

Random Rectangular Networks:

Theory and Applications

Matthew Sheerin

Department of Mathematics and Statistics

University of Strathclyde, Glasgow

July 2, 2018

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Acknowledgements

I would first like to thank my PhD supervisor Professor Ernesto Estrada for the enormous amount of help and support he has provided. The MM409 class on network theory in the final year of my undergrad was delivered by Ernesto, as well as Phil Knight, and it was such an enjoyable class that I became interested in trying to do a PhD studying networks. Ernesto had a very interesting idea, and was happy to give me the opportunity to explore it. He has since contributed more ideas and his vast knowledge, and has helped me develop as both a researcher and a person. I will always be grateful for everything he has done for me, and the experiences I have enjoyed along the way.

I would also like to thank EPSRC and the Weir Group for providing the funding that made my PhD possible. Thanks also to E. Santiago for providing the datasets of rock fracture networks that are a key part of this work.

Finally, many thanks to my mum Mary, sister Sarah, brother Chris, and fiancée Lorna for being there to support me, especially during the more stressful moments. Without them, completing this journey would have been much more difficult.

Abstract

The world is full of complex systems, with many interconnected parts interacting in some way, and the whole system cannot be understood simply by looking at each part individually. Instead, it is necessary to consider the system as a whole, which may exhibit emergent behaviour as a result of the many interconnections. Network theory is a very powerful tool for analysing complex systems, and has been applied to a wide range of phenomena with great success.

This work is concerned with spatial networks, which are the ones that are naturally embedded in physical space in some way. For example, wireless sensor networks in a geographical region such as a city where the flow of information is essential, or plant populations in a crop field where it is important to understand and try to limit the spread of diseases. Another example is the rocks found deep underground in the Gulf of Mexico that are highly fractured; these fractures can clearly be thought of as spatially embedded networks through which fluids such as oil and gas can flow. There is great commercial interest in efficiently extracting the oil and gas from the rocks, and there are also serious efforts to use the depleted rocks as a means of carbon sequestration to help combat the problem of greenhouse gases. Therefore, it is clearly important to understand the nature of the structure and dynamical properties of these real-world networks.

It is intuitive that the topological and dynamical properties of spatial networks depend on the shape of the space in which they are embedded. In this work we discuss the generalisation of two spatially-defined random graph models to consider nodes located in a unit rectangle. We generalise the random geometric graph (RGG) to the random rectangular graph (RRG), and the relative neighbourhood graph (RNG) to the

Abstract

rectangular relative neighbourhood graph (RRNG). We found an analytic expression for the expected value of the average node degree of RRGs, as well as useful bounds for the diameter, the average path length, and the algebraic connectivity, and approximations to the degree distribution, connectivity, and clustering coefficient. For the RRNGs we found an approximation to the average node degree and a bound on the diameter and algebraic connectivity.

Using this generalisation, we examine the behaviour of diffusion in RRGs and find that increasing the elongation causes diffusive particles to spread more slowly. We also discuss some results relating to epidemics in crop fields, where we show that elongating the field makes it more difficult for a disease to become epidemic. Finally, we find that the relative neighbourhood graphs work well to mimic the properties of the rock fracture networks compared to other null models, and we are able to optimise the value of the elongation in the model for each fracture network.

Publications and Presentations

The following papers were published in the course of the work contained in this thesis.

- E Estrada and M Sheerin. Random rectangular graphs. *Phys. Rev. E*, 91:042805, Apr 2015.
- E Estrada and M Sheerin. Consensus dynamics on random rectangular graphs. *Physica D: Nonlinear Phenomena*, 323-324:20–26, 2016.
- E Estrada, S Meloni, M Sheerin, and Y Moreno. Epidemic spreading in random rectangular networks. *Phys. Rev. E*, 94:052316, 2016.
- E Estrada and M Sheerin. Random neighborhood graphs as models of fracture networks on rocks: Structural and dynamical analysis. *Applied Mathematics and Computation*, 314:360–379, 2017.

Some parts of this work have also been presented at the following

- E Estrada and M Sheerin. Random Rectangular Graphs. Poster session presented at: *International School and Conference on Network Science (NETSCI) 2015*.
- M Sheerin. Random Rectangular Graphs. Applied Analysis seminar, Mathematics and Statistics Department, University of Strathclyde, 12th November 2015.

Contents

| | |
|---|------------|
| Acknowledgements | ii |
| Abstract | iii |
| Publications and Presentations | v |
| List of Figures | x |
| List of Tables | xi |
| Notation | xii |
| Introduction | 1 |
| I Theory | 6 |
| 1 Networks and Graphs | 7 |
| 1.1 Definitions and Concepts | 7 |
| 1.1.1 The Structure of Graphs | 8 |
| 1.1.2 Matrices | 9 |
| 1.1.3 Spectral Graph Theory | 11 |
| 1.1.4 Global Properties | 12 |
| 1.2 Random Models | 20 |
| 1.2.1 Erdős-Rényi | 20 |
| 1.2.2 Barabási-Albert | 20 |

Contents

| | | |
|----------------------------------|--|-----------|
| 1.2.3 | Random Geometric Graph | 21 |
| 1.2.4 | Proximity Graphs and Relative Neighbourhood Graphs | 25 |
| 1.3 | Dynamics on Networks | 28 |
| 1.3.1 | Diffusion (Consensus) Process | 29 |
| 1.3.2 | Epidemics on Networks | 33 |
| 1.3.3 | The Flow of Fluids | 36 |
| 1.4 | Summary | 37 |
| II Results and Discussion | | 38 |
| 2 | Structure and Connectivity of Random Rectangular Graphs | 39 |
| 2.1 | Definition of RRGs | 40 |
| 2.2 | Topological Properties of RRGs | 41 |
| 2.2.1 | Average Node Degree | 41 |
| 2.2.2 | Degree Distribution | 49 |
| 2.2.3 | Diameter and Average Path Length | 51 |
| 2.2.4 | Clustering Coefficient | 55 |
| 2.2.5 | Algebraic Connectivity | 60 |
| 2.2.6 | Spectral Radius | 60 |
| 2.2.7 | The λ_{-1} Eigenvalue | 62 |
| 2.2.8 | Connectivity | 69 |
| 3 | Structure of Rectangular Relative Neighbourhood Graphs | 75 |
| 3.1 | Definition of RRNGs | 75 |
| 3.2 | Topological Properties of RRNGs | 77 |
| 3.2.1 | Average Node Degree | 77 |
| 3.2.2 | Diameter and Algebraic Connectivity | 81 |
| 4 | Applications | 85 |
| 4.1 | Diffusion (Consensus) in RRGs | 85 |
| 4.2 | Epidemics in RRGs | 89 |

Contents

| | | |
|----------|---|------------|
| 4.3 | Rock Fracture Networks | 93 |
| 4.3.1 | Description of Fracture Networks | 93 |
| 4.3.2 | Similarity between Fracture Networks and RRNGs | 95 |
| 4.3.3 | Fluid Diffusion on RFNs and RRNGs | 101 |
| 4.3.4 | Theoretical analysis | 102 |
| 5 | Conclusions | 108 |
| 5.1 | Future Directions | 111 |
| 5.1.1 | Topological properties of RRGs and RRNGs | 111 |
| 5.1.2 | Epidemics in crop fields | 112 |
| 5.1.3 | Influence of fracture aperture in RFNs | 112 |
| 5.1.4 | Influence of long-range hops of diffusive particles | 114 |
| 5.1.5 | Influence of 3D environments | 117 |
| | Appendix | 120 |
| | A Small Subgraphs | 120 |
| | Bibliography | 121 |

List of Figures

| | | |
|------|--|----|
| 1 | Comparison of differently shaped cities | 3 |
| 1.1 | Example RGG | 22 |
| 1.2 | β -skeleton examples | 27 |
| 1.3 | Small graph to demonstrate dynamics. | 29 |
| 1.4 | Example of diffusion process | 32 |
| 2.1 | Illustration of an RRG | 40 |
| 2.2 | Derivation of average node degree: quarter circles | 43 |
| 2.3 | Average node degree of RRGs | 45 |
| 2.4 | Derivative of average node degree of RRGs | 46 |
| 2.5 | Degree distributions of RRGs | 51 |
| 2.6 | Diameter and average path length of RRGs | 53 |
| 2.7 | Clustering coefficient vs elongation of RRGs | 58 |
| 2.8 | Clustering coefficient vs connection radius of RRGs | 59 |
| 2.9 | Algebraic connectivity vs elongation of RRGs | 61 |
| 2.10 | Spectral radius vs average node degree in RRGs | 62 |
| 2.11 | Chain structure contributing to the multiplicity of $\lambda = -1$ | 65 |
| 2.12 | Hexagon structure in RRNGs | 67 |
| 2.13 | Comparison of m_{-1} and \tilde{G}_{-1} in different random graphs | 68 |
| 2.14 | Symmetric nodes and neighbourhood chains in an RRG | 69 |
| 2.15 | An extension to the concept of neighbourhood chains | 70 |
| 2.16 | Connectivity of RRGs | 74 |

List of Figures

| | | |
|------|--|-----|
| 3.1 | Example RRNGs | 76 |
| 3.2 | Construction of RRNG for average node degree proof | 78 |
| 3.3 | Average node degree of RRNGs | 80 |
| 3.4 | Example of edge violating RRNG average node degree assumptions | 81 |
| 3.5 | Properties of RRNGs vs elongation | 84 |
| 4.1 | Consensus dynamics in RRGs | 87 |
| 4.2 | Contour plots of diffusion dynamics in RRGs | 88 |
| 4.3 | Stationary states of epidemics | 90 |
| 4.4 | Comparison of theoretical and simulated epidemic threshold | 91 |
| 4.5 | Illustration of extraction of RFN from rock image | 94 |
| 4.6 | Similarity between RRNG model and RFNs | 99 |
| 4.7 | Comparison of different random graph models to RFNs | 100 |
| 4.8 | Illustration of different diffusion scenarios | 102 |
| 4.9 | Comparisons of RFN diffusion and RRNG properties | 106 |
| 4.10 | Correlation of fragments 1, 3, and 4 | 107 |
| 5.1 | Diffusion in weighted vs unweighted RFNs | 114 |
| 5.2 | Consensus time in RFNs with long-range hops | 117 |
| 5.3 | 3D RNGs | 118 |
| 5.4 | Comparison of 2D and 3D RNGs | 119 |
| A.1 | Illustration of small subgraphs | 120 |

List of Tables

| | | |
|-----|--|-----|
| 1 | Notation (English) | xiv |
| 2 | Notation (Greek) | xiv |
| 4.1 | RFN data | 94 |
| 4.2 | Network properties | 96 |
| 4.3 | Regression analysis of diffusion in RFNs | 103 |

Notation

| | |
|---------------|--|
| A | Adjacency matrix, general matrix |
| A_{ij} | (i, j) th entry of matrix A |
| $A(x)$ | Area within a distance r of point x lying inside \mathcal{R} |
| a | Rectangle elongation |
| $B(p, r)$ | Open ball at point p of radius r |
| BA | Barabási–Albert graph |
| b | Rectangle height |
| $b_s(G)$ | Spectral bipartivity |
| C_n | Cycle graph of length n |
| $ C_3 $ | Number of triangles |
| $ C_{3,i} $ | Number of triangles incident to node v_i |
| \bar{C}_d | Asymptotic clustering coefficient of d -dimensional RGG |
| C_i | Clustering coefficient of node v_i |
| $\bar{C}(G)$ | Average (Watts–Strogatz) clustering coefficient |
| $C(G)$ | Global (Newman) clustering coefficient |
| $CS(G)$ | Collatz–Sinogowitz index |
| D | Distance matrix $D_{ij} = d(v_i, v_j)$ |
| d | Dimension |
| $d(v_i, v_j)$ | Graph metric |
| $d_{max}(G)$ | Diameter |
| E | Set of edges |
| \mathbf{E} | Expectation |
| $EE(G)$ | Estrada index |
| EE_i | Subgraph centrality of node v_i |
| EE_i^{odd} | Odd subgraph centrality of node v_i |
| EE_i^{even} | Even subgraph centrality of node v_i |
| ER | Erdős–Rényi graph |
| $f(a, b, r)$ | Function related to average node degree of RRG |
| $F(G)$ | Free energy |
| G | Graph $G(V, E)$ |
| G' | A subgraph of G |

Notation

| | |
|--------------------------|---|
| G_{ij} | Communicability between v_i and v_j |
| \tilde{G}_{-1} | Measures concentration of spectrum around -1 |
| GG | Gabriel Graph |
| I_n, I | Identity matrix |
| $Kf(G)$ | Kirchhoff index (resistance distance) |
| K_n | Complete graph |
| K | Diagonal matrix of node degrees |
| k_i | Degree of node v_i |
| $\bar{k}(G)$ | Average node degree |
| $k_{max}(G), k_{min}(G)$ | Maximum, minimum node degree |
| \mathcal{L} | Laplacian matrix |
| $\hat{\mathcal{L}}$ | Normalised Laplacian matrix |
| $\bar{l}(G)$ | Average path length |
| m | Number of edges |
| m_{-1} | Multiplicity of λ_{-1} |
| $N'(v_i)$ | (Closed) neighbourhood of node v_i |
| n | Number of nodes |
| P | Probability |
| P_{fc} | Probability of being connected |
| P_n | Path graph of length n |
| $ P_{n,i} $ | Number of path graphs of length n centred on node v_i |
| $ P_n $ | Number of path graphs of length n |
| $p = \min(a, r)$ | Notation useful for average node degree integrals |
| $p(k)$ | Degree distribution |
| $q = \min(b, r)$ | Notation useful for average node degree integrals |
| R | Region of influence in construction of β -skeleton |
| \mathcal{R} | Rectangle |
| RFN | Rock Fracture Network |
| RGG | Random Geometric Graph |
| RNG | Relative Neighbourhood Graph |
| RRG | Random Rectangular Graph |
| RRNG | Rectangular Relative Neighbourhood Graph |
| r | Connection radius of RRG |
| $S(G)$ | Entropy |
| SIS | Susceptible-Infected-Susceptible model |
| SIR | Susceptible-Infected-Recovered model |
| t | Time |
| t_c | Time for consensus |
| t_c^- | Approaching time for consensus |
| $\langle t_c \rangle$ | Average time for consensus |
| $\mathbf{u}(t)$ | State vector for dynamical process |

Notation

| | |
|--------------------|--|
| \mathbf{u}_i | i th element of \mathbf{u} |
| \mathbf{u}_0 | Initial condition of state vector \mathbf{u} (for example) |
| $\mathbf{u}_{0,i}$ | i th element of \mathbf{u}_0 |
| v_i | i th node (for example) |
| WSN | Wireless Sensor Network |

Table 1: Notation (English)

| | |
|--------------------|--|
| β | Parameter of β -skeleton model, infection rate of epidemics |
| Δ | Spectral gap |
| δ | Tolerance for consensus process to be completed |
| $\delta(G)$ | Density of a graph |
| ϵ | Step size for discrete-time model |
| $\varepsilon(v_i)$ | Eccentricity of node v_i |
| θ_{ij} | Communicability angle between v_i and v_j |
| $\bar{\theta}(G)$ | Average communicability angle |
| Λ | Diagonal matrix of eigenvalues of adjacency matrix A |
| λ_i | i th eigenvalue of adjacency matrix A |
| λ_{-1} | Eigenvalue equal to -1 |
| M | Diagonal matrix of eigenvalues of Laplacian matrix \mathcal{L} |
| μ | Recovery rate of epidemics |
| μ_i | i th eigenvalue of Laplacian matrix \mathcal{L} |
| ξ_{ij} | Communicability distance between v_i and v_j |
| $\bar{\xi}(G)$ | Average communicability distance |
| $\rho(t)$ | Fraction of infected nodes at time t |
| $\rho(A)$ | Spectral radius |
| τ | Epidemic threshold |
| Φ | Matrix of eigenvectors ϕ_i of adjacency matrix A as columns |
| ϕ_i | i th eigenvector of adjacency matrix A |
| $\phi_{i,j}$ | j th element of i th eigenvector of adjacency matrix A (for example) |
| Ψ | Matrix of eigenvectors ψ_i of Laplacian matrix \mathcal{L} as columns |
| ψ_i | i th eigenvector of Laplacian matrix \mathcal{L} |

Table 2: Notation (Greek)

Introduction

Network theory has received a lot of attention in recent years, though it is difficult to say precisely when it was established. An early example of the mathematics that would eventually develop into network theory is the origin of graph theory in 1736 when Euler tackled the famous bridges of Königsberg problem. He needed a mathematical tool that captured how certain discrete objects were linked to each other in order to understand the topology of the problem and answer certain questions, and that tool was graph theory. In order to say if it is possible to use each bridge connecting the four landmasses exactly once, the exact geographical layout is not important, but rather which pieces of land are connected to which other pieces of land, and by how many bridges. By considering the problem in this way, it is easier to tackle the problem, and Euler managed to show that such a traversal, now known as an *Eulerian trail*, was not possible. It is amusing that the origin of network theory involved ignoring spatial information and only considering how the different areas are connected, but in this Thesis the geometric setting of the networks is the central idea.

One can think of a large number of examples of situations where there are many individuals or components that are connected or interacting in some way, and the overall structure of the connections is fundamental to understanding the system. Network theory is the ideal tool to analyse these systems [118, 48]. For example, consider people in a social network, neurons in the brain, proteins in a cell, webpages on the Internet, or road intersections, and the links as friendships, neural connections, interactions within a cell, hyperlinks, or roads respectively. Networks can capture the structure of these systems and provide a very useful framework for analysing their properties, which is why network theory is now being put to use on a wide range of real-world problems

Introduction

across the various sciences with a lot of success [33, 12, 120]. Graph theory alone is not enough to fully understand these systems, so network theory also relies on topics such as statistical mechanics, geometry, topology, and dynamical systems.

Many real-world networks are naturally embedded in space and are therefore inherently geometrical, so-called *spatial networks* [13]. These include infrastructural networks such as road networks [160] and airport transportation networks [106, 158, 27], certain biological networks such as brain networks or the networks representing the proximity of cells in a biological tissue [90, 48], networks of patches and corridors in a landscape [150], the networks of galleries in animal nests [130, 24], wireless sensor networks (WSNs) and peer-to-peer networks [74, 62, 132, 101, 67, 42, 11, 128], geographical populations where epidemics can spread [91, 42, 148, 114, 96, 76], and the networks of fractures in rocks [104, 1, 7, 89, 25, 83, 153], among others.

This Thesis is a step forward in the analysis of real spatial networks. It is clear that spatial networks come in a variety of shapes, and the shape of the space in which the network is embedded should have a significant impact on the properties of the network. Previous work on spatial networks has given little consideration to these shapes however. For example, the random geometric graph model is defined in the unit hypercube [127, 69], and so in a 2D application it is assumed that the system being modelled is square-shaped. However, in general this assumption is not a good one. For example, Fig. 1 shows two cities: San Francisco is approximately an 11 km by 11 km square, whereas Manhattan is approximately a 21.6 km by 3.7 km rectangle. Then, in one case the assumption of having the network embedded in a square region is reasonable, and in the other it is clearly far from accurate. Therefore, instead of assuming all spatially embedded networks are square, we introduce a generalisation of the models used to consider rectangles to account for networks that are more elongated, which is a useful step in making models that more accurately reflect reality.

We consider two different types of geometrically defined random graph models, and generalise the models to distribute the nodes in rectangles instead of just squares. The random geometric graphs (RGGs) [127] are generalised to the random rectangular graphs (RRGs) and the relative neighbourhood graphs (RNGs) to the rectangular rel-

Introduction

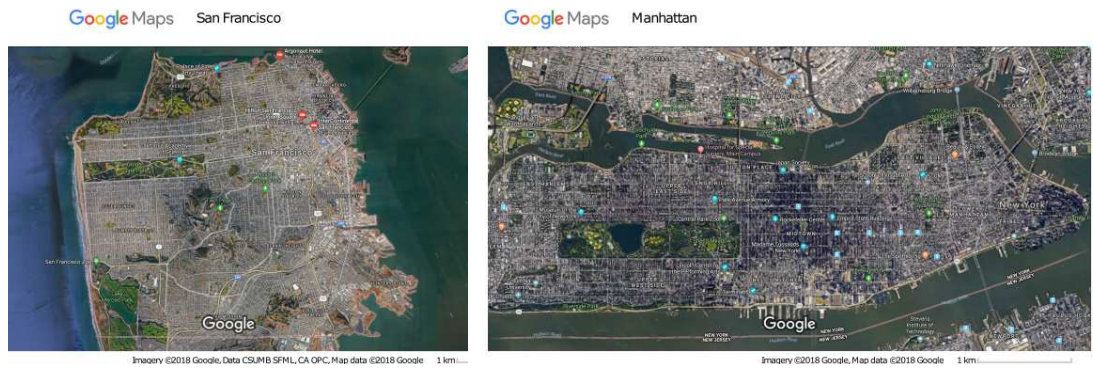


Figure 1: San Francisco (left) is approximately square but Manhattan (right) is approximately rectangular. Images from Google Maps.

ative neighbourhood graphs (RRNGs), where the RNG model is a special case of the more general β -skeleton graphs [149, 95]. We discuss different topological properties of these models, then we consider their use in three different applications.

First, we discuss diffusion in RRGs. We use a simple heat equation, which we may also consider as a consensus process [110]. The interpretation as a consensus process is used in the study of WSNs, for example, where the RGG model has been applied [101]. A WSN consists of sensors (nodes) distributed in some location and they are each collecting information such as air temperature. They need to be able to communicate this wirelessly through forming a network with each other, such that information from every node can be extracted efficiently. However, if we consider WSNs in the cities of San Francisco and Manhattan as above, we find that the square domain of the RGG could be a good approximation to the true shape of the city, but for Manhattan the generalisation to the RRG model is important since it more closely resembles an elongated rectangle. In this section we consider the behaviour of diffusion in RRGs in general, with the understanding that the results are applicable to various real-world scenarios [101, 42, 91, 74, 114].

The second application we consider is that of epidemiological models on networks [17, 115, 116, 129, 159]. It is predicted that climate change will have a significant impact on crops, including susceptibility to diseases [145, 84]. Therefore, it is very important to develop better crop-growth models to gain more understanding of such problems in order to take effective action. There has been recent progress on the analy-

Introduction

sis of plant ecosystems using network theory, partially due to more data being available [136, 70, 137]. When studying the propagation of diseases in plants there is an important factor that needs to be taken into account. It is obvious that plants are not as mobile as humans and animals, thus they reach lower levels of mixing in a given population. The immediate consequence of this lack of mobility is that the shape of the plot or field in which the plants are distributed may significantly affect disease dynamics [126, 157, 93, 65, 21, 113, 112, 164, 63], with square fields being more prone to the spread of diseases than elongated ones of the same area [126, 157, 93, 65]. We give some discussion of using RRGs to model this phenomenon, and we find that in simulations the elongation of the rectangle does indeed increase the epidemic threshold. We also give some theoretical arguments to further justify this conclusion, including an analytical result bounding the epidemic threshold $\frac{1}{\lambda_1}$.

Finally, we consider oil reservoirs containing rocks that have been fractured by geological stress. These fractures in the rocks contain fluids of petrochemical interest, and can be considered as a network through which the fluids can flow. Then, understanding the flow of these fluids requires understanding the structure of these networks. These fractures have long been studied in hydrocarbon geology and hydrogeology due to the role that these fractures play on the evaluation of potential oil reservoirs [1, 7, 89, 25, 83, 153]. The analysis of rock fracture networks (RFNs) plays a fundamental role in determining the nature and disposition of heterogeneities appearing in petroliferous formations to determine the capability for the transport of fluid through them [18, 75, 77, 162, 94]. In many of the analyses described in the literature the use of synthetic fracture networks facilitates the analysis due to the sometimes scarce availability of real-world data [14, 121, 36, 163, 117, 143, 68]. In contrast, Santiago et al. have published a series of papers [138, 139, 140] in which they used real fracture networks derived from original hand-sampled images of rocks extracted from a Gulf of Mexico oil reservoir. These works have used a graph-theoretic analysis of these real-world networks in order to extract information about the topological (static) characteristics of this set of rocks. These analyses are relevant even after the oil has been extracted, since once an oil reservoir has been depleted it can be used to store

Introduction

other fluids, such as carbon dioxide [86, 147, 85, 35, 142]. Rock fractures have also been studied in a more general sense for their applications to both oil reservoirs and other fluid flows within rocks such as groundwater, examining properties such as fractal scaling and anomalous diffusion [15, 20, 43, 155, 88, 87].

A discrete version of the diffusion equation is used to study the diffusion of a fluid through the channels of the real RFNs. Although the Boussinesq equation is the natural way to model diffusion on porous media, it has been found that the simple diffusion equation performs very well compared to the experimental results for modelling oil diffusion on rocks. The reason for that resides in the high pressure and temperature at which the flow occurs deep below the Earth's surface. The diffusion through the real RFNs is compared to diffusion on the synthetic models, showing that the RRNGs reproduce not only the most important structural properties of the real networks but also their diffusive properties. As in the series of papers by [138, 139, 140], two-dimensional cuts of rocks that show a fracture network embedded into the rock sides are considered. Then, a potential criticism to these works is the fact that rock networks are three-dimensional [6, 105, 103], and that inferring these 3D networks from 2D information is hard. However, as has been previously documented, the analysis of 2D rock fractures identifies important parameters that allow the characterisation of real rock samples [92, 141, 138, 139, 140]. In addition, note that the generalised proximity graphs that are introduced in this work can be easily extended to the 3D case. Thus, 3D rock fracture networks extrapolating the topological information that is obtained here from the analysis of 2D samples can be easily generated.

The first chapter of this Thesis discusses existing theory and general concepts in network theory, the notation that will be used throughout this work, and the random graph models that are relevant here. The second chapter contains our theoretical results for the RRG model, and the third chapter similarly contains our theoretical results for the RRNG model. The fourth chapter covers the three applications we have analysed: diffusion on RRGs, epidemics on RRGs, and RFNs. The fifth chapter is where we give our conclusions and discuss future work. Finally, we have one appendix followed by the list of references used in this work.

Part I

Theory

Chapter 1

Networks and Graphs

1.1 Definitions and Concepts

The study of networks is related to graph theory, and so we will sometimes use the terms graph and network interchangeably. However, network theory relies not only on graph theory, but also makes use of topics such as statistical mechanics, geometry, topology, and dynamical systems. A graph is the pair of sets $G(V, E)$, where V is the set of nodes and $E \subseteq V \times V$ is the set of edges or links. That is, E consists of those pairs (v_i, v_j) for which there is a link from v_i to v_j . The sizes of these sets are denoted as $n = |V|$ and $m = |E|$. A useful way to represent the structure of a graph is the *adjacency matrix*. This is an $n \times n$ matrix that is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (1.1)$$

If we were to consider instead a *weighted graph* where each link is assigned the value of some real number, then each entry $A_{ij}, (v_i, v_j) \in E$, is the weight of the link from v_i to v_j . A loop is a link from a node to itself, and multiple edges refers to a pair of nodes with more than one link between them. A graph with no loops or multiple edges is said to be *simple*. If when $(v_i, v_j) \in E$ we also have $(v_j, v_i) \in E$, the graph is said to be *undirected*, otherwise it is *directed*. For the majority of this Thesis and unless stated otherwise we will deal only with simple undirected graphs. Therefore, we

1. Networks and Graphs

present the background theory as it applies to simple undirected graphs and note that some concepts require modification when applied to other types of graphs.

1.1.1 The Structure of Graphs

Here we introduce some basic terminology that we will need to describe further concepts. More in-depth discussions of the basics of network theory can be found in many places, such as [48], and we provide references for the introduction of the various properties where applicable.

- *Adjacent, incident*: We describe the edges connected to a node as being incident to that node. Two nodes are adjacent if there is an edge joining them, and two edges are adjacent if there is a node they are both incident to.
- *Neighbourhood*: The neighbourhood of a node v_i is the set $\{v_j | (v_i, v_j) \in E\}$
- *Node degree*: The degree of a node v_i is the number of nodes to which it is connected, and is written as k_i . The diagonal matrix $K_{ii} = k_i$ is the degree matrix of the graph.
- *Walk, path*: A walk is a sequence of edges $(v_1, v_2), (v_2, v_3), \dots, (v_k, v_{k+1})$, the length of the walk is the number of edges; a path is a walk without repeated nodes. A *closed walk* is one in which the first and last nodes are the same.
- *Connected*: A graph is connected if there is a path between any pair of nodes. If it is not connected, we may consider it as the union of several connected *components*.
- *Subgraph*: A graph G' is a subgraph of a graph G ($G' \subseteq G$) if it is considered as a subset of the nodes and edges of G . This may be G itself or the null graph (containing no nodes), but it is usually somewhere in between.
- *Complete graph*: A complete graph K_n has n nodes that are pairwise connected.
- *Triangle*: A triangle is a subgraph of 3 nodes in a graph that are pairwise connected.

1. Networks and Graphs

- *Bipartite*: A graph is bipartite if the set of nodes V can be split into two disjoint subsets V_1 and V_2 , such that all edges link a node in V_1 to a node in V_2 . That is, there are no edges linking two nodes in the same subset.
- *Regular*: A graph is regular if every node has the same degree.

1.1.2 Matrices

Since it is convenient to store information about networks in matrices, we use some of the mathematics of matrices in order to manipulate the information for our purposes, and so we recall here some properties of matrices. We will be using $n \times n$ real-valued matrices unless stated otherwise.

There is a fundamental link between powers of the adjacency matrix and walks in the graph. For example, $(A^2)_{ij} = \sum_{k=1}^n A_{ik}A_{kj}$ counts the number of nodes v_k such that there is a path $v_i \rightarrow v_k \rightarrow v_j$, which is the number of walks of length 2 that start at node v_i and end at node v_j . Note that $(A^2)_{ii}$ is then the number of closed walks of length 2 starting at node v_i , which is just the node degree k_i . In general, $(A^l)_{ij}$ counts the number of walks of length l from v_i to v_j . Walks, and in particular closed walks, are related to subgraphs and this relationship will be considered in more detail later. The identity matrix $I^{n \times n}$ has $I_{ii} = 1$ and all other entries equal to 0, and we usually denote it simply by I if the size is obvious in context. Note that $A^0 = I$ for any square matrix A .

We will also make use of matrix functions. In general, extending the definition of functions from scalar variables to matrices can be subtle and complicated [82]. In most cases, primary matrix functions are desirable, and these can be defined via the Jordan canonical form, Hermite interpolation, or the Cauchy integral. However, we consider here only functions that can be easily defined as a Taylor series, which is compatible with the other definitions. Given some matrix A , we can write a matrix polynomial

$$p(A) = a_0I + a_1A + \cdots + a_kA^k. \quad (1.2)$$

As with polynomials of scalar variables, it is possible to extend this to an infinite

1. Networks and Graphs

sum that will converge. As a motivating example, the scalar differential equation

$$\frac{dx}{dt} = \alpha x, \quad x(0) = x_0 \quad (1.3)$$

has solution $x(t) = e^{t\alpha} x_0$. This extends naturally to a system of ordinary differential equations

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (1.4)$$

that has a solution that can be written in the same form as the scalar version, $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$, using the matrix exponential. The matrix exponential can be defined as a series in the expected way, which always converges:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}. \quad (1.5)$$

This also works for other well-known functions such as sinh and cosh. It is even possible to extend properties of these functions from the scalar case, for example we have

$$\cosh(A) = \frac{e^A + e^{-A}}{2}, \quad \sinh(A) = \frac{e^A - e^{-A}}{2} \quad (1.6)$$

Some other matrices that will be used within this work are the Laplacian \mathcal{L} and the normalised Laplacian $\hat{\mathcal{L}}$. The motivation for the Laplacian matrix is given in Subsection 1.3.1 later in this chapter, where we begin to discuss dynamical processes on graphs and find that it arises naturally.

$$\mathcal{L}_{ij} = \begin{cases} -1 & \text{if } (v_i, v_j) \in E, \\ k_i & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \quad ; \quad \hat{\mathcal{L}}_{ij} = \begin{cases} -\frac{1}{\sqrt{k_i k_j}} & \text{if } (v_i, v_j) \in E, \\ 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (1.7)$$

We also use the distance matrix D

$$D = [d_{ij}]_{n \times n}, \quad (1.8)$$

1. Networks and Graphs

where the graph metric d_{ij} is the number of edges in the shortest path connecting v_i and v_j .

1.1.3 Spectral Graph Theory

One method for obtaining information about the structure of a network is to use *spectral graph theory*. Given some matrix A , we may consider the equation $A\mathbf{x} = \lambda\mathbf{x}$. The pairs of scalars λ and vectors \mathbf{x} that satisfy this equation are incredibly important, and are known as *eigenvalues* and *eigenvectors* respectively. As discussed previously, there are several matrices that can be associated with a given network, and the eigenvalues and eigenvectors of these matrices can contain important information about the structure of the network.

The adjacency matrix A has eigenvalues denoted by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and corresponding eigenvectors $\phi_1, \phi_2, \dots, \phi_n$. The spectral gap $\Delta = \lambda_1 - \lambda_2$ is quite important, as a high value indicates a lack of structural bottlenecks. We can construct a matrix $\Phi = [\phi_1 \phi_2 \dots \phi_n]$, an orthonormal matrix (meaning $\Phi\Phi^T = \Phi^T\Phi = I$, where Φ^T is the transpose of Φ) containing the eigenvectors, and a diagonal matrix of the eigenvalues $\Lambda_{ii} = \lambda_i$, then we can write the adjacency matrix as its spectral decomposition $A = \Phi\Lambda\Phi^T$. Since the adjacency matrix is real valued and symmetric, all the eigenvalues are real. Since it is also non-negative, there is also an eigenvalue $\lambda_1 = \rho(A)$, the spectral radius, due to Perron's theorem. We will see later that λ_1 is closely related to the dynamics of epidemics on networks.

The Laplacian matrix has eigenvalues $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$ with corresponding eigenvectors $\psi_1, \psi_2, \dots, \psi_n$. The quantity μ_2 is often called the *algebraic connectivity* and is closely connected to the behaviour of dynamic processes taking place on the network [64, 37]. The corresponding eigenvector ψ_2 is known as the *Fiedler vector* and is also important. It is known that the multiplicity of the 0 eigenvalue is equal to the number of connected components of a graph, and so if we have a connected graph we have that $\mu_2 > 0$. The spectral decomposition of the Laplacian is written as $\mathcal{L} = \Psi M \Psi^T$, where $\Psi = [\psi_1 \psi_2 \dots \psi_n]$ is an orthonormal matrix containing the eigenvectors and $M_{ii} = \mu_i$ is a diagonal matrix of the eigenvalues. Since $\mathcal{L}\mathbf{1} = \mathbf{0}$ gives

1. Networks and Graphs

us $\mu_1 = 0$, and since Ψ is orthonormal, we get that $\psi_1 = n^{-1/2}\mathbf{1}$.

It is sometimes helpful to use the spectral decomposition of a matrix instead of the matrix itself. If we want to understand the effect of a function on a matrix, it can be easier to understand its effect in terms of the eigenvalues and eigenvectors if something is known about these. In network theory it is common to be interested in the diagonal elements of a matrix, and it can be shown that

$$\mathrm{tr}(A^p) = \sum_{j=1}^n \lambda_j^p, \quad (1.9)$$

and since we consider matrix functions as powers of the adjacency matrix, we can write

$$\mathrm{tr}(f(A)) = \sum_{j=1}^n f(\lambda_j). \quad (1.10)$$

As an example, we can then deduce that the value of $\mathrm{tr}(e^A)$ will be dominated by the contribution from the largest eigenvalues. If the spectral gap is large, this will be dominated by just the largest eigenvalue: $\mathrm{tr}(e^A) \approx e^{\lambda_1}$.

1.1.4 Global Properties

Many different concepts have been proposed to describe and compare the structural properties of networks. In practice, it is common to use matrices to represent the network in some way, and then these are used to calculate quantities of interest that try to capture some information about the network. Here we describe some of these quantities relating to the global properties of networks that will be used in this work.

We start with some important concepts related to the node degrees. One basic statistic is the average node degree $\bar{k} = \frac{1}{n} \sum_{i=1}^n k_i$. It can be helpful to consider the density $\delta(G) = \bar{k}/(n-1)$ in some situations. We can also consider the degree distribution of a network. We write $p(k)$ for the probability that a node selected uniformly at random will have degree k , and we can visualise this by plotting $p(k)$ against k . The shape of this distribution is useful and can say many things about the network, such as giving hints about how the network was formed or developed, though

1. Networks and Graphs

it does not uniquely determine the full structure of the network [48]. Similarly, we can look at the distribution of node degrees for random graph models, and different models will typically have different distributions.

