
Peer reviewed version

Link to published version (if available):
10.1093/comnet/cnx022

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via [insert publisher name] at [insert hyperlink] . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
We study symmetric motifs in random geometric graphs. Symmetric motifs are subsets of nodes which have the same adjacencies. These subgraphs are particularly prevalent in random geometric graphs and appear in the Laplacian and adjacency spectrum as sharp, distinct peaks, a feature often found in real-world networks. We look at the probabilities of their appearance and compare these across parameter space and dimension. We then use the Chen-Stein method to derive the minimum separation distance in random geometric graphs which we apply to study symmetric motifs in both the intensive and thermodynamic limits. In the thermodynamic limit the probability that the closest nodes are symmetric approaches one, whilst in the intensive limit this probability depends upon the dimension.

Keywords: Random geometric graph, spectrum, motif, Chen-Stein method.

2010 Math Subject Classification: 90B15, 47A10, 05C50, 05C80

1. Introduction

Many physical systems like social networks, biological networks, transport networks and technological infrastructures can be modelled using the graph concept of a set of nodes connected by edges. For an introduction see for example [23, 24]. In the study of complex networks, a popular technique is to randomly generate graphs in such a way as to capture certain features of the topology and dynamics of the real-life systems of interest. Some common examples of these are the Erdos-Renyi random graph [11], the Barabási-Albert scale-free network generator [4] and the Watts-Strogatz small-world network generator [34]. A powerful tool for analysing the topology and dynamics of these networks is their graph spectra [7] [15] [21]. In particular, important subgraphs called symmetric motifs can be seen through the presence of sharp peaks in the spectra indicative of the multiplicity of particular eigenvalues, see figure 1 for an illustration of this in the real-world network of the high-voltage power grid in the Western States of the United States of America [34] and see [19] for an analysis of the symmetric motifs in this particular network. Networks such as this are important for an understanding of efficiency and robustness of power-grids [28]. It is known that the presence of symmetric motifs is important for other real-world complex systems [20] and has been shown to influence synchronisation processes [1], [9] and dynamical stability [3], [10] and is also related to redundancy and network stability [19]. However, the non-spatial random graph generators mentioned above rarely contain these important subgraphs. It is known, however, that they do occur in a spatial random graph model the random geometric graph (RGG). RGGs were first introduced by Gilbert as a way of modelling wireless networks [16]. In a RGG
the nodes are distributed throughout a given domain uniformly at random and are connected by an edge when they are within a given range of each other. See [26] and [31] for introductions and see figure 2 for an illustration of a RGG.

RGGs are often used to model networks in which the node location is an important factor, so called spatial networks (See Ref.[5] for a review). In particular RGGs have been used in modelling wireless networks [27], [17], [29], [14], in the study of epidemics [32], [22], [30], in the study of city development [33], in modelling the vulnerability of infrastructure [35] and biological protein-protein interaction networks [18]. In addition the properties of RGGs such as synchronisation [12],[9], consensus dynamics [13], connectivity [8] and spectral properties of RGGs [25], [6] have all been studied.

Here we study the symmetric motifs in RGGs. Our aim is to understand how properties like density and dimension of the RGG affect their appearance. We consider the probability of finding symmetric nodes as a function of the connection radius and in two high density limit cases. In particular, we analyse the symmetric motifs in the intensive limit of fixed connection radius, and the thermodynamic limit of fixed mean degree. It turns out that the expected number of symmetric motifs is non-trivially dependent on these system parameters.

We will study the binomial model of RGGs. That is, we distribute $N$ points representing $N$ nodes of a network uniformly on the unit torus. A connection is made between two nodes if their toral distance is less than some given range $r$. This is known as the unit disc connection model, but note there are many other connection models in which the links are random with probability depending on the inter-node distance [8].

In section 2 we will look at the adjacency and Laplacian spectrum of some RGG ensembles and discuss in particular the sharp peaks we find in these spectra. We will explain how the symmetric motifs which give rise to these sharp peaks and study the probability of finding them in RGGs. In section 3 we will then use the Chen-Stein method [2] to derive the scaling of the minimum separation distance in RGGs. The Chen-Stein method is a tool for obtaining a bound on the total variation distance between a stochastic process which contains dependent random variables and a corresponding independent process. Here, we apply it to remove the dependence of inter-node distances, and hence obtain the variation of the symmetric motifs with dimension in the above limits. Section 4 contains a summary.
2. Spectrum

We first look at the spectrum of the graph adjacency matrix $A$ and Laplacian matrix $L$. The adjacency matrix is the zero-one adjacency matrix whose entries $a_{ij} = 1$ if there is a connection between nodes $i$ and $j$ and zero otherwise. The Laplacian matrix $L = D - A$ where $D$ has entries $d_{ij} = k_i \delta_{ij}$, $k_i$ the degree of vertex $i$. Again see figure 1 for an illustration of the spectrum of $A$ and $L$ in the real-world network of the high-voltage power grid in the Western States of the United States of America [34].

