HIERARCHICAL GRAPHS AND OSCILLATOR DYNAMICS.

H J DORRIAN

PhD 2015

HIERARCHICAL GRAPHS AND OSCILLATOR DYNAMICS.

HENRY JOSEPH DORRIAN

A thesis submitted in partial fulfilment of the requirements of The Manchester Metropolitan University, for the degree of Doctor of Philosophy

Department of Computing, Mathematics and Digital Technology
The Manchester Metropolitan University
February 2015

Abstract

In many types of network, the relationship between structure and function is of great significance. This work is particularly concerned with community structures, which arise in a wide variety of domains. A simple oscillator model is applied to networks with community structures and shows that waves of regular oscillation are caused by synchronised clusters of nodes. Moreover, we demonstrate that such global oscillations may arise as a direct result of network topology. We also observe that additional modes of oscillation (as detected through frequency analysis) occur in networks with additional levels of hierarchy and that such modes may be directly related to network structure. This method is applied in two specific domains (metabolic networks and metropolitan transport), demonstrating the robustness of the results when applied to real world systems.

A topological analysis is also applied to the real world networks of metabolism and metropolitan transport using standard graphical measures. This yields a new artificial network growth model, which agrees closely with the graphical measures taken on metabolic pathway networks. This new model demonstrates a simple mechanism to produce the particular features found in these networks.

We conclude that (where the distribution of oscillator frequencies and the interactions between them are known to be unimodal) the observations may be applicable to the detection of underlying community structure in networks, shedding further light on the general relationship between structure and function in complex systems.

Declarations

Parts of this thesis have appeared in print; specifically, the work presented in Chapter 5 appeared in the following article: 'Henry Dorrian, Jon Borresen, and Martyn Amos. Community structure and multi-modal oscillations in complex networks. PLOS ONE, 8(10):e75569, 2013.' [53]

Acknowledgements

I would like to extend my warmest gratitude to my supervisor, Dr Jon Boressen, for all the insight and help he has given me throughout my project. By giving me the freedom to work independently has made this a truly enjoyable experience. I am also very grateful to my director of studies, Professor Martyn Amos, for all of his guidance and support. Without this I would not have finished on time! Their differences in approach have amounted to a great supervisory team and have helped shape my research into something I am proud of.

A big thanks to Dr Kieran Smallbone, for not only providing me with data, but giving me an insight into how these abstract concepts can be applied to real problems and helping me to see the possible applications of my work. A big thanks to all my friends and family in Manchester for all their support throughout this seemingly endless work, and listening to my exciting tales of networks and oscillators over the many beers.

I thank the Dalton Research Institute, Faculty of Science and Engineering, Manchester Metropolitan University, for financial support via a Ph.D. studentship. And finally, I would like to thank the fireflies, without whom, none of this would be possible.

Contents

Chapter 1

Introduction

The study of complex networks is a vibrant and interesting field with a wide range of applications [54, 83, 93, 116]. Unravelling the structure of complex networks gives a significant understanding into many real world systems [116]. Although complex networks from a wide array of fields have been extensively studied, there is still much debate on the structure of some of the most important real world networks [15, 54, 124, 129, 152], which this work will address. Combining standard topological analysis with newly derived methods for determining structure will yield a greater insight into these networks and clarify which of the previous literature agrees with contemporary data. This work is particularly concerned with the real world networks of metabolism and underground railway networks, due to the importance of understanding how these systems function.

Arguably the most important finding presented in this work is a novel method of deciphering topological hierarchies in networks. These networks consist of

small groups of highly connected nodes connecting together to form ensembles, in which there are more connections within the group than to other groups. These ensembles are then connected to each other in the same manner, forming a hierarchical structure.[124]. Networks exhibiting a hierarchical structure are topologically self similar, a property found in many real world networks, which is described in detail in Section 2.3. We also describe a new artificial network growth model which shares the salient features of metabolic networks. Producing a generic growth model which shares features of a particular real world network allows for a more in depth study into their properties. This is achieved by allowing for multiple networks of any size to be generated which share features of the network under study, which can then be tested. For example, if we wanted to understand how perturbations affect such a network we could generate multiple artificial networks and then perform the particular perturbation under investigation [16]. The benefit of using an artificial network here is that multiple versions can be generated of the same size, with the same properties, rather than just the single network that is under study. These artificial networks can also help in the design of man made networks [26] in which desirable properties found in a particular network can be replicated.

This thesis consists of two main topics, usually considered as separate mathematical fields, which this work will bring together. The first is the topological analysis of real world networks, in which artificial network growth models are derived which share many characteristics with these complex networks. The second area in that of oscillator dynamics, which in this work is used to derive structural properties of networks by replacing each node with an oscillator. Although these two

fields may seem unrelated previous work has demonstrated that oscillator dynamics can be used to determine graphical properties [25, 103, 117, 153].

Combining methods from different areas of mathematics has always been fruitful in developing new mathematical techniques [135]. Applying techniques from fields of research which seem far apart can help to not only shed light on a specific problem but also help to gain insight into seemingly unrelated problems. In this work a method using oscillator dynamics is described and applied as a topological analysis, giving a new insight into the structure of some real world networks. This introduction will explain the motivation for the work and how the thesis is presented.

As well as gaining a more general topological understanding of real world networks, this work also pays particular attention to detecting pseudo-hierarchy in networks. A pseudo-hierarchical network is one with an underlying hierarchical organisation but with a less regular structure in which modules within the network overlap and contain edges which connect sub-networks of the graph which would not be connected in a truly hierarchical network. This can be thought of as a hierarchical network which has been rewired to reduce the regularity of the structure, as well as the hierarchical properties. Many real world networks are thought to have this type of composition, such as the networks of neurons in the brain and the networks of metabolism, which is thought to affect their properties [124, 163]. The networks of metabolism are one such class of network which have been suggested to be of this structure, which gives them a particular robust structure [124].

In the field of network theory it is often useful to produce an artificial network growth model which shares the fundamental characteristics of specific classes of real world networks. A simple growth model can consist of simply adding a node to a network, connecting to existing nodes according to some probability [14], although many other methods can be used [56, 124]. Previous work has attempted to categorise networks into different classes which can allow for similarities in their structure to be uncovered and demonstrate similar constraints on the way in which they develop [65]. Once similarities between networks are detected, theoretical network models can be produced which share these characteristics, suggesting possible mechanisms as to how these complex systems emerge. These characteristics are usually in the form of graphical measures taken on the networks. The importance of these models is firstly that it allows for mechanisms in which these networks may grow to be uncovered. These models also allow for a greater understanding of how to control these networks. For example, if a real world network is said to be scale-free it is drastically affected by targeted attacks, which can then be used to disrupt or protect the real world network [116]. The reason for this is due to the structure consisting of few very highly connected nodes with many less well-connected nodes. By attacking the highly connected nodes the structure is drastically affected, and therefore the function (This will be explained in more detail in Section 2.4). These models can also help in the design of artificial networks, in that their properties can be easily controlled if they are designed to be of a certain topological type.

In this analysis a network growth model is derived which agrees more closely with the structure of certain real world networks. Whilst conducting this analy-

sis it became clear that standard topological tools, whilst giving a large amount of information about the network, do not capture the full picture. To address this problem we conduct a dynamical analysis which allows for the presence of pseudo-hierarchy to be detected in complex networks. Networks of this nature are of great importance in current research, being thought to exist in many classes of complex networks [124]. The dynamical analysis provides a novel method for detecting hierarchies which uses oscillator dynamics to uncover these features which utilise the global behaviour of a network of oscillators. Although using oscillator dynamics to uncover topology has been previously performed [133], in this work a novel method of hierarchy detection is derived which uses a different oscillatory behaviour to previous methods.

Applying this novel method to real world networks helps to extract information about these networks and also demonstrates significant differences in their structure. It should be noted that it is not the aim of this work to dynamically model the behaviour of particular networks, rather the aim is to uncover structural properties which influence more general dynamical behaviours.

There are many applications in unravelling the structure of complex networks which are applicable to various fields. Understanding the structure of metabolic pathways can help in drug targeting. Similarly in transportation networks the effects of altering the network can be modelled before being implemented which, for example, could indicate which subway line to close to cause minimum disruptions. As well as these network specific applications a more comprehensive understanding of the structure and properties of networks can also help to uncover

similarities in how different types of networks have grown, suggesting underlying principles governing their development. It may also be useful to uncover the differences between man made and natural networks. It is clear, for example, that biological networks have evolved to be extremely efficient for the task they perform [93, 95, 139]. It may therefore be useful to produce man made networks, such as technological networks with the same underlying principles to help increase their efficiency and robustness.

As well as detecting hierarchy in networks whose local structure is known, the presence of oscillatory behaviour in networks with an unknown structure can also be used to determine how these networks are structured. This process would be very useful in uncovering the structure (or at least, structural properties) of large, complex networks where the structure can not be determined in a more traditional manner. For example, in the complex network of neurons in the brain, it is currently infeasible to actually model the network neuron by neuron due to the vast number of neurons and connections. By determining the structure of the brain from the oscillating waves it produces, (i.e. theta, gamma etc.) abnormalities and changes in development could be detected and related back to a particular change in network structure. The advantages of this type of detection would be its non invasive nature, allowing for the naturally produced waves to be detected and analysed, revealing the network structure.

This thesis covers two different topics, network theory and oscillator dynamics, each of which will be introduced in more depth in their respective chapters. The work on oscillator dynamics is used to give a new insight into the more traditional

approaches from network theory, providing a link between these two areas of research. We first explore previous work on network theory and its applications to real world networks. The main areas explored are in measures that can be taken on networks and in what way they describe a network. Previous work in the field of oscillator dynamics are then explored with models of oscillators being introduced and how these can be used to model certain behaviours seen in real world systems.

A topological analysis of the real world networks of metabolism and transportation networks is then presented. As previous work has shown [123, 129, 37], there are some disparities in properties found in real world networks and the generative models suggested to possess similar characteristics. Using these findings a new generative growth model is derived which agrees with the topological properties of metabolic pathways. The second analysis chapter derives a new method of hierarchy detection utilising oscillator dynamics which is first applied to generated networks to demonstrate its use. This is applied to both regular and irregular structures to illustrate the robustness of the method. This method is then used to analyse the real networks described earlier to derive more information about these complex systems, as well as to further demonstrate the ability of the method to work in networks exhibiting a degree of noise. The final chapter presents a discussion and gives some suggestions of further work which may help to develop this method of detecting structure, as well as further avenues of research in the topological analysis of networks.

Chapter 2

Review of Modelling Complex Real World Networks

In this chapter we present an overview of the current research in the field of complex networks. Section 2.1 gives an introduction to network theory and its applications, leading onto Section 2.2 which details the various graphical measures used in the analysis of complex networks. The concept of hierarchy in networks is then explained in Section 2.3. Section 2.4 gives an overview of artificial networks models leading on to Section 2.5 which gives details of some real world networks.

2.1 An Introduction to Networks

A *complex network* is a network which has a structure which is non-trivial. These are networks which are not globally connected or of a random structure but have complex topological features controlling their function. Many real world systems can be modelled using network theory and are usually examples of complex networks [21, 33, 134, 149]. Complex networks with similar properties can also be generated via algorithms, sharing some of the features seen in the real world networks they are used to model. The advantage of considering such systems in this way is that it allows for the underlying structure to be investigated. As the structure of a system directly effects the function this can allow us to understand why certain behaviour occurs. For example, in epidemiology diseases which are spread from physical contact can be seen to be dependent on the interactions of the population, or a network of interactions [21]. This helps us to understand how and why epidemics occur, and more importantly how best to control them. This abstraction also allows us to design and test systems before they are created allowing us to determine if, for example, a subway system will be able to cope with the increased usage as populations increases. This section gives a brief introduction as to how networks (or graphs) can be used to represent real world systems which can help to gain an understanding as to how these systems behave [32].

A *graph*, or *network*, is an ordered pair, $G = (V,E)$ consisting of sets satisfying $E \subseteq |V|^2$, in that the elements of E are a two element subset of V [51]. That is to say, a graph or network consists of nodes representing a particular item in the system and edges connecting them if a specified relationship occurs between the

two [116]. A simple example of this is that of a social network, where the nodes represent people and an edge connects two people if they are friends. So, if a social group consisted of four people where person a is friends with persons b, c and d, and persons b, c and d were not friends with each other this could be represented as the network in Figure 2.1:

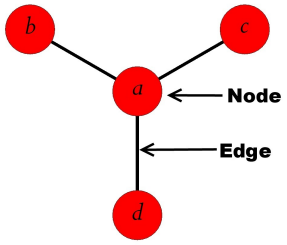

Figure 2.1: **A simple undirected network with nodes labelled.** *In this example nodes represent people and an edge connects the two if they are friends.*

This *undirected* network representation is one of the most basic in network theory [116]. In certain situations, further complexity can also be added by using a *directed* network, where $ab \not\Rightarrow ba$. That is to say, a path from a to b does not imply that there is a path from b to a, with an arrow shown on the edge demonstrating the direction of the relationship. Social networks are thought by most to be undirected networks, this is very apparent on social networking websites such as Facebook, where both parties must agree to be friends for a connection to be made [98]. However in some cases social networks can be directed, as one person may consider another person as a friend but the converse may not be true. Making an alteration to the earlier example in that person a considers d a friend, but d does not consider a a friend this could be represented as the directed network in

Figure 2.2.