We can also look at other aspects of the node degrees. Recall that a regular graph is one in which all the nodes have the same degree, so then we might try to quantify the irregularity of a network, such as by considering the variance of the node degrees [146]. The Collatz-Sinogowitz index [156] $CS(G) = \lambda_1 - \bar{k}$ can also be used to capture the regularity of the network, which is because λ_1 is a kind of ‘average-degree’ of the network, and is equal to \bar{k} only when the network is completely regular, and in general $k_{max} \geq \lambda_1 \geq \bar{k}$. Another is the degree heterogeneity [47], $\rho = \sum_{i,j} \hat{\mathcal{L}}_{ij}$, which uses the normalised Laplacian matrix.

The assortativity [119] is a measure of the degree correlations between nodes that are linked. An assortative network is one in which nodes tend to be connected to nodes of a similar degree, in contrast a disassortative network has a tendency for nodes with high degree to link to nodes with a low degree, which can be found in a network with hubs.

Next, we recall the distance matrix D and consider the usual graph metric, that is the distance function d where $d(v_i, v_j)$ is the length of the shortest path from v_i to v_j . In the case of an unweighted graph this is the number of edges in the path and for a weighted graph it is the smallest sum of edge weights in a path from v_i to v_j . This suggests some structural parameters of interest. The *diameter* of a graph is the length of the longest of all the shortest paths,

$$d_{max}(G) = \max_{i,j,i \neq j} d(v_i, v_j), \quad (1.11)$$

which is simply the largest element of D . The *eccentricity* $\varepsilon(v_i)$ of a node v_i is the maximum value of $d(v_i, v_j)$ over all $j \neq i$, and so the diameter can also be considered as the largest eccentricity. We can also define the *average path length*, which is the

1. Networks and Graphs

average of all the shortest path lengths between all pairs of nodes

$$\bar{l}(G) = \frac{1}{n(n-1)} \sum_{i,j,i \neq j} d(v_i, v_j). \quad (1.12)$$

We can gain clues about how a network was formed by looking at the occurrences of small subgraphs. For example, in a social network it is intuitively clear that if person A is friends with both person B and person C, there is a relatively high chance that then person B and person C will also be friends. This results in a lot of triangles in the network and is a characteristic of social networks. An excess of a particular small subgraph is known as a *network motif*. There are 18 small subgraphs that will be considered in this work (see [48]), which are shown in Appendix A.

The occurrences of triangles are commonly of particular interest, and so there are ways to calculate the tendency of the network to form triangles. For a given node, the *clustering coefficient* [161] or *transitivity* of the node is the ratio of the number of triangles incident to that node to the number of paths of length 2 that are centred on that node, $|P_{2,i}|$. This follows from the fact that every pair of edges incident to a node is a potential triangle, since triangles are precisely those instances where the two nodes these edges link to are themselves joined by an edge. The number of triangles incident to a node is calculated as

$$|C_{3,i}| = \frac{(A^3)_{ii}}{2}, \quad (1.13)$$

since the cube of the adjacency matrix captures paths of length three, and its diagonal then counts closed walks of length 3. However, each triangle may be traversed in each of two directions and so we double count, and therefore we have to divide by two to correct for this. The number of paths of length 2 centred on a node v_i is calculated as

$$|P_{2,i}| = \binom{k_i}{2} = \frac{k_i(k_i - 1)}{2}, \quad (1.14)$$

which is simply the number of ways of picking two edges from the k_i edges incident to

1. Networks and Graphs

the node. Then, the clustering coefficient of the node is

$$C_i = \frac{|C_{3,i}|}{|P_{2,i}|} = \frac{|C_{3,i}|}{k_i(k_i - 1)}, \quad (1.15)$$

where we define $C_i = 0$ if $k_i < 2$. There are two ways to calculate the overall clustering of the network. The one that will be used in this work is the average clustering coefficient, also known as the Watts-Strogatz clustering coefficient, which is simply the average value of C_i

$$\bar{C}(G) = \frac{1}{n} \sum_{i=1}^n C_i. \quad (1.16)$$

The other option is the Newman clustering coefficient, which considers the total numbers of triangles $|C_3|$ and paths of length 2, $|P_2|$:

$$C(G) = \frac{3|C_3|}{|P_2|}. \quad (1.17)$$

Perhaps surprisingly, these two clustering coefficients are not necessarily correlated, so a graph may have a high Watts-Strogatz clustering coefficient and a low Newman clustering coefficient, or vice versa.

An idea that was first considered in the context of social networks is centrality, which is the intuition that some nodes are more important or influential than others [22]. Many different ways of characterising the importance of each node have been developed, which are not necessarily correlated with each other. A simple approach is node degree centrality, which considers that nodes with a higher degree are more important. There is certainly some merit to this definition, but it only uses information about nearest neighbours and so it is quite limited in the information that it captures. More sophisticated approaches can be used that also consider nodes that are at a larger distance from the node to determine how influential that node is, thereby capturing more of the structure of the network as a whole. In this work we consider undirected networks where the in-degree equals the out-degree for every node, but in a directed graph such as the world wide web this is not true, and concepts such as the Page-Rank centrality exploit this.

1. Networks and Graphs

One way to obtain centrality measures is to consider weighed sums of the adjacency matrix,

$$f(A) = \sum_{l=0}^{\infty} c_l A^l, \quad (1.18)$$

which corresponds to considering walks in the graph and assigning some weight based on the length of the walks. The coefficients must be chosen carefully so that the series converges, though out of the choices we discuss here only one is not guaranteed to converge, but this exception will converge if it satisfies a particular straightforward condition. All the coefficients should be non-negative, and larger weights should be given to smaller powers of the adjacency matrix and thus to shorter walks. The diagonal entries of A^l count the number of closed walks from each node to itself of length l , and so these values are related to subgraphs of the network. Then, by giving a larger weight to smaller powers of A we are giving more weight to smaller subgraphs. We give the definitions of a few useful subgraph centralities [46, 57]:

$$EE_i = \left(\sum_{l=0}^{\infty} \frac{A^l}{l!} \right)_{ii} = (e^A)_{ii} \quad (1.19)$$

$$EE_i^{odd} = \left(\sum_{l=0}^{\infty} \frac{A^{2l+1}}{(2l+1)!} \right)_{ii} = (\sinh(A))_{ii} \quad (1.20)$$

$$EE_i^{even} = \left(\sum_{l=0}^{\infty} \frac{A^{2l}}{(2l)!} \right)_{ii} = (\cosh(A))_{ii}. \quad (1.21)$$

The quantity $EE(G) = \sum_{j=1}^n EE_j$ is known as the *Estrada index* [46]. The EE centrality measure considers all possible walks in the graph, while EE^{odd} only considers closed walks of odd length and EE^{even} only considers closed walks of even length. Since a bipartite graph has no cycles of odd length, such a graph will have $EE = EE^{even}$. In practice, we may have a graph that is close to being bipartite, which means that we can make it bipartite by removing some small number of edges, and therefore we are interested in quantifying how close the network is to being bipartite. One obvious method for this would be to calculate the smallest set of edges that can be removed to make the graph bipartite, but in general this is very computationally expensive. Instead, we can use the properties of the subgraph centralities. A good way to do this

1. Networks and Graphs

is to consider the difference between the number of closed walks of even length and odd length, and normalise by the total number of closed walks. This gives the so-called *spectral bipartivity* [57]

$$b_s(G) = \frac{\text{tr}(\cosh(A)) - \text{tr}(\sinh(A))}{\text{tr}(\cosh(A)) + \text{tr}(\sinh(A))} = \frac{\text{tr}(\exp(-A))}{\text{tr}(\exp(A))} = \frac{\sum_{j=1}^n \exp(-\lambda_j)}{\sum_{j=1}^n \exp(\lambda_j)}. \quad (1.22)$$

We note that there are other notable centrality measures that we do not find to be useful in this context. There are several commonly used centrality measures based on, for example, node degree, betweenness, and the main eigenvector [22]. Other centrality measures with a similar derivation to the Estrada index include the Katz centrality, which converges when $\alpha < 1/\rho(A)$ [100]

$$K_i = \left[\left(\alpha^0 A^0 + \alpha^1 A^1 + \alpha^2 A^2 + \dots + \alpha^k A^k + \dots \right) \mathbf{1} \right]_i = \left[(I - \alpha A)^{-1} \mathbf{1} \right]_i, \quad (1.23)$$

as well as more recently proposed variations such as the double factorial subgraph centrality [61]

$$\Gamma_i = \left(\sum_{l=0}^{\infty} \frac{A^l}{l!!} \right)_{ii}. \quad (1.24)$$

Since networks are used as models of real-world phenomena, it makes sense to apply concepts from such physical systems to network analysis. While these may superficially not make sense if the network we are considering is not obviously related to the physical systems from which we take these concepts, they can still be used to extract useful information about the network. The concepts can be seen as an analogy, and may be able to be interpreted in terms of some feature of the network being considered. We consider here some analogies taken from statistical physics [54]. The entropy and the free energy are calculated as

$$S(G) = - \sum_{j=1}^n p_j \ln p_j, \quad (1.25)$$

$$F(G) = - \ln EE, \quad (1.26)$$

respectively, where $p_j = \frac{\exp(\lambda_j)}{EE}$. We also consider an analogy from electrical circuits,

1. Networks and Graphs

which was introduced by Kirchhoff when considering the resistance in the circuits and is known as the *Kirchhoff index* or *resistance distance* [102]:

$$Kf(G) = \sum_{i < j} \sum_{k=2}^n \frac{1}{\mu_k} (\psi_{k,i} - \psi_{k,j})^2. \quad (1.27)$$

Another important concept in the study of networks is communicability, which is a measure of how well information can flow around the network. The communicability between two nodes is characterised by the number of walks between them, with a larger weight given to shorter walks. We can define the communicability between v_i and v_j as

$$G_{ij} = \sum_{k=0}^{\infty} c_k (A^k)_{ij}. \quad (1.28)$$

We impose the same restrictions on the coefficients c_k as we did for the subgraph centrality (Eq. 1.18). If we choose $c_k = 1/k!$ we obtain

$$G_{ij} = (e^A)_{ij}. \quad (1.29)$$

Note that subgraph centrality of a node v_i is G_{ii} . We refer to Eq. 1.29 as the communicability function, though of course there are other useful choices of coefficients that result in other communicability functions. In the famous experiment by Milgram, people were instructed to try to send a letter to a certain destination by only passing the letter to someone they knew, to try to discover how many steps separate people in social networks. In each case, either the letter reached its target, or it got lost somewhere. In general, if we are trying to transmit some information through a network we can consider the information transmitted from v_i that reaches v_j , and the ‘disruption’ in the communication is the information that ends up returning to v_i . The quantity G_{ij} represents the information that is successfully transmitted from v_i to v_j , and if the communication is bidirectional, also from v_j to v_i ; then G_{ii} is the disrupted information travelling from v_i and G_{jj} the disrupted information leaving from v_j . We define an index that considers the successfully transmitted information and the disrupted information

1. Networks and Graphs

as

$$\xi'_{ij} := G_{ii} + G_{jj} - 2G_{ij}. \quad (1.30)$$

Minimising the value of ξ'_{ij} corresponds to increasing the information successfully transmitted. It can be shown that

$$\xi_{ij} = \sqrt{\xi'_{ij}} \quad (1.31)$$

is a Euclidean distance, and so ξ_{ij} is called the *communicability distance* [49]. It is useful to define the average communicability distance

$$\bar{\xi}(G) = \frac{\sum_{i \neq j} \xi_{ij}}{n(n-1)}. \quad (1.32)$$

A related concept is the *communicability angle* [55], which is defined as

$$\theta_{ij} = \arccos \left(\frac{G_{ij}}{\sqrt{G_{ii}G_{jj}}} \right). \quad (1.33)$$

Again, it is useful to consider the average value of the communicability angle

$$\bar{\theta}(G) = \frac{\sum_{i \neq j} \theta_{ij}}{n(n-1)}. \quad (1.34)$$

1.2 Random Models

1.2.1 Erdős-Rényi

The idea of random graphs was first described by Erdős and Rényi [45, 134]. If one uses the phrase “*random graph*”, it usually refers to this model, though it is also commonly referred to as the *Erdős-Rényi* (ER) model. There are two parameters in the graph, the number of nodes n and a connection probability p . Each pair of nodes is then connected independently with probability p . Alternatively, some number of edges m is specified and links are added uniformly at random until m edges have been added. It is widely used as a null model because it is randomly-generated - the properties of an ER random graph demonstrate what happens if a network is constructed randomly. Therefore, if a real-world network has properties that differ significantly from this we can conclude that there is some mechanism guiding its formation, and gain clues about what that mechanism might be. This makes it a crucial tool for analysing networks.

This model has been studied in great depth and many of its properties are known. The expected node degree is $\bar{k} = (n - 1)p$, and it has a Poisson degree distribution $p(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}$. The average clustering coefficient is $\bar{C} = p$, and the average path length is $\bar{l} = \frac{\ln n - \gamma}{\ln(pn)} + \frac{1}{2}$. As the value of p is increased, the size S of the largest component grows in a known way, in three stages.

- *Subcritical* $\bar{k} < 1$: all components are small, and the largest is of size $O(\ln n)$.
- *Critical* $\bar{k} = 1$: there is a single largest component of size $O(n^{2/3})$.
- *Supercritical* $\bar{k} > 1$: $\mathbf{P}((f - \epsilon)n < S < (f + \epsilon)n) \rightarrow 1$ as $n \rightarrow \infty$, where $f = f(\bar{k})$ is the positive solution to $e^{-\bar{k}f} = 1 - f$.

1.2.2 Barabási-Albert

Another widely-used type of random graph is the *Barabási-Albert* (BA) model [10]. The main distinguishing feature of the BA model is that it has a power-law degree

1. Networks and Graphs

distribution with exponent -3 , and therefore it displays scale-free properties. That is, its degree distribution extends over several orders of magnitude, and therefore has features at various scales. It is constructed using the idea that many real-world networks, such as webpages on the internet linking to each other or actors working together on movies, grow over time. When new nodes are added to these networks, i.e. a new webpage is created or an actor appears in their first movie, they are more likely to connect to high-degree nodes rather than nodes with a small degree that are more obscure, which is known as *preferential attachment*. A BA random graph is created by taking some small ‘seed’ graph (usually an ER random graph), and adding one node at a time; each time a node is added, it is connected to each existing node with a probability depending on the degree of those nodes. Nodes that already have a high degree are more likely to have new nodes connect to them, thus making their degree even larger. This results in ‘hubs’ that are connected to a large number of other nodes, and peripheral nodes that are poorly connected. When a large number of nodes has been added, the effect of the seed graph becomes quite small and unimportant. Many real-world networks have, or have been claimed to have, scale-free properties, and so it is useful to have a random graph model that produces scale-free graphs in order to analyse how such graphs behave. The degree distribution is of the form $p(k) \approx \alpha k^{-3}$.

The preferential attachment causes high-degree nodes to be more important, and this heterogeneity in the importance of the nodes only grows as more nodes are added. In contrast, in the spatial networks we consider in this Thesis, such as WSNs or crop fields, all of the nodes are established at once and there is neither growth nor preferential attachment. Then, we should not expect such networks to be scale-free with a power-law degree distribution.

1.2.3 Random Geometric Graph

One problem with the above models in certain applications is that they do not consider any spatial information. There are several random graph models currently considered in the literature that are defined geometrically. They consider nodes distributed in some metric space, and the distance between the nodes in a specified metric is used to decide

1. Networks and Graphs

which pairs of nodes are connected by edges and which are not. Such graphs are referred to in general as *proximity graphs* [95]. Typically, the space is some subset of \mathbb{R}^d , the nodes are distributed uniformly at random, and the metric is the Euclidean distance. We first consider the random geometric graph (RGG) model that was introduced by Gilbert [69].

The RGG is defined by distributing uniformly and independently n points in the unit d -dimensional cube $[0, 1]^d$ [127]. Then, two points are connected by an edge if their Euclidean distance is at most r , which is a given fixed number known as the *connection radius*. That is, we create a disk of radius r centred at each node, and every node inside that disk is connected to the central node. An example of an RGG is shown in Fig. 1.1. This disk plays the role of the area of influence of a given node, such as the area of coverage of a mobile or wireless sensor. A variation of the model is sometimes considered where the probability of two nodes being connected is some function of the distance between them. Then, we are simply considering the case where this probability is 1 if the distance between the nodes is less than r and 0 otherwise.

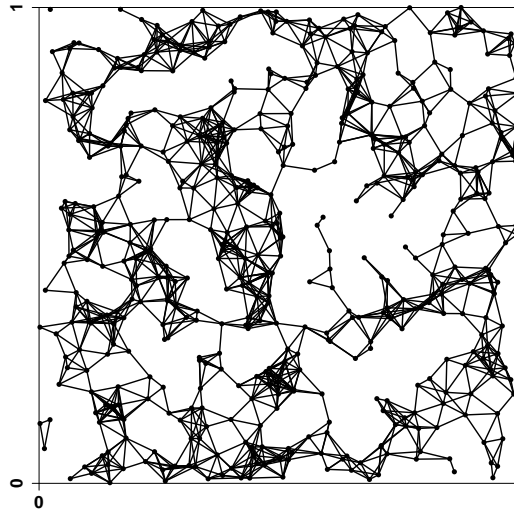


Figure 1.1: Illustration of an RGG created with 500 nodes, the nodes are connected if they are at a Euclidean distance smaller than or equal to $r = 0.07$.

The RGG model has been widely used in the study of wireless sensor networks (WSNs) and peer-to-peer networks [74, 62, 132], where the problem of consensus has

1. Networks and Graphs

received great attention due to the fact that it allows tasks to be achieved with a minimum overhead of communication [101, 67, 42, 11, 128]. In the consensus protocols, as they are known in technological applications, the problem consists of making the scalar states of a set of agents converge to the same value under local communication constraints [122, 110]. Thus, since the communication requires only local information there is no congestion due to network traffic. RGGs are also used to model populations that are geographically constrained in a certain region, like a city. This scenario is important, for instance, for the analysis of epidemic spreading in such populations [91, 42, 148, 114]. In this sense Riley et al. [135] have remarked that RGGs “*provide a nice way of escaping the lack of local correlation and clustering that are implicit properties of the configuration graphs often used to explore epidemic dynamics*”. In a similar fashion, RGGs can be used to model structured populations in which opinions, instead of viruses, are propagated. In this case the RGGs also capture very well the geographic constraints of the population and, in comparison with other models [166], they “*are more realistic for a number of reasons: (i) RGG is isotropic (on average) while regular lattice is not; (ii) the average degree for an RGG can be set to an arbitrary positive number, instead of a small fixed number for the lattice; (iii) RGGs closely capture the topology of random networks of short-range-connected spatially-embedded artificial agents*”. RGGs have even been used to model systems that are less obviously geometrical, such as protein-protein interaction networks [81].

Some work has been done to determine the properties of the RGG model. Taking toroidal boundary conditions for the RGG greatly simplifies the analysis since the area that lies within the connection radius r of each node is always a circle, and so this is frequently done. The expected average node degree is very easy to compute in this regime as

$$\mathbf{E}(\bar{k}) = (n - 1)\pi r^2 \approx n\pi r^2. \quad (1.35)$$

In the case of closed (non-toroidal) boundary conditions, Eq. 1.35 is approximately true if r is very small, otherwise the boundary starts to have a significant effect.

The RGG model has been tackled analytically mostly in terms of some infinite limit, and in the case of the general hypercube. By letting the number of nodes tend

1. Networks and Graphs

to infinity, it is possible to find the smallest value of the connection radius such that in the limit there is some nonzero fraction of the nodes in the largest cluster. Considering toroidal boundary conditions, the analysis is simplified to find a simple expression for the node degree and clustering coefficient, and obtain this *percolation threshold*. This concept of continuum percolation has been used to explain the stretched exponential relaxation in certain systems such as polymers and glasses [97]. Consideration of the limit as $n \rightarrow \infty$ has also been shown to give analytical results on sharp thresholds where a certain property of the RGG appears, such as connectivity [72].

The average clustering coefficient of a RGG has been obtained by Dall and Christensen [34] for $r^2 = \frac{\log n + \alpha}{n\pi}$ when $n \rightarrow \infty$ and $\alpha \rightarrow \infty$, where $\alpha \in \mathbb{R}$ is a constant for a given number of nodes:

$$\bar{C}_d = \begin{cases} 1 - H_d(1) & d \text{ even} \\ \frac{3}{2} - H_d(1/2) & d \text{ odd,} \end{cases} \quad (1.36)$$

where d is the dimension of the hypercube in which the nodes are embedded and

$$H_d(x) = \frac{1}{\sqrt{\pi}} \sum_{i=x}^{d/2} \frac{\Gamma(i)}{\Gamma(i + \frac{1}{2})} \left(\frac{3}{4}\right)^{i + \frac{1}{2}}, \quad (1.37)$$

where $\Gamma(i)$ is the Gamma function. Thus, for $d = 2$, $\bar{C}_2 = 1 - \frac{3\sqrt{3}}{4\pi} \approx 0.5865$ and for $d = 1$, $\bar{C}_1 = 3/4 = 0.75$.

In contrast, there has been relatively little work considering finite RGGs with boundary in 2D. There has been some work considering the connectivity of such RGGs [31, 30], that is, under what conditions the graphs are likely to be connected. This is important because when considering some dynamical process taking place on a graph, nodes can have no interaction with other nodes when there are no paths between them, and it is therefore helpful to be in a regime where there is a single component. The idea is that the boundary will cause a heterogeneity in the node degrees such that nodes that are near the boundary, especially near a vertex of the domain, will have a lower degree since they are surrounded by fewer nodes. Then, to a first-order approximation,

1. Networks and Graphs

the cause of an almost-connected RGG to be disconnected is an isolated node that is likely to be found near a vertex. This has also been explored in fractal regions [40] and convex right prisms [32], which are notable for considering more interesting shapes than squares or circles.

Some consideration has also been given to the RGG in the unit ball [44], and hyperbolic space [125]. Furthermore, as a result of our introduction of the RRG model, a recent paper considers spectral and eigenfunction properties of RRGs [4], and there has been separate work by Estrada on the synchronisability of RRGs [52].

1.2.4 Proximity Graphs and Relative Neighbourhood Graphs

The ‘classical’ β -skeleton graphs $G_\beta(V)$ are a family of spatially defined graphs constructed from a point set V with parameter $\beta \geq 0$, and are described by [95, 149]. For $\beta \geq 1$ the lune-based definition of the β -skeleton model is used and described here as it is more suitable for our needs, an alternative is the circle-based definition that we do not examine here. Though we mostly consider the 2D case, the construction detailed below also works in higher dimensions. We first discuss the construction then demonstrate examples of β -skeleton graphs in Fig. 1.2.

Let p_i and p_j be two arbitrary points that are separated by a Euclidean distance L , and let $B(p, r)$ denote the open ball located at point p with radius r . Two circles, $B((1 - \frac{\beta}{2})p_i + \frac{\beta}{2}p_j, \frac{\beta}{2}L)$ and $B((1 - \frac{\beta}{2})p_j + \frac{\beta}{2}p_i, \frac{\beta}{2}L)$ are constructed, and let R be the intersection of the circles. It is obvious that the area of R increases as β is increased. Although it was previously stated that we are only considering the lune-based definition for $\beta \geq 1$, the case of $0 < \beta < 1$ is also defined. In this case two circles of radius $L/(2\beta)$ that pass through both points p_i and p_j are instead constructed and the intersection is again denoted by R , and note that the construction of the circles differs from the case of $\beta \geq 1$.

Then, if there is no other point p_k included in the region R , the points p_i and p_j are connected by a segment of line, otherwise the points are not connected. By considering this process for all pairs of points, a graph $G_\beta = (V, E)$ is constructed in which the set of vertices V is formed by the points p_i and the set of edges E is formed by the

1. Networks and Graphs

segments of lines connecting pairs of vertices. Obviously, for small values of β , e.g., for $0 < \beta < 1$, the chances that there is a point in the region R associated with p_i and p_j is very small, and there is a high probability that these two points are connected. As a consequence of this, the resulting graphs are very dense, containing a large number of triangles. It can be seen that if $\beta = 0$, the resulting graph is just the complete graph. Special cases that are commonly considered in the computational geometry literature are when $\beta = 1$ or $\beta = 2$, which corresponds to the Gabriel graph (GG) and relative neighbourhood graph (RNG) respectively. We note here that the $\beta = 2$ case (RNG) will be especially important in this Thesis.

The β -skeletons are monotone in β , that is if $\beta_1 > \beta_2$ then $G_{\beta_1}(V) \subset G_{\beta_2}(V)$. The β -skeletons are related to other well-known proximity graphs such minimum spanning trees (MST) and Delauney triangulations (DT). It has been shown that in \mathbb{R}^2 with the L_p norm, $MST(V) \subset G_{\beta}(V) \subset DT(V)$ for $1 \leq \beta \leq 2$. This implies that for $1 \leq \beta \leq 2$ the β -skeleton is both connected and planar. It also gives bounds on the number of edges of these graphs of $n - 1 \leq m \leq 3n - 6$, where the lower bound comes from the connectedness and the upper bound comes from the planarity of $DT(V)$.

It is clear that the literature for the relative neighbourhood graphs and other β -skeletons is mostly concerned with efficient computation of the graphs in various circumstances [152, 98, 99], their use in identifying clusters of points [151], and broad properties of the resulting graphs. This is because it has mostly been of interest in computational geometry and pattern recognition, and less so as a random graph model, and so it has received relatively little discussion in terms of its properties and how they change as the parameters of the model are changed.

1. Networks and Graphs

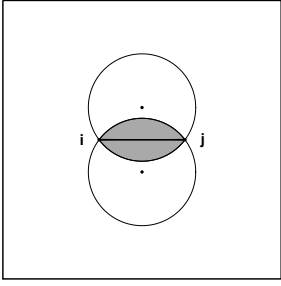
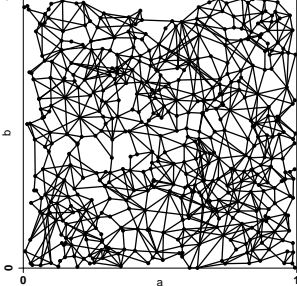
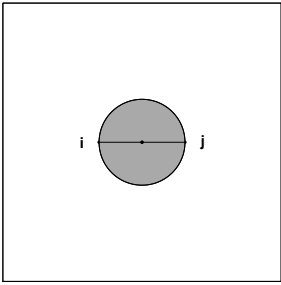
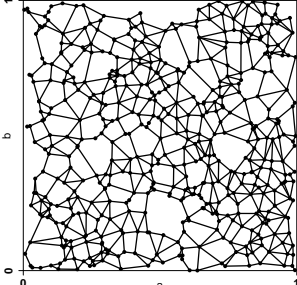
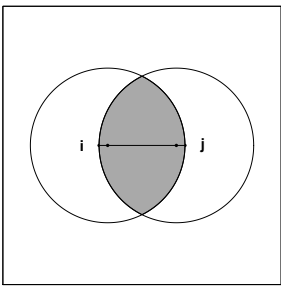
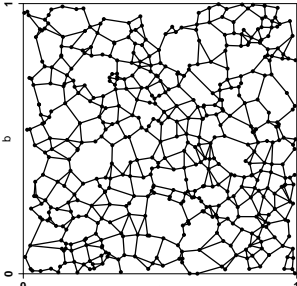
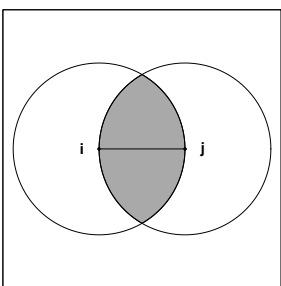
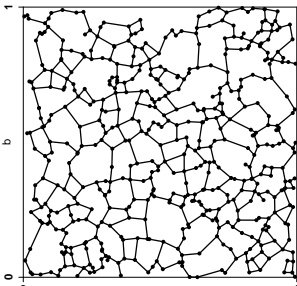
| β Value | Construction of β -skeleton | Example of β -skeleton |
|-----------------|---|--|
| $\beta < 1$ |  |  |
| $\beta = 1$ |  |  |
| $1 < \beta < 2$ |  |  |
| $\beta = 2$ |  |  |

Figure 1.2: Illustration of the construction of β -skeletons in a unit square for different values of β , where the region R is shaded grey, with an example graph with $n = 500$ nodes for each. From top to bottom: $\beta = 0.8, 1.0, 1.8, 2.0$.

1.3 Dynamics on Networks

Many real-world networks are not static in nature. There may be some dynamical process taking place on the objects represented by the nodes, and also the network itself may be evolving over time. Understanding the behaviour of these processes and how they are influenced by the topology of a particular network is of great practical interest in many areas. There are two possibilities for interpreting the network representation of the system. Either the nodes represent some individual that is interacting with other individuals in some way, or they represent possible locations for the flow of information, people, particles, etc. In either case, we can consider some variable for each node that contains the information about its state.

For example, in social groups information such as opinions will propagate through people interacting. Alternatively, we may be interested in the propagation of some disease that is being transmitted by people in close proximity to one another. Understanding the dynamics of these systems will help in understanding, predicting and influencing the spread of disease in order to minimise the risk of epidemics occurring. This kind of epidemic spreading also applies to other situations, such as plants in fields where a disease may spread to another plant if it is within close enough proximity.

Another example is the flow of information in a wireless sensor network. Wireless sensors may be distributed over a geographical area in order to make some measurements, e.g. for measuring air pollution. They can only communicate with nearby sensors without causing interference and excessive power consumption. This means that it is crucial to ensure that information can spread through the entire network, and that it can do so efficiently.

We model a dynamical process on a network by assigning a value to each node. These are independent variables that we identify with some quantity of interest in the real system. Therefore, we can describe the state of the system at a time $t \geq 0$ by the vector $\mathbf{u}(t) = (\mathbf{u}_1(t), \mathbf{u}_2(t), \dots, \mathbf{u}_n(t))^T$. We clarify that \mathbf{u}_i is the i th entry of \mathbf{u} , in contrast to eigenvectors where we recall that $\boldsymbol{\psi}_i$ is the i th eigenvector of \mathcal{L} , and

1. Networks and Graphs

we will write $\psi_{i,j}$ for the j th entry of the i th eigenvector. The vector \mathbf{u} contains the information of the state of every node as the dynamics evolves in time, as nodes interact with each other along the links of the network.

1.3.1 Diffusion (Consensus) Process

Here we discuss diffusion processes in graphs, how they are related to consensus protocols, and show a derivation of the Laplacian matrix that is crucial to this work. We consider first a simple diffusion process taking place on the nodes of the graph, which is governed by the following equation.

$$\frac{d\mathbf{u}_i(t)}{dt} = \dot{\mathbf{u}}_i(t) = \sum_{j:(v_i,v_j) \in E} (\mathbf{u}_j(t) - \mathbf{u}_i(t)), \quad i = 1, 2, \dots, n. \quad (1.38)$$

To visualise this, consider the small graph shown in Fig. 1.3. For each of the neighbours of node v_i , we consider that its contribution to the rate of change of the state node v_i is proportional to the difference in the states of the two nodes. Then, the rate of change of the state of a node is just the sum of these differences over all adjacent nodes as in Eqs. 1.39. This is equivalent to a simple consensus process (also discussed as a consensus protocol) where every node is trying to reach an agreement with its neighbours [110, 122, 123, 38, 124], though other consensus protocols exist.

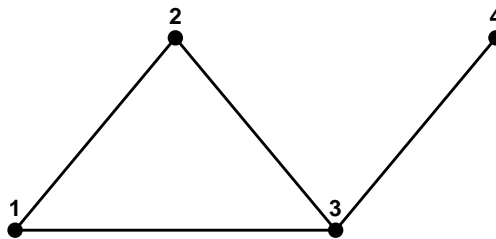


Figure 1.3: A small graph to help demonstrate the equations governing dynamics.

1. Networks and Graphs

$$\begin{aligned}
\dot{\mathbf{u}}_1 &= (\mathbf{u}_2 - \mathbf{u}_1) + (\mathbf{u}_3 - \mathbf{u}_1) \\
&= \mathbf{u}_2 + \mathbf{u}_3 - 2\mathbf{u}_1 \\
\dot{\mathbf{u}}_2 &= (\mathbf{u}_1 - \mathbf{u}_2) + (\mathbf{u}_3 - \mathbf{u}_2) \\
&= \mathbf{u}_1 + \mathbf{u}_3 - 2\mathbf{u}_2 \\
\dot{\mathbf{u}}_3 &= (\mathbf{u}_1 - \mathbf{u}_3) + (\mathbf{u}_2 - \mathbf{u}_3) + (\mathbf{u}_4 - \mathbf{u}_3) \\
&= \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_4 - 3\mathbf{u}_3 \\
\dot{\mathbf{u}}_4 &= (\mathbf{u}_3 - \mathbf{u}_4).
\end{aligned} \tag{1.39}$$

More concisely and in general, this can be written for each node v_i as

$$\begin{aligned}
\dot{\mathbf{u}}_i(t) &= - \sum_{j=1}^n A_{ij} (\mathbf{u}_i(t) - \mathbf{u}_j(t)) \\
&= \sum_{j=1}^n A_{ij} \mathbf{u}_j(t) - k_i \mathbf{u}_i(t).
\end{aligned} \tag{1.40}$$

This suggests that we may write the set of equations for our example as

$$\begin{pmatrix} \dot{\mathbf{u}}_1 \\ \dot{\mathbf{u}}_2 \\ \dot{\mathbf{u}}_3 \\ \dot{\mathbf{u}}_4 \end{pmatrix} = - \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \mathbf{u}_4 \end{pmatrix} \tag{1.41}$$

It is obvious now that we can write Eq. 1.38 by using the Laplacian matrix of the graph as in Eq. 1.41,

$$\dot{\mathbf{u}}(t) = -\mathcal{L}\mathbf{u}(t), \quad \mathbf{u}(0) = \mathbf{u}_0 \tag{1.42}$$

where in a slight abuse of notation we have some initial condition $\mathbf{u}(0) = \mathbf{u}_0$, and write $\mathbf{u}_{0,i}$ for the i th entry of \mathbf{u}_0 . This shows how the Laplacian matrix arises naturally when considering dynamical processes on graphs. In Eq. 1.42 the Laplacian matrix is acting over the vector $\mathbf{u}(t)$; this is simply the heat equation on a graph, and then the interpretation is that each node has a temperature and will exchange heat with its

1. Networks and Graphs

neighbours until thermal equilibrium is reached. The solution of this equation is

$$\mathbf{u}(t) = e^{-t\mathcal{L}}\mathbf{u}_0. \quad (1.43)$$

Recalling that $0 = \mu_1 < \mu_2 \leq \dots \leq \mu_n$ are the eigenvalues and $\psi_1, \psi_2, \dots, \psi_n$ are the eigenvectors of the Laplacian matrix, the solution of the diffusion equation on the graph is given by

$$\mathbf{u}(t) = e^{-t\mu_1}(\psi_1 \cdot \mathbf{u}_0)\psi_1 + e^{-t\mu_2}(\psi_2 \cdot \mathbf{u}_0)\psi_2 + \dots + e^{-t\mu_n}(\psi_n \cdot \mathbf{u}_0)\psi_n, \quad (1.44)$$

where $\mathbf{x} \cdot \mathbf{y}$ represents the inner product of the corresponding vectors.

For the sake of simulations it is sometimes useful to consider the discrete-time version of the equation, which can be written as [110, 122]:

$$\mathbf{u}_i(k+1) = \mathbf{u}_i(k) + \epsilon \sum_{j=1}^n A_{ij}(\mathbf{u}_j(k) - \mathbf{u}_i(k)), \quad (1.45)$$

where $0 < \epsilon < k_{max}^{-1}$ is the time step for the simulation and A_{ij} is the (i, j) th entry of A . Equation 1.45 can also be written in matrix form as

$$\mathbf{u}(k+1) = (I - \epsilon\mathcal{L})\mathbf{u}(k). \quad (1.46)$$

The matrix $(I - \epsilon\mathcal{L})$ is usually known as the Perron matrix. An example of a diffusion process taking place on a graph is shown in Fig. 1.4.