In [25] the Laplacian spectra of one-dimensional RGGs is studied. The authors show that the ensemble-averaged Laplacian spectrum of RGGs on the circle consist of a continuous part and a discrete part consisting of peaks at integer values. We numerically obtained the Laplacian spectrum for an ensemble of RGGs. The ensemble-averaged spectral density $\rho_l(\lambda)$ is illustrated in figure 3. In [6] the spectra of the adjacency matrix is studied for RGGs. They find that the ensemble averaged spectral density has a discrete part consisting of a peak at $-1$. We numerically obtained the adjacency spectrum of an ensemble of RGGs. This is illustrated in figure 4.

In both [25] and [6] they identify the presence of symmetric nodes as the structural phenomenon which gives rise to the multiplicities in the eigenvalues which characterise the spectral densities. These Motifs or graph orbits are subgraphs whose nodes are invariant under permutation of the indices. That is two nodes are symmetric in this sense when they are connected to the same set of nodes. Eigenvectors localise on these symmetric nodes and give rise to integer eigenvalues. To see this let $n_1$ and $n_2$ be symmetric nodes and $x$ be a vector with $x_1 = 1$ and $x_2 = -1$, all other entries equal to zero. Considering the adjacency matrix we then have

![Random geometric graph on the torus. An illustration of a 400 node RGG with a connection range of $r = 0.112$. The nodes (black dots) have been distributed randomly, uniformly and are connected by an edge (straight black line) when they are within a Euclidean distance of $r$ with periodic boundary conditions.](image)
FIG. 3. Ensemble-averaged Laplacian spectral density. An illustration of the ensemble-averaged spectral density of $10^3$ node RGGs with a connection range of $r = 0.09375$ (a) and $r = 0.3$ (b). The ensembles consist of $10^4$ RGGs. Inset in (b) shows detail of peaks at integer values.

FIG. 4. Ensemble-averaged adjacency spectral density. An illustration of the ensemble-averaged spectral density of $10^3$ node RGGs with a connection range of $r = 0.09375$ (a) and $r = 0.3$ (b). The ensembles consist of $10^4$ RGGs.

\[
\begin{pmatrix}
0 & 1 & \cdots \\
1 & 0 & \cdots \\
1 & 1 \\
0 & 0 \\
\vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
1 \\
-1 \\
0 \\
0 \\
\vdots
\end{pmatrix} =
\begin{pmatrix}
1 \\
-1 \\
0 \\
0 \\
\vdots
\end{pmatrix}
\] (2.1)

if $n_1$ and $n_2$ are connected (called a Type-I symmetry). If they are not connected we get eigenvalue 0 (a Type-II symmetry). For the Laplacian we get

\[
\begin{pmatrix}
k & -1 & \cdots \\
-1 & k & \cdots \\
-1 & -1 \\
0 & 0 \\
\vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
1 \\
-1 \\
0 \\
0 \\
\vdots
\end{pmatrix} = (k+1)
\begin{pmatrix}
1 \\
-1 \\
0 \\
0 \\
\vdots
\end{pmatrix}
\] (2.2)
if \( n_1 \) and \( n_2 \) are connected. If they are not connected we get eigenvalue \( k \). Here \( k \) is the degree of the vertices \( n_i \). We note that the Laplacian distinguishes between different degrees in these motifs whilst the adjacency does not. Note that for a set of \( s \) symmetric nodes, there are \( s-1 \) independent, orthogonal eigenvectors that lead to a multiplicity of \( s-1 \) for the eigenvalue.

