Coventry University

DOCTOR OF PHILOSOPHY

Complex networks: shape, topology and spatial embedding

De Regt, Robin

Award date:
2018

Awarding institution:
Coventry University

COVENTRY UNIVERSITY

DOCTORAL THESIS

Complex Networks: topology, shape and spatial embedding

Supervisors:

Dr. Christian VON FERBER

Prof. Yurij HOLOVATCH

Prof. Ralph KENNA

Dr. Thierry PLATINI

Author:

Robin DE REGT

A thesis submitted in fulfillment of the requirements

for the degree of Doctor of Philosophy

in the

School of Computing, Electronics and Mathematics

April 11, 2018

Declaration of Authorship

I, Robin DE REGT, declare that this thesis titled, "Complex Networks: topology, shape and spatial embedding" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: April 11, 2018

"In retrospect, Euler's unintended message is very simple: Graphs or networks have properties, hidden in their construction, that limit or enhance our ability to do things with them. For more than two centuries the layout of Konigsberg's graph limited its citizens' ability to solve their coffeehouse problem. But a change in the layout, the addition of only one extra link, suddenly removed this constraint"

Albert Barabasi

COVENTRY UNIVERSITY

School of Computing, Electronics and Mathematics

Doctor of Philosophy

Complex Networks: topology, shape and spatial embedding

by Robin DE REGT

Abstract

Complex networks have recently become a vibrant branch of complexity science. Their ability to describe a variety of interacting systems, appearing in the natural and man-made environment, has led to its increased popularity in multiple scientific fields such as mathematics, physics, biology, computer science, sociology, epidemiology and many others. This thesis aims to further broaden the scope of complex networks by investigating physical models at a microscopic, mesoscopic and macroscopic scale. First we study the shapes of tree-like polymers comparing three different numerical and analytical methods (Wei, Benhamou and Monte Carlo methods). We find excellent agreement across all methods, indicating that increased branching generates more spherical objects. Secondly, UK Public transport networks are studied. Topological measures of robustness are investigated via the Molloy-reed parameter. Using fractal properties we extract information on the serviceability of stations and their efficiency. Thirdly, large scale structures of the universe are investigated. Here, we generate a network of the cosmic web to study its topological properties. Our main results indicate a correlation between clustering coefficient and the astrophysical properties of colour index and stellar mass.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Complexity Science is a relatively new term which has become quite popular over the last few decades. Often associated to multi-disciplinary research topics, Complexity Science does not refer to one particular area, but could be defined by the set of theories, and concepts from a variety of fields. One should mention that the definition of complexity itself is not unique, and strongly dependent on the scientific field. For example, in computation theory, time and space complexity refers to the amount of time and memory space needed to reach the solution of a given problem. However, in physics, applied-mathematics, economics, biophysics and mathematical biology, Complexity Science regroups, research focused on so called complex systems. Such systems are composed of elements (agents) connected to each other via relationships (interactions). The sum of these interactions often lead to a non-trivial collective behaviour which emerge naturally from the underlying rules governing each element. In physics, the study of self-organisation and critical phenomena gives a perfect illustration of what emerging collecting behaviour refers too. In particular, we talk about "emergence" when referring to macroscopic phenomenon, induced by the interaction of a large number of microscopic elements.

Many complex systems are perfectly described within the complex network formalism. The notion of a complex network itself has become one of the central notions of modern scientific discourse [1]. It is straightforward to find many examples of complex systems (natural or man-made), which one can present to the reader. The Internet is, most likely, one of the most cited examples. To cite just a few others, one can mention the brain (connection between neurons), social networks (interaction between individual), gene

co-expression networks (relation between genes involved in protein production), public transport networks, ...

It appears that many real world systems can quite simply be mapped onto a complex network. Under the right assumptions, elements of a given problem are reduced to nodes of a network while their interactions are mapped onto edges. While in some cases, interactions between agents may be tangible (the wires connecting electrical circuits or the synaptic nerves joining neurons), they may also reflect more abstract relations (the ties in the social networks or the trades and exchanges between countries). In epidemiology for example, the problem of diffusion of an infectious disease can be mapped onto a simple two-states (susceptible or infected) model evolving on a complex network (reflecting the structure of the social interactions).

The theory of complex network science can be traced back to the birth of graph theory and the famous seven bridges of Königsberg problem [2]. More than 250 years later, this theory has began to flourish. Receiving notable contributions from the fields of statistical physics, information theory and non-linear dynamics. Complex networks have successfully been applied to describe numerous natural and artificial systems. These include but are not limited to the study of protein interactions [3], mythological narratives [4], [5], archeology [6], scientific collaborations [7] and zoological systems [8].

Given the versatility and ability of Complexity Science to describe a range of interacting systems, it is unsurprising that the field has continued to grow, finding new applications, and often facilitating inter-disciplinary research [9]. This thesis offers further such examples. Applying the theory of networks to identify alternative perspectives from which to view a range of complex problems within the Sciences and beyond. Usually, the primary goal of representing a complex system as a network is to single out its topological features and analyse the impact of these features on the behaviour of the system as a whole. However, recently, emphasis has been placed in analysing the interplay between a systems topology and its geometry, i.e. its properties in Euclidean (usually 2D or 3D) space. In this thesis we aim to further contribute to this field. Studying the topology, shape and spatial embedding of several network-like systems of different origin, covering all scales from the microscopic to the macroscopic. In particular in this

thesis we will investigate the following properties of three different complex network systems:

- Polymers: Shape properties such as g-ratio, asphericity, prolateness, and form factors, of ideal comb and dendritic polymers.

- Public transport networks (PTN): The topological and geographical features of public transport networks in London, Manchester, the West Midlands and Bristol, together with the UK coach and rail networks.

- Large scale structures of the universe: Correlations that exist between the topological features of a network generated of the cosmic web and the astrophysical properties of galaxies.

This work aims to highlight the adaptiveness of complex network science and identify ways to further broaden its scope. The case studies presented exist on vastly different scales, and are formed via both natural and man-made phenomena. On the microscopic scale, natural or synthetic polymers are defined by the connection of multiple monomer units (via covalent bonds) to form long chain-like structures. At the opposite end of the spectrum, on a macroscopic scale, is the cosmic web, a conglomeration of many galaxies formed under the influence of gravity, dark energy and dark matter. Between these two extremes we consider public transport networks, a man-made mesoscopic system that is governed by economic forces stemming from the demand to facilitate large-scale human movement.

The layout of the rest of this thesis is as follows:

- In Chapter 2, we start with an introduction on complex network theory. We define fundamental tools such as the adjacency and Kirchhoff matrices, the average degree, degree distribution, as well as other notions such as assortativity, centrality, preferential attachment and small worldedness.

- In Chapter 3, we study the shape properties of several polymer macromolecules of different topology. The shape of polymers is known to significantly affect many physical properties such as their viscosity and melting point. We consider comb and dendritic polymers in the ideal regime in both 2D and 3D. Applying both

analytical and numerical approaches we compute shape parameters that appear to be in excellent agreement with each other. We confirm that comb polymers with a complete set of interior branches display a more spherical shape. In addition, dendrimers with increased branching also tend to be more spherical.

- In Chapter 4, we investigate public transport networks PTN using data derived from the National Public Transport Data Repository. We study the topological and spatial features of public transport networks including London, Manchester, West Midlands, Bristol, national coach and rail. We are able to rank PTNs in relation to their stability using methods from statistical physics. Moreover studying the fractal nature of these networks, allows for useful interpretations regarding the serviceable area of stations.

- In Chapter 5, we study the cosmic web; the network describing the large-scale structure of the Universe. Our goal is to find correlations between galaxy astrophysical (colour index and stellar mass) and topological (defined by location in the network) properties. We study 2D projections of galaxies spatial distributions in different redshift slices. For different galaxy samples, our results show a high similarity level in topology. We discuss statistically significant correlation found when selecting galaxies according to different modes of clustering coefficient.

Personal contribution of the researcher

The work presented was carried out by the author except in the case of collaborative research, as outlined below:

- The Monte Carlo simulations for shape properties of comb and dendritic polymers in section 3, were carried out by Marvin Bishop, John Stone, Adam Barillas and Tylor Borgeson at Manhatten College, Manhattan, New York, United States.

- The astrophysical interpretation of network properties in section 5, were carried out by Stepan Apunevych and Bohdan Novosyadlyj at the Ivan Franko National University of Lviv, Ukraine.

Research connection with scientific programs, plans and themes

This thesis has been prepared at Coventry University under the support the following projects: PhD program Collège Doctoral "Statistical Physics of Complex Systems" Leipzig-Lorraine-Lviv-Coventry ($\mathbb{L}^4$), FP7 EU IRSES projects 269139 "Dynamics and Cooperative Phenomena in Complex Physical and Biological Media" (DCP-PHYSBIO), 295302 "Statistical Physics in Diverse Realizations" (SPIDER), 612707 "Dynamics of and in Complex Systems" (DIONICOS).

Thesis approbation

The results of this thesis have been reported and discussed at the following scientific meetings:

- "Research Symposium"(Coventry, 6th May, 2014);

- VIII Scientific Conference "Selected Issues of Astronomy and Astrophysics (Lviv, Ukraine, 17th-20th Oct 2016);

- Joint Steering Committee meeting of the IRSES projects DIONICOS and STREVCOMS (Lviv, Ukraine, 14th-16th June 2017);

- 80th Annual meeting and spring meeting of the German Physical Society (Regensburg, Germany, 6th-11th March 2016);

- Cost meeting (Sofia, Bulgaria, 13th-14th March) and three seminars at the Statistical Physics of Complex Systems (ICMP, Lviv, Ukraine).

Publications

Material in this thesis has been published in: four papers, two in a peer reviewed journal [10], [11] and two on ArXive as preprints [12], [13] for which [12] has been accepted for publication in the journal: *Monthly Notices of the Royal Astronomical Society* and [13] is currently under going review at the journal: *Transportmetrica A: Transport Science*; and in five conference abstracts [14]–[18].

Chapter 2

Complex networks: a brief introduction

In this thesis we consider three different case studies, namely: the shape properties of comb and dendritic polymers; public transportation networks in the UK; and the COSMOS galaxy field all viewed as a complex networks. The methodology employed to investigate these three objects, is the common feature linking them. For this reason, this chapter focuses on the theory of complex networks. Here, we present the tools and main characteristics used to investigate their properties. Initially, in section 2.1 we give a brief historic account of significant events and contributions to the theory. In section 2.2 we introduce the main network observables and some classical models that have been developed, to describe complex networks.

2.1   Main historic events in complex network development

Complex network science can be traced back to 1741, when Leonhard Euler famously solved the seven bridges of Königsberg conundrum [2]. The problem consists of determining whether a path existed, such that, all seven bridges of Königsberg could be traversed once and once only (see Figure 2.1). Euler solved the puzzle by reducing it to a set of nodes and edges, showing that for such a path to exist at least one of two conditions must hold:

- either all nodes must have an even number of adjacent edges, so that a closed path can be drawn;

- only two nodes can possess three adjacent edges while all other nodes have an even number of connections. Under this constraint, the path must start and end on the two "odd nodes" (with an odd number of adjacent edges).

In so doing Euler proved that, in Königsberg, no such path existed. In general, a path satisfying the first or second condition is called a Eulerian path or semi-Eulerian path respectively. This historical event marked the beginning of a new form of discrete mathematics called graph theory.

Subsequently, what began as trivial pursuit started to find applications in the nineteen century when, in 1845, Gustav Kirchhoff, whilst attending the university of Könisgberg, formulated the laws that govern the flow of electricity within circuits. Kirchhoff's laws have had a significant impact to the field of electrical engineering and are still taught in undergraduate classes to this day. In 1861, graph theory also became one of the founding blocks of algebraic chemistry when Alexander Brown published his thesis entitled 'Theory of Chemical Combination' [19]. Here, he linked valence electrons to the number of edges adjacent a node (atom) and used this to determine the number of configurations of a chemical compound.

In the twentieth century a number of other scientific disciplines began to contribute and apply the theory of networks within other fields. Sociologist, started to use the concept of networks in their research. The most famous example being Stanley Milgram's experiment, in 1967, when he applied the concepts of networks to determine how connected people were to the rest of the world [20]. A decade earlier, in 1959, the famous mathematician, Erdös with his colleague Rényi [21] first published their work on random graphs. Such objects have been the subject of numerous studies ever since.However, it was not until the 1990s, after a number of complex networks had been studied, that scientist started to realise that rare events in complex networks occurred far more often than randomly expected. This realisation sparked the interest of the physics community and subsequently two more famous models of network generation were postulated. The first model, suggested by Watts and Strogatz [22], in 1998, offered insight into how a probabilistic model could generate networks that were both locally, highly correlated and globally, compactly connected at the same time. The second model, suggested by Albert and Barabási [23], in 1999, reproduced the power law

FIGURE 2.1: From left to right: Euler's first figure in Ref.[2], where he depicts the seven bridges of Königberg; Euler's graph representation of the seven bridges of Königsberg where nodes and edges reflect land masses and bridges respectively.

degree distributions, otherwise known as scale free behaviour, very often seen in complex networks.

During the course of the development of complex network science a large number of network observables have been developed. In the next section the network observables applied within this thesis are introduced and discussed in detail.

2.2 Network Observables

Let us start by defining a graph, G as a pair (V, E) where V and E are the sets of nodes and edges respectively. We write $N = |V|$ the total number of nodes. Note that E is a subset of $\Omega = V^2 = \{(u,v)|u, v \in V\}$, with $|\Omega| = N^2$ so that $|E| \leq N^2$. One should mention that in some cases, E is a multiset (instead of a set). This re-definition of E allows for the consideration of multi-edges (see next section for details) and for the construction of networks with a high number of edges i.e. $|E| > N^2$. In general, a network can be represented by various matrices, such as the adjacency matrix ($\hat{K}$) and the Kirchhoff matrix ($\hat{L}$). Both matrices and typical networks characteristics are defined in the following sub-sections. On numerous occasions we will refer to the book of Newman, "Networks: An introduction" [1]. This book provides extensive and detailed information on the fundamentals of network theory. It also includes many empirical studies on real complex networks. It is an ideal starting point for the reader looking for a more comprehensive and detailed overview of complex network theory.

2.2.1 Adjacency Matrix

A graph, G with N nodes can be represented by an $N \times N$ adjacency matrix $\hat{A}$ with elements

$$\hat{A}_{i,j} = \begin{cases} w_{i,j}, & \text{if } (i,j) \in E, \\ 0, & \text{if } (i,j) \notin E, \end{cases}$$

where $w_{i,j}$ is the weight associated to the edge (i,j). In the most general case, a network can be i) directed, ii) weighted, iii) accept multi-edges and iv) accept loops.

i) We talk about directed networks (or directed edges) when an edge (i,j) is said to be pointing from node i to node j. For such networks (i,j) and (j,i) are two distinct connections and the total number of edges is given by $M = |E|$. However, when the direction of an edge does not matter, we talk about undirected graphs. In this case, (i,j) and (j,i) represent the same link so that we have $(i,j) \in E \Leftrightarrow (j,i) \in E$. It follows that the total number of (undirected) edges is $M = |E|/2$, when self-edges (loops) are not permitted.

ii) Networks are said to be weighted, when a weight is associated to each edge. Artificial neural networks, some social networks, and numerous others fall into this category. The weights are usually positive numbers but, in some cases, negative weights are considered. For example, in the study of social networks generated from epic narratives, friendly and hostile interactions are associated with positive and negative weights respectively [5]. However, when all edges carry identical weight $w_{i,j} = 1$ we talk of unweighted networks.

iii) A network is said to have multi-edges when two nodes i and j can be connected by more than one single edge. In this case the set E should be extended to a multi set, allowing for the count of multi-edges. It follows that $w_{i,j}$ represents the number of multi-edges between nodes i and j. A graph is usually considered as not having multi-edges unless stated otherwise.

iv) Loops or self-edges appear in graph for which a node can connect to itself. The loop attached to node i is represented by the couple (i,i). Once again, unless

explicitly mentioned a graph is assumed to have no loops.

Simple networks (or simple graphs) are defined as unweighted, undirected graphs, presenting no multi-edges and no loops. For simple graphs, elements $\hat{A}_{i,j}$ (and $w_{i,j}$) are either 0 or 1. Moreover, the adjacency matrix is symmetric and presents zeros along the diagonal. In our work we will consider both undirected, weighted and unweighted networks.

The number of edges adjacent to a given node i is known as the node degree, k_i. With the help of the adjacency matrix we can now define the node degree and its distribution. We should mention that for directed networks, it is useful to define an in- and out-degree, associated to the number of edges pointing to or from a given node respectively. However, restraining ourself to undirected networks, the node degree (defined in terms of the adjacency matrix) reads:

$$k_i = \sum_{j=1}^{N} \hat{A}_{i,j}. \tag{2.1}$$

The mean of k^n, for arbitrary n, is defined by averaging of over all nodes in the network:

$$\langle k^n \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^n. \tag{2.2}$$

In particular it is easy to show that $\langle k \rangle = 2M/N$. The variance is defined in the usual way:

$$\mathrm{Var}(k) = \langle (k - \langle k \rangle)^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} (k_i - \langle k \rangle)^2. \tag{2.3}$$

Together $\langle k \rangle$ and $\langle k^2 \rangle$ (or $\mathrm{Var}(k)$) offer a description (up to second order) of the distribution of interactions within the network.

2.2.2 Degree distribution

The degree distribution and cumulative degree distribution are ubiquitous quantities in network theory. To introduce the degree distribution $P(k)$ it is convenient to rewrite Eq.(2.2) as

$$\langle k^n \rangle = \sum_{k} P(k) k^n, \tag{2.4}$$

where the sum over k spans all possible values of the node degree. $P(k)$ satisfies $\sum_k P(k) = 1$ and is given by $P(k) = N_k/N$, with N_k being the number of nodes with degree k. From here, the cumulative distribution is defined by:

$$P_{\geq}(k) = \sum_{q \geq k} P(q) \,. \tag{2.5}$$

In simple words, the distribution $P_{\geq}(k)$ gives the probability of selecting a node of degree greater than or equal to k. In fact, the degree distribution of complex networks, generated by real world data, often exhibits a significant amount of noise. This noise can usually be smoothed out by considering the cumulative distribution.

Investigations of the degree distribution of real networks has lead to the observation that numerous networks present a positively skewed degree distributions [1], [24]–[26]. This behaviour, for large degree k, signals the existence within a given network of high-degree nodes (also called hubs). In the presence of such a distribution, we commonly refer to the degree distribution as fat tailed.

2.2.3 Functional forms of the degree distribution

Many different functional forms have been used to describe the degree distribution of real networks. Amongst them; the Poissonian distribution and power law distribution (also called scale free) are described in this section. The reader may often encounter other distributions such as: truncated power-law distributions, exponential or Gaussian distributions, log-normal distributions, stretched exponential and Weibull distributions [1].

The Poisson distribution

In 1959, two models for random network generation were published: the Erdös-Rényi model, $G(N, M, p)$ [21], and the Gilbert model, $G(N, p)$ [27]. In the $G(N, M, p)$ model, one starts to generate a fixed set of nodes and proceeds by:

1. Connecting two nodes, which are randomly selected,

2. Iterating the previous step until a total of M edges are generated.

Gilbert's model ($G(N, p)$) has a slightly different algorithm. Starting with a fixed set of nodes, we first:

1. Connect two nodes with probability p,

2. Repeat the previous step, iterating through all pairs of nodes once and once only.

While in the first model, the number of links is fixed to M, in Gilbert's model this number can fluctuate between 0 and $N - 1$. However, when averaged over many configurations both models present a Poisson distribution. One should mention that other network characteristics such as average path length ($l \simeq \ln(N)$) and average clustering coefficient ($C \simeq p$) are also common to these two models. The latter quantities will be defined later in this chapter.

In Gilbert's model we can easily evaluate the degree distribution $P(k)$. We first note that the probability for a vertex to have degree k is proportional to $p^k(1 - p)^{N-1-k}$. Keeping count of all arrangements of k links, amongst $N - 1$ we have

$$P(k) = C_k^{N-1} p^k (1 - p)^{N-1-k}, \tag{2.6}$$

where C_k^{N-1} is the binomial coefficient. From the degree distribution one can easily estimate the mean and variance:

$$\langle k \rangle = p(N - 1) \tag{2.7}$$

$$\text{Var}(k) = p(1 - p)(N - 1). \tag{2.8}$$

It is when taking the limit N large, that the Poisson distribution arrises. Using the Stirling's approximation $N! \simeq \sqrt{2\pi} N^{N+1/2} e^{-N}$ on the binomial coefficient, leads to

$$P(k) \simeq \frac{(Np)^k}{k!} e^{-Np}, \tag{2.9}$$

with $\langle k \rangle \simeq pN$.

The Power laws distribution

In 1999, Albert and Barabási (AB) published a model reproducing the scale free behaviour prevalent in many complex systems [23]. At first, they tried growth models followed by preferential attachment models to no avail. It was not until they included both features of growth and preferential attachment into one model that scale free behaviour was observed. The algorithm developed is as follows:

1. Start with an initial configuration m_0 nodes.

2. Add a node u and from u generate m_0 links with the existing nodes.

3. Add a node $u + 1$. The probability to attach $u + 1$ to an existing node i is now chosen to be proportional to the degree k_i.

4. Iterate step three until the desired number of nodes is reached.

Figure 2.2 presents a network generated following the AB algorithm where $N = 20$ and $m_0 = 2$. The value of each node represents the order which nodes enter the network. As a result of preferential attachment it can be seen that earlier nodes tend to have higher degree nodes. A network generated under these rules of construction present a power law degree distribution of the form

$$P(k) \propto k^{-\alpha} . \tag{2.10}$$

In general, we say that a network has a power laws distribution (or is a scale free network) when its degree distribution can be approximated with $P(k) \sim Ck^{-\alpha}$, for some constant C. This is a common characteristic of a large number of networks for which typical values of α fall in the range $2 \leq \alpha \leq 3$ [24], [28]. We should mention that for a given network, the functional form is not expected to fit all k values (especially not small values of k) [29]. Usually, the expression $Ck^{-\alpha}$ is fitted to the tail of the data. Given $P(k) \propto k^{-\alpha}$, we easily see by integration, that the cumulative distribution is also a power law distribution,

$$P_{\geq}(k) \propto k^{1-\alpha} . \tag{2.11}$$
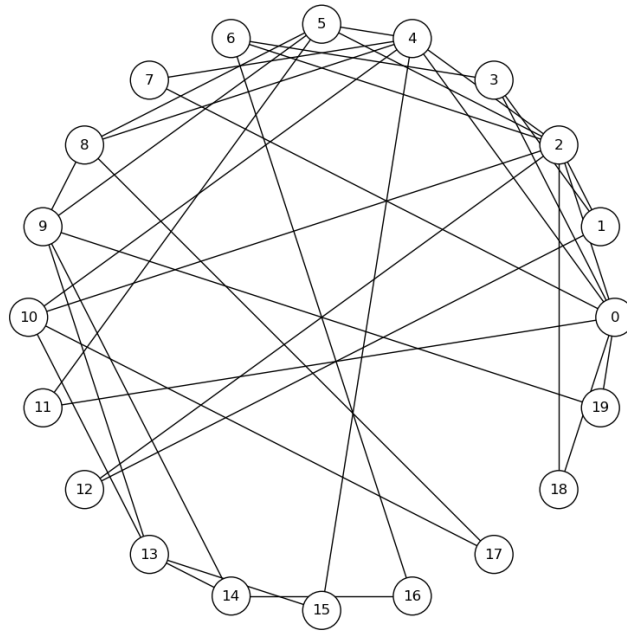
FIGURE 2.2: Example of Albert-Barabási preferential attachment network of twenty nodes and initialising with two nodes ($m_0 = 2$). The values for each node represents the sequence which these nodes where added to the network.

2.2.4 Kirchhoff matrix for simple graphs

For simple graphs the Kirchhoff matrix is defined as: $\hat{K} = \hat{D} - \hat{A}$, where $\hat{D}$ is diagonal ($D_{i,j} = 0$ for $i \neq j$) with elements given by $\hat{D}_{i,i} = k_i = \sum_j \hat{A}_{i,j}$. The Kirchhoff matrix is another useful representation of a network. The Kirchhoff matrix appears in many problems including random walks, resistor networks, graph partitioning. Its eigenvalue spectrum gives insight into some of the properties of networks. Written λ_i ($i \in \{1, 2, ..., N\}$) the eigenvalues of $\hat{K}$ are conventionally sorted in decreasing order $\lambda_1 \geq \lambda_2 \geq ...\lambda_n \geq 0$. For simple graphs, the following properties hold:

- The number of sub connected components s in G is given by $s = \sum\limits_{i=1}^{N} \delta(\lambda_i)$ with $\delta(x) = 1$ if $x = 0$ and zero otherwise.

- $\sum\limits_{i=1}^{N} \hat{K}_{i,i} = \sum\limits_{i=1}^{N} \lambda_i$,

- $|\hat{K}| = \prod\limits_{i=1}^{N} \lambda_i$,

- The eigenvalues of $\hat{K}^q$ (for positive integer values of q) are λ_i^q,

- For a fully connected graph G, the number of spanning trees t is given by $t = \frac{1}{N} \left(\prod_{i=1}^{N-1} \lambda_i \right)$, where t is a subgraph that is a tree which includes all nodes of G.

2.2.5 Clustering coefficient

There exist two definitions of the so called clustering coefficient, one is a local quantity while the other is defined globally. The global clustering coefficient is often called transitivity and is commonly used in sociology [30]. This quantity is defined by the ratio of total number of triangles N_T and the total number of triplets N_t (a triplet being defined by the nodes on a path of length 2) in the network: $C_T = 3N_T/N_t$. In this section we will focus on the local definition, the reader interested in learning more on the global clustering coefficient is invited to read [1]. The clustering coefficient can be understood as the measure of the average probability that two neighbours of a vertex are themselves neighbours. Let us consider a node i and all its nearest neighbours. The clustering coefficient C_i is defined by

$$C_i = \frac{2y_i}{k_i(k_i - 1)}, \qquad k_i \geq 2, \tag{2.12}$$

where y_i is the number of links, between the k_i nearest neighbours of the node i [22]. If neighbours of i are fully connected then $y_i = k_i(k_i - 1)/2$ so that $C_i = 1$ whereas if none of them are connected we have $C_i = 0$. Once all C_i, are known for all nodes, the average clustering coefficient is defined as

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i. \tag{2.13}$$

The clustering coefficient for a given graph is often compared with the clustering coefficient derived for a classical Erdös-Rényi random graph [21], [22] of the same size. The latter quantity being defined by:

$$C_{\mathrm{r}} = \frac{\langle k \rangle}{N - 1}. \tag{2.14}$$

Note that in the limiting case of a fully connected graph one has $\langle k \rangle = N - 1$ leading to $C_{\mathrm{r}} = 1$ while for an empty graph we have $\langle k \rangle = 0$ so that $C_{\mathrm{r}} = 0$. Often, real social

networks appear to present a high clustering coefficient. This characteristic is induced by the underlaying rules governing the formation of social networks. In Ref.[31], Newman studies the network of collaborations between physicists and shows how social dynamical rules tend to favour the creation of triangles. However, other networks, such as non-social networks (governed by rules other than human interactions) are believed to have a relatively lower clustering coefficients. The Internet's clustering coefficient has been evaluated [1] to be about 0.012, which is far lower than the "expected" value which would have been obtained if connections were random (0.84).

2.2.6 Assortativity

Agents in many complex systems tend to have similar attributes, this feature is often summarised in the well known saying "Birds of a feather flock together". This tendency is also known as homophily [32]. The degree assortativity r (or simple assortativity) provides a mathematical measure for the tendency to connect nodes of similar degree. This is given by the Pearson correlation coefficient [1] and is calculated as:

$$r = \frac{\sum_{i,j} \hat{A}_{i,j}(k_i - E[k])(k_j - E[k])}{E[k^2] - E[k]^2}, \tag{2.15}$$

where the expectation $E[.]$ is defined by:

$$E[k^n q^m] = \frac{1}{2M} \sum_{i,j} \hat{A}_{i,j} k_i^n q_j^m. \tag{2.16}$$

It is important to note that $E[k] \neq \langle k \rangle$. In fact one can show that $E[k] = \langle k^2 \rangle / \langle k \rangle$. A better understanding can be reached by expressing the assortativity as a function of $E[kq]$:

$$r = \frac{E[kq] - E[k]E[q]}{E[k^2] - E[k]^2}. \tag{2.17}$$

It is then clear that a positive assortativity will be associated to network for which $E[kq] > E[k]E[q]$. For such networks nodes of similar degree tend to associate (share edges) with one another. A negative assortativity will appear when $E[kq] < E[k]E[q]$. Such networks are said to be disassortative; nodes of high degree tend to link with nodes of low degree. Finaly, if one considers, uncorrelated networks, for which one can

approximate $\hat{A}_{i,j} \propto k_i k_j$, it is easy to show that $E[k^n] = E[q^n]$ and $E[kq] = E[k]E[q]$ leading to a vanishing assortativity where $r = 0$.