In general, the best structural parameter describing the diffusion time is the algebraic connectivity, μ_2 . In order to understand this relation, we consider the solution of the diffusion equation (Eq. 1.42). Since we can consider this to be a consensus process we now use some corresponding terminology and results. When the time tends to infinity every node tends to the state dictated by the average of the values of the initial condition. This state is usually known as the *consensus set* [110] and it can be formally defined as the set $\mathcal{A} \subset \mathbb{R}^n$, which is the subspace $span\{\mathbf{1}\}$, i.e.,

$$\mathcal{A} = \{\mathbf{u} \in \mathbb{R}^n | \mathbf{u}_i = \mathbf{u}_j, \forall i, j \in \{1, \dots, n\}\}. \quad (1.47)$$

1. Networks and Graphs

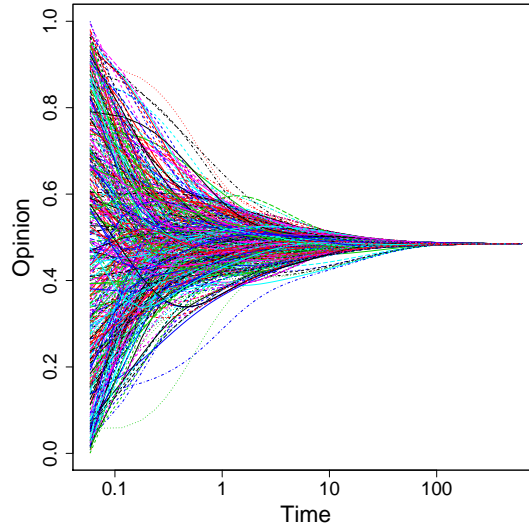


Figure 1.4: Illustration of the diffusion dynamics in an RGG. The simulations were carried out using a discrete time diffusion model (see Eq. 1.46) with a random allocation of initial states for the nodes.

The following is a well-known result in the theory of consensus dynamics on networks.

Theorem 1.3.1. ([110] p. 46) *Let G be a connected graph. Then, the consensus dynamics converges to the agreement set with a rate of convergence that is dictated by μ_2 .*

Proof. As $t \rightarrow \infty$

$$\mathbf{u}(t) \rightarrow (\boldsymbol{\psi}_1 \cdot \mathbf{u}_0) \boldsymbol{\psi}_1 = \frac{\mathbf{1} \cdot \mathbf{u}_0}{n} \mathbf{1}, \quad (1.48)$$

and hence $\mathbf{u}_t \rightarrow \mathcal{A}$. As μ_2 is the smallest positive eigenvalue of the graph Laplacian, it dictates the slowest mode of convergence in Eq. 1.44.

To understand this in more detail, consider that the time tends to the time of consensus $t \rightarrow t_c$, where t_c is the time at which $|\mathbf{u}_i(t) - \mathbf{u}_j(t)| \leq \delta$ for all i, j . Denote this time by t_c^- . Note that we refer to the time of consensus as the time of diffusion or diffusion time when we are discussing diffusion. Then,

$$\mathbf{u}_p(t_c^-) = \frac{1}{n} \sum_{q=1}^n \mathbf{u}_{0,q} + \sum_{j=2}^n \left(\boldsymbol{\psi}_{j,p} e^{-t_c^-(p)\mu_j} \sum_{q=1}^n \boldsymbol{\psi}_{j,q} \mathbf{u}_{0,q} \right), \quad (1.49)$$

1. Networks and Graphs

where here $t_c^-(p)$ means the time at which the node p is close to reaching the diffusion state. Let $\langle \mathbf{u}_0 \rangle = \frac{1}{n} \sum_{q=1}^n \mathbf{u}_{0,q}$ and write Eq. 1.49 as follows

$$\mathbf{u}_p(t_c^-) - \langle \mathbf{u}_0 \rangle = \sum_{j=2}^n \left(\psi_{j,p} e^{-t_c^-(p)\mu_j} \sum_{q=1}^n \psi_{j,q} \mathbf{u}_{0,q} \right). \quad (1.50)$$

A node v_p is selected such that $\psi_{2,p}$ has the same sign as $\psi_2 \cdot \mathbf{u}_0$. Since μ_2 is the smallest eigenvalue in the sum on the right hand of the expression, this terms tends to 0 slower than the terms for the other values of j . This means that, if δ is sufficiently small, the values of t_c and thus t_c^- will be very large. Thus, it is possible to ensure that the left side of the equation is small enough such that $\sum_{j=3}^n \left(\psi_{j,p} e^{-t_c^-(p)\mu_j} (\psi_j \cdot \mathbf{u}_0) \right) < 0$. This implies that

$$(\mathbf{u}_p(t_c^-) - \langle \mathbf{u}_0 \rangle) < \psi_{2,p} e^{-t_c^-(p)\mu_2} (\psi_2 \cdot \mathbf{u}_0). \quad (1.51)$$

Now, because $|\mathbf{u}_p(t_c^-) - \langle \mathbf{u}_0 \rangle| \geq \delta$,

$$\delta \leq |\mathbf{u}_p(t_c^-) - \langle \mathbf{u}_0 \rangle| < \left| \psi_{2,p} e^{-t_c^-(p)\mu_2} (\psi_2 \cdot \mathbf{u}_0) \right|. \quad (1.52)$$

Then, the time at which the diffusion is reached $t_c(p)$ is bounded by

$$t_c(p) \geq t_c^-(p) \geq \frac{1}{\mu_2} \ln \left| \frac{\psi_{2,p} (\psi_2 \cdot \mathbf{u}_0)}{\delta} \right|. \quad (1.53)$$

Finally, the average time of diffusion is bounded by

$$\langle t_c \rangle \geq \frac{1}{\mu_2 n} \sum_{p=1}^n \ln \left| \frac{\psi_{2,p} (\psi_2 \cdot \mathbf{u}_0)}{\delta} \right|. \quad (1.54)$$

□

1.3.2 Epidemics on Networks

Networks are very appealing for the application of epidemiological models in ecology on the different spatial and temporal scales. The discovery of the fact that networks with

1. Networks and Graphs

fat-tailed degree distribution do not display an epidemic threshold in the asymptotic limit is a relevant example of how the connectivity pattern of interacting agents can dramatically change the course of an epidemic [19, 28]. The use of network theory in epidemiological models provides a way to incorporate the individual-level heterogeneity necessary for a mechanistic understanding of the spread of infectious disease. However, in a recent review about the implications of modelling disease using networks for plant sciences, Jeger et al. [96] has recognized that there has been “*relatively little use of network theory in plant epidemiology*” and claimed that models of epidemics on networks “*might work better for animal or human than for plant diseases*”. Among the models that have found applications to study plant diseases, those using spatial features for characterising the structure of populations in heterogeneous landscapes have gained recent interest [96]. In such models it is possible to consider spatial networks that treat interactions as a continuous variable that decays with increasing distance or by distributing randomly and independently a set of vertices on the Euclidean plane to represent the relative spatial location of individual hosts or habitat patches. The second kind of model is based on RGGs.

Brooks et al. [23] have used RGGs to model the interactions between the anther smut fungus and fire pink using a temporal data that span 7 years of field studies. They have concluded that the use of spatially explicit network models “*can yield important insights into how heterogeneous structure can promote the persistence of species in natural landscapes*”. Since plants are not mobile, the shape of the plot or field in which the plants are distributed may have a significant impact on the spread of disease. In fact, there is both empirical and theoretical evidence that supports this hypothesis [126, 157, 93, 65, 21, 113, 112, 164, 63]. In general, it has been suggested that square plots and fields favoured higher spreading of plant diseases than elongated ones of the same area [126, 157, 93, 65]. It is important to remark that the area of the field also plays a fundamental role, with larger plots and fields better favouring the spread of diseases [113, 164, 63]. Also, the orientation of elongated fields may affect the disease propagation with orientations perpendicular to prevalent winds suppressing epidemic progression [157, 65]. All in all, for plots and field of the same area and

1. Networks and Graphs

orientation there is empirical and theoretical evidence that elongated shapes decreases the impact of epidemics on plant populations. It is worth noting that the theoretical models [112, 164, 63] used in the previously mentioned studies do not use network theory as a tool for the study of epidemic spreading.

When using networks to model infectious diseases spreading between individuals of a population, nodes correspond to individuals and edges correspond to contact between them through which the infection can spread [9, 5]. There are three states individuals can be in: they are *susceptible* (S) when they can be infected by the disease, *infected* (I) when they can spread the pathogen, and *recovered* (R) when they have been previously infected and are now immune. There are two common models that are used, depending on the nature of the pathogen. The Susceptible-Infected-Recovered (SIR) model is appropriate when a previous infection provides immunity to further infection, and Susceptible-Infected-Susceptible (SIS) when this does not happen. The dynamics of both SIS and SIR models are governed by two parameters. The per contact *infection rate* β , which is the rate at which a susceptible individual will catch the infection from an infected individual, and the *recovery rate* μ , which is the rate at which an infected individual recovers. Let s_i , x_i and r_i be the probabilities that the node v_i belongs to S, I, or R respectively. The SIS model is then written as

$$\begin{aligned}\dot{s}_i &= -\beta s_i \sum_j A_{ij} x_j + \mu x_i, \\ \dot{x}_i &= \beta s_i \sum_j A_{ij} x_j - \mu x_i,\end{aligned}\tag{1.55}$$

while the SIR model is written as

$$\begin{aligned}\dot{s}_i &= -\beta s_i \sum_j A_{ij} x_j, \\ \dot{x}_i &= \beta s_i \sum_j A_{ij} x_j - \mu x_i, \\ \dot{r}_i &= \mu x_i.\end{aligned}\tag{1.56}$$

The ratio β/μ is crucial to the dynamics of the infection. The *epidemic threshold* (or

1. Networks and Graphs

basic reproduction number) $\tau = (\beta/\mu)_c$ is defined as the critical value of the transition and depends on the topology of the network. There are two phases, an absorbing phase $\beta/\mu < \tau$ where the infection dies out before infecting a large number of nodes, and an active phase $\beta/\mu > \tau$ where it succeeds in spreading through a large fraction of the nodes. When $\beta/\mu = \tau$ the infection becomes endemic in the population. The transition from the absorbing to the active phase strictly resembles a non equilibrium second order phase transition in statistical physics [109, 79]. In particular, it has been shown that τ is inversely related to the spectral radius [29, 73, 154]:

$$\tau = \frac{1}{\lambda_1}. \quad (1.57)$$

1.3.3 The Flow of Fluids

The flow of fluids through porous media is frequently described by the Boussinesq equation, which is also known as the porous media equation:

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) = \frac{\partial}{\partial x} (\mathbf{u}(x, t) \frac{\partial}{\partial x} \mathbf{u}(x, t)), \quad \mathbf{u}(x, 0) = \mathbf{u}_0, \quad (1.58)$$

where $\mathbf{u}(x, t)$ is a non-negative scalar function on $x \in \Omega = [0, 1]$ and time $t \geq 0$.

Suppose that the fluid is flowing through a capillary of length L and height h_0 , and suppose that the capillary is much longer than it is thick: $L \gg h_0$. Then, according to [78] (see especially Fig. 4) and [131], the Boussinesq equation can be very well approximated by a simple heat equation

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) = \frac{\partial^2}{\partial x^2} \mathbf{u}(x, t), \quad \mathbf{u}(x, 0) = \mathbf{u}_0. \quad (1.59)$$

Obviously, the channels produced by the fractures of rocks are less than a millimetre thick and a few centimetres long. Thus, $L \gg h_0$ is always true and the use of the heat equation is justified for modelling the diffusion of oil and gas through the channels formed by the network of rock fractures. In the case of diffusion through the edges of

1. Networks and Graphs

a network the previous equation can be written as

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) = -\mathcal{L}\mathbf{u}(x, t), \quad \mathbf{u}(x, 0) = \mathbf{u}_0. \quad (1.60)$$

Therefore, we recover the same model as for the diffusion (consensus) process.

1.4 Summary

In this part of the Thesis we introduced the background theory relevant to this work, and established the notation and concepts that will be required. A table of useful notation may be found on page xii. We first gave an overview of some basic graph theory and network theory, then we discussed several random graph models that are used within this work, and finally we discuss some dynamical processes on networks.

In the next part of this Thesis we first present some theoretical results on a generalisation of the spatially defined random graph models we have introduced. Then, we consider three applications of these random graph models, with an emphasis on the three dynamical processes we have mentioned here.

Part II

Results and Discussion

Chapter 2

Structure and Connectivity of Random Rectangular Graphs

In all the real-world scenarios where the RGG model has been applied, the shape of the location in which the nodes of the graph are distributed may play a fundamental role in the topological and dynamical properties of the resulting graphs. That is, it is intuitive to think that the connectivity, distance, clustering and other fundamental topological properties of the graphs are affected if we, for instance, elongate the unit square in which the points are distributed. We present a new model that generalises the RGG by allowing the embedding of the nodes in a unit rectangle instead of a unit square. Our main goal is to investigate how the elongation of a unit square influences the topological properties of the graphs generated by the model. We call these graphs the *random rectangular graphs* (RRGs) and have published several results [58, 59, 56], see also [52]. In this work we study the influence of the length of the rectangle on the topological properties of the graphs emerging on them, such as their average degree, connectivity, degree distribution, average path length and clustering coefficient. In particular, we find analytical expressions and bounds for all of them and provide computational evidence of the tightness of the bounds for relatively large RRGs. Then, we consider dynamical processes taking place on the RRGs and study how the behaviour of such processes changes with the elongation of the rectangle. Throughout this work, R was used to perform most of the numerical calculations and to produce most of the figures. All

2. Structure and Connectivity of Random Rectangular Graphs

other calculations and figures used MATLAB.

2.1 Definition of RRGs

Let us now define a unit hyperrectangle as the Cartesian product $[0, x_1] \times [0, x_2] \times \cdots \times [0, x_d]$ where $x_i > 0 \in \mathbb{R}$, and $\prod_{1 \leq i \leq d} x_i = 1$. Hereafter we will restrict ourselves to the 2-dimensional case, the unit rectangle, except to discuss future work. The RRG is defined by uniformly and independently distributing n points in the unit rectangle $\mathcal{R} = [0, a] \times [0, b]$ where $a \geq 1, b = 1/a$, and then connecting two points by an edge if their (Euclidean) distance is at most r . Though of course we may write all results in terms of a , we may also make use of b explicitly when it is useful to do so. It is evident that the only change we have introduced here is to consider a rectangle of unit area instead of the analogous square. The rest of the construction process remains the same as for the RGG. This means that $\text{RRG} \rightarrow \text{RGG}$ as $a \rightarrow 1$. In this sense we can say that the RRG is a generalisation of the RGG. In Fig. 2.1 we illustrate an RRG.

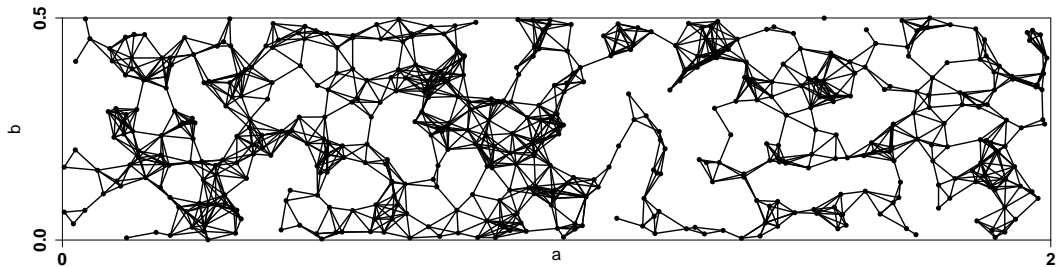


Figure 2.1: Illustration of a random rectangular graph with $a = 2$ ($b = 0.5$). It has 500 nodes and $r = 0.07$, which is the same as Fig. 1.1.

An interesting question is what happens at the other extreme, when a becomes very large. In this case $b \approx 0$, which means that the n points behave as though they are uniformly and independently distributed on the straight line. Let us consider a disk of radius $r > 0$ centred at each of these points and connect every point to the other points that lie inside its disk. The resulting graph resembles a one-dimensional RGG, that is a graph created by placing the n points uniformly and independently on the interval $[0, 1]$ and then connecting pairs of nodes if they are at a (Euclidean) distance smaller than or equal to a certain connection radius r (see for instance [39, 66, 71]).

2.2 Topological Properties of RRGs

In order to understand the behaviour of the RRG model, determine that it is indeed different from the RGG model, and obtain useful insight into the behaviour of dynamical processes on the graphs that model some real-world scenario, we need to examine various key topological properties. As well as simply examining the behaviour, we would like to find bounds on the values of these properties for different parameter values, and if possible, find an analytical expression for the expected value of these properties. In this section, we present the results we have found for several important structural parameters of RRGs.

2.2.1 Average Node Degree

We start the study of the topological properties of RRGs by considering an analytical expression for the average degree \bar{k} .

Theorem 2.2.1. *The expected value of the average node degree $\bar{k}(G_{RRG})$ of an RRG is [58]*

$$\mathbf{E}(\bar{k}) = (n - 1)f, \quad (2.1)$$

where f is given by

$$f = \begin{cases} 0 \leq r \leq b & \frac{1}{2}r^4 - \frac{4}{3}(a + b)r^3 + \pi r^2, \\ b \leq r \leq a & -\frac{4}{3}ar^3 - b^2r^2 + \frac{1}{6}b^4 + (\frac{4}{3}ar^2 + \frac{2}{3}b)\sqrt{r^2 - b^2} \\ & + 2r^2 \arcsin\left(\frac{b}{r}\right), \\ a \leq r \leq \sqrt{a^2 + b^2} & -\frac{1}{2}r^4 - (a^2 + b^2)r^2 + \frac{1}{6}(a^4 + b^4) \\ & + (\frac{4}{3}ar^2 + \frac{2}{3}b)\sqrt{r^2 - b^2} + (\frac{4}{3}br^2 + \frac{2}{3}a)\sqrt{r^2 - a^2} \\ & + 2r^2(\arcsin\left(\frac{b}{r}\right) - \arccos\left(\frac{a}{r}\right)). \end{cases} \quad (2.2)$$

Proof. To start with, let us consider that for a given node, there are $n - 1$ nodes distributed in the rest of the rectangle. Define $A(x)$ to be the area within the radius r of a

2. Structure and Connectivity of Random Rectangular Graphs

point x that lies within the rectangle. Since the nodes are uniformly and independently distributed, the expected degree of a node v_i is $\mathbf{E}(k_i) = (n-1)A(v_i)$, where in a slight abuse of notation $A(v_i)$ uses the point where node v_i is located. This is because dividing the nodes between the area within distance r and the rest of the rectangle gives rise to the Binomial distribution $\text{Bin}(n-1, A(v_i))$. Averaging this over all possible node locations $x \in \mathcal{R}$ gives

$$\mathbf{E}(\bar{k}) = \int_{\mathcal{R}} (n-1)A(x) dx = (n-1) \int_{\mathcal{R}} A(x) dx. \quad (2.3)$$

We simply write the integral with respect to dx for now, and will do so again for similar integrals, but note that this is indeed a 2D integral and we deal with a specific parametrisation separately to solve this as a double integral. Recalling that $\delta(G) = \bar{k}/(n-1)$, let $f(a, b, r) = \mathbf{E}(\delta(G))$ be the area within radius r of a point that lies in the rectangle, integrated over all points, i.e., $f(a, b, r) = \int_{\mathcal{R}} A(x) dx$. We find that we need to consider the following three regions: $0 \leq r \leq b$, $b \leq r \leq a$ and $a \leq r \leq \sqrt{a^2 + b^2}$, recalling that $a \geq b$. We call these cases 1, 2 and 3, respectively. Thus, the function $f(a, b, r)$ takes different forms f_i for each case i . This means that we can write

$$\mathbf{E}(\bar{k}) = (n-1)f, \quad (2.4)$$

with

$$f = \begin{cases} f_1 & 0 \leq r \leq b, \\ f_2 & b \leq r \leq a, \\ f_3 & a \leq r \leq \sqrt{a^2 + b^2}. \end{cases} \quad (2.5)$$

We now find analytical expressions for each f_i by constructing suitable integrals.

We consider the rectangle in Fig. 2.2, which shows 3 quarter circles of different radii (each corresponding to one of the three cases) as they intersect the interior of the rectangle. We exploit the symmetry of the problem to simplify analysis by considering only quarter circles instead of circles, then we quadruple the result at the end. For each of these quarter circles, we divide them into vertical rectangular strips of width Δx that will approximate the areas of the intersection between the quarter circles and

2. Structure and Connectivity of Random Rectangular Graphs

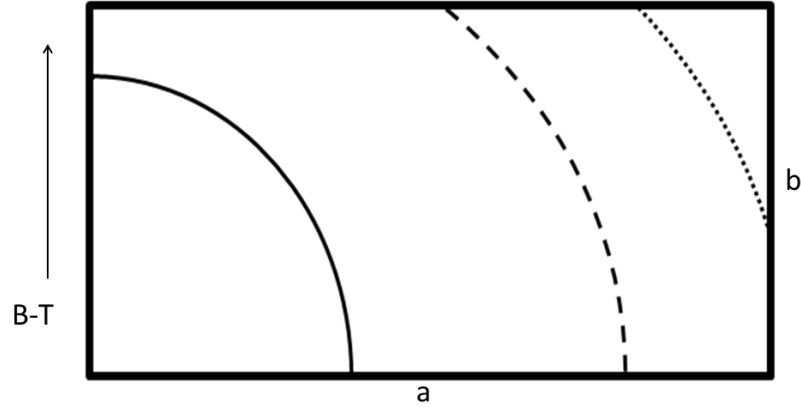


Figure 2.2: Illustration of three different quarter circles in the rectangle corresponding to $0 \leq r \leq b$ (solid line), $b \leq r \leq a$ (broken line) and $a \leq r \leq \sqrt{a^2 + b^2}$ (dotted line). The direction used for displacing the circles is represented as bottom-top (B-T) with an arrow in the graphic.

the full rectangle; this will of course become exact in the limit as $\Delta x \rightarrow 0$ to form an integral. We now consider several possibilities for these strips.

First, the strips may approximate an area that is not rectangular, which occurs when the height of the strip is smaller than the height b of the rectangle. For a strip of distance x from the left of the rectangle, this corresponds to $0 \leq x \leq r$ for the smaller quarter circle (case 1), $\sqrt{r^2 - b^2} \leq x \leq r$ for the medium quarter circle (case 2), and $\sqrt{r^2 - b^2} \leq x \leq a$ for the largest quarter circle (case 3). Setting $p = \min(a, r)$, $q = \min(b, r)$, we have $\sqrt{r^2 - q^2} \leq x \leq p$.

Since we need to integrate these areas over all possible quarter circles, given a fixed radius, we wish to know how we can translate the quarter circle in the rectangle and preserve a particular strip. That is, for a particular strip of distance x from the left of the rectangle, we may find a corresponding strip on the other quarter circles of the same radius. Since we have a rectangular strip of width Δx , height $\sqrt{r^2 - x^2}$, and distance x from the centre of the (quarter) circle, we may find this strip in any of the $(a - x - \Delta x)$ positions horizontally, and $(b - \sqrt{r^2 - x^2})$ vertically. Thus, we can use integration to

2. Structure and Connectivity of Random Rectangular Graphs

find the total area of all these strips by multiplying $(a - x - \Delta x)(b - \sqrt{r^2 - x^2})$ by the area of the strip $\sqrt{r^2 - x^2} \Delta x$, and taking the limit to obtain

$$\begin{aligned} I_1 &= \int_{\sqrt{r^2 - q^2}}^p (a - x)(b - \sqrt{r^2 - x^2})\sqrt{r^2 - x^2} dx \\ &= \int_{\sqrt{r^2 - q^2}}^p (a - x)(b\sqrt{r^2 - x^2} - (r^2 - x^2)) dx. \end{aligned} \quad (2.6)$$

Secondly, we note that if these strips are translated far enough in the bottom-top (B–T) direction, they become truncated by the top of the rectangle. For a particular truncated strip, we may still find a corresponding strip in any of the $(a - x - \Delta x)$ positions horizontally, and the truncated height t of a strip may be any value between 0 and the full height of the strip:

$$\begin{aligned} I_2 &= \int_{\sqrt{r^2 - q^2}}^p (a - x) \int_0^{\sqrt{r^2 - x^2}} t dt dx \\ &= \int_{\sqrt{r^2 - q^2}}^p \frac{1}{2}(a - x)(r^2 - x^2) dx. \end{aligned} \quad (2.7)$$

Alternatively, we may have $\sqrt{r^2 - b^2} > b$, in which case the rectangular strip is exact and of height b . In this case, the only contribution is from the truncated strips. We note that this applies for $0 \leq x \leq \sqrt{r^2 - q^2}$ by a similar argument as before, and we integrate as follows

$$\begin{aligned} I_3 &= \int_0^{\sqrt{r^2 - q^2}} (a - x) \int_0^b t dt dx \\ &= \int_0^{\sqrt{r^2 - q^2}} \frac{1}{2}(a - x)b^2 dx. \end{aligned} \quad (2.8)$$

Thus, we have the expression for f as four times the sum of the above integrals

$$\begin{aligned} f &= 4(I_1 + I_2 + I_3) \\ &= \int_0^{\sqrt{r^2 - q^2}} 2(a - x)b^2 dx + \int_{\sqrt{r^2 - q^2}}^p (a - x)(4b\sqrt{r^2 - x^2} - 2(r^2 - x^2))dx. \end{aligned} \quad (2.9)$$

Evaluation of these integrals yields the closed-form expression given in the theorem.

2. Structure and Connectivity of Random Rectangular Graphs

□

Now we demonstrate this result computationally. In Fig. 2.3(a) we plot the values of the average degree observed for RRGs with three different values of the rectangle side length. These observed values (represented by solid squares, circles and triangles) are the average of 100 random realisations of RRGs with 1500 nodes. The solid lines represent the expected values according to the expressions (2.2). The Pearson correlation coefficients for the linear regression between the observed and expected values are larger than 0.9999 in all three cases. We enlarge the region of small radii for the case $a = 30$ (see Fig. 2.3 (b)) where it can be seen that it is a perfect fit also for this region with Pearson correlation coefficient as good as for the general case. As per the discussion of the node degree in the RGG with toroidal boundary (Eq. 1.35), for small radii we see that the average node degree looks quadratic in r , since $\bar{k} \approx n\pi r^2$, but here we precisely account for the effect of the boundary.

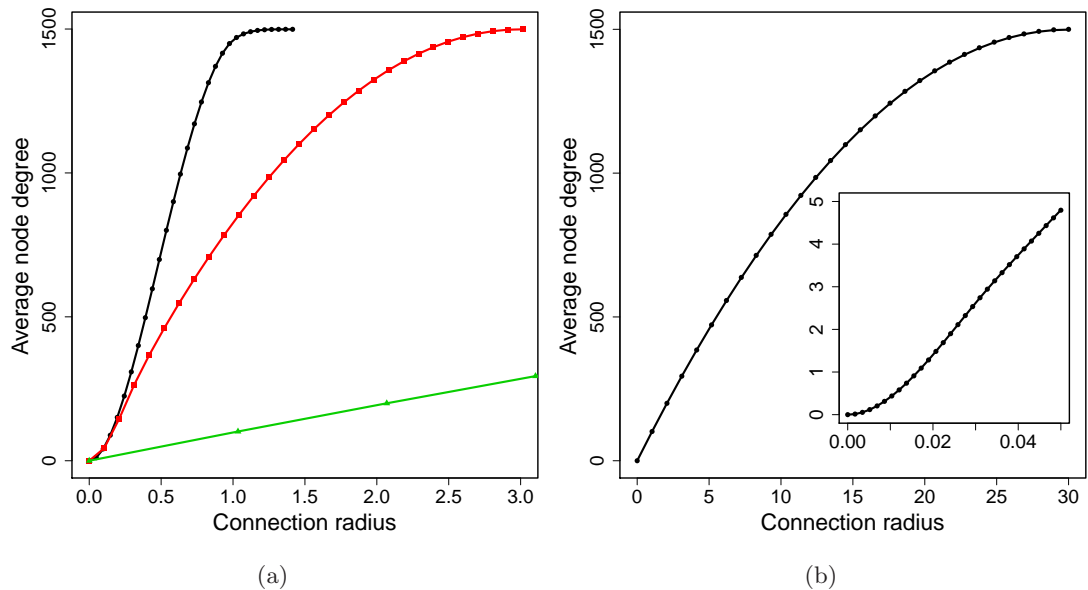


Figure 2.3: (a) Illustration of the fit between the observed (black circles ($a = 1$), red squares ($a = 3$), and green triangles ($a = 30$)) and expected (solid line) values of the average degree for RRGs with different side lengths of the rectangle. (b) Wider range of radii for $a = 30$ (zooming in for small radii in the inset).

We now consider the change in the average node degree as the rectangle elongation

2. Structure and Connectivity of Random Rectangular Graphs

varies. Specifically, we want the partial derivative $\partial f/\partial a$ to be negative, or at least never positive, so that the average node degree will generally decrease as we increase the elongation. In Fig. 2.4 we demonstrate the behaviour of this partial derivative, both for varying the elongation for several values of connection radius ($r = 0.05, 0.01, 0.15, 0.20$), and for varying the connection radius for $a = 1.5$. We find that the value is 0 when $r = 0$, $r = \sqrt{a^2 + b^2}$, or $a = 1$, and negative otherwise, and we then provide a proof that this holds in general.

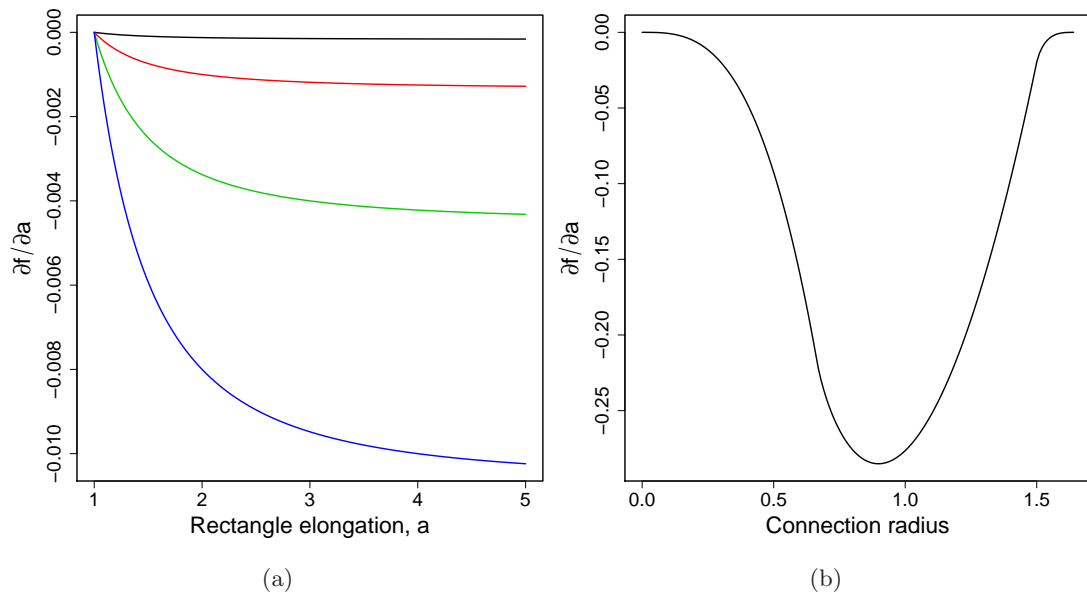


Figure 2.4: Illustration of the partial derivative $\partial f/\partial a$ (a) versus a for $r = 0.05$ (black), $r = 0.10$ (red), $r = 0.15$ (green), $r = 0.20$ (blue), (b) versus r for $a = 1.5$.

Theorem 2.2.2. *The average node degree $\bar{k}(G_{RRG}) = (n - 1)f$ is a non-increasing function of rectangle elongation a , that is,*

$$\frac{\partial f}{\partial a} \leq 0, \quad (2.10)$$

with equality only when $a = 1$, $r = 0$ or $r = \sqrt{a^2 + b^2}$

Proof. Result follows from Lemmas 2.2.3, 2.2.4 and 2.2.7. \square

Let $f = f_i(a, r)$, and for convenience we sometimes use the subscript notation f_a

2. Structure and Connectivity of Random Rectangular Graphs

to indicate the partial derivative of f_i w.r.t. a , where we do not write the i for clarity since it is obvious from context.

Lemma 2.2.3. *For $r \in [0, a^{-1}]$ (case 1), we have that $f_a \leq 0$ with equality only when $a = 1$ or $r = 0$.*

$$\frac{\partial f_1}{\partial a} = -\frac{4}{3}r^3(1 - a^{-2}) \leq 0, \quad (2.11)$$

Proof. Follows immediately since $(1 - a^{-2}) \geq 0$, with equality only when $a = 1$ or $r = 0$. \square

Lemma 2.2.4. *For $r \in [a^{-1}, a]$ (case 2), we have that $f_a \leq 0$ with equality only when $a = 1$*

$$\frac{\partial f_2}{\partial a} = -\frac{4r^3}{3} + \frac{2r^2}{a^3} - \frac{2}{3a^5} + \frac{4(a^2r^2 - 1)^{3/2}}{3a^3} \leq 0. \quad (2.12)$$

Proof. Follows from the Lemma 2.2.5 and Lemma 2.2.6, since $\frac{\partial f_2}{\partial a} \leq 0$ at both ends of the interval $[a^{-1}, a]$ and $\frac{\partial^3 f_2}{\partial a \partial r^2} > 0$ on this interval, we can conclude that $\frac{\partial f_2}{\partial a} \leq 0$ over the entire interval. \square

Lemma 2.2.5. *For $r \in [a^{-1}, a]$ (case 2), we have that $f_a \leq 0$ at the endpoints of the interval, with equality only when $a = 1$*

$$\frac{\partial f_2}{\partial a}(a, a^{-1}) = \frac{4(a^{-5} - a^{-3})}{3} \leq 0, \quad (2.13)$$

$$\frac{\partial f_2}{\partial a}(a, a) = \frac{-4a^8 + 6a^4 - 2 + (4a^6 - 4a^2)\sqrt{a^4 - 1}}{3a^5} \leq 0. \quad (2.14)$$

Proof. The case of $r = a^{-1}$ is trivial, especially since the partial derivative is continuous and so lemma 2.2.3 applies, and is it straightforward to see that $f_a(1, r) = 0$ here. Next, we investigate $r = a$, and we establish that for $a > 1$ it is negative for at least one value

2. Structure and Connectivity of Random Rectangular Graphs

of a and has no roots. First,

$$\frac{\partial f_2}{\partial a} \left(2^{1/4}, 2^{1/4} \right) = \frac{2^{5/4}}{3} - \frac{2^{3/4}}{2} \approx -0.048 < 0. \quad (2.15)$$

Now, we set $f_a(a, a) = 0$, and after some algebraic manipulation we get

$$(a^4 - 1)^2 = 0, \quad (2.16)$$

which indeed has no roots for $a > 1$, and the result follows. \square

Lemma 2.2.6. *For $r \in [a^{-1}, a]$ (case 2), we have that $f_{a,r} > 0$*

$$\frac{\partial^3 f_2}{\partial a \partial r^2} = \frac{8a^4 r^2 - 4a^2 + (4 - 8a^3 r) \sqrt{a^2 r^2 - 1}}{a^3 \sqrt{a^2 r^2 - 1}} > 0. \quad (2.17)$$

Proof. We simplify the analysis by first multiplying by the denominator, and dividing by 4 to simplify the constants: call this $y^*(a, r)$. The denominator equals 0 when $r = a^{-1}$, but this is not a problem as the expression simply tends to infinity as $r \rightarrow a^{-1}$ and the denominator is non-negative otherwise. We set

$$\alpha(a, r) = \sqrt{a^2 r^2 - 1}, \quad (2.18)$$

$$\beta(a, r) = 2a^4 r^2 - a^2 - 2a^3 r \sqrt{a^2 r^2 - 1}, \text{ then} \quad (2.19)$$

$$y^* = \alpha + \beta. \quad (2.20)$$

It is obvious that $\alpha \geq 0$, now we show that $\beta > 0$. After some algebraic manipulation, we find that $\beta = 0 \Rightarrow a^4 = 0$, which has no roots for $a > 1$ and therefore β cannot change sign, and $\beta(\sqrt{2}, 1) = 6 - 4\sqrt{2} \approx 0.3432 > 0$. Therefore, α is non-negative and β is positive, so $y^* > 0$ and we have the required result. \square

Lemma 2.2.7. *For $r \in [a, \sqrt{a^2 + a^{-2}}]$ (case 3), we have that $f_a \leq 0$ with equality only*

2. Structure and Connectivity of Random Rectangular Graphs

when $a = 1$ or $r = \sqrt{a^2 + a^{-2}}$

$$\frac{\partial f_3}{\partial a} = 2r^2 \left(\frac{1}{a^3} - a \right) + \frac{2}{3} \left(a^3 - \frac{1}{a^5} \right) + \frac{4(a^2 r^2 - 1)^{3/2}}{3a^3} - \frac{4(r^2 - a^2)^{3/2}}{3a^2} \leq 0, \quad (2.21)$$

Proof. Follows from lemma 2.2.8, since it is easily verified that $f_a(1, r) = 0$ and

$$\frac{\partial f_3}{\partial a}(a, \sqrt{a^2 + a^{-2}}) = 0. \quad (2.22)$$

□

Lemma 2.2.8. For $r \in [a, \sqrt{a^2 + a^{-2}}]$ (case 3), we have that $f_{a,r} \geq 0$ with equality only when $a = 1$ or $r = \sqrt{a^2 + a^{-2}}$

$$\frac{\partial^2 f_3}{\partial a \partial r} = -\frac{4r(a^4 - 1 - a^2 \sqrt{a^2 r^2 - 1} + a \sqrt{r^2 - a^2})}{a^3} \geq 0 \quad (2.23)$$

Proof. We demonstrate that this has no roots inside the interval $(a, \sqrt{a^2 + a^{-2}})$ and is positive somewhere. First we note that

$$\frac{\partial^2 f_3}{\partial a \partial r}(1.1, 1.2) \approx 0.184 > 0. \quad (2.24)$$

For finding the roots, after setting $f_{a,r} = 0$ and some manipulation, we get

$$(a - 1)^2 (a^4 - a^2 r^2 + 1)^2 = 0, \quad (2.25)$$

which only has positive roots at $a = 1$ and $r = \sqrt{a^2 + a^{-2}}$. □

2.2.2 Degree Distribution

In the RRG the n nodes are distributed uniformly and independently on the unit rectangle. Then, the degree distribution can be easily estimated by considering the probability density function of a node v_i having degree k_i given that there are $n - 1$

2. Structure and Connectivity of Random Rectangular Graphs

other nodes uniformly distributed in the unit rectangle (see for instance [8]). This gives rise to the binomial distribution, which when $(n - 1) \approx n$ is given by

$$\mathbf{P}(k_i = k) = \binom{n}{k} A(v_i)^k (1 - A(v_i))^{n-k}. \quad (2.26)$$

When n is large and r is sufficiently small, $A(v_i)$ takes the same value for most of the nodes and this binomial distribution approaches a Poisson distribution of the form [58]

$$p(k) \approx \frac{\bar{k}^k \exp(-\bar{k})}{k!}. \quad (2.27)$$

As we have previously obtained an analytic expression for the average degree \bar{k} we can easily compute the degree distribution for RRGs. We select RRGs with 5000 nodes and radius of connection equal to 0.025. Then, we obtain the degree distribution for different values of the rectangle side length and take the average of 100 random realisations. In Fig. 2.5(a) we also plot the expected distribution using the equation (2.27) in which we have plugged the values of the expected average degree obtained previously. As can be seen, independently of the side length of the rectangle the RRG displays Poisson degree distributions. That is, the elongation of the rectangle does not affect the degree distribution of the nodes from being Poisson.