### 2.1 Symmetry probability

The symmetry of a node is dependent upon the nodes in its neighbourhood and their respective neighbourhoods. The neighbourhood of a node \( n_i \) is defined as \( B_r(n_i) = \{ x : |x-n_i| \leq r \} \), that is the region within the connection range of \( n_i \). Two nodes have a shared neighbourhood \( \mathcal{N}(n_i,n_j) = B_r(n_i) \cap B_r(n_j) \) which is given by the intersection of their respective neighbourhoods. An additional important concept is the excluded neighbourhood, \( \mathcal{N}_e(n_i,n_j) = (B_r(n_i) \cup B_r(n_j)) \backslash (B_r(n_i) \cap B_r(n_j)) \) which for two vertices is the region that is within the range of one of the vertices but not both.

A given node \( n_i \) is type-I symmetric when there is a node \( n_j \) within its neighbourhood and when the excluded neighbourhood of \( n_i \) and \( n_j \) is empty. This ensures that \( n_i \) and \( n_j \) are connected and that they are connected to the same set of nodes. If they are not within the neighbourhood of each other but have an empty excluded neighbourhood then they are type-II symmetric.

We calculated the probabilities of finding type-I and type-II symmetric nodes from ensembles of RGGs. The results are illustrated in figure 5. For one-dimensional RGGs the probability of type-I symmetric nodes quickly approaches a constant value as a function of connection radius whilst type-II quickly approaches zero. For two and three-dimensional RGGs we see entirely different behaviour. Most interestingly there is an optimal value of \( r \) for finding type-I symmetric nodes.

### 3. Dimension and limiting probabilities

In order to look closer at the differences across dimension, we have looked at the adjacency spectral density of both 2D and 3D RGGs with similar connection probability. The results are illustrated in figure 6 where we see a noticeable difference in that for the 3D RGGs the peak at \(-1\) is far less prominent.

To understand this property we look at the probability that nearest neighbour vertices are symmetric (these are the most likely to be symmetric). We will analyse the minimum separation distance between nodes in a RGG and derive how it scales as the number of nodes increases.

We consider \( N \) nodes \( n_1, n_2, \ldots, n_N \), uniformly distributed on \( d \)-dimensional torus and choose an index set \( I \) which consists of all pairs of nodes \( I = \{ \alpha \subset \{1,2,\ldots,N\} : |\alpha| = 2 \} \). For each \( \alpha = \{i,j\} \) in the index set \( I \) we let \( X_\alpha \) be the indicator random variable of event \(|n_i-n_j| \leq x\). That is, the event that the nodes \( n_i, n_j \) are separated by a distance less than \( x \). Note that \( P(X_\alpha = 1) = c_D x^D = p_\alpha \) where \( c_D \) is the volume of the unit ball in \( D \) dimensions. Furthermore, for each \( \alpha \in I \) we choose a \( B_\alpha \subset I \) with \( \alpha \in B_\alpha \) such that \( B_\alpha \) is a neighbourhood of dependence. That is for \( \beta \in B_\alpha \), \( X_\alpha \) and \( X_\beta \) are dependent. For each \( \alpha \) we let

\[
B_\alpha = \{ \beta \in I : \alpha \cap \beta \neq \emptyset \}.
\]

Consider the sum

\[
W = \sum_{\alpha \in I} X_\alpha,
\]

and note that

\[
P(W = 0) = P(s_{\text{min}} > x).
\]
Fig. 5. Symmetry probability. Illustrated here is the probability $P_{sym}$ of a node being symmetric as calculated from ensembles of $10^3$ node RGGs as a function of connection radius $r$. We have calculated these as a function of dimension $D$ and intensity (number of nodes) $N$.

Fig. 6. 3D RGG. Illustrated here is the adjacency spectral density of an ensemble of 2D (left) and 3D RGGs. The RGGs consist of 1000 nodes whilst the connection radius is 0.348569 and 0.45 for the 2D and 3D graphs respectively. This ensures that the connection probability is the same for both.

where $s_{min}$ is the minimum distance of any pair of nodes in the RGG. The expectation of $W$, is given by

$$
E(W) = \binom{N}{2} c_D x^D
$$

We now apply the Chen-Stein method described in the Introduction to remove the dependency of the
inter-node distances. From Theorem 1 in Ref.[2] we have that
\[ |P(W = 0) - e^{-w}| \leq (b_1 + b_2 + b_3) \left( \frac{1 - e^{-w}}{w} \right) \] (3.5)

where \( w = \mathbb{E}(W) \) and \( b_1, b_2, b_3 \) are constants given by
\[ b_1 = \sum_{\alpha \in I} \sum_{\beta \in B} p_{\alpha \beta} \] (3.6)
\[ b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B} p_{\alpha \beta}, \quad p_{\alpha \beta} = \mathbb{E}(X_{\alpha}X_{\beta}). \] (3.7)
\[ b_3 = \sum_{\alpha \in I} \mathbb{E}[\mathbb{E}(X_{\alpha} - p_{\alpha}|\sigma(X_{\beta}: \beta \notin B_{\alpha})] \] (3.8)