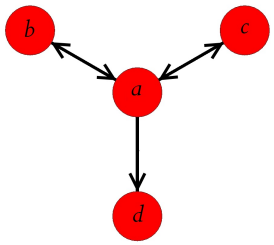

Figure 2.2: **A simple directed network with nodes labelled.** *In this example nodes represent people and an edge connects the two if they are friends with an arrow showing the direction of the relationship.*

These networks can now be used to represent more complex situations where connections have a direction. A further example of this is the World Wide Web [116], where one website may link to another via a hyper-link but the converse may not be true. In simple networks like these the length of the edges does not necessarily represent distance, usually the length of the edge does not represent any particular value and is just chosen to best visualise the network. A classic example of this is the London underground tube map, where the distances between stations are chosen for aesthetic reasons, as opposed to their actual distances [67]. This representation allows for the highly concentrated stations at the centre of London to be spread out so that they can easily be distinguished from one another.

Further information can be added to these simple networks by including values for the *weight* of the edges to produce a weighted network. For example, on a map of railways, where stations and railway lines are represented as nodes and edges respectively with the weights giving the length of the track. These weights

can also be more abstract, for example, the weight could represent the number of flights between two airports [100], or in the case of the network of scientific collaborations the weights represent the number of times a collaboration occurs [99]. These weighted networks can be both directed and undirected, depending on the real world system being modelled, as well as the complexity required for the investigation. In the analysis of weighted networks an important problem is that of the *The Travelling Salesman Problem (TSP)* [118]. In this problem the aim is to find a route which visits each node once and returns to the start in the minimum distance, inspired by the idea of a salesman having to visit a certain number of customers and return home. Although a seemingly simple problem no general method of solution is known and the problem is *NP-Hard*. Although not part of this work, the TSP demonstrates the complexity in solving a seemingly simple problem, and the wide ranging applications of such a solution.

For the purpose of this work unweighted networks will be used due to more tools being available for their analysis, the reasons for their suitability for each system being analysed will be explained in Section 2.5.

2.2 Graphical Measures

Once a system has been represented as a network the next step is to apply some analysis using graphical measures, which can help to derive information about these systems and allow for a greater understanding of how these systems behave. When looking at graphical measures, the first step is to convert the network into

a usable form, rather than a representational form. To achieve this an *adjacency matrix* is produced which contains all the information relating to the network's structure. The adjacency matrix, $A = (a_{ij})_{n\times n}$ of graph G is defined by Equation 2.2.1 [51].

$$a_{ij} := \begin{cases} 1 \text{ if } v_i v_j \in E \\ 0 \text{ otherwise} \end{cases} \tag{2.2.1}$$

That is to say, the adjacency matrix is an $n \times n$ matrix (n rows and columns) with n being equal to the number of nodes in the network. Element (a,b) is given a value of 1 if node a is connected to node b, and 0 otherwise. For an undirected network the matrix is symmetrical (about the leading diagonal), since if a is connected to b then the converse is true. Directed networks, on the other hand, are usually asymmetrical. If the adjacency matrix of a directed network is symmetrical, then it can be represented as an undirected network without loss of complexity. (for a weighted network the weights are given in the adjacency matrix, as opposed to values of 1 [81, page 18]). The adjacency matrix, denoted A for the undirected network in Figure 2.1 is given in Equation 2.2.2.

$$A_1 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{2.2.2}$$

As the network in Figure 2.1 is undirected matrix 2.2.2 is symmetrical, where (m,n)=(n,m). For the directed network in Figure 2.2 the adjacency matrix, A, is

given in Equation 2.2.3.

$$A_2 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.2.3}$$

As the network in Figure 2.2 is directed, matrix 2.2.3 is asymmetrical, as node d is not connected to nodes a (element (4,1)). These matrices convert a network into a usable form, removing any unnecessary information, such as the relative position of nodes, to allow for analysis to be conducted. The first measure to be considered here is the set of *degrees* of each node in the network. The degree of each node is the number of edges incident to it. For an undirected network each node has a single degree, for a directed network each node has an 'in' and 'out' degree. The in-degree is the number of edges going into that node and the out-degree is the number of edges leaving that node [116]. For networks of a small size this is done by simply counting the number of edges on each node, so for the network in Figure 2.1 the degree set is $D = [3,1,1,1]$ for nodes a,b,c,d respectively. The network in Figure 2.2 would have an in-degree of $D_{in} = [2,1,1,1]$ and out-degree of $D_{out} = [3,1,1,0]$ (again, for a,b,c,d respectively). As with all graphical measures, the degree can be calculated from the adjacency matrix. For a directed network the columns are summed to give the in-degree and the rows are summed to give the out-degree. The *degree sequence* can then be taken from the set of degrees which is the number of nodes having each degree, in ascending order. For the network in Figure 2.1 the degree sequence would be [3,0,1], i.e. three nodes of degree one, zero of degree two and one of degree three. The *degree distribution*

is then usually calculated, which is the probability distribution of the degrees over the network, in the case of 2.1, this would be [0.75,0,0.25]. These distributions have been found to follow various forms such as Gaussian, normal, binomial, exponential and power law [116]. The form which the degree distribution takes can be used to classify networks into different types with networks following the same distribution tending to have similar properties. For degree distributions following a *power law* the *decay rate*, γ, of the distribution is usually calculated [14].

Clauset *et el* [38] have demonstrated a method of detecting the presence of a power law behaviour in empirical data combining maximum likelihood with goodness-of-fit test based on the Kolmogorov-Smirnov statistic and likelihood ratios. Their method in the case of discrete distributions can be outlined as follows, for the derivation of this method please see [38]. The degree distribution, x, obeys a power law if it is drawn from the probability distribution

$$p(x) \propto x^{-\gamma} \tag{2.2.4}$$

where γ is a constant known as the exponent or scaling parameter. It should be noted that the α is used to denote the exponent by Clauset *et el* but has been changed to γ here to be consistent with the literature considered in this work. In empirical data this usually only applies above some minimum $x_{\min}$ which needs to be estimated, along with γ to determine if the hypothesis of the degree distribution following a power law is plausible. To estimate the value of $x_{\min}$ the Kolmogorov-Smirnov (KS) statistic used, which is the maximum distance between the CDF of

the data and the fitted model,

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)| \tag{2.2.5}$$

Where $S(x)$ is the complementary cumulative distribution function (CDF) of the observations with value at least $x_{\min}$. $P(x)$ is the CDF for the power law model that best fits the data in the region $x \geq x_{\min}$. The value of $x_{\min}$ which minimises D is our estimate, $\hat{x}_{\min}$. The estimate for γ is then found by the direct maximization of the likelihood function,

$$\mathscr{L} = -n \ln \zeta(\gamma, x_{\min}) - \gamma \sum_{i=1}^{n} \ln x_i \tag{2.2.6}$$

A goodness of fit test is then implemented to test whether the distribution follows a power law. If a power law distribution is plausible then we have our fitted γ value, if this is not the case then we can reject the hypothesis that the degree distribution follows a power law. If a power law is found then the decay rate can then used to classify networks and also demonstrates the growth rate of the network [116]. For the purpose of this work the decay rate of the degree distribution will not be considered as it has been demonstrated that the networks under investigation here do not follow a power law [38]. However, it is important to understand what this measure is due to it being used extensively in previous work.

The subsequent analysis will only be concerned with undirected networks so the

descriptions of measures will only be explained for this type of network (most of the measures have to be modified or are unsuitable for directed networks). Another measure to be investigated is the *eigenvalue spectrum* of the *laplacian* which has been linked to network structure [106, 121]. The first step is calculating the laplacian matrix from the adjacency matrix. The laplacian matrix, L, is calculated as $L = D - A$, where A is the adjacency matrix and D is the *degree matrix* [111]. The degree matrix has as many rows and columns as there are nodes. The matrix is all zeros, except element (n,n) which contains the degree of node n. That is, the leading diagonal contains the degree of each node in the network with all other elements being zero. The laplacian for the simple four node network shown in Figure 2.1 is calculated as follows:

$$L = D - A \tag{2.2.7}$$

$$Ł = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & -1 & -1 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix},$$

From this laplacian matrix the eigenvalues are then calculated. From 'The Algebraic Eigenvalue Problem' [156] the eigenvalues, λ, of matrix A are solutions to the equation $Ax = \lambda x$ where x is the corresponding eigenvector. In other words, the eigenvalue problem is the determination of the values of λ for which $Ax = \lambda x$ has a non-trivial solution [156]. The equation has a non-trivial solution when $det(A - \lambda I) = 0$, which gives the set of eigenvalues, λ. To find the eigenvalues,

the laplacian matrix is first produced, with λ being taken from each element in the leading diagonal. The determinant of this matrix is then calculated and solved as being equal to zero. The solutions of λ give the eigenvalues of the laplacian.

The eigenvalue spectrum contains as many eigenvalues as there are nodes in the network. There is always at least one zero, with the number of zeros being equal to the number of components within the network [59]. The lowest non-zero eigenvalue is known as the *algebraic connectivity* and gives an indication of how connected the network is as a whole as well as the robustness of the network [59]. In the field of oscillator dynamics, the algebraic connectivity indicates how fast a network will synchronise, due to the speed at which information can spread across the network [160]. As the spread of information is an important aspect of real world networks it is an important measure to consider when designing networks where the speed of information transfer is integral to their design.

Another important measure is the *geodesic path length*. The geodesic path length between two nodes in a network is the shortest route from one to the other, in terms of the number of edges. [31, page 58]. For example, taking the network in Figure 2.3, the minimum path from node *a* to node *c* would be two. Obviously for larger networks this becomes difficult to ascertain without using some calculations on the adjacency matrix. It is simple to show that squaring an adjacency matrix yields a matrix containing all the paths of length two, cubing yields all paths of three etc. [31, page 200], this is used to find the set of geodesic path lengths. The adjacency matrix, A, gives all the paths of length one, but as the minimum path from one node to itself is zero, the leading diagonal is set to zero to give all

the geodesic lengths of 1, P_1. To find the geodesic lengths of two, the adjacency matrix is first squared to give all the paths of length two, A^2. All non zero elements are then set to one, which gives a matrix containing a value of one for each pair of nodes that have a distance of two between them. The leading diagonal is set to zero, as the minimum distance from one node to itself is zero and any element in P_1 which has a value of one is set to zero in the matrix A^2 (as these node pairs have a minimum path of 1 between them) giving P_2, the geodesic paths of length two. This is continued for increasing powers (so for paths of length three where the adjacency matrix is cubed, the non zero elements are set to one, the leading diagonal to zero and any non-zero elements in P_1 or P_2 are set to zero) until P_i is a zero matrix and therefore the maximum geodesic length is $i-1$. Each P matrix is then summed along the row to give a sum vector, which is then summed again. This value is then halved (as the matrix contains paths from a to b and b to a, which are the same) to give the number of each geodesic path. These totals can then be used to calculate the average (mean) and maximum geodesic path lengths as well as being plotted in ascending order to give the geodesic path length spectrum. This method is demonstrated in the example below:

Taking the network in Figure 2.3 :

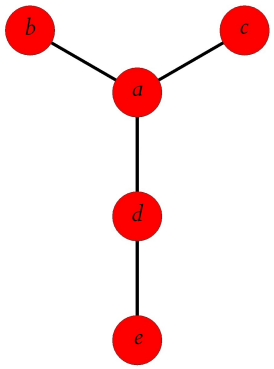

Figure 2.3: **A simple undirected network with five nodes.**

The corresponding connectivity matrix is:

$$A = P_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.2.8}$$

The leading diagonal is already zero, so this is equal to the geodesic lengths of one. Squaring the adjacency matrix, A, gives:

$$A^2 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 3 & 0 & 1 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \tag{2.2.9}$$

All the non zero elements in this matrix are first set to one. The leading diagonal is then set to zero as the shortest path from one node to itself is zero and any non-

zero elements from matrix P_1 are set to 0 in this new matrix, to give all geodesic lengths of two:

$$P_2 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{2.2.10}$$

To calculate the paths of length three, A is cubed:

$$A^3 = \begin{pmatrix} 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 3 & 0 & 1 \\ 3 & 3 & 0 & 4 & 0 \\ 0 & 0 & 4 & 0 & 2 \\ 1 & 1 & 0 & 2 & 0 \end{pmatrix} \tag{2.2.11}$$

Matrix A^3 then contains all paths of length three and again all non zero elements are first set to 1. The leading diagonal is then set to zero and all non-zero elements from the previous networks, P_1 and P_2 are set to zero in this new matrix, to give all the geodesic lengths of three:

$$P_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{2.2.12}$$

To calculate the paths of length four, A is put to the power of four:

$$A^4 = \begin{pmatrix} 3 & 3 & 0 & 4 & 0 \\ 3 & 3 & 0 & 4 & 0 \\ 0 & 0 & 1 & 0 & 4 \\ 4 & 4 & 0 & 6 & 0 \\ 0 & 0 & 4 & 0 & 2 \end{pmatrix} \tag{2.2.13}$$

Matrix A^4 then contains all paths of length four and again all non zero elements are first set to one. The leading diagonal is then set to zero and all non-zero elements from the previous networks, P_1, P_2 and P_3 are set to zero in this new matrix, to give all the geodesic lengths of four:

$$P_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.2.14}$$

As the matrix P_4 is all zeros, all geodesic lengths have been found with a maximum geodesic length of three. To obtain the set of paths from the matrices P_1, P_2 and P_3 each matrix is summed across the rows and then this sum vector is then summed. This value is then halved giving four minimum paths of length one, four minimum paths of length two and two minimum paths of length three. Therefore, any node can be reached from any other by traversing at most three paths.

Another measure to be considered is the *clustering coefficient* which is a measure of *cliqueness* [155]. The problem of finding cliques of given size is known to be an NP complete problem, where a clique is defined as a subset of the nodes of a graphs in which every pair of nodes in said subset are connected by an edge [1].