However, if r is not small, then the distribution may not be well approximated by a Poisson distribution. Then, we may naively compute the degree distribution numerically as

$$p(k) = \int_{\mathcal{R}} \binom{n-1}{k} A(x)^k (1 - A(x))^{n-k-1} dx. \quad (2.28)$$

This is reasonable to compute in practice, and we demonstrate the resulting distribution for an RGG in Fig. 2.5(b). It is clear that this is not a Poisson distribution as it is bimodal. The shape of the distribution may be understood from the fact that, if r is large, there will be relatively many nodes that lie within distance r of the boundary and will have a lower degree on average than the nodes in the bulk, where the bulk is the area of the rectangle that does not lie within distance r of the boundary. The degree distribution of nodes in the bulk will have a peak at a larger value of k than for nodes near the boundary, and both make significant contributions to the overall degree

2. Structure and Connectivity of Random Rectangular Graphs

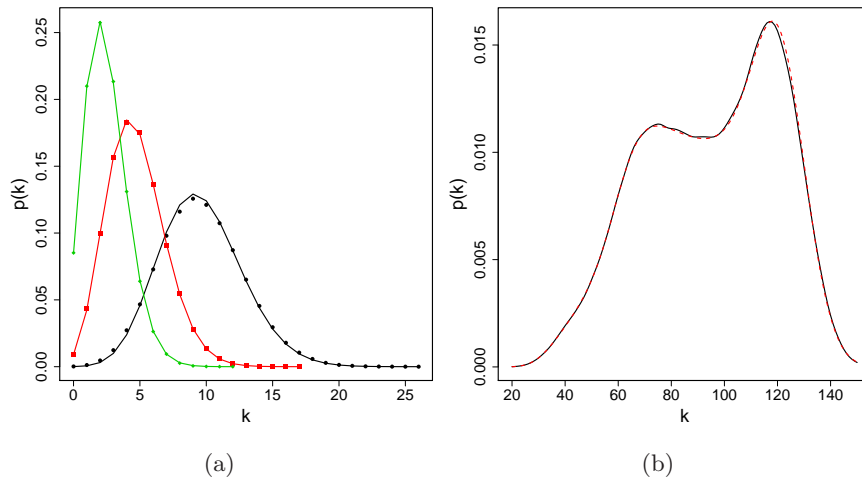


Figure 2.5: (a) Degree distribution of RRGs with $n = 5000$, connection radius $r = 0.025$, and rectangle side lengths $a = 1$ (black circles), $a = 50$ (red squares), $a = 100$ (green diamonds). The circles, squares, and diamonds correspond to the average of 100 random realisations for the given network. The solid lines correspond to the shape of the Poisson distribution (2.27) with the corresponding average degree obtained from Eq. (2.2). (b) Degree distribution of an RRG with $n = 500$ and connection radius $r = 0.28$. The black line corresponds to the observed degree distribution and the red dashed line corresponds to the numerical approximation from Eq. 2.28

distribution. As a result, the degree distribution has a shape that is non-trivial and clearly distinct from the Poisson distribution.

2.2.3 Diameter and Average Path Length

We now consider the diameter d_{max} , which is of course closely related to the average shortest path length \bar{l} . This bound will be useful when describing the algebraic connectivity of the RRG, which in turn is very important for describing the behaviour of diffusive processes. We are careful here to ensure that r is large enough that the RRG is connected. For given values of a and n , there is a critical value of the connection radius which ensures that the RRG is connected with high probability. Since we only want to examine the diameter in the connected regime, we only consider values of r above this threshold so that the RRGs are connected. Connectedness of RRGs is discussed further in subsection 2.2.8.

2. Structure and Connectivity of Random Rectangular Graphs

Theorem 2.2.9. *The diameter $d_{max}(G_{RRG})$ of an RRG is bounded from below as [58]*

$$d_{max}(G_{RRG}) \geq \frac{\sqrt{a^2 + b^2}}{r}. \quad (2.29)$$

Proof. The nodes of the RRG are uniformly and independently distributed in the unit rectangle. We consider the expected eccentricity of a node v_c located exactly at a corner, and let $c = \sqrt{a^2 + b^2}$. To reach a node located at the opposite corner of the rectangle, a path must take at least $\lceil c/r \rceil$ steps since each adjacent pair of nodes in the path has length at most r , which is a lower bound on the expected eccentricity of v_c and therefore the diameter. \square

The tightness of this bound is demonstrated in Fig. 2.6(a). We use logarithmic axes to help illustrate the behaviour, as it demonstrates that the bound is then a straight line. As expected, the diameter of the RRG decreases as we increase the radius and increases as we increase the elongation of the rectangle, and the bound seems to perform well. In the worst case here, for very small r and $a = 1$, the bound is about 28, which is quite close to the correct value of about 35. For larger values of r the bound is even better. Figure 2.6(b) shows the behaviour of the diameter and the bound for small RRGs with only $n = 100$ nodes, and so we conclude that the bound is tight. We note the interesting fact that as the connection radius is increased, the diameter does not always vary smoothly, but rather forms a step-like shape especially when the diameter is small.

The average shortest path length is related to the diameter, and so the bounds are constructed in a similar way.

Theorem 2.2.10. *Let $c = \sqrt{a^2 + b^2}$ be the length of the diagonal of the rectangle \mathcal{R} , $c_r = \lceil c/r \rceil$ be the smallest multiple of r that exceeds the length c , and $A_c(s)$ be the area that lies within the rectangle and is within distance s of a specified corner. Then, the*

2. Structure and Connectivity of Random Rectangular Graphs

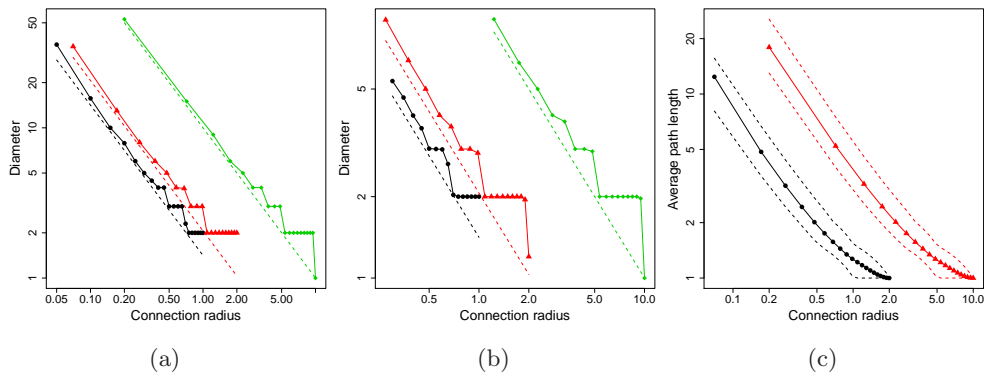


Figure 2.6: Change in the diameter of RRGs with $a = 1$ (black circles) and $a = 2$ (red triangles) and $a = 10$ (green diamonds), with (a) $n = 100$ and (b) $n = 1500$ nodes, for both observed values (solid lines) and bounds (dashed lines), as the connection radius r is varied (log-log scale). The number of random realisations used for each choice of parameter values was 20 and 100 respectively. (c) Change in the average path length for RRGs with $n = 1500$ nodes for $a = 1$ (black circles) and $a = 2$ (red triangles), for both observed values (solid lines) and bounds (dashed lines), as the connection radius r is varied (log-log scale). 20 random realisations were used for each choice of parameter values.

expected value of the average path length \bar{l} of an RRG is bounded from above as

$$\mathbf{E}(\bar{l}) \leq c_r - \sum_{j=1}^{c_r-1} A_c(jr). \quad (2.30)$$

Proof. We again consider that the n points are distributed homogeneously in the rectangle, and that we are in a regime where the RRGs are connected. We expect the main contribution to the average path length bound to be from nodes located near the corners of the rectangle. Thus, we approximate the expected average path length l_c of a node located exactly at a corner v_c , by assuming that the length of the shortest path between two nodes v_c and v_i is equal to the smallest multiple of r that exceeds the Euclidean distance between the nodes $|v_c - v_i|$. This gives a bound on the average path length \bar{l} . Since we are already overestimating the value of \bar{l} by considering the largest contribution possible from a node at a vertex, the error introduced from the

2. Structure and Connectivity of Random Rectangular Graphs

approximation is not problematic. Then we can write

$$\begin{aligned}
\mathbf{E}(\bar{l}) &\leq \mathbf{E}(l_c) \\
&= \mathbf{E}\left(\sum_{j=1}^{c_r} j \mathbf{P}((j-1)r < |v_c - v_i| < jr)\right) \\
&= \sum_{j=1}^{c_r} j(A_c(jr) - A_c((j-1)r)) \\
&= c_r A_c(c_r r) - \sum_{j=1}^{c_r-1} A_c(jr) \\
&= c_r - \sum_{j=1}^{c_r-1} A_c(jr).
\end{aligned} \tag{2.31}$$

□

By a similar reasoning, we can obtain a lower bound on the average path length by considering the average path length l_m of a node placed in the middle of the rectangle v_m with coordinates $(a/2, b/2)$. Let $m_r = \lceil c/(2r) \rceil$ and define $A_m(s)$ to be the area that lies within the rectangle and is within distance s of the centre of the rectangle.

Theorem 2.2.11. *The expected value of the average path length \bar{l} of an RRG is bounded from below as*

$$\mathbf{E}(\bar{l}) \geq m_r - \sum_{j=1}^{m_r-1} A_m(jr). \tag{2.32}$$

Proof. The proof is analogous to the upper bound. We consider a node v_m located in the centre of the rectangle and such a node has the lowest expected average path

2. Structure and Connectivity of Random Rectangular Graphs

length, and proceed as expected

$$\begin{aligned}
\mathbf{E}(\bar{l}) &\geq \mathbf{E}(l_m) \\
&= \mathbf{E}\left(\sum_{j=1}^{m_r} j \mathbf{P}((j-1)r < |v_m - v_i| < jr)\right) \\
&= \sum_{j=1}^{m_r} j (A_m(jr) - A_m((j-1)r)) \\
&= m_r A_m(m_r r) - \sum_{j=1}^{m_r-1} A_m(jr) \\
&= m_r - \sum_{j=1}^{m_r-1} A_m(jr).
\end{aligned} \tag{2.33}$$

□

In Fig. 2.6(c) we illustrate the variation of \bar{l} with the connection radius for $a = 2$ and $a = 10$. In the same plot we illustrate the values of the upper bound using Eq. 2.30, and the lower bound obtained with Eq. 2.32. It can be seen that the bounds are very close to the average shortest path obtained for these RRGs. Particularly, for large values of r the observed values are almost identical to those of the upper and lower bounds.

2.2.4 Clustering Coefficient

Here we are interested in an expression that accounts for the variations of the clustering coefficient with both the connection radius and the rectangle side length, with a finite number of nodes. Since finding a good analytical approximation here is difficult, we instead present two observations which capture some crucial aspect of the behaviour of the clustering coefficient, and we hope that these can give some useful insights and help guide future work.

Observation 2.2.12. The average clustering coefficient $\bar{C}(G_{RRG})$ may be approximated as [58]

$$\bar{C} \approx \frac{2r^2 \arccos\left(\frac{\delta}{2r}\right) - \frac{1}{2}\delta\sqrt{4r^2 - \delta^2}}{\pi r^2}, \delta = an^{-\frac{a}{a+b}}. \tag{2.34}$$

2. Structure and Connectivity of Random Rectangular Graphs

Since this is undefined for $r < \delta/2$, we set $\bar{C} \approx 0$ in this range.

Justification. Our strategy is similar to the one used in [34]. That is, let v_i and v_j be two connected nodes in a RRG that are separated at a Euclidean distance δ from each other. Let us draw two circles of radius r centred respectively at v_i and v_j . Because v_i and v_j are connected, any point in the area formed by the overlap of the two circles will form a triangle with the nodes v_i and v_j . In addition, any node inside the two circles that is not in the overlapping area forms a path of length two with the nodes v_i and v_j . Thus if we quantify the ratio of the overlapping area to the total area of the circle we account for the ratio of the number of triangles to open triads in which the nodes v_i and v_j take place, i.e., the clustering coefficient. The boundary will have an effect here, since these areas may lie partially outside of the rectangle, but this is difficult to account for and so we ignore it here for simplicity. This ratio is given by

$$\frac{2r^2 \arccos\left(\frac{\delta}{2r}\right) - \frac{1}{2}\delta\sqrt{4r^2 - \delta^2}}{\pi r^2}. \quad (2.35)$$

At this point we only need an estimation of the length δ between two connected nodes in a RRG. Since this is just a crude estimate we use a simple approach based on the following intuition. Let us start by considering n nodes in a square in such a way that they form a regular square lattice. Then, δ is proportional to the length of the side of the rectangle a divided by the number of circles along this side. As we have a square, the number of points along the side of length a is the same as that for the other side. Consequently, $\delta \sim n^{-1/2}$. If we elongate the rectangle to $a \rightarrow \infty$, which resembles a straight line, we will have that the separation between the two points is just the length of the straight line divided by the number of nodes, $\delta \sim an^{-1}$. For a general rectangle the separation between two points in a line along the edge side of length a is given by $\delta \sim an^{-\gamma}$, where $\gamma \sim a(a+b)^{-1}$. Notice that when $a = 1$ ($b = 1$) we have $\delta \sim n^{-1/2}$ and when $a \rightarrow \infty$ ($b \rightarrow 0$) we have $\delta \sim an^{-1}$. \square

Observation of the clustering coefficient in RRGs gives some surprising results. In Fig. 2.7(a) we illustrate the dependence of \bar{C} on a for different radii based on Eq. 2.34,

2. Structure and Connectivity of Random Rectangular Graphs

with the estimated value of δ given before. Notably, the clustering coefficient is predicted to change non-monotonically with the rectangle side length. Instead, for small values of a the clustering coefficient is predicted to increase to a maximum value and after it the clustering decreases linearly. In addition, according to this model, as the connection radius increases the clustering coefficient is expected to increase for the same value of a . It is this qualitative behaviour, especially the non-monotonicity, that we are trying to capture here. We expect a more sophisticated version of this approach to capture this behaviour more accurately, such as by taking into account boundary effects and pair-distance statistics.

In order to demonstrate these findings we compute the average clustering coefficient of RRGs and compare these to the analytical formulas. The results of the variation of the clustering with the rectangle elongation are illustrated in Fig. 2.7(a). As can be seen the clustering increases up to a maximum, whose location depends on the connection radius, and then decays with the increase of the elongation of the rectangle. We have not been able to capture the dependence of δ with the radius in our previous reasoning, but we have captured the behaviour of the clustering of having a non-monotonic change with a . Also, these experiments show that the increase of the connection radius increases the average clustering coefficient as predicted by our analytical results. As can be seen in Fig. 2.7(b) for $a = 1$ and small radius the average clustering coefficient is $\bar{C} \approx 0.61$, which is very close to the expected value for the 2-dimensional RGG according to [34]. When $a = 30$ and the radius is relatively large, the average clustering coefficient is $\bar{C} \approx 0.75$, which coincides with the exact value expected for the one-dimensional RGG according to [34]. Consequently, the RRG generalises the values of the clustering coefficient of both the one- and two-dimensional RGG, for $a = 1$ and $a \rightarrow \infty$, respectively. In addition, it provides a series of intermediate values of the clustering coefficient for intermediate values of the side length of the rectangle.

We now further explore the relation between the radius r and the clustering for RRGs with different side lengths. We consider graphs with $n = 1500$ nodes and $a = 1, 5, 10, 30$. As the radius increases the graph is becoming more and more dense, which means that the clustering coefficient is characterised by an abrupt increase at the

2. Structure and Connectivity of Random Rectangular Graphs

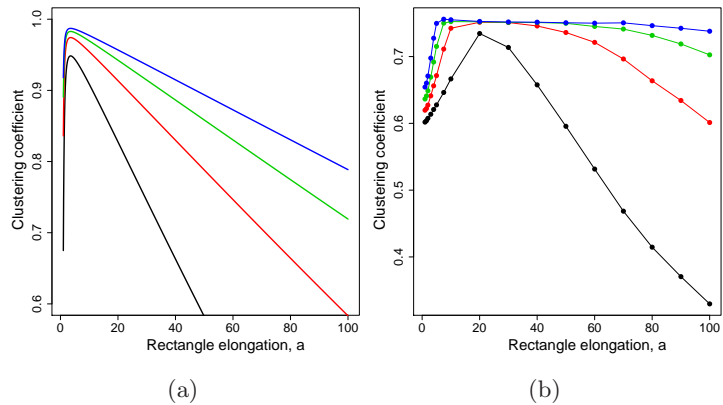


Figure 2.7: Illustration of the dependence of the clustering coefficient with the rectangle side length for different connection radii: $r = 0.05$ (black), $r = 0.1$ (red), $r = 0.15$ (green) $r = 0.20$ (blue). In plot (a) we show the analytical results and in plot (b) the numerical ones. Both results are obtained for RRG with $n = 1500$ nodes, and for the observed results we averaged the results of 100 random realisations.

beginning of the plot and then a mostly linear increase until the value of $\bar{C} = 1$ is reached for the complete graph (see Fig. 2.8).

This guides our second observation, which is to try to approximate the clustering coefficient using the observed trends in a piecewise fashion. Recalling Eq. 1.36 (page 24),

$\bar{C}_2 = 1 - \frac{3\sqrt{3}}{4\pi} \approx 0.5865$ and $\bar{C}_1 = 0.75$, and we refer to our previous approximation of Eq. 2.34 as \bar{C}_{nm} for the non-monotonic motivation behind it. We see that for $a = 1$ the clustering increases sharply as we increase r until $r = r^*$ where $\bar{C} \approx \bar{C}_2$, and \bar{C}_{nm} is a reasonable approximation in this regime. Then, it increases in an almost linear way until $\bar{C} \approx 1$ at $r \approx a = 1$.

For larger values of a , the initial sharp increase is almost identical. Then, it increases almost linearly to $\bar{C} \approx \bar{C}_1$ at $r \approx b$, and then to $\bar{C} \approx 1$ at $r \approx a$.

Observation 2.2.13. The average clustering coefficient $\bar{C}(G_{RRG})$ may be approximated

2. Structure and Connectivity of Random Rectangular Graphs

as

$$\bar{C} \approx \begin{cases} 0 & r \leq \delta/2 \\ \bar{C}_{nm} & \delta/2 \leq r \leq r^* \\ \bar{C}_2 + \frac{(r - r^*)(1 - \bar{C}_2)}{a - r^*} & r^* \leq r \leq a \\ 1 & a \leq r \end{cases} \quad (2.36)$$

for $a = 1$, and

$$\bar{C} \approx \begin{cases} 0 & r \leq \delta/2 \\ \bar{C}_{nm} & \delta/2 \leq r \leq r^* \\ \bar{C}_2 + \frac{(r - r^*)(\bar{C}_1 - \bar{C}_2)}{b - r^*} & r^* \leq r \leq b \\ \bar{C}_1 + \frac{(r - b)(1 - \bar{C}_1)}{a - b} & b \leq r \leq a \\ 1 & a \leq r \end{cases} \quad (2.37)$$

for large values of a . We note that this does not provide a smooth transition from $a = 1$ to large values of a , and has the same problem as \bar{C}_{nm} for very small values of r . If these shortcomings can be addressed then we expect this to be a practical way to approximate the value of the clustering coefficient for RRGs.

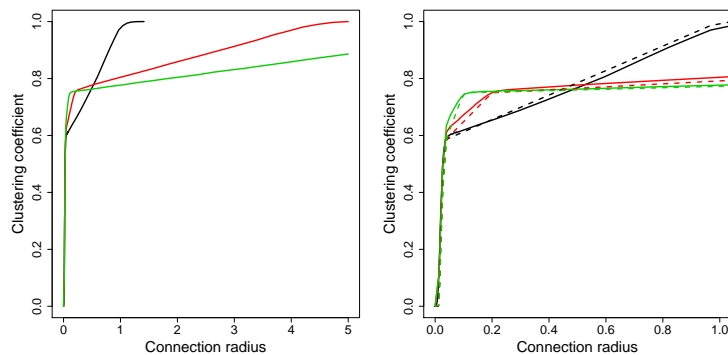


Figure 2.8: (a) Variation of the clustering coefficient with the radius for RRGs with $n = 1500$ nodes and $a = 1$ (black), $a = 5$ (red) and $a = 30$ (green). Every point is the average of 100 random realisations. (b) Demonstration of the same data with approximation of Observation 2.2.13 (dashed lines), with a smaller range of radii for clarity.

2. Structure and Connectivity of Random Rectangular Graphs

2.2.5 Algebraic Connectivity

Since the algebraic connectivity is such a useful concept for network dynamics, we provide an upper bound on its expected value.

Theorem 2.2.14. *The algebraic connectivity $\mu_2(G_{RRG})$ of an RRG is bounded as [59]*

$$\mu_2(G_{RRG}) \leq \frac{8(n-1)r^2}{a^2+b^2} \log_2^2 n. \quad (2.38)$$

Proof. In order to prove Theorem 2.2.14 we need the following result by Alon and Milman for the algebraic connectivity of any simple graph.

Theorem 2.2.15. *From [3]: the second smallest eigenvalue of the Laplacian matrix of any graph is bounded as*

$$\mu_2(G) \leq \frac{8k_{max}}{d_{max}^2} \log_2^2 n. \quad (2.39)$$

Then, by substituting 2.29 into 2.39 we have

$$\mu_2(G_{RRG}) \leq \frac{8k_{max}}{d_{max}^2} \log_2^2 n \leq \frac{8k_{max}r^2}{a^2+b^2} \log_2^2 n \leq \frac{8(n-1)r^2}{a^2+b^2} \log_2^2 n, \quad (2.40)$$

where the last inequality uses the fact that for any simple graph $k_{max} \leq n-1$. \square

In Fig. 2.9 we demonstrate this bound. We find that while it is not tight, it does appear to follow the same trend as the observed values. Then, we conclude that if we elongate the rectangle, which causes the diameter to increase, the algebraic connectivity will decay in the expected way.

2.2.6 Spectral Radius

We consider the following well-known bounds for the largest eigenvalue of the adjacency matrix of a simple graph λ_1 , which we recall is also the spectral radius

$$k_{min} \leq \bar{k} \leq \lambda_1 \leq k_{max}. \quad (2.41)$$

2. Structure and Connectivity of Random Rectangular Graphs

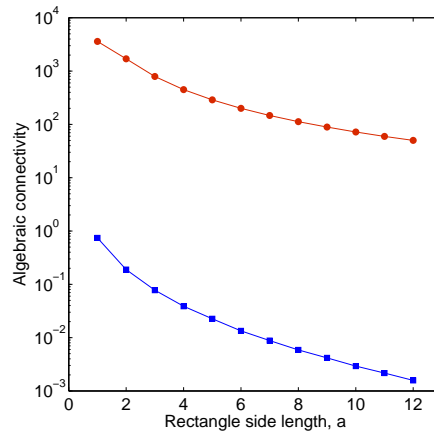


Figure 2.9: Variation of the algebraic connectivity with elongation for RRGs with $n = 500$ nodes and $r = 0.15$: observed values (blue squares) and bound of Eq. 2.38 (red circles). Every point is the average of 100 random realisations.

Let us now consider what happens to λ_1 when we elongate the rectangle. We have already proven in Theorem 2.2.2 that when a increases, \bar{k} does not increase, and in most cases decreases. Consequently, the spectral radius of the adjacency matrix λ_1 also decreases. Strictly speaking, proving that the average degree decreases with the elongation of the rectangle is not a proof that the spectral radius also decreases. However, it is possible to infer such a relation between the elongation and the spectral radius as follows. In particular, we are interested here in showing whether the average degree and the spectral radius of RRGs show the same trend when the rectangle is elongated. In Fig. 2.10 we illustrate the plot of the spectral radius versus the average degree for RRGs with $a = 1$ (left), $a = 30$ (centre) and $a = 1, 2.5, 5, 7.7, 15, 20, 25, 30$ (right) for different values of the connection radius. As can be seen in all cases, and particularly in the last one, the trend of the spectral radius and the average degree is exactly the same and indeed they are very highly linearly correlated. Thus, our conclusion here is that proving that the average degree has a certain behaviour when the rectangle is elongated can be directly extrapolated to the behaviour of the spectral radius.

2. Structure and Connectivity of Random Rectangular Graphs

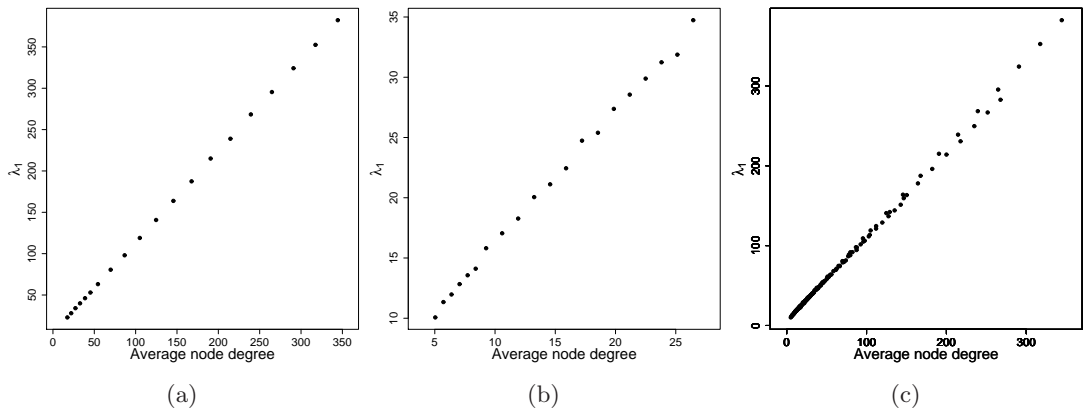


Figure 2.10: Scatter plots of the spectral radius versus the average degree for RRGs with $a = 1$ (a), $a = 30$ (b) and $a = 1, 2.5, 5, 7.5, 10, 15, 20, 25, 30$ (c) for different values of the connection radius.

2.2.7 The λ_{-1} Eigenvalue

It has been noted that the spectrum $\sigma(A)$ of the adjacency matrix of an RGG has an obvious peak at -1 , and in particular a high multiplicity of $\lambda = -1$. We denote this multiplicity as m_{-1} . It has been proven that in the infinite limit of $n \rightarrow \infty$ there is a singularity in the spectrum at -1 [16]. This is attributed to the presence of symmetric subgraphs [107, 108, 41]. In particular, pairs of nodes that are connected and which may be permuted without changing the structure of the graph contribute to m_{-1} . It follows readily from the definition that the property of a pair of nodes having this symmetry is transitive, that is if v_p and v_q are symmetric and v_p and v_r are symmetric, then v_q and v_r are symmetric. Then, the subgraph induced by all the nodes that have this symmetry form disjoint cliques. If there are n_i nodes in clique i , then it contributes $(n_i - 1)$ to the value of m_{-1} . This is shown by a simple construction. Let nodes v_1 and v_2 be symmetric, then we construct a vector \mathbf{v} with entries

$$\mathbf{v}_j = \begin{cases} 1 & \text{if } j = p \\ -1 & \text{if } j = q \\ 0 & \text{otherwise} \end{cases} \quad (2.42)$$

By construction, $(A\mathbf{v})_p = \mathbf{v}_q = -\mathbf{v}_p$, $(A\mathbf{v})_q = \mathbf{v}_p = -\mathbf{v}_q$, and $(A\mathbf{v})_i = 0$ otherwise

2. Structure and Connectivity of Random Rectangular Graphs

since every other node is connected to both p and q or neither, and so, $A\mathbf{v} = -\mathbf{v}$. Then, we have constructed an eigenvector corresponding to an eigenvalue of -1 . In clique i of mutually symmetric nodes we may construct $(n_i - 1)$ linearly independent eigenvectors in this manner, with each having only two non-zero entries corresponding to a pair of symmetric nodes, each of these contributes to the multiplicity of the eigenvalue. Note that the complete graph K_n is well-known to have spectrum $[n - 1, (-1)^{n-1}]$, which is explained by the above construction and the fact that the graph is regular with all node degrees equal to $(n - 1)$. In general, m_{-1} is more subtle as other, higher-order structures containing more than two nodes also contribute to its value. For now, we give a general construction, then we use an example of such a structure that is relevant to the case of the RRG to demonstrate some of the characteristics of these structures.

Lemma 2.2.16. *Let $V = V_0 \cup V_1 \cup V_2$ be a partition of the nodes of a graph into three disjoint subsets, such that only V_0 may be empty. Such a partition allows a construction of an eigenvector with eigenvalue -1 if the following holds*

$$\sum_{j \in V_1} A_{ij} = \sum_{k \in V_2} A_{ik} - \delta(i), \text{ for each } v_i \in V, \text{ where} \quad (2.43)$$

$$\delta(v_i) = \begin{cases} 0 & \text{if } v_i \in V_0 \\ 1 & \text{if } v_i \in V_1 \\ -1 & \text{if } v_i \in V_2 \end{cases}. \quad (2.44)$$

Proof. We construct a vector with entries $\mathbf{v}_i = \delta(i)$. By construction, $A\mathbf{v} = -\mathbf{v}$, and m_{-1} is at least the number of linearly independent eigenvectors that can be constructed. \square

Remark 2.2.17. We get structures relating to the multiplicity of $\lambda = 0$, the nullity of the graph, if we take the above construction and modify the condition to be

$$\sum_{j \in V_1} A_{ij} = \sum_{k \in V_2} A_{ik}, \text{ for each } v_i \in V. \quad (2.45)$$

From the arguments we present relating to m_{-1} , we may expect to find such struc-

2. Structure and Connectivity of Random Rectangular Graphs

tures in an RRG since it also deals with a symmetry between nodes. However, they appear to be less frequent.

Figure 2.11 shows a simple example of such a higher-order structure, with the three subsets of nodes highlighted in different colours, which we can observe in RRGs. We refer to this and similar structures as a *neighbourhood chain*, which we describe and explain gradually. We now discuss why $m_{-1} = 2$ in this example. Nodes v_1 and v_2 are symmetric and so we have the eigenvector $\mathbf{v}_1 = (1, -1, 0, 0, 0, 0)^T$, and the chain highlighted in the figure gives an eigenvector $\mathbf{v}_2 = (1, 0, -1, 0, 1, -1)^T$. Therefore, we have two linearly independent eigenvectors and $m_{-1} \geq 2$. Of course, we may instead include node v_2 in V_1 instead of node v_1 giving a third eigenvector \mathbf{v}_3 , but it is clear that this is not linearly independent of the other two since $\mathbf{v}_3 = \mathbf{v}_2 - \mathbf{v}_1$. Therefore, we can completely explain each contribution to the value of m_{-1} in this case. In an RRG where the value of r is not too large and it is disconnected, we can often observe this subgraph as a component of the graph. Note that nodes v_1 and v_2 form a complete graph K_2 , and node v_6 can be considered as K_1 . Then, we may replace these with any complete graphs K_{n_1} and K_{n_2} , and we observe that in such a graph we have $m_{-1} = (n_1 - 1) + (n_2 - 1) + 1 = n_1 + n_2 - 1$, as expected by considering the two subsets of symmetric nodes and the neighbourhood chain we have described here. Also, we may include more nodes that have the same topological role as node v_4 , that is they connect to nodes v_3 and v_5 and no other nodes where \mathbf{v}_2 is non-zero. We may also introduce any number of nodes that connect to nodes v_1 and v_6 in a similar way; the cyclic graph C_6 has $m_{-1} = 2$ for this reason, taking into account symmetry and the requirement for linear independence of the eigenvectors.

This neighbourhood chain is not unrelated conceptually to symmetric nodes. Let $N'(v_i)$ be the closed neighbourhood of v_i , that is the set of nodes adjacent to v_i and including v_i itself. Then, for symmetric nodes we have $N'(v_i) = N'(v_j)$, and so we can consider that the eigenvalue -1 is capturing some correlation between the neighbourhoods of adjacent nodes. Now, note that in Fig. 2.11 we have $N'(v_3) \setminus N'(v_1) = N'(v_5) \setminus N'(v_6) = v_4$, where $X \setminus Y$ is the set difference operator. This is not a coincidence, and considering the product $A\mathbf{v}_2$ in terms of the nodes of the graph make this

2. Structure and Connectivity of Random Rectangular Graphs

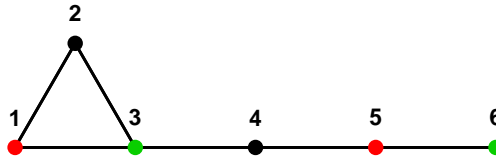


Figure 2.11: Example of a small graph with a neighbourhood chain that contributes to m_{-1} , where V_0 is black, V_1 is red and V_2 is green.

clear: the value of the vector $A\mathbf{v}_2$ at a given node is equal to the sum of the values of \mathbf{v}_2 over all adjacent nodes. Then, there is still some correlation between the neighbourhoods of adjacent nodes, but also a longer-range correlation between the nodes that are almost symmetric, and frustrated by the same node (or nodes, in general). Both symmetric nodes and these correlations between neighbourhoods of adjacent nodes across larger scales are what $\lambda = -1$ seems to be detecting.