From Eq.(3.1) we have
\[ b_1 = |I||B_{\alpha}|(c_{Dx}D)^2 \]
\[ = \left( \frac{N}{2} \right) \left( \left( \frac{N}{2} \right) - \left( \frac{N - 2}{2} \right) \right) (c_{Dx}D)^2 \]
\[ = \frac{N^3}{2} - \frac{3N^2}{4} + \frac{3N}{2} (c_{Dx}D)^2. \] (3.9)

Considering \( b_2 \) we note that the \( X_{\alpha} \) are pairwise independent, this implies \( \mathbb{E}(X_{\alpha}X_{\beta}) = \mathbb{E}(X_{\alpha})\mathbb{E}(X_{\beta}) = (c_{Dx}D)^2 \), therefore
\[ b_2 = |I||B_{\alpha} - 1|(c_{Dx}D)^2 \]
\[ = \left( \frac{N}{2} \right) \left( \left( \frac{N}{2} \right) - \left( \frac{N - 2}{2} \right) - 1 \right) (c_{Dx}D)^2 \]
\[ = \frac{N^3}{2} - \frac{3N^2}{4} + \frac{3N}{2} (c_{Dx}D)^2. \] (3.10)

Finally we note that \( X_{\alpha} \) is independent of all \( X_{\beta} \) where \( \beta \notin B_{\alpha} \), hence \( b_3 = 0 \). Combining Eqs.(3.9,3.10) via Eq.(3.5) we have the following
\[ |P(W = 0) - e^{-w}| \leq (4N - 7)c_{Dx}D \left( 1 - e^{-\frac{N^2}{2} - \frac{3N}{2} (c_{Dx}D)^2} \right). \] (3.11)

We note that from Eq.(3.4) we have that \( e^{-w} \) is of order one when \( N^2x^D \sim 1 \), that is when \( x \sim N^{-\frac{2}{D}} \). Furthermore in this regime from Eq.(3.11)
\[ |P(W = 0) - e^{-w}| \to 0, \quad N \to \infty. \] (3.12)

Hence we see that \( P(W = 0) \) is order one when \( x \sim N^{-\frac{2}{D}} \). From this we derive that \( s_{\text{min}} \) scales like \( N^{-\frac{2}{D}} \) for large \( N \).

For \( D = 1 \), two points separated by a distance \( s \) have an excluded neighbourhood of length \( 2s \). The probability that they are type-I symmetric with \( s = s_{\text{min}} = C_1N^{-2} \) for constant \( C_1 \) is then
\[ \mathbb{P}(N(\mathcal{N}_{\alpha}) = 0) = \left( 1 - \frac{2C_1}{N^2} \right)^{N-2} \]
\[ \to 1, \quad N \to \infty. \] (3.13)
So the closest nodes will be type-I symmetric in this limit. We obtain the same result in the thermodynamic limit, where the mean degree is kept constant with \( r = CN^{-\frac{1}{2}} \).

For \( D = 2 \), two points separated by a distance \( s \) have an excluded neighbourhood equal to two times the area of a circle with radius \( r \), minus four times the area of the circular segment with height \( r - s/2 \), that is

\[
||N_{ex}|| = 2\pi r^2 - 4r^2 \cos^{-1}\left(\frac{s}{2r}\right) + 2s\sqrt{r^2 - \frac{s^2}{4}}
\]

\[
= 4r^2 \sin^{-1}\left(\frac{s}{2r}\right) + 2s\sqrt{r^2 - \frac{s^2}{4}}
\]

(3.15)

Using \( s = s_{\text{min}} = C_2 N^{-1} \) for some constant \( C_2 \) the probability of type-I symmetry in two dimensions is then

\[
P(N(N_{ex}) = 0) = \left(1 - 4r^2 \sin^{-1}\left(\frac{C_2}{2rN}\right) - \frac{2C_2}{rN} \sqrt{r^2 - \frac{C_2^2}{4N^2}}\right)^{N-2}
\]

\[
= \left(1 - \frac{2C_2}{rN} \left(r + \sqrt{r^2 - \frac{C_2^2}{4N^2}}\right) - O(N^{-3})\right)^{N-2}
\]