The problem of finding cliques in a network is related to the problem of finding hierarchies, in that nodes within one level of a hierarchy will tend to form cliques with one another, as opposed to with nodes in a different hierarchy. A standard measure which gives a measure of cliqueness, but does not find the actual cliques is the clustering coefficient, which is described as follows; If a vertex v has k_v neighbours then at most $k_v(k_v-1)/2$ edges can exist between them. Letting c_v denote the fraction of these allowable edges that occur, c is calculated as the average of c_v over all v [155]. This measure helps to classify networks into different types and show similarities in the structure and properties between various networks. Previous work has linked high clustering coefficients with a hierarchical network structure [124], however the presence of high clustering coefficients alone is not enough to classify a network as such. All these measures are used to describe the structure of networks which can help to describe their function. When used together they can help to build up a picture of a network and allow a suitable model to be determined, which can provide simple mechanisms to explain how particular features can be replicated [14].

The clustering coefficient of a network is a measure of transitivity - how the nodes in the network tend to cluster together. For this analysis a network average global clustering coefficient [155] is considered, of the form.

$$< c >= \frac{1}{n}\sum_{i=1}^{n} c_i, \tag{2.2.15}$$

where n is the number of nodes in the network and

$$c_i = \frac{2e_i}{k_i(k_i - 1)}, \tag{2.2.16}$$

where k_i is the degree of node i and e_n is the number of connected pairs between the nodes to which node i is connected.

Taking a simple four node network as in Figure 2.4:

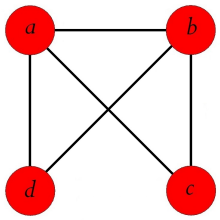

Figure 2.4: **A simple undirected four node network.** *This network has all but one possible connections, with the edge from d to c being the only connection not present.*

The local clustering coefficients are first calculated for each of the respective nodes, before being averaged. For node a, its neighbours (i.e. the nodes it is connected to) are b, c and d.

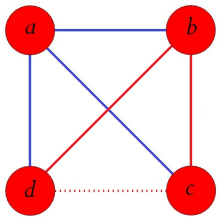

Figure 2.5: **A simple undirected four node network.** *The red and blue edges represent edges present in the network. The blue edges connect neighbours of node a to node a. The red edges show the edges connecting a's neighbours to each other, with the dotted line illustrating missing edges.*

From Figure 2.5 we see that the neighbours of node a (nodes b, c and d) have realised two out of the possible three connections (red edges), with the missing third illustrated with a dotted line. Therefore, the local clustering coefficient of a, C_a is $\frac{2}{3} = 0.667$.

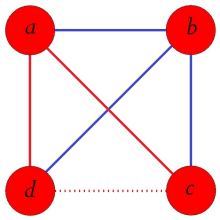

Figure 2.6: **A simple undirected four node network.** *The red and blue edges represent edges present in the network. The blue edges connect neighbours of node b to node b. The red edges show the edges connecting b's neighbours to each other, with the dotted line illustrating missing edges.*

From Figure 2.6 we see that the neighbours of node b (nodes a, c and d) have realised two out of the possible three connections (red edges), with the missing third illustrated with a dotted line. Therefore, the local clustering coefficient of b,

C_b is $\frac{2}{3} = 0.667$.

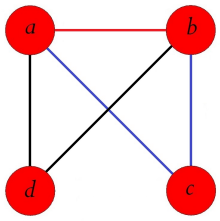

Figure 2.7: **A simple undirected four node network.** *The red, black and blue edges represent edges present in the network. The blue edges connect neighbours of node c to node c and the red edges show the edges connecting c's neighbours to each other. The black edges represent other connections in the network, which neither connect c to its neighbours or connect c's neighbours to each other.*

From Figure 2.7 we see that the neighbours of node c (nodes a and b) have realised one out of the possible one connection, in this case there are no further connections that c's neighbours could have. Therefore, the local clustering coefficient of c, C_c is $\frac{1}{1} = 1$.

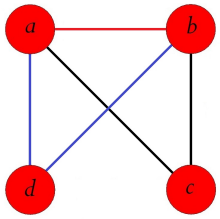

Figure 2.8: **A simple undirected four node network.** *The red, black and blue edges represent edges present in the network. The blue edges connect neighbours of node d to node d and the red edges show the edges connecting d's neighbours to each other. The black edges represent other connections in the network, which neither connect d to its neighbours or connect d's neighbours to each other.*

From Figure 2.8 it can be seen that the neighbours of node d (nodes a and b)

have realised one out of the possible one connection, in this case there are no further connections that d's neighbours could have. Therefore, the local clustering coefficient of d, C_d is $\frac{1}{1} = 1$.

To calculate the global clustering coefficient these four local clustering coefficients are averaged, giving $C = \frac{C_a+C_b+C_c+C_d}{4} = \frac{\frac{2}{3}+\frac{2}{3}+1+1}{4} = \frac{10}{12} = 0.833$. Hence the global clustering coefficient of the network shown in Figure 2.4 is 0.833.

The final measure to be considered in this work is *Q modularity* which is used as a measure of community structure in networks.

Q modularity is defined as [97]:

$$Q = \frac{1}{4m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})s_i s_j \tag{2.2.17}$$

where the expected number of edges between vertex i and j if the edges were randomly placed is $\frac{k_i k_j}{2m}$, k_i and k_j are the degrees of nodes i and j respectively and m is the total number of edges in the network. A_{ij} the adjacency matrix, and $s_i = 1$ if vertex i belongs to group 1 and $s_i = -1$ if it belongs to group 2. The factor of $\frac{1}{4m}$ is included to normalise the measure to make it compatible with other measures of modularity. The partitioning vector, s can be found using various different algorithm so as to maximise the value of Q. i.e the partitioning which gives the highest modularity is found. This process is repeated to subdivide the network into further modules, maximising Q. In the case of this work the eigenvector based method outlined by Newman [115] will be used. This method has been chosen as neither the number of modules nor the number of nodes in each module need

to be known before hand. As well as this clear advantage, this method has also been shown to yield better results then using greedy algorithms, although does take more computation time. As the network examined here a relatively small, the computation time is of no real concern. This optimisation works as follows, taken from [115]:

The equation in 2.2.17 can be written in matrix form as

$$Q = \frac{1}{4m} s^T \mathbf{B} \mathbf{s} \tag{2.2.18}$$

with s being the column vector of s_i and $\mathbf{B}$ the modularity matrix, which is a real symmetric matrix with elements

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m} \tag{2.2.19}$$

Taking equation 2.2.18 and writing s as a linear combination of the normalised eigenvectors u_i of $\mathbf{B}$ so that $\mathbf{s} = \sum_{i=1}^{n} a_i \mathbf{u}_i$ with $a_i = \mathbf{u}_i^T \cdot \mathbf{s}$ we have,

$$Q = \frac{1}{4m} \sum_i a_i \mathbf{u}_i^T \mathbf{B} \sum_j a_j \mathbf{u}_j = \frac{1}{4m} \sum_{i=1}^{n} (\mathbf{u}_i^T \cdot \mathbf{s})^2 \beta_i \tag{2.2.20}$$

Where β_i is the eigenvector of matrix $\mathbf{B}$ corresponding to eigenvector $\mathbf{u}_i$. The eigenvectors are labelled in decreasing order, $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$. We now want to maximise modularity by choosing the index vector $\mathbf{s}$, which is achieved by choosing the leading eigenvalue and corresponding eigenvector. The network is then divided into two groups according to the signs of this eigenvector. This process is

then repeated for each of these groups to divide the network into further modules. Once a proposed split will not increase the total modularity the corresponding subgroup is not divided any further, and once no more splits will increase modularity the algorithm is complete. This gives Q in the range of $[-1/2, 1]$ where a higher value corresponds to a more modular structure.

2.3 Hierarchy in networks

Many real world networks such as the language network, the World Wide Web and metabolic pathway networks have been shown to have features of a *hierarchical* network [123, 124, 17]. A hierarchical network consist of many small groups of highly connected nodes known as modules. These modules are grouped together in an ensemble with the modules being less well connected to one another than within their module. These ensembles of modules can also be connected to other ensembles, forming a self similar network in which a portion of the network has similar properties to the network a whole [124]. The modules within a hierarchical network represent functional entities [92]. In a social network these entities could be a family within the wider social networks where the family members are highly connected to one another, compared to the rest of the network. Clearly, in this type of network entities would overlap, many people in one family may all know many people in another family.

Networks exhibiting an underlying hierarchy are known as pseudo-hierarchical networks [147]. These networks have an underlying hierarchical organisation but

with a less regular structure in which modules within the network overlap and contain edges which connect sub-networks of the graph which would not be connected in a truly hierarchical network.

Networks with a hierarchical structure can be generated via algorithms which create this self-similar modular structure. The model outlined by [123] is explained in the next section as well as a selection of other important artificial networks.

2.4 Artificial network models

In the analysis of real world networks an important step is producing an artificial network which shares some features with the network under study, such as high clustering or low average geodesic path length [140]. These artificial networks enable us to study how processes spread over the network and give indications as to how the networks can be made more resilient or can be disrupted [39]. They can also be used to design networks with similar properties to real world networks which have some desired features such as fast information spread or robustness [62].

An important artificial network model is the *random network* model outlined by Erdös and Rényi [56]. This random network model is produced by first taking n nodes and then determining randomly out of all possible edges which edges to include, with the probability of an edge existing being p. The value of p can be used to tune the network's connectivity. Alternatively, the number of edges, e, to include can be chosen with random pairs of nodes selected until the required

number of edges is reached. Figure 2.9 shows such a network with $n = 200$ grown to have the same number of edges as in Figure 2.11.

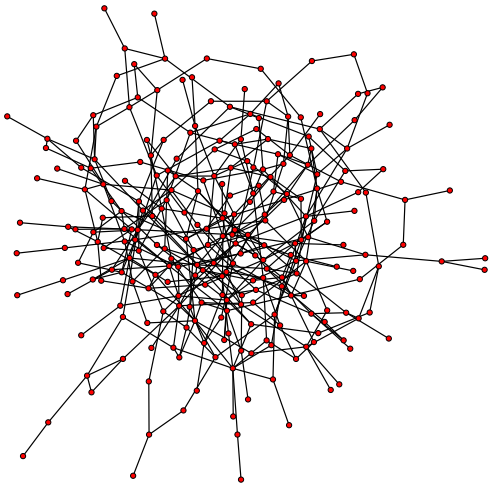

Figure 2.9: **Erdös and Rényi random network** *with n=256, e=412. The number of edges were chosen to be equal to the scale-free network in Figure 2.11 for a comparison.*

Random networks as in Figure 2.9 are known to follow a Poisson degree distribution, with low clustering coefficients, and low average geodesic path lengths compared to other network growth models [116]. In this specific case, the clustering coefficient is 0.00772 and the degree distribution following a Poisson distribution. The average and maximum geodesic geodesic path lengths are 4.96 and 11 respectively and the algebraic connectivity is 0.158. Although a very simple model, it demonstrates the concept of generating a network using a probability based algorithm and examining how the network changes as it grows. This is an important aspect of network theory as understanding and perhaps even manipulating how a network grows could be very useful in controlling certain real world systems.

One important class of networks is those possessing what is known as a *small-world* topology. A network with a small-world topology is one in which the average geodesic length is small relative to the total number of nodes in the network [116]. Social networks are one such network said to have this property. This was demonstrated by Milgram's social experiments [108]. In these experiments Milgram sent parcels to randomly selected people, which contained instructions to forward the parcel to a friend who they thought may bring the parcel closer to a specific individual. The recipient of this parcel would be given the same instructions, which Milgram was able to monitor via tracer postcards. This resulted in the parcels passing through chains of friends, showing the distance between people in the network. This demonstrated that social networks are small-world, with hubs (popular people) providing short cuts between chains of people with the average person to person distance being six. This concept has been popularised by the game *Six degrees of Kevin Bacon* in which the goal is to connect a particular actor to Kevin Bacon in as few steps as possible, where an actor is connected to another if they appeared in the same film [41]. This is exactly the same concept as Milgram's social network but instead of a generic social network, the network is the film actor network. Small world networks can be generated using the Watts and Strogatz model [155]. The model begins with a ring lattice in which each node is connected to a certain number of neighbours. In the simplest case a ring lattice consists of a ring of nodes with each connected to one neighbour on each side forming a ring of nodes, or more connected rings can be formed by connecting each node to more than one neighbour on each side. From this lattice each node is taken and each edge is rewired with probability p to a randomly selected node out of all possible nodes which avoid self loops and double edges. These generated

networks exhibit the low average geodesic path length and high clustering found in many real world networks such as the film actor network, the power grid and the neural network of a nematode worm [155]. Small-world network generated in this manner can be seen to lie between a fully regular and completely random structure. By comparing the clustering coefficients of these generated networks, the transition from regular to small-world networks is quite slight, whereas the transition from small-world to random networks is large. This demonstrates that the clustering coefficients can be used to classify networks and compare randomness (or regularity) [155].

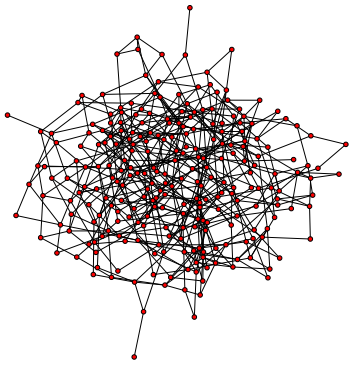

Figure 2.10: **Watts and Strogatz small-world network** *with n=256, k=4 and p=0.5.*

The Wattz and Strogatz small-world network in Figure 2.10 has a clustering coefficient is 0.174 with the degree distribution following a Poisson distribution similar to that of a random graph. The average and maximum geodesic path lengths are 4.64 and 8 respectively and the algebraic connectivity is 0.288.