Further, this can be easily extended to similar structures with more non-zero elements in the eigenvector. In the path graph P_{3k-1} for $k > 1$, we can have the vector $\mathbf{v} = (1, -1, 0, \dots, 0, 1, -1)$, and clearly $A\mathbf{v} = -\mathbf{v}$ so this is an eigenvector with eigenvalue -1 . Additional nodes may be added to such a path graph as discussed in the previous paragraph, and these structures may indeed be found in RRGs. This is because, by construction, nodes that are in close proximity are connected, and there is clearly some correlation between the other nodes they are connected to. It is then common that either their neighbourhoods are exactly the same and the nodes are therefore symmetric, or they are very similar. If their neighbourhood is only similar, we may find another pair of nodes not too far away that also have similar neighbourhoods, with a small set of nodes lying between these two pairs and frustrating the symmetry of both. Therefore, we have explained the nature of the structures that contribute to m_{-1} and why they are so common in RRGs. We take this further, and demonstrate this excess in neighbourhood correlations in more detail.

We consider the peak at -1 in the spectrum of the adjacency matrix of an RRG in two ways. First, we can simply count the multiplicity of $\lambda = -1$. Secondly, we

2. Structure and Connectivity of Random Rectangular Graphs

introduce an index that will take a large value if the spectrum of a graph is concentrated at -1 and a small value if it is not. Recall that the Estrada index $EE(G)$ takes the trace of the exponential of the adjacency matrix, which can be written as the sum of the exponentials of the eigenvalues. This sum is dominated by the most positive eigenvalues. The work in [51] considers an index that is similar to the Estrada index, but instead computes

$$\tilde{G} = \text{tr}(\exp(-A^2)) = \sum_{i=1}^n \exp(-\lambda_i^2). \quad (2.46)$$

This is dominated by the part of the spectrum near 0, since of course the largest value $\exp(-\lambda_i^2)$ can take is 1 when $\lambda = 0$, and so the idea is to ‘fold’ the spectrum around 0. It is then simple to consider

$$\tilde{G}_x = \text{tr}(\exp(-(A - xI)^2)) = \sum_{i=1}^n \exp(-(\lambda_i - x)^2), \quad (2.47)$$

and we are of course interested here in \tilde{G}_{-1} . We compute the value of this index for the RRG for various elongations and across the range of possible connection radii, with the average taken over many random realisations, and show the results in Fig. 2.13. We also do this for the ER graph with p varying from 0 to 1, and also for the BA and RRNG graphs. A similar analysis is performed for the multiplicity of $\lambda = -1$. We consider RRNGs in this analysis since it is mainly concerned with RRGs, rather than having a small separate discussion in the next chapter where we focus on RRNGs.

It is clear that the RRG has on average significantly larger values of m_{-1} and \tilde{G}_{-1} , which we expect from our previous analysis. The ER and BA graphs do not have any spatial aspect, and so do not have the correlations between neighbourhoods of nodes that we find in the RRG. It might be expected that the RRNG, which is also spatially defined, would also have a large value of m_{-1} from symmetric pairs of nodes, but this is not what is observed. The reason for this is that when two adjacent nodes v_i and v_j have the same neighbourhood and are connected to at least one other node v_k , then v_i , v_j and v_k form a triangle. Since RRNGs are connected and we do not concern ourselves with RRNGs on only two nodes, any two adjacent nodes that have the same

2. Structure and Connectivity of Random Rectangular Graphs

neighbourhood must take part in at least one triangle. However, RRNGs constructed on a set of randomly distributed points almost never contain a triangle, and therefore symmetric nodes are practically impossible. Note, however, that higher order structures contributing to m_{-1} are more plausible but not abundant. For example, we may find a chordless 6-cycle in which only one pair of opposite nodes is connected to any other node in the graph, as in Fig. 2.12.

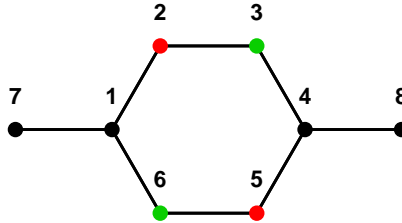


Figure 2.12: Example of a subgraph based on C_6 that contributes to m_{-1} in RRNGs, where V_0 is black, V_1 is red and V_2 is green as per Fig. 2.11.

We show in Fig. 2.13 that the values of m_{-1} and \tilde{G}_{-1} are higher in the RGG than in other random graph models, where we put density on the x -axis to make a comparison fair. We also show that in the RRG, these values typically increase as the rectangle is elongated. This can be explained by the fact that if a pair of adjacent nodes are not symmetric, the nodes that frustrate this will be near the boundaries of the circles of radius r centred at each node. As the rectangle is elongated, there is more boundary, which means there are more nodes near the boundary, and therefore more nodes where the boundary of their circle of radius r intersects the boundary of the rectangle. Then there is a smaller chance that some node will frustrate the symmetry between that node and an adjacent node, and we expect more pairs of symmetric nodes.

In Fig. 2.14 we show examples of the structures we have discussed in an RRG. In this example with $n = 500$, $a = 1.5$, and $r = 0.065$, we find that $m_{-1} = 47$. We highlight one pair of symmetric nodes (labelled ‘pair’) and one neighbourhood chain (labelled ‘chain’), where we colour the nodes with red and green colours as before, and also circle the nodes to highlight them. The value of m_{-1} indicates that there are several other such structures in this graph.

2. Structure and Connectivity of Random Rectangular Graphs

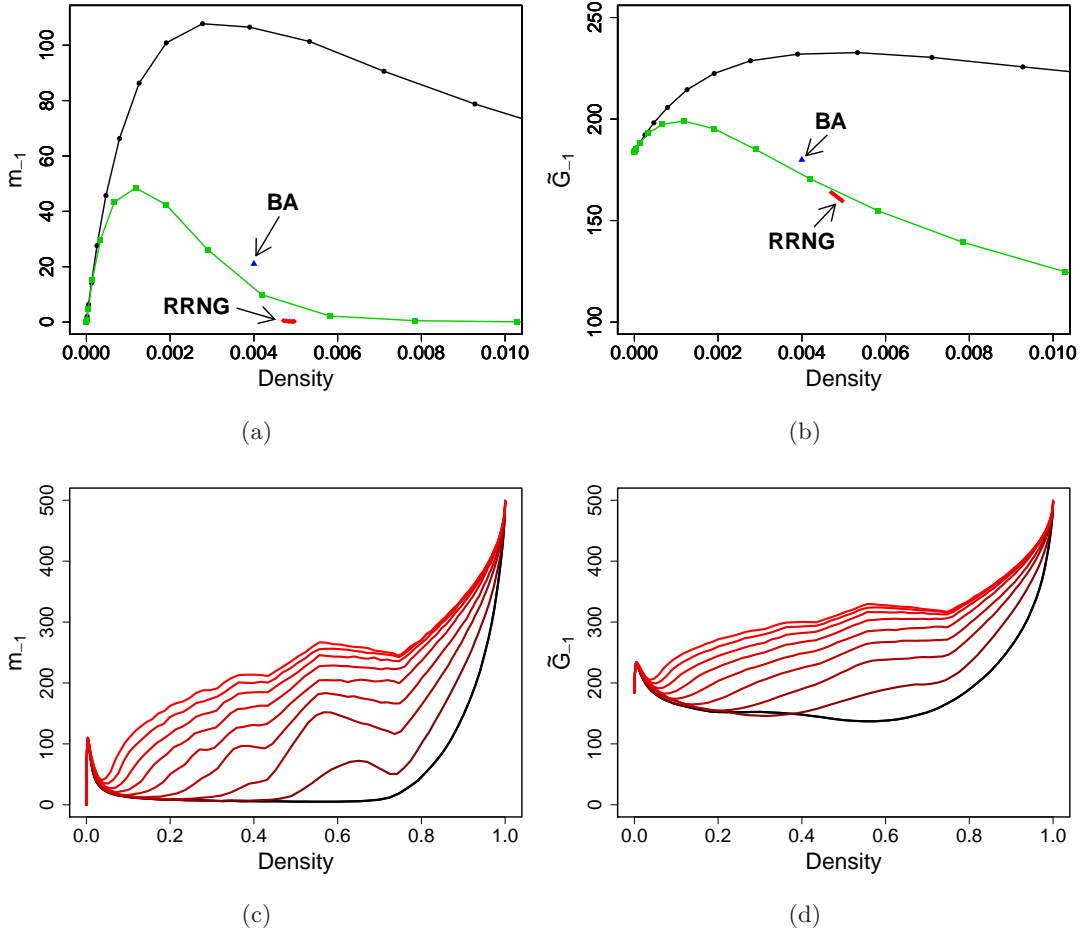


Figure 2.13: Plot of (a) m_{-1} and (b) \tilde{G}_{-1} against density for the RGG (black circles), RRNG (red diamonds), ER (green squares) and BA (blue triangle). Plot of (c) m_{-1} and (d) \tilde{G}_{-1} against density for RRGs for $a = 1$ (black) through $a = 5$ (red) in steps of 0.5. In all cases $n = 500$.

To conclude this discussion, we briefly mention some further investigations that could be made. The value of m_{-1} cannot be fully explained by the presence of symmetric nodes and neighbourhood chains. As a simple example to prove this fact, consider the following graph with 6 nodes in Fig. 2.15 with the same colouring rules as before. It is not difficult to construct even larger examples. We do not know the relative abundance of these structures larger than pairs of nodes in RRGs or other graphs, but presume they are rare.

However, perhaps symmetric nodes, neighbourhood chains, and the generalisation

2. Structure and Connectivity of Random Rectangular Graphs

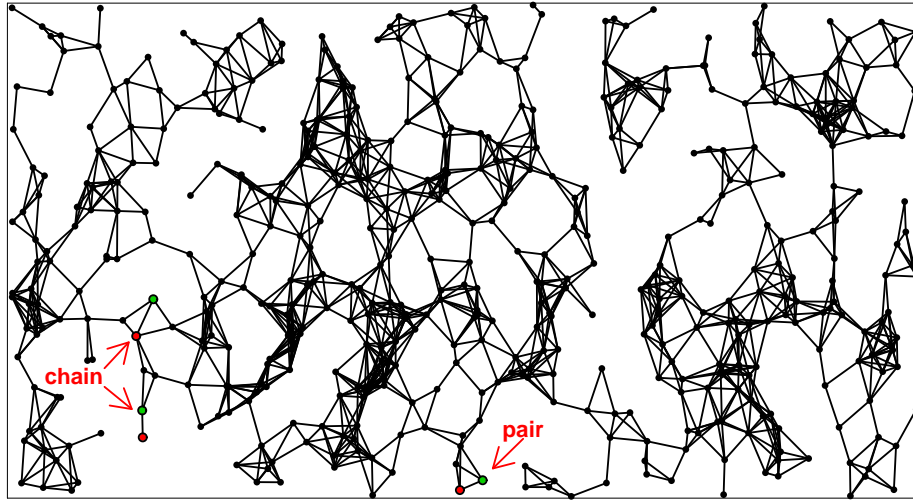


Figure 2.14: Example of a pair of symmetric nodes and a neighbourhood chain in an RRG, where V_0 is black, V_1 is red and V_2 is green for each structure as per Fig. 2.11.

of the neighbourhood chains to include larger structures as well as pairs of nodes are enough. Since these all give eigenvectors with entries in $\{0, 1, -1\}$, this gives us the following conjecture.

Conjecture 2.2.18. *A basis of the eigenspace corresponding to the eigenvalue -1 of a simple undirected graph can be chosen such that all the components of each eigenvector lie in the set $\{0, 1, -1\}$.*

2.2.8 Connectivity

In this section we examine the connectivity of the RRG. This has been investigated in the general case by Coon et al. [30], and we demonstrate their result for the RRG and also present a modification of their result. In an RRG it is harder to have connected nodes located near the boundary of the rectangle and particularly near the corners. Then, if an RRG is not connected, we could expect that the lack of connectivity is due to the existence of some isolated node, probably near a corner of the rectangle. We then draw a circle of radius r around the point. Let A be the fraction of the area of the circle that is inside the rectangle. For example, a point not near the boundary has $A = \pi r^2$ since the full circle is inside the rectangle, but for a point exactly on

2. Structure and Connectivity of Random Rectangular Graphs

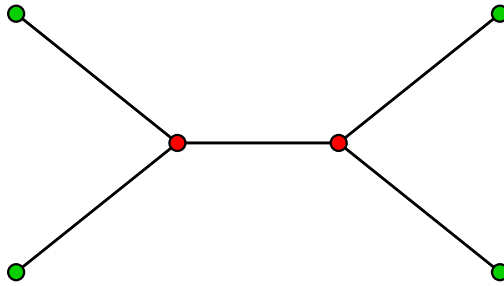


Figure 2.15: Example of a larger structure similar in nature to the pairs of nodes found in neighbourhood chains, where V_1 is red and V_2 is green as per Fig. 2.11.

a corner of the rectangle $A = \frac{1}{4}\pi r^2$ since only one quarter of the circle is inside the rectangle. For a general point near the boundary, it will take some value between these two extremes because some part of the circle will lie outside the rectangle. We can use this observation to approximate the probability that the RRG is connected.

In Fig. 2.16(a) we illustrate the variation of the connectivity of RRGs with the change of the connection radius for different values of the rectangle elongation. As can be seen the probability that the RRG is connected changes as a sigmoid function with the increase of the connection radius in a similar way as in the case of the RGGs. However, as the elongation of the rectangle increases it is more difficult for the graph to be connected and the critical radius guaranteeing that the graph is connected increases significantly with a .

The work by Coon et al. [30] considers an expansion around the point of becoming fully connected. We consider the first-order approximation for the RRG, which we rewrite in the relevant notation for our purposes:

$$P_{fc} \approx 1 - n \int_{\mathcal{R}} (1 - A(x))^{n-1} dx. \quad (2.48)$$

The intent of this approach is to have a good approximation in the region of the parameter space near the critical radius where the RRG is connected with high probability, which is the most important region for establishing the connectivity of the RRG. We may use this to find the approximate value of the connection radius r such that the

2. Structure and Connectivity of Random Rectangular Graphs

RRG is connected with high probability. For example, we may use a numerical root-finding algorithm to find r such that $\mathbf{P}(\text{RRG is connected}) = 0.999$. Note that trying to find $\mathbf{P}(\text{RRG is connected}) = 1$ should not work well since in theory this should only equal 1 when $r = \sqrt{a^2 + b^2}$, so a small tolerance is used for practical purposes. Using too small a tolerance is likely to be numerically unstable since the curve is very flat in this region.

The approach in Eq. 2.48 gives a very poor approximation of the probability of being connected when the connection radius is much smaller than the critical radius, even taking negative values, which is to be expected as it considers an expansion about the critical radius. Then, we use similar reasoning in order to approximate the connectivity over all values of the connection radius such that we capture the sigmoid shape better. In particular, we make a simplifying assumption that the node degrees are independent, which is known to be false but the approximation does not suffer too much as a result. Then, we reason as follows:

$$\begin{aligned}
 \mathbf{P}(\text{connected}) &\approx \mathbf{P}(k_i > 0, \forall i \in V) \\
 &= \prod_{i=1}^n \mathbf{P}(k_i > 0) \\
 &\approx \mathbf{P}(k_i > 0)^n \\
 &= (1 - \mathbf{P}(k_i = 0))^n \\
 &= (1 - \int_{\mathcal{R}} \mathbf{P}(\text{Bin}(n-1, A(x)) = 0) dx)^n \\
 &= (1 - \int_{\mathcal{R}} \binom{n-1}{0} A(x)^0 (1 - A(x))^{n-1} dx)^n \\
 &= (1 - \int_{\mathcal{R}} (1 - A(x))^{n-1} dx)^n.
 \end{aligned} \tag{2.49}$$

Again, this involves computing an integral, which can be done numerically.

In panel (b) of Fig. 2.16 we illustrate the way in which we determine the critical radii. For a given value of a we find the minimum value of r for which $\mathbf{P}(\text{connected}) = 1$. Although in all cases we use the values obtained from the simulations we can see that both theoretical bounds for $\mathbf{P}(\text{connected})$ produce very similar results near the critical radius. The approximation of Eq. 2.48 does provide a closer approximation for most

2. Structure and Connectivity of Random Rectangular Graphs

of the region where the RRG is only sometimes connected, but when the connectivity is near zero we see that the approximation becomes negative. The approximation of Eq. 2.49 tends towards zero off the left side of the plot, thus forming a sigmoid shape, but not providing as tight an approximation as we would have hoped for from this approach. Since it is just a first-order approximation, the second-order approximation may provide a much closer fit to the numerical results while maintaining the sigmoid shape, thus providing a better overall approximation. A simpler improvement would be to take the maximum of Eq. 2.48 and 0.

We then plot the values of the connectivity radius versus the elongation of the rectangles in Fig. 2.16(c). The curve joining the points of this plot makes a separation between the RRGs that are connected (upper triangular part) from those that are disconnected (lower triangular part). That is, the curve represents the critical radii versus critical elongation, and it gives the critical region indicating the connectivity of the RRGs. It can be read in two different ways. You can fix a value of a and then determine the critical radius for which the network will be connected. For instance, for a rectangle with longer side $a = 15$ it is necessary to use $r > 0.17$ to make the RRGs connected. Alternatively, we have a fixed radius of connection and we can find what elongation of the rectangle disconnects the network. For instance, if the connection radius is fixed to $r = 0.35$ every RRG is connected for $a < 30$. Finally, in Fig. 2.16(d) we show the outage near full connectivity, where $\text{outage} = 1 - \mathbf{P}(\text{connected})$ is the complement of connectedness, with the estimated connectivity using Eq. 2.48. We plot the outage since we can then use a log-scale on the y -axis to get a better view of the behaviour. We see that the expected outage is very similar for both $a = 1$ and $a = 2.5$, and the observed values appear to follow this trend reasonably well. Note that for the larger values of the connection radius than the ones shown on the plot we calculated an outage of 0, and so these points are necessarily omitted.

2. Structure and Connectivity of Random Rectangular Graphs

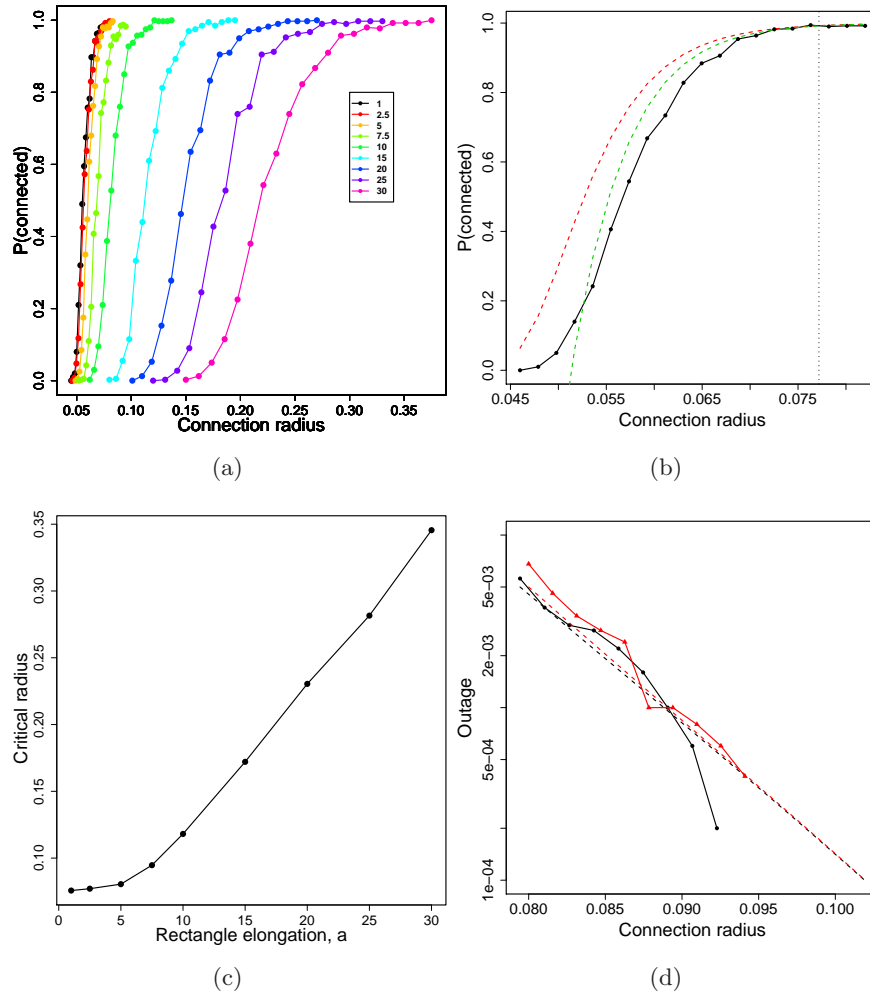


Figure 2.16: (a) Change of the connectivity of RRGs with the change of the connection radius for different values of the rectangle elongation. (b) Connectivity of an RRG with $a = 2.5$, with the upper bounds of Eq. 2.49 (red dashed line), Eq. 2.48 (green broken line) and critical radius (vertical dotted line) illustrated. (c) Plot of the critical versus the rectangle elongation. (d) Plot of outage near full connectivity for $a = 1$ (black) and $a = 2.5$ (red): observed values (circles and triangles) and approximation of Eq. 2.48 (dashed lines), with logarithmic y -axis. All RRGs studied here have $n = 1000$ nodes. Plots (a) and (c) use the average of 20 random realisations, plot (b) uses 500, and plot (d) uses 5000.

Chapter 3

Structure of Rectangular Relative Neighbourhood Graphs

In the previous chapter we discussed RRGs, the extension of the RGG model to rectangles; here we perform an analysis for rectangular relative neighbourhood graphs (RRNGs), which is the analogous extension of relative neighbourhood graphs to rectangles [60].

3.1 Definition of RRNGs

In this section, we consider the properties of the β -skeleton graphs when they are embedded into unit rectangles in an analogous way to the extension of the RGG model to the RRG model. Figure 3.1 illustrates two rectangular β -skeleton graphs with $\beta = 2$ and different values of the rectangle side length a and the same number of nodes. In the first case when $a = 1$ the graph corresponds to the classical β -skeleton graph in which the nodes are embedded into a unit square. The second case corresponds to $a = 2$, which is a slightly elongated rectangle.

Since we are mostly interested in β -skeletons in the context of rock fracture networks, we wish to determine what the best parameter of β is to best replicate these networks. We find that for all the RFNs we consider this is approximately $\beta = 2$, and so from here onwards we consider only rectangular RNG (RRNG) graphs which

3. Structure of Rectangular Relative Neighbourhood Graphs

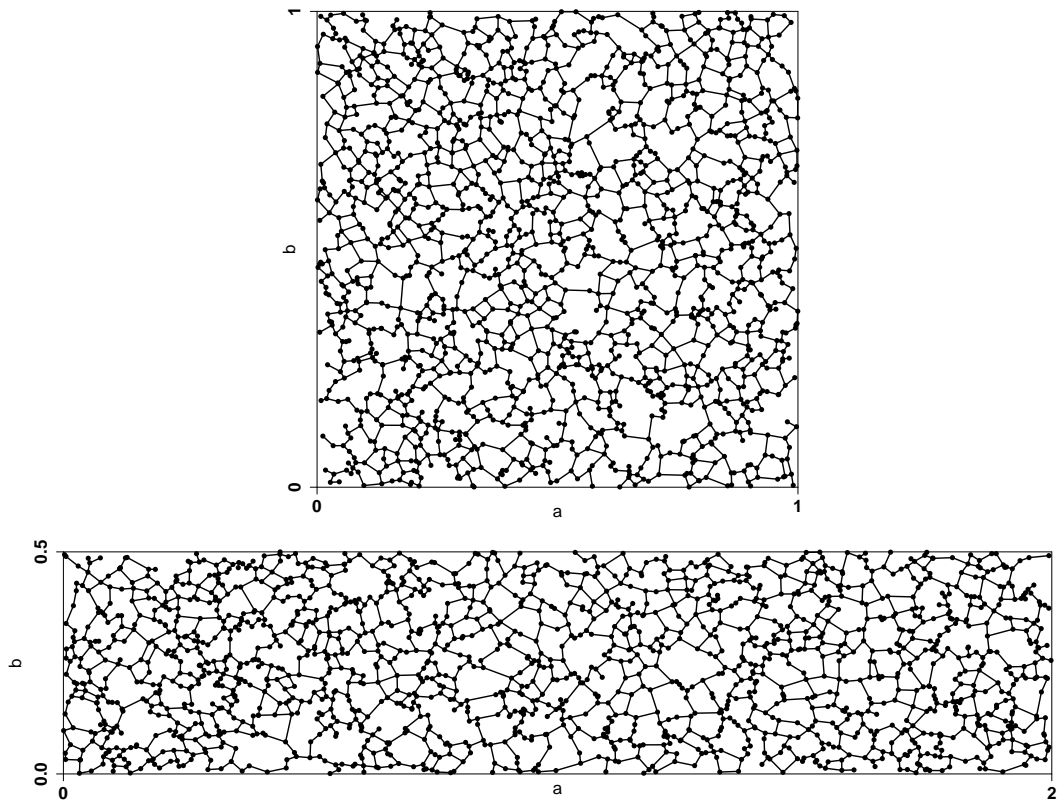


Figure 3.1: Example RNGs for $a = 1$ (top) and $a = 2$ (bottom). Each has $n = 1686$ nodes, which corresponds to a particular RFN.

correspond to $\beta = 2$. Note that if $\beta > 2$ then the graph would not be guaranteed to be connected.

3. Structure of Rectangular Relative Neighbourhood Graphs

3.2 Topological Properties of RRNGs

Here we show how some of the properties of the RRNG change with the elongation parameter, to show why the model can be of practical interest. We let $G_{RRNG}(n, a)$ be an *RRNG* with n nodes and embedded in a unit rectangle $[0, a] \times [0, b = a^{-1}]$. We focus on just a few of the most important structural parameters here: the average node degree, the diameter, and the algebraic connectivity μ_2 .

3.2.1 Average Node Degree

The average node degree of the RRNG model is more complicated than for the RRG model, so we assume that n is not too small and a is not too large, and find what the expected value of the average node degree should be in this regime. Fully understanding the behaviour for more general parameter values is left as future work. First, we derive the form that the formula should take. Then, we estimate the unknown constants using numerical simulations. We then look at the behaviour of the average node degree numerically, before concluding with a brief discussion justifying why we can make simplifying assumptions when n is large and a is small and what happens when these assumptions do not hold.

Theorem 3.2.1. *Assuming that n is large and the elongation a is small, the expected value of the average node degree $\bar{k}(G_{RRNG})$ of an RRNG takes the form*

$$\mathbf{E}(\bar{k}) \approx \alpha + \beta (a + a^{-1}) n^{-1/2}. \quad (3.1)$$

Proof. Consider the following alternative construction of an RRNG, for which the motivation will become clear. We assume that n is large and a is small. In this discussion we call the usual RRNG we construct G_0 , which has n nodes and m_0 edges.

1. Distribute n nodes uniformly at random in the rectangle \mathcal{R} .
2. Distribute nodes uniformly at random with the same density in a sufficiently large region containing \mathcal{R} .

3. Structure of Rectangular Relative Neighbourhood Graphs

3. Construct an RRNG G_1 in the usual way using all nodes.
4. Remove all edges E_b which intersect the boundary of \mathcal{R} .
5. Remove all nodes and edges which lie outside of \mathcal{R} .
6. Check whether each node has all the edges it would have in the usual RRNG constructed on these n nodes, and add any missing edges E_a .

We illustrate this process in Fig. 3.2, in particular highlighting the edges E_b removed in step 4 in red and the edges E_a added in step 6 in green, which are the main importance of this construction. In step 2, the choice of region should ensure that all nodes inside the rectangle are not subject to boundary effects in the construction of G_1 .

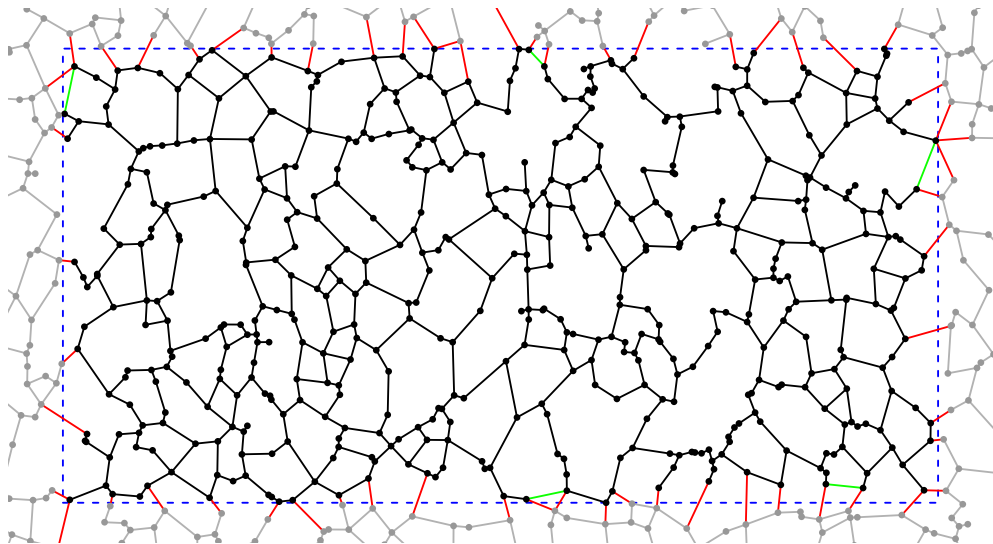


Figure 3.2: RRNG construction for average node degree proof. Nodes within the rectangle (blue dashed line) are part of G_0 and are black, those outside are in G_1 only and are grey. Edges common to G_0 and G_1 are black, edges in G_1 only are grey, edges in E_b are red, edges in E_a are green.

Let m_1 be the number of edges attached to the n nodes inside \mathcal{R} in G_1 (total of black and red edges in Fig. 3.2), m_b be the number of edges E_b of G_1 that intersect the boundary of \mathcal{R} , m_a be the number of edges E_a added in step 6, \hat{d} be the average distance along the boundary of \mathcal{R} between two adjacent intersections with edges in E_b , and $p = 2(a + a^{-1})$ be the perimeter of \mathcal{R} . The number of intersections is $m_b = p/\hat{d}$.

3. Structure of Rectangular Relative Neighbourhood Graphs

Considering how lengths scale, we can write $\hat{d} = c_d n^{-1/2}$, where c_d is a constant, so then

$$m_b = \frac{pn^{1/2}}{c_d}. \quad (3.2)$$

We are now in a position to state our simplifying assumptions. First, nodes losing or gaining edges at one edge of the boundary do not interact with such nodes at the opposite edge of the boundary. In other words, the effect of removed nodes is ‘localised’. Secondly, we assume that added edges occur as a result of removed edges, at a constant average rate of $\gamma = m_a/m_b \ll 1$. Note that the second assumption relies on the first one. We then have

$$m_0 = m_1 - m_b + m_a = m_1 - (1 - \gamma)m_b. \quad (3.3)$$

Finally, we let $\bar{k}_0 = m_0/n$ and $\bar{k}_1 = m_1/n$, then after multiplying Eq. 3.3 through by $2/n$ and rewriting m_b using Eq. 3.2, we find that the expected value of the average node degree is

$$\begin{aligned} \mathbf{E}(\bar{k}(G_{RRNG})) &= \mathbf{E}(\bar{k}_0) = \mathbf{E}\left(\bar{k}_1 - \frac{4(1 - \gamma)(a + a^{-1})n^{-1/2}}{c_d}\right) \\ &= \alpha + \beta(a + a^{-1})n^{-1/2}. \end{aligned} \quad (3.4)$$

□

Remark 3.2.2. We have found computationally that $\alpha \approx 2.557$ and $\beta \approx -0.898$, so then

$$\mathbf{E}(\bar{k}) \approx 2.557 - 0.898(a + a^{-1})n^{-1/2}, \quad (3.5)$$

where we use $n = 400, 500, \dots, 1000$ and $a = 1, 1.5, 2, 2.5$ in an attempt to stay in the region of the parameter space where we expect the assumptions to hold. Though these choices are somewhat arbitrary, experimenting with different choices leads us to believe that they are sensible.

It is clear that as a becomes large the RRNG tends to the path graph P_n , which will dictate the behaviour at larger elongations, and Fig. 3.3 (a) demonstrates this.

3. Structure of Rectangular Relative Neighbourhood Graphs

As the elongation increases, the average node degree decreases since the perimeter increases and m_b is proportional to p ; with further elongation the value will approach $\bar{k} = (2 - 2/n) \approx 2$. Figure 3.3 (b) demonstrates the effectiveness of Eq. 3.5 for $1 \leq a \leq 5$ in steps of 0.5 and $n = 100, 500, 1000$. We see that when n is larger the average node degree is slightly larger, around 2.50 for 1000 nodes in the square compared to 2.37 for $n = 100$. This is because m_b is proportional to $n^{1/2}$, so $m_b/n \rightarrow 0$ as $n \rightarrow \infty$ and thus m_b has a diminishing effect as n is increased – indeed in the limit of n tending to infinity it is easy to see that \bar{k} approaches the constant \bar{k}_1 . We also note that our approximation fits the curve very well for the larger values of n , but does not fit the shape of the curve for $n = 100$ with the same accuracy. This is to be expected, since we assumed that we are in the regime where n is large, and therefore we do not fully capture the behaviour of the average node degree when there are relatively few nodes. However, it appears to still give a good approximation, with an absolute error of only 0.8% for the case of $n = 100$ and $a = 5$, which is where we expect the approximation to be the least reliable out of all the choices of parameter values used here.

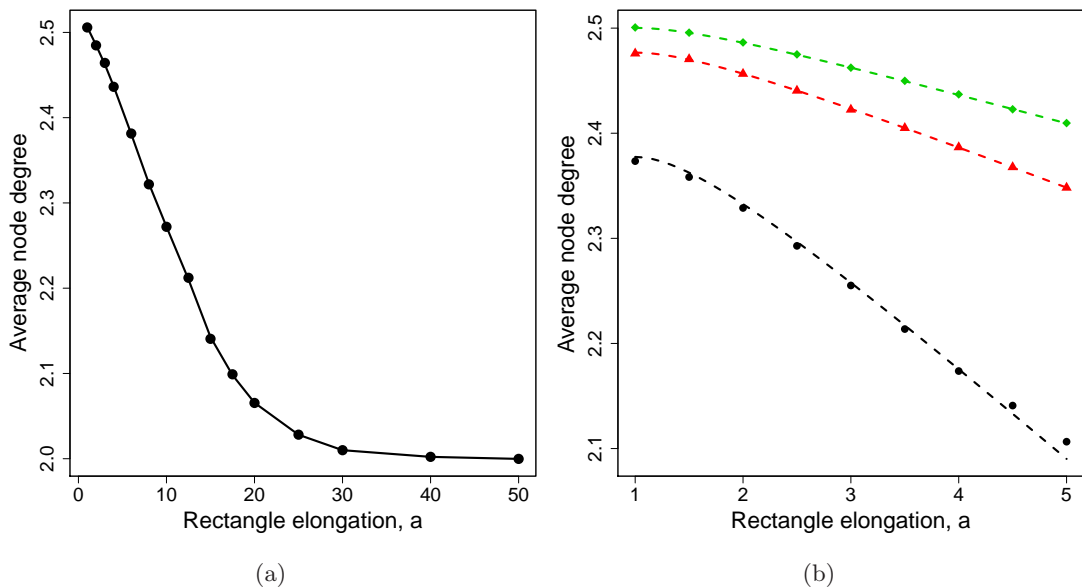


Figure 3.3: (a) Average node degree of RRNGs with $n = 1000$ nodes for a varying between 1 and 50 to demonstrate the long-term behaviour. (b) Average node degree with $a = 1, \dots, 5$ for $n = 100$ (black circled), $n = 500$ (red triangles), and $n = 1000$ (green diamonds). The dashed lines are the corresponding approximations from Eq. 3.5.