(3.16)

\[
\rightarrow e^{-4C_2 r}, N \rightarrow \infty \quad (3.17)
\]

where we use \( \sin^{-1}(x) = x + O(x^3) \). In the thermodynamic limit with \( r = CN^{-\frac{1}{2}} \)

\[
P(N(N_{ex}) = 0) = \left(1 - \frac{4C_2^2}{N} \sin^{-1}\left(\frac{C_2}{2CN^{0.5}}\right) - \frac{2C_2}{N} \sqrt{\frac{C_2^2}{N^2} - \frac{C_2^2}{4N^2}}\right)^{N-2}
\]

\[
= \left(1 - \frac{2C_2C}{N^{1.5}} \left(1 + \sqrt{1 - \frac{C_2^2}{4C_2^2 N^2}}\right) - O\left(N^{-3/2}\right)\right)^{N-2}
\]

(3.18)

\[
\rightarrow 1, N \rightarrow \infty ,
\]

this probability goes to one.

With \( D = 3 \), two points separated by a distance \( s \) have an excluded neighbourhood equal to two times the volume of a sphere minus four times the volume of the spherical cap with height \( r - s/2 \), that is

\[
||N_{ex}|| = \frac{8\pi r^3}{3} - \frac{4\pi}{3} \left(r - \frac{s}{2}\right)^2 \left(2r + \frac{s}{2}\right)
\]

\[
= 2\pi sr^2 - \frac{\pi s^3}{6}
\]

(3.19)

Using \( s = s_{\text{min}} = C_3 N^{-\frac{3}{2}} \) for constant \( C_3 \),

\[
||N_{ex}|| = 2\pi r^2 C_3 N^{-\frac{3}{2}} - \frac{\pi}{6} C_3 N^{-2}
\]

(3.20)
Hence the probability of seeing no nodes in this excluded neighbourhood is

\[
P(N(A_{ex}) = 0) = \left( 1 - 2\pi r^2 C_3 N^{-\frac{3}{2}} + \frac{\pi}{6} C_3 N^{-2} \right) N^{-2} \rightarrow 0 \quad N \to \infty.
\]

(3.21)

In the thermodynamic limit with \( r = CN^{-\frac{1}{3}} \)

\[
P(N(A_{ex}) = 0) = \left( 1 - 2\pi C^2 C_3 N^{-\frac{4}{3}} + \frac{\pi}{6} C_3 N^{-2} \right) N^{-2} \rightarrow 1 \quad N \to \infty.
\]

(3.22)

The difference we see is that in one and two dimensions the probability that points separated by the minimal distance are symmetric approaches a constant value in the intensive limit of large \( N \) whilst in three dimensions this probability goes to zero. In the thermodynamic limit this probability goes to one.

4. Summary

We have looked at the appearance of symmetric motifs in random geometric graphs. These subgraphs are of interest as they are prevalent in many real world networks and random geometric graphs but not in other random graph models. We looked at how the probability of finding symmetric nodes is dependent on the connection radius and the dimension of the random geometric graph. We found that in one dimensional random geometric graphs this probability is close to being independent of the connection radius and the density in that it quickly approaches a constant value as \( r \) is increased. In two and three dimensions in contrast we found that there is a value of \( r \) at which the probability attains a maximum value. In the thermodynamic limit we found that the closest nodes will be symmetric almost surely, irrespective of the dimension. Whilst in the intensive limit, in three dimensions this probability goes to zero.

Future work in this direction will look to analytically understand the numerical results we presented here on the probability of finding symmetric nodes as a function of connection radius. In addition, it will be possible to study how generalisations of the hard disc connection function affect the appearance of symmetric nodes.

Acknowledgment

This work was supported by the EPSRC grant number EP/N002458/1 for the project Spatially Embedded Networks. The authors are grateful to an anonymous referee for correcting a few of the equations in Section 3.

REFERENCES

population density. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical
d and Engineering Sciences*, 367 (1901), 3321–3329.

Graph. in *The Ninth International Symposium on Operations Research and Its Applications*, Lecture Notes in

440–442.

in *2011 IEEE International Conference on Communications Workshops (ICC)*, pp. 1–6.