2.2.7 Paths

Let us start by defining three notions which, though being similar, have different meanings: a walk, a trail and a path. A walk in a network is a sequence of edges for which edges are not necessarily distinct. A trail however, is a walk where all edges are distinct, while a path is a trail in which all nodes are distinct. The number of walks of length q between two nodes i and j can easily be obtained via the element $(\hat{A}^q)_{i,j}$. The shortest walk between two nodes is also the shortest path. Thus the shortest path between nodes i and j can be found by observing successive powers of $\hat{A}$. As we increase the power from q to $q + 1$, the shortest path is given by $\ell_{i,j} = q + 1$ when $\hat{A}_{i,j}$ jumps from 0 to 1.

We should point out that many complex networks tend to be fragmented into disconnected sub components. The largest of these sub components is known as the giant connected component (GCC). Let us use the symbol $\mathcal{C}_1$ to refer to the set of nodes in the GCC. All other components will be denoted by $\mathcal{C}_j$ ($j > 1$). We should point out that between nodes of distinct sub-components the shortest path can not be defined. For this reason the mean shortest path length $\langle \ell \rangle_m$ is defined on the component $\mathcal{C}_m$ as the average number of steps along the shortest path $\ell_{i,j}$ for all possible pairs of nodes i, j:

$$\langle \ell \rangle_m = \frac{2}{|\mathcal{C}_m|(|\mathcal{C}_m| - 1)} \sum_{i,j \in \mathcal{C}_m} \ell_{i,j}, \tag{2.18}$$

where $|\mathcal{C}_m|$ is the number of nodes in component $\mathcal{C}_m$, with the summation running over nodes belonging to the $\mathcal{C}_m$ component only. Note that if $i = j$ one has $\ell_{i,j} = 0$ which simply does not contribute to the summation. Measuring how closely connected nodes are within the GCC is often a point of interest as it gives an indication of a network's connectivity. In fact, the measure of shortest path are often restricted to the giant component, only. Hence, $\langle \ell \rangle$ simply refers to $\langle \ell \rangle_1$: the shorted path in the GCC.

For a given network, the shortest path is often compared to average path length of a $G(N,p)$-random network $\langle \ell_r \rangle$ of the same size [33]:

$$\langle \ell_r \rangle = \frac{\ln N - \alpha}{\ln(Np)} + \frac{1}{2} \, , \tag{2.19}$$

where $\alpha \approx 0.5772$ is the Euler-Mascherroni constant. The average path length can be extended to weighted networks. Here the shortest path is not always the shortest sequence of distinct edges but rather the path carrying the least weight. Finally, the diameter D of a network is defined as the largest shortest path in the GCC.

2.2.8 Degree, Betweenness, and Closeness Centrality

In a network the importance of a node can be defined in different ways depending on the property we investigate. In fact, a number of centrality measures have been developed. In the following section we give a brief definition of the degree, betweenness and closeness centrality which has been well documented in Ref.[34].

Degree Centrality

One measure of the importance of a node is to consider the number of other nodes it is directly associated with. The degree centrality of a node is defined by,

$$C_d(j) = \frac{k_j}{N-1}. \tag{2.20}$$

This can be seen as the amount of influence a node has within its local vicinity and provides a certain measure of its importance within the network.

Betweenness centrality

The betweenness of a node determines a nodes importance by considering how often this node appears on the shortest paths between other nodes in the network. It is calculated by

$$C_b(i) = \sum_{j,k} \frac{\rho(j,i,k)}{\rho(j,k)}, i \neq j \neq k \, , \tag{2.21}$$

where $\rho(j,k)$ is the number of shortest paths between nodes j and k and $\rho(j,i,k)$ is the number of shortest path that go through i.

Closeness centrality

The closeness centrality reveals how central a node is in the network. It is defined as:

$$C_c(j) = \frac{\mathcal{C}_m - 1}{N - 1} \frac{\mathcal{C}_m - 1}{\sum_{t=1}^{\mathcal{C}_m - 1} \ell_{j,t}}, \tag{2.22}$$

given $j, t \in \mathcal{C}_m$. If the network is disconnected, as is the case for many networks, the first term will act to normalise the centralities for each fully connected subcomponent.

2.2.9 Small world model

A network is said to be "small world" if the typical distance between two random nodes scales approximately according to $\log(N)$ and the network is highly correlated (i.e. $C \gg C_r$) [22]. The notion of a small world was first introduced by a writer, called Frigyes Karinthy, in the 1920s when he published the short story "Chain-Links" [35]. Here, he theorised of a world that was becoming more inter-connected due to advances in communication. It was only in the 1960s that this hypothesis was tested by Stanley Milgram in his famous experiment [20]. In this experiment, people selected at random in the US cities of Omaha and Wichita where asked to send a letter directly to a person (once again selected at random) in Boston if and only if they knew them on a first name basis. If they did not know the designated person they were asked to send the letter to a person they thought more likely to know the person of interest. The aim of this experiment was to determine the average path length in social networks. Milgram found this to be six which was extremely small considering the population of the US and geographical divide between cites. This resulted in the coining of the term "Six degrees of separation". The random networks hypothesised by Erdös-Rényi and Gilbert have the feature of a relatively small average path length, seen in Milgram's experiment, however these models failed to explain the high correlation observed on a local scale in many complex networks.

FIGURE 2.3: Watts and Strogatz Small world: Clustering coefficient C and average path length $\langle \ell \rangle$ as a function of p for $N = 100$ and $M = 1500$ [22].

Three decades later, in 1998, Watts and Strogatz (WS) developed a model that could account for both of these features [22], where simultaneously $\langle \ell \rangle \approx \log(N)$ and $C \gg C_r$. The WS model is defined by the following algorithm:

1. Begin by setting the initial condition of a ringed lattice where all neighbours are linked.

2. Pick an edge (u, v) and with probability p randomly rewire the edge to (u, x).

3. Apply step two, iterating through pairs of edges once and once only, until all $n(n-1)/2$ pairs of edges are exhausted.

Figure 2.3 shows the clustering coefficient C and mean path length $\langle \ell \rangle$ as a function of the rewiring probability p. If p equals zero the initial configuration remains unchanged and as the parameter p increases C remains large but $\langle \ell \rangle$ becomes relatively small, thus displaying the characteristics for a small world network. If however p continues to increase, $p \to 1$, the network becomes completely random showing small C and $\langle \ell \rangle$, the features of a random network. One should note that the WS model does not produce the scale free behaviour observed in the degree distribution of many complex networks.

2.3 Conclusion

In this chapter we discussed some of the contributing events to the development of complex network science and the main theory applied in this thesis. Complex networks are useful in quantifying processes in complex systems involving interacting agents. It has become a large field of science which has brought together many researchers from areas of science such as mathematics, physics, computer science, biology and sociology [9] amongst others. Whilst being a well established field it is still developing and evolving. In this thesis we have chosen to further contribute to this field of science by considering three different object all studied under the umbrella of complex networks. In the chapters that follow the methods discussed here will be applied to polymers, transport public networks and large scale structures of the universe.

Chapter 3

Shape properties of ideal, ramified tree-branched polymers

It is well known that the branching structure affects the behaviour of polymer systems [36], [37]. For example, branching can affect the shape of a polymer which will, in turn, affect properties such as the viscosity and melting points of polymer conglomerations. This has a significant effect on the strength and durability of polymers. Knowledge of the shape properties of different polymeric configurations can aid in the development of more purpose built materials. One specific form of branched polymer is dendrimers (see Figure 3.1). Shaped like concentric branching trees, these structures, have the ability to fold up to form a molecular cage [38]. Dendrimers have attracted attention within medical sciences as they can be used to "carry" drugs to areas of the body that are typically hard to reach [39], [40].

In this chapter we investigate the shape parameters of two types of polymers: comb polymers with two, three, four and five junctions and first and second generation dendrimers in 3D (see Figure 3.1). The shape parameters investigated include: g-ratio, asphericity, prolateness and form factor, analysed by three different techniques. In the first method (M1), we apply a Monte Carlo Growth Algorithm (MCGA), where polymers are grown randomly on square and triangular lattices. Method two (M2), is derived from a method developed by Wei [41], [42], where only the topology of polymers is required to determine the shape properties. This is done by representing the polymer as a graph $G(V, E)$ where V and E are the monomers and covalent bonds between monomers respectively. The third method (M3), the Benhamou method [43], is used to
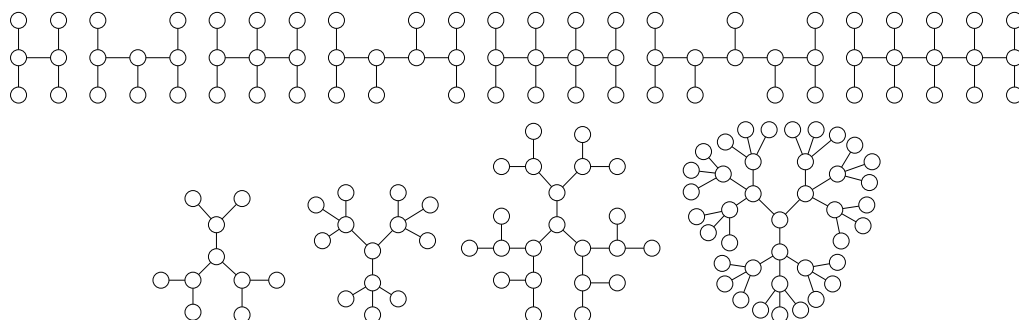
FIGURE 3.1: Ramified polymeric networks. The first row shows the two, three, four, and five junction comb polymers. The second row displays the dendrimers structures.

calculate the exact form factor of polymers employing Debeye scattering techniques for Gaussian chains.

In Ref.[41], [42], Wei analytically determines various universal shape parameters for certain types of comb-like structures which are precise in 2D and very accurate in 3D. Here, however we redesign this method to extrapolate results by generating shape parameters for finite size systems and then extrapolating to an infinite size system. The advantage of this new approach is that the shape properties of any tree-like structure, whereby the Kirchoff matrix can be generated, can be semi-analytically evaluated in a quick and efficient manner. These results are then compared with the more traditional methods of Monte Carlo simulations.

The chapter is laid out with the following structure. In section 3.1 the gyration tensor and the resulting shape properties are defined and discussed. In section 3.2, methods applied in the analysis are explained. In section 3.3, we present our results for two, three and four junction combs in 2D (subsection 3.3.1), 5-junction combs in 2D and 3D (subsection 3.3.2) and first and second generation polymers (subsection 3.3.3). Finally, some concluding remarks are offered in section 3.4.

As we have shown in this study, all three methods show excellent agreement with each other and available theory. The main results of this work have been published in Refs.[11], [13].

3.1 Gyration tensor and shape of complex structures

Here, ramified comb and dendrimer polymers of various topologies are studied. A ramified polymer can be described as a tree network, connected by long flexible chains where each chain has the same degree of polymerisation. If m is the number of monomers in a branch (or chain) and b is the number of branches, then there are a total of $N = bm + 1$ monomer units in the network.

The size of an ideal linear polymer chain can be measured by studying its end to end distance, R_e. The mean end to end distance $\langle R_e \rangle$ for an ideal chain constructed of N monomers with a diameter σ can be easily calculated by considering a random walk in one dimension. In this scenario, $\langle R_e \rangle = 0$ because the mean number of steps in the left direction will equal the means number of steps in right direction. The variance can then be determined in the usual way, $\text{Var}(R_e) = \langle R_e^2 \rangle - \langle R_e \rangle^2$. This leads to $\text{Var}(R_e) = \langle R_e^2 \rangle$ as $\langle R_e \rangle = 0$ which then gives a mean end to end distance,

$$\langle R_e^2 \rangle = N\sigma^2 \, , \tag{3.1}$$

for an ideal linear chain in one dimension. This can be extended into higher dimensions, where $\langle R_{e_x}^2 \rangle = \langle R_{e_y}^2 \rangle = \langle R_{e_z}^2 \rangle = \frac{N\sigma^2}{3}$. Finally leading to an end to end distance distribution,

$$P(R_e) = \left(\frac{3}{2\pi N\sigma^2} \right)^{3/2} e^{-\frac{3\langle R_e^2 \rangle}{2N\sigma^2}} \, . \tag{3.2}$$

So the size of a linear chain can be measured via an end to end distance. However, for more complex polymeric structures this measure is not well defined. For example branched polymers will have many ends and ringed polymers no ends at all. Hence when measuring polymers a different mechanism for measuring the size of a polymer is required.

A measure that is well defined to determine the size of all kinds of polymers is the mean-square radius of gyration, $\langle S^2 \rangle$. For a polymer, $\langle S^2 \rangle$ can be determined using the gyration tensor Q. If $X_j^{(\alpha)}$ denotes the α component of the position vector of the j-th

monomer, then the center of mass coordinates $(X_{CM}^{(\alpha)})$ for a given configuration is given by

$$X_{CM}^{(\alpha)} = \frac{1}{N} \sum_{j=1}^{N} X_j^{(\alpha)} \text{ , for } \alpha = 1, 2 \text{ or } 3 \tag{3.3}$$

and the matrix components of the gyration tensor (Q) may be written in the form

$$Q_{\alpha\beta} = \frac{1}{N} \sum_{j=1}^{N} (X_j^{(\alpha)} - X_{CM}^{(\alpha)})(X_j^{(\beta)} - X_{CM}^{(\beta)}). \tag{3.4}$$

The square radius of gyration of this configuration is calculated in three dimensions as

$$S^2 = Q_{11} + Q_{22} + Q_{33} = e_1 + e_2 + e_3, \tag{3.5}$$

and in two dimensions as

$$S^2 = Q_{11} + Q_{22} = e_1 + e_2. \tag{3.6}$$

It is well known [44] that for large polymers, with or without branches, $\langle S^2 \rangle$ follows the scaling law :

$$\langle S^2 \rangle = C N^{2\nu}. \tag{3.7}$$

The coefficient, C, is a model dependent amplitude but the exponent, 2ν, is universal and equal to 1 for all ideal polymers, irrespective of topology. This leaves,

$$\langle S^2 \rangle = C N , \tag{3.8}$$

remembering the end to end distance defined in Eq.(3.1), $\langle S^2 \rangle$ can then be defined as,

$$\langle S^2 \rangle = C \langle R_e^2 \rangle , \tag{3.9}$$

Thus highlighting the relation that exists between the end to end distance and the radius of gyration of polymeric structures. It can easily be shown that $C = \sigma^2/6$ for both linear and acyclically branched polymers.

The overall shape of a polymer can be determined using the gyration tensor Q. If the

eigenvalues of Q are ordered by magnitude: $e_i \leq e_j$ for $i < j$. These are the principal moments of gyration along the principal orthogonal axes [45]. Rudnick and Gaspari [46], [47] have defined the average asphericity, $\langle A \rangle$, of a polymer in d dimensions as

$$\langle A \rangle \;=\; \left\langle \frac{\sum_{i>j}^{d}(e_i - e_j)^2}{(d-1)(\sum_{i=1}^{d} e_i)^2} \right\rangle . \tag{3.10}$$

The g-ratio is a dimensionless parameter that measures the compactness of a polymer. It is defined by the radius of gyration of a branched polymer $\langle S^2 \rangle_b$ and a linear polymer chain $\langle S^2 \rangle_l$, containing the same number of monomers,

$$g = \frac{\langle S^2 \rangle_b}{\langle S^2 \rangle_l} . \tag{3.11}$$

Casassa and Berry [48] obtained a general equation for the g-ratio of uniform, ideal comb polymers, where n is the number of junctions with three branches adjacent, which are regularly spaced along the backbone:

$$g = 1 - h - \frac{h^2(1-h)}{n+1} + \frac{2h(1-h)^2}{n} + \frac{(3n-2)(1-h)^3}{n^2} . \tag{3.12}$$

Here, h is the ratio of the number of units in the comb backbone to the total number of units in the polymer. Table 3.1 displays the h, n and g-ratio values obtained for uniform junction combs.

Structure	h	n	g-ratio
5-branch	3/5	2	0.7120
7-branch	4/7	3	0.6676
9-branch	5/9	4	0.6379
11-branch	6/11	5	0.6168

TABLE 3.1: g-ratios for uniform ideal comb polymers with a junction functionality of three for 2, 3, 4 and 5 junctions.

The g-ratios for other non uniform junction combs and dendrimers studied here are derived from the form factor $S(k)$ (later defined in this chapter). In Ref.[49], von Ferber et al. found g-ratios values of $37/64$ (0.5781) and $683/1331$ (0.5131) for eight and eleven branch combs, respectively.

The shape of a two (three) dimensional linear polymer can vary from a fully extended rod in which e_2 completely vanishes leaving $\langle A \rangle = 1$, to a circle (sphere) for which all the eigenvalues e are equal. In the latter case $\langle A \rangle$ is zero. In between the extremes of a rod and a circle (sphere), a polymer configuration can be imagined as approximately enclosed inside an ellipse (ellipsoid).

In addition to the asphericity, other properties can be calculated to identify the overall shape of a polymer. In two (respectively three) dimensions, the shape factor parameter $\langle \delta_1 \rangle$ (Prolateness $\langle P \rangle$) is used:

$$\langle \delta_1 \rangle = \frac{\langle e_1 \rangle}{\langle S^2 \rangle} , \tag{3.13}$$

and

$$\langle P \rangle = \left\langle \frac{27(e_1 - \bar{e})(e_2 - \bar{e})(e_3 - \bar{e})}{(\sum_{i=1}^{3} e_i)^3} \right\rangle , \tag{3.14}$$

where, $\bar{e}$ is

$$\bar{e} = \frac{e_1 + e_2 + e_3}{3} . \tag{3.15}$$

Note that $\langle A \rangle$ and $\langle P \rangle$ involve an average of a ratio whereas $\langle \delta_1 \rangle$ involves a ratio of averages.

Another important structural property is the scattering function, $S(k)$ which provides information on the spatial distribution of monomers. When a laser beam is directed at a polymer solution most of the photons from the incident beam go straight through. However, some of the molecules in the path of the beam scatter a small fraction of the photons. The scattering vector **k** is given by the change in the wave vector upon scattering,

$$\mathbf{k} = \mathbf{k_s} - \mathbf{k_i} \tag{3.16}$$

where $\mathbf{k_i}$ if the wave vector of the incident beam which has a magnitude of $\frac{2\pi}{\lambda_l / n_{sol}}$ (λ_l / n_{sol} is the wave length of light in a solution of refractive index n_{sol}) and $\mathbf{k_s}$ is the vector of the scattered beam. For static light scattering (i.e when monomers are assumed

to be static in a solution) the magnitude of $\mathbf{k_i}$ and $\mathbf{k_s}$ are equal. Thus the magnitude or change in scattering wave vector, $|\mathbf{k}| = k$ is given by,

$$k = \frac{4\pi n_{sol}}{\lambda_l}\sin(\theta/2) \,, \tag{3.17}$$

where θ is the scattering angle. We can see that as θ increases for $0 \leq \theta \leq 180$, k also increases in magnitude.

As the incident beam $\mathbf{k_i}$ propagates forward its orthogonal electric field $\mathbf{E_i}$ will oscillate with an angular frequency and thus $\mathbf{E_i}$ will depend on the position $\mathbf{r}$ and time t according to:

$$\mathbf{E_i} = \mathbf{E_{i0}}e^{\mathbf{i}(\mathbf{k_i}\cdot\mathbf{r}-\omega t)} \,, \tag{3.18}$$

where $\mathbf{E_{i0}}$ is the complex amplitude of the field and ω is the angular frequency.

As the laser beam travels through a polymer solution photons will collide with monomers polarising them at a given position $\mathbf{r_i}$ and time t_i and thus the scattering radiation $\mathbf{E_{si}}$ for each monomer will be different and described by,

$$\mathbf{E_{si}} = \mathbf{E_{sm}}e^{\mathbf{i}(\mathbf{k_s}\cdot\mathbf{r}-\mathbf{r_i})-\omega(t-t_i)} \,, \tag{3.19}$$

where $\mathbf{E_{sm}}$ is the complex amplitude of the beam scattered by a monomer propagated in direction of the scattering wave vector $\mathbf{k_s}$.

By summing $\mathbf{E_{si}}$ for all monomers and with some simple derivation

Flory and Volkenstein, in Ref.[50], formally define the Fourier transform for the density-density autocorrelation function:

$$S(k) = \frac{1}{N^2} \sum_{m,n}^{N} \langle e^{\,i\mathbf{k}\cdot(\mathbf{r}_n-\mathbf{r}_m))} \rangle \,. \tag{3.20}$$

Here, N is the number of monomers in the polymer, $\mathbf{k}$ is the scattering vector, and $\mathbf{r}_m$ and $\mathbf{r}_n$ are the positions of the m-th and n-th monomers. The values given by this

function depend on the dimension that the polymer occupies. This is a laser beam has a large coherence meaning that photons perpendicular to $\mathbf{k_i}$ are all assumed to be in phase. So in two dimensions the phase angle for polarisation will always be the same whereas in three dimension this changes. After averaging over the angles in two dimensions, the scattering function is defined as the zeroth order Bessel function, $J_0(x)$. Whereas in three dimensions, the scattering function is defined as $\sin(x)/x$ where $x^2 = k^2\langle S^2\rangle$. To compute $S(k)$ using the generated MC configurations, we first average Eq.(3.20) over the angle between $\mathbf{k}$ and $(\mathbf{r}_n - \mathbf{r}_m)$ to find that in two dimensions

$$S(k) = \frac{1}{N^2}\sum_{m,n}^{N}\langle J_0[\mathbf{k}(\mathbf{r}_n - \mathbf{r}_m)]\rangle \,, \tag{3.21}$$

and in three dimensions

$$S(k) = \frac{1}{N^2}\sum_{m,n}^{N}\left\langle \frac{\sin[\mathbf{k}(\mathbf{r}_n - \mathbf{r}_m)]}{[\mathbf{k}(\mathbf{r}_n - \mathbf{r}_m)]}\right\rangle \tag{3.22}$$

In an ideal polymer, the configurations have a Gaussian distribution. Casassa and Berry [48] used this fact to obtain the form factor for uniform ideal combs with single branches attached to the backbone:

$$S(k) = \frac{2(A + B + C)}{x^2} \,, \tag{3.23}$$

where

$$A = x - 1 + e^{-xh}, \tag{3.24}$$

and

$$B = \left(1 - e^{-x(1-h)/n}\right)\left(n + 2\frac{1 - e^{-xhn/(n+1)}}{1 - e^{xh/(n+1)}}\right) \tag{3.25}$$

and

$$C = \left(1 - e^{-x(1-h)/n}\right)^2 \frac{(n-1)(e^{xh/(n+1)} - 1) - (1 - e^{-xh(n-1)/(n+1)})}{(1 - e^{xh/(n+1)})^2} \tag{3.26}$$

In these equations $x = k\langle S^2\rangle_\ell$ where $\langle S^2\rangle_\ell$ is the radius of gyration of an ideal linear polymer with N units. $\langle S^2\rangle_\ell$ is related to $\langle S^2\rangle_b$ by the g-ratio. The form factor of a uniform junction comb polymer is easily obtained from these equations.

In the next section we discuss the methods that have been applied to determine the shape parameters introduced here.

3.2 Numerical and analytical approaches to quantify shape properties of tree like polymer networks

3.2.1 Monte Carlo growth algorithm

For the MC simulations, linear chains are grown on simple square and triangular lattices in 2D and cubic lattices in 3D, to generate ideal polymers. For two-dimensional structures, where polymers are grown on a regular are triangular grid, there are four (North, South, East, West) or six (East, West, North East, North West, South East, South West) directions which could be chosen. In three dimensions, there also are six (North, East, South, West, up, Down) possible directions. Once a direction is randomly selected, a new monomer is placed in that direction. For the ideal regime considered here, overlapping beads are allowed. This process is repeated m times to generate a branch.

All polymers studied are generated by implementing the above linear polymer growth algorithm, with slight modifications. For all combs, a three branched star is grown by placing the first monomer at the origin $(0, 0)$. The first branch is grown from the origin to include m monomer units. Then, two other branches of the same size are grown again starting at the origin. The origin is then moved to the end of the third branch, where new branches are subsequently grown. From here the process for generating polymers with a complete and an incomplete set of internal branches differs:

- Complete set: The process of generating a three branched star is repeated until $b - 2$ branches exist in the network. At this point only two additional branches are subsequently added.

- Incomplete set: After a three branched star is generated, two branches are grown from the new origin where the origin is then moved to the end of the second branch. This iterative process continues until b-branches exist in the comb polymer.

In the case of dendrimers, initially a linear branch is generated. Then, all connecting branches are grown from the junction at the end of this branch. In the case of the second-generation twenty one branch and thirty nine branch structures, additional branches are then grown from the end junctions of the inner branches. Once a complete multi branch component is grown, the algorithm cycles back to the initial seed monomer to repeats the process two more times.

After each polymer is completely constructed, the shape properties discussed in section 3.1 are calculated for each configuration. In the simulated systems for combs, the number of monomers N ranged from 1001 to 7001 monomers. For dendrimers, N ranged from 55 to 1342 monomers. To determine the mean and the standard deviation for each simulated system, 1.6×10^5 and 10^5 samples were generated for comb and dendrimers polymers respectively.

For the computation of $S(k)$, the number of monomers N has a smaller range (from 127 to 701), due the computational complexity. Moreover, each form factor calculation involves two simulation runs: one to determine $\langle S^2 \rangle$, and a second to calculate $S(k)$. As we demonstrate, later in this chapter, these smaller systems are still large enough to probe the asymptotic regime. The set of configurations for $S(k)$ was further averaged over 10^4 samples.

3.2.2 Wei method

In 1995, Wei designed a method that determines the universal shape parameters of Gaussian macromolecules. In general this method can be applied to any polymer network where the Kirchhoff matrix (defined in section 2.2.4) and its corresponding eigenvalues can be determined. In Refs.[41], [42], Wei analytically derives the shape parameters for star and simple combs structures. These results are exact in 2D and give a extremely good approximation in 3D. A particular component of the present execution of this approach is that we determine the shape parameters by extrapolating in the limit $N \to \infty$. This is done by following a simple algorithm. First, we construct a ramified comb or dendrimer network that has $m = 64$ monomers per chain. Here the initial number of monomers N depends on the polymer architecture. We then determine the

shape parameters for a polymer of size N. In the next step, we increase the size of each chain in the network by a factor of $\sqrt{2}$ and then calculate the new shape parameters for the same polymer architecture that is now $\sqrt{2}N$ larger. For combs and 9-branch dendrimers this process is repeated until a chain size of 4096 monomers is reached. However, due to the computational complexity of computing eigenvalues for large Kirchhoff matrices larger dendrimer networks could only be determine for smaller N. The smallest number of polymers per branch considered was 724 monomers for the 39-branch dendrimer. From here we can extrapolate in the thermodynamic limit. This is shown in Figure 3.2, where we show the extrapolation process for the shape parameter asphericity for nine and eleven branch combs in 2D.

