plots, the first plot levelling off is the one which begins at cluster 1. Each of the lines then level off according to their beginning vertex, so the second plot that levels off is at vertex 36, the next is at 76 and so on.

Once they all level off, the spectral gaps fluctuate around a similar value. They converge at about the vertex 250, but then fluctuate again over a wider range before converging again at the last vertex point. All ten of the spectral gaps show a good lower bound after levelling off, meaning the graph is both sparse and highly connected no matter which cluster in the region we start from. It would be interesting to see what happens after 350 vertices, to see whether it would continue to converge or fluctuate again, but unfortunately, for this range, we have a limited number of clusters.

In the range  $-0.4 \leq \nu \leq -0.2$ , we have over 700 vertices, Figure 5.5 shows the spectral gaps for different starting vertices for this range of  $\nu$ .

From the figure, it is clear that the plots all differ for few vertices, i.e. under 100 vertices, with some having an immediate rise in the value of the spectral gap, whilst others have a low value until the main rise occurs. All ten iterations have the main rise and then all follow the same pattern, just as before. The first plot to rise comes from a starting vertex at cluster 1, the second from the vertex 85, continuing to rise until the final plot which comes from a starting vertex of 765.

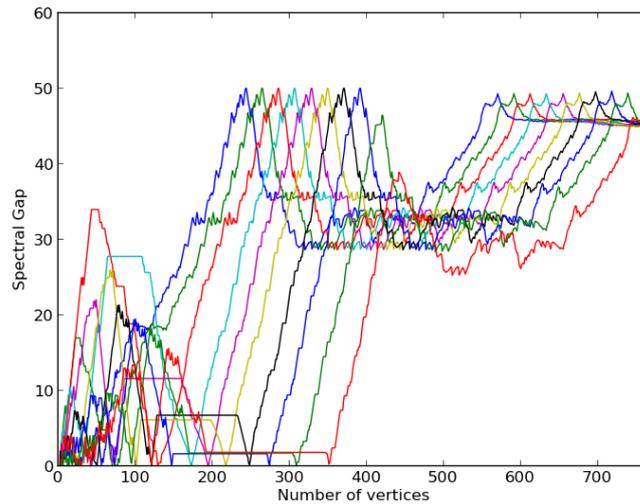


Figure 5.5: Spectral gap for the graph in the range  $-0.4 \leq \nu \leq -0.2$ , starting at the vertices 1, 85, 169, 232, 341, 425, 509, 593, 677 & 765.

All of the iterations for this range level off at about the same value over 400-600 vertices and they all show very similar fluctuations around this level. In the Figure 5.4, for the range  $0 \leq \nu \leq 0.2$ , once the graph levels off, the plots fluctuate randomly not following a particular pattern, whereas in the range  $-0.4 \leq \nu \leq -0.2$ , the plots all have relatively similar fluctuations, they do not bounce randomly like before but they follow the same pattern of fluctuating.

In the Figure 5.4, for  $0 \leq \nu \leq 0.2$ , the plots seem to converge, but then split again. We thought this might be because of the limited amount of vertices. By using this second range between  $-0.4 \leq \nu \leq -0.2$ , there are more vertices and we can see that this allows the plots to almost converge on a single value. The first plot, from vertex 1, settles around 580 vertices and the last looks like it would rest after the last vertex available. Using the large amount of vertices has been advantageous, as it has shown us that changing the starting vertex has not changed any results taken from the spectral gap, and that there might be a final value for which it does eventually converge to.

Using the spectral gap on a set of finite clusters will always have disadvantages. There will always be a limit on the amount of vertices available and therefore we can only predict what the spectral gap would have done next, whether it stays fixed at a certain point or diverges again. But from the graphs produced, whether we change the starting vertex or from the bootstrap technique, we know that changing the local properties will not alter the global ones, therefore the spectral gap is a very stable variable forming the same conclusions no matter the change.

# Chapter 6

## Results and Conclusions

The main result of this work is the translation of random field from the ISM into the world of graph theory. By taking the clusters from the gas density in the warm phase of the simulated interstellar gas and transforming them into vertices of a graph, we can now analyse these complicated regions. Using the spectral gap, we have found that even though there are a few different ways of representing the clusters as vertices, they all give very similar graphs. The spectral gap was also useful in finding the best radial distance to connect up these clusters. The clusters can be transformed into vertices by using the centres of mass, and edges between clusters that are within 400 parsecs of each other. The ideas of graph theory have been used in both 2D and 3D regions.