Another useful measure in network theory is the algebraic connectivity, which gives an indication as to how connected a network is [59]. Simple networks such

as rings, stars and fully connected have an algebraic connectivity taking a specific predictable value. For more complex networks it has been demonstrated that the algebraic connectivity will be within particular bounds for different types of networks [59, 111]. This demonstrates that the algebraic connectivity is a useful tool in deciphering the structural properties of a network. As well as the first non zero eigenvalue, the algebraic connectivity, the whole spectrum of eigenvalues can be used to derive more information from a network, with a dynamical analysis demonstrating that it can also be used to detect hierarchy in artificially generated networks [34].

Another feature of networks used for classification is the degree distribution. Networks where the degree distribution follows a power law are known as *scale-free* [46] which are characterised by the presence of hubs in the network. Many real world networks have been suggested to have a scale-free structure, such as the World Wide Web [15, 29, 57], social networks [21, 108] and biological networks [58, 77, 161]. In a scale-free model the probability, $P(k)$, of a node in the network having k connections follow a power law distribution: $P(k) \sim k^{-\gamma}$ [14]. This is different from a random network, which is known to follow a binomial or Poisson distribution [56] but is similar to the network Milgram [108] described. Scale-free networks can be generated using simple algorithms, with the resulting networks having similar properties to some real world networks [108, 155]. The most commonly used algorithm is the Barabási and Albert model of *preferential attachment* in which the network is grown by starting with a single node and on each iteration a new node is added where connections are more likely to be made to nodes with a higher degree [14]. The probability of this new node connecting to an ex-

isting node i is shown in Equation 2.4.1. By using probability based attachment the network generated has very similar properties to those of certain real world networks.

$$P_i = \frac{K_i}{\sum_j (K_j)} \tag{2.4.1}$$

where P_i is the probability the new node joins an existing node, i, K_i is the degree of node i and $\sum_j (K_j)$ is the degree sum of all nodes in the network

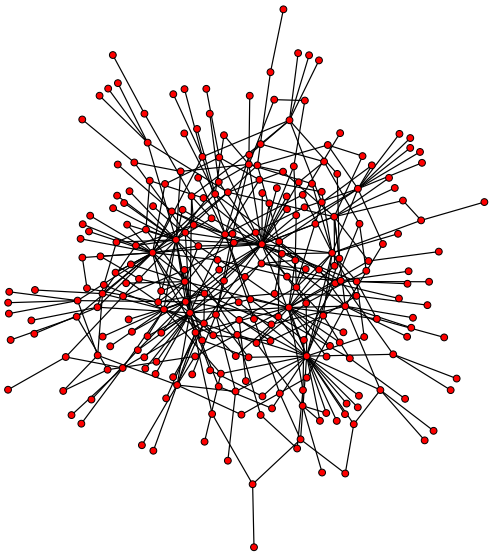

Figure 2.11: **Barabási and Albert scale-free network of preferential attachment** *with n=256 grown according to Equation 2.4.1*

Scale-free networks as in Figure 2.11 follow a power law degree distribution with higher clustering coefficient and lower average path lengths than those of a random network.

In this specific case, the clustering coefficient is 0.042 and the degree distribution following a power law, with $\gamma = 1.95$. The average and maximum geodesic path lengths are 3.77 and 7 respectively and the algebraic connectivity is 0.272.

Although, as mentioned earlier, many networks have been shown to be examples of scale-free networks, some networks are more closely modelled with a hierarchical model. Algorithms to produce these networks generally begin with a highly (or fully) connected sub unit repeated numerous times and then connected together, with this larger unit being copied and connected to others. This produces highly clustered networks, across which information passes rapidly, making them useful in man made information networks. Ravasz and Barabási [123] have proposed a hierarchical model which begins with four nodes being globally coupled, which is then replicated three times. The central node of the first cluster is then connected to the outer three nodes of each of the three new clusters. The central node of each of the four clusters are then all connected to each other. This process is then repeated, replicating this sixteen node network three times, and connecting them in the same manner as before. This processes can be repeated to produce large hierarchical network. Figure 2.12 shows such a network, consisting of 256 nodes.

This hierarchical model produces networks with higher clustering coefficients than that of a BA scale-free network, a power law degree distribution, a low average geodesic length and high algebraic connectivity [123]. For a hierarchical network of 256 nodes, as in Figure 2.12, the Clustering coefficient is 0.801, the average and maximum geodesic path lengths are 2.88 and 5 respectively and the algebraic connectivity is 0.360, with the degree distribution following a power law distribution with $\gamma = 2.39$.

Table 2.1 gives a summary of these measures on different artificially generated networks. Two different sizes of networks are given to demonstrate how these

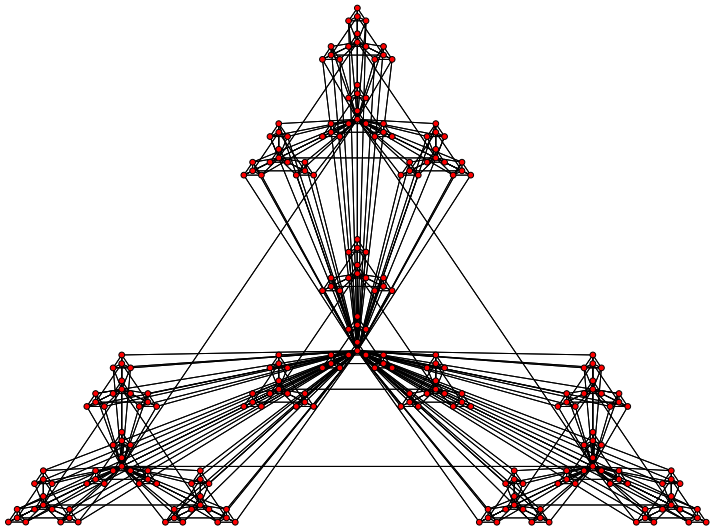

Figure 2.12: **Barabási and Albert hierarchical network** *with n=256*

variables change as the network size increases. For the Wattz and Strogatz small-world network, two versions for each size are used, to demonstrate how the probability used in the model alters the graphical properties.

Network	Nodes	Maximum Geodesic Length	Mean Geodesic Length	Algebraic Connectivity	Clustering Coefficients	γ
Barabási and Albert Hierarchical	64	3	2.25	0.603	0.833	2.28
Barabási and Albert Hierarchical	256	5	2.88	0.360	0.801	2.39
Wattz and Strogatz Small-World, p=0.2	64	7	3.63	0.346	0.268	NA
Wattz and Strogatz Small-World, p=0.4	64	6	3.37	0.438	0.132	NA
Wattz and Strogatz Small-World, p=0.2	256	11	5.70	0.137	0.313	NA
Wattz and Strogatz Small-World, 0=0.4	256	8	4.64	0.288	0.174	NA
Barabási and Albert Scale-Free	64	5	2.97	0.301	0.0745	1.97
Barabási and Albert Scale-Free	256	7	3.77	0.271	0.0415	1.93
Erdös and Rényi Random,	64	8	3.62	0.239	0.0548	NA
Erdös and Rényi Random	256	11	4.96	0.158	0.00772	NA

Table 2.1: **Various measures for different types of networks**

These artificial network models have been shown to share features with various types of real world networks [123, 15] and provide simple mechanisms for how

particular topological features can be produced. These models can aid in the creation of man made networks as well as testing their behaviour. The next section will focus on two particular important real world network, that of metabolic pathways and underground railway networks.

2.5 Real World Networks

2.5.1 Introduction

Many real world problems can be represented as a network theory problem, which allows for information about these systems to be discovered. A simple example would be representing stations in a railway network as nodes with edges connecting pairs of stations if a track connected the two. This simple representation can allow for an analysis which will demonstrate weaknesses in the network and suggest improvements in their design [79]. This type of graphical analysis has been conducted on many real world problems, and has shown that research in fields which seem very far apart such as the World Wide Web [15, 29, 57], social networks [21, 108] and biological networks [58, 77, 161] can all be investigated in a similar manner. Analysing these networks can shed light on how the systems have developed and allow predictions to be made about their behaviour. This section focuses on the previous work on modelling networks and the eventual application to metabolic pathways and transportation networks as understanding their structure and behaviour could have many applications in controlling these processes [65].

2.5.2 Metabolic Pathway Networks

Metabolic pathways are the set of biochemical reactions which occur within biological cells, the driving force behind life [128]. The chemical reactions that occur utilize catalysts and enzymes and allow for the conversion of certain chemicals into more useful products. These reactions allow for energy to be extracted from food as well as the synthesis of molecules required to sustain life [107].

Metabolic pathway networks can be analysed in much the same way as any complex network. They can be converted to an adjacency matrix where nodes represent chemicals and connected nodes represent chemicals that react together in some way [77]. Although metabolic pathway networks are usually considered to be directed networks, it can be argued that undirected networks are a better representation since reversibility of a reaction is not what determines whether a reaction will take place [152]. Using this representation Wagner and Fell [152] compared various metrics for a particular set of metabolic pathways to that of the random network. The metabolic pathways examined in their analysis were shown to have a higher standard deviation in degree than random networks of a similar size, illustrating that the random network does not model these types of network [152].

It has been suggested that metabolic pathway networks are examples of scale-free networks, with a degree distribution of 2.2 [77]. An important feature of scale-free networks is the small network diameter, which makes these networks robust against accidental failures which is due to the inhomogeneous topology. An accidental failure can be seen as the random removal of nodes in the network, in a

scale-free network this randomly selected node will usually be a low degree node due to the large number of these in the network compared to the number of highly connected nodes (or hubs). Removing these nodes will not disrupt the network significantly due to the low number of connections they have [18]. Conversely, this scale-free structure also gives networks an inherent weakness to targeted failures, in that knocking out a highly connected hub will drastically affect the network's functionality. As metabolic pathway networks have been suggested to have this structure this concept can be applied to drug design and administration. In effect, knocking out a specific metabolite can completely break down the network's functionality, and so target sites for drugs to act can be chosen from a topological perspective [77]. This also leads to the idea that the resistance to random errors of generated scale-free networks manifests itself in metabolic networks as a resistance to mutation and is a function of the network's structure [83].

An alternative model to preferential attachment has been proposed which uses an algorithm of *preferential depletion* inspired by the rewiring capability of the brain [129]. This method gives lower decay rates than preferential attachment and clustering coefficients which are higher and decrease less with system size, both being features found in metabolic networks [129]. To allow for the high clustering found in metabolic pathway networks Ravasz and Barabási [123] have proposed a hierarchical network model (as described in Section 2.4). This model is presented as a more accurate representation of metabolic pathway networks and may exist as a result of the way in which these systems have evolved over time [124]. Although fitting some of the metrics, this model does not produce some of the features seen in these networks such as long chains of metabolites and irregular structures.

The purpose of analysing metabolic pathways is to obtain as much information about these fundamental biochemical networks from the vast amounts of data becoming readily available, without such analysis the data is of little use. As well as simply classifying networks, the various tools available can be used to determine functional modules, in essence finding the networks within the network. This demonstrates that metabolic networks, and all complex networks for that matter, need to be investigated as a whole, not just by looking at small parts of the network as they all interact with each other in some way [65].

The network models described above tend to focus on the idea that metabolic pathways follow a power law degree distribution. However the work carried out by Clauset *et el* [38] (see Section 2.2 for an outline of this method) has demonstrated that this is not the case. For this reason this work will focus on other graphical measures to produce a model that fits closely with these attributes.

An application of the many analyses conducted on metabolic pathways is the ability to classify lifeforms into different groups [104]. By taking various measures from different metabolic pathways and examining the difference between discrete groups (ie aerobic and non-aerobic etc.), graphical differences between these groups become apparent which can then be used to classify any organism. These comparisons also suggest that species with a more complex lifestyle, where their environment is more demanding and less static, have developed bigger and more complex metabolic pathways. This demonstrates the varied amount of information that can be deduced from these chemical networks.

There are many methods for deriving metabolic pathway networks, one of which

is an energy and mass balance method in which all possible reactions are derived and the more plausible reactions are used [150]. Another method to derive these networks uses methods from network theory where process graphs (bipartite directed graphs) are used to derive feasible reaction pathways using a set of elementary reactions [101]. Other methods include using flux balance analysis and energy balance analysis to produce network representations of metabolic pathways which give an understanding of how these complex systems behave [20]. Although it is important to understand how these pathways can be derived, this work will use existing metabolic pathways from the BiGG database, a Biochemical Genetic and Genomic knowledge-base of large scale metabolic reconstructions [127].

2.5.3 Transport Networks

Historically, the study of transport systems and network theory have always been intrinsically linked. In fact, the *Seven Bridges of Königsberg problem* led to the beginning of graph theory as a mathematical field. In this classic problem one has to determine whether there exists a walk through the city of Königsberg, where the land masses are separated be a river, by crossing each of the seven bridges exactly once [7, 23]. The original problem can be seen in Figure 2.13, with the seven bridges connecting the different parts of the city.