To calculate shape parameters after constructing the polymer network, we first determine the non-zero eigenvalues $\lambda_1, \ldots, \lambda_{N-1}$ of the $N \times N$ Kirchhoff matrix $\hat{K}$ for a Gaussian polymer with N monomers. The diagonal matrix is defined as,

$$\Lambda_N(y) = \begin{pmatrix} \lambda_1 + y & & 0 \\ & \ddots & \\ 0 & & \lambda_{N-1} + y \end{pmatrix} \tag{3.27}$$

With this matrix at hand a reduced variant of the characteristic polynomial of $\hat{K}$ can be expressed as,

$$P_N(x) = \text{Det}[\Lambda_N(-x)] = \prod_{j=1}^{N-1} (\lambda_j - x) , \tag{3.28}$$

where Det[.] represents the determinant of a matrix. The $N - 1$ zeros of the polynomial P_N are the non-zero eigenvalues of $\hat{K}$. We are now in the position to define the functions $D_N(x)$, $S_{1,N}(x)$ and $S_{2,N}(x)$ essential for the calculation of the shape parameters

$\langle A \rangle$, $\langle \delta_1 \rangle$ and $\langle P \rangle$. Using the notation $y = x/N$ we define

$$
\begin{aligned}
D_N(x) &= \text{Det}[\Lambda_N^{-1}(0)]\text{Det}[\Lambda_N(y^2)] \\
&= P_N^{-1}(0)P_N(-\frac{x^2}{N^2}) \\
&= \prod_{j=1}^{N-1} \frac{\lambda_j + y^2}{\lambda_j} ,
\end{aligned}
\tag{3.29}
$$

and

$$
S_{1,N}(x) = \frac{1}{N^2} \text{Tr}[\Lambda_N^{-1}(y^2)] = \frac{1}{N^2} \sum_{j=1}^{N-1} (\lambda_j + y^2)^{-1} ,
\tag{3.30}
$$

where Tr is the trace of a matrix. Following the work of Wei, the functions for $S_{k,N}$ for $k = 2, 3, ..$ correspond to traces of higher powers of Λ_N^{-1},

$$
S_{2,N}(x) = \frac{1}{N^4} \sum_{j=1}^{N-1} (\lambda_j + y^2)^{-2} ,
\tag{3.31}
$$

and

$$
S_{3,N}(x) = \frac{1}{N^6} \sum_{j=1}^{N-1} (\lambda_j + y^2)^{-3} .
\tag{3.32}
$$

The two-dimensional shape parameter $\delta_{1,N} = \frac{1}{2} - \chi_{1,N}$ is calculated with help of

$$
\chi_{1,N} = S_{1,N}(0)^{-1} \int_0^\infty |xD_N(x+ix)|^{-1}\text{Im}(S_{1,N}(x+ix))dx .
\tag{3.33}
$$

Using Eq.(3.33) and the substitution $y = x/N$ we find after some algebra

$$
\chi_{1,N} = -2 \sum_{i=0}^{N-1} \lambda_j \int_0^\infty \prod_{j=1}^{N-1} \frac{\lambda_j}{\sqrt{\lambda_j^2 + 4y^4}} \sum_{k=0}^{N-1} \frac{1}{\lambda_k^2 + 4y^4} dy.
\tag{3.34}
$$

The asphericity parameters for $d = 2, 3$ can then also be derived using

$$\langle A_d \rangle = \frac{d(d+2)}{2} \int_0^\infty x^3 D_N^{-d/2}(x) S_{2,N}(x) \mathrm{d}x \tag{3.35}$$

$$= \frac{d(d+2)}{2} \int_0^\infty \mathrm{d}y \sum_{j=1}^{N-1} \frac{y^3}{(\lambda_j + y^2)^2} \left[\prod_{k=1}^{N-1} \frac{\lambda_k}{\lambda_k + y^2} \right]^{d/2}. \tag{3.36}$$

The prolateness parameter in $d = 3$ dimensions gives

$$\langle P_3 \rangle = \frac{105}{8} \int_0^\infty x^5 D_N^{-3/2}(x) S_{3,N}(x) \mathrm{d}x \tag{3.37}$$

$$= \frac{105}{8} \int_0^\infty \mathrm{d}y \sum_{j=1}^{N-1} \frac{y^5}{(\lambda_j + y^2)^3} \left[\prod_{k=1}^{N-1} \frac{\lambda_k}{\lambda_k + y^2} \right]^{3/2}. \tag{3.38}$$

Finally, independent of the dimension the g-ratio is given by

$$g = S_{1,N}^{\mathrm{branch}}(0) / S_{1,N}^{\mathrm{chain}}(0)$$
$$= \sum_{j=1}^{N-1} \lambda_j^{-1} / \sum_{k=1}^{N-1} \tilde{\lambda}_k^{-1} \tag{3.39}$$

where λ_j and $\tilde{\lambda}_k$ is the non-zero eigenvalues of the Kirchhoff matrix of a branched and linear chain with N monomers.
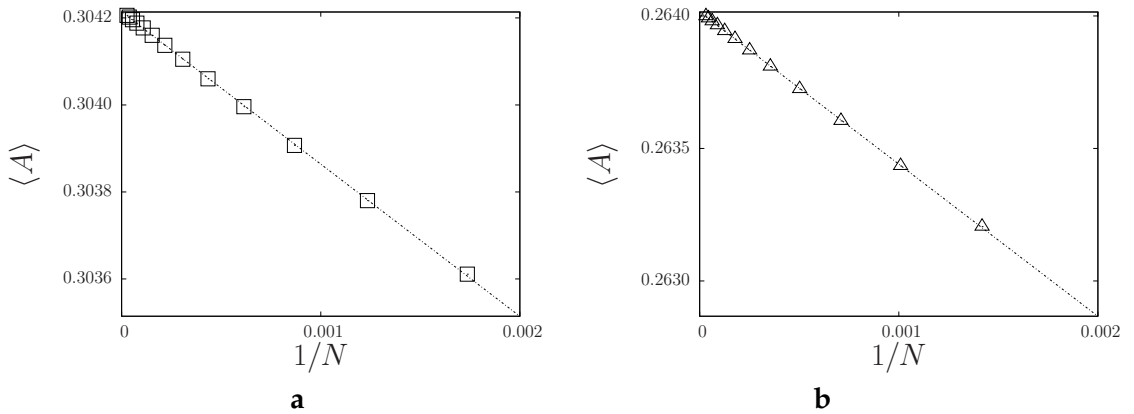
FIGURE 3.2: Extrapolation for the two dimensional asphericities $\langle A \rangle$ of **a**: nine branch and **b**: eleven branch ideal ramified comb polymers as a function of $1/N$. Here markers are the actual values of $\langle A \rangle$ calculated applying the Wei method and the dotted lines are their respective fits.

3.2.3   Benhamou method

The form factor of an ideal ramified polymeric network can be derived by applying a method developed by Benhamou [43]. Here, Benhanou applies Debye scattering techniques to tree networks, where each chain is formed by an ideal linear polymer. For these networks the form factor can be decomposed into intra-chain and inter-chain contributions, each involving two distinct chains in a network.

The contribution of the intra-chain component $S_1(k)$ is directly related to the scattering function of an ideal linear polymer $S_l(k)$:

$$S_1(k) = \frac{1}{b} S_l(k) \, , \tag{3.40}$$

for which the Debye scattering function of a linear polymer is given by

$$S_l(k) = \frac{2}{x^2} (x - 1 + e^{-x}) \, , \tag{3.41}$$

where $x = k \langle S^2 \rangle_\ell$ and $\langle S^2 \rangle_\ell = \sigma^2 N / 6$.

The inter-chain contribution mainly depends on the distance between chains in the network. More specifically, by this we mean the distance between chains refers to the length of a unique path that exists and includes any combination of chains in the polymer network. As we are only dealing with tree-like structures in this study, a unique path will always exist between all combinations of chains. The inter-chain contribution of all these combinations of chains is found to lead to

$$S_2(k) = \frac{2(e^{2x} - 2e^x + 1)}{bx^2} \sum_{\zeta \subset G; |\zeta| \geq 2} \left(e^{-|\zeta|x} \, , \tag{3.42}\right.$$

where ζ is the length of a connected path between two branches in the network G, with $|\,\zeta\,| \geq 2$.

3.2. Numerical and analytical approaches to quantify shape properties of tree like polymer networks

These two intra and inter contributions then combine to give the exact form factor equation for ideal ramified polymers with a tree branching topology

$$S(k) = \frac{2}{bx^2}\left[x - 1 + e^{-x} + \frac{(e^{2x} - 2e^x + 1)}{b}\sum_{\zeta \subset G; |\zeta| \geq 2} e^{-|\zeta|x}\right]. \tag{3.43}$$

It is shown that in the case of comb polymers Eq.(3.43) can be expressed as a function of the internal and external junctions. If b and $b-1$ is the number branches adjacent the inner and outer junctions of a comb polymers respectively. The form factor is derived as

$$\begin{aligned}
S_{b,b-1}^{(n)}(k) &= \Bigg(\bigg(\left(b^2 - 3b + 2\right)ne^{(n+1)x} - 2(b-1)(b+n-3)e^{(n+2)x} \\
&\quad + (b-1)(bn-4)e^{(n+3)x} + 2(b-2)^2 e^{2x} + 4(b-2)e^{3x} + 2e^{4x}\bigg)\bigg(\\
&\quad \times \frac{e^{-(n+3)x}}{x^2((b-1)n-1)^2} + \frac{2\left(x + e^{-x} - 1\right)}{x^2((b-1)n-1)},
\end{aligned} \tag{3.44}$$

where n is the number of junctions of a comb polymer.

For combs, where all junctions have the same number of branches b, the form factor may be written as

$$\begin{aligned}
S_{b,b}^{(n)}(k) &= \bigg(-(b-2)ne^{(n+2)x} + bne^{(n+4)x} - 2(b+n-1)e^{(n+3)x} + 2(b-1)e^{3x}\bigg)\bigg(\\
&\quad \times \frac{(b-1)e^{-(n+4)x}}{x^2((b-1)n+1)^2} + \frac{2\left(x + e^{-x} - 1\right)}{x^2((b-1)n+1)}.
\end{aligned} \tag{3.45}$$

In the special case of the fourteen branch combs, when we normalize $x \to x/14$, we obtain

$$S(k) = \left(\frac{2}{x^2}\right)\bigg(10 + x - 4e^{-x/14} - 15e^{-x/7} + 4e^{-3x/7} + 4e^{-5x/14} + e^{-2x/7}\bigg), \tag{3.46}$$

which can be used to determine the g-ratio. This is done by dividing this expression for the fourteen branch comb with $S(k)$ for a linear chain. To do this we write the first order Taylor expansion for a fourteen branch comb

$$S(k) = 1 - \frac{163}{1029}x + \dots, \tag{3.47}$$

and linear chain

$$S(k) = 1 - \frac{1}{3}x + ... \,, \tag{3.48}$$

leading to a g-ratio of 163/343 (0.47522). Applying the same method enables exact g-ratios to be obtained for other structures (see results section for theoretical g-ratios).

3.3 Results

In both MC and Wei methods the results are computed for a finite number of monomers N. Whereas the theoretical results are valid in the limit $N \to \infty$. For the MC method infinite N shape parameter values are obtained by fitting a scaling law

$$S = S_\infty \left(1 - \frac{K}{N^\triangle} \right) \Big(\,, \tag{3.49}$$

where S_∞ is the value of a shape parameter for an infinite size polymer and $\triangle$ is the finite-size scaling exponent. The value for S_∞ is then determined by fitting a weighted least-squares [51] line in $1/N^\triangle$ for MC data, as shown in Zweier and Bishop [52].

The Wei method implements a nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm [53] to extrapolate shape parameters in the thermodynamic limit. The error bars of the Wei method indicate the root mean-square (RMS) of the residuals corresponding to the errors of the fitted data. For the results reported in this section, the number in parenthesis denotes one standard deviation in the last displayed digits.

3.3.1 Two, three and four junction combs polymers in 2D

The MC simulation results are contained in Tables 3.2, 3.3, 3.4, 3.5 and 3.6. In all cases, $\langle S^2 \rangle$ vs $(N-1)^{2\nu}$ gave $2\nu = 1.00 \pm 0.01$, in excellent agreement with the well-known result of $2\nu = 1$, for random walks.

The growth MC simulation results for asphericity $\langle A \rangle$ contained in the tables 3.2, 3.3, 3.4, 3.5 and 3.6 display only a weak dependence on N. As expected, the results for extrapolated values of $\langle A \rangle$ (Table 3.7) indicate that polymers become more circular in

	Property	$N = 1001$	$N = 1501$	$N = 2001$	$N = 2501$
	$\langle e_1 \rangle$	92.83(15)	139.14(22)	185.58(29)	232.37(36)
	$\langle e_2 \rangle$	25.48(3)	38.26(5)	51.03(7)	63.84(8)
Sq	$\langle S^2 \rangle$	118.31(16)	177.39(23)	236.61(31)	296.20(39)
	$\langle A \rangle$	0.310(1)	0.310(1)	0.310(1)	0.310(1)
	$\langle S^2 \rangle_l$	166.71(27)	249.39(40)	333.72(53)	416.95(66)
	$\langle e_1 \rangle$	92.85(14)	139.65(22)	186.13(29)	233.04(36)
	$\langle e_2 \rangle$	25.58(3)	38.40(5)	51.21(7)	64.03(8)
Tri	$\langle S^2 \rangle$	118.43(15)	178.05(23)	237.34(31)	297.07(39)
	$\langle A \rangle$	0.310(1)	0.310(1)	0.310(1)	0.310(1)
	$\langle S^2 \rangle_l$	166.81(26)	250.13(40)	334.29(53)	418.45(66)

TABLE 3.2: Properties of 5-branch combs, for different number of monomers N, on a square (Sq) and triangular (Tri) lattice.

their shape as the structure changes to higher branching and a complete set of interior branches.

The MC g-ratios in Table 3.7 have been calculated from the radius of gyration data and the errors in these quantities have been computed from the standard equation relating the error in a ratio to the errors in the numerator and denominator. These extrapolated g-ratios for ideal systems are compared with other findings in Table 3.7. Both Wei's method and the MC simulations are in excellent agreement with each other and the theoretical predictions. The g-ratios of the eight and eleven branch (four junction) combs, which have a complete set of interior branches, have a relatively lower value than that found for the five, seven and nine branch combs.

The shape factor, $\langle \delta_1 \rangle$ is contained in Table 3.7. The values for both lattices give similar results to each other and the predicted values of the Wei method. As expected, the results again indicate that comb polymers become more circular in their shape (less stretched out) as the structure changes to higher branching and a complete set of interior branches.

Two dimensional ideal five branch combs were also investigated by Perrelli and Bishop [54] and Gorry and Bishop [55]. Perrelli and Bishop [54] employed a MC growth algorithm on a square lattice but for a smaller range of N (100 to 800) and for a much smaller number of simulations (10^4). They found that $g = 0.704 \pm 0.005$, $\langle A \rangle = 0.311 \pm 0.002$, and $\langle \delta_1 \rangle = 0.785 \pm 0.005$. These values are well within two standard deviations of the mean or in the 95% confidence interval compared to the results reported here.

	Property	$N = 1401$	$N = 2101$	$N = 2801$	$N = 3501$
Sq	$\langle e_1 \rangle$	121.91(19)	182.69(29)	244.15(38)	305.17(48)
	$\langle e_2 \rangle$	33.53(4)	50.33(6)	67.19(8)	84.08(10)
	$\langle S^2 \rangle$	155.44(21)	233.02(31)	311.34(41)	389.26(51)
	$\langle A \rangle$	0.306(1)	0.305(1)	0.306(1)	0.306(1)
	$\langle S^2 \rangle_l$	233.37(37)	349.77(56)	466.76(74)	584.39(93)
Tri	$\langle e_1 \rangle$	122.05(19)	183.14(29)	244.04(38)	305.42(48)
	$\langle e_2 \rangle$	33.71(4)	50.52(6)	67.45(8)	84.18(10)
	$\langle S^2 \rangle$	155.75(21)	233.66(31)	311.49(41)	389.60(51)
	$\langle A \rangle$	0.304(1)	0.305(1)	0.305(1)	0.305(1)
	$\langle S^2 \rangle_l$	233.96(37)	351.00(56)	468.44(74)	584.08(93)

TABLE 3.3: Properties of 7-branch combs, for different number of monomers N, on a square (Sq) and triangular (Tri) lattice.

	Property	$N = 1601$	$N = 2401$	$N = 3201$	$N = 4001$
Sq	$\langle e_1 \rangle$	117.76(17)	176.49(26)	235.72(34)	294.43(43)
	$\langle e_2 \rangle$	36.13(4)	54.21(6)	72.26(8)	90.55(11)
	$\langle S^2 \rangle$	153.89(19)	230.71(28)	307.98(37)	384.99(46)
	$\langle A \rangle$	0.274(1)	0.274(1)	0.275(1)	0.274(1)
	$\langle S^2 \rangle_l$	266.39(42)	400.36(64)	534.05(85)	667.03(106)
Tri	$\langle e_1 \rangle$	117.74(17)	177.07(26)	235.70(34)	295.05(43)
	$\langle e_2 \rangle$	36.19(4)	54.41(6)	72.64(9)	90.61(11)
	$\langle S^2 \rangle$	153.93(18)	231.48(28)	308.34(37)	385.65(46)
	$\langle A \rangle$	0.274(1)	0.274(1)	0.273(1)	0.274(1)
	$\langle S^2 \rangle_l$	267.04(42)	401.71(64)	534.70(85)	667.85(106)

TABLE 3.4: Properties of 8-branch combs, for different number of monomers N, on a square (Sq) and triangular (Tri) lattice.

	Property	$N = 1801$	$N = 2701$	$N = 3601$	$N = 4501$
Sq	$\langle e_1 \rangle$	150.23(24)	224.80(36)	300.54(48)	375.90(60)
	$\langle e_2 \rangle$	40.97(5)	61.45(7)	82.27(10)	102.83(12)
	$\langle S^2 \rangle$	191.20(26)	286.25(38)	382.81(51)	478.72(64)
	$\langle A \rangle$	0.305(1)	0.305(1)	0.304(1)	0.305(1)
	$\langle S^2 \rangle_l$	299.81(48)	450.31(71)	600.03(95)	750.46(119)
Tri	$\langle e_1 \rangle$	150.16(24)	225.68(36)	300.95(48)	375.78(60)
	$\langle e_2 \rangle$	41.06(5)	61.71(7)	82.29(10)	102.77(12)
	$\langle S^2 \rangle$	191.22(25)	287.39(38)	383.24(51)	478.55(63)
	$\langle A \rangle$	0.304(1)	0.305(1)	0.305(1)	0.305(1)
	$\langle S^2 \rangle_l$	300.33(48)	451.13(72)	601.47(96)	752.39(12)

TABLE 3.5: Properties of 9-branch combs, for different number of monomers N, on a square (Sq) and triangular (Tri) lattice.

	Property	$N = 2201$	$N = 3301$	$N = 4401$	$N = 5501$
	$\langle e_1 \rangle$	143.30(21)	214.72(32)	287.01(42)	358.52(52)
	$\langle e_2 \rangle$	44.68(5)	67.05(7)	89.52(10)	112.02(12)
Sq	$\langle S^2 \rangle$	187.98(23)	281.77(34)	376.53(45)	470.54(57)
	$\langle A \rangle$	0.265(1)	0.265(1)	0.265(1)	0.265(1)
	$\langle S^2 \rangle_l$	366.87(58)	550.73(87)	733.82(117)	918.03(146)
	$\langle e_1 \rangle$	143.45(21)	214.80(31)	287.14(42)	358.85(53)
	$\langle e_2 \rangle$	44.83(5)	67.13(7)	89.67(10)	112.24(12)
Tri	$\langle S^2 \rangle$	188.28(23)	281.93(34)	376.81(45)	471.09(57)
	$\langle A \rangle$	0.265(1)	0.265(1)	0.265(1)	0.264(1)
	$\langle S^2 \rangle_l$	367.58(58)	550.83(88)	734.66(117)	918.83(146)

TABLE 3.6: Properties of 11-branch combs, for different number of monomers N, on a square (Sq) and triangular (Tri) lattice.

Property	Branch	$\square$ Lattice	$\triangle$ Lattice	Wei	Theory
	5-branch	0.710(2)	0.712(2)	0.7120(11)	0.7120_a
	7-branch	0.666(2)	0.666(2)	0.6676(2)	0.6676_a
g-ratio	8-branch	0.576(2)	0.577(2)	0.5781(7)	0.5781_b
	9-branch	0.638(2)	0.636(2)	0.6379(4)	0.6379_a
	11-branch	0.514(2)	0.514(2)	0.5131(5)	0.5131_b
	5-branch	0.310(2)	0.310(2)	0.3095(5)	
	7-branch	0.306(1)	0.306(2)	0.3047(4)	
$\langle A \rangle$	8-branch	0.275(1)	0.274(2)	0.2736(3)	
	9-branch	0.304(2)	0.306(2)	0.3042(2)	
	11-branch	0.265(2)	0.265(2)	0.2640(2)	
	5-branch	0.783(1)	0.784(2)	0.7838(2)	
	7-branch	0.784(1)	0.783(2)	0.7835(1)	
$\langle \delta_1 \rangle$	8-branch	0.765(1)	0.764(2)	0.7645(1)	
	9-branch	0.784(2)	0.785(2)	0.7850(1)	
	11-branch	0.762(2)	0.762(2)	0.7615(1)	

TABLE 3.7: Comparison of g-ratios, $\langle A \rangle$ and $\langle \delta_1 \rangle$ for extrapolated Square Lattice MC $\square$, Triangular Lattice MC $\triangle$ and Wei method also including g-ratios for theoretical infinite monomers values reported in the literature. For values with subscript a and b see reference [48] and [49] respectively.

Gorry and Bishop [55] used a pivot MC algorithm, where the initial polymer configuration is constructed by linking together tangent circular units with a diameter of one. Here the polymer is not restricted to a fixed lattice. A random number is used to select one of the monomers as a pivot and a second random number is employed to generate a random angle between $0°$ and $360°$. All the monomers further along the branch containing the pivot monomer are rotated by this angle. In this MC method successive samples are not independent and it is necessary to both discard the beginning phase of the simulation and to collect data at sufficiently large intervals so as to avoid correlation effects. Gorry and Bishop [55] found that the g-ratio = 0.713 ± 0.002 and $\langle A \rangle =$ 0.310 ± 0.001, which concur with the current results.

3.3.2 Five junction comb polymers in 2D and 3D

The $\langle S^2 \rangle$ data is reported in Tables 3.8 and 3.9. It was found that 2ν had the value of 1.00 ± 0.02 for both eleven and fourteen branched combs in two dimensions and $1.00 \pm$ 0.01 for both branched combs in three dimensions. These results are perfectly consistent with the theoretical value ($\nu = 0.5$).

The MC g-ratios have been calculated from the radius of gyration data in tables 3.8 and 3.9. These extrapolated g-ratios for the comb systems are compared to other findings in Table 3.10. Both Wei's method and the MC simulations produce similar results to each other and the theoretical predictions derived from form factor calculation when applying the Benhamou method. The g-ratios of the fourteen branch comb, which has a complete set of interior branches, have a relatively lower value than those found for the eleven branch combs.

The MC simulation results for asphericity $\langle A \rangle$ are also contained in Tables 3.8 and 3.9. The asphericity data display only a weak dependence on N. In Table 3.10 the extrapolated values found for $\langle A \rangle$ of ideal eleven and fourteen branch comb polymers concur with the theoretical prediction of the Wei method. As expected, the results indicate that the polymers become more symmetric in their shape as the structure changes to higher branching and a complete set of interior branches. Table 3.10 also demonstrates the similarity found between the analytical and MC results for the shape parameters $\langle \delta_1 \rangle$

	Property	$N = 2201$	$N = 3301$	$N = 4401$	$N = 5501$
	$\langle e_1 \rangle$	177.87(29)	266.89(43)	356.33(58)	444.52(72)
	$\langle e_2 \rangle$	47.86(6)	71.81(8)	95.75(11)	119.75(14)
11-branch	$\langle S^2 \rangle$	225.73(31)	338.69(46)	452.08(61)	564.27(76)
	$\langle A \rangle$	0.307(1)	0.307(1)	0.307(1)	0.307(1)
	$\langle \delta_1 \rangle$	0.788(1)	0.788(1)	0.788(1)	0.788(1)
	$\langle S^2 \rangle_l$	366.87(58)	550.73(87)	733.8(12)	918.0(15)
	Property	$N = 2801$	$N = 4201$	$N = 5601$	$N = 7001$
	$\langle e_1 \rangle$	169.70(25)	255.02(38)	339.39(51)	423.65(64)
	$\langle e_2 \rangle$	52.34(6)	78.50(9)	104.75(11)	130.79(14)
14-branch	$\langle S^2 \rangle$	222.04(27)	333.52(41)	444.14(54)	554.45(60)
	$\langle A \rangle$	0.265(1)	0.265(1)	0.265(1)	0.264(1)
	$\langle \delta_1 \rangle$	0.764(1)	0.764(1)	0.764(1)	0.764(1)
	$\langle S^2 \rangle_l$	466.76(74)	701.5(11)	935.1(15)	1170.0(19)

TABLE 3.8: Properties of 11-branch, 14-branch and linear chains, for different numbers of monomers N, in two dimensions.

and prolateness $\langle P \rangle$. These quantities further indicate that the 14-branch structures are more compact than the 11-branch ones.

The form factor $S(k)$ results are presented in Figures 3.3, 3.4 and 3.5. The reciprocal of the form factor is plotted to emphasize differences at higher values of **k**. Figure 3.3 illustrates the number dependence of the three dimensional 14-branch MC simulation. The solid curve is the exact result for an infinite polymer whereas the squares, the diamonds, and the circles are the MC values when N = 127, 267, and 687, respectively. Clearly, comb systems with $N = 687$ are large enough to match the infinite system prediction.

The three dimensional results are given in Figure 3.4a. The dashed curve, solid curve and dotted curve are the exact results for the 14-branch combs, the 11-branch combs, and linear chains, respectively. The circles, the diamonds, and the triangles are the MC values for the 14-branch combs when $N = 687$, the 11-branch combs when $N = 683$, and the linear chains when $N = 680$. The exact predictions and the MC simulations follow approximately the same function, except at the largest values of x, where these larger values of x examine smaller values of distance. At this scale, the detailed structure of a polymer has a significant effect on the form factor. Increased crowding is clearly seen as the polymer is changed from a linear to an 11-branch and then to a 14-branch comb.

	Property	$N = 2201$	$N = 3301$	$N = 4401$	$N = 5501$
	$\langle e_1 \rangle$	159.38(23)	238.09(34)	317.71(46)	397.06(58)
	$\langle e_2 \rangle$	46.87(5)	70.16(7)	93.49(9)	116.80(12)
	$\langle e_3 \rangle$	20.75(2)	31.04(3)	41.37(4)	51.75(5)
11-branch	$\langle S^2 \rangle$	226.99(25)	339.28(37)	452.58(50)	565.61(62)
	$\langle A \rangle$	0.299(1)	0.299(1)	0.299(1)	0.299(1)
	$\langle P \rangle$	0.308(1)	0.308(1)	0.308(1)	0.308(1)
	$\langle S^2 \rangle_l$	366.98(47)	550.18(71)	732.75(94)	915.0(12)
	Property	$N = 2801$	$N = 4201$	$N = 5601$	$N = 7001$
	$\langle e_1 \rangle$	148.73(20)	223.14(30)	297.59(40)	371.26(51)
	$\langle e_2 \rangle$	49.65(5)	74.44(7)	99.26(9)	123.97(11)
	$\langle e_3 \rangle$	23.74(2)	35.59(3)	47.40(4)	59.30(5)
14-branch	$\langle S^2 \rangle$	222.15(22)	333.18(33)	444.25(44)	554.43(55)
	$\langle A \rangle$	0.256(1)	0.256(1)	0.256(1)	0.255(1)
	$\langle P \rangle$	0.240(1)	0.240(1)	0.240(1)	0.239(1)
	$\langle S^2 \rangle_l$	466.46(60)	700.32(90)	933.4(12)	1166.4(15)

TABLE 3.9: Properties of 11-branch, 14-branch and linear chains, for different numbers of monomers N, in three dimensions.

Similar behaviour in two dimensions is displayed in Figure 3.4b. The curves represent the exact equations as in Figure 3.4a. The exact expressions for ideal systems do not depend upon the spatial dimension. Here the circles are the MC values for the 14-branch combs when $N = 701$, the diamonds are the MC values for the 11-branch combs when $N = 694$, and the triangles are the MC values for the linear chains when $N = 680$. As was the case in three dimensions, similar behaviour is observed between the MC simulations and the exact predictions.

Figure 3.5 presents a direct comparison of the MC data in two and three dimensions. Here, lines represent the three dimensional data and markers the two dimensional data. The results indicate that indeed the spatial dimension does not effect the behaviour of ideal polymers.

Wei's and Benhamous's methods and a Monte Carlo growth algorithm have been used to investigate branched five junction comb polymers in the ideal regime. The g-ratio, the asphericities (with their respective error bars) and the form factor have been determined for a wide range of N. It is found that the values obtained by of all the techniques are in excellent agreement with each other and the available theory.