3. Structure of Rectangular Relative Neighbourhood Graphs

We now give some discussion of when our assumptions do not hold. We can say that

$$\lim_{a \rightarrow \infty} \mathbf{E}(\bar{k}_0) = \lim_{a \rightarrow \infty} \left(\bar{k}_1 - \frac{4(1-\gamma)(a+a^{-1})n^{-1/2}}{c_d} \right) = 2. \quad (3.6)$$

Since $\lim_{a \rightarrow \infty} ((1-\gamma)(a+a^{-1})) = \lim_{a \rightarrow \infty} ((1-\gamma)a)$, we then have

$$\lim_{a \rightarrow \infty} ((1-\gamma)a) = \frac{(\bar{k}_1 - 2)c_d n^{1/2}}{4} \Rightarrow (1-\gamma) \sim n^{1/2} a^{-1}. \quad (3.7)$$

This means that the $(1-\gamma)$ term decreases as n is decreased or a is increased, which means that γ is increasing. We assumed that $\gamma \ll 1$, which would mean that $(1-\gamma) \approx 1$ is essentially constant, and now we see that this assumption must fail when n is too small or a is too large. We demonstrate in Fig. 3.4 one reason why our assumptions can fail when there are relatively few nodes in paths from the top edge of the rectangle to the bottom. We had assumed that nodes which lose edges to opposite sides of the rectangle do not interact and so the effect of steps 4 and 6 were localised, but here we find two such nodes that become connected by an added edge.

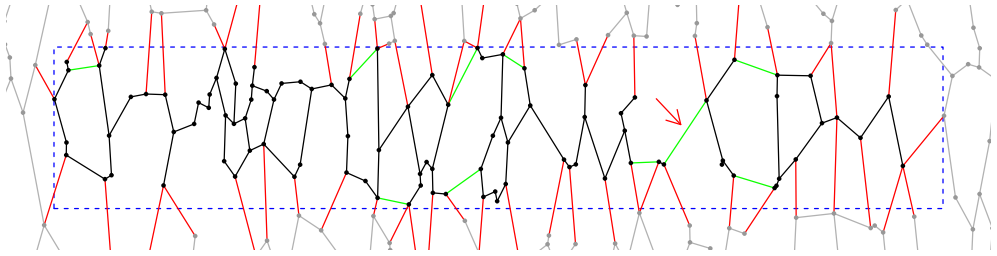


Figure 3.4: RRNG construction similar to Fig. 3.2 but with $n = 100$ and $a = 4$. We find an edge, highlighted with a red arrow, which demonstrates our assumptions being broken.

3.2.2 Diameter and Algebraic Connectivity

First we prove the following result about the diameter of the RRNG.

Theorem 3.2.3. *The diameter d_{max} of an RRNG $G = G_{RRNG}(n, a)$ is bounded as*

$$d_{max}(G) \geq \sqrt{(a^2 + b^2)n}. \quad (3.8)$$

3. Structure of Rectangular Relative Neighbourhood Graphs

Proof. The nodes of the RRNG are uniformly and independently distributed in the unit rectangle. Then, let us assume that the n points are equally spaced in the area of the rectangle separated by a Euclidean distance L . The longest path should lie roughly along the main diagonal of the rectangle. If the length of the main diagonal is c there are $\frac{c}{L}$ connected nodes in this line. Thus, the maximum shortest path distance in the RRNG is $\frac{c}{L}$. For a connected RRNG this is the shortest the diameter can be, because if two points in the main diagonal are separated at a Euclidean distance larger than L , then the diameter of G will be larger than $\frac{c}{L}$. Now, the main problem here is to place an upper bound on the value of L for the average separation of two points in the RRNG. Here we consider that the n points can be distributed in the square ($a = 1$) forming a regular square lattice. In this case there should be \sqrt{n} rows of \sqrt{n} points equally spaced in the square. Thus, the separation between two points in the grid is $L = 1/\sqrt{n}$. In the case of the rectangle we follow a naive approach of considering that a rectangle of major length $a > 1$ can be obtained just by cutting the square a times in the direction of the y -axis and pasting the cut rectangle along the x -axis. In this way it is guaranteed that the separation between two points in the original square remains the same in the elongated rectangle. That is, $L \approx 1/\sqrt{n}$. Note that for very large values of a the nodes will be placed on a line and the construction fails, but in that regime the RRNG is barely 2D and therefore uninteresting for our purposes.

Since the nodes are placed randomly and not on a grid, we note that by construction the points in the RRNG prefer to attach to nearby nodes, and any such connections from say node v_i to node v_j will prevent node v_i from attaching to slightly further away nodes near v_j . Therefore, there is a clear bias towards shorter edge lengths, and we have an upper bound on the value of L . Since we consider a lower bound on the diameter, this proves the result. \square

We note the following construction that gives a different insight into why the diameter should be proportional to \sqrt{n} . First, construct an RRNG G_1 on a rectangle with n_1 nodes in a rectangle $[0, a] \times [0, b]$. The expected value of the diameter of this graph is the expected number of steps required to travel distance c through an RRNG. Now, we consider the subset of nodes of cardinality n_2 in one quarter of the

3. Structure of Rectangular Relative Neighbourhood Graphs

rectangle, say $[0, a/2] \times [0, b/2]$ and construct an RRNG G_2 on these nodes. Clearly, $d_{max}(G_1) \approx 2d_{max}(G_2)$ and $n_1 \approx 4n_2$, which implies that $d_{max} \sim \sqrt{n}$.

Obviously, we can fit the values of the right-hand-side of (3.8) to the actual values of the diameter in RRNGs of different sizes. That is, we can obtain an empirical relation of the kind $d_{max}(G) \approx \alpha \sqrt{(a^2 + b^2)n}$, where here α is a fitting parameter. By doing so for RRNGs with sizes $n = 500, \dots, 1000$ we found that $\alpha \approx 1.414$ and the correlation coefficient between the observed and the calculated diameter is larger than 0.999. The diameter of RRNGs is shown in Fig. 3.5(a), and the fitted value is demonstrated in Fig. 3.5(b).

The previous results is very important because it allows us to bound the algebraic connectivity of the RRNG. Then, we next prove the following result.

Lemma 3.2.4. *The second smallest eigenvalue of the Laplacian matrix of the RRNG $G = G_{RRNG}(n, a)$ is bounded as*

$$\mu_2(G) \leq \frac{8k_{max}}{a^2 + b^2} \frac{\log_2^2 n}{n}, \quad (3.9)$$

Proof. We simply substitute our bound for the diameter of the RRNG into the Alon–Milman bound for the algebraic connectivity of any graph (Eq. 2.39), as we did for the RRG. □

It should be noticed that in RRNGs the value of k_{max} is typically quite small. In all our simulations it is not larger than 6. In order to see these effects in practice we develop a series of simulations considering RRNGs with $n = 1000$ nodes, $1 \leq a \leq 5$ in steps of 0.5, which we find to be a sufficiently large range of elongations for the analyses we perform. These two theoretical results clearly indicate that the properties of the RRNG are significantly and non-trivially affected by the elongation of the rectangle. The diameter of the RRNGs increases almost linearly with the elongation of the rectangle. On the other hand, the elongation of the rectangle in the RRNG makes the graphs drastically less connected. As predicted by our theoretical results, the diameter of the RRNGs increases almost linearly with the elongation of the rectangle and the algebraic connectivity decays in a nonlinear fashion with it. Therefore, we conclude

3. Structure of Rectangular Relative Neighbourhood Graphs

that as the elongation of the RRNG model is increased, the structural parameters also change, including those closely connected to dynamical processes such as the algebraic connectivity, which is shown in Fig. 3.5(c).

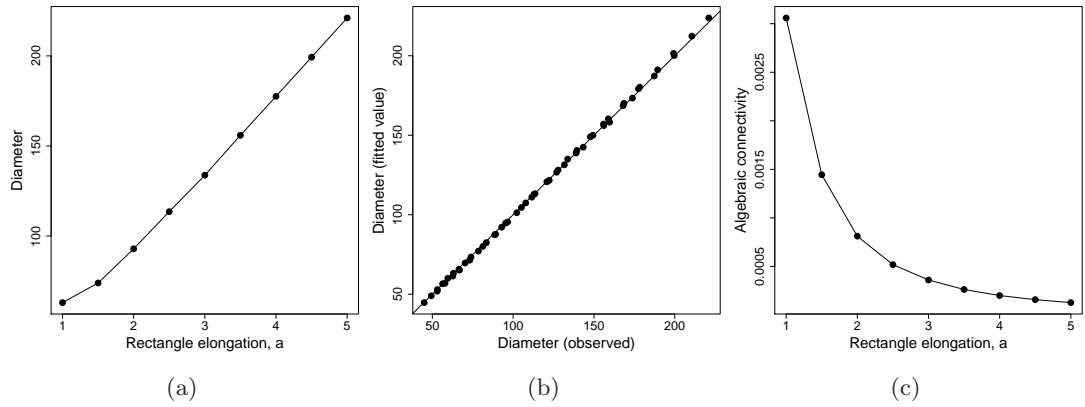


Figure 3.5: Plots of how several structural parameters of the RRNG model change as the elongation of the rectangle varies from $a = 1$ to $a = 5$ for $n = 1000$ nodes: (a) diameter, (b) observed vs. fitted diameter, (c) algebraic connectivity.

Chapter 4

Applications

In this chapter we discuss three applications of the RRG and RRNG models. First, we consider diffusion (consensus) processes on RRGs; secondly, we consider epidemic models on RRGs. For the third application we study how the models can be used to model RFNs. The third application receives the most attention here because we were able to get data for several RFNs, and therefore we were able to perform a detailed study of the models in a real-world situation.

4.1 Diffusion (Consensus) in RRGs

Since RRGs have been used to model several real-world networks as we have previously discussed, it is of practical interest whether the dynamical behaviour changes as we generalise from RRGs to RRNGs. In this section we discuss in more detail the dynamics of diffusion in RRGs [59].

The importance of Theorem 1.3.1 is that when $\mu_2 \rightarrow 0$ the time of diffusion grows to infinity. We have already proved in Theorem 2.2.14 that the elongation of a random geometric graph with a given number of nodes and a fixed connection radius will make the algebraic connectivity go asymptotically to zero. The immediate consequence of this result is that the diffusion time grows to infinity in RRGs when $a \rightarrow \infty$ due to the inverse relation between the diffusion time and the algebraic connectivity.

We now study the influence of the rectangle elongation on the diffusion dynamics on RRGs. We take care with the elongation process so that the graphs do not become

4. Applications

disconnected. First we compare the discrete-time evolution of two RRGs with $n = 500$ nodes and $r = 0.09$, but one having $a = 1$ and the other having $a = 2$. In Fig. 4.1(a) we plot the time evolution of the diffusion dynamics as the maximum pairwise difference in the node states (compare Fig. 1.4, which gives a more in-depth illustration of the diffusion process), where it can be seen that the time of diffusion for the graph embedded into the unit square is at least 10 times shorter than that for the elongated RRG. Because these plots are the results of only one random realisation we perform a systematic variation of the rectangle side length and report the average of the time of diffusion after 100 random realisations for each value of a with a stopping criterion of $\delta = 10^{-4}$. The results are illustrated in Fig. 4.1(b), where we plot the values of the average time for diffusion versus the rectangle side lengths (blue squares). As can be seen the time for diffusion increases with the elongation of the rectangle. The best fit for this correlation is a 4th order polynomial: $\langle t_c \rangle \approx 0.1885a^4 - 1.651a^3 + 19.59a^2 - 37.06a + 30.59$; the fit has a Pearson correlation coefficient of 0.9997. Using this model we can obtain a more precise estimation of the average time for diffusion of the random realisation illustrated in Fig. 4.1(a). For a value of $\delta = 10^{-4}$ the diffusion is reached for $a = 1$ at a time of 11.66, while for $a = 10$ at a time of 1853.

The estimated times for diffusion obtained from Eq. 1.54 on page 33 are also plotted in Fig. 4.1(b) (red circles), where it can be seen that they follow the same trend as the observed values. Indeed, the plot of the observed values versus those expected from Eq. (1.54) (see Fig. 4.1(c)) indicates a perfect linear correlation between the two with a Pearson correlation coefficient of 0.9999.

Finally, we plot in Fig. 4.2 the dependence of the time of diffusion with respect to both the connection radius and the rectangle side length. Visually, the line that divides the region of relatively fast diffusion (deep blue region in the contour plot) from that of relatively slow one is given by $a = \kappa r - 1.5$, where $\kappa = 28$ for the analytical and $\kappa = 26$ for the observed results. Thus, a condition for fast diffusion in RRG with $n = 500$ can be simply approximated by

$$\frac{a + 1.5}{r} < \kappa. \quad (4.1)$$

We briefly explore some of the consequences that our results have on the study of

4. Applications

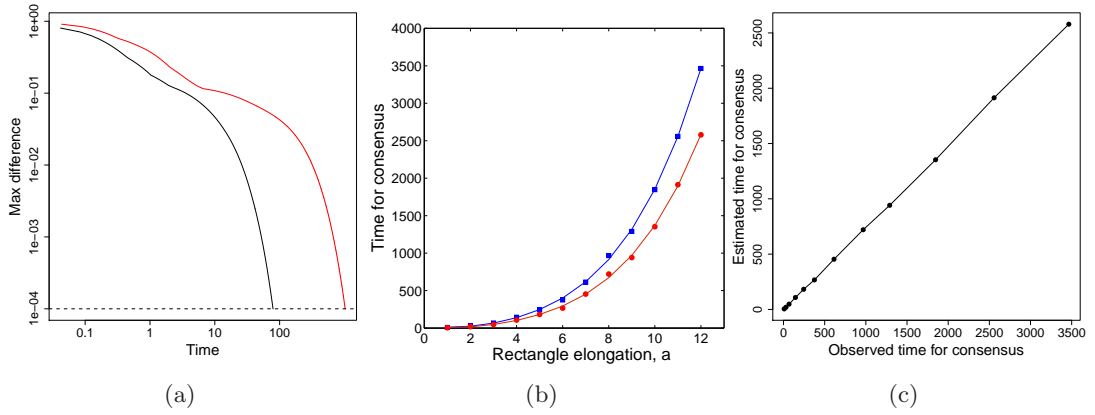


Figure 4.1: (a) Illustration of the difference in diffusion dynamics for RRGs with $a = 1$ (black) and $a = 2$ (red) in log-log scale, the y -axis is the maximum difference in the state of any two nodes and diffusion is reached if this value is small enough. The simulations were carried out using a discrete time diffusion model (see Eq. 1.45) with a random allocation of initial states for the nodes. Both networks have $n = 500$ nodes and the connection radius is $r = 0.09$. The elongated rectangle takes over 10 times longer to finish diffusion. (b) Dependence of the time for diffusion with the length of the side of the rectangle. Here the squares represent the average values of the 100 simulations and the circles are the values obtained from Eq. 1.54. The solid lines are the lines of best fit, which were obtained using 4th order polynomials. (c) Linear plot of the observed and estimated (using Eq. 1.54) time of diffusion of the RRGs with $n = 500$ nodes and $r = 0.15$.

diffusion in real-world situations. A city like Manhattan has dimensions that resemble a rectangle more than a square. That is, Manhattan is 21.6km long and 3.7km wide. This can be represented as a unit rectangle of dimensions $a \approx 2.42$ and $b \approx 0.41$. Using our fitted model, and considering that we embed 500 nodes, e.g., wireless sensors to monitor the city, we obtain the expected time for diffusion on this RRG, which is 38.3. This time is 3.3 times longer than the one expected if the network is considered to be embedded into a unit square. That is, we would be underestimating the time for diffusion of the sensors by a factor of three. Also, according to Eq. 4.1 we can estimate that a fast diffusion can be reached in this network only if $r > 0.157$.

4. Applications

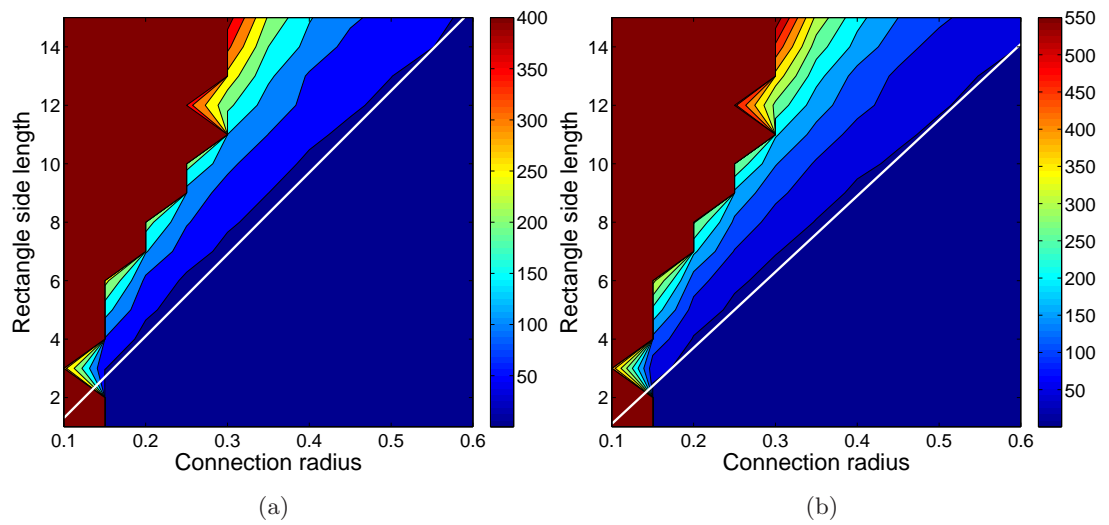


Figure 4.2: Contour plot showing the dependence of the time of diffusion with the connection radius and the rectangle side length in RRGs with 500 nodes: (a) Analytical results from Eq. 1.54 and (b) Observed results from the simulations. The diagonal white line corresponds to the equations $a = \kappa r - 1.5$, where $\kappa = 28$ for the analytical and $\kappa = 26$ for the observed results.

4.2 Epidemics in RRGs

We concentrate more on the phenomenology of the process than in deriving analytical results about the dependence of the epidemic threshold with the topological parameters of the RRGs. Equation 1.57 on page 36 tells us that λ_1 is what determines the epidemic threshold τ . Recalling that $\bar{k} \leq \lambda_1$, it is straightforward to realise that $\tau \geq \bar{k}^{-1}$. Then, by substituting in Eq. 2.1 we can bound the epidemic threshold of an RRG as

$$\tau \geq \frac{1}{(n-1)f}. \quad (4.2)$$

This result generalises the one obtained by Preciado and Jadbabaie [133] for the RGG to the case where we have a rectangle of any elongation and where we consider explicitly the border effects of the rectangle. We have shown already that as elongation increases, λ_1 decreases, which implies that the epidemic threshold grows monotonically.

Here we consider extensive numerical simulations of the SIS dynamics for different values of the elongation a and fixed radius r with the goal of checking the goodness of the bound in Eq. 4.2 and to illustrate how the elongation of the rectangle in the RRG model changes the epidemic dynamics. In the simulations a small fraction $\rho_0 = 0.01$ of the nodes were seeded with the infection, and the SIS dynamics were allowed to evolve for $5 \cdot 10^4$ time-steps. At this point, we let the simulations run for an additional 10^3 time-steps, and calculated the fraction of infected nodes ρ as the average of $\rho(t)$ over this period. For each choice of parameters 250 random realisations were used, each with different initial conditions. The final value of ρ is the average over all repetitions. Finally, we note that the resemblance of SIS and SIR epidemiological models translates into identical expressions for the epidemic threshold [165]. We therefore anticipate that our results will be also valid in a SIR framework, which could possibly be more relevant for plant diseases.

Figure 4.3 shows the fraction of infected nodes in the stationary state against the infection rate β for $a = 1, 10, 20, 30$.

4. Applications

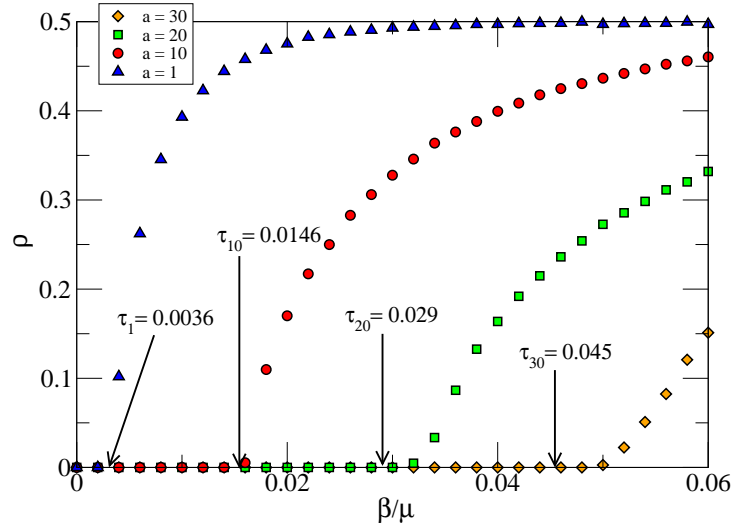


Figure 4.3: Fraction of infected nodes at the stationary state ρ as a function of the infection rate β for $a = 1, 10, 20, 30$; $a = 1$ represents the first case ($0 \leq r \leq a^{-1}$) of Eq. 2.2 while $a = 10, 20, 30$ fall into the second case ($a^{-1} \leq r \leq a$). Other parameters are: $n = 1000$, $r = 0.35$ and $\mu = 1.0$. Each point is an average over 250 random realisations.

To have a more detailed picture of the behaviour of the epidemic threshold, in Fig. 4.4(a) we compare the theoretical bound with the epidemic threshold obtained via the numerical simulations. This comparison covers two of the three cases of Eq. 2.2: $0 \leq r \leq a^{-1}$ and $a^{-1} \leq r \leq a$. We have then calculated the relative error for the values of the epidemic threshold observed in the simulations respect to the bound obtained in Eq. 4.2. The average relative error for all the RRGs having $1 \leq a \leq 35$ is 2.93% and in no case is the relative error larger than 10%. Also, we observe no trend in the relative error related to the elongation of the rectangle.

Finally, in Fig. 4.4(b) we tested the third case of Eq. 2.2, $a \leq r \leq \sqrt{a^2 + a^{-2}}$ for $a = 3$ and $r = 3.01$. In all cases, as expected, the theoretical and the simulation results show that increasing the elongation of the rectangle produces an increase of the epidemic threshold. In other words, the elongation of the rectangle retards the disease progress through the nodes embedded in the rectangle.

In the case of disease propagating on plants, these results—both analytical and

4. Applications

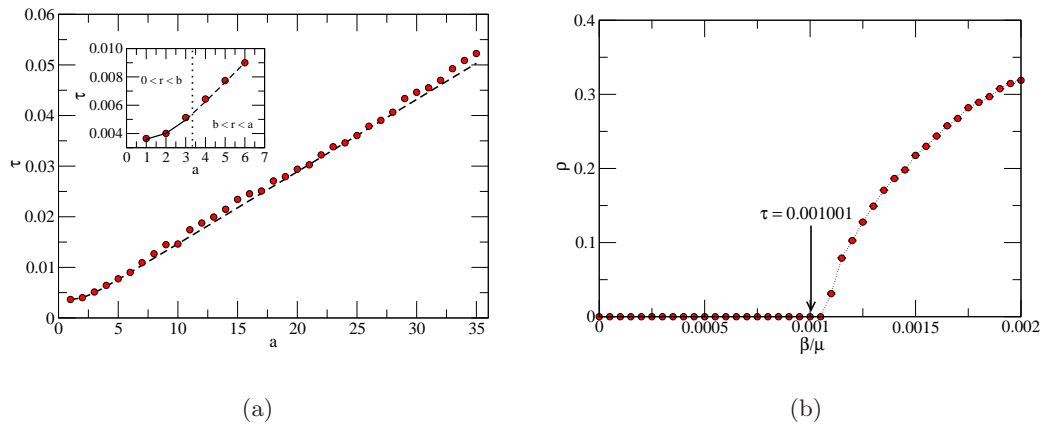


Figure 4.4: (a) Comparison between the theoretical bound and the epidemic threshold obtained via numerical simulations. Line represents the theoretical prediction of Eq. 4.2 while points represent the results of simulations. The inset shows a zoom for the first case of Eq. 2.2 $0 \leq r \leq a^{-1}$ (full line) and the second case $a^{-1} \leq r \leq a$ (dashed line). Other parameters are: $n = 1000$, $r = 0.35$ and $\mu = 1.0$. Each point is the average over 250 random realisations. (b) Fraction of infected nodes at the steady state ρ as a function of the infection rate β for $a \leq r \leq \sqrt{a^2 + a^{-2}}$. In the simulations $a = 3$ and $r = 3.01$. Other parameters are: $n = 1000$ and $\mu = 1.0$. Each point is the average over 250 random realisations.

simulations—coincide with the field observations and simulations using stochastic models [126, 157, 93, 65, 21, 113, 112, 164, 63], which suggest that square plots and fields favoured higher spreading of plant diseases than elongated ones of the same area [126, 157, 93, 65].

Our analytical and simulation results point to the fact that under the same conditions, the propagation of an epidemic on a rectangular plot/field is much harder than on a square one because a larger number of infected individuals is needed for the disease to become epidemic. Indeed, Fleming et al. [65] had very long ago concluded that “*decreasing field size and elongating field shape can retard disease progress and thus reduce yield losses*” when studying the spread of crop diseases. Here we have kept the size of the field constant by considering unit rectangles in our analysis. However, it is important to consider that other factors, such as the orientation of the field, play a fundamental role in the propagation of a disease on plants.

For instance, if the rectangular plots are placed perpendicular to the direction of

4. Applications

the prevalent winds the disease will not propagate as a consequence of this factor. To account for this one option would be to replace the circle around each node with an ellipse elongated in the downwind direction. Another option would be to somehow introduce bias in the direction of the wind in the equations governing the time evolution of each node, such that an infected node may more easily infect a node directly downwind compared to a node located crosswind.

4. Applications

4.3 Rock Fracture Networks

This section describes the dataset of real-world networks consisting of the channels and their intersections produced by fractures in rocks of petrophysical interest. The procedures described hereafter are based on the analysis developed by [138]. These authors have considered a series of rocks extracted from wells in the Gulf of Mexico. We have made our own analysis of the data extracted from these rocks, and compared the topology of the fractures to different random graph models, especially the RRG and the RRNG, to determine which is the most suitable, and whether changing the parameter of elongation is indeed useful [60]. This section is significantly longer than the previous two because we had access to data to perform a detailed investigation of real-world networks, and we had no such data for the other applications discussed in this work.

4.3.1 Description of Fracture Networks

The rocks are cut into two halves and images are taken of one of the rock halves, which show the fractures in the corresponding rock. An algorithm is then used to find the skeleton of these fractures and construct a network representation of it, which is stored as an adjacency matrix. In this work we worked only with the adjacency matrices that resulted from this process, and did not have data from any prior step. The nodes of the network correspond to where fractures intersect or terminate and an edge between nodes corresponds to a channel between those points in the skeleton of the rock fracture. A sketch of the process is illustrated in Fig. 4.5. Of course, in a practical setting it is important to understand not just the fluid flow within the individual rocks where the oil is trapped, but also in between the rocks. We only consider the former in this work since we lack data on the latter.

In total 29 rock samples are considered here, kindly provided by [138, 140, 139]. The number of nodes n and edges m in the 29 networks studied here are provided in Table 4.1 together with the labels used in the subsequent analysis in this work.

4. Applications

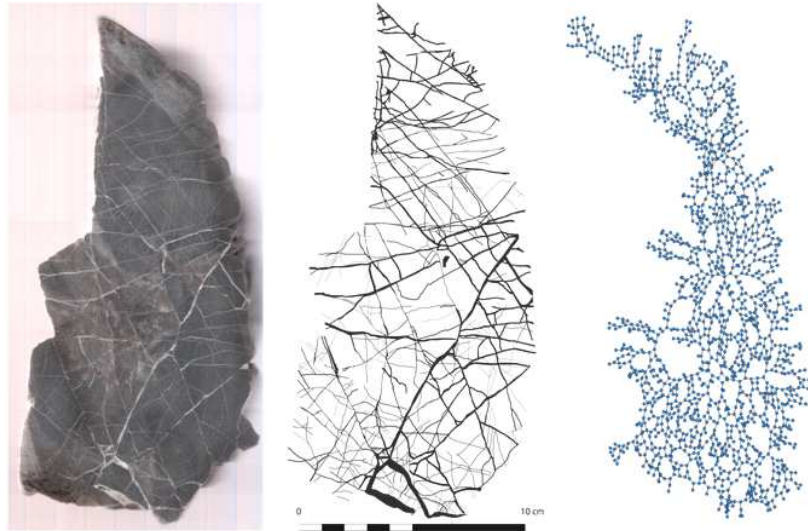


Figure 4.5: Illustration of the process of creating rock fracture networks from rock samples. (Left) A rock sample from the Gulf of Mexico showing one of the halves of the rock. (Centre) Digital image of the rock illustrating the fractures existing in it. (Right) Rock fracture network created from the digital image by taking the intersection of channels as the nodes and channels as the edges of the network. This corresponds to M20 in Table 4.1. Image is courtesy of E. Santiago [140].

| No. | n | m | a | | No. | n | m | a |
|-----|-----|-----|-----|--|-----|------|------|-----|
| M1 | 93 | 109 | 1.0 | | M16 | 779 | 887 | 1.5 |
| M2 | 46 | 54 | 1.5 | | M17 | 633 | 728 | 1.5 |
| M3 | 71 | 74 | 2.0 | | M18 | 3585 | 4183 | 1.5 |
| M4 | 109 | 124 | 1.5 | | M19 | 97 | 101 | 2.0 |
| M5 | 346 | 380 | 3.5 | | M20 | 1567 | 1996 | 1.0 |
| M6 | 85 | 89 | 2.5 | | M21 | 70 | 74 | 2.0 |
| M7 | 55 | 60 | 1.0 | | M22 | 1686 | 1905 | 2.0 |
| M8 | 87 | 94 | 1.0 | | M23 | 396 | 396 | 4.5 |
| M9 | 44 | 46 | 1.0 | | M24 | 181 | 180 | 5.0 |
| M10 | 296 | 336 | 1.5 | | M25 | 394 | 418 | 4.0 |
| M11 | 47 | 48 | 2.0 | | M26 | 808 | 813 | 4.5 |
| M12 | 46 | 47 | 1.5 | | M27 | 223 | 222 | 4.0 |
| M13 | 215 | 233 | 2.5 | | M28 | 297 | 305 | 4.0 |
| M14 | 132 | 144 | 1.5 | | M29 | 363 | 365 | 5.0 |
| M15 | 40 | 40 | 2.0 | | | | | |

Table 4.1: Rock fracture networks studied in this work, their number of nodes n , and number of edges m .

4. Applications

4.3.2 Similarity between Fracture Networks and RRNGs

We now compare the real-world rock fracture networks with their random analogues created by using the β -skeleton approach. To achieve this goal, random rectangular neighbourhood graphs are created with a value of $\beta = 2$ and having the same number of nodes and edges as the corresponding real-world fracture network. The elongations of the rectangle are varied in the range $1 \leq a \leq 5$ with a step size of 0.5. We selected the value of $\beta = 2$ based on empirical observations of the dataset of rock fracture networks under analysis. First, these real-world networks are always planar and have a relatively low number of triangles. These two characteristics are well-reproduced by the relative neighbourhood graphs corresponding to $\beta = 2$. Furthermore, these RNGs have been widely considered in the literature and can be constructed more easily and quickly than an a β -skeleton for some arbitrary value of β .

Then, for each graph, we construct a $k \times 1$ vector consisting of the structural properties listed in Table 4.2 and the 18 small subgraphs given in Appendix A. That is, every network is represented in a k -dimensional space ($k = 37$) in which each coordinate represents a structural parameter, e.g., average degree, clustering coefficient, etc. A number of random constructions of the RRNG are realised for each elongation and the value averaged over all of them. The number of random realisations varies depending on the size of the network for reasons of computational difficulty, with effort made to ensure a large enough number of realisations such that the variance is not too large.

The similarity between the real-world rock fracture networks and their RRNG analogues can then be calculated; we may also refer to the dissimilarity, which is simply a different view of the same quantity. This similarity is quantified by simply using the Euclidean distance between the corresponding points in the k -dimensional property space in which they are represented, and thus the similarity may also be referred to as the ‘distance’ when convenient. A potential problem arising here is the fact that the values of the properties calculated lie in a very wide range of numerical values. Then, the values of each property are normalised to lie in the range between 0 and 1. Such

4. Applications

| No. | Property | No. | Property |
|-----|---------------------------|-----|---------------------------------------|
| 1 | Average degree | 11 | Largest eigenvalue of \mathcal{L} |
| 2 | Largest eigenvalue of A | 12 | Algebraic connectivity |
| 3 | Degree variance | 13 | Estrada index |
| 4 | Collatz-Sinogowitz index | 14 | Spectral bipartivity |
| 5 | Degree heterogeneity | 15 | Average communicability distance |
| 6 | Degree assortativity | 16 | Average communicability angle |
| 7 | Clustering coefficient | 17 | Entropy |
| 8 | Diameter | 18 | Free energy |
| 9 | Average path length | 19 | Kirchhoff index (resistance distance) |
| 10 | Spectral gap | | |

Table 4.2: List of network properties used for comparing RFNs to random graph models.

normalisation is carried out as follows. Let $r_{i,0}$ be the property vector for the i th RFN, and $r_{i,j}$ be the property vector for the RRNG analogous to the corresponding RFN created with the j th elongation a of the rectangle, $1 \leq j \leq 9$, and for each elongation this is averaged over all random realisations to obtain a single vector for each. Each vector $r_{i,j}$ is then normalised as follows for $0 \leq j \leq 9$

$$\hat{r}_{i,j}(p) = \frac{r_{i,j}(p) - \min_j r_{i,j}(p)}{\max_j r_{i,j}(p) - \min_j r_{i,j}(p)}, \quad (4.3)$$

where $r_{i,j}(p)$ represents the p th entry of the $r_{i,j}$ vector. That is, for a given RFN and property, the value of this property is normalised for the RFN and all corresponding elongations of RRNG so that they lie between 0 and 1, with the smallest value mapped to 0 and the largest mapped to 1. We have centred the data on the minimum for simplicity since the goal is to ensure each range is the same length, and our results do not depend on the choice of the endpoints of the interval we map to.

We observe that for each of the rock fracture networks there is a minimum in the plot of the dissimilarity versus the rectangle elongation (see Fig. 4.6(a)), which indicates that there is an optimal elongation for each RRNG that makes it most similar to the real-world fracture network. In Fig. 4.6(b), the frequency with which the maximum similarity occurs for a given value of the rectangle elongation a is plotted. It can be seen that the histogram is two-peaked with the first maximum corresponding to elongations between $a = 1$ and $a = 2$ and the second one for elongations around $a = 4$. The

4. Applications

first peak clearly corresponds to rock fracture networks that are better reproduced by almost square neighbourhood graphs. However, the second group of real-world fracture networks are better reproduced by elongated rectangles in which one of the sides of the rectangle is about 16 times longer than the other. A more detailed statistical analysis of the histogram of the optimal elongations shows that the modal value of elongation is $a = 1.5$, and elongations between $a = 1$ and $a = 2$ are the most common, accounting for about $2/3$ of the rocks. The rest of the values lie between $a = 2.5$ and $a = 4.5$, except for a couple of rocks that have minimum at $a = 5$.

The most interesting observation carried out in this analysis is the following. Most of the rock fracture networks that are better reproduced by almost square RRNGs correspond to the smallest ones, while those that are better reproduced by elongated RRNGs are those having the largest number of nodes. These results are illustrated in Fig. 4.6(c and d) where the networks are split into two groups, those with $n \leq 150$, and those with $n > 150$. As can be seen in Fig. 4.6(c), which corresponds to the first group, the maximum similarity occurs for $1 \leq a \leq 2$. However, for the largest networks the maximum similarity occurs for values of $3 \leq a \leq 4$. In closing, the rock fracture networks are better described by the RRNGs depending on the size of the networks, with almost square RRNG describing better the smallest RFNs and more elongated RRNGs describing better the largest RFNs. In other words, it is more plausible that larger RFNs are those coming from elongated rocks, and consequently better reproduced by RRNGs with $a > 1$ which better reproduce this characteristic. The smallest RFNs appear to come from more rounded rocks, which are better reproduced by almost-square RRNGs.