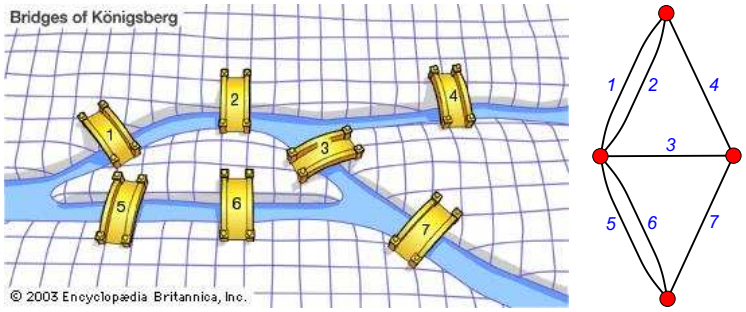

Figure 2.13: **The Seven Bridges of Königsberg** *Pictorial representation of the Seven Bridges of Königsberg problem taken from the Encyclopedia Britannica [35] (left) along with an abstracted representation of the same problem. In the abstracted version each land mass is represented by a node, with an edge representing each of the seven bridges. The edges are labelled with the bridge they correspond to. It should be noted that it is typically the nodes in a network that are labelled, with edge labels defining weights, however in this example it is the number of bridges that defines the problem.*

The importance of the solution to this problem was the abstraction Euler used. By representing the areas of land as nodes and the bridges as edges connecting them, Euler simplified the problem while still keeping all the salient features. The simplified network representation of this problem can be seen in Figure 2.13, showing the four nodes representing the four separate areas of land, with seven edges representing the seven bridges. This type of path, in which every edge is traversed exactly once, is known as a *Eulerian path*. To determine whether a solution exists, one has to simply count the degree of each node. For a Eulerian path to exist there can be at most two nodes of odd degree [81]. This is because all nodes except the starting and finishing nodes need to be entered and exited the same number of times, so need to have an even degree. For the seven bridges problem all four nodes have an odd degree, hence no Eulerian path exists. If a network is determined to have a Eulerian path the path can be found via various algorithms such as

Fleury's or Hierholzer's algorithms [157]. By identifying the important features of the problem, in this case the nodes and edges connecting them, a way to tackle the problem becomes apparent, demonstrating the importance of extracting all the salient information in a problem.

As transport networks are man-made (although geography does influence their topology) the underlying network needs to be understood to aid their design. In this analysis the *mass transit* networks of underground railways are analysed due to the fact they are an important feature of many cities around the world. Understanding more about how they function can help with the design of new underground rail systems, as well as improving older networks. Figure 2.14 shows maps of two such networks; the London Underground and the New York subway system.

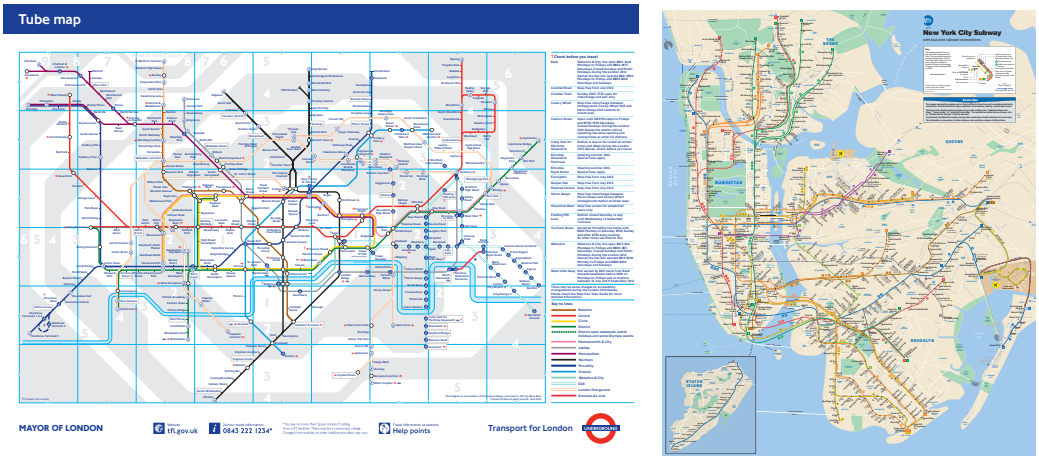

Figure 2.14: **Underground railway network maps** *Underground railway maps for the London Underground (left) and New York Subway (right). Both of these maps are the standard map used by customers taken from the 'Transport For London' [146] and 'Metropolitan Transportation Authority' [114] websites respectively.*

As with all real world networks the salient information needs to be extracted and converted to a usable form. Again, this is usually done by producing an adjacency matrix or a list of connected node pairs. For this analysis the former will be used. When representing a system of this kind as a network, nodes represent stations and edges represent a direct connection between pairs of stations. Previous work has used both weighted and unweighted representations and has demonstrated that when analysed as weighted networks the networks exhibit high efficiency of information exchange [93]. It has been shown that underground railway networks have different topological properties to other real world networks. This is due to the fact they are not a closed system, stations can also be reached by bus or walking. This makes the fault tolerance of the network less important, which is why the global efficiency of information exchange is allowed to be less than that found in other real world networks which are considered closed systems. Even with this difference to other real world networks they still have the similarity with many other networks of being small-world [93, 94, 132]. Networks which are constrained by geography, such as the power grid network and transport networks can also exhibit a fractal structure [43]. As these networks do not follow the structure of other real world networks, a different generative model has been suggested which grows the network via long chains [8]. This not only gives similar graph metrics, but in most cases (for example, the London Underground and New York Subway), also agrees with how these systems are known to develop, in that long chain of stations were added, originally by separate companies which were eventually joined together to make one large system [146]. This would suggest that railway networks not built in this fragmented manor may not share this structure, such as railways networks which have been designed as a whole rather than as

individual sections.

The network structure has also been shown to have a strong correlation with the number of people using the system [48], with usage increasing with measures such as directness and connectivity. This suggests that a subway network with high usage need to have certain topological properties which need to be considered in their design. Another application in analysing these networks is in finding possible vulnerabilities in the network. An extreme example of this is in highlighting targets that terrorists may attack. It is highly likely that the London tube bombings of 2005 were chosen using a topological analysis, in which the most effective targets were chosen to maximise disruption [79].

From the existing research it would seem subway networks are small-world networks [93, 94, 132], with the London Underground also being shown to exhibit an underling fractal structure [43]. Although growth models agree with certain measures examined in previous work [8] it is not clear whether these models agree with the fractal structure demonstrated in other studies. Analysing these networks in more depth with a dynamic analysis could be expected to provide a greater insight into their structure and therefore how they function.

2.6 Conclusion

In this chapter previous work on the analysis of various real world networks has been explored, giving an insight into the importance of analysing these complex systems. It has also been shown that in the current literature there is still some

uncertainty as to the structure that these networks have.

This chapter has presented previous work on modelling real world networks, beginning with simple networks and graphical measures building up to artificial network growth models used to model these complex networks. Metabolic pathway networks have been suggested to share properties with a few different topological models, which have been mainly based on the presence of a power law degree distribution. However by applying more rigorous testing the existing of a power law degree distribution has been disproved [38].

Judging by the existing literature, it seems that algebraic connectivity, while used in network theory has not been applied to metabolic pathways, suggesting they may be useful in analysing them or developing a more accurate model [28]. The various investigations into complex biological systems so far give insights into their evolution [159] and will eventually have many applications [16]. Given their importance, it seems apparent that a new model for metabolic networks needs to be produced which provides a more accurate model for these complex networks.

Previous work on transport networks has demonstrated the important applications in analysing these real world networks [48, 79]. Various studies have demonstrated that their underlying topological structure has small-world properties, but being unlike most other real world networks [93, 94, 132]. A more in depth topological analysis can help gain an insight into how these networks function and can also help in improving the function of existing networks, as well as in the design of completely new systems.

In the following chapter a review of oscillator dynamics will be presented which will be used in the subsequent analysis to derive graphical features of complex networks.

Chapter 3

Review of Synchronisation in Coupled Oscillator Models

A graphical analysis, as shown earlier gives a great insight into the structure of a system but does not demonstrate how they behave (although it does give indications). In this chapter a review on *oscillator* models is presented which will be used in Chapter 5 to provide a novel method of hierarchy detection.

3.1 Introduction

An oscillation is the repetitive change in some measure over time. A simple example of this would be the tide, over time it goes from being in to being out. Measuring the distance from the point where the sea meets the shore to some fixed

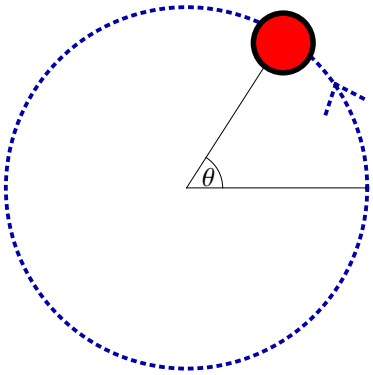

Figure 3.1: **A phase oscillator** *The oscillator is depicted as the red circle rotating around a fixed point along the blue circular path. The position of the oscillator at a given time is described by the angular co-ordinate, θ (known as the phase).*

point would yield a regular oscillation over time. Oscillations occur in many natural and man made systems from circadian clocks to electronic oscillators [141]. Figure 3.1 shows a simple oscillator which is rotating around a fixed point.

A phase oscillator, as shown in Figure 3.1 is described by it's phase, θ, which change over time. If the change in phase over time is known then the frequency, which is the number of rotations in a certain amount of time, can the be calculated. The change in the phase over time can be shown as a simple oscillation by taking the sine of the phase, as in Figure 3.2. From Figure 3.2 the period can be seen as the distance from one peak (or trough) to the next, with the frequency being the number of periods occurring within a specific length of time. To analyse such a signal a *Fourier transform* is often extremely useful [84].

The Fourier transform takes a signal in the time domain and transforms it to the frequency domain. This is achieved due to the fact that any signal can be broken down into an infinite sum of simple sinusoids (sine and cosine functions) of

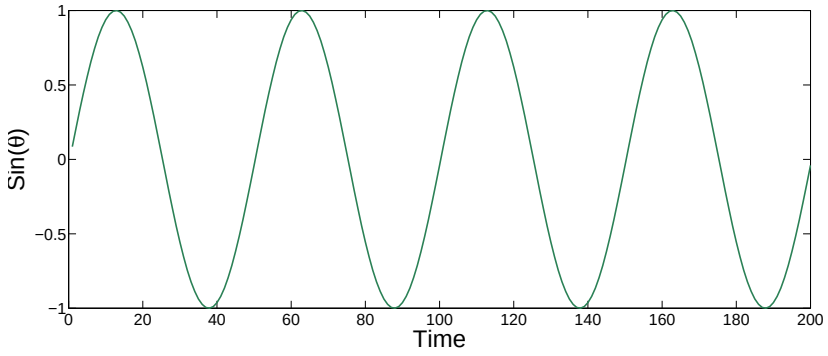

Figure 3.2: **Change in phase for a single oscillator** *Sine of the phase from a single oscillator as in Figure 3.1.*

different frequencies, shown in Equation 3.1.1 [84].

$$F(\nu) = \int_{-\infty}^{\infty} f(t)(\cos(2\pi\nu) - i\sin(2\pi\nu))\mathrm{dt} \tag{3.1.1}$$

where ν is the frequency and t is time. Which can be expressed in terms of exponents using Euler's identity, $e^{-i\theta} = \cos\theta + i\sin\theta$, as

$$F(\nu) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\nu t}\mathrm{dt} \tag{3.1.2}$$

In the case of signal processing the Discrete Fourier Transform (Equation 3.1.3) is used as an approximation to the Fourier Transform [151], which is applicable as both the time and frequency are discrete (assuming the analysis uses a digital computer).

$$F(k) = \sum_{n=0}^{N-1} f(n)e^{-2\pi ink/N},\ k = 0,1,1,...N-1 \tag{3.1.3}$$

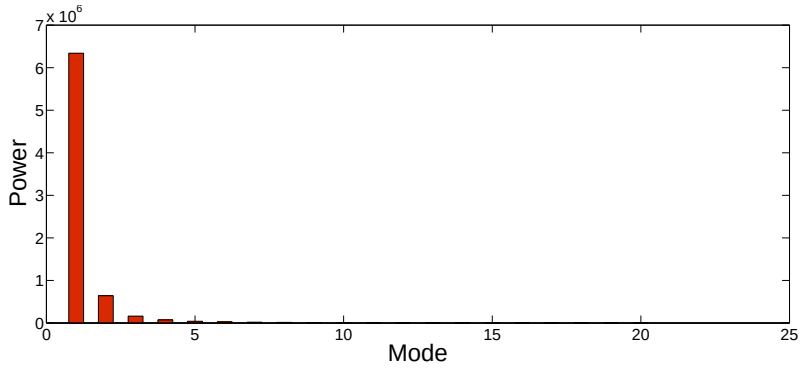

Figure 3.3: **Fourier transform for the phase of a single oscillator** *Fourier transform of the signal in Figure 3.2 showing a strong peak representing the mode of oscillation.*

where N is the number of samples, $f(n)$ is the signal sampled at n and $F(k)$ is the kth Fourier component. Squaring the series of complex numbers, $F(k)$ give the power of each component. Plotting this against k, the frequency or mode, produces a Fourier spectrum [24]. Taking the Fourier spectrum of the signal in Figure 3.2 yields the profile in Figure 3.3.