	Property	MC 2D	Wei 2D	Exact	MC 3D	Wei 3D
11-branch	g-ratio	0.616(2)	0.61683(0)	0.61683	0.617(1)	0.61683(0)
	$\langle A \rangle$	0.307(2)	0.30616(0)		0.299(1)	0.29919(9)
	$\langle \delta_1 \rangle$	0.788(2)	0.78750(8)			
	$\langle P \rangle$				0.308(2)	0.15401(5)[a]
14-branch	g-ratio	0.473(2)	0.47521(9)	0.47522	0.475(2)	0.47521(9)
	$\langle A \rangle$	0.264(2)	0.26441(0)		0.255(2)	0.25566(6)
	$\langle \delta_1 \rangle$	0.764(2)	0.76393(6)			
	$\langle P \rangle$				0.239(2)	0.11976(5)[a]

TABLE 3.10: Results for eleven and fourteen branch comb polymers in two and three dimensions. Note that the Wei method defines $\langle P \rangle$ with an additional factor of $1/2$.
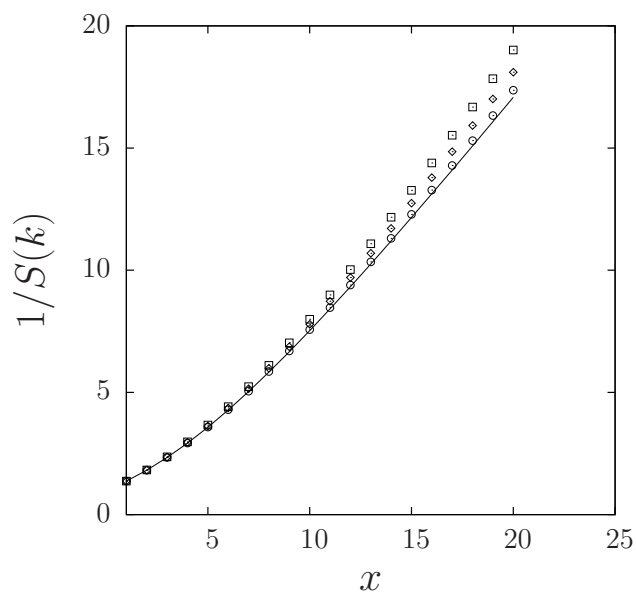
FIGURE 3.3: Inverse of the form factor $S(k)$ as a function of $x = k\langle S^2 \rangle_\ell^{1/2}$ for different values of N for a fourteen branch comb. The solid curve is the exact result for infinite N. The squares, diamonds and circles are the MC values for N equal to 167, 267 and 687 respectively.

3.3.3 Dendrimers

Here, we present the results found for first and second generation dendrimers, in 3D considering the shape parameters: g-ratio, asphericity and prolateness. These structures have a different branching system to that of the comb polymers as described in subsection 3.2.1 and depicted in Figure 3.1. For all dendrimers the scaling exponent, 2ν was found to be: 0.992(1), 0.986(1), 0.992(1), and 0.992(1) for nine, twelve, twenty one and thirty nine branches, respectively.

In Tables 3.12 and 3.11, we present the MC simulations results for the shape parameters: g-ratio, $\langle A \rangle$ and $\langle P \rangle$ for finite N. The extrapolated results for these parameters are then

(a) 2D

(b) 3D
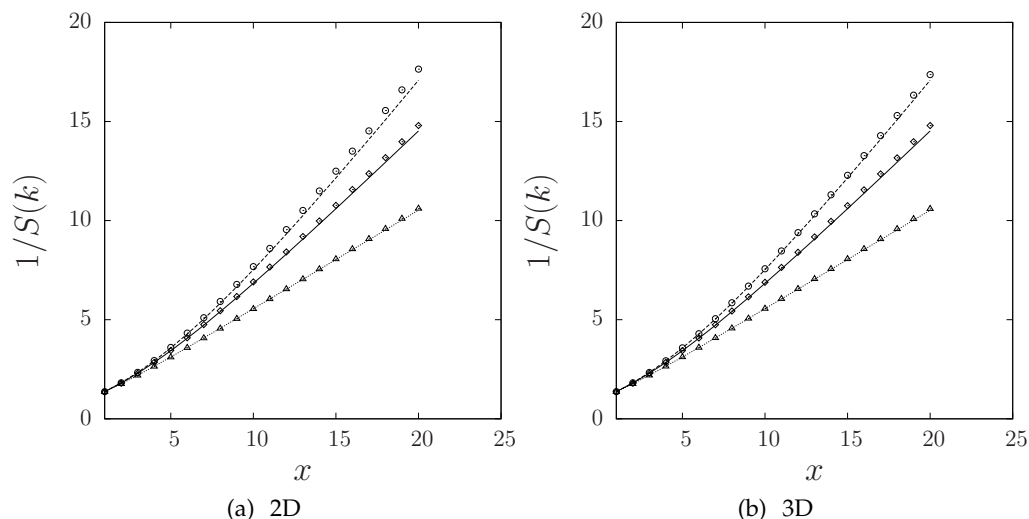
FIGURE 3.4: Comparison of the two and three dimensional MC simulation for the reciprocal of the form factor to the exact results. In both: a) 2D and b) 3D, the dashed curve is the exact result for the 14-branch combs, the solid curve is the exact result for the 11-branch combs, and the dotted curve is the exact result for linear chains. The circles are the MC values for the 14-branch combs when N = 687(2D) and 701(3D), the diamonds are the MC values for the 11-branched combs when N = 683(2D) and 694(3D), and the triangles are the MC values for the linear chains when N = 680 for both 2 and 3D.

given in Table 3.13. As was the case with the combs structures we find all calculated values for various shape parameters to be in good agreement with each other. All shape parameters values decrease as the number of generations and branches per junction are increased. These results indicate that dendrimers with more generations and branches per junction tend to be more symmetric and compact in nature.

3.4 Conclusions

We have applied the Wei and Benhamou method in conjunction with Monte Carlo simulations on square, triangular and cubic lattices to investigate two and three dimensional branched comb and dendrimers polymers in the ideal regime. The g-ratios, the asphericities, shape factors in two and three dimensions including their respective error bars and form factors have been calculated for a wide range of number of monomers N. It is found that the extrapolated values of all the techniques are in excellent agreement with each other and the available theory.
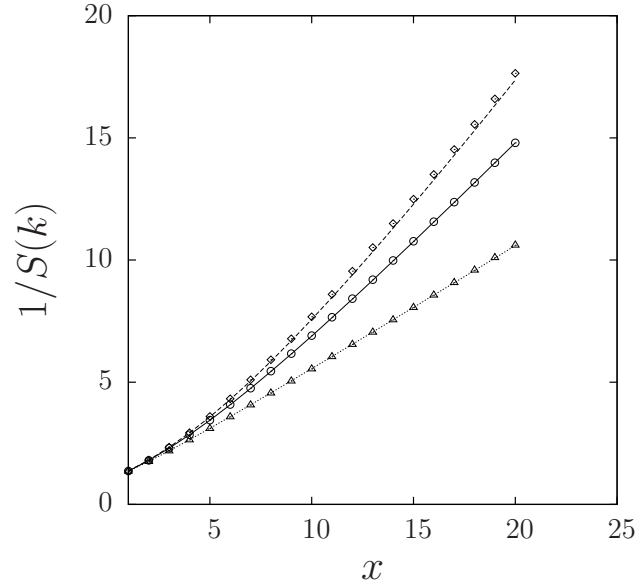
FIGURE 3.5: Comparison of the three and two dimensional MC simulations for the reciprocal of the form factor to the exact results. The dashed curve is the 3D MC results for the 14-branch combs, the solid curve is the 3D MC result for the 11-branch combs, and the dotted curve is the 3D MC result for linear chains (Debye equation). The circles are the 2D MC values for the 14-branch combs, the diamonds are the 2D MC values for the 11-branched combs, and the triangles are the 2D MC values for the linear chains.

	Property	$N = 55$	$N = 181$	$N = 442$	$N = 892$	$N = 1342$
	$\langle e_1 \rangle$	3.76(1)	12.22(2)	29.73(5)	60.07(9)	90.27(14)
	$\langle e_2 \rangle$	1.39(1)	4.40(1)	10.64(2)	21.38(3)	32.14(5)
	$\langle e_3 \rangle$	0.57(1)	1.81(1)	4.33(1)	8.71(1)	13.08(2)
9-branch	$\langle S^2 \rangle$	5.73(1)	18.43(2)	44.71(5)	90.17(11)	135.50(16)
	$\langle A \rangle$	0.255(1)	0.262(1)	0.264(1)	0.266(1)	0.266(1)
	$\langle P \rangle$	0.215(1)	0.226(1)	0.230(1)	0.233(1)	0.233(1)
	$\langle S^2 \rangle_l$	9.16(1)	30.16(5)	73.60(12)	148.32(24)	223.33(4)
	Property	$N = 148$	$N = 232$	$N = 1030$	$N = 2080$	$N = 3120$
	$\langle e_1 \rangle$	6.88(1)	10.73(1)	47.31(6)	95.50(13)	143.53(20)
	$\langle e_2 \rangle$	2.93(1)	4.53(1)	19.77(2)	39.84(5)	59.85(8)
	$\langle e_3 \rangle$	1.34(1)	2.08(1)	8.99(1)	18.07(2)	27.13(3)
21-branch	$\langle S^2 \rangle$	11.16(1)	17.34(2)	76.07(8)	153.41(16)	230.51(24)
	$\langle A \rangle$	0.203(1)	0.206(1)	0.210(1)	0.211(1)	0.211(1)
	$\langle P \rangle$	0.143(1)	0.147(1)	0.152(1)	0.153(1)	0.153(1)
	$\langle S^2 \rangle_l$	24.66(4)	38.61(6)	171.24(28)	345.21(56)	521.73(85)

TABLE 3.11: Properties of 9-branch, 21-branch and linear chains, for different numbers of monomers N, in three dimensions.

	Property	$N = 109$	$N = 289$	$N = 409$	$N = 589$
	$\langle e_1 \rangle$	5.87(1)	15.40(2)	21.77(3)	31.27(5)
	$\langle e_2 \rangle$	2.22(1)	5.72(1)	8.06(1)	11.54(2)
	$\langle e_3 \rangle$	1.00(1)	2.54(1)	3.57(1)	5.12(1)
12-branch	$\langle S^2 \rangle$	9.08(1)	23.66(3)	33.40(4)	47.93(6)
	$\langle A \rangle$	0.236(1)	0.242(1)	0.243(1)	0.244(1)
	$\langle P \rangle$	0.195(1)	0.203(1)	0.205(1)	0.206(1)
	$\langle S^2 \rangle_l$	18.15(3)	48.23(8)	68.20(11)	97.91(16)
	Property	$N = 547$	$N = 937$	$N = 1912$	$N = 2887$
	$\langle e_1 \rangle$	14.39(2)	24.52(3)	49.90(6)	75.34(9)
	$\langle e_2 \rangle$	6.72(1)	11.41(1)	23.18(2)	34.90(4)
	$\langle e_3 \rangle$	3.52(1)	5.97(1)	12.06(1)	18.16(2)
39-branch	$\langle S^2 \rangle$	24.63(2)	41.91(4)	85.14(8)	128.40(12)
	$\langle A \rangle$	0.162(1)	0.163(1)	0.164(1)	0.165(1)
	$\langle P \rangle$	0.101(1)	0.102(1)	0.102(1)	0.104(1)
	$\langle S^2 \rangle_l$	90.89(15)	155.45(25)	317.37(52)	480.86(78)

TABLE 3.12: Properties of 12-branch, 39-branch and linear chains, for different numbers of monomers N, in three dimensions.

	Property	MC	Wei	Exact
	g-ratio	0.606(1)	0.60493(0)	0.6049
9-branch	$\langle A \rangle$	0.266(1)	0.26583(9)	
	$\langle P \rangle$	0.233(1)	0.11613(0)[a]	
	g-ratio	0.486(1)	0.48611(0)	0.4861
12-branch	$\langle A \rangle$	0.211(1)	0.21076(9)	
	$\langle P \rangle$	0.208(1)	0.10438(0)[a]	
	g-ratio	0.442(1)	0.44152(0)	0.4415
21-branch	$\langle A \rangle$	0.246(1)	0.24560(1)	
	$\langle P \rangle$	0.154(1)	0.07633(0)[a]	
	g-ratio	0.267(1)	0.26586(0)	0.2658
39-branch	$\langle A \rangle$	0.169(1)	0.16484(4)	
	$\langle P \rangle$	0.109(1)	0.05193(0)[a]	

TABLE 3.13: Results for nine, twelve, twenty one and thirty nine branch dendrimers in three dimensions. Note that the Wei method defines $\langle P \rangle$ with an additional factor of $1/2$.

The branching of the ideal polymer strands does not affect the universal two-dimensional ideal (non interacting) scaling behaviour. Thus indicating the end-end distances within strands, as well as for the branched structure as a whole, will scale with the scaling exponent $\nu = 1/2$. In three dimensions this has been verified by Kosmas et al. [56] who found that the interior branches and the overall four junction comb structure had the same ideal exponent in the ideal regime.

This behaviour will break, however, if we take into account an environment with correlated disorder. Under the influence of such a correlated (disordered) environment the polymer strands, even if they display no self-interaction, will change their conformation due to the disordered background [57].

Chapter 4

Public transportation networks in the UK: topology and spatial embedding

Over the last few decades society has become increasingly dependent on public transport networks (PTN) to facilitate commuters and the movement of commodities on both a local and global scale. With transport playing such an significant role in the economy of cities and countries, it is becoming increasingly important to better understand the underlying structure and properties within these complex systems. A deeper knowledge will assist in the development of cost effective methods to evaluate the efficiency and robustness of existing PTNs. This chapter is dedicated to studying these properties.

This is achieved by applying methods in network science and percolation theory to estimate the stability of a network to random failure and employing fractal analysis to determine the serviceability of stations. Here, PTNs on both a local and global scale, within the UK are investigated. These include the local municipal areas of Bristol, Greater Manchester, West Midlands, Greater London and the national rail and coach networks of mainland, UK.

One approach to study these networks is offered through complex network science, a recently established research field with a firm theoretical background and a broad range of applications. It has successfully explained numerous phenomena that have

emerged in natural and man made systems involving separate agents connected via various types of interactions (Albert and Barabási, 2002; Dorogovtsev and Mendes, 2003; Barrat et al., 2008; Newman, 2010).

The purpose of this chapter is to investigate the properties of PTN in the UK using both complex networks (i.e. topological) as well as spatial descriptions in order to gain useful insights into robustness and efficiency of PTN. Currently, there exists a relatively broad literature on the application of a complex networks for public transport analysis (see, in particular, the discussion below). Here, we add to the existing analysis and further explore certain topological measures which can be used to classify PTN with respect to their stability to random failures. We add to this analysis by studying the fractals properties of these systems which offer insight into the serviceability and efficiency of PTN. Together both the topological and spatial features studied here may contribute to a better understanding of the underlying mechanisms governing PTN growth and modeling.

Out of those, two PTN operate on an nation-wide scale (national coach and rail networks) and the remaining four are PTN of Bristol, Manchester, West Midlands and Greater London. By this choice we attempted to have examples of areas of different geographical and economical scales. In turn, this enables one to seek for general (universal) characteristics of transportation system as a whole.

Our main practical results include the link found between topology, stability and efficiency of networks. This highlights the interplay that exists between the efficiency and stability, for not only PTNs but networks in general. The fractal methods applied here are able to extrapolate useful information in terms of the serviceable area of stations.

The layout of the chapter is as follows: In the next section 4.1, an overview of the previous analysis of PTN where a complex network approach is applied is given; section 4.2 discusses the data set; section 4.3 considers the topological properties of PTN including measures of robustness; section 4.4 covers fractals methods in determining the serviceability of stations; and finally, in section 4.5 some concluding remarks are given.

The main results of this chapter have been published in [13].

4.1 Complex network and Statistical physics approaches in quantifying the behaviour of public transport

Complex network analysis of PTNs began comparatively recently when, in 2002, Latora and Marchiori published their work analysing the topological properties of the Boston subway [58]. Subsequently, similar analysis has been performed for many other PTNs around the world. This can be seen in Figure 4.1, where black dots indicate the approximate geographical positions of PTNs that have been analysed via a complex network approach. The size of these PTNs have ranged from 152 to 45000 stations. The types of PTNs investigated include the subway [58], [59], bus [60]–[63], rail [64], air [65]–[67] and various combinations of these [68]–[71].

Thus far, a number of different topological representations of a network have been developed, by attributing different constituents of a network to vertices and edges. This is for the purpose of extracting different information about the PTN. For example one can represent each station as a node. Then link together all nodes that belong to a particular route, to form a complete subgraph (clique). Different subgraphs will then be joined via common stations, that share different routes. This representation is called **P**-space and has been applied in many studies [59], [60], [64], [68], [69], [72]. It is useful, in particular for determining the mean number of vehicle changes when traveling between any two points on a PTN service network.

In the so-called **B**-space [68], [73] one constructs a bipartite graph that contains nodes of two types: node-stations and node-routes. Only nodes of different types can be linked: a node-station is linked to the node-route if it belongs to that route. One can pass from such representation to a graph where only nodes of one type are present. This is achieved by the so-called single-mode projection, when all nodes of similar type that are linked to the common node of another type are represented as a complete subgraph. Naturally, the single-mode projection of the **B**-space graph to the nodes-stations leads to **P**-space. In turn, an analogous projection to the nodes-routes leads to the so-called **C**-space [68]. Here, one considers how routes are connected to each other and describes how routes are linked throughout the network. In **C**-space if any two routes service the same station they are obviously linked.

FIGURE 4.1: Map indicating various locations of the world where PTN have been analysed within the complex network science framework. Topological characteristics of some of them are further displayed in Table 4.2.

In this study we use the so-called **L**-space [58], [60], [68], [69], [74]. As we will see below it most naturally describes the properties of PTNs we are interested in. In this representation, a node in a graph corresponds to a PTN station. The **L**-space representation is constructed following a simple process. If two stations are adjacent in a route, a link is formed between the two stations, as it can be seen in Figure 4.2. However, if there are multiple routes going through the same two stations, **L**-space will not reflect this as it does not permit multiple links. This topology is ideal for studying the connectivity of networks. For example calculating metrics like mean shortest path length $\langle \ell \rangle$, Giant Connected Component (GCC) size and other similar metrics. This space is probably the most commonly used topology and has been applied in many different studies of real world networks [68], [69]. The **L**-space topology can further be extended to a weighted network. Where, weights indicate a distance or time between stations [58]. This added dimension, can often give a more realistic view of real world systems.

Over the years, features such as 'harness' behaviour have become of interest. This concept describes how different routes tend to follow similar paths for a certain number of stations. The harness distribution $P(r, s)$ is defined as the number of sequences of consecutive stations s, serviced by r parallel routes. This distribution has been analysed in Refs.[68], [75]. A similar feature has also been studied for weighted networks

Some materials have been removed due to 3rd party copyright. The unabridged version can be viewed in Lancester Library - Coventry University.

a

b

c

d
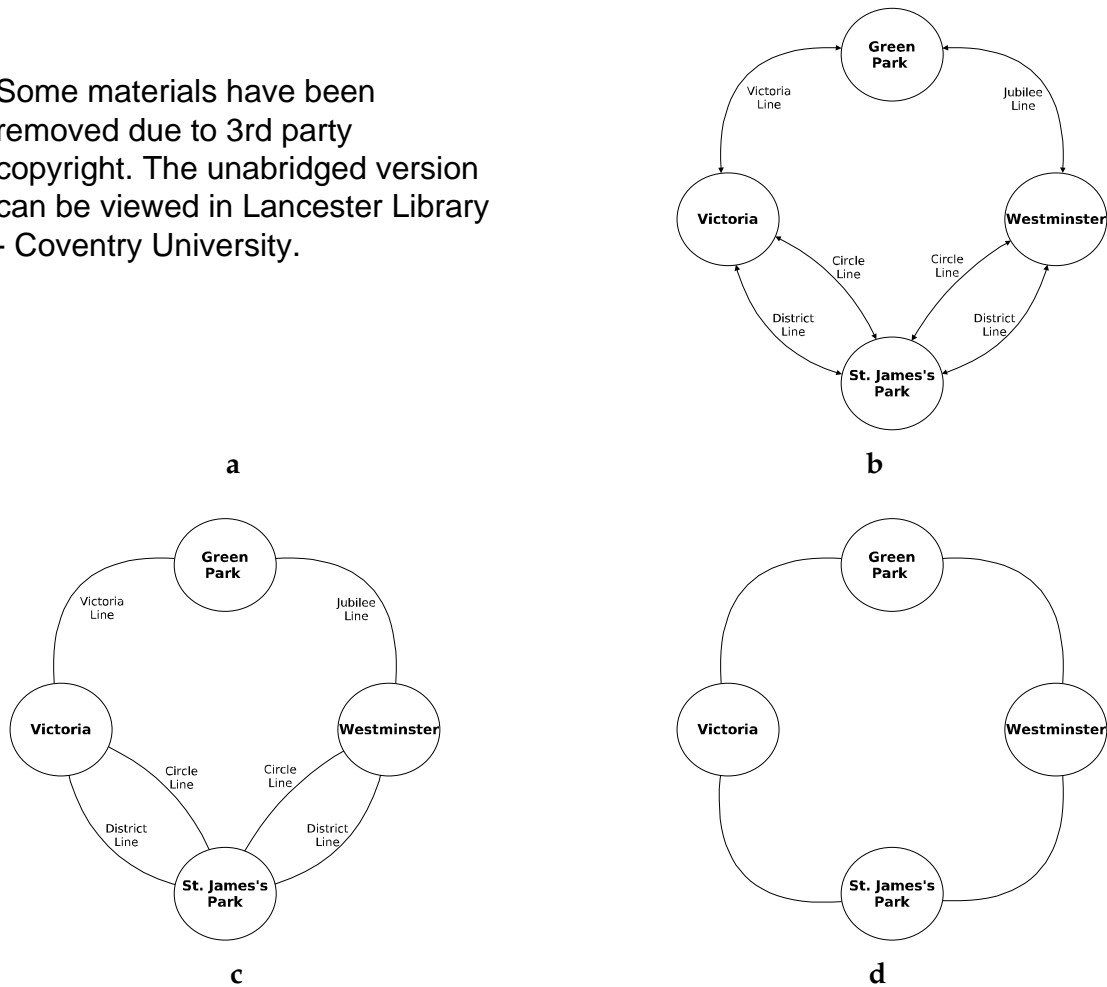
FIGURE 4.2: A fragment of London PTN and its representations in the form of a graph. **a**: a piece of the city map that includes several PTN stations (such representation will be called the geo-space onwards). Lines of different colour on the map correspond to different PTN routes; **b**: the stations on map **a** shown as graph nodes with links indicating out- and - incoming routes; **c**: the same as **b**, but directions of routes are not shown; **d**: the same as **c** but multiple links are reduced to single ones. This is the **L**-space used in our analysis.

in Ref.[60]. In both studies the harness distributions produced, exhibited scale free behaviour.

The question of PTN robustness to random break down or targeted removal of their constituents has also been considered in Refs.[76]–[78]. One of the goals of such studies is to present criteria, that allow priori quantification of the stability of real world correlated networks of finite size. Moreover, to investigate how these criteria correspond to analytic results available for infinite uncorrelated networks. The analysis focused on the effects that defunct or removed constituents (stations or joining links) have on the

properties of PTNs. Simulating different directed attack strategies, vulnerability criteria have been derived that result in minimal strategies that have a substantial impact on these systems.

The above empirical research has revealed that PTNs constructed in cities with different geographical, cultural and historical background share a number of basic common topological properties. They appear to be strongly correlated structures with high values of clustering coefficients and comparatively low mean shortest path values. Moreover, their node degree distributions are often found to follow exponential or power-law decay.

In turn, collected empirical data has lead to the development of a number of simulated growth models for PTNs. In Ref.[75] interacting self avoiding walks on a 2D lattice, with preferential attachment rules are applied to produce similar statistics to those found in PTNs. In Ref.[62], PTNs are grown one route at a time by employing an ideal n-depth clique topology. In Ref.[61] the optimised growth of a route is studied by considering two competing factors: investors and clients. Clients want routes to be as direct as possible to save time. Whereas, investors want routes to meander, in order to collect as many passengers to maximising profits.

Thus far, the main subject of analysis concerned topology. This analysis has lead to substantial progress in understanding the collective phenomena taking place on PTN networks. Another essential ingredient in the analysis of PTNs is the spatial embedding of networks.

There have been a few studies that directly consider PTN spatial analysis which mainly focus on modes of transport such as railway and subway [61], [68], [79]–[83] and with the availability of data improving more studies are sure to follow. In general, the fractal dimension, d_f (see section 4.4.1 and Eq.4.2 for a more detailed explanation of d_f) of a PTN in these studies were determined by counting the number of stations as a function of the radius from the centre of the city, $N(r)$. One of the earliest studies performed was by Thibault in Ref.[79] where in three Lyon regions the rail, bus and drainage networks were analysed and shown to have fractal dimension ranging between 1.64-1.88, 1-1.45 and 1.21-1.79 respectively. These fractal dimensions all show that as the radius from

the centre of a city increases the density of rail, bus and drainage decreases. Rail displays the largest values of fractal dimensionality with the least variance thus indicating that the length of track decreases more slowly than the number of bus stations as the distance is increased from the centre of the city. In other studies the Stuttgart railway fractal dimension d_f was found to be $d_f = 1.58$ [84]. The Paris railway was estimated to have $d_f = 1.47$ [82]. In Ref.[80], exponents of $d_f = 1.70 \pm 0.05$ for the Moscow and Rhinetowns railways were found. The Paris metro was also found to produce a value of $d_f = 1.80 \pm 0.05$. In Ref.[85], the Seoul transportation network had the value of $d_f = 1.5$ for stations and $d_f = 1.35$ for the railway track.

The majority of the above mentioned research has considered either mainly the topological or spatial properties of PTN. A particular feature of the research we present below is a cumulative analysis of both the topological and geographical characteristics of PTN. To this end we have chosen to consider six GB PTN using the data available on the National Transport Data Repository [86]. From these, two PTN operate on an nation-wide scale (national coach and rail networks) and the remaining four are local PTN for Bristol, Manchester, West Midlands and Greater London. In the next section we explain the origin of the data and how it will be used in our analysis.

4.2   Description of PTN database

The data for this study originates from the National Transport Data Repository (NTDR) website [86]. The website has an Open Government License meaning it is open to the public. It contains information on public transport travel and facilities throughout the UK for the years 2004 to and including 2011. The information provided, is a yearly snapshot of the public transport network for a sample week in each year. The week on which the data is usually recorded is either the first or second week in October to avoid recording during school holidays or other seasonal variations which are at a minimum during this period according to NTDR.

The data is collected and assembled following a decentralised system where individual regional travel lines (RTL) are responsible for recording the travel within their allocated districts. These records are then sent to the NTDR to be collated into one comprehensive

database. There are 11 RTLs that provide the NTDR with data, these are: Scotland, North East and Cumbria, North West, Yorkshire, Wales, West Midlands, East Midlands, East Anglia, South East, South West and London. The data for national coach and rail are the only data sets to be compiled centrally. Using a decentralised method for data retrieval may have benefits especially when it comes to efficiency, however, it does create more opportunity for errors. For example duplication of routes and stations on routes that span borders of two or more RTLs. Other complications, result from slight differences occurring in the formatting of the data sent to the NTDR. However, to prevent such errors the NTDR has an explicit document detailing the format of the data. Nevertheless, there remain slight differences in the format which need to be taken into account when analysing the data. Figure 4.3 is a snapshot of data taken from the Bristol bus network in its raw form.

The data set includes transport modes for national coach and rail which span UK's mainland. Bus networks for all cities in the UK are covered. Metro systems for larger Metropolis areas such as: London, Greater Manchester and the West Midlands. Some of these networks are subsets of others i.e. one PTN might cover a county and another a city within that particular county.

```
ATCO-CIF0510    Bristol, City of - BUS        ATCOPT20050420185520
QSNBL  C7C9620040926      1111100 X178              I
QO01000053210 0720  T1F0
QI0100BRP9037307220722B  T0F
QI0100BRA1079107250725B  T0F
QI0100BRA1079507260726B  T0F
QLN01000053210 Bus Station Bay 12              E0057160
QBN01000053210 358927  173514              Kingsdown
QLN0100BRP90373The Centre (Cp)              E0057160
QBN0100BRP90373358656  173009              Bristol City Centre
```

FIGURE 4.3: A snapshot of data from the Bristol bus network in its raw form. Each line starting with QS represents a route, each line beginning with QO, QI, QT represent the start, intermediate and terminal station respectively for route. The following 12 characters on these lines are the unique station identifiers which is directly followed by the times station arrive and leave. Lines starting with QL and QB provide information on spatial coordinates.