### 6.1 Graphs in 2D

By transforming the random density field in the simulated ISM into graphs, using the method presented in Chapter 3, we can begin to analyse a 2D slice of the field. Figure 6.1 shows the graph for the range  $0 \leq \nu \leq 1$  in the mid-plane.

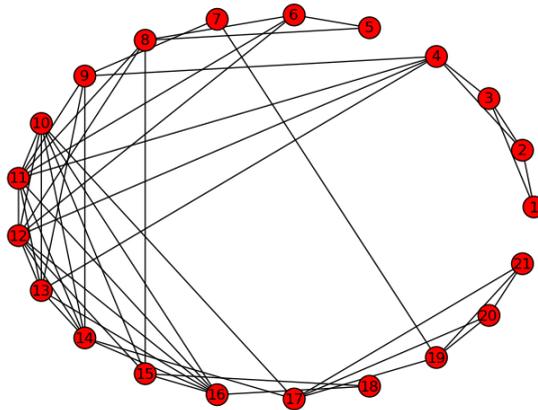


Figure 6.1: Graph for the range  $0 \leq \nu \leq 1$ , vertices from the centre of mass, with vertex 1 at quarter past 3 on a clock face.

The graph has 21 clusters and is connected with a radius of 400 parsecs. There are groups of vertices which are almost all connected to each other, such as the vertices 1 to 4. Figure 6.1 has several of these groups of clusters, and they could be labelled as a single cluster or vertex to see the graphs they would produce.

Transforming the 2D random fields into graphs can also be used to compare different ranges of density. The histogram in Figure 6.2, shows the spread of the density values over a range of  $\nu$  in the 2D plane at the the mid-plane. Most of the data is near the centre of the range, so two ranges to compare will be  $-1 \leq \nu \leq 0$  and  $0 \leq \nu \leq 1$ .

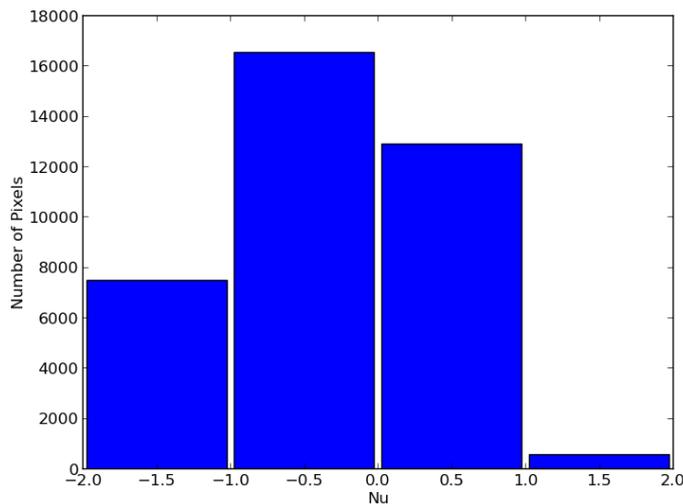


Figure 6.2: Histogram of  $\nu$  in the  $x$ - $y$  plane, at the mid-plane.

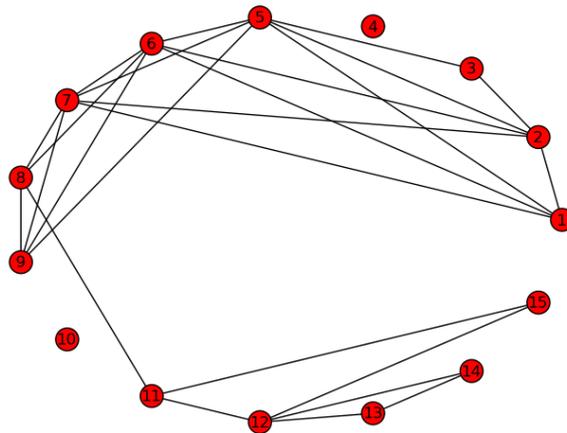


Figure 6.3: Graph for the range  $-1 \leq \nu \leq 0$ , in the  $x$ - $y$  plane at the mid-plane.