The RRNGs reproduce relatively well the main structural properties of real-world RFNs of different sizes. This conclusion is reached by the fact that the dissimilarity between the RFNs and the RRNGs is generally smaller than the dissimilarity between the RFNs and other commonly used random graph models. In Fig. 4.7 we compare the use of the RRNGs and the Erdős–Rényi (ER), Barabási–Albert (BA), and random rectangular graph (RRG) models. Notice that all the random graphs must be connected in order to calculate some of the structural parameters. For the ER model the value

4. Applications

of $p = \frac{1}{n} \log n$ was chosen, which is the critical value at which the large connected component appears, with discarding of any random realisations in which the networks were disconnected. For the RRG model, again elongations from $a = 1$ to $a = 5$ in steps of 0.5 were used, and the best elongations were selected for each rock individually in exactly the same manner as for the RRNG model. In every case, the radius r was selected to ensure the RRG was connected with high probability, since selecting r to best match the number of edges in the RFN would almost always give a disconnected graph. In each case, several random realisations were used and the results averaged, with fewer realisations used in the case of a large number of nodes for reasons of computational difficulty.

With this data, the models were compared to see which of them works the best. Let $r_{i,0}$ be the property vector for the i th RFN, and $r_{i,j}$ be the property vector for the j th model, $1 \leq j \leq 4$, where for the RRNG and the RRG only the best elongation is used. Then, the normalisation takes exactly the same form as Eq. 4.3. After this normalisation, dissimilarities between the RFNs and each model were calculated. As can be seen in Fig. 4.7(a) the RRG model is the worst, with a dissimilarity larger than 5 for all RFNs, which is more than double the largest dissimilarity for any other model. To better compare the remaining models, the RRG model was removed from the analysis and the dissimilarities were recomputed with the remaining three models, since the presence of the data from the RRG model affects the normalisation values. The results are illustrated in Fig. 4.7(b) where we observe that the ER model is much worse than the other two, so this model was also removed. The result is the comparison of the best two models, the RRNG and the BA, which is illustrated in Fig. 4.7(c). In the large majority of cases the RRNG is observed to be the best model to reproduce the topological properties of the rock fracture networks due to a smaller value of dissimilarity.

4. Applications

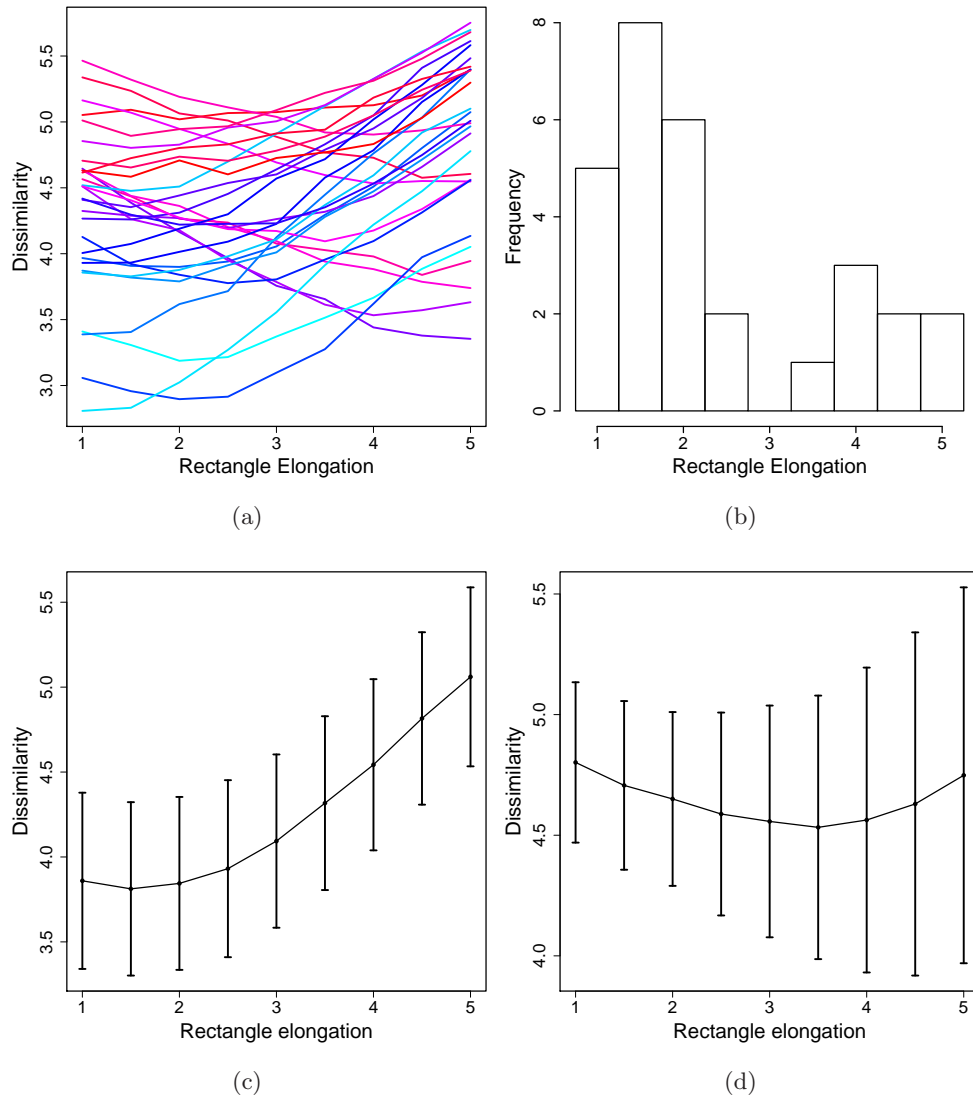


Figure 4.6: (a) Variation of the dissimilarity between RFNs and the corresponding RRNGs for different values of the rectangle elongation a . Each curve corresponds to one RFN-RRNG pair. (b) Number of graphs for which the minimum dissimilarity is a given value of the rectangle elongation a . Average variation of the dissimilarity between RFNs and the corresponding RRNGs for different values of the rectangular elongation a for RFNs with (c) less than $n = 150$ nodes and (d) more than $n = 150$ nodes. The vertical bars represent the standard deviation.

4. Applications

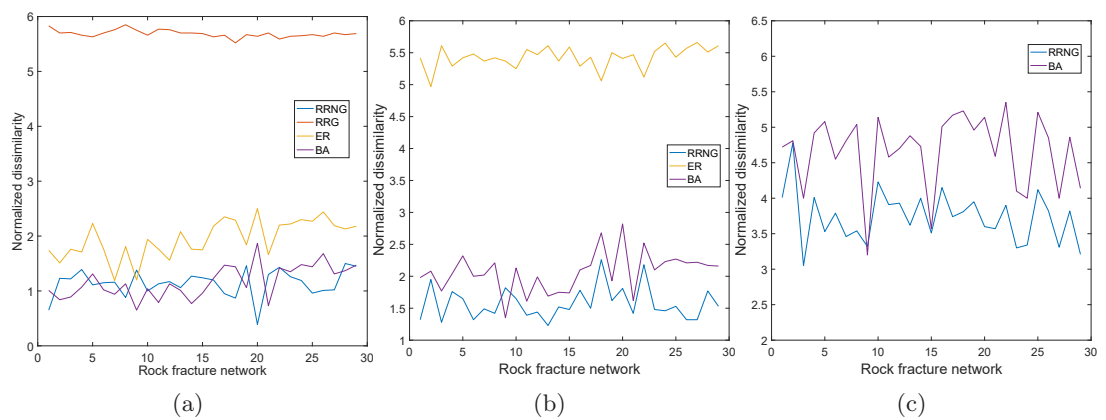


Figure 4.7: Relative normalised dissimilarity (the dissimilarity after the data has been normalised as described in the main text) between the RFNs and four theoretical models: random rectangular neighbourhood networks (RRNG), random rectangular graphs (RRG), Erdős–Rényi (ER) and Barabási–Albert (BA). From left to right the iterative process described in the main text is shown where the least similar model is eliminated until the two best models remain.

4. Applications

4.3.3 Fluid Diffusion on RFNs and RRNGs

We now consider the process of diffusion on RFNs and RRNGs. In order to select the initial condition vector u_0 two different scenarios are considered. The first one considers the case in which only a few nodes of the rock fracture network are in contact with the reservoir as illustrated in Fig. 4.8(a). This initial condition is used mainly to study the influence of the flow directionality on the diffusion through the rock fracture network and will be referred to as directional diffusion. In this case the vector u_0 is constructed such that the entry $u_{0,i}$ is a random number for those nodes considered to be in contact with the reservoir, and all random numbers used in the initial condition are i.i.d. uniform random variables from $[0, 1]$. The nodes are randomly selected from the set of nodes of the graph and should have the condition that they are close in space to each other. This set of points is selected to include one of the two extremes of the longest path (diameter) of the graph and nearby nodes. The rest of the entries of the initial condition vector are set to zero. The second scenario is based on the assumption that the RFN is in contact with the reservoir from many different positions as illustrated in Fig. 4.8(b), and will be called isotropic diffusion. In this case the entries $u_{0,i}$ of the initial condition vector are selected randomly for each of the nodes of the graph. The results of these two types of initial conditions are highly correlated, with a Pearson correlation coefficient of 0.999. The directional diffusion takes longer, since the oil must spread along the diameter of the rock from one end in this case, whereas in the isotropic case there is already oil spread (unevenly) throughout the rock at time $t = 0$. On average, the directional diffusion takes 1.44 times longer than the isotropic one. Due to these similarities, hereafter only the case where the whole rock is in contact with the reservoir is considered (Fig. 4.8(b)).

We consider that the diffusion process has taken place if $|u(i, t) - u(j, t)| \leq \delta$ for all pairs of nodes v_i and v_j in the graph when $t \rightarrow \infty$. In this work a threshold of $\delta = 10^{-5}$ is selected. This means that if the ‘concentration’ of the diffusive particle in one node does not differ from that in any other node by more than $\delta = 10^{-5}$ it is

4. Applications

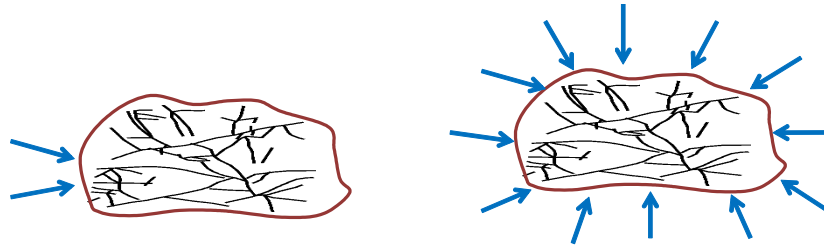


Figure 4.8: Illustration of the two scenarios used for modelling the diffusion of oil and gas from a reservoir through the fracture network of a rock. (a) Only a small region of the rock is in contact with the reservoir (directional). (b) The whole rock is in contact with the reservoir (isotropic).

considered that the diffusion process has ended. Then, the time at which this threshold is reached is recorded, and is called the diffusion time. Due to the fact that many realisations of the same process are carried out the average of this time is taken over all these realisations. Once the values of the average time of diffusion are obtained for each RFN, we calculate the correlations with the structural parameters considered in this work in order to see which of them influence the diffusive process on the RFNs. Table 4.3 reports the Pearson correlation coefficients of these relations, where the entries in red are those that are more significant from a statistical point of view. The next subsection will analyse the theoretical foundations for these findings.

Figure 4.9 illustrates the correlation between the average diffusion time obtained by simulation of the diffusive process on the RFNs versus the same process simulated on the analogous RRNGs. In addition, it provides evidence that the average diffusion time simulated on the RFNs is correlated to some of the most important structural parameters of the RRNGs. This means that the real-world RFNs can be replaced by their analogous random models in order to simulate the diffusion processes taking place on the rocks.

4.3.4 Theoretical analysis

As can be seen from Eq. 1.54 the average diffusion time in a network inversely correlates with the second smallest eigenvalue of the Laplacian matrix (see also the negative Pearson correlation coefficient in Table 4.3). This analysis clearly indicates that μ_2 can

4. Applications

| structural parameter | correlation | log | structural parameter | correlation | log |
|----------------------|---------------|-----|----------------------|--------------|-----|
| 1 | 0.026 | | F_1 | 0.815 | Y |
| 3 | -0.307 | | F_2 | 0.415 | |
| 4 | 0.225 | | F_3 | 0.775 | Y |
| 5 | -0.613 | | F_4 | 0.764 | Y |
| 6 | 0.142 | | F_5 | 0.423 | |
| 7 | -0.127 | | F_6 | 0.418 | |
| 8 | 0.809 | Y | F_7 | 0.085 | |
| 9 | 0.745 | Y | F_8 | 0.387 | |
| 10 | -0.495 | Y | F_9 | 0.419 | |
| 11 | 0.449 | | F_{10} | 0.435 | |
| 12 | -0.995 | Y | F_{11} | 0.416 | |
| 13 | 0.841 | Y | F_{12} | 0.409 | |
| 14 | 0.104 | | F_{13} | 0.381 | |
| 15 | 0.328 | | F_{14} | 0.424 | |
| 16 | 0.872 | | F_{15} | 0.389 | |
| 17 | 0.876 | | F_{16} | 0.407 | |
| 18 | -0.845 | | F_{17} | 0.000 | |
| 19 | 0.923 | Y | F_{18} | 0.412 | |

Table 4.3: Results of the regression analysis between the diffusion time in RFNs and structural parameters of the same networks. The numerical values correspond to the Pearson correlation coefficient, with entries larger than 0.7 in magnitude highlighted in red, and ‘log’ indicates whether the correlation is in a log-log scale. The numbers used for the structural parameters are given in Table 4.2 and the structure of the fragments is given in Appendix A.

be used as an estimator of the rate of diffusion of oil and gas in rock fracture networks. Using our previous results in which we found a bound for the diameter and for the algebraic connectivity of RRNGs, we see why the diffusion time correlates relatively well with the diameter of the graph. That is, as the elongation increases, the algebraic connectivity decays as a consequence of the increase of the diameter of the graph, which make the diffusion time increase. In other words, increasing the elongation of the RRNGs makes the diffusion process take longer to finish.

A lower bound for the algebraic connectivity has also been reported by [111] in terms of the average path length $\bar{l}(G)$ of the graph, which explains the correlation

4. Applications

obtained between the diffusion time and $\bar{l}(G)$

$$\mu_2(G) \geq \frac{4}{2(n-1)\bar{l}(G) - (n-2)}. \quad (4.4)$$

On the other hand, the high correlation observed between the diffusion time and the so-called Kirchhoff index can also be understood by using the relation in Eq. 1.54 because the Kirchhoff index is defined by

$$Kf = \sum_{i < j} \sum_{k=2}^n \frac{1}{\mu_k} (\psi_{k,i} - \psi_{k,j})^2, \quad (4.5)$$

from which it can easily be seen that the largest contribution is made by the second smallest eigenvalue of the Laplacian matrix μ_2 and its corresponding eigenvector (the Fiedler vector).

The somehow unexpected relations are those observed between the diffusion time and Estrada index, the entropy, the free energy and the average communicability angle, which are based on the adjacency instead of on the Laplacian matrix of the graph. These relations can be understood through the structural interpretation of the Estrada index in term of subgraphs of the graph. This index can be written as

$$EE(G) = tr(A^0) + tr(A) + \frac{1}{2}tr(A^2) + \frac{1}{6}tr(A^3) + \frac{1}{24}tr(A^4) + \frac{1}{120}tr(A^5) + \dots \quad (4.6)$$

Clearly, $tr(A^0) = n$ and $\frac{1}{2}tr(A^2) = m$ (notice that the adjacency matrix is traceless due to the lack of any self-loops in the networks). It is well-known that $\frac{1}{6}tr(A^3) = F_2$ is the number of triangles in the graph. In a similar way $tr(A^4) = 2m + 4F_1 + 8F_5$ and $tr(A^5) = 30F_2 + 10F_6 + 10F_8$. Consequently, the Estrada index of a graph can be written as

$$EE = n + \frac{13}{12}m + \frac{1}{6}F_1 + \frac{5}{4}F_2 + \frac{1}{3}F_5 + \frac{1}{12}F_6 + \frac{1}{12}F_8 + \dots, \quad (4.7)$$

which indicates that this index can be expressed as a weighted sum of the number of small fragments in the graph. The correlation coefficients between the diffusion time

4. Applications

of the RFNs and their number of nodes and edges are 0.863 and 0.846, respectively. Similarly, there are high correlations with the fragments F_1 , F_3 , and F_4 . Because the RFNs are not very dense and as previously observed they do not contain a large number of triangles (indeed there are a few RFNs that are triangle-free), this can be approximated as

$$EE \approx an + bm + cF_1 + d, \quad (4.8)$$

where a, b, c, d are coefficients. Linear regression analysis makes an estimation of these parameters as

$$EE \approx 0.85n + 0.93m + 0.40F_1 + 1.56, \quad (4.9)$$

with a Pearson correlation coefficient $r > 0.99999$ when including all 29 RFNs studied here. Then, we conclude that the correlation observed between the Estrada index and the diffusion time in RFNs is due to the fact that the diffusion time is well described by a few small fragments of the graphs, namely the number of nodes, edges and paths of length 2 (F_1), which are the main contributors to the Estrada index in these graphs. The correlations with F_3 , and F_4 observed in Table 4.3) can be explained by the fact that these fragments are numerous and are related to the fragment F_1 , since they contain it as a subgraph with one fewer node, and we confirm this fact in Fig. 4.10. Finally, the relatively high correlations observed for the entropy, the free energy and the average communicability angle can be explained the fact that all of these measures are in some way related to the Estrada index of the graph.

4. Applications

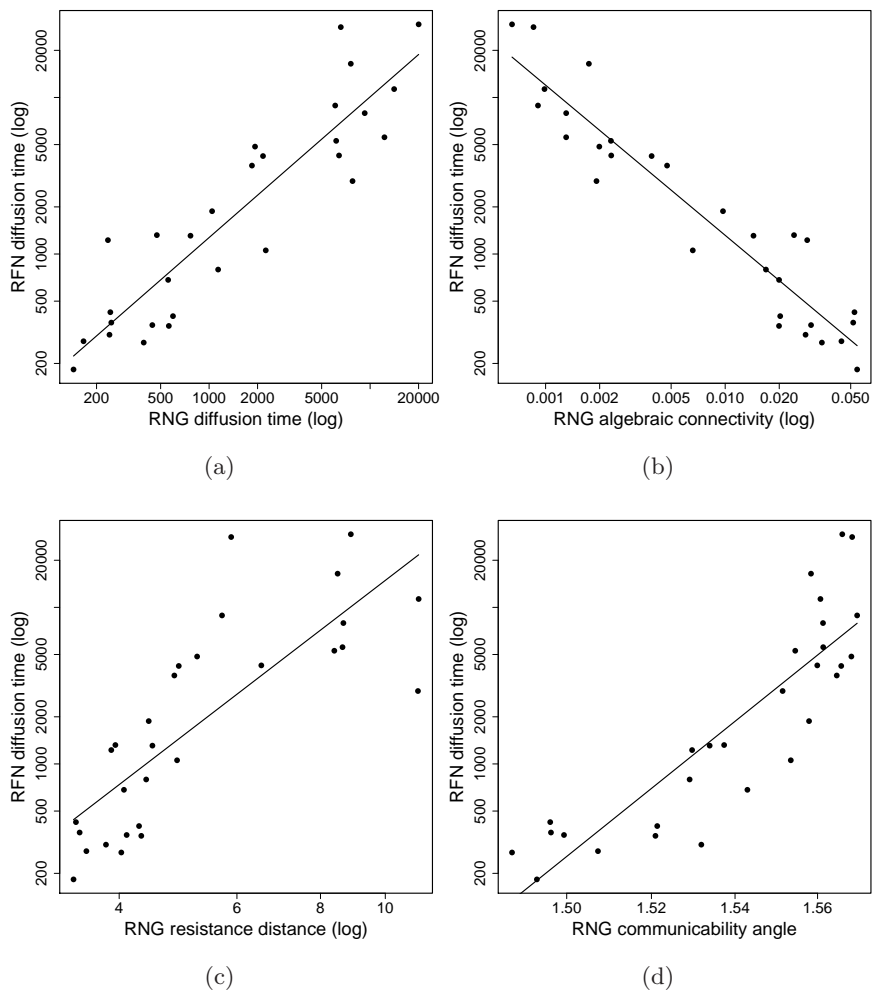


Figure 4.9: Plots of the average diffusion time of the RFNs against different properties of RRNGs: (a) diffusion time, (b) algebraic connectivity, (c) resistance distance (Kirchhoff index), and (d) average communicability angle.

4. Applications

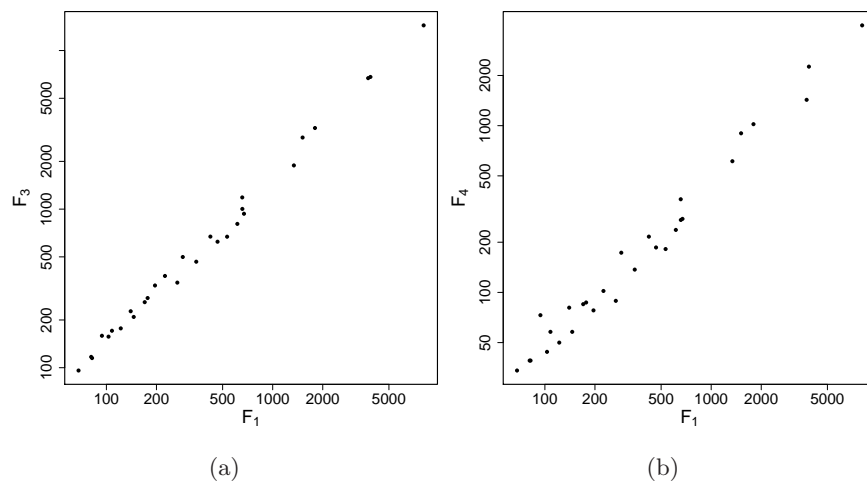


Figure 4.10: Plots of the correlations between fragment F_1 and (a) F_3 , (b) F_4 , in RFNs (log-log scale).

Chapter 5

Conclusions

We have introduced a generalisation of two classes of proximity graphs to explicitly consider the shape of the region in which the nodes are distributed. Instead of taking a unit square, or in general a unit hypercube, we considered a simple generalisation to unit rectangles or hyperrectangles. In particular, we considered the generalisation of the random geometric graph model and the relative neighbourhood graph model.

For the RRG model, we found an analytic expression for the expected value of the average node degree in terms of the elongation of the rectangle and the connection radius. We also found useful bounds for the diameter, the average path length, and the algebraic connectivity, as well as approximations to the degree distribution, connectivity, and clustering coefficient. As the elongation is increased, the diameter and average path length increase, which in turn contributes to a decrease in the algebraic connectivity. We proved that the average node degree is a non-increasing function of the rectangle elongation, and since the spectral radius is highly correlated with the average node degree, this causes the spectral radius to decrease as the elongation is increased. We discussed the relevance of the peak at -1 in the spectrum of the adjacency matrix of the RRG, which is not observed in the other random graph models we used for comparison, in terms of structures relating to neighbourhoods of nodes found within the graphs as a result of the way in which they are constructed.

Similarly, we find that in RRNGs increasing the elongation of the rectangle causes a decrease in the average node degree and an increase in the diameter. Then, we

5. Conclusions

can confidently state that the elongation parameter in these models is something new and useful that has a significant impact on both important topological properties and dynamical properties of these graphs.

The first application we discussed was diffusion in RRGs. Since the algebraic connectivity is closely related to the behaviour of diffusive processes and this decreases as the elongation is increased, we find that increasing the elongation parameter increases the time taken for diffusive particles to spread out evenly. We give numerical results demonstrating how the behaviour of diffusive processes in RRGs change as the parameters of the model are varied.

Secondly, we have studied the propagation of diseases in plants using the RRG model, deriving analytically a lower bound of the epidemic threshold for a SIS or SIR model running on these networks. This model is appropriate for the simulation of disease spreading on plants allocated on plots and fields of varied shapes. RRGs account for the spatial distribution of nodes allowing the variation of the shape of the unit square commonly used in RGGs. We have shown here by using analytical results and extensive numerical simulations of the SIS dynamics for different values of the elongation a and a fixed radius r that elongating the plots/fields in which the nodes (plants) are distributed makes the network more resilient to the propagation of epidemics. This is due to the fact that the epidemic threshold increases with the elongation of the rectangle. These results agree with a large accumulation of empirical evidence about the role of plot/field elongation on the propagation of diseases on plants. This model represents a new way to analyse disease propagation on plants or similar scenarios, by combining the heterogeneities introduced at individual level by networks with the influence produced by the shape variation of the plots and fields where the plants are growing.

Finally, we have then applied these two models to the problem of finding the most suitable model for capturing the topological properties of RFNs using a comparison of many important structural parameters. After fitting the models to each of the RFNs, especially the elongation parameter of the RRG and RRNG models, the results clearly show that the RRNG model is the best at replicating the properties of the RFNs. It

5. Conclusions

seems that the RRG model does not have the ideal characteristics even though it is spatially defined. The RRNG model is also better than the ER and BA models, which we expected since those models are not geometric in nature, and so cannot capture the spatial information of fracture networks or the varying elongations of the rocks. We also observed strong correlations between the diffusion time in RFNs and key structural properties of RRNGs, which is also an indication that the model performs very well for mimicking the properties of RFNs.

We found that the properties of small rock fracture networks are well-described by RRNGs with small values of the rectangle elongation. In contrast, larger RFNs are better described by RRNGs with significantly longer elongations. This means that small RFNs are embedded into more spherical rocks than the larger ones, which are mainly embedded into rocks with a higher aspect ratio. The most important characteristic of the RRNG is that it makes it possible to study a large variety of structural and dynamical processes by changing the parameters of the model. This allows us to be more independent of the existence of appropriate datasets of real-world RFNs, which in many cases are scarce.

We note that the heat equation adapted to consider the graph as the space in which the diffusive process is taking place is an appropriate tool for studying the diffusion of oil and gas on RFNs. We studied the relationship between the structure of rock networks and the diffusion dynamics taking place on them. In particular, a set of a few structural parameters were obtained that describe well the diffusive process taking place on the networks. Of special interest is the algebraic connectivity, which shows a correlation coefficient larger than 0.99 with the average diffusion time on RFNs. This index is then a very good predictor of the capacity of a given network to diffuse substances through the channels produced by the fractures in the rock. All in all, we consider that the newly proposed model based on random rectangular neighbourhood graphs is very flexible and can be adapted to specific simulation requirements, and is therefore appropriate to model RFNs in a variety of scenarios.

Overall, we can say that the RRG model and the RRNG model are useful generalisations of previous models for understanding various phenomena that can be represented

5. Conclusions

by a spatially embedded network.

5.1 Future Directions

Although we have performed a detailed study of these models in different applications, there are of course potential improvements that could be made, especially if more data were available. Here we discuss some of the ways in which future work could build upon the analyses presented here. As before, we have more to say about the modelling of RFNs since we had access to the data for several such networks.

5.1.1 Topological properties of RRGs and RRNGs

In this work we have opened a new research area by investigating some of the important topological properties of RRGs and RRNGs. We succeeded in finding the expected value of the average node degree of RRGs. We also found approximations or bounds for the degree distribution, diameter, average path length, clustering coefficient, algebraic connectivity, spectral radius and connectivity of RRGs, and for the average node degree, diameter and algebraic connectivity for RRNGs. We recall that the clustering coefficient and connectivity of RRNGs are trivial. Then, there is potential for improving these approximations or bounds of these important topological properties, perhaps even finding an exact analytical formula for the expected value.

For example, the clustering coefficient of RRGs is difficult to pin down precisely in an analytical manner, but we have presented two observations which are a step towards finding a useful approximation. The connectivity of RRGs is also difficult to calculate precisely - we saw that a known method provides a good estimate but the wrong trend and a modification that has the correct trend but unfortunately a slightly worse approximation, though both fare well in the important region near the critical radius. We also note that the bound given for the algebraic connectivity of RRGs has the correct trend, but is not tight.

Also, there are properties of these models for which we do not discuss any theoretical results, such as the spectral gap or the maximum node degree, and they could be the

5. Conclusions

subject of future work.

5.1.2 Epidemics in crop fields

We have considered the SIS model in RRGs to numerically model the spread of a disease through a crop field. Though there are reasons to expect that the SIR model is qualitatively similar to the SIS model, it may be worth performing the numerical simulations to examine this in more detail. Furthermore, we note that the model may be extended to make it more realistic, such as accounting for the prevailing wind direction by altering the shape of the connection disc around each node. This should be important since in an elongated field a wind blowing along the long axis of the field will help an infection at the upwind end of the field propagate and infect all the crops, whereas a wind blowing perpendicularly to this should help to limit the spread of the disease.

5.1.3 Influence of fracture aperture in RFNs

Any model is always a simplification of the reality made on the basis of a series of physical assumptions, empirical observations and availability of mathematical and computational tools to solve it. In this work we have used a few of these simplifications to produce a simple but effective modelling of the diffusion of fluids through rock fracture networks. However, there are a few areas in which improvements can be implemented in order to gain more realistic description of the physical processes taking place. One of our assumptions in this model is that all the fractures have the same aperture. This assumption is of course far from real, but it was used due to the lack of data about such apertures for the rock fracture networks that we studied in this work. However, such data is not difficult to obtain experimentally and we will give here some hints about how to implement this parameter and how it would affect the modelling results.

When assuming that all fracture apertures were identical we used unweighted graphs to represent the RFNs, i.e., every edge in the graph received an identical unit weight. If information concerning the aperture of the fractures were available we can represent it in our model by transforming the graphs representing RFNs into weighted graphs, in

5. Conclusions

which every edge receives a weight corresponding to the aperture of that fracture. As we do not have such data for the RFNs studied here we assume that such apertures are randomly and independently distributed from $[0, 1]$ over the fractures of a network. We then repeat our simulations for these weighted graphs representing RFNs with different, randomly distributed, apertures. We report here the average of the diffusion times for 10 random realisations.

In Fig. 5.1 we show the correlation between the diffusion time in the weighted RFNs and the unweighted RFNs in a log-log scale. It is clear that there is a strong correlation between the two results, with a correlation coefficient of 0.870. It is obvious that the diffusion times will change in dependence of the type of aperture that we use. For instance, the average times using apertures in the range $[0, 1]$ is larger than that when using all weights equal to one. If apertures of larger magnitude were used, an acceleration of the diffusion process should be observed, with significantly smaller diffusion times than those observed for the unweighted case. However, what is most important here is that a power-law relation exists between the diffusion time in the unweighted and the randomly weighted networks. That is,

$$\langle t_c \rangle \sim \langle \tilde{t}_c \rangle^\kappa, \quad (5.1)$$

where \tilde{t}_c is the time of diffusion in the weighted RFN and $\kappa \approx 0.92$ is a fitting parameter obtained by using nonlinear regression analysis for the data displayed in Fig. 5.1. More work is needed to show whether this kind of power-law relation exists in general between these two parameters, which would indicate some kind of universal scaling between the network with identical apertures for all the fractures and that with apertures randomly and independently distributed. In the meantime, we can assert that based on the current results knowing the behaviour of diffusion on the unweighted RFNs also gives information about the behaviour of the diffusion when the aperture of the fractures is randomly and independently distributed.

5. Conclusions

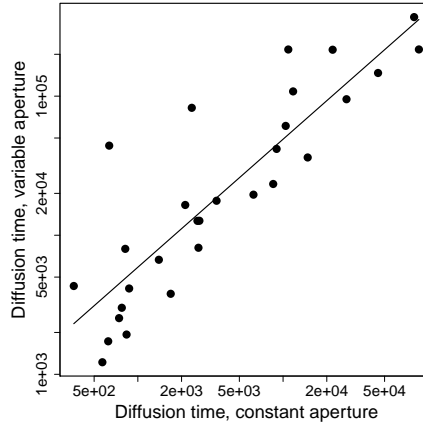


Figure 5.1: Plot of the diffusion time for the weighted RFNs against the diffusion time with all edge weights equal to 1 (log-log scale).

5.1.4 Influence of long-range hops of diffusive particles

A potential criticism of the analysis of RFNs is that it only considers normal diffusion. However, there is currently a vast volume of literature suggesting the lack of homogeneity in many unconventional naturally fractured reservoirs of oil [2]. These studies point out to the existence of complex combinations of connected and isolated pores, combinations of regions with discrete and continuous fractures and variable properties of the hydrocarbon properties across the reservoir [26]. These geological and petrophysical complexities cannot be described by using the simple diffusion equation and much more sophisticated models have emerged, which exploit the fractal nature of such irregularities. It has been well documented that such inhomogeneities in the properties of the systems to be modelled play a major role in the diffusion processes taking place, which are quite similar to anomalous diffusion in disordered media. Then, it is normal to model the geostatistics of these reservoirs by a fractional Brownian motion and fractional Lévy motion [80, 144].

In this work we will follow a different path that connects in a natural way with our previous model based on the normal diffusion equation on graphs. We consider here a generalisation of this equation using the so-called d -path Laplacian operators introduced in [50]. In a recent work, Estrada et al. [53] have proved analytically the existence of anomalous diffusion—superdiffusive and ballistic behaviour—when this model is used

5. Conclusions

in infinite one-dimensional systems. This work also shows that in the case of finite graphs, the biggest possible acceleration of diffusion is obtained for certain parameters of the model in any graph, independently of its topology.

Let us now define the d -path Laplacian matrices that account for the hopping of the diffusive particle to non-nearest-neighbours in the graph. Let $P_l(v_i, v_j)$ denote a shortest-path of length l between v_i and v_j . The nodes v_i and v_j are called the endpoints of the path $P_l(v_i, v_j)$. Because there could be more than one shortest path of length l between the nodes v_i and v_j we introduce the following concept. The irreducible set of shortest paths of length l in the graph is the set $P_l = \{P_l(v_i, v_j), P_l(v_i, v_r), \dots, P_l(v_s, v_t)\}$ in which the endpoints of every shortest-path in the set are different. Every shortest-path in this set is called an irreducible shortest-path. Now, we have generalised the Laplacian matrix to the so-called d -path Laplacian matrices that are defined as follows.

Let $d \leq d_{max}$. The d -path Laplacian matrix, denoted by $\mathcal{L}_d \in \mathbb{R}^{n \times n}$, of a connected graph of n nodes is defined as

$$\mathcal{L}_d(i, j) = \begin{cases} k_d(i) & \text{if } i = j, \\ -1 & \text{if } d_{ij} = d, \\ 0 & \text{otherwise,} \end{cases} \quad (5.2)$$

where $k_d(i)$ is the number of irreducible shortest-paths of length d that are originated at node v_i , i.e., its d -path degree. We can now define the generalised diffusion equation in which the d -path Laplacian operators are transformed by certain coefficients that make that the hopping probability of the diffusive particle decay with the distance that the particle is going to hop. Estrada et al. have analysed mathematically three different transforms of the d -path Laplacian operators and proved that for the infinite linear chain there is superdiffusive behaviour when the operators are transformed by using the Mellin transform with $1 < s < 3$. Hereafter we adopt this generalised diffusion equation that can be written in the following way:

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) = - \left(\sum_{d=1}^{d_{max}} d^{-s} \mathcal{L}_d \right) \mathbf{u}(x, t), \quad \mathbf{u}(x, 0) = \mathbf{u}_0. \quad (5.3)$$

5. Conclusions

Obviously, when $s \rightarrow \infty$ the terms $d^{-s} \rightarrow 0$ for all $d > 1$, and we recover the normal diffusion equation (1.60). In this framework we should expect just the normal diffusion to take place. However, when $s \rightarrow 0$, the system behaves as a fully-connected graph in which the diffusive particle can hop to any other node with identical probability.

We have calculated the diffusion time using the generalised diffusion equation (5.3) with $s = 1$ for all the RFNs studied here and compare the generalised diffusion times with those of the normal diffusion process. The normal diffusion time averaged for all RFNs is 12 315, while that for the generalised process is 0.69. If we consider the ratio of both times, i.e., the diffusion time under the normal conditions $\langle t_c \rangle$ and the diffusion time under long range hops of the diffusive particle $\langle \hat{t}_c \rangle$, we see that on average it is 42 784. In other words, the time for diffusion on the RFNs decreases 40 thousand times when we consider long-range hops. In Fig. 5.2 we illustrate the ratio $\langle t_c \rangle / \langle \hat{t}_c \rangle$ as the effect of long-range hops for the 29 real-world fracture networks studied in this work. The most interesting thing in these results is the fact that this ratio is dramatically larger than the average for 3 of the RFNs studied. In these three cases the ratio $\langle t_c \rangle / \langle \hat{t}_c \rangle$ is ten times larger than the mean of this value for all the networks. Although this is not a signature of superdiffusion it is worth further investigation to determine whether superdiffusive behaviour is observed for these three networks under the long-range hop scheme. This, however, is out of the scope of the current work and we leave it for a further and more complete investigation of this phenomena.