For each mode of transport that a city or county offers, which could be any combination of coach, train, metro, and ferry, a separate file is held in the records. For each station

the type of information that can be extracted is the following: the location of the station within a particular route; first, intermediate or last; number of times a station is visited throughout the day; geographical coordinates, using an Easting and Northing reference system; whether the route is incoming or outgoing and which routes these nodes belong to.

In the analysis we consider only outgoing routes. The reason for this is that in general the incoming and outgoing stations are usually on opposite sides of the road or very nearby. So instead of having a directed network one can assume both incoming and outgoing stations are the same and reduce the network to an undirected network. This approach allows a more intuitive interpretation of the network statistics. For example if two stations are next to each other where one is on the incoming and the other on the outgoing line. Then in a directed network they are actually far apart as the passenger would have to travel all the way to the beginning of the line and return on corresponding opposite route to reach the station across the road. This is avoidable in the case of an undirected network. Using this method would obviously cause problems if these incoming and outgoing stations were not close to each other, but we discard such situations as highly improbable.

There are a few errors in the data that do require attention and there is some missing data that needs to be considered. However, in general the database provides a rich platform which we intend to use to analyse the topological and spatial aspects of PTNs in the UK.

4.3   Topological properties of PTNs

With the available data on each PTN connectivity, it is straightforward to get their representation in the form of graphs in the **L**-space. As we will demonstrate below, such graphs posses a number of highly non-trivial features and thus can be considered as complex networks [1], [24]–[26]. Each network can be uniquely described in terms of its adjacency matrix $\hat{A}$, as described in subsection 2.2.1. In turn, based on the adjacency matrix constructed for each PTN under consideration, we are in the position to extract the main observables that are commonly used to quantify network properties.

Networks	N	M	$\hat{R}$	$\langle k \rangle$	$\langle k^2 \rangle$	k_{max}	GCC	$\langle \ell \rangle$	$\langle \ell_r \rangle$	ℓ_η	D	$\langle \ell_t \rangle$	r	C	C_r
Coach	2499	3228	258	2.58	12.17	43	96.1	23.1	7.22	3.2	121	359.2	0.27	0.075	1×10^{-3}
Rail	2575	4450	504	3.46	20.03	31	99.9	11.82	5.76	2.1	48	162.9	0.24	0.309	1×10^{-3}
Bristol	2580	3101	172	2.40	6.58	10	99.8	27.7	7.92	3.5	122	40.0	0.26	0.034	9×10^{-4}
Manchester	10742	12275	862	2.29	5.87	21	100	48.7	10.10	4.8	238	61.0	0.20	0.028	2×10^{-4}
West Mid	11743	15034	521	2.56	7.75	20	100	52.8	9.06	5.8	168	55.4	0.13	0.035	2×10^{-4}
London	16397	18496	767	2.26	5.56	11	99.8	53.5	10.82	4.9	206	72.5	0.21	0.014	1×10^{-4}

TABLE 4.1: General statistics for the PTN under consideration and main topological features of their representations in a form of complex networks. N, M: number of nodes and links of a PTN network with $\hat{R}$ routes. $\langle k \rangle$, $\langle k^2 \rangle$, k_{max}: mean, mean square and maximal node degree. GCC: size of the giant connected component in percents to the general network size. $\langle \ell \rangle$, $\langle \ell_r \rangle$: mean shortest path length of a PTN and of a random graph of the same size, ℓ_η: $\langle \ell \rangle / \langle \ell_r \rangle$, D: diameter (maximal shortest path length). $\langle \ell_t \rangle$: mean shortest travel time, in minutes. r: assortativity. C, C_r: mean clustering coefficient of a PTN and of a random graph of the same size.

The first two columns in Table 4.1 give the number of nodes N and edges M for each network.[1] The number of nodes directly corresponds to the number of PTN stations. The number of edges in **L**-space gives a reduced value of real linkage between the stations, see Figure 4.2. In the table we also display the number of routes $\hat{R}$ for each PTN, this does not have its counterpart in network topology in the **L**-space.

Table 4.1 gives mean $\langle k \rangle$, mean square $\langle k^2 \rangle$ and maximal values $k_{\max}$ of k for each of the networks. It can be seen that $\langle k \rangle$ for rail is slightly higher than for the other PTNs. Which is a result of railway networks tending to possess more high degree nodes, also known as hubs, than other modes of public transport studied here.

One of the indicators to measure the distance between nodes, providing a useful measure of the efficiency of a PTN, is given by the mean shortest path length $\langle \ell \rangle$. It is instructive to compare properties of the networks under consideration with those of the Erdös-Rényi classical random graph of the same size, i.e. when the same number of nodes N are randomly linked together by M links. To proceed, we simply calculate $\ell_\eta = \langle \ell \rangle / \langle \ell_r \rangle$, where a larger value of ℓ_η represent a less efficient network. It can be seen in Table 4.1 that larger PTNs tend to be less efficient than their smaller counterparts. It is instructive also to calculate the mean shortest path for weighted networks. For national networks, on average, it takes twice as long to get to any other station within the network on coach as it does on rail. One has to bear in mind a limitation of this measure when relating to PTN efficiency. As PTN routes will not necessarily follow the shortest path between two points.

Local correlations, often present in real-world complex networks are measured by the clustering coefficient C, see Eq.(2.13). The comparison of data for PTN clustering coefficients C with that of the random graph of the same size C_r gives undoubted evidence of strong correlations in PTN networks: $C/C_r \sim 10^2$ for almost all networks.

Much attention has recently been paid to small-world networks [22] characterised by strong local correlation and small topological distance. We know that PTNs are highly correlated (see previous paragraph). However, in Table 4.2, comparatively large mean

---

[1]Data for London's PTN slightly differs from those presented in Ref.[68]. This is because an updated database has been used in our study.

shortest path lengths are found when comparing to random networks. Therefore, caution is to be taken when attributing small world properties to PTNs. This may be understandable as many nodes of degree two exist in PTNs.

Correlation between degrees of neighbouring nodes in a network are usually measured in terms of the assortativity r, defined by Eq.(2.17). The values of r found in this study, although small, clearly indicate assortative mixing, see from table 4.2, where $0.1 \leq r \leq 0.3$, for the PTNs under consideration. This means that links tend to connect nodes of similar degree. Let us note, that it is not always the case for PTN. As analysed in Ref.[68], the PTN of Düsseldorf, Moscow, Paris, São Paolo, show no preference in linkage between nodes with respect to node degrees ($r \simeq 0$). So in this respect PTN analysed in our study belong to a group which includes the cities of Berlin, Los Angeles, Rome, Sydney, Taipei ($0.1 \leq r \leq 0.3$) [68].

It is worth noting from the above observations, although PTNs studied span over different distances in the geographic space, their topological features manifest striking similarities! Indeed, all the networks considered in this study possess comparatively low values of the mean node degree, high clustering coefficients, they are disassortative with respect to node-node correlations. Moreover, the presence of high clustering in these networks is not accompanied by a low value of the mean shortest path length, as is usually expected for small world networks.

4.3.1 Topological measures of robustness

Obviously, network integrity plays a crucial role for various processes occurring on the network. In particular, transportation can not be maintained between nodes belonging to different network fragments that are not joined together. As one can see from Table 4.1, the largest connected component of each PTN, also known as the giant connected component GCC, includes almost all nodes, making any location on the network reachable from any other location.

The analysis of topological features of real-world networks aids in the prediction of their behaviour under removal of their constituents [24], [29], [87], [88]. A useful criterion, that originates from percolation theory, which allows one to evaluate network

City	Type	n	$\hat{R}$	κ	α
Dallas	B	5366	117	2.35	5.49
London	B	16397	767	2.46	4.25
West Midlands	B	11743	521	2.56	3.10
Manchester	B	10742	862	2.56	4.36
Istanbul	BST	4043	414	2.69	4.04
Los Angeles	B	44629	1881	2.73	4.85
Bristol	B	2580	172	2.74	3.56
Berlin	BSTU	2992	211	3.16	4.30
Düsseldorf	BST	1494	124	3.16	3.76
Hamburg	BFSTU	8084	708	3.26	4.74
Rome	BT	3961	681	3.67	3.95
Taipei	B	5311	389	4.02	3.74
Sydney	B	1978	596	4.37	4.03
Hong Kong	B	2024	321	5.34	2.99
Saõ Paolo	B	7215	997	5.95	2.72
Paris	BS	3728	251	6.93	2.62
Moscow	BEST	3569	679	7.91	3.22

TABLE 4.2: Molloy-Reed parameter κ (4.1) for several PTNs of the UK cities (our data: first three lines of the table) in comparison with PTNs of some other cities of the world, as obtained in [76]. The type of transport taken into account are: Bus:B; Electric trolleys:E; Ferry:F; Subway:S; Tram:T; and Urban train:U. PTN size is given in terms of number of stations N and of routes $\hat{R}$. The last column gives an exponent α for a the power law (2.10) fit for node degree distribution.

vulnerability is known as the Molloy-Reed criterion [89]. It states that in any uncorrelated network the GCC is present if:

$$\kappa = \langle k^2 \rangle / \langle k \rangle \geq 2\,. \tag{4.1}$$

The Molloy-Reed parameter κ allows one to evaluate network stability to random failures. The higher the value of κ, the more stable a network, i.e. a larger number of nodes are required to be removed to destroy a given GCC. Although Eq.(4.1) has been obtained for infinite uncorrelated networks, its utility for PTN analysis has been proven recently by analysing the vulnerability of PTNs of different cities in the world [76], [77], [83]. In particular, it has been shown, that networks with higher values of κ are more stable to random removal of their nodes. In table 4.2 we compare values of κ for several UK city PTNs where κ is obtained by us with other city PTNs from around the world. The obtained values give a measure of stability of the networks considered.

It is also instructive to determine the values of κ for the PTN that cover larger geographic areas of the UK. In this case we get, $\kappa = 4.72$, and $\kappa = 5.79$ for national coach
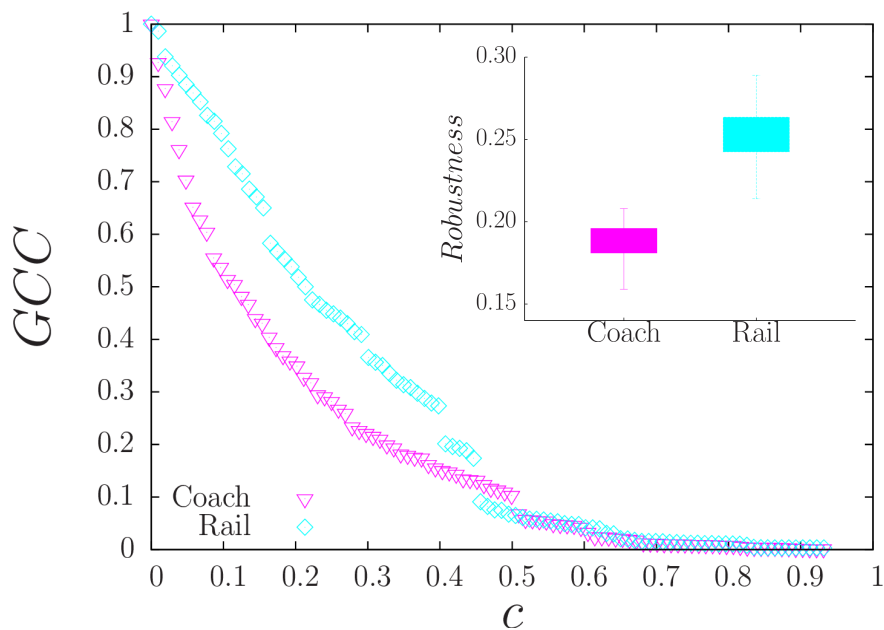
FIGURE 4.4: The normalized size of the largest connected component GCC of the coach and rail PTN as function of the share c of randomly removed stations. The insert represents the distribution of network robustness, measured as the area under the curve, for 100 simulations of random failure.

and rail networks, respectively. This signals an even higher stability of these networks in comparison to PTNs of some UK cities reported in Table 4.2. It will be interesting to check these values against their counterparts for networks covering larger geographic space in other regions of the world.

To further support the claim that the Molloy-Reed parameter is indeed a good indicator of network stability we performed a numerical simulation. Here, we simulate 100 random failures on both national, coach and rail PTNs to determine their mean robustness. For each simulation we begin with the original network, following this algorithm:

1. Determine the size of the GCC,

2. Randomly select 1% of the nodes in the network,

3. Remove the nodes selected in step 2, such that, when a node is removed the adjacent edges are also removed,

4. Calculate the new GCC,

5. Repeat steps 2-4 until all nodes are exhausted.

In Figure 4.4, we plot the normalised size of the largest connected component GCC, for coach and rail PTNs, as function of the fraction of randomly removed stations c. The area under the curve, acts as an analogue for robustness, where a larger area indicates a more robust network. Qualitatively, for this particular simulation, rail is more robust than coach. This measure of robustness can be quantified by simply integrating over the function c. As previously mentioned, we performed 100 simulations, where for each simulation the value for robustness is calculated. The distributions of this robustness measure, for both coach and rail, are plotted in the inset of Figure 4.4. Here, it can be seen that the robustness for rail is larger than its national counterpart, thus indicating rail as a more robust PTN. This confirms the original prediction of the Molloy-Reed parameter where values of $\kappa = 4.72$, and $\kappa = 5.79$ are obtained for coach and rail respectively. Thus supporting the idea that this easily determined metric can be employed to measure the stability of a PTN.

4.3.2 Degree distribution

As already highlighted in section 2 of this Thesis, cumulative distributions generally have smoother functions and their functional dependence enables a more accurate determination of $P(k)$. Hence, the corresponding cumulative distributions for all PTNs are plotted in Figure 4.5 in both log-linear and double logarithmic scales. On inspection it seems that the degree distributions of these networks show clear preference with respect to the power law decay $P(k) \propto k^{-\alpha}$ rather than an exponential decay $P(k) \propto e^{-\xi k}$. For confirmation, using a nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm [53], we have produced the fits for these distributions and display the fitted values of α and ξ in Table 4.3. As it follows from our analysis, the node degree distributions are better fitted by the power-law than by the exponential decay (for a detailed explanation of the method used to determine the best fit, see Appendix A).

Scale-free networks are characterised by unique properties, in particular with respect to their stability [90], [91]. Networks that have smaller values of the exponent α should
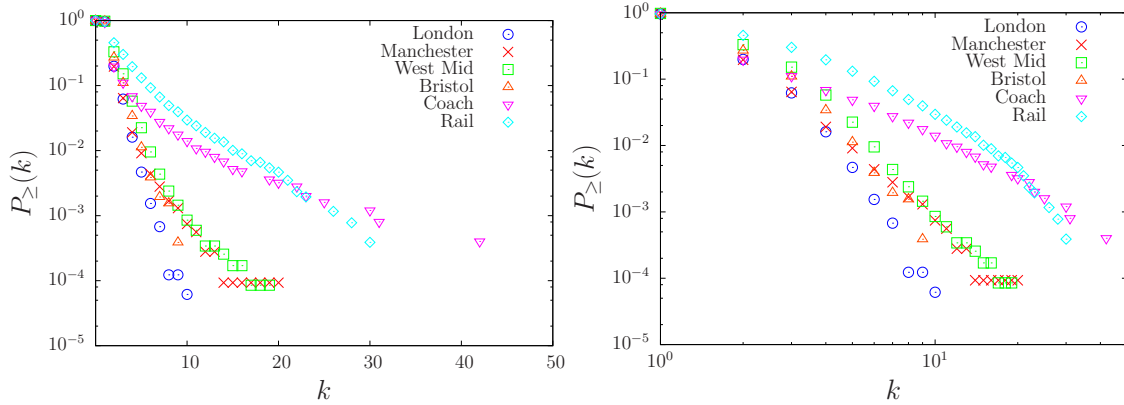
FIGURE 4.5: Cumulative degree distribution $P_{\geq}(k)$ for six PTNs under consideration in log-lin (left) and log-log (right) scales.

	Coach	Rail	Bristol	Manchester	West Mid	London
α	5.06 ± 0.15	2.5 ± 0.07	3.56 ± 0.09	4.36 ± 0.04	3.10 ± 0.07	4.25 ± 0.05
ξ	2.08 ± 0.09	0.90 ± 0.06	1.37 ± 0.07	1.73 ± 0.05	1.15 ± 0.05	1.67 ± 0.05

TABLE 4.3: Fitted degree distribution exponents α and ξ. For all PTN considered here, the $P(k)$ dependency is much better fitted by the power-law than by the exponential decay.

manifest stronger stability with respect to the removal of their constituents, see also Table 4.2. A prominent example, follows from the comparison of the UK national coach and rail networks: the α exponent for the rail PTN is only half that of its coach counterpart. This brings about higher stability in the former, under random removal of its constituents. It can seen in Figure 4.6 that a negative correlations exists between the Molloy-Reed parameter and degree distribution.
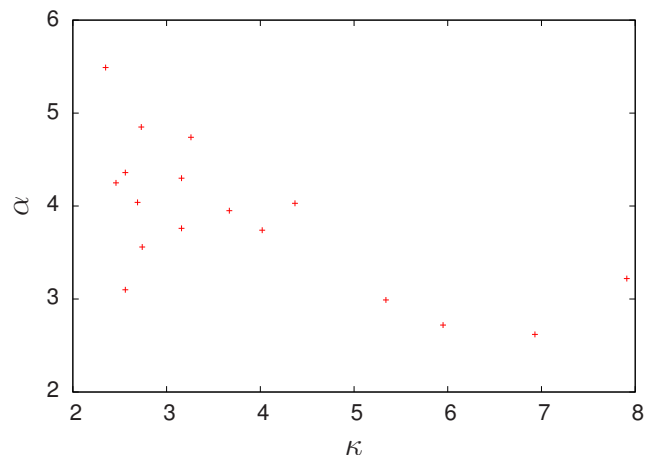
FIGURE 4.6: Correlation between Molloy-Reed parameter κ and degree distribution exponent α for PTNs in Table 4.2.

4.4   Geospatial properties of PTN

Thus far we have studied the properties of a PTN that originate from their topology. Such analysis is usual for the complex networks, even though the spatial embedding plays an important role for some, as in the case of PTN [92]. Very often data about network topology is not accompanied by their location in the embedding, Euclidean space. The advantage of the database we are using in our analysis is that it contains the geographic coordinates of stations. This gives us the unique possibility to complete analysis of topological properties by examining properties in the Euclidean two-dimensional ($d = 2$) space, which will be called geospace from here onwards. This neglects the slight curvature in the earth but does not effect calculations over the area considered in our analysis. Unveiling and understanding such properties may help in future modeling of transport networks. Insets in Figures 4.7, 4.8 display positions of PTN in the geospace. It is the distribution of these positions that will be of interest in this section.

4.4.1   Mass fractals: critical radius for homogeneous service

The notion of fractal dimension is often used to quantify development and growth of cities and of their communication and transportation systems. City growth has been shown to exhibit self-similar behaviour, an observation that might imply a universality of processes that drive city agglomeration and clustering [93], [94]. Moreover, several physical growth processes such as, percolation or diffusion limited aggregation, are known to lead to such geometry have been exploited to explain city growth [93]–[96].

Correspondingly, transportation lines of some cities have been shown to behave as fractal structures too. Usually, to this end one measures the number of stations $N(R)$ that are located inside a circle of radius R. If a scaling of,

$$N(R) \sim R^{d_f} , \qquad (4.2)$$

is observed with non-integer value in the exponent d_f, the exponent is associated with the fractal dimension of the network. Indeed, if the stations in the PTN were equidistantly distributed along straight lines, this would correspond to the exponent $d_f = 1$.

Likewise, constant station density (number of station per unit area) would lead to $d_f = 2$. The two cases described above, correspond to one- and two-dimensional systems. However, the real-world transportation networks were reported to be characterised by different results, as discussed in section 4.1.
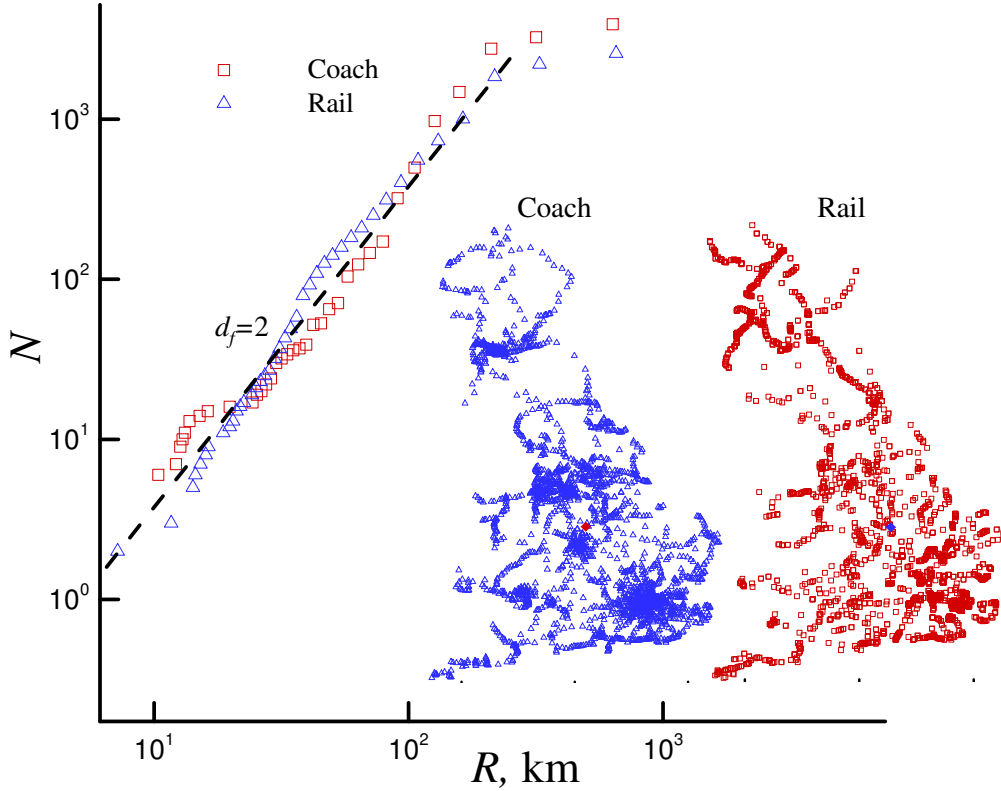
FIGURE 4.7: Number of stations $N(R)$ in the circle of radius R for the UK national coach and train networks. The fit of the dependence to the straight line on the log-log plot brings about the exponent close to $d_f = 2$ (dashed line). The inset shows the networks in geospace.

In order to determine the mass fractals we first find the centre of mass for a PTN (see Eq.(3.3)). Then moving out in concentric circles of radius R, the number of stations as a function of the radius $N(R)$ is calculated, see Eq.(4.2). For national networks $1 \text{ km} \leq R \leq 600 \text{ km}$, whereas for local networks $100 \text{ km} \leq R \leq 100 \text{ km}$. The outcome of this analysis is plotted in Figures 4.7 and 4.8 for both the national and local PTNs respectively.

One can see from Figure 4.7, that the fractal dimension of national networks, in the range of distances $1 \text{ km} \leq R \leq 200 \text{ km}$ is close to $d_f = 2$. Thus indicating that these networks tend to uniformly cover this region they are servicing. The outcome for the same analysis for local PTNs is shown in Figure 4.8. These networks tend to also uniformly
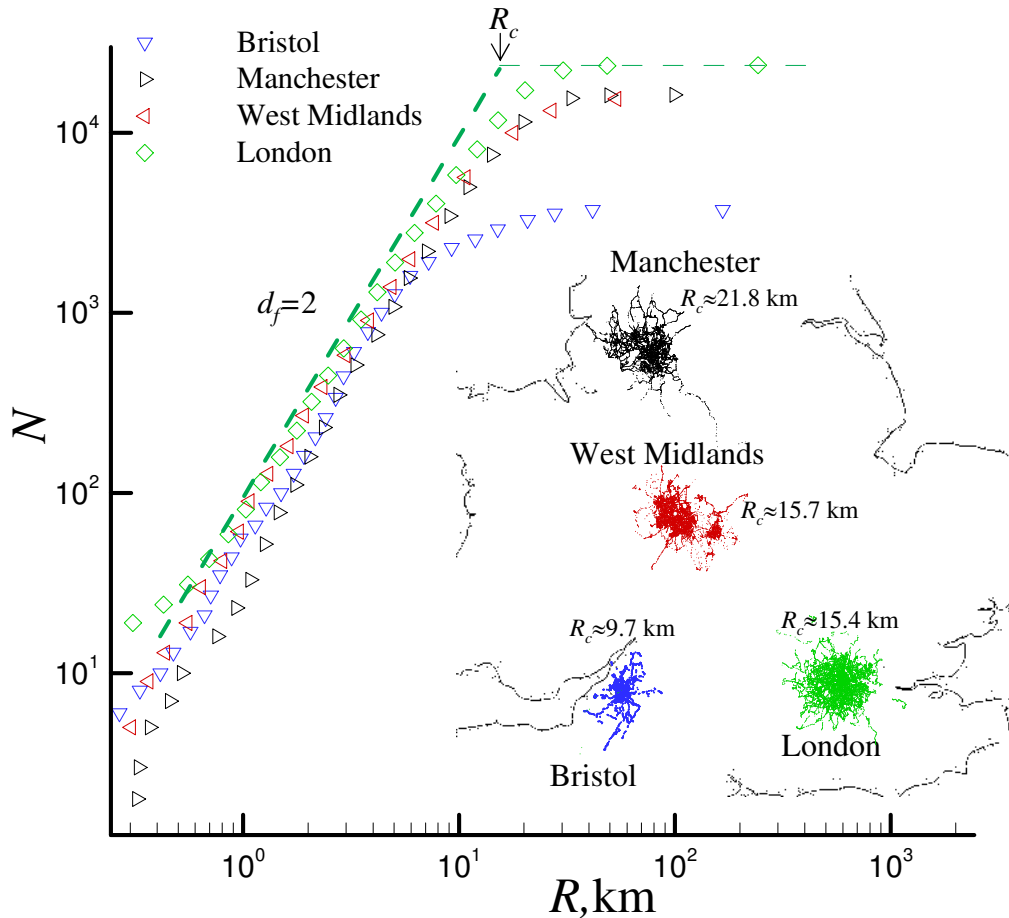
FIGURE 4.8: Number of stations $N(R)$ in the circle of radius R for the PTNs of Bristol, Greater Manchester, West Midlands, and Greater London. The bold dashed line (green online) corresponds to the exponent $d_f = 2$. The inset shows the networks in geospace, together with a part of the UK coastline. The radius R_c corresponds to the transition from the compact central area to the rarefied space with $d_f < 2$. The value of R_c is shown for each network.

cover the central area with what we call a critical radius R_c, see Figure 4.9, where R_c for the London PTN is shown. The radius R_c corresponds to the transition from a compact central area to the rarefied space with $d_f < 2$. In the inset of Figure 4.8, the value of R_c is shown for each local network. This transition can be interpreted as the point at which a PTN ceases to provide uniform access to commuters using public transport. This is an important consideration when modeling PTN. If R_c, is too large, this would be a waste of resources and this would scale rather penally as R^2. Alternatively, if R_c, is too small, the PTN would be neglecting many citizens living on the periphery of the city. Either of the two cases mentioned above are important to avoid when modeling PTN.
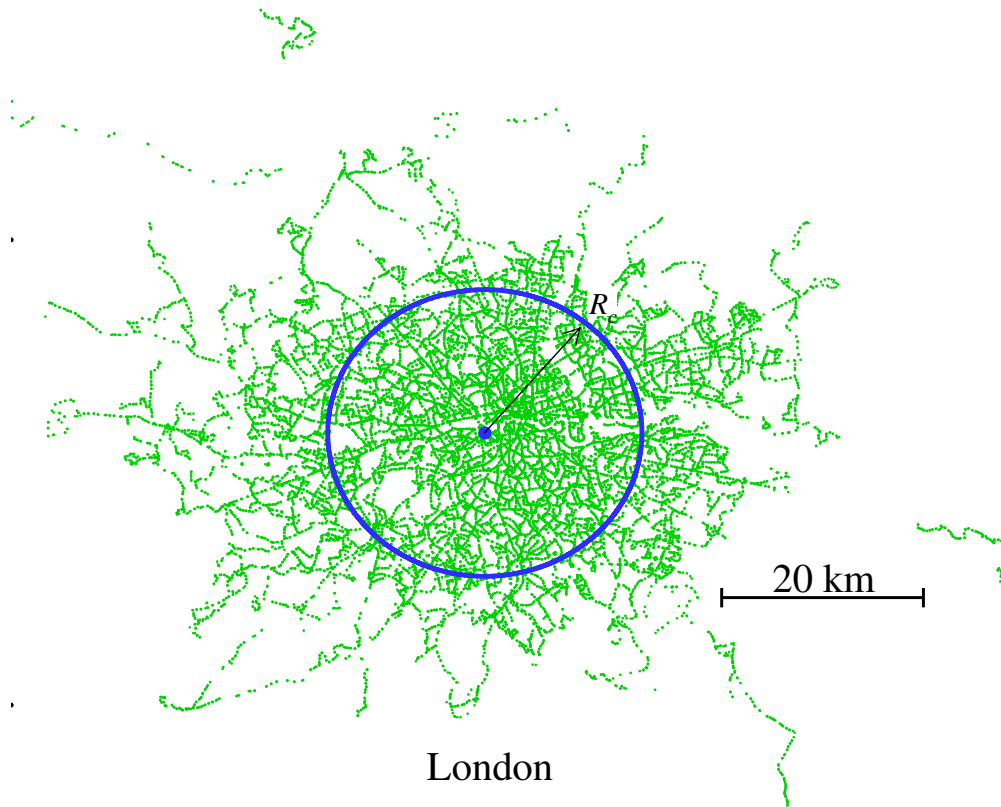
FIGURE 4.9: Example of the PTN of Greater London. The radius R_c ($\simeq 15.4$ km) corresponds to the transition from the compact central area to the rarefied space with $d_f < 2$.