Figure 6.3 shows the graph for the range  $-1 \leq \nu \leq 0$ . There are some very obvious differences between the graphs in Figures 6.1 and 6.3. The graph in Figure 6.3 has less edges than the first. Consequently the graph for the second region is more sparse,

implying that it is not as well connected as the first. In Figure 6.3, many clusters will be near one another to create groups of clusters and isolated clusters will form further away.

The histogram in Figure 6.2 suggests that there are more data points in the range  $-1 \leq \nu \leq 0$ , but the graph has fewer vertices. This indicates that the density data in the first range is more uniformly spread over the region, whereas the points in the second range are grouped together more. However, there are exceptions, such as the  $x$ - $y$  slice at  $z = -100$  parsecs.

The graphs for 2D slices can also be used to compare the clusters at various positions along the  $z$ -axis. The graphs for the range  $0 \leq \nu \leq 1$  at  $z = 100$  parsecs and  $z = -100$  parsecs are shown in Figure 6.4.

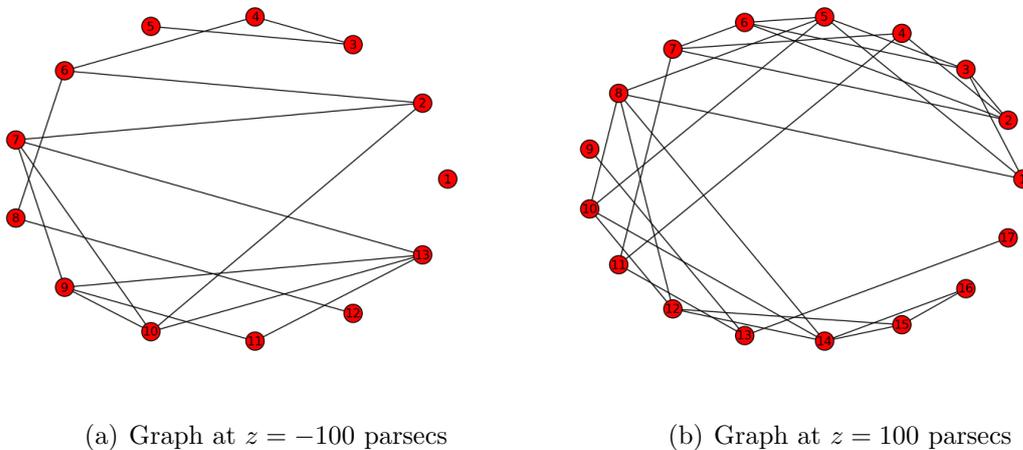


Figure 6.4: Graphs in the range  $0 \leq \nu \leq 1$ , in the  $x$ - $y$  plane at different values of  $z$  with vertex 1 at quarter past the hour on a clock.

Comparing them in this way, some significant differences are noticed immediately. Even though the two graphs are at the same distance from the mid-plane, the two graphs are very different. The graph at  $z = 100$  parsecs has more vertices and more edges. This means that the data in this slice has more clusters spread over the region, but they are close to one another to create as many connections as it has. The graph at  $z = -100$  parsecs has fewer edges, allowing vertex 1 to be isolated. This does not necessarily mean that there is less density in this range, but the data points are more bunched together. Therefore, there are less clusters formed although they are larger in size. These clusters are also more dispersed over the region so there are fewer connections between them. If a graph has more vertices, this does not necessarily mean that there are more density points. In fact, the slice at  $z = -100$  parsecs has more points, but forms less clusters.

## 6.2 Graphs in 3D

As mentioned in Chapter 4, the graphs for the 3D regions contain so many connections that it would be impractical to display the whole graphs. As a result of this, we came to consider the eigenvalues and the spectral gap. The spectral gap is usually used on

graphs where every vertex has the same degree, also known as a *regular* graph. However, it can be used on irregular graphs to find the sparsity and provide an upper bound on the degree of vertices.

Although the spectral gap is usually used on graphs with an infinite number of vertices, so that more vertices can always be added, our graphs only have a finite number of vertices. Figure 6.5 shows the spectral gap for two different ranges of  $\nu$ . The first range  $-0.4 \leq \nu \leq -0.2$ , shown in Figure 6.5(a), contains 765 vertices and the second range  $0.2 \leq \nu \leq 0.4$ , in Figure 6.5(b), contains 297 vertices.