From Figure 3.3 it can be seen that the regular oscillation of the signal in Figure 3.2 yields a strong peak in the Fourier spectrum. This single peak represents one mode of oscillation in the original signal. The modes of oscillation can be thought of as the number of component signals which the signal is composed of. Taking a bimodal oscillation, in this case taken from the sum of two cosine functions (shown in Figure 3.4) and calculating the Fourier transform gives the two peaked transform shown in Figure 3.5. This demonstrates the ability of the transform to extract the modes of oscillation, which will also work in noisy signals. In fact, among the many applications of the Fourier transform is for the removal of noise from a signal [19]. By breaking down the signal into components the unwanted

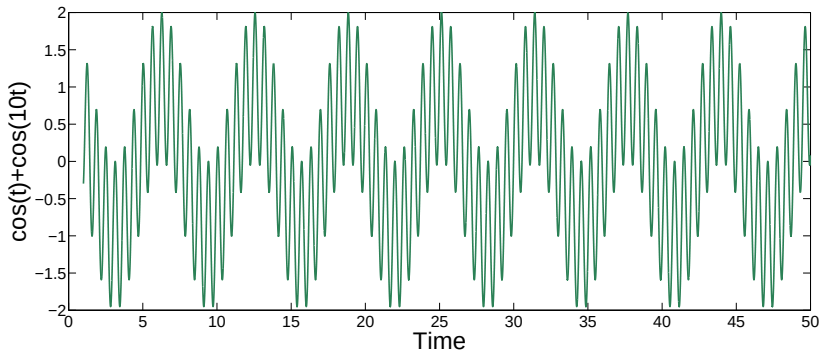

Figure 3.4: **Sum of two cosine functions.** *Time series for function of* $\cos t + \cos 10t$. *This function produces a bimodal oscillation.*

noise component can be removed, and can transformed back to the time domain, using the inverse Fourier transform. Digital signal processing relies heavily on Fourier transforms, for example in the task of compressing sound signals. One of the ways this is applied is to take the Fourier transform and then remove any high frequencies in the signal, which would be inaudible to the human ear [162].

For this analysis the fast Fourier transform will be used [84], which reduces the computation time and will be computed using MATLAB.

Clearly the behaviour of one oscillator is of little interest. However, various models have been proposed consisting of *coupled* oscillators, where the oscillators are connected in some way [141, 110, 61]. One of the ways in which this is done is to have the phase of each oscillator being dependent on some function of the phase of the oscillators which it is coupled to. These types of coupled oscillators exhibit synchronisation, in which, over time, all the oscillators settle down to having the same phase and frequency. A classic example of this phenomenon in the natural world is that of fireflies synchronising their flashing, the emergence

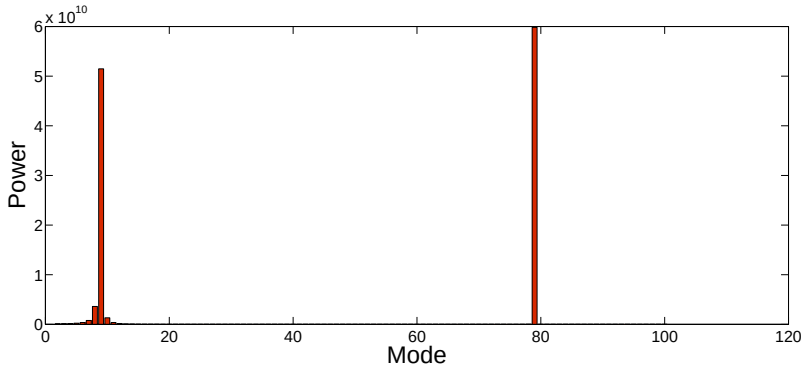

Figure 3.5: **Fourier transform of a bimodal oscillation.** *Fourier transform of the signal in Figure 3.4 showing two strong peaks representing the modes of oscillation. It can be seen that the mode of the second peak is 10 times that of the first, due to the multiple of t in the equation that produces the bi-modal oscillation.*

of which has been modelled using coupled oscillators [110]. Many more natural synchronisation systems have also been investigated including human reproductive cycles, sleep cycles and a possible mechanism for formation of memory [141]. This demonstrates that this phenomenon is a large part of the natural world and is an important field of study. For this analysis the synchronisation model is not used to demonstrate how a real system behaves but is instead used as a tool to decipher structure.

3.2 Systems Exhibiting Synchronisation

Arguably the most important oscillator model to exhibit synchronisation is that of the *Hodgkin-Huxley* model. Hodgkin and Huxley's Nobel prize winning work explained how action potentials arise in the giant axon of the squid [70]. Differential Equations are used to model how the change in ion concentrations give rise

to an action potential in nerve cells. In this model the change in potassium and sodium ions give rise to action potentials. The axon is modelled in terms of the lipid bilayer, voltage gated ion channels and leak channels. These are represented as capacitance, electrical conductance and linear conductance respectively. The system of equations representing this system are shown in Equations 3.2.1, 3.2.2, 3.2.3 and 3.2.4.

$$I = C_m \frac{dV_m}{dt} + \bar{g}_k n^4 (V_m - V_K) + \bar{g}_{Na} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_l) \tag{3.2.1}$$

where I is the total membrane current density and C_m is the membrane capacity per unit area. V_i and g_i are the reversal potential and conductance of the i^{th} ion channel respectively. K, Na and l represent potassium, sodium and leak channels respectively, with n, m and h being dimensionless quantities and t being time.

$$\frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n \tag{3.2.2}$$

where n is a dimensionless quantity, t is time and α_i and β_i are both rate constants for the i^{th} ion channel.

$$\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m \tag{3.2.3}$$

where m is a dimensionless quantity, t is time and α_i and β_i are both rate constants for the i^{th} ion channel.

$$\frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h \tag{3.2.4}$$

where h is a dimensionless quantity, t is time and α_i and β_i are both rate constants for the i^{th} ion channel.

From the set of four Equations 3.2.1, 3.2.2, 3.2.3 and 3.2.4 the change in I can be plotted over time, which allows for a model for an spiking neuron, shown in

Figure 3.2. This behaviour demonstrates that it is the change in ion concentrations within the axons which result in an action potential across the cell.

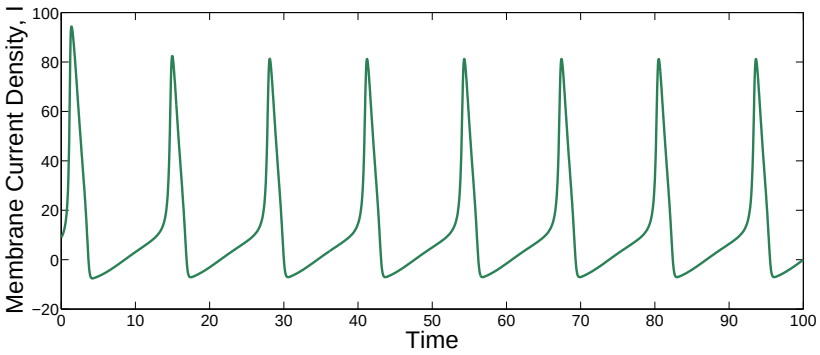

Figure 3.6: **Hodgkin-Huxley model of action potential in squid axons.** *From the set of four Equations 3.2.1, 3.2.2, 3.2.3 and 3.2.4 the total membrane current, I is shown to spike in the same manner as a neuron. This demonstrated that the differential equations can be used to model how action potentials arise in neurons.*

Further work has coupled these neural models together, allowing them to interact in some way and allow for more complex behaviours to be demonstrated, such as synchronisation in both the frequency and phase [36, 64, 154]. Although an important model, many adjustments to the model have been made which greatly simplify the model yet still allow for certain desirable behaviours. An important example of this is the *FitzHugh-Nagumo* model [61]. In this model the four dimensional Hodgkin-Huxley model is reduced to two dimensions by combining variables with similar properties. This reduction drastically simplifies the model, yet still allows it to model the spiking behaviour observed in neurons (see Equation 3.2.5).

$$\frac{dV}{dt} = v - \frac{V^3}{3} - w + I \tag{3.2.5}$$
$$\frac{dw}{dt} = \frac{1}{t}(V + a - bw)$$

where a and b are positive dimensionless parameters, V is the activation variable, W is the refractoriness of the system, I is the total membrane current density and t is time.

The behaviour of the variables v and w from Equation 3.2.5 is shown in Figure 3.2, both variables exhibit regular oscillations.

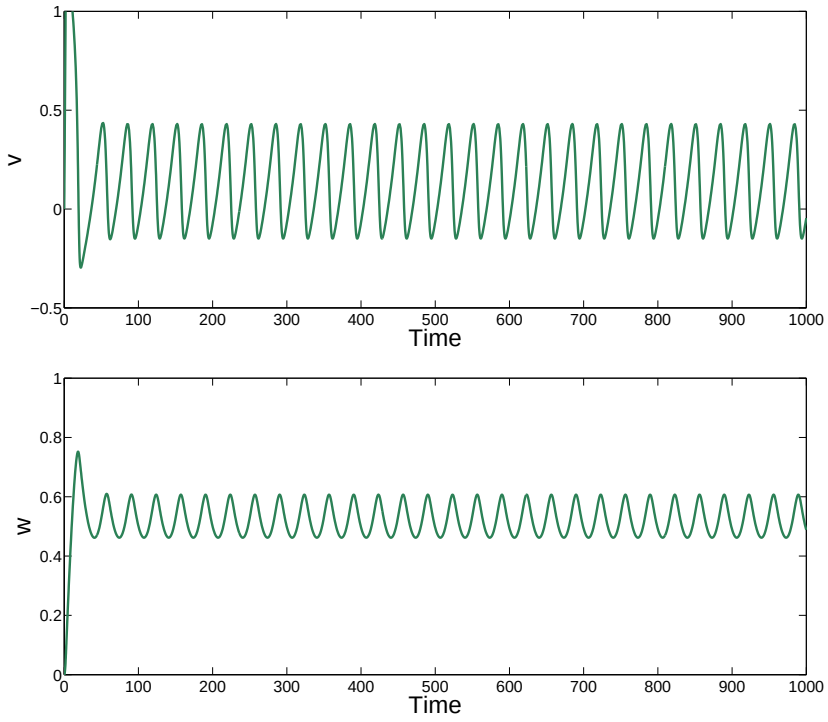

Figure 3.7: **Time dependent variables v and w of the FitzHugh-Nagumo model plotted against time.** *Both v and w from Equation 3.2 exhibit oscillatory behaviour.*

Combining the variables v and w in Equation 3.2.5 producing a phase portrait

(Figure 3.2) demonstrates the oscillator-like behaviour exhibited by the model. The phase portrait of the set of ordinary differential equations converge to a limit cycle, or an oscillator.

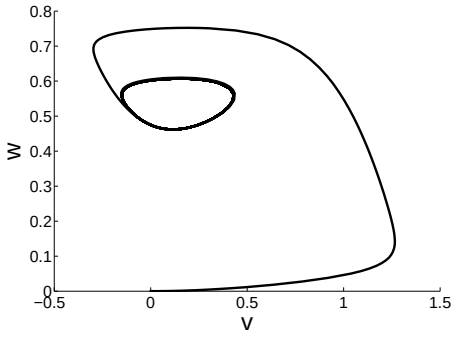

Figure 3.8: **Phase portrait of the FitzHugh-Nagumo model.** *It can be seen that the phase portrait of the time dependent variables v and w from Equation 3.2 form a limit cycle, behaving as an oscillator.*

Many other models have been proposed which capture the synchronisation phenomenon observed in many real world systems. One method to model and study synchronisation is using *Rössler* equations for the oscillators [120]. Equation 3.2.6 shows a single Rössler oscillator.

$$\begin{aligned} \dot{x} &= -y - z, \\ \dot{y} &= x + ay, \\ \dot{z} &= b + z(x - c), \end{aligned} \tag{3.2.6}$$

where x, y and z are dynamical variables defining the phase space of the oscillator and a, b and c are parameters. [126].

These oscillators can be coupled together, as in Equation 3.2.7, with the system being shown to exhibit synchronisation under certain conditions [120].

$$\begin{aligned} \dot{x}_i &= -\omega_i y_i - z_i + \varepsilon X, \\ \dot{y}_i &= \omega_i x_i + a y_i, \\ \dot{z}_i &= b + z_i(x_i - c), \end{aligned} \tag{3.2.7}$$

where x_i, y_i and z_i are dynamical variables defining the phase space of oscillator i, with ε being the coupling strength. a, b and c are parameters of the model, with values of 0.15, 0.2 and 10 respectively being shown to allow for phase synchronisation, although other values have also been shown to accomplish this. The mean field is $X = N^{-1}\sum_1^N x_i$, where N is the number of elements in the ensemble [120]. In the model shown in Equation 3.2.7, w_i is chosen from a Gaussian distribution and the coupling strength is the tunable variable, with the observable variable being X. These systems have also been used to model metabolic and neural networks and have helped to gain an insight into the periodic behaviour of metabolism [73]. In such a model the coupling between cells is performed by a metabolite that all cells produce with models exhibiting the asynchronous as well as synchronous behaviour [158], demonstrating how these seemingly abstract systems can be used to model real world systems.

Another model which is simpler than the ones described above but can show some of the dynamical properties described is proposed by Field [60]. The model is motivated by analogue computers which use cells coupled using *patch-cords* to their correct output which may provide a simpler way to analyse these complex real world systems and could be used to investigate metabolic networks. There are many ways in which molecular networks can be modelled dynamically and investigated, the main ones being, biochemical kinetic, generalised models of regulation, functional object-oriented databases, exchange languages, approaches based on formal methods and integrated frameworks with GUI [122]. Priami and Quaglia explain how although these methods are useful, they all have limitations in describing complex networks and so suggests the process algebra approach which does not have the same limitations, such as sensitivity to numerical parameters. This approach allows for hundreds of experiments to be conducted, in the form of computer simulations, without using reactants or animals. This can help to determine which experiments are viable, which can then be carried out as wet experiments, drastically reducing cost and time (once the system is developed and tuned). This has many applications such as drug design, gene therapy and environmental studies.