4.4.2 Surface fractals: serviceable area of stations

In this section, we consider an alternative method to investigate the fractal properties of a PTN. Now we will calculate its fractal dimension by considering a box counting method where circles of different radius's are used to cover the object of interest, see inset in Figure 4.10. Obviously, the fractal dimensionality d_f^s calculated within this method depends on the size of the circles r_s, used to cover the object. As one can see in Figure 4.10, the fractal dimensionality changes from $d_f^s \simeq 1.28$ to $d_f^s = 1.91$ as r_c is increased. An interesting interpretation of the fractal dimensionality, as determined by this method, can be achieved by considering the size of a box as an area serviced by separate public transportation stations. When boxes are small one ends up with the structure where $d_f^s < 2$: effectively, the service area of all network is smaller than the dimensionality of the geospace $d = 2$. In turn, increasing the service area of each station (i.e. increasing of the box size) leads to an increase of d_f^s, finally leading to a $d_f^s \simeq 2$. The derivative of this function determines the rate at which d_f reaches its asymptote.

This may offer some insight into how evenly distributed PTN stations are across its serviceable area. This is an important consideration as more evenly distributed PTNs can provide fairer access to cities commuters, when should be taken into account when planning and modeling public transport.
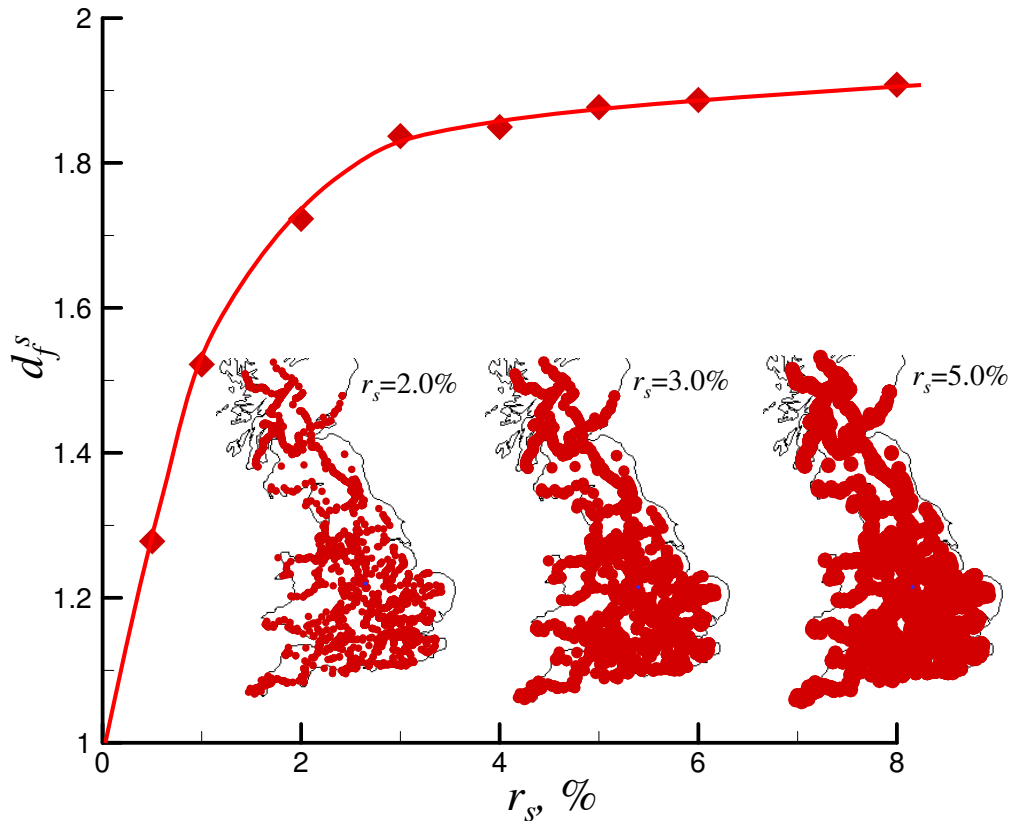
FIGURE 4.10: Fractal dimension d_f^s of the UK coach network calculated by considering a boxing method where circles of different size r_s (% of maximum size of the object) are used to cover the object of interest. Depending on the boxes size, one gets the value that spans the region $d_f^s = 1.28 - 1.91$. The inset shows examples of the networks in geospace at different values of r_s.

4.5 Conclusions

A feature of this analysis is that although the samples chosen included both local and national public transport networks, we show that they share a lot of common properties. The main topological features are summarized in Table 4.1. Comparison of data for PTNs with that of a classical random graph of the same size gives significant evidence that these networks are strongly correlated assortative structures with comparatively

small typical mean shortest path length (although caution is to be made when attributing to them small world properties). Their node degree distributions are well described by the power law decay, which brings about their scale-free properties, at least for a certain range of node degree values k.

Analysis of PTN topological features also aids us to predict their behaviour under random failures. Table 4.2 shows the Molloy-Reed parameter for UK networks, which may serve as a measure of PTN stability in comparison with that for some other cities in the world. To the best of our knowledge, it has never been calculated so far for large-scale transportation networks. In this sense our data for the UK national rail and coach networks provide the first example of such calculations and we wait for their comparison with their counterparts for the networks covering larger geographic space in other regions of the world.

One of the corner stones of modern complexity science is making analogies between statistical properties of systems of interacting agents of different nature. In particular, to study the sensitivity of such systems to changes in their parameters (as in the mentioned above case random failures), to analyse emergent collective phenomena, to shed light on the origin of power laws that very often govern the statistics of such systems (for a recent review see Ref.[95] and references therein). These features are very often reflected in the application of concepts and methods borrowed from physics in the out-of-physical fields. Examples from our analysis are given by using concepts of fractal dimension to quantify PTN properties in geospace. We believe that further work in this direction will be useful for better understanding the complex structure of PTNs and its modeling.

_____

Chapter 5

Network analysis of the COSMOS galaxy field

In this chapter we investigate how network science can be applied in the analysis of the large scale structures in the universe. Here, we search for correlations between metrics in network science and astrophysical properties. We further investigate the possibility of applying network science to define important structures that can be observed in the cosmic web: filaments, clusters and backbones.

In comparison with already existing methods developed for Cosmic Web analysis, network analysis has a number of potential benefits: (a) it is not built on some ad-hoc assumptions on the nature of the data, e.g. the existence of a continuous density field; (b) it's computationally effective in treating discrete data, as no density estimator or Hessian is computed; (c) it is capable of describing and quantifying the content of data at an adjustable level of detail and complexity; (d) it's equally applicable to the results of simulations and real observational data, thus allowing for direct comparison between them; (e) it can go beyond the classification of environments as clusters or filaments, by providing a more holistic view on the topology of the multi-scale phenomenon of the Cosmic Web. Thus, network analysis can complement other methods and effectively integrate them into a framework capable of investigating the complexity of large-scale structures of the Universe.

The research presented here allows one to approach probably the most important problem in cosmology, the mapping of the observable distribution of luminous matter to the

underlying dark matter distribution, sometimes called the problem of biasing. The results here are derived from real-world observational data, so they are not just a description of the spatial structure, they encode information of extremely complex processes of star formation, gas and radiation transfer in different environments. So, our findings on the common behaviour in the evolution of stand-alone galaxies and cliques bring important confirmation for the Cosmic Web Detachment model Argon2016, identifying the events of detachment in real observations.

Moreover, the application of complex networks to the Cosmic Web analysis places the research into a more general context of complex systems thus creating opportunities to search for analogies between different phenomena that occur in systems of interacting agents of various nature.

The main results of this study include the tentative classification of filaments, clusters and backbones within the cosmic web, with the use of network science. Another notable result is the identification of correlation found between the topological metric of clustering coefficient and the astrophysical metrics of stellar mass and colour index.

The chapter is laid out in the following way. Initially, in section 5.1 a brief description of the problem is given. In section 5.2 the origins of the data are discussed. Section 5.3 discusses the methods employed to construct the network of the comic web. Finally, in sections 5.4 and 5.5 the results of analysis are given and discussed in section 5.6.

5.1 Cosmic web: large-scale structure of the universe viewed as a complex network

The large scale structure of the Universe is generally explored by analysing the observations of spatial distributions and motions of galaxies. It appears to be rich in a variety of shapes and topological features. We can identify structures such as voids, walls, clusters and filaments. Voids represent large areas of empty space, typically between 10-100 mega parsecs (1 parsec $\sim$ 3.3 light years) in diameter, where either no or only a few galaxies exist. Whereas, walls are regions in space that contain a typical mean density of matter abundance and can be split into two sub-structures: filaments

and clusters. Here, filaments are branch-like structures, often tens of mega parsecs in length, which join clusters, which are high galaxy density areas, together. Altogether, they comprise the cosmic web, a term coined in Ref.[97], to be hereafter treated as a graph or complex network, see Refs.[1], [24], [26], [98] and references therein.

Numerous approaches have been devised in an attempt to properly describe and analyse the geometry and topology of large scale structures of the universe, see for example the list of recent studies [99]–[110]. Methods and approaches of complex network science have recently entered cosmology [103], [111], [112]. In astrophysics, complex network methods are believed to assist in solving various open problems. These are related to the evolution of galaxies, the geometry and topology of large scale structures and the nature and properties of dark matter and dark energy. This is performed by promoting a better understanding of the formation of the phase-space distribution of dark and luminous matter, and its impact on environments of galaxies [113], [114].

Here, we aim to study the cosmic web with the aid of complex network theory. First, by estimating the global and local properties of galaxy samples. Then investigating for correlations between observable astrophysical properties of galaxies and the local and global features of the cosmic web. We follow the pioneering research of Hong and Dey [103], where three network measures of topological importance (degree centrality, closeness centrality and betweenness centrality) have been derived from the COSMOS (COSMOlogical evolution Survey) catalogue by Ilbert et al. [115], to estimate the relationship of galaxy parameters to their topological environment. Hong and Dey [103], in turn, follow Scoville et al. [116], where the same problem was addressed using "traditional" methods with the same data source.

We extend this analysis by investigating how the above mentioned properties are evolving, by considering neighbouring redshift values. We further complement these investigations, by introducing other network metrics of interest like the number of edges, mean node degree, size of the giant connected component, clustering coefficient, assortativity, average path length and diameter. Moreover, we assess the applicability, restrictions and accuracy of network analysis with respect to real observational data.

5.2 COSMOS catalogue

The dataset is based on the catalogue built by Ilbert et al. [115] for studying the mass assembly of galaxies in the COSMOS field. This data has been used in previous studies [103], [116]. Thus, this data set provides a benchmark for different kinds of large scale structure analysis.

The COSMOS Collaboration is a huge astronomical effort to integrate the data produced by a variety of space and ground-based telescopes. The survey is aimed at analysing galaxy evolution and is designed to collect essentially all possible objects in the field of view, so it is meant to be as deep as possible, meanwhile covering an area of celestial sphere large enough to mitigate for the influence of cosmic variance. For this reason the volume surveyed is restricted to a box region of $1\overset{\circ}{.}0$ in height and width, which corresponds to ~ 54 mega parsecs in co-moving space. This region of sky is given by the coordinates: right ascension (R.A.) $149\overset{\circ}{.}4 - 159\overset{\circ}{.}4$ and declination (Decl.) $1\overset{\circ}{.}7 - 2\overset{\circ}{.}7$. While the height and width of this box is measured in degrees. The depth is measured in redshift values z. Here, z is defined by

$$1 + z = \frac{a_{\mathrm{obsv}}}{a_{\mathrm{emit}}} \, , \tag{5.1}$$

where a_{emit} is the wavelength of the light emitted by an object and a_{obsv} is the observed wavelength. Larger values of z represent a deeper measurement within this box which relates to an earlier time in the age of the universe.

The catalogue is selected using the UltraVISTA ultra-deep near-infrared survey, data release one [117]. It includes directly observable quantities, such as celestial coordinates for galaxies and photometric magnitudes for a number of broad bands. This data is complemented by quantities that are indirectly estimated from photometric data, as described in Ref.[115]. It also includes the redshifts for galaxies (colour corrected for dust extinction), galaxy classifications according to colour (quiescent or star-forming), stellar mass and other physical galaxy parameters estimated as best-of-fit to some model.

To investigate how the structure is changing in time and to estimate the robustness of

the method, our analysis includes three galaxy slices at neighbouring ranges of redshifts, $0.88 \leq z \leq 0.91$, $0.91 \leq z \leq 0.94$ and $0.94 \leq z \leq 0.97$, hereafter named z_1, z_2 and z_3 respectively. By this choice we extend the data analysed by Hong and Dey [103] for redshift z_2 to include neighbouring redshift values z_1 and z_3. Note that the slices sample completely different galaxies and do not reflect the evolution of the same galaxies in time. The redshift ranges for the slices are chosen because there are a similar number of galaxies in each sample, which provide a dense enough populations of the various types of galaxies, including a large fraction of early-type (red) galaxies. Moreover, the central slice recreates the one used in Ref.[116], where it was shown that when $z > 1$, the relation of galaxy properties within a local environment abruptly diminishes.

The elaborated analysis of multi-band photometry data estimates the redshifts of galaxies with high accuracy (at 1% level), but it is still not accurate enough to treat as a full 3D picture of galaxy spatial distribution. The errors in z are comparable to the slices of redshift z_1, z_2 and z_3 and so we analyse these slices in 2D.

5.3 Constructing the cosmic web

Contrary to the data coming from computer science, industrial databases and social networks, data in cosmology is inherently non-networked and contains a substantial amount of noise. Hence, a network must be constructed from the data set (catalogue) using appropriate criteria and methodology. This is equivalent to the transformation of data from an unstructured representation to a structured network representation (nodes and edges). As with any transformation, this one can not always be performed in a lossless manner.

The task is to encode as much information of interest as possible, in this case the existence of structure over a random distribution of galaxies. This aspect of the problem resembles the studies of clustering, commonly found in cosmology. There is no universal technique to construct the network from the data for any situation, however the major steps to consider are the following:

1. Capture similarity between data points.

2. Adopt some rules based on a similarity function for establishing the links between data points.

3. Implement some criteria to judge whether the network is properly built, analogous to a "goodness-of-fit" approximation procedure.

Initially, Hong and Dey in Ref.[103] introduced a similarity parameter called "linking length" which they used to generate the cosmic network. With this method one starts with a network of N galaxy-nodes with no links (i.e. no edges). Then by considering the spatial position of galaxies-nodes an undirected network is generated by adding edges between galaxy-nodes, if and only if, the euclidean distance between two galaxy-nodes is less or equal to the prescribed linking-length.

As mentioned, the aim here is to generate a network whereby the networks topological features can classify some of the structural and astrophysical features of the cosmic web. If the linking length l is to small, no edges will be generated leaving the network as a set of disconnected nodes. Whereas, if l is to large, the network will tend towards a complete graph. In both cases this will result in no useful information being encoded into the topology of the network. So one has to vary l until useful information can be encoded into the topology of the network. For example in sparser regions of space a larger linking length would be required to generate networks with enough structure to extract useful information. Such a method appears to be robust as Hong and Dey generate a network where its properties can classify various structural features of the cosmic web.

In a subsequent study various other similarity parameters for constructing complex networks from galaxy survey data were discussed [111]. Here they considered similarity parameters such as:

- Linking length: as describe above,

- Nearest neighbour: galaxies are linked to their nearest n neighbours,

- Mass: linking length varies for each galaxies according to its mass,

- Relative velocity: linking length are related to a galaxies relative velocity to other galaxies.

In this study it was concluded that the linking length criterion seemed to be the most relevant method for galaxy property studies. This simple recipe for analysing clustering was used for decades as "top-hat filtering" [118] and is closely related to the "friend-of-friend" algorithm [119], which is used in cosmology for the study of the large-scale structure of the Universe. It is however not universal, as different samples may require a different linking length to generate meaningful statistics. Thus, the choice of linking length l, especially when comparing samples, is crucial and requires some measure of consideration to ensure unbiased sampling.

In our analysis we extend Hong and Dey's network generation approach by determining the critical point at which the complex network "equilibrates" in terms of topological structure. This can be done by considering clustering coefficient, as defined in subsection 2.2.5 by Eq.(2.13), as this quantity essentially gives a measure of how correlated a network is at local scale. If plotted as a function of linking length clustering coefficient may offer some insight into the best choice of l. In Figure 5.1 we plot C as a function of l for all redshift slices. Observing Figure 5.1, it can seen that as l is increased and at some value of l the clustering coefficients (local strucutre) equilibrates and stabilises. It is at this point that the network should offer sufficient topological structure to encode useful information on the astrophysical properties of the cosmic web. For these samples we can see that see clustering coefficient stabilises at $l \approx 0\overset{\circ}{.}0216$, the value chosen to construct the cosmic web (see Figure 5.2), this corresponds to $C_1 = 0.604$, $C_2 = 0.612$, $C_3 = 0.603$ for z_1, z_2 and z_3 accordingly. It is at this value of l, C for all networks stabilise. This is approach allows for a more systematic and unbiased approach to determining linking length. This supports what appears to be an optimal value of l as shown first in Ref.[103] and further confirmed here.

So, hereafter a fixed linking length is predefined to be equal to $0\overset{\circ}{.}0216$, this corresponds to a linear scale of 1.2 Mega parsecs in standard ΛCDM cosmology (also known as the standard model of Big Bang cosmology). This value was derived by Hong and Dey [103] by considering Poissonian node degree distributions.

To summarise, we generate three neighbouring cosmic networks over the different redshifts, using the same linking length as the clustering coefficient for all three samples stabilise at approximately the same value of l. This is done to have consistency between
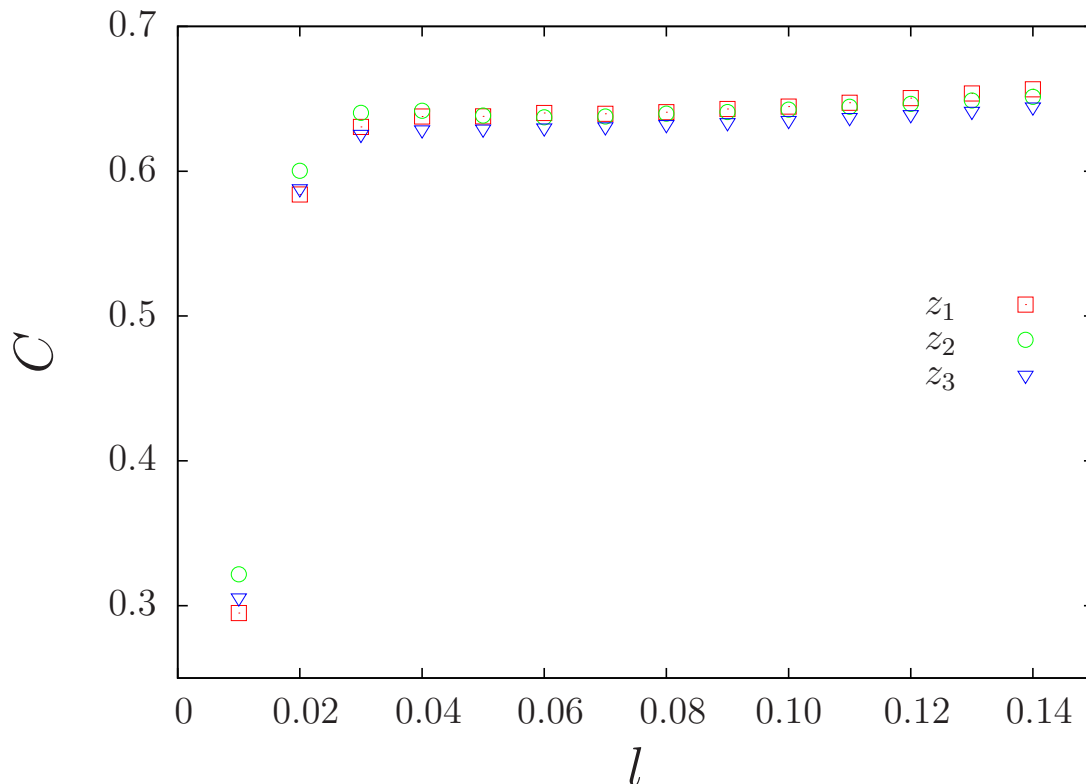
FIGURE 5.1: Clustering coefficient C as a function of linking length l for the red shifts z_1, z_2 and z_3 .

samples and enable tracking of the evolution of galaxy populations over the time. In Figure 5.2 we show the cosmic networks generated using this prescribed linking length for each redshift.

5.4 General network properties

Our results for different network metrics are listed in the Table 5.1 for the three networks visualised in Figure 5.2. From a network perspective, we can confirm that we have a robust network generation method that creates a network providing sufficient structure, where relevant information can be obtained from the network metrics. To show this we can compare the average path length $\langle \ell \rangle$ and the average clustering coefficient C, defined in subsections 2.2.7 and 2.2.5, with their random counterparts $\langle \ell \rangle_r$ and C_r. When comparing these metrics one can see from Table 5.1, that the networks generated are similarly, highly correlated structures. Similar results can be seen when comparing $\langle \ell \rangle$ with $\langle \ell \rangle_r$. This provides evidence of some regular structure within the giant connected component GCC. These comparisons point to a method of network

FIGURE 5.2: Complex networks constructed on the base of the photometric redshift slices $0.88 \leq z \leq 0.91$, $0.91 \leq z \leq 0.94$, and $0.94 \leq z \leq 0.97$ (top to bottom) from the COSMOS field using linking length of $0\overset{\circ}{.}0216$ degrees. The middle figure recovers that formerly obtained in Ref.[116].

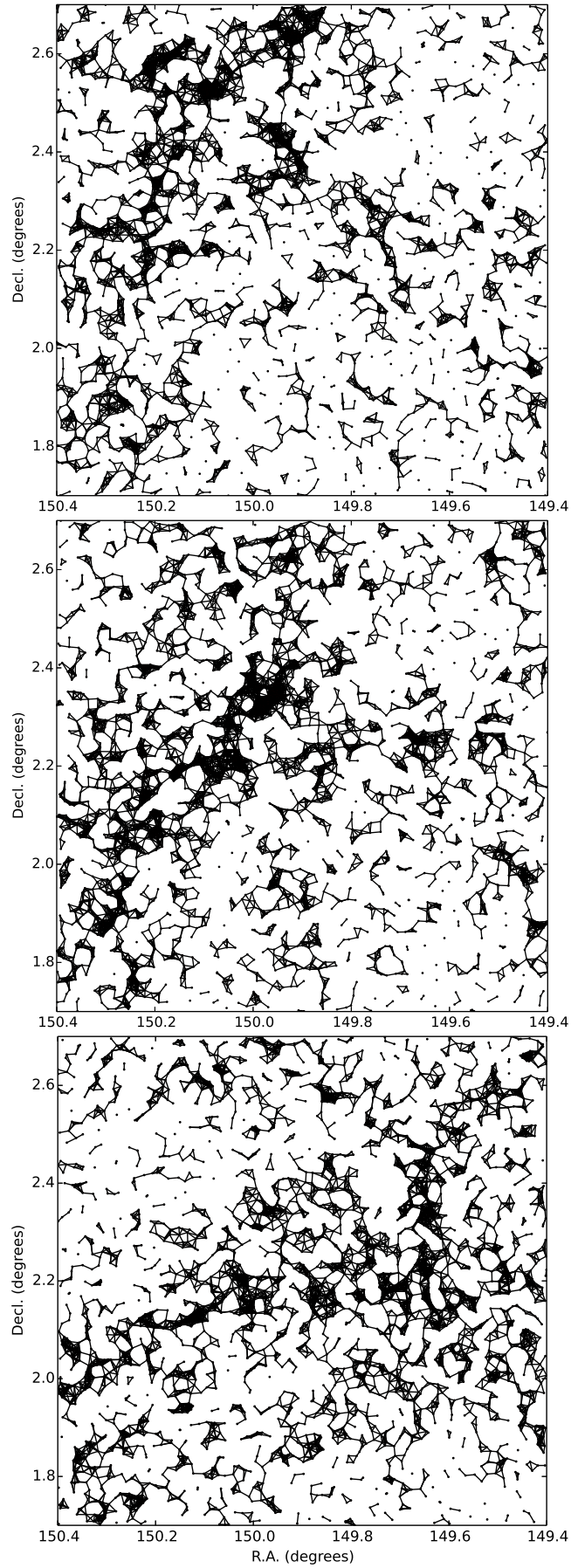|	$0.88 \leq z < 0.91$		$0.91 \leq z \leq 0.94$		$0.94 < z \leq 0.97$	
	Mean	$[1\sigma, \text{SE}]$	Mean	$[1\sigma, \text{SE}]$	Mean	$[1\sigma, \text{SE}]$
N	3318		3678		3606	
M	11747		14317		12206	
r	0.85		0.86		0.80	
D	116		113		117	
GCC	2079		2369		2828	
$\langle k \rangle$	7.08	[5.02, 0.087]	7.79	[5.68, 0.093]	6.77	[4.36, 0.071]
$\langle \ell \rangle$	37.5	[22.63, 0.011]	33.6	[19.00, 0.008]	39.8	[21.31, 0.008]
$\langle \ell_r \rangle$	3.06		3.00		3.12	
C	0.6040	[0.263, 0.0048]	0.6120	[0.261, 0.0043]	0.6030	[0.264, 0.0044]
C_r	0.0021		0.0021		0.0019	
C_{c1}	0.0019	[0.0001, 0.00003]	0.0028	[0.0018, 0.00003]	0.0018	[0.0013, 0.00005]
C_{c2}	0.0180	[0.0041, 0.00009]	0.0206	[0.0052, 0.00009]	0.0210	[0.0052, 0.00010]
C_d	0.0021	[0.0015, 0.00003]	0.0021	[0.0015, 0.00003]	0.0019	[0.0012, 0.00002]
C_b	0.0043	[0.0146, 0.00025]	0.0037	[0.0097, 0.00016]	0.0066	[0.0172, 0.00029]
$Colour_1$	0.64	[0.66, 0.012]	0.63	[0.68, 0.012]	0.61	[0.67, 0.012]
$Colour_2$	4.02	[0.54, 0.033]	4.20	[0.61, 0.032]	4.13	[0.66, 0.036]
$\log M_{stellar}$	9.29	[0.67, 0.012]	9.50	[0.69, 0.011]	9.44	[0.66, 0.011]

TABLE 5.1: Mean values and 1σ range of network characteristics of the COSMOS galaxy samples for three redshift slices. N is number of nodes, M is number of edges, $\langle k \rangle$ is mean node degree, GCC is number of nodes in the Giant Connected Component, C is mean clustering coefficient, $\langle \ell \rangle$ is mean shortest path of the network, r is assortativity, D is diameter (maximal shortest path length), C_d is Degree centrality, C_b is Betweenness centrality, C_{c1} is Closeness centrality for the distribution of fragment clusters, C_{c2} is Closeness centrality for the distribution of GCC, $\langle \ell_r \rangle$ is the mean shortest path and mean clustering coefficient of the random graph of corresponding size. $Colour_1$ and $Colour_2$ are mean colour indexes for both modes of bimodal distributions shown in the top panels of Fig. 5.7; $\log M_{stellar}$ is the logarithm of mean stellar mass (in units of solar one) of galaxies for the distributions shown in the bottom panels of Fig. 5.7. Information in brackets represents [Standard deviation, Standard error]. Where there are no brackets this indicates that we do not have a distribution to measure.

generation that is unbiased and yet able to provide sufficient structure to retrieve useful information about the distribution of these galaxies.