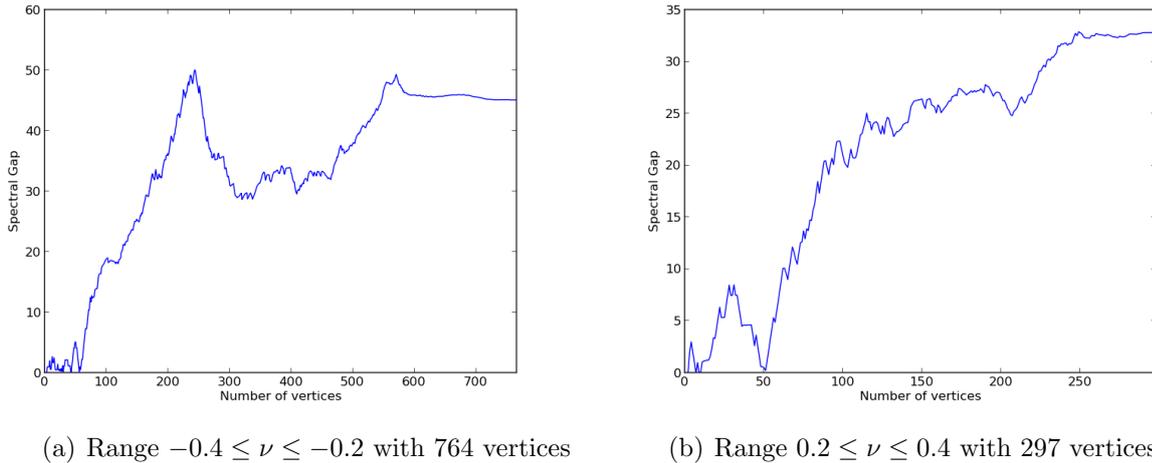


Figure 6.5: Spectral gap for two ranges of  $\nu$ , with a varying number of vertices.

In the first range the spectral gap levels off and fluctuates around an asymptotic value, whilst in the second range it seems to rise steadily after the initial main rise, although it does eventually seem to level off at a large number of vertices. The first range also has weaker fluctuations towards the end of the curve, almost settling on a single value. Comparing Figure 6.5(a) to Figure 4.4 also shows that there are less fluctuations towards the end of the plot. If there are more vertices in a graph, it appears that the spectral gap tends to a single value. But, as shown for graphs in 2D, if there are more vertices in a graph this does not mean the spectral gap has a better defined single value. Comparing the two plots in Figures 6.5, Figure 6.5(a) shows a clearer spectral gap with an obvious lower bound than the second. However, when comparing this plot with the one for the original range  $0 \leq \nu \leq 0.2$ , this is not the case.

As a result of this, we can say that the actual values of the spectral gap do not tell us very much, but the plot of the spectral gap is related to how well connected the graph is. So, for a plot of the spectral gap with a clear lower bound and a steady main trend, the graph will be well connected. If the graph shows a clear lower bound, but still fluctuates over a large range of values, the graph will be well connected, but will not have all the properties desired.

After using the spectral gap to confirm that the graph is both highly connected and sparse, the eigenvalues or the spectrum of the graph can then be used to analyse it further. For a regular graph, where the degree of every vertex is the same, the first eigenvalue would be equal to that degree. But as our graphs do not carry that property, the first two eigenvalues can be used to place a bound on the average degree of the vertices.

Table 6.1: The first two eigenvalues,  $\lambda_1$  and  $\lambda_2$ , and average degree of the vertices in different ranges of  $\nu$ .

Range of $\nu$	$\lambda_1$	$\lambda_2$	Average Degree
$-0.4 \leq \nu \leq -0.2$	170.2	125.08	130.07
$-0.2 \leq \nu \leq 0$	108.21	88.28	90.26
$0 \leq \nu \leq 0.2$	87.7	67.79	75.29
$0.2 \leq \nu \leq 0.4$	95.98	63.14	80.22

From Table 6.1, we can see that the average degree of the vertices always lies between the first two eigenvalues. Table 6.1 only considers four ranges of density. To confirm that these eigenvalues can be used as the upper and lower bounds, we checked that this holds for all the ranges between  $-1 \leq \nu \leq 1$ , of length 0.2. From the first two eigenvalues, we can determine the average number of clusters which will be connected to any other cluster in any region.