Another important characteristic of the results obtained in this subsection of the work is the lack of correlation between $\langle t_c \rangle$ and $\langle \hat{t}_c \rangle$ (graphic not shown). In contrast to what we have observed for the case of fracture aperture where a power-law relation was observed between the time considering random apertures and that using a fixed one, such a relation does not exist here. This lack of correlation between $\langle t_c \rangle$ and $\langle \hat{t}_c \rangle$ could be indicative of a different physical nature of the processes of diffusion under normal conditions and the diffusion under long range hops of the diffusive particle. This could show that the multi-hop approach to diffusion captures some phenomenology not captured by the normal diffusion process, which may include the case of superdiffusive behaviour.

5. Conclusions

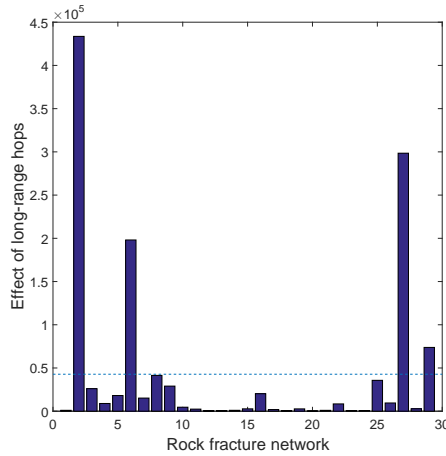


Figure 5.2: Bar plot of the values of the ratio $\langle t_c \rangle / \langle \hat{t}_c \rangle$ used to account for the effect of long-range hops on the 29 real-world fracture networks studied in this work. The horizontal broken line represents the average value of this ratio for the 29 RFNs.

5.1.5 Influence of 3D environments

Another characteristic of the current model that can be easily incorporated into the study of diffusive processes is its extension to higher-dimensional environments. We have considered here the 2-dimensional problem only due to the fact that they are suitable for the systems we consider, or the data we have. We consider WSNs spread over a geographical region, plants in rectangular crop fields, and RFNs corresponding to 2-dimensional rock slices. However, it is obvious that the fracture networks cover the 3-dimensional space of the rock and extend over its volume, and other spatial networks such as neurons in a brain are 3-dimensional.

For modelling purposes we should remark that now the number of parameters increases and that more kinds of shapes emerge. While for the case of the 2D scenario we can have only square-like or elongated rectangle-like frameworks, in 3D we have the following three main possibilities of environments: (i) a cube, (ii) an elongated cuboid, (iii) a flat cuboid. We show them in Fig. 5.3. There are of course many possible choices of the parameters in these models, which will cover a large variety of shapes in 3D space.

To give a flavour of the differences between the 2- and 3-dimensional RRNGs we study here the change in the average degree, the diameter, the algebraic connectivity

5. Conclusions

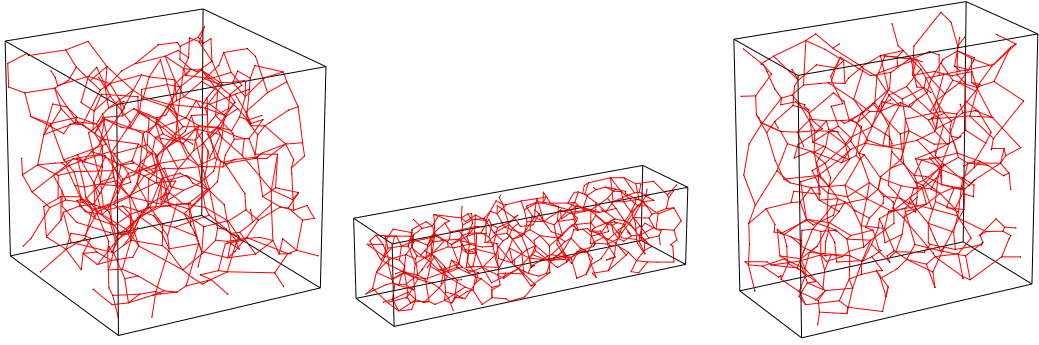


Figure 5.3: Illustration of some of the possibilities for modelling rock fracture networks in 3-dimensional space.

and the diffusion time with the elongation in the 2- and 3-dimensional RRNGs (see Fig. 5.4). To make things comparable we study here only the case one side of the cube changes, say $1 \leq a \leq 5$ in steps for 0.5, keeping $b = 1/a$ and $c = 1$. As for the case of the 2D graphs, the elongation of the cube produces a decay of the average degree, an increase of the diameter and a decrease of the algebraic connectivity. The resulting effects on the diffusion is that elongation delays the diffusion process. It happens, as expected, that the average degree of the 3D RRNG is larger than that of the 2D analogue. This is due to the fact that nodes can now be connected in three directions of space instead of two. More interestingly, the increase of the diameter of the 3D model is much slower than that in the case of the 2D one. For instance, the diameter increases almost linearly with the elongation according to $d_{max} \approx 14.8 + 40.66a$ for the 2D case, while it is $d_{max} \approx 11.1 + 13.53a$ for the 3D case. That is, the growth of the diameter in the 3D model studied is almost four times slower than in the 2D case for similar elongations. These results can be understood by adapting our previous bound for the 3D case. In this case we should consider that the separation between the points in the 3D cube is given by $1/\sqrt[3]{n}$, such as

$$d_{max}(G) \geq \sqrt[3]{n} \sqrt{a^2 + b^2}. \quad (5.4)$$

Then, it is clear that the diameter of the 3D RRNGs grows more slowly than that of the 2D ones. This of course produces a dramatic decay of the algebraic connectivity

5. Conclusions

with the elongation in the 3D model ($\mu_2 = 0.0246a^{-1.87}$), where this parameter drops much faster with the elongation than in the 2D case ($\mu_2 = 0.00307a^{-1.92}$). The main consequence of this elongation effect is observed in the slower growth of the diffusion time for the cuboid model than for the rectangular one. While in the 2D case the diffusion time increases as $\langle t_c \rangle \approx 0.064 + 0.093a$ with the elongation, in the 3D case it grows as $\langle t_c \rangle \approx 0.059 + 0.028a$. That is, the elongation of the cuboid seems to have an effect on the diffusion time that is three times smaller than the effect of elongation of the rectangle.

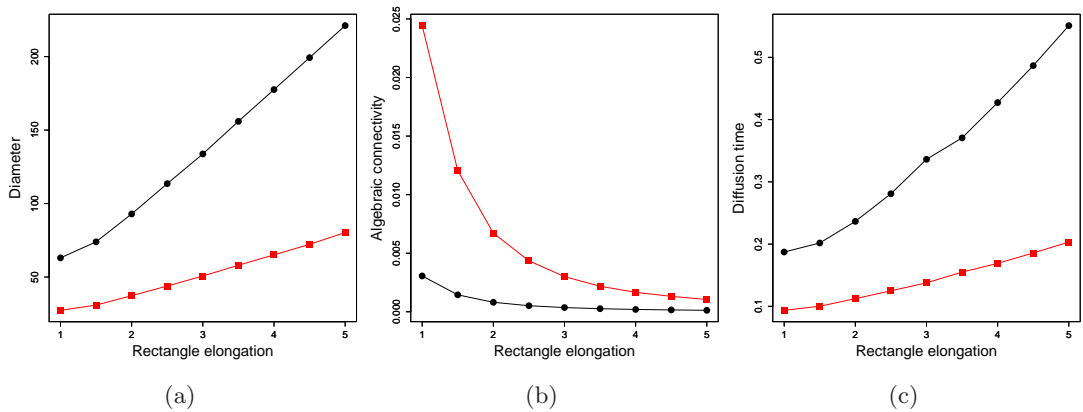


Figure 5.4: Comparison of the effects of elongation on (a) the diameter, (b) the algebraic connectivity, and (c) diffusion time in 2- (black dots) and 3-dimensional (red squares) RRNGs having $n = 1000$ nodes.

In closing we can think that the combination of the three effects studied here as potential extensions of our model will make the diffusion of oil and gas much faster than what we have predicted using the normal 2D diffusion model with constant apertures of the fractures. That is, if we consider apertures larger than one, include long-range hops of the diffusive particles and consider a 3D space instead of a 2D one, we will observe super-fast diffusion on the rock fracture networks studied. We should consider such combinations when some experimental data is available, which permit us to compare our theoretical predictions with reality. Also, the extension of the models to cuboids in 3-dimensions should be of great interest to the study of other spatial networks.

Appendix A

Small Subgraphs

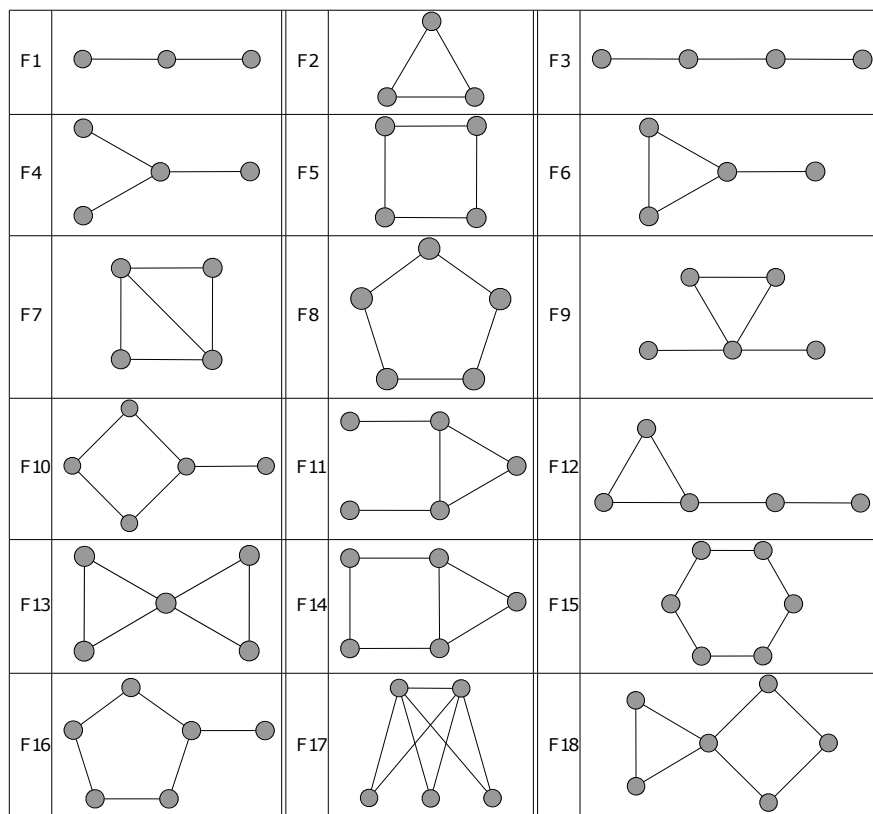


Figure A.1: Illustration of the structure of the small subgraphs used in this work.

Bibliography

- [1] P M Adler and J F Thovert. *Fractures and Fracture Networks. Theory and Applications of Transport in Porous Media*. Springer, Netherlands, 1999.
- [2] A Albinali, R Holy, H Sarak, and E Ozkan. Modeling of 1D anomalous diffusion in fractured nanoporous media. *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles*, 71(4):56, 2016.
- [3] N Alon and V D Milman. λ_1 , Isoperimetric inequalities for graphs, and super-concentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.
- [4] L Alonso, J A Méndez-Bermúdez, A González-Meléndrez, and Yamir Moreno. Weighted random-geometric and random-rectangular graphs: spectral and eigenfunction properties of the adjacency matrix. *Journal of Complex Networks*, 2017.
- [5] R M Anderson, R M May, and B Anderson. *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library, 1992.
- [6] J Andersson and B Dverstorp. Conditional simulations of fluid flow in three-dimensional networks of discrete fractures. *Water Resources Research*, 23(10):1876–1886, 1987.
- [7] C A Andresen, A Hansen, R Le Goc, P Davy, and S Mongstad Hope. Topology of fracture networks. *Frontiers in Physics*, 1:7, 2013.
- [8] A Antonioni and M Tomassini. Degree correlations in random geometric graphs. *Phys. Rev. E*, 86:037101, 2012.

Bibliography

- [9] N T J Bailey. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [10] A-L Barabási and R Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [11] A Baronchelli and A Díaz-Guilera. Consensus in networks of mobile communicating agents. *Phys. Rev. E*, 85:016113, Jan 2012.
- [12] A Barrat, M Barthélemy, and A Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [13] M Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [14] C C Barton. *Fractal Analysis of Scaling and Spatial Clustering of Fractures*, pages 141–178. Springer US, Boston, MA, 1995.
- [15] B Berkowitz. Characterizing flow and transport in fractured geological media: A review. *Advances in Water Resources*, 25(8-12):861–884, 2002.
- [16] P Blackwell, M Edmondson-Jones, and J Jordan. *Spectra of adjacency matrices of random geometric graphs*. University of Sheffield. Department of Probability and Statistics, 2007.
- [17] S Boccaletti, V Latora, Y Moreno, M Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [18] D Bogatkov and T Babadagli. Characterization of fracture network system of the midale field. In *Canadian International Petroleum Conference*. Petroleum Society of Canada, 2007.
- [19] M Boguñá, R Pastor-Satorras, and A Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.*, 90:028701, Jan 2003.

Bibliography

- [20] E Bonnet, O Bour, N E Odling, P Davy, I Main, P Cowie, and B Berkowitz. Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39(3):347–383, 2001.
- [21] F Bonnot, H De Franqueville, and E Lourenço. Spatial and spatiotemporal pattern analysis of coconut lethal yellowing in Mozambique. *Phytopathology*, 100(4):300–312, 2010.
- [22] S P Borgatti and M G Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.
- [23] C P Brooks, J Antonovics, and T H Keitt. Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *The American Naturalist*, 172(2):149–159, 2008.
- [24] J Buhl, J Gautrais, R V Solé, P Kuntz, S Valverde, J L Deneubourg, and G Theraulaz. Efficiency and robustness in ant networks of galleries. *The European Physical Journal B - Condensed Matter and Complex Systems*, 42(1):123–129, 2004.
- [25] M C Cacas, E Ledoux, G de Marsily, A Barbreau, P Calmels, B Gaillard, and R Margritta. Modeling fracture flow with a stochastic discrete fracture network: Calibration and validation: 2. the transport model. *Water Resources Research*, 26(3):491–500, 1990.
- [26] R Camacho Velazquez, M A Vasquez-Cruz, and G Fuentes-Cruz. Recent advances in dynamic modeling of naturally fractured reservoirs. Offshore Technology Conference, 2012.
- [27] A Cardillo, M Zanin, J Gómez-Gardeñes, M Romance, A J García del Amo, and S Boccaletti. Modeling the multi-layer nature of the European air transport network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics*, 215(1):2333, 2013.
- [28] C Castellano and R Pastor-Satorras. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.*, 105:218701, Nov 2010.

Bibliography

- [29] D Chakrabarti, Y Wang, C Wang, J Leskovec, and C Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10(4):1–26, Jan 2008.
- [30] J Coon, C P Dettmann, and O Georgiou. Full connectivity: Corners, edges and faces. *Journal of Statistical Physics*, 147(4):758–778, 2012.
- [31] J Coon, C P Dettmann, and O Georgiou. Impact of boundaries on fully connected random geometric networks. *Phys. Rev. E*, 85:011138, Jan 2012.
- [32] J P Coon, O Georgiou, and C P Dettmann. Connectivity in dense networks confined within right prisms. In *2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 628–635, May 2014.
- [33] L da F Costa, O N Oliveira Jr, G Travieso, F A Rodrigues, P R Villas Boas, L Antigueira, M P Viana, and L E Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.
- [34] J Dall and M Christensen. Random geometric graphs. *Phys. Rev. E*, 66:016121, Jul 2002.
- [35] K Damen, A Faaij, and W Turkenburg. Health, safety and environmental risks of underground CO₂ storage – overview of mechanisms and current knowledge. *Climatic Change*, 74(1):289–318, 2006.
- [36] B Damjanac and P A Cundall. Validation of lattice approach for rock stability problems. In *47th US Rock Mechanics/Geomechanics Symposium*. American Rock Mechanics Association, 2013.
- [37] N M M de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423(1):53–73, 2007.
- [38] M H Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

Bibliography

- [39] M Desai and D Manjunath. On the connectivity in finite ad hoc networks. *IEEE Communications Letters*, 6(10):437–439, 2002.
- [40] C P Dettmann, O Georgiou, and J P Coon. More is less: Connectivity in fractal regions. In *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pages 636–640, Aug 2015.
- [41] C P Dettmann and G Knight. Symmetric motifs in random geometric graphs. *arXiv preprint arXiv:1704.00640*, 2017.
- [42] A Díaz-Guilera, J Gómez-Gardeñes, Y Moreno, and M Nekovee. Synchronization in random geometric graphs. *International Journal of Bifurcation and Chaos*, 19(02):687–693, 2009.
- [43] Y Ederly, S Geiger, and B Berkowitz. Structural controls on anomalous transport in fractured porous rock. *Water Resources Research*, 52(7):5634–5643, 2016.
- [44] R B Ellis, J L Martin, and C Yan. Random geometric graph diameter in the unit ball. *Algorithmica*, 47(4):421–438, 2007.
- [45] P Erdős and A Rényi. On the evolution of random graphs. *Selected Papers of Alfréd Rényi*, 2:482–525, 1976.
- [46] E Estrada. Characterization of 3D molecular structure. *Chemical Physics Letters*, 319(5-6):713–718, 2000.
- [47] E Estrada. Quantifying network heterogeneity. *Phys. Rev. E*, 82:066102, Dec 2010.
- [48] E Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Oxford, 2011.
- [49] E Estrada. The communicability distance in graphs. *Linear Algebra and its Applications*, 436(11):4317–4328, 2012.

Bibliography

- [50] E Estrada. Path Laplacian matrices: Introduction and application to the analysis of consensus in networks. *Linear Algebra and its Applications*, 436(9):3373–3391, 2012.
- [51] E Estrada, A A Alhomaiddhi, and F Al-Thukair. Exploring the “Middle Earth” of network spectra via a Gaussian matrix function. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(2):023109, 2017.
- [52] E Estrada and G Chen. Synchronizability of random rectangular graphs. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):083107, 2015.
- [53] E Estrada, E Hameed, N Hatano, and M Langer. Path Laplacian operators and superdiffusive processes on graphs. I. one-dimensional case. *Linear Algebra and its Applications*, 523:307–334, 2017.
- [54] E Estrada and N Hatano. Statistical-mechanical approach to subgraph centrality in complex networks. *Chemical Physics Letters*, 439(1-3):247–251, 2007.
- [55] E Estrada and N Hatano. Communicability angle and the spatial efficiency of networks. *SIAM Review*, 58(4):692–715, 2016.
- [56] E Estrada, S Meloni, M Sheerin, and Y Moreno. Epidemic spreading in random rectangular networks. *Phys. Rev. E*, 94:052316, 2016.
- [57] E Estrada and J A Rodríguez-Velázquez. Spectral measures of bipartivity in complex networks. *Phys. Rev. E*, 72:046105, Oct 2005.
- [58] E Estrada and M Sheerin. Random rectangular graphs. *Phys. Rev. E*, 91:042805, Apr 2015.
- [59] E Estrada and M Sheerin. Consensus dynamics on random rectangular graphs. *Physica D: Nonlinear Phenomena*, 323-324:20–26, 2016.
- [60] E Estrada and M Sheerin. Random neighborhood graphs as models of fracture networks on rocks: Structural and dynamical analysis. *Applied Mathematics and Computation*, 314:360–379, 2017.

Bibliography

- [61] E Estrada and G Silver. Accounting for the role of long walks on networks via a new matrix function. *Journal of Mathematical Analysis and Applications*, 449(2):1581-1600, 2017.
- [62] D Estrin, R Govindan, J Heidemann, and S Kumar. Next century challenges: Scalable coordination in sensor networks. In *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, MobiCom '99*, pages 263–270, New York, NY, USA, 1999. ACM.
- [63] Francis J Ferrandino. The explicit dependence of quadrat variance on the ratio of clump size to quadrat size. *Phytopathology*, 95(5):452–462, 2005.
- [64] M Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.
- [65] R A Fleming, L M Marsh, and H C Tuckwell. Effect of field geometry on the spread of crop disease. *Protection Ecology (Netherlands)*, 4(2):81–108, 1982.
- [66] C H Foh, G Liu, B S Lee, B-C Seet, K-J Wong, and C P Fu. Network connectivity of one-dimensional MANETs with random waypoint movement. *IEEE Communications Letters*, 9(1):31–33, 2005.
- [67] M Franceschelli, A Giua, and C Seatzu. Distributed averaging in sensor networks based on broadcast gossip algorithms. *Sensors Journal, IEEE*, 11(3):808–817, March 2011.
- [68] T T Garipov, M Karimi-Fard, and H A Tchelepi. Discrete fracture model for coupled flow and geomechanics. *Computational Geosciences*, 20(1):149–160, 2016.
- [69] E N Gilbert. Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543, 1961.
- [70] A Girón, H Saiz, F S Bacelar, R F S Andrade, and J Gómez-Gardeñes. Synchronization unveils the organization of ecological networks with positive and negative interactions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(6):065302, 2016.

Bibliography

- [71] E Godehardt and J Jaworski. On the connectivity of a random interval graph. *Random Structures & Algorithms*, 9(1-2):137–161, 1996.
- [72] A Goel, S Rai, and B Krishnamachari. Sharp thresholds for monotone properties in random geometric graphs. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 580–586, New York, NY, USA, 2004. ACM.
- [73] S Gómez, A Arenas, J Borge-Holthoefer, S Meloni, and Y Moreno. Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *EPL (Europhysics Letters)*, 89(3):38009, 2010.
- [74] P Gupta and P R Kumar. *Critical Power for Asymptotic Connectivity in Wireless Networks*, pages 547–566. Birkhäuser Boston, Boston, MA, 1999.
- [75] J J Han, T H Lee, and W M Sung. Analysis of oil production behavior for the fractured basement reservoir using hybrid discrete fractured network approach. *Advances in Petroleum Exploration and Development*, 5(1):63–70, 2013.
- [76] T P Handford, F J Pérez-Reche, S N Taraskin, L da F Costa, M Miazaki, F M Neri, and C A Gilligan. Epidemics in networks of spatially correlated three-dimensional root-branching structures. *Journal of The Royal Society Interface*, 2010.
- [77] J Hansford and Q Fisher. The influence of fracture closure from petroleum production from naturally fractured reservoirs: A simulation modelling approach. *The Oral Presentation at AAPG Annual Convention, Denver, Colorado, USA*, 2009.
- [78] A Hasan, B Foss, and S Sagatun. Flow control of fluids through porous media. *Applied Mathematics and Computation*, 219(7):3323–3335, 2012.
- [79] M Henkel, H Hinrichsen, S Lübeck, and M Pleimling. *Non-equilibrium phase transitions*, volume 1. Springer, 2008.

Bibliography

- [80] T A Hewett. Fractal distributions of reservoir heterogeneity and their influence on fluid transport. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 1986.
- [81] D J Higham, M Rašajski, and N Pržulj. Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics*, 24(8):10931099, 2008.
- [82] N Higham. *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.
- [83] A M Hitchmough, M S Riley, A W Herbert, and J H Tellam. Estimating the hydraulic properties of the fracture network in a sandstone aquifer. *Journal of Contaminant Hydrology*, 93(1-4):38–57, 2007.
- [84] N M Holden, A J Brereton, R Fealy, and J Sweeney. Possible change in Irish climate and its impact on barley and potato yields. *Agricultural and Forest Meteorology*, 116(3):181196, 2003.
- [85] S Holloway. Storage of fossil fuel-derived carbon dioxide beneath the surface of the earth. *Annual Review of Energy and the Environment*, 26(1):145–166, 2001.
- [86] S Holloway. Underground sequestration of carbon dioxide—a viable greenhouse gas mitigation option. *Energy*, 30(11):2318–2333, 2005.
- [87] S Hope, P Davy, J Maillot, R Le Goc, and A Hansen. Topological impact of constrained fracture growth. *Frontiers in Physics*, 3:75, 2015.
- [88] S Hope, S Kundu, C Roy, S Manna, and A Hansen. Network topology of the desert rose. *Frontiers in Physics*, 3:72, 2015.
- [89] O Huseby, J F Thovert, and P M Adler. Geometry and topology of fracture systems. *Journal of Physics A: Mathematical and General*, 30(5):1415, 1997.
- [90] D E Ingber. Tensegrity II. how structural networks influence cellular information processing networks. *Journal of Cell Science*, 116(8):1397–1408, 2003.

Bibliography

- [91] V Isham, J Kaczmarska, and M Nekovee. Spread of information and infection on finite random networks. *Phys. Rev. E*, 83:046128, Apr 2011.
- [92] A Jafari and T Babadagli. Estimation of equivalent fracture network permeability using fractal and statistical network properties. *Journal of Petroleum Science and Engineering*, 92-93:110–123, 2012.
- [93] W C James and C S Shih. Size and shape of plots for estimating yield losses from cereal foliage diseases. *Experimental Agriculture*, 9(1):63–71, Jan 1973.
- [94] Y H Jang, T H Lee, J H Jung, S I Kwon, and W M Sung. The oil production performance analysis using discrete fracture network model with simulated annealing inverse method. *Geosciences Journal*, 17(4):489–496, 2013.
- [95] J W Jaromczyk and G T Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9):1502–1517, Sep 1992.
- [96] M J Jeger, M Pautasso, O Holdenrieder, and M W Shaw. Modelling disease spread and control in networks: implications for plant sciences. *New Phytologist*, 174(2):279–297, 2007.
- [97] P Jund, R Jullien, and I Campbell. Random walks on fractals and stretched exponential relaxation. *Phys. Rev. E*, 63:036131, Feb 2001.
- [98] J Katajainen. The region approach for computing relative neighbourhood graphs in the L_p metric. *Computing*, 40(2):147–161, 1988.
- [99] J Katajainen, O Nevalainen, and J Teuhola. A linear expected-time algorithm for computing planar relative neighbourhood graphs. *Information Processing Letters*, 25(2):77–86, 1987.
- [100] L Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):3943, Mar 1953.
- [101] H Kenniche and V Ravelomananana. Random geometric graphs as model of wireless sensor networks. In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, volume 4, pages 103–107, Feb 2010.

Bibliography

- [102] D J Klein and M Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.
- [103] K Koike, C Liu, and T Sanga. Incorporation of fracture directions into 3D geostatistical methods for a rock fracture system. *Environmental Earth Sciences*, 66(5):1403–1414, 2012.
- [104] M Kowaluk. Planar β -skeletons via point location in monotone subdivisions of subset of lunes. *ArXiv e-prints*, Nov 2014.
- [105] J C S Long and D M Billaux. From field data to fracture network modeling: An example incorporating spatial structure. *Water Resources Research*, 23(7):1201–1216, 1987.
- [106] O Lordan, J M Sallan, P Simo, and D Gonzalez-Prieto. Robustness of the air transport network. *Transportation Research Part E: Logistics and Transportation Review*, 68:155–163, 2014.
- [107] B D MacArthur and R J Sánchez-García. Spectral characteristics of network redundancy. *Phys. Rev. E*, 80:026117, Aug 2009.
- [108] B D MacArthur, R J Sánchez-García, and J W Anderson. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531, 2008.
- [109] J Marro and R Dickman. *Nonequilibrium phase transitions in lattice models*. Cambridge University Press, 2005.
- [110] M Mesbahi and M Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton Series in Applied Mathematics. Princeton University Press, 2010.
- [111] B Mohar. Eigenvalues, diameter, and mean distance in graphs. *Graphs and Combinatorics*, 7(1):53–64, 1991.
- [112] C C Mundt and L S Brophy. Influence of number of host genotype units on the effectiveness of host mixtures for disease control: a modeling approach. *Phytopathology (USA)*, 1988.

Bibliography

- [113] C C Mundt, L S Brophy, and S C Kolar. Effect of genotype unit number and spatial arrangement on severity of yellow rust in wheat cultivar mixtures. *Plant Pathology*, 45(2):215–222, 1996.
- [114] M Nekovee. Worm epidemics in wireless ad hoc networks. *New Journal of Physics*, 9(6):189, 2007.
- [115] F M Neri, A Bates, W S Führtbauer, F J Pérez-Reche, S N Taraskin, W Otten, D J Bailey, and C A Gilligan. The effect of heterogeneity on invasion in spatial epidemics: From theory to experimental evidence in a model system. *PLOS Computational Biology*, 7(9):1–9, 09 2011.
- [116] F M Neri, F J Pérez-Reche, S N Taraskin, and C A Gilligan. Heterogeneity in susceptible-infected-removed (SIR) epidemics on lattices. *Journal of The Royal Society Interface*, 2010.
- [117] S P Neuman. *Stochastic Continuum Representation of Fractured Rock Permeability as an Alternative to the REV and Fracture Network Concepts*, pages 331–362. Springer Netherlands, Dordrecht, 1988.
- [118] M Newman. *Networks: An Introduction*. OUP Oxford, 2010.
- [119] M E J Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, Oct 2002.
- [120] M E J Newman. Random graphs as models of networks. *arXiv preprint cond-mat/0202208*, 2002.
- [121] D D Nolte, L J Pyrak-Nolte, and N G W Cook. The fractal geometry of flow paths in natural fractures in rock and the approach to percolation. *pure and applied geophysics*, 131(1):111–138, 1989.
- [122] R Olfati-Saber, J A Fax, and R M Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, Jan 2007.
- [123] R Olfati-Saber and Richard M M. Consensus protocols for networks of dynamic agents. In *Proceedings of the 2003 American Controls Conference*, 2003.

Bibliography

- [124] R Olfati-Saber and R M Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, Sept 2004.
- [125] F Papadopoulos, D Krioukov, M Boguna, and A Vahdat. Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces. In *2010 Proceedings IEEE INFOCOM*, page 19, March 2010.
- [126] R E Paysour and W E Fry. Interplot interference: A model for planning field experiments with aerially disseminated pathogens. *Phytopathology*, 73(7):1014–1020, 1983.
- [127] M Penrose. *Random geometric graphs*, volume 5. Oxford University Press Oxford, 2003.
- [128] S S Pereira and A Pages-Zamora. Consensus in correlated random wireless sensor networks. *Signal Processing, IEEE Transactions on*, 59(12):6279–6284, Dec 2011.
- [129] F J Pérez-Reche, S N Taraskin, L da F Costa, F M Neri, and C A Gilligan. Complexity and anisotropy in host morphology make populations less susceptible to epidemic outbreaks. *Journal of The Royal Society Interface*, 7(48):1083–1092, 2010.
- [130] A Perna, S Valverde, J Gautrais, C Jost, R Solé, P Kuntz, and G Theraulaz. Topological efficiency in three-dimensional gallery networks of termite nests. *Physica A: Statistical Mechanics and its Applications*, 387(24):6235–6244, 2008.
- [131] P Y Polubarinova-Kochina. Theory of ground water movement Princeton university press. *Princeton, NJ*, 1962.
- [132] G J Pottie and W J Kaiser. Wireless integrated network sensors. *Commun. ACM*, 43(5):51–58, May 2000.
- [133] V M Preciado and A Jadbabaie. Spectral analysis of virus spreading in random geometric networks. In *Proceedings of the 48th IEEE Conference on Decision and*

Bibliography

- Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 4802–4807, Dec 2009.
- [134] A Rényi and P Erdős. On random graphs. *Publicationes Mathematicae*, 6(290-297):5, 1959.
- [135] S Riley, K Eames, V Isham, D Mollison, and P Trapman. Five challenges for spatial epidemic models. *Epidemics*, (0), 2014.
- [136] H Saiz, J Gómez-Gardeñes, J P Borda, and F Maestre. The structure of plant spatial association networks increases plant diversity in global drylands. *bioRxiv*, 2017.
- [137] H Saiz, J Gómez-Gardeñes, P Nuche, A Girón, Y Pueyo, and C L Alados. Evidence of structural balance in spatial ecological networks. *Ecography*, 40(6):733741, 2017.
- [138] E Santiago, M Romero-Salcedo, J X Velasco-Hernández, L G Velasquillo, and J A Hernández. *An Integrated Strategy for Analyzing Flow Conductivity of Fractures in a Naturally Fractured Reservoir Using a Complex Network Metric*, pages 350–361. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [139] E Santiago, J X Velasco-Hernández, and M Romero-Salcedo. A methodology for the characterization of flow conductivity through the identification of communities in samples of fractured rocks. *Expert Systems with Applications*, 41(3):811–820, 2014. *Methods and Applications of Artificial and Computational Intelligence*.
- [140] E Santiago, J X Velasco-Hernández, and M Romero-Salcedo. A descriptive study of fracture networks in rocks using complex network metrics. *Computers & Geosciences*, 88:97–114, 2016.
- [141] S Sarkar, M N Toksoz, and D R Burns. Fluid flow simulation in fractured reservoirs. Technical report, Massachusetts Institute of Technology. Earth Resources Laboratory, 2002.

Bibliography

- [142] T Sarkus, M Tennyson, and D Vikara. Geologic carbon storage. *Fossil Fuels: Current Status and Future Directions*, 1:49, 2016.
- [143] S Seifollahi, P A Dowd, C Xu, and A Y Fadakar. A spatial clustering approach for stochastic fracture network modelling. *Rock Mechanics and Rock Engineering*, 47(4):1225–1235, 2014.
- [144] A T Silva, E K Lenzi, L R Evangelista, M K Lenzi, and L R da Silva. Fractional nonlinear diffusion equation, solutions and anomalous diffusion. *Physica A: Statistical Mechanics and its Applications*, 375(1):65–71, 2007.
- [145] P Skelsey, D E L Cooke, J S Lynott, and A K Lees. Crop connectivity under climate change: future environmental and geographic risks of potato late blight in Scotland. *Global Change Biology*, 22(11):37243738, 2016.
- [146] T A B Snijders. The degree variance: An index of graph heterogeneity. *Social Networks*, 3(3):163–174, 1981.
- [147] J E Streit and R R Hillis. Estimating fault stability and sustainable fluid pressures for underground storage of CO₂ in porous rock. *Energy*, 29(9):1445–1456, 2004.
- [148] Z Toroczkai and H Guclu. Proximity networks and epidemics. *Physica A: Statistical Mechanics and its Applications*, 378(1):68–75, 2007. Social network analysis: Measuring tools, structures and dynamics Social Network Analysis and Complexity.
- [149] G T Toussaint. The relative neighbourhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268, 1980.
- [150] D Urban and T Keitt. Landscape connectivity: a graph-theoretic perspective. *Ecology*, 82(5):1205–1218, 2001.
- [151] R Urquhart. Graph theoretical clustering based on limited neighbourhood sets. *Pattern Recognition*, 15(3):173–187, 1982.
- [152] R B Urquhart. Algorithms for computation of relative neighbourhood graph. *Electronics Letters*, 16(14):556–557, 1980.

Bibliography

- [153] L Valentini, D Perugini, and G Poli. The “small-world” topology of rock fracture networks. *Physica A: Statistical Mechanics and its Applications*, 377(1):323–328, 2007.
- [154] P Van Mieghem, J Omic, and R Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14, Feb 2009.
- [155] J N Vevatne, E Rimstad, S M Hope, R Korsnes, and A Hansen. Fracture networks in sea ice. *Frontiers in Physics*, 2:21, 2014.
- [156] L Von Collatz and U Sinogowitz. Spektren endlicher grafen. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 21(1):63–77, 1957.
- [157] P E Waggoner. Weather, space, time, and chance of infection. *Phytopathology*, 52(11):1100, 1962.
- [158] J Wang, H Mo, F Wang, and F Jin. Exploring the network structure and nodal centrality of China’s air transport network: A complex network approach. *Journal of Transport Geography*, 19(4):712–721, 2011.
- [159] P Wang and M C González. Understanding spatial connectivity of individuals with non-uniform population density. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1901):3321–3329, 2009.
- [160] D Watanabe. A study on analyzing the grid road network patterns using relative neighborhood graph. In *The Ninth International Symposium on Operations Research and Its Applications*, pages 112–119. Citeseer, 2010.
- [161] D J Watts and S H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [162] T H Wilson, V Smith, and A Brown. Developing a model discrete fracture network, drilling, and enhanced oil recovery strategy in an unconventional naturally fractured reservoir using integrated field, image log, and three-dimensional seismic data. *AAPG Bulletin*, 99(4):735–762, 2015.