The GCC, is analogous to the largest cluster and the diameter D, is analogous to the spine or backbone, of this largest cluster. From Table 5.1 we see that all networks have slightly different GCCs, with similar backbones. This indicates a variance in the largest cluster size between samples, with z_3 having the largest cluster and z_2 the smallest.

We have computed the centrality measures for betweenness, closeness and degree that Hong and Dey [103] consider in their paper and estimated their standard errors for three galaxy samples z_1, z_2 and z_3. They are presented in Table 5.1 and their distributions are shown in Figure 5.3.
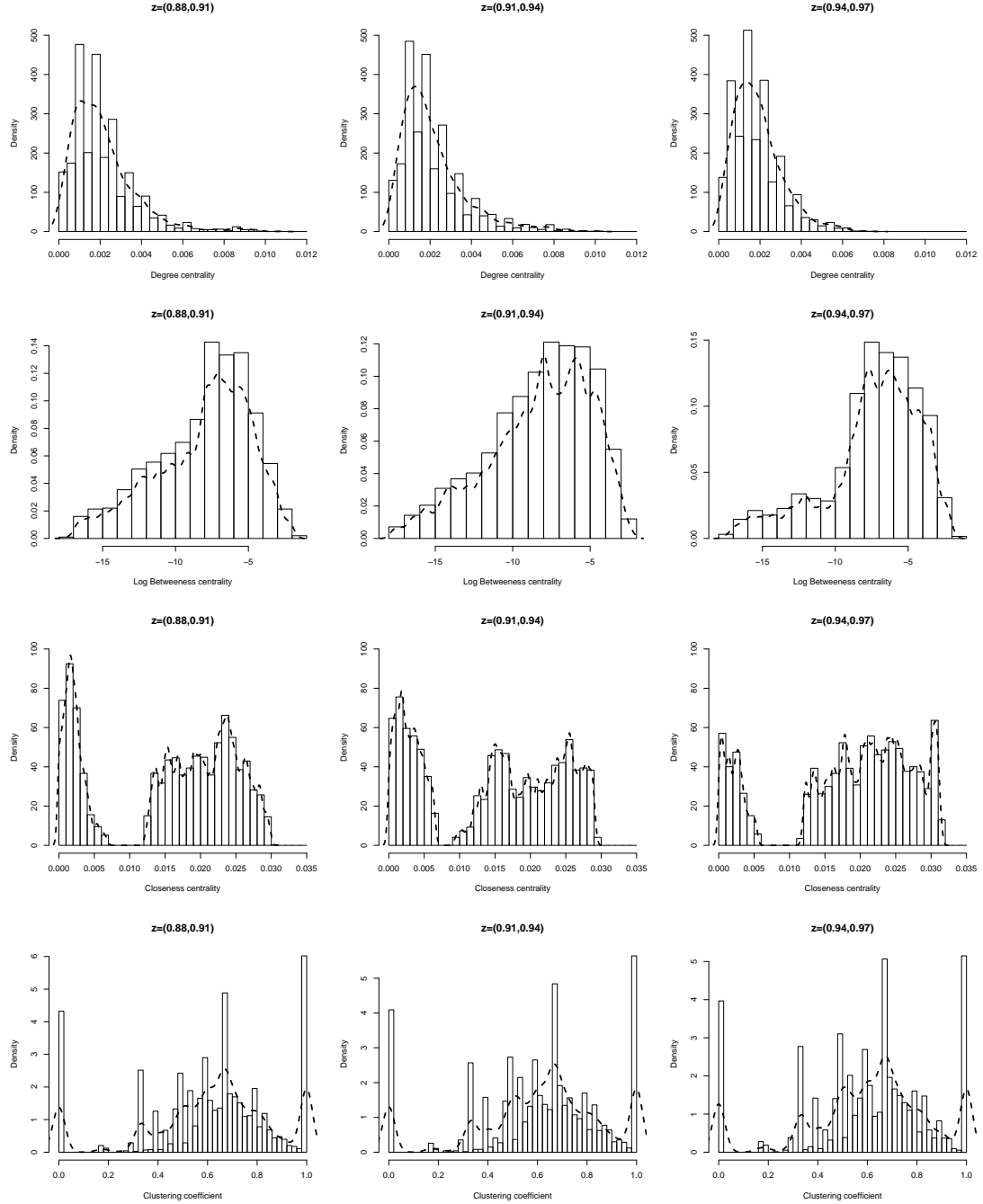
FIGURE 5.3: The distributions of the centralities for degree C_d, betweenness C_b, closeness C_c and clustering coefficient C (from top to bottom) for three redshift intervals $0.88 \leq z < 0.91$, $0.91 \leq z \leq 0.94$ and $0.94 \leq z \leq 0.97$.

5.4.1 Degree Centrality: Geometry of clusters

Degree centrality distribution $C_d(j)$, as defined in subsection 2.2.8, describes the node degree distribution. The mean of $C_d(j)$ similar to the average degree $\langle k \rangle$, defined by Eq.(2.1) and listed in Table 5.1 for each redshift. We can also see by inspecting Figure 5.3 (top row), the degree distribution for each redshift is Poissonian in nature, with z_1 and z_2 slices having more extended tails than z_3. An extended tail is an indication of high degree nodes which means that z_1 and z_2 have some tightly packed galaxies within clusters. Whereas in z_3, within the clusters the distances between galaxies are more evenly distributed.

5.4.2 Betweenness Centrality: Filaments

Betweenness centrality $C_b(j)$, as defined in subsection 2.2.8, measures the importance of a node in terms of communication between other nodes. A node that is involved in the most number of shortest paths will be the most important node according to this centrality (see Eq.(2.21)). By this definition nodes that join two clusters (object with a large number of galaxies) will have a high betweenness centrality. This is because many nodes will exist in each cluster, meaning that many paths will have to go through the nodes joining these clusters. This would not be the case however, if one of the clusters was significantly smaller than the other. Galaxies that display a high betweenness centrality are then galaxies that reside in these branch like structure joining clusters. Remembering that this is the very definition of a filament, this centrality offers a viable method from which to search for filamentary structures in the cosmic web. From Figure 5.3 (second row) we can see that the distribution is positively skewed in a log-log scale indicating only a few high betweenness galaxies. The surrounding area of these galaxies might be classified, in astrophysical terms, as filaments that link the larger segments of clusters together. For example, in Figure 5.4, we show the position of nodes representing galaxies with betweenness centrality greater then 0.02 by red squares and others by blue circles. We see that there are only a few nodes where $C_b > 0.02$ and they all tend to be galaxies that form paths between the larger segments within the GCC.
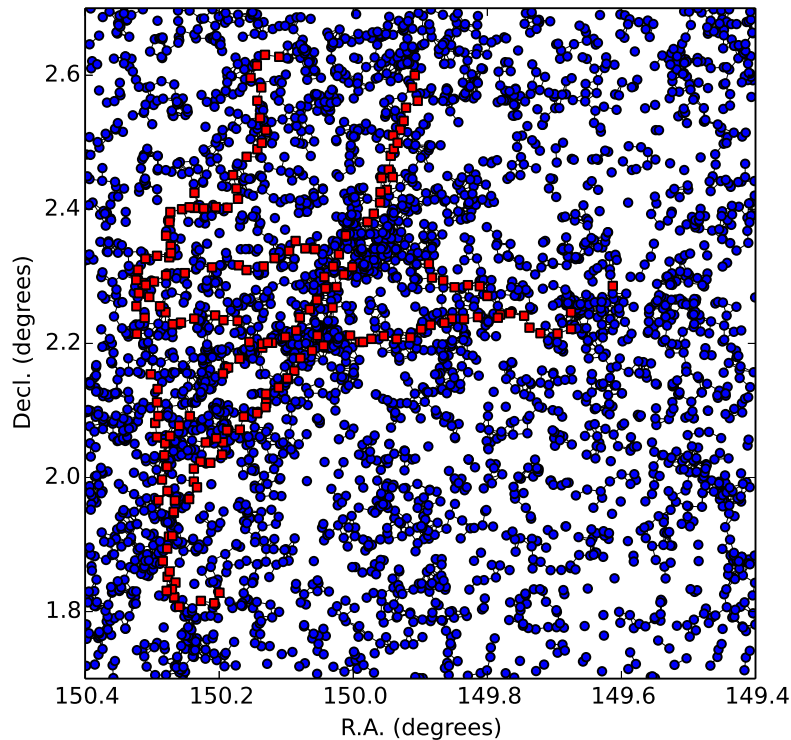
FIGURE 5.4: Galaxies in z_2-slice with betweenness centrality greater than 0.02 are red squares and galaxies with lower value are noted by blue circles.

5.4.3 Closeness Centrality: Cluster distributions

The closeness centrality $C_c(j)$, as defined in subsection 2.2.8, measures how close nodes are to each other in terms of path length ℓ. Its distribution is shown in Figure 5.3 (third row) for three different redshift values. As one can see from the figures, these distributions are bimodal, with two peaks centered about values ≈ 0.002 and ≈ 0.02. The more precise values together with intervals are given in Table 5.1. Note that the two peaks are characterised by different widths, in turn, leading to different variance of the distributions as given in Table 5.1. As it follows from a more thorough analysis of the data, the population of galaxies that belong to the second peak corresponds to the Giant connected component of the network (GCC). In turn, the nodes in the centre of the GCC are characterised by shorter distances to the rest of the nodes, leading by Eq.(2.22), to larger values of $C_c(j)$. The periphery nodes are characterised by larger distances to the rest of the nodes, therefore they have smaller values of $C_c(j)$. In a similar way, one

can identify the population of galaxies that give rise to the second peak in the closeness centrality distribution. These are the galaxies that belong to the smaller clusters, that are not attached to the GCC. Here, the central nodes of the clusters correspond to the right wing of the first peak and the periphery nodes are those, contributing to the left wing. The possibilities to find two distinct populations in the distribution is caused by the difference in sizes of the GCC and the other smaller sub components in the network. The larger the difference, the more distinct the peaks. Indeed, as one can see from Table 5.1, the largest size of GCC (78%), corresponds to the largest redshift value $0.94 < z \leq 0.97$. It is in this case that the gap between two peaks is most pronounced.

5.4.4 Clustering Coefficient: Galaxy selections

The clustering coefficient, as defined in section 2.2.5 by Eq.(2.12), is a useful index to measure the correlation on a local scale. It provides information about the elementary substructures (patterns) that exist in a network. The high values of clustering coefficient obtained for all redshifts ($C_1 = 0.604$, $C_2 = 0.612$, $C_3 = 0.603$, see Table 5.1) brings about the presence of real pattern-groups of tightly connected galaxies on different sites, as can be seen from Figure 5.2.

The local clustering of each node can also be considered in an effort to help elaborate a more robust method of defining structures in the cosmic web. From the histograms of clustering coefficient in Figure 5.3 it can be seen that it has the complex discrete distribution with three main peaks at 0, 0.66 and 1. In most cases, galaxies with a clustering coefficient less than 0.1 will have less then two neighbours. This intuitively indicates that these galaxies are most likely located in some sparse region of space. There, the mean distance between galaxies is larger than the prescribed linking length, and this selection can be called "stand-alone" galaxies. The nodes with clustering coefficient ranging between $0.1 - 0.9$, describe galaxies that are tightly packed next to one another. Galaxies, with a clustering coefficient larger than 0.9 tend to highlight small clusters in these large voids, or in other words participate in some "cliques".

In Figure 5.5 we show the selections of galaxies according to the clustering coefficient on the spatial distributions of galaxies, for three redshift slices.
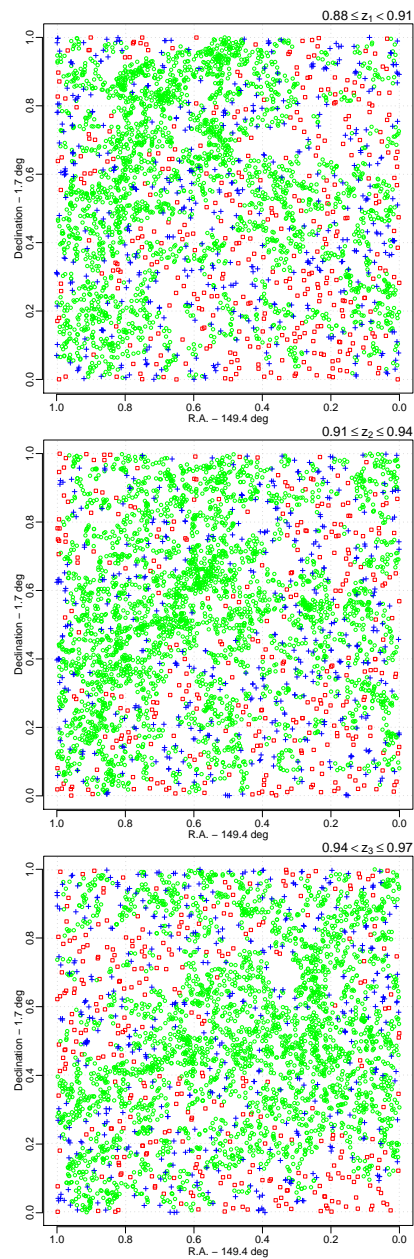
FIGURE 5.5: The spatial coordinates of galaxies from different selections according to clustering coefficient. Red squares denote the "stand-alone" galaxies, green circles denote the galaxies with interim values of clustering coefficient, and small "cliques" are denoted by blue crosses.

The reason for nodes within denser clusters not exhibiting higher clustering than their sparser counterparts is mainly due to the fixed linking length. In these large clusters, node i will link to all nodes within the prescribed linking length, including for example node j on the edges of linking length. However, due to linking length being smaller than cluster of galaxies, node j will link to other nodes in this cluster that are outside the range of node i. Meaning not all of its neighbour will be linked to node i. Hence the reason, counter intuitively, smaller clustering coefficients are found among highly clustered galaxies rather than in the smaller groups of galaxies found in voids.

5.4.5 Average path length: Large world

When calculating $\langle \ell \rangle$, as defined in section 2.2.7 by Eq.(2.18), in cosmic network the GCC can only be used. This is because disconnected nodes will have no path and hence an infinite path which cannot be calculated. From Table 5.1 we can see that $\langle \ell \rangle$ for all redshifts range between 33 and 40. This can then be compared with the $\langle \ell_r \rangle$, defined in section 2.2.7 by Eq.(2.19), of a random network of the same size. In network theory the idea of small worldedness, as discussed in section 2.2.9, has received significant amounts of attention [22]. The cosmic network, which we analyse, does not display the small world characteristics so often seen in complex networks. Here, the first condition, where $C \gg C_r$ holds. However, on the second, where $\langle \ell \rangle > \langle \ell_r \rangle$, the cosmic web fails. Thus, the cosmic network in network science can be considered a "large world", in this context. This is almost certainly as a result of the geometric constraint imposed by linking length, as it restricts galaxies outside a certain range from forming edges.

5.4.6 Assortative nature of the cosmic web

For a disassortative network the value of r, as described by Eq.(2.17) in subsection 2.2.6, would be negative. Thus indicating that nodes of low degree tend to associate with nodes of high degree. In turn, when this value is positive this indicates an assortative network where nodes of similar degree link with one another. Figure 5.6 provides a qualitative perspective, where it can be clearly seen that the cosmic network displays positive correlation and this can be further confirmed quantitatively in Table 5.1 with r
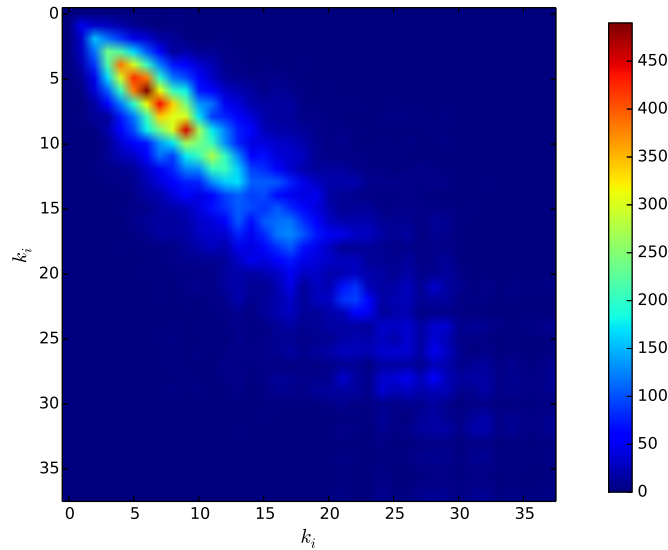
FIGURE 5.6: Node degree correlations for z_2 redshift slice: If a node of degree 1 is linked to a node of degree 10 the data point plotted will be (1,10). The Colour map indicates the amount a particular combination occurs. It can be clearly seen that the network displays positive correlation: nodes of similar degrees tend to be linked to each other. This is further confirmed quantitatively with assortativity r being 0.86.

for all redshifts being ≥ 0.80. This indicates that in the cosmic network, galaxies with a similar number of links tend to be connected to one another.

5.5 Astrophysical quantities vs Topology

The principal goal of this chapter is to search for relationships that exist between the topological environment generated by the cosmic web and the astrophysical properties of galaxies. For this we consider the topological selections of clustering coefficient and the astrophysical properties of colour index and stellar mass. Colour index measures the colour of a galaxy. From this measure the temperature of galaxies can be derived, where a smaller value of colour index represents a hotter galaxy. Stellar mass gives the mass of a galaxy which is usually normalised according to the mass of the sun.

When searching for these correlations there are a number of limitations to consider. One fundamental limitation is due to the non-Gaussian nature of the galaxy parameter distributions, studied here. This invalidates classical statistical methods of correlation analysis such as: Pearson coefficient, t-tests, regression analysis and ANOVA.
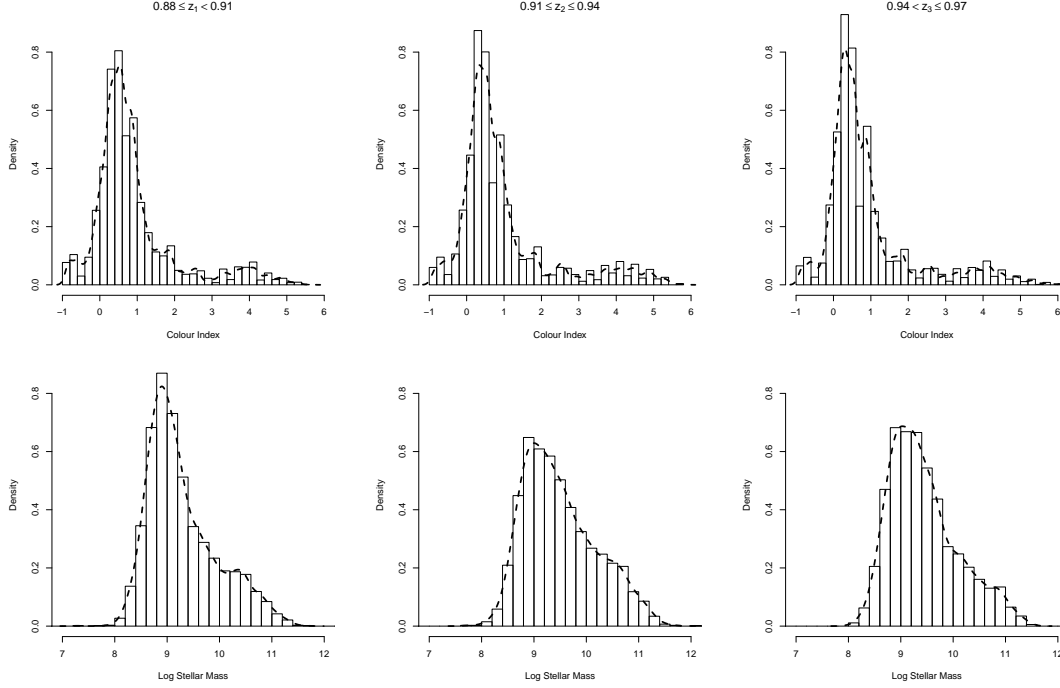
FIGURE 5.7: The statistical distributions of galaxy parameters for three ranges of z: $0.88 \leq z_1 < 0.91$, $0.91 \leq z_2 \leq 0.94$, and $0.94 < z_3 \leq 0.97$.

The framework of null hypothesis significance testing, however, allows us to run non-parametric two-sample tests, to evaluate, with a certain degree of confidence, whether the distributions are similar or not. In this case, the null hypothesis assumes both samples are drawn from the same distribution.

Usually, the Kolmogorov-Smirnov test [120], [121] is used as a non-parametric test, as in Ref.[103]. Although this test is widely applied, it can have problems with complex distributions and thus should be cross-validated by other approaches, like the Anderson-Darling [122] or Mann-Whitney-Wilcox [123], [124] tests.

5.5.1 Distributions of galaxy parameters

First we analyse the distributions of colour and stellar masses over the redshift slices, as seen in Figure 5.7. The means and standard deviations of these distributions are included in Table 5.1. Applying the Hartigans' dip test [125], we have found that the bimodality of the colour index distributions are statistically significant i.e. the null hypothesis of unimodality is rejected with p-value $\ll 0.05$.

For colour index, the hypothesis of samples, originating from a common distribution is strongly rejected when comparing z_1 and z_3 subsamples, mildly rejected for z_1 and z_2 subsamples, and mildly accepted for z_2 and z_3 subsamples. Therefore, the tests clearly indicate a weak yet still significant evolutionary trend for colour index, over the redshift span.

The distribution of stellar mass indicate the strong and clear evolutionary trend over redshift values, albeit more mild for z_2 and z_3. Note that colour index is derived directly from observed photometric measurements, meanwhile the stellar mass of galaxies is computed from the same photometric data with a number of approximations and elaborate modeling of spectral energy distributions of galaxies.

5.5.2 Selections by clustering coefficient

To find the correlations between the clustering coefficient and colour index or stellar mass we should compare the probability distributions of galaxy properties for different samples, drawn according to their clustering coefficient values. These probability distributions are visually inspected by building histograms, whereas the estimations and significance testing are usually performed by analysing the empirical distribution function.

Here, we split the galaxy samples into three selections according to local clustering coefficient: selection I (stand-alone galaxies) $C = 0$; selection II (intermediately packed galaxies) $0 < C < 1$; selection III (compact cliques of galaxies) $C = 1$. We then run two-sample tests over the different selections, within each redshift slice.

The cumulative distributions of colour index and stellar mass for different populations according to local clustering coefficient I (red squares), II (green circles) and III (blue crosses), for redshift sample z_2 are shown in Figure 5.8. The colour index and stellar mass distributions for I and III selections are similar for all z-slices. Qualitatively, the distribution for stellar mass for selection II differs from selection I and III for all z-slices, while the distribution of colour index differs only for the z_2-slice.
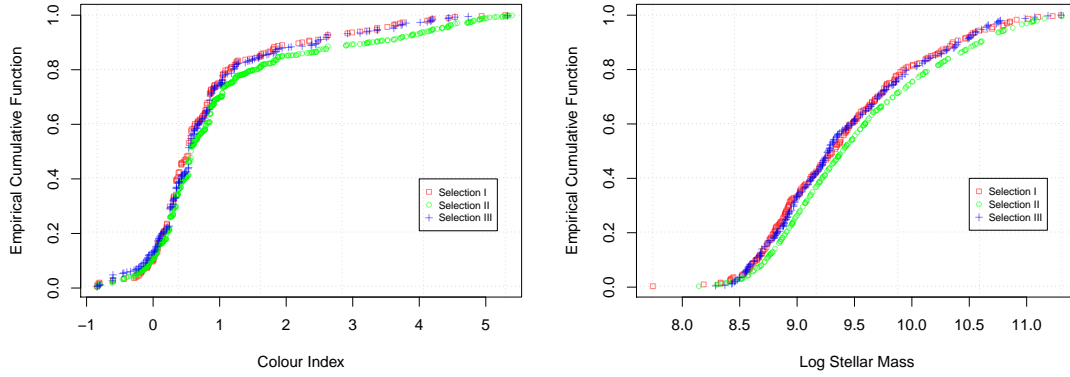
FIGURE 5.8: The cumulative distributions of colour index (left) and stellar mass (right) for different populations according to local clustering coefficient, coloured in red, green and blue for clustering coefficients selections I, II and III respectively for redshift sample z_2.

Quantitatively, these distributions were tested to determine if they, do indeed, differ from each other by applying a non-parametric Anderson-Darling test. In Table 5.2, results of these tests are presented. Here, a significance value of 0.05 is chosen. With this, the following conclusions can be deduced: the selections I and III (stand-alone and densely packed in small-groups) are coming from non-distinguishable samples; the clearest distinction appears to be between selections I and II (stand-alone and with interim clustering coefficient); colour distributions are not the same, only for cases associated with the z_2 slice; the z_3 slice reveals different behaviour in comparison with other slices. The smallness of the evolutionary effects is understandable since the age differences of the nearest and farthest sample of galaxies, do not exceed 400 million years.

	Colour			Stellar Mass		
	z_1	z_2	z_3	z_1	z_2	z_3
I vs II	0.062	0.018	0.31	$5 \cdot 10^{-6}$	0.0048	0.00023
II vs III	0.29	0.025	0.37	0.0032	0.0018	0.014
I vs III	0.74	0.79	0.49	0.19	0.91	0.18

TABLE 5.2: The results of Anderson-Darling tests (p-value) for colour and stellar mass distributions for selections: I $C = 0$; II $0 < C < 1$; III $C = 1$.

We have to bear in mind the caution already expressed in section 5.2. One of the limitations of the database is that it does not allow, high enough precision of coordinates for galaxies in 3D. Indeed, the 2D slices of the real-world pictures result from the projection of their 3D counterparts. According to Scoville et al. [116], the binning matched the accuracy of the redshifts, thus providing optimal signal-to-noise ratio. For the density estimation, the 2D projections are linearly related to a 3D volume. Whereas, for the topological environment, this might not be the case. Despite this obvious limitation, one can still retrieve information on the correlations, we are interested in.

5.6 Conclusions

In this chapter we have presented the complex network analysis of some part of the cosmic web. For this purpose, we have used the data from the COSMOS catalogue of galaxies [115]. A distinct feature of this study is that we analyse galaxy samples in the same region $1° \times 1°$ of the celestial sphere with coordinates R.A.=$149°.4 \div 150°.4$ and Decl.=$1°.7 \div 2°.7$ for three neighbouring redshift intervals $0.88 \leq z < 0.91$, $0.91 \leq z \leq 0.94$ and $0.94 \leq z \leq 0.97$, marked by z_1, z_2 and z_3 accordingly. Where networks are generated using a linking length of $0°.0216$, proposed by Hong and Dey [103].

We have calculated the local complex network measures of degree centrality $C_d(j)$, closeness centrality $C_c(j)$, betweenness centrality $C_b(j)$, clustering coefficient $C(j)$, node degree $k(j)$ as well as the global ones (average path length $\langle \ell \rangle$, diameter D, average clustering coefficient C, number of nodes g and diameter D of the giant connected component GCC, mean node degree $\langle k \rangle$, assortativity r.

Analyses of the distributions of the local complex network measures and corresponding global measures, show that all three galaxy samples are similar and variations of mean values of network measures are in the range of their standard deviations. We have not seen apparent signs of evolutionary changes of network measures, maybe because the difference of cosmological ages of the samples of galaxies are not sufficient.

The comparison of the computed measures of our networks with corresponding measures of random ones, give us some global characteristics of the cosmic web, in the

context of complex network theory in that they are "large worlds". The GCC and diameter are found to be viable analogues to the super clusters and backbones found in cosmology. The high value of assortativity coefficient $r \sim 0.80 \div 0.86$ indicates that in the cosmic network, galaxies with similar number of links tend to be connected to one another.

We have calculated various centrality measures. All of which offer more insight into the structure of the cosmic web: degree centrality provide some information on how clusters are geometrically distributed; the closeness centrality is useful for determining cluster size distribution and the betweenness centrality measure, if turned correctly, can describe the filamentary structures in the cosmic web.

Most of the local network measures have non-Gaussian distributions, often bi- or multi-modal ones (Figure 5.3). The local clustering of each node $C(j)$, in the cosmic network, shows a three mode distribution, which allows for the discrimination between singlets and dumbbells of galaxies ($C = 0$) on the one hand and cliques of galaxies ($C = 1$) on the other. So, the network metrics analysed here allow for discrimination between topologically different structures.