Another use of the eigenvalues, along with the adjacency matrix, is to find the Laplacian matrix. We can form the matrix  $D$  as the matrix of vertex degrees, where the diagonal of the matrix contains the degree of each vertex, and the rest of the values are zero. Then the Laplacian matrix is given by  $L = D - A$ , hence

$$L_{ij} = \begin{cases} \text{degree}(v_{ij}), & \text{if } i = j, \\ -1, & \text{if } i \neq j \text{ and vertex } v_i \text{ is adjacent to } v_j, \\ 0, & \text{otherwise.} \end{cases} \quad (6.1)$$

The Laplacian matrix is symmetric along the diagonal and the elements in the rows and columns sum to zero. The smallest non-zero eigenvalue of  $L$  is equal the spectral gap of  $A$ , which was found earlier, and the number of eigenvalues that equal zero represents the number of components, or vertices, in the graph. The Laplacian matrix is the discrete analog of the Laplacian operator in multivariate calculus and can embed the graph onto a plane.

### 6.3 Conclusion

Using the graph theory as proposed in this paper is a great way to start analysing a random field in a new and exciting way. The methods used to transform the structures into a graph has been shown to be a valuable and effective way of simplifying the structure whilst still capturing most of the important information. The eigenvalues, from the adjacency matrix, and the spectral gap have been shown to be sensitive characteristics of a random field and are used to quantify the relative graph.

The statistical properties of the spectral gap have been explored, along with its fluctuations. By using several random variations of the underlying random field, and the way the graph is constructed, the spectral gap has been shown to be a stable and useful property to use on these graphs.

To improve how these vertices are formed, we can take the shape of each cluster into

consideration by moving into a 5-dimensional space, using the filamentary and planarity as variables. The last two co-ordinates will show whether the cluster is long and thin or uniform in shape, to make sure that all aspects of the cluster are addressed during the transformation.

The next step for using graph theory is to consider the appropriate graphs as a topological space. By looking at the graphs as a set of vertices,  $V$ , and edges,  $E$ , the graph can be embedded onto a topological surface. We can consider the use of graph embeddings to prove structural results and allow a new way to interpret the graphs in physical terms.

# Bibliography

- [1] E. Vanmarcke *Random Fields, Analysis and Synthesis*, World Scientific Publishing, pg 1-3 (2010)
- [2] <http://www-ssg.sr.unh.edu/ism/what1.html>, retrieved on 19/03/15
- [3] S. Jha, J. Ching *Simulating spatial averages of stationary random fields using the Fourier series method*, J. Eng. Mech., pg 594 - 605 (2013)
- [4] A. Lang, J. Potthoff, *Fast simulation of Gaussian random fields* (2013) at <http://arxiv.org/pdf/1105.2737.pdf>
- [5] P. Welch, *The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms*, Transactions on Audio and Electroacoustics, Vol Au-15, No 2, pg. 1 (1967)
- [6] F. Gent, *Supernovae Driven Turbulence in the Interstellar medium*, PHD thesis, Newcastle University (2012)
- [7] J. Taylor, *Random Fields*, Lecture notes, Stanford University, pg. 1-45 (2009)
- [8] [http://en.wikipedia.org/wiki/Correlationfunction\(astronomy\)](http://en.wikipedia.org/wiki/Correlationfunction(astronomy)), retrieved on 31/03/15
- [9] K. Ruohonen, *Graph Theory*, Lecture notes, University of Turku (2013)
- [10] H. Ueda, M. Itoh, *A graph-theoretical approach for quantifying galaxy distributions in a cold dark matter universe*, The Astrophysical Journal, pg. 560-561 (1999)
- [11] <http://cs.stackexchange.com/questions/7690/why-we-do-isomorphism-automorphism-and-homomorphism>, retrieved on 09/04/15
- [12] <http://stackoverflow.com/questions/12599143/what-is-the-distinction-between-sparse-and-dense-graphs>, retrieved on 09/04/15
- [13] F.R.K Chung, *Spectral graph theory*. American mathematical Society, pg. 2 (1997)