Although both the Hodgkin-Huxley and FitzHugh-Nagumo model capture a wide range of dynamic behaviour observed in real world systems, much simpler models have been proposed which can capture the collective dynamics of both these models. The *Kuramoto* model is one such model which uses coupled oscillators interacting in a manner which allows for synchronisation to occur [85]. The Ku-

ramoto model is applied to a network by replacing each node with an oscillator in which the change in phase is dependent on the phase of the nodes it is coupled to (Equation 3.2.8). With larger systems, computers are utilised, which, by comparing analytical solutions to those of a computer, have been shown to give accurate results [86]. Further work has also demonstrated that integration yields accurate results for the Kuramoto model [4].

$$\frac{\partial \theta_i}{\partial t} = \omega_i + \frac{k}{N} \sum_{j=1} \sin(\theta_j - \theta_i), i = 1....N \tag{3.2.8}$$

where N is the number of nodes in the network, ω_i is the natural frequency of oscillator i, k is the coupling strength between connected oscillators and θ_i is some oscillatory phase $\in [0, 2\pi]$.

This original model of Kuramoto assumes mean-field interactions. In the absence of any external noise, the global dynamics are determined by the coupling strength k, the distribution of natural frequencies $\sigma(\omega)$ and the connectivity within the underlying network. In general, the coupling strength k acts to synchronise the oscillators, the wider the distribution of ω, the harder it is for the oscillators to synchronise and higher connectivity within the graph also serves to cause the oscillators to synchronise (i.e. all to all coupling will synchronise more easily than sparsely coupled networks).

In order to measure the coherence of the oscillators the Kuramoto order parameter is often used:

$$\chi = \frac{1}{N}\left|\sum_{j=1}^{N} e^{i\theta_j}\right|, \tag{3.2.9}$$

This is the average phase of all oscillators within the network. For a fully synchronised networks $\chi = 1$ and for networks where the phases of all oscillators are equally distributed around $[0, 2\pi]$, $\chi = 0$. For all other states, χ is in the range $(0, 1)$. This is demonstrated in Figure 3.9, which shows that as the phase of the nodes move towards synchronisation, the Kuramoto order parameter moves towards 1.

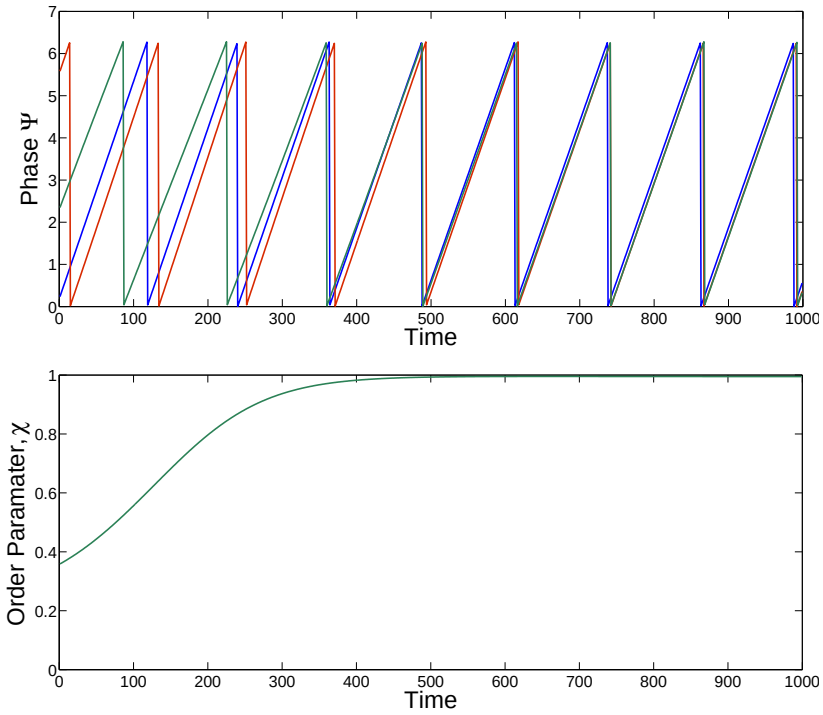

Figure 3.9: **Phase of oscillators and Kuramoto order parameter of three coupled oscillators** *Phase of the oscillators in a three node network as the model is iterated through, which can be seen to be synchronising with one another (top). Order Parameter of the phases of the three oscillators in the top figure, as the oscillators synchronise the order parameter moves towards one (bottom).*

For the analysis conducted in this work the Kuramoto model will be used. This is mainly due to the fact that the model is relatively simple to implement yet can demonstrate a wide range of dynamical behaviour. The conclusions drawn using the Kuramoto model will also be applicable to other oscillator models as their interacting nature that dictates the behaviours they exhibit, rather than the actual form of the oscillator.

3.3 Modelling Systems and Applications

The applications of synchronisation and dynamics, as touched on earlier, are very wide. One practical application is using these models, in this case from a system of Bloch equations, is to interpret anomalies observed in Nuclear Magnetic Resonance (NMR) experiments [148]. Another neuroscience application is in determining the structure of the brain from the different brain waves which can be picked up by an EEG [80]. Josic and Rubin explore this idea to use the dynamics observed on networks to infer their structure which will be a very useful tool in detecting illness as well as understanding the structure of the huge network of the brain.

Another area where synchronisation and dynamics can be applied is in understanding the circadian clocks which not only control sleep cycles but are also known to affect metabolic cycles. Huang *et al.* [74] explain that when people have a circadian misalignment, for example if they work night shifts, they have increased incidents of diabetes and heart disease showing the relationship between circadian clocks and metabolism and the importance of understanding the mechanisms of circadian clocks and metabolism. Metabolism synchronisation has also been observed, Bianchi [22] explains how yeast cells have been shown to synchronise their metabolism with each other, from alkaline phase to acid phase. Another influence the circadian clock has is on glucocorticoids production, animals with faulty circadian clock genes have defects in cell proliferation, which is controlled by glucocorticoids [50]. Dickmeis and Foulkes explain that this effect of the cir-

cadian clock could be used to improve cancer treatments, as glucocorticoids are used for certain treatments, administering at certain parts of the circadian clock could improve treatment. This gives a very important application where understanding the dynamics of the circadian clock can have very obvious benefits.

The synchronisation effect seen in oscillator models can also be applied to laser arrays [4]. The effect being that when an array of lasers are phase locked a large output power is achieved by a high number of low powered lasers. Similarly, coupling Josephson Junctions so that they operate in synchrony allows them to achieve a large power output, due to the fact that power is proportional to the number of Josephson junctions squared [4]. The benefit of this is that a larger power is achieved than if they were operating independently.

There has also been some work in the more abstract application of investigating structure of complex networks. By examining how quickly different nodes in the network synchronise, clusters within the network can be detected, and so structure and hierarchy inferred [9]. This method can also be used to find missing links in a network, which is useful when not all the information is available for the network [37], for example if the network is a biological one not all the connections may be known at present.

3.4 Conclusion

The field of dynamics and synchronisation have been used to uncover the mathematical basis behind many complex natural phenomenon. There are various models which can be used to model the synchronisation effect known to occur [60, 69, 120, 122], each with its own limitations and advantages. This work will use the Kuramoto model [85], as it is simple to implement but can still simulate complex behaviour. There are many applications to this work, some of which are outlined above, such as understanding epilepsy and improving cancer treatments and no doubt many more will become apparent as this field is studied in more depth and applied to more problems.

This chapter has explored some of the previous research in the field of oscillator dynamics, examining the various oscillator models known to exhibit synchronisation and the applications in this field of research. Chapter 5 will derive a novel method of detecting topological hierarchies which utilise a particular dynamical behaviour. This method is applied to various artificially generated network to demonstrate how the behaviour of the oscillators is used to extract topological features. The model is also applied to real world networks, giving a deeper understanding into the structure of these complex networks than a standard topological analysis can yield.

In the following chapter a standard topological analysis is performed on real world networks utilizing the methods outlined in Chapter 2, producing a network growth model which agrees closely with the topological features of the networks

of metabolism.

Chapter 4

Topological Analysis of Real World Networks

This chapter presents an analysis of the real world networks of metabolism and underground railway networks. This analysis yields a novel network growth model which captures the structural features of the networks of metabolism. Producing a network growth model with the same features as these real world networks can help in the design of man made networks such as transportation networks which we may want to design with similar properties to networks found in nature [16]. These artificial networks can then also be used to test particular properties, allowing for numerous examples to be tested with the same properties as the system under investigation [26].

4.1 Metabolic Pathway Network Analysis

4.1.1 Introduction

The graphical structure of metabolic pathways has been extensively studied [77, 89, 131, 152] and has led to a greater understanding of their functionality [138]. Many of the salient features of such networks have been investigated, particularly any scale-free nature [15], although this model is currently the subject of some debate [82, 143]. Of particular interest are the high clustering coefficients observed in metabolic pathways which have previously been explained using concepts of topological hierarchy [123] and modularity [144].

The once widely accepted notion that metabolic pathways follow a power law degree distribution [77, 152] has recently been challenged by applying a more rigorous test for this topological property [38]. This demonstrates that the BA scale-free model is not an accurate model for these networks, as it is the power law property that is the main focus of this network model. In this analysis we will derive a more accurate network model which agrees closely with topological features found in metabolic networks.

We first conduct a graphical analysis on eight bacterial metabolic pathways, concentrating on their clustering coefficients, average (mean) geodesic lengths, maximum geodesic lengthsand algebraic connectivity. Here microorganisms have been chosen, as their metabolic pathways are less complex than higher organisms allowing greater illustration of the concept [104]. Further to this, a growth model in

which the rate of growth decays as a function of network size is used to replicate some of the structural features of metabolic networks.

4.1.2 Methodology

The purpose of this analysis is to produce a simple network growth model which has similar characteristics to metabolic pathway networks. The first step in producing such a model is to take various graphical measures on these networks and determine what this suggests about their structure. An artificial network growth model can then be derived which shares these characteristics.

Microbial metabolic pathways from the following species were considered:

i *Escherichia Coli*

ii *Escherichia Coli iAF1260*

iii *Escherichia Coli iJR904*

iv *Helicobacter Pylori*

v *Methanosarcina Barkeri*

vi *Staphylococcus Aureus*

vii *Mycobacterium Tuberculosis*

viii *Saccharomyces Cerevisiae*

Three formulations of *E Coli* are chosen to ensure the results are independent of the methodology used to initially determine the network. *E Coli* are a gram-negative bacteria which are considered a model organism in microbiology. They are commonly found in the intestines of mammals with some strains being known to cause food poisoning [42]. *H Pylori* are another gram-negative bacteria which are commonly found in the stomach of mammals and are known to cause gastric ulcers [49]. *S Aureus* are a gram-positive bacteria which are known to cause skin infections and respiratory disease [44]. *M Tuberculosis* is another gram-positive bacteria and is known to infect the lungs [119]. The selection of bacteria were chosen due to the importance of them in the field of medicine and the large amount of research on these particular microbes. *M Barkeri* is an example of an archaea which is a distinct type of prokaryote to bacteria and has the unusual feature of showing primitive cellular differentiation [78]. This example was chosen due to these unusual features and the large amount of research that has been carried out on this microbe. The final microbe investigated is *S Cerevisiae* which is a model eukaryote organism commonly known as yeast [125]. This was chosen due to the importance it has in modern food production (being used to make bread and alcohol), as well as the large amount of research that has already been conducted.

The eight metabolic reconstructions are downloaded in SBML format [75] from the BiGG database [127]. The models were imported to Matlab using libSBML [27].

For this analysis each metabolic pathway is represented as an undirected network with the adjacency matrix of the network being the boolean representation of the chemical interactions. For each model, nodes i and j are defined as adjacent if

metabolite i appears as a reactant and metabolite j as a product, or i as a product and j as a reactant in any reaction. Although this is a highly simplified representation this has been shown to be a useful tool in analysing such systems [130].

Note: There are a variety of network constructions available, both including and excluding sub-cellular compartmentalisation, and with/without considering the role of water and protons. In this analysis two versions of each of these pathways will be considered both including and excluding the presence of water, without compartmentalisation, as this most accurately represents the nature of the biochemical processes within the cells [77, 89, 131, 152]. The results obtained here are applicable to the other available formulations with some small modification to the size dependent decay constant α described in Section 4.1.3.

A selection of graphical measures are considered in this analysis, namely the clustering coefficients, average and maximum geodesic length and the algebraic connectivity which are explained in Section 2.2. These measures were chosen as they give indications as to how fast information is spread through the network. This is a particular important feature of networks as it helps to deduce how these networks function. For example, a network with low average geodesic lengths and high clustering given the connectivity of said network will have a particular efficiency of information spread. When designing artificial networks it is of particular importance to replicate or avoid these features to give the network certain properties.

4.1.3 Results

The geodesic path length, as mentioned earlier, is the shortest path between any pair of nodes, in terms of the number of edges that need to be traversed. For the metabolic pathways investigated here, the average geodesic path length is surprisingly low given both the size of the networks and their average degree being $D_{mean} = 6.42$ in the range of $[5.03, 8.08]$. The average value of $G_{mean} = 2.99$, in the range $[2.61, 3.43]$, demonstrates a very strong small-world effect [108]. Conversely, the highest of the geodesic path lengths are greater than one would expect, given the average of these is relatively low. For instance, the *H. Pylori* metabolic network with 481 nodes has an average geodesic path length of 2.83 and a maximum geodesic path length of 7. For an undirected network of this size, such a high maximum path length suggests structural qualities not in keeping with other small-world networks. The clustering coefficients, which give a sense of cliqueness in the network, were found to vary significantly between metabolic pathways and Barabási-Albert scale-free networks. For the metabolic pathways investigated here the clustering coefficients were found to have an average value of $c = 0.22$ with a range of $[0.16, 0.32]$. The maximum geodesic length was found to have an average value of $G_{max} = 7.56$ in the range $[5, 11]$. The final measure examined here is the algebraic connectivity, which demonstrates the global connectivity of the network and was found to have an average value of $\lambda = 0.25$ in the range of $[0.09, 0.53]$. The results for the measures taken on all these networks can be seen in Table 4.1.

Metabolic Pawthay Network	Nodes	Maximum Geodesic Length	Mean Geodesic Length	Algebraic Connectivity	Clustering Coefficients
E. Coli with H_20	74	5	2.61	0.531	0.263
E. Coli without H_20	70	6	2.77	0.278	0.209
E. Coli iAF1260 with H_20	1339	7	2.92	0.254	0.261
E. Coli iAF1260 without H_20	1335	10	3.43	0.145	0.159
E. Coli iJR904 with H_20	764	8	2.87	0.092	0.251
E. Coli iJR904 without H_20	752	9	3.32	0.187	0.158
H. Pylori iIT341 with H_20	481	7	2.83	0.253	0.247
H. Pylori iIT341 without H_20	477	11	3.27	0.139	0.165
M. Barkeri iAF692 with H_20	622	8	2.78	0.182	0.286
M. Barkeri iAF692 without H_20	608	8	3.16	0.176	0.208
M. Tuberculosis iNJ661 with H_20	822	6	2.62	0.329	0.317
M. Tuberculosis iNJ661 without H_20	818	9	3.15	0.185	0.218
S. Aureus iSB619 with H_20	653	6	2.77	0.441	0.253
S. Aaureus iSB619 without H_20	629	7	3.17	0.301	0.167
S. Cerevisiae iND750 with H_20	755	6	2.82	0.348	0.254
S. Cerevisiae iND750 without H_20	751	8	3.30	0.204	0.166

Table 4.1: **Various measures for different examples of metabolic pathway networks**

A Size Dependent Generative Model

In this section we modify the BA model of preferential attachment to allow for the high clustering and low average path lengths found in the networks of metabolism. Although the BA model (Equation 2.4.1) does not provide good fits to the metabolic networks, a generative model using preferential attachment would seem to have much to offer when considering the growth of metabolic networks. Here metabolites which are present in higher numbers of reactants will be more likely to react with a new metabolite joining the network.

As such, a generative model similar to the BA model appears as a strong candidate for describing some of the features which develop in the growth of metabolic networks. However, as a variety of studies have demonstrated [123, 124, 129], the scale-free model alone is not sufficient to fully describe the networks when a range of metrics are considered.

By taking the BA model and simply increasing the probability of making a connection it was found that although this produced the high clustering found in these networks the average geodesic path length became significantly lower. The algebraic connectivity was also found to be significantly higher than found in the metabolic pathways. Here we present a model with non-static growth that allows for the high clustering yet also agrees with the algebraic connectivity and average geodesic path length found in the metabolic pathway networks.

Networks are grown according to the following:

$$P_i = \frac{K_i}{\sum_j (K_j)} \times \frac{\alpha}{n} \tag{4.1.1}$$

where P_i is the probability the new node attaches to the existing node, i, K_i is the degree of node i, $\sum_j (K_j)$ is the degree sum of all nodes in the network, n is the number of nodes and α is a constant.

Such a model will produce a globally connected network for $n^2 + 1 \leq \alpha$. For $n^2 + 1 > \alpha$ the preferential attachment model prevails, however, the probability of attachment is initially high. As the network grows and $\sum j \sim \alpha$, new nodes attach in a manner identical to the original BA model and for $\sum j > \alpha$ the network ceases to grow any further. (Note, this model assumes that nodes are not self connected and restricts the probability of any new attachment to $P \leq 1$, for $n^2 + 1 \leq \alpha$).

The effect of modifying the BA model has two effects: Firstly, the probabilities of attachment are not static and are rescaled as each new node enters the network. Secondly the probability that any new node will attach is decreasing as the network grows. Figure 4.1 shows the total number of edges in the network as nodes

are added to the network for the size dependent model along with the original BA model. It can be seen that for the size dependent network connections are made rapidly at the start which then begins to slow whereas the BA model makes connections in a linear fashion.

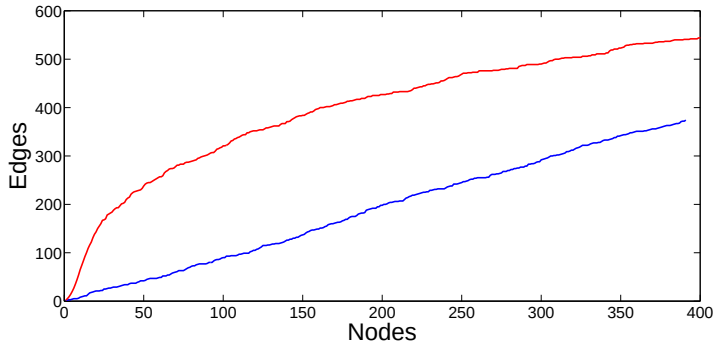

Figure 4.1: **Total number of connections as network growth model is iterated through.** *The red line shows the total number of edges as nodes attempt to be added to the network for the size dependent model in Equation 4.1.1 with* $\alpha = 125$ *grown to 250 nodes. The blue line shows the standard BA scale-free model also grown to 250 nodes.*

The size to which any network will grow in finite time, is essentially determined by the value α. When $n > \alpha$, the probability of new nodes attaching is very small and these new nodes form no edges with existing nodes, thereby not joining the network. Simulations were conducted until the network had grown to a specified size, attempting to attach new nodes until this was successful.

Repeated simulations for networks grown according to Equation 4.1.1 were performed for varying values of the constant α. Figures 4.2, 4.3, 4.4 and 4.5 show how graphical measures vary for a high α value of twice the network size and a low α value of a tenth of network size along with the metabolic pathway networks.

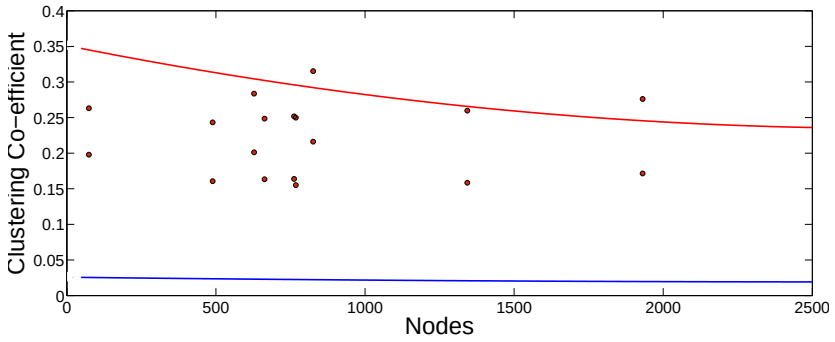

Figure 4.2: **Clustering coefficients for metabolic and artificially generated networks.** *Clustering coefficients for metabolic pathway networks (circles) and those of artificially generated size-dependent networks with both high (red) and low (blue) α values. It can be seen that the metabolic pathways mostly fit well within the upper and lower bounds for the artificially generated network*

It can be seen that the measures considered do not show a correlation with the network size. It can also be seen that the range of values the clustering coefficients and algebraic connectivity can take in the metabolic pathways is more varied than that of the average and maximum geodesic path lengths, which is echoed in the range of values the artificial model can take for high and low α values. These figures demonstrate that the range of values found for these networks fit within the bounds of a high and low α value for the artificial network model. This shows that the artificial model can be tuned to fit the various measures considered providing an artificial network model with similar topological features to these real world networks.

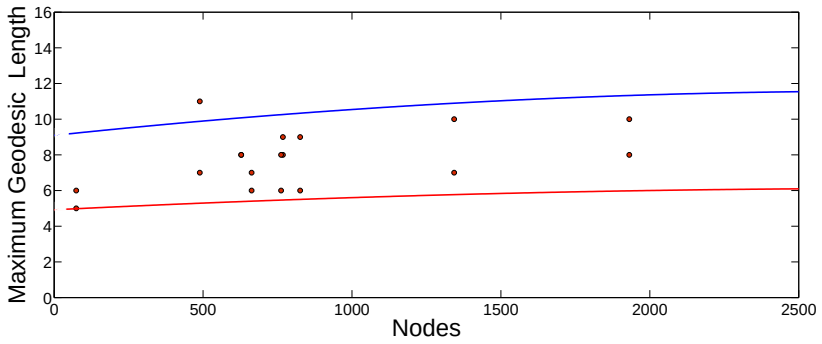

Figure 4.3: **Maximum geodesic path lengths for metabolic and artificially generated networks.** *Maximum geodesic length for metabolic pathway networks (circles) and those of artificially generated size-dependent networks with both high (red) and low (blue) α values. It can be seen that the metabolic pathways mostly fit well within the upper and lower bounds for the artificially generated network*

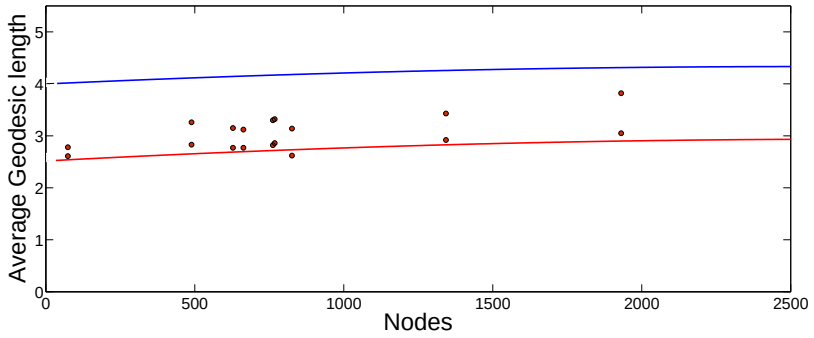

Figure 4.4: **Average geodesic path lengths for metabolic and artificially generated networks.** *Average geodesic length for metabolic pathway networks (circles) and those of artificially generated size-dependent networks with both high (red) and low (blue) α values. It can be seen that the metabolic pathways mostly fit well within the upper and lower bounds for the artificially generated network*

The overall effect of such a size dependent modification to the original BA model is to initially produce a highly connected 'hub' which then grows via preferential attachment giving rise to high clustering coefficients and a very strong small-world effect resulting in low average geodesic path lengths. In the size dependent

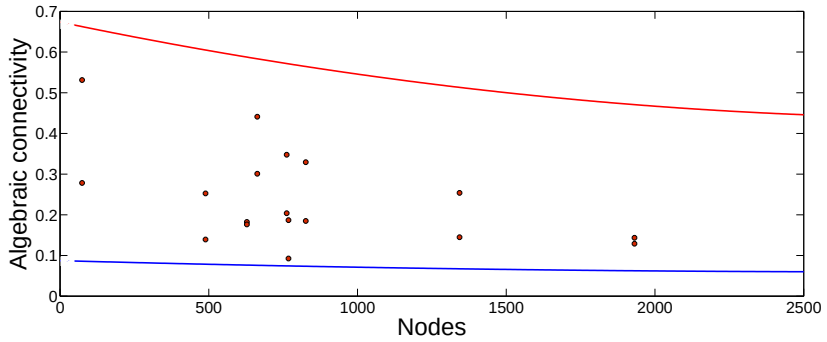

Figure 4.5: **Algebraic connectivity for metabolic and artificially generated networks.** *Algebraic connectivity for metabolic pathway networks (circles) and those of artificially generated size-dependent networks with both high (red) and low (blue) α values. It can be seen that the metabolic pathways mostly fit well within the upper and lower bounds for the artificially generated network*

model initially as n is small the probability of attachment is higher than the BA model and more highly connected nodes are formed. This initial component of highly connected nodes serves as the seeds for clusters with nodes continuing to connect to these with high probability throughout the growth. After this initial surge the new nodes still attach to a large number of others before this drops which produces a more highly clustered model than from the original BA scale-free model.

It is straightforward to amend the size dependent model 4.1.1 such that no globally connected initial stage occurs by rescaling the size dependent modification in Equation 4.1.1 to $C/n+\sqrt{\alpha}$. This produces a network which has a more pronounced scale-free structure, while retaining the high clustering coefficients of the original formulation.

4.1.4 Example - *S. Aureus*

In order to demonstrate the similarities between the metabolic pathways and the size dependent network growth model a specific comparison is now provided. It should be noted that all of the networks considered in this study were analysed at this level with each showing a strong similarity with the model. The analysis on the bacterium *Staphylococcus Aureus* gives the results of a metabolic network consisting of 629 nodes, a clustering coefficient of $c = 0.167$, algebraic connectivity of $\lambda = 0.301$, a maximum geodesic length of 7 and an average geodesic length of 3.17. Comparing this to a BA scale-free network of the same size, averaged over 100 trials gives a clustering coefficient of $c = 0.018$, algebraic connectivity of $\lambda = 0.150$, a maximum geodesic length of 9.70 and an average geodesic length of 4.35. Clearly this scale-free model does not fit any of the graphical measures well.

In order to produce a size dependent network as defined in Equation 4.1.1 the network size is fixed to that of the specific network being modelled and the value of α is tuned to allow the four measures considered to agree with that found in the metabolic pathways. Essentially the model is fitting four graphical measures using one parameter, α for a network of a specific size.

The (mean) average values for 100 trials of a network of 741 nodes, generated according to Equation 4.1.1 with $\alpha = 450$ are: clustering coefficient of $c = 0.160$, algebraic connectivity of $\lambda = 0.320$, maximum geodesic path length of $G_{max} = 7.80$ and mean geodesic length of $G_{mean} = 3.20$. The distribution of geodesic lengths over the whole of the *S Aureus* network and an example size dependent

network are shown in Figure 4.6 in addition to the degree distributions.

Figure 4.7 shows a graphical representation of the BA model, *S. Aureus* and the size dependent growth model to give an understanding of the structure of these network. It should be noted, however, that here are many possible methods for drawing these networks and using these to compare networks structure is of little use for networks of this complexity.