Another goal of our study was to analyse the impact of galaxy surroundings on their astrophysical properties, in particular the colour indices and stellar masses of galaxies. Doing so, besides studying the obvious impact of a galaxies closest neighbourhood (which can be and is done by means of other methods too) we present here, an elaborated way to study the subtle topological features of galaxy distribution.

The mean values and distributions of colour indices and stellar masses of galaxies (Table 5.1) in the redshift slices z_1, z_2 and z_3 (Figure 5.7) are found to be similar. Meanwhile, the non-parametric tests used for distributions of colour indices of galaxies in z_1, z_2 and z_3 slices, show a weak evolutionary trend over the redshift span 0.88-0.97. Using the non-parametric Anderson-Darling test for the distributions of astrophysical characteristics, the statistically significant differences and correlations have been indicated in (Table 5.2 and Figure 5.8), for different selections defined by a different mode of clustering coefficient distribution. In particular, it was shown that stand-alone galaxies with $C(j) = 0$ (I) and galaxies densely packed in small cliques with $C(j) = 1$ (III)

are not distinguishable by colour indices and stellar masses distributions. Stellar mass distributions for galaxies with interim clustering coefficient (selection II) differ from the corresponding distributions in selections I and III. This difference holds for all redshift slices. The analogous difference in colour index distributions holds, however, only in the z_2 redshift slice. The latter z_2-sample has been intensively studied by other methods in the papers [116] and [103].

The presented results demonstrate the promising use of complex network theory in the study of the Cosmic Web. We hope that in future, with improving accuracy in defining redshift values for galaxies, this will allow the cosmic network to be determined in 3D which will, in turn, provide more accurate results.

Chapter 6

Conclusions

One of the corner stones of modern complexity science is to form analogies between statistical properties of different systems of interacting agents. Complex networks is one avenue from which this can be achieved. It has successfully explained many phenomena that have emerged in natural and man-made systems. A relatively new field of science, its ability to describe different types of interacting systems has sparked its rapid growth. In this thesis, we have sought to add to the body of knowledge on complex networks by studying the properties of three different networks: ideal polymer macromolecules, public transportation networks and Cosmic Web.

Complex network formalism have played a central role in explaining collective phenomena, that have emerged in all three case studies. Another common feature of networks considered, in this thesis, is their spatial embedding. This, in turn, opened a perspective of analysing the interplay between network, topological features and their properties in Euclidean (2D or 3D) space. The main conclusions of our analysis are given in sections 3.4, 4.5 and 5.6. Here, we briefly summarise our main findings.

In chapter 3, we present a study on the shape properties of ideal comb and dendritic polymers. To produce high precision numerical shape parameters, we redesign and apply a scheme proposed by Wei [41], [42]. Generating a polymer network of nodes (monomers) and edges (covalent bonds) this method takes advantage of the eigenvalue spectrum of the Kirchhoff matrix to determine the shape properties: g-ratio, aphericity $\langle A \rangle$, shape factor $\langle \delta_1 \rangle$, prolateness $\langle P \rangle$. We also determine the form factors $S(k)$ applying a scheme proposed by Benhamous [43]. These results are compared to Monte

Carlo simulations on both square and triangular lattices. The extrapolated values produced (for all of these methods) are in excellent agreement with each other. We confirm that comb polymers with a complete set of interior branches display a more spherical shape. This is also reflected in dendritic polymers as increased generations and branching per junction also generate more spherical structures. The redesigned Wei method has shown itself to be an accurate model in determining the shape parameters of comb and dendritic polymers. Here, more scope exists to further investigate other tree-like polymeric structures to determine their shape parameters.

In chapter 4, we investigate the topological and spatial features of UK public transport networks, using data derived from the National Public Transport Data Repository. We study the bus networks of London, Manchester, West Midlands, Bristol and national networks of rail and coach for 2011. We are able to rank networks in relation to their stability using the so called, Molloy-Reed parameter. Moreover, studying the fractal nature of these networks allows for useful interpretations regarding the serviceable area of stations. These models can be used to determine the accessibility, robustness and efficiency of networks. The next step in transport analysis is to consider their dynamical features. Answering questioning like how robustness is affected as a function of load? Are these systems correlated in time and space? and do dynamical properties correlate with topological properties? All these question remain open to future research.

Finally, in chapter 5, we present a complex network analysis for a region of the Cosmic Web, using data from the COSMOS catalogue [115]. We analyse galaxy samples in the same region $1° \times 1°$ of the celestial sphere with coordinates R.A.=$149°.4 \div 150°.4$ and Decl.=$1°.7 \div 2°.7$, for three neighbouring redshift ranges: $0.88 \leq z < 0.91$, $0.91 \leq z \leq 0.94$ and $0.94 \leq z \leq 0.97$. Results show the distribution of the network metrics for all galaxy samples are similar, showing no evolutionary change. We show how centrality measures can describe certain structures seen in the cosmic web. For example the degree centrality provides information on how clusters are geometrically distributed; the closeness centrality is useful for determining cluster size distribution and the betweenness centrality measure, if turned correctly, can describe the filamentary structures in the cosmic web. Our most significant finding is the correlation found between the astrophysical quantities: colour index and stellar mass and the network metric clustering

coefficient. These promising results open the door to future studies of the cosmos, using complex network approaches.

Appendix A

Goodness-of-fit method

Here we show the method we use to determine the function that best describes a given set of data. So for example when determining whether data is best described by a power law $\sim ak^{-\alpha}$ or an exponential $\sim be^{-\xi/k}$. First, we determine the standard errors for the free parameters (i.e. the coefficient and exponent in the above cases) applying a nonlinear least-squares (NLLS) Marquardt-Levenberg algorithm [53].

The standard errors are given in different scales (i.e log-lin and log-log), making them difficult to compare. Instead, we consider functions in linear space and apply integrals to determine the best fit. We find the difference in area, for each function, where standard errors give the upper and lower bound in area. From example below we give the areas calculated for a power law A_p and exponential function A_e

$$A_\mathrm{p} = \int\limits_{k_\mathrm{min}}^{k_\mathrm{max}} (a + \delta)k^{-(\alpha-\sigma)} - \int\limits_{k_\mathrm{min}}^{k_\mathrm{max}} (a - \delta)k^{-(\alpha+\sigma)} \, , \qquad (\mathrm{A.1})$$

and

$$A_\mathrm{e} = \int\limits_{k_\mathrm{min}}^{k_\mathrm{max}} (b + \delta)e^{-(\xi-\sigma)/k} - \int\limits_{k_\mathrm{min}}^{k_\mathrm{max}} (b - \delta)e^{-(\xi+\sigma)/k} \, , \qquad (\mathrm{A.2})$$

where δ and σ is the standard error for the prefactors and exponents respectively. The function that gives the least area is then considered the best fit for the given data.

Bibliography

[1] M. Newman, *Networks: An Introduction.* Oxford university press, 2010.

[2] L. Euler, "Solutio problematis ad geometriam situs pertinentis", *Commentarii academiae scientiarum Petropolitanae*, vol. 8, pp. 128–140, 1741.

[3] N. Pržulj and D. Higham, "Modelling protein–protein interaction networks via a stickiness index", *Journal of the Royal Society Interface*, vol. 3, pp. 711–716, 2006.

[4] P. Mac Carron and R. Kenna, "Universal properties of mythological networks", *EPL (Europhysics Letters)*, vol. 99, p. 28 002, 2012.

[5] J. Yose, R. Kenna, P. Mac Carron, T. Platini, and J. Tonra, "A networks-science investigation into the epic poems of ossian", *Advances in Complex Systems*, vol. 19, p. 1 650 008, 2016.

[6] C. Knappett, T. Evans, and R. Rivers, "Modelling maritime interaction in the aegean bronze age", *Antiquity*, vol. 82, pp. 1009–1024, 2008.

[7] M. Newman, "The structure of scientific collaboration networks", *Proceedings of the National Academy of Sciences*, vol. 98, pp. 404–409, 2001.

[8] D. Lusseau, "The emergent properties of a dolphin social network", *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 270, pp. 186–188, 2003.

[9] A. Motter and R. Albert, "Networks in motion", *arXiv preprint:1206.2369*, 2012.

[10] R. de Regt, M. Bishop, A. Barillas, T. Borgeson, and C. von Ferber, "The two dimensional shapes of simple three and four junction ideal comb polymers", *Physica A: Statistical Mechanics and its Applications*, vol. 458, pp. 391–398, 2016.

[11] M. Bishop, J. Stone, C. von Ferber, and R. de Regt, "The shapes of ideal five junction comb polymers in two and three dimensions", *Physica A: Statistical Mechanics and its Applications*, vol. 484, pp. 57–65, 2017.

[12] R. de Regt, S. Apunevych, C. von Ferber, Y. Holovatch, and B. Novosyadlyj, "Network analysis of the cosmos galaxy field", *arXiv preprint:1707.00978*, 2017.

[13] R. de Regt, C. von Ferber, Y. Holovatch, and M. Lebovka, "Public transportation in uk viewed as a complex network", *arXiv preprint:1705.07266*, 2017.

[14] S. Apunevych, R. de Regt, C. von Ferber, Y. Holovatch, and B. Novosyadlyj, "Using network analysis to clarify the relation of galaxy evolution to topological environment", in *Book of abstracts of VIII Scientific Conference Selected Issues of Astronomy and Astrophysics*, (Lviv, Ukraine), 2016.

[15] S. Apunevych, B. Novosyadlyj, R. de Regt, C. von Ferber, and Y. Holovatch, "Complex networks for modeling large-scale structure of universe.", in *Book of abstracts of joint steering committee meeting of the IRSES projects DIONICOS and STREVCOMS*, (Lviv, Ukraine), ICMP, 2017.

[16] R. de Regt and C. von Ferber, "Evolution of london public transport network", in *Book of abstracts of Coventry University Research Symposium*, (Coventry, UK), 2015.

[17] R. de Regt, M. Bishop, A. Barillas, T. Borgeson, and C. von Ferber, "Two and three dimensional shapes of simple of three and four junction comb polymers", in *Book of abstracts of 80th annual meeting and spring meeting of the German Physical Society (DPG)*, (Regenburg, Gemany), DPG, 2016.

[18] R. de Regt, C. von Ferber, Y. Holovatch, and M. Lebovka, "Topological and spatial aspects of public transportation in uk viewed as a complex network", in *Book of abstracts of COST meeting*, (Sofia, Ukraine), 2017.

[19] D. Larder, "Alexander crum brown and his doctoral thesis of 1861", *Ambix*, vol. 14, pp. 112–132, 1967.

[20] J. Travers and S. Milgram, "The small world problem", *Phychology Today*, pp. 61–67, 1967.

[21] P. Erdös and A. Rényi, "On the evolution of random graphs", *Bull. Inst. Internat. Statist*, vol. 38, pp. 343–347, 1961.

[22] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks", *nature*, pp. 440–442, 1998.

[23] A. Barabási and R. Albert, "Emergence of scaling in random networks", *science*, vol. 286, pp. 509–512, 1999.

[24] R. Albert and A. Barabási, "Statistical mechanics of complex networks", *Reviews of modern physics*, vol. 74, p. 47, 2002.

[25] S. Dorogovtsev and J. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2013.

[26] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge university press, 2008.

[27] E. Gilbert, "Random graphs", *The Annals of Mathematical Statistics*, vol. 30, 1141–1144, 1959.

[28] S. Strogatz, "Exploring complex networks", *nature*, vol. 410, p. 268, 2001.

[29] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabási, "The large-scale organization of metabolic networks", *Nature*, vol. 407, pp. 651–654, 2000.

[30] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994.

[31] M. Newman, "Scientific collaboration networks. i. network construction and fundamental results", *Physical review E*, vol. 64, p. 016 131, 2001.

[32] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a feather: homophily in social networks", *Annual review of sociology*, vol. 27, pp. 415–444, 2001.

[33] A. Fronczak, P. Fronczak, and J. Hołyst, "Average path length in random networks", *Physical Review E*, vol. 70, p. 056 110, 2004.

[34] U. Brandes, "A faster algorithm for betweenness centrality", *Journal of mathematical sociology*, vol. 25, pp. 163–177, 2001.

[35] F. Karinthy, "Chain-links: everything is different", *Recuperado de https://djjr-courses. wdfiles. com/local–files/soc180% 3Akarinthy-chain-links/Karinthy-Chain-Links_1929. pdf*, 1929.

[36] J. Douglas, J. Roovers, and K. Freed, "Characterization of branching architecture through "universal" ratios of polymer solution properties", *Macromolecules*, vol. 23, pp. 4168–4180, 1990.

[37] J. Freire, "Conformational properties of branched polymers: theory and simulations", in *Branched Polymers II*, Springer, 1999, pp. 35–112.

[38] S. Rangou, P. Theodorakis, L. Gergidis, A. Avgeropoulos, P. Efthymiopoulos, D. Smyrnaios, M. Kosmas, C. Vlahos, and T. Giannopoulos, "Synthesis, molecular characterization and theoretical study of first generation dendritic homopolymers of butadiene and isoprene with different microstructures", *Polymer*, vol. 48, pp. 652–663, 2007.

[39] K. Madaan, S. Kumar, N. Poonia, V. Lather, and D. Pandita, "Dendrimers in drug delivery and targeting: drug-dendrimer interactions and toxicity issues", *Journal of pharmacy & bioallied sciences*, vol. 6, p. 139, 2014.

[40] W. de Jong and P. Borm, "Drug delivery and nanoparticles: applications and hazards", *International journal of nanomedicine*, vol. 3, p. 133, 2008.

[41] G. Wei, "Exact shapes of random walks in two dimensions", *Physica A: Statistical Mechanics and its Applications*, vol. 222, pp. 152–154, 1995.

[42] G. Wei, "New approaches to shapes of arbitrary random walks", *Physica A: Statistical Mechanics and its Applications*, vol. 222, pp. 155–160, 1995.

[43] M. Benhamou, "Scattering from ramified polymeric systems", *Condensed Matter Physics*, vol. 7, p. 179, 2004.

[44] P. de Gennes, *Scaling concepts in polymer physics*. Cornell university press, 1979.

[45] K. Šolc, "Shape of a random-flight chain", *The Journal of Chemical Physics*, vol. 55, pp. 335–344, 1971.

[46] J. Rudnick and G. Gaspari, "The aspherity of random walks", *Journal of Physics A: Mathematical and General*, vol. 19, p. 191, 1986.

[47] J. Rudnick and G. Gaspari, "The shapes of random walks", *science*, vol. 237, pp. 384–389, 1987.

[48] E. Casassa and G. Berry, "Angular distribution of intensity of rayleigh scattering from comblike branched molecules", *Journal of Polymer Science Part B: Polymer Physics*, vol. 4, pp. 881–897, 1966.

[49] C. von Ferber, M. Bishop, T. Forzaglia, and C. Reid, "Form factor of simple three and four junction comb polymers", *Macromolecules*, vol. 46, pp. 2468–2473, 2013.

[50] P. Flory and M. Volkenstein, *Statistical mechanics of chain molecules*, 1969.

[51] P. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*. McGraw Hill Book Co, New York, 1969.

[52] S. Zweier and M. Bishop, "The shapes of h-comb polymers", *The Journal of chemical physics*, vol. 131, p. 116 101, 2009.

[53] K. Levenberg, "A method for the solution of certain non-linear problems in least squares", *Quarterly of applied mathematics*, vol. 2, pp. 164–168, 1944.

[54] M. Perrelli and B. Marvin, "Monte carlo simulations of ideal two dimensional h-combs", *Computers and Education*, vol. 25, p. 25, 2014.

[55] D. Gorry and M. Bishop, "Modelling and simulation of two dimensional h-comb polymers with the pivot algorithm", *Computers and Education*, pp. 95–98, 2008.

[56] M. Kosmas, T. Forzaglia, and M. Bishop, "End-to-end distances in four-junction polymers", *Macromolecular Theory and Simulations*, vol. 20, pp. 806–813, 2011.

[57] V. Blavatska, C. von Ferber, and Y. Holovatch, "Universal features of polymer shapes in crowded environments", *Physics Letters A*, vol. 374, pp. 2861–2864, 2010.

[58] V. Latora and M. Marchiori, "Is the boston subway a small-world network?", *Pysica A*, vol. 314, pp. 109–113, 2002.

[59] K. Seaton and L. Hackett, "Stations, trains and small-world networks", *Physica A*, vol. 339, pp. 635–644, 2004.

[60] X. Xu, J. Hu, F. Liu, and L. Liu, "Scaling and correlations in three bus-transport networks of china", *Physica A: Statistical Mechanics and its Applications*, vol. 374, pp. 441–448, 2007.

[61] Y. Sui, F. Shao, R. Sun, and S. Li, "Space evolution model and empirical analysis of an urban public transport network", *Physica A: Statistical Mechanics and its Applications*, vol. 391, pp. 3708–3717, 2012.

[62] X. Yang, G. Chen, B. Sun, S. Chen, and W. Wang, "Bus transport network model with ideal n-depth clique network topology", *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 4660–4672, 2011.

[63] L. Guo, Y. Zhu, Z. Luo, and W. Li, "The scaling of several public transport networks in china", *Fractals*, vol. 21, p. 1 350 010, 2013.

[64] P. Sen, S. Dasgupta, A. Chatterjee, P. Sreeram, G Mukherjee, and S. Manna, "Small-world properties of the indian railway network", *Physical Review E*, vol. 67, p. 36 106, 2003.

[65] R. Guimera and L. A. N. Amaral, "Modeling the world-wide airport network", *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, pp. 381–385, 2004.

[66] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles", *Proceedings of the National Academy of Sciences of USA*, vol. 102, pp. 7794–7799, 2005.

[67]   M. Guida and F. Maria, "Topology of the italian airport network: a scale-free small-world network with a fractal structure?", *Chaos, Solitons & Fractals*, vol. 31, pp. 527–536, 2007.

[68]   C. von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Transport networks: empirical analysis and modeling", *The European Physical Journal B*, vol. 68, pp. 261–275, 2009.

[69]   J. Sienkiewicz and J. Hołyst, "Statistical analysis of 22 public transport networks in poland", *Physical Review E*, vol. 72, p. 46 127, 2005.

[70]   H. Soh, S. Lim, T. Zhang, X. Fu, G. Lee, T. Hung, P. Di, S. Prakasam, and L. Wong, "Weighted complex network analysis of travel routes on the singapore public transportation system", *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 5852–5863, 2010.

[71]   L. Alessandretti, M. Karsai, and L. Gauvin, "User-based representation of time-resolved multimodal public transportation networks", *Open Science*, vol. 3, 160156, 2016.

[72]   S. Ghosh, A. Banerjee, N. Sharma, S. Agarwal, A. Mukherjee, and N. Ganguly, "Structure and evolution of the indian railway network", in *Summer Solstice International Conference on Discrete Models of Complex Systems (SOLSTICE)*, 2010.

[73]   H. Chang, B. Su, Y. Zhou, and D. He, "Assortativity and act degree distribution of some collaboration networks", *Physica A: Statistical Mechanics and its Applications*, vol. 383, pp. 687–702, 2007.

[74]   P. Angeloudis and D. Fisk, "Large subway systems as complex networks", *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 553–558, 2006.

[75]   B. Berche, C. von Ferber, and T. Holovatch, "Network harness: bundles of routes in public transport networks", in *AIP Conference Proceedings*, AIP, vol. 1198, 2009, pp. 3–12.

[76]   B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks", *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 71, pp. 125–137, 2009.

[77]   B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Transportation network stability: a case study of city transit", *Advances in Complex Systems*, vol. 15, p. 1 250 063, 2012.

[78] C. von Ferber, B. Berche, T. Holovatch, and Y. Holovatch, "A tale of two cities", *Journal of Transportation Security*, vol. 5, pp. 199–216, 2012.

[79] S Thibault and A Marchand, "Rśeaux et topologie", *Technical report, Laboratoire Méthodes, Institut National Des Sciences Appliques de Lyon (INSA), Villeurbanne, France*, 1987.

[80] L. Benguigui and M. Daoud, "Is the suburban railway system a fractal?", *Geographical Analysis*, vol. 23, pp. 362–368, 1991.

[81] L. Benguigui, "A fractal analysis of the public transportation system of paris", *Environment and Planning A*, vol. 27, pp. 1147–1161, 1995.

[82] L. Benguigui, "The fractal dimension of some railway networks", *Journal de Physique I*, vol. 2, pp. 385–388, 1992.

[83] C. von Ferber and Y. Holovatch, "Fractal transit networks: self-avoiding walks and lévy flights", *arXiv preprint:1209.2590*, 2012.

[84] P. Frankhauser, "Aspects fractals des structures urbaines", *Espace géographique*, vol. 19, pp. 45–69, 1990.

[85] K. Kim, L. Benguigui, and M. Marinov, "The fractal structure of seoul's public transportation system", *Cities*, vol. 20, pp. 31–39, 2003.

[86] *National transport data repository*, http://data.gov.uk/dataset/nptdr, Accessed: 2014-09-30.

[87] Y. Tu, "How robust is the internet?", *Nature*, vol. 406, pp. 353–354, 2000.

[88] R. Sole and M. Montoya, "Complexity and fragility in ecological networks", *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, pp. 2039–2045, 2001.

[89] M. Molloy and B. Reed, "Further algorithmic aspects of the local lemma", in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, ACM, 1998, pp. 524–529.

[90] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Resilience of the internet to random breakdowns", *Physical review letters*, vol. 85, p. 4626, 2000.

[91] D. Callaway, M. Newman, S. Strogatz, and D. Watts, "Network robustness and fragility: percolation on random graphs", *Physical review letters*, vol. 85, p. 5468, 2000.

[92] M. Barthélemy, "Spatial networks", *Physics Reports*, vol. 499, p. 1, 2011.

[93] M. Batty and Y. Xie, "From cells to cities", *Environment and planning B: Planning and design*, vol. 21, pp. 31–48, 1994.

[94] M. Batty, "The size, scale, and shape of cities", *science*, vol. 319, pp. 769–771, 2008.

[95] Y. Holovatch, R. Kenna, and S. Thurner, "Complex systems: physics beyond physics", *European Journal of Physics*, vol. 38, p. 23 002, 2017.

[96] H. Makse, S. Havlin, and H. Stanley, "Modelling urban growth patterns", *nature*, vol. 377, pp. 608–612, 1995.

[97] J. Bond, L. Kofman, and D. Pogosyan, "How filaments are woven into the cosmic web", *arXiv preprint:9512141*, 1995.

[98] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2013.

[99] M. Cautun, R. van de Weygaert, B. J. Jones, and C. S. Frenk, "Evolution of the cosmic web", *Monthly Notices of the Royal Astronomical Society*, vol. 441, pp. 2923–2973, 2014.

[100] Y. Chen, S. Ho, P. Freeman, C. Genovese, and L. Wasserman, "Cosmic web reconstruction through density ridges: method and algorithm", *Monthly Notices of the Royal Astronomical Society*, vol. 454, pp. 1140–1156, 2015.

[101] Y.-C. Chen, S. Ho, J. Brinkmann, P. E. Freeman, C. R. Genovese, D. P. Schneider, and L. Wasserman, "Cosmic web reconstruction through density ridges: catalogue", *Monthly Notices of the Royal Astronomical Society*, vol. 461, pp. 3896–3909, 2016.

[102] O. Hahn, "Collisionless dynamics and the cosmic web", *arXiv preprint:1412.5197*, 2014.

[103] S. Hong and A. Dey, "Network analysis of cosmic structures: network centrality and topological environment", *Monthly Notices of the Royal Astronomical Society*, vol. 450, pp. 1999–2015, 2015.

[104] F. Leclercq, G. Lavaux, J. Jasche, and B. Wandelt, "Comparing cosmic web classifiers using information theory", *Journal of Cosmology and Astroparticle Physics*, vol. 2016, p. 027, 2016.

[105] J. Lee and G. Yepes, "Turning around along the cosmic web", *The Astrophysical Journal*, vol. 832, p. 185, 2016.

[106] F. Pace, M. Manera, D. Bacon, R. Crittenden, and W. Percival, "The importance of the cosmic web and halo substructure for power spectra", *Monthly Notices of the Royal Astronomical Society*, vol. 454, pp. 708–723, 2015.

[107] P. Pranav, H. Edelsbrunner, R. van de Weygaert, G. Vegter, M. Kerber, B. Jones, and M. Wintraecken, "The topology of the cosmic web in terms of persistent betti numbers", *Monthly Notices of the Royal Astronomical Society*, vol. 465, pp. 4281–4310, 2016.

[108] N. Ramachandra and S. Shandarin, "Topology and geometry of the dark matter web", *arXiv preprint:1608.05469*, 2016.

[109] C. Zhao, F. Kitaura, C. Chuang, F. Prada, G. Yepes, and C. Tao, "Halo mass distribution reconstruction across the cosmic web", *Monthly Notices of the Royal Astronomical Society*, vol. 451, pp. 4266–4276, 2015.

[110] N. Libeskind, R. van de Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel, M. Alpaslan, M. Aragoon-Calvo, J. Forero-Romero, R. Gonzalez, *et al.*, "Tracing the cosmic web", *arXiv preprint:1705.03021*, 2017.

[111] B. Coutinho, S. Hong, K. Albrecht, A. Dey, A. Barabási, P. Torrey, M. Vogelsberger, and L. Hernquist, "The network behind the cosmic web", *arXiv preprint:1604.03236*, 2016.

[112] S. Hong, B. Coutinho, A. Dey, A. Barabási, M. Vogelsberger, L. Hernquist, and K. Gebhardt, "Discriminating topology in galaxy distributions using network analysis", *Monthly Notices of the Royal Astronomical Society*, vol. 459, pp. 2690–2700, 2016.

[113] M. Brouwer, M. Cacciato, A. Dvornik, L. Eardley, C. Heymans, H. Hoekstra, K. Kuijken, T. McNaught-Roberts, C. Sifón, M. Viola, *et al.*, "Dependence of gama galaxy halo masses on the cosmic web environment from 100 deg2 of kids weak lensing data", *Monthly Notices of the Royal Astronomical Society*, vol. 462, pp. 4451–4463, 2016.

[114] T. Kuutma, A. Tamm, and E. Tempel, "From voids to filaments: environmental transformations of galaxies in the sdss", *Astronomy & Astrophysics*, vol. 600, p. 6, 2017.

[115] O. Ilbert, H. McCracken, O. Le Fèvre, P. Capak, J. Dunlop, A. Karim, M. Renzini, K. Caputi, S. Boissier, S. Arnouts, *et al.*, "Mass assembly in quiescent and star-forming galaxies since z 4 from ultravista", *Astronomy & Astrophysics*, vol. 556, p. 55, 2013.

[116] N. Scoville, S. Arnouts, H. Aussel, A. Benson, A. Bongiorno, K. Bundy, M. Calvo, P. Capak, M. Carollo, F. Civano, *et al.*, "Evolution of galaxies and their environments at z= 0.1-3 in cosmos", *The Astrophysical Journal Supplement Series*, vol. 206, p. 3, 2013.

[117] H. McCracken, B. Milvang-Jensen, J. Dunlop, M. Franx, J. Fynbo, O. Le Fèvre, J. Holt, K. Caputi, Y. Goranova, F. Buitrago, *et al.*, "Ultravista: a new ultra-deep near-infrared survey in cosmos", *Astronomy & Astrophysics*, vol. 544, p. 156, 2012.

[118] J. Bardeen, J. Bond, N. Kaiser, and A. Szalay, "The statistics of peaks of gaussian random fields", *The Astrophysical Journal*, vol. 304, pp. 15–61, 1986.

[119] W. Press and M. Davis, "How to identify and weigh virialized clusters of galaxies in a complete redshift catalog", *The Astrophysical Journal*, vol. 259, pp. 449–473, 1982.

[120] A. Kolmogorov, "On the empirical determination of a distribution function", in *Breakthroughs in statistics*, Springer, 1992, pp. 106–113.

[121] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions", *The annals of mathematical statistics*, vol. 19, pp. 279–281, 1948.

[122] T. Anderson and D. Darling, "A test of goodness of fit", *Journal of the American statistical association*, vol. 49, pp. 765–769, 1954.

[123] H. Mann and D. Whitney, "On a test of whether one of two random variables is stochastically larger than the other", *The annals of mathematical statistics*, pp. 50–60, 1947.

[124] F. Wilcoxon, "Individual comparisons by ranking methods", *Biometrics bulletin*, vol. 1, pp. 80–83, 1945.

[125] J. Hartigan and P. Hartigan, "The dip test of unimodality", *The Annals of Statistics*, pp. 70–84, 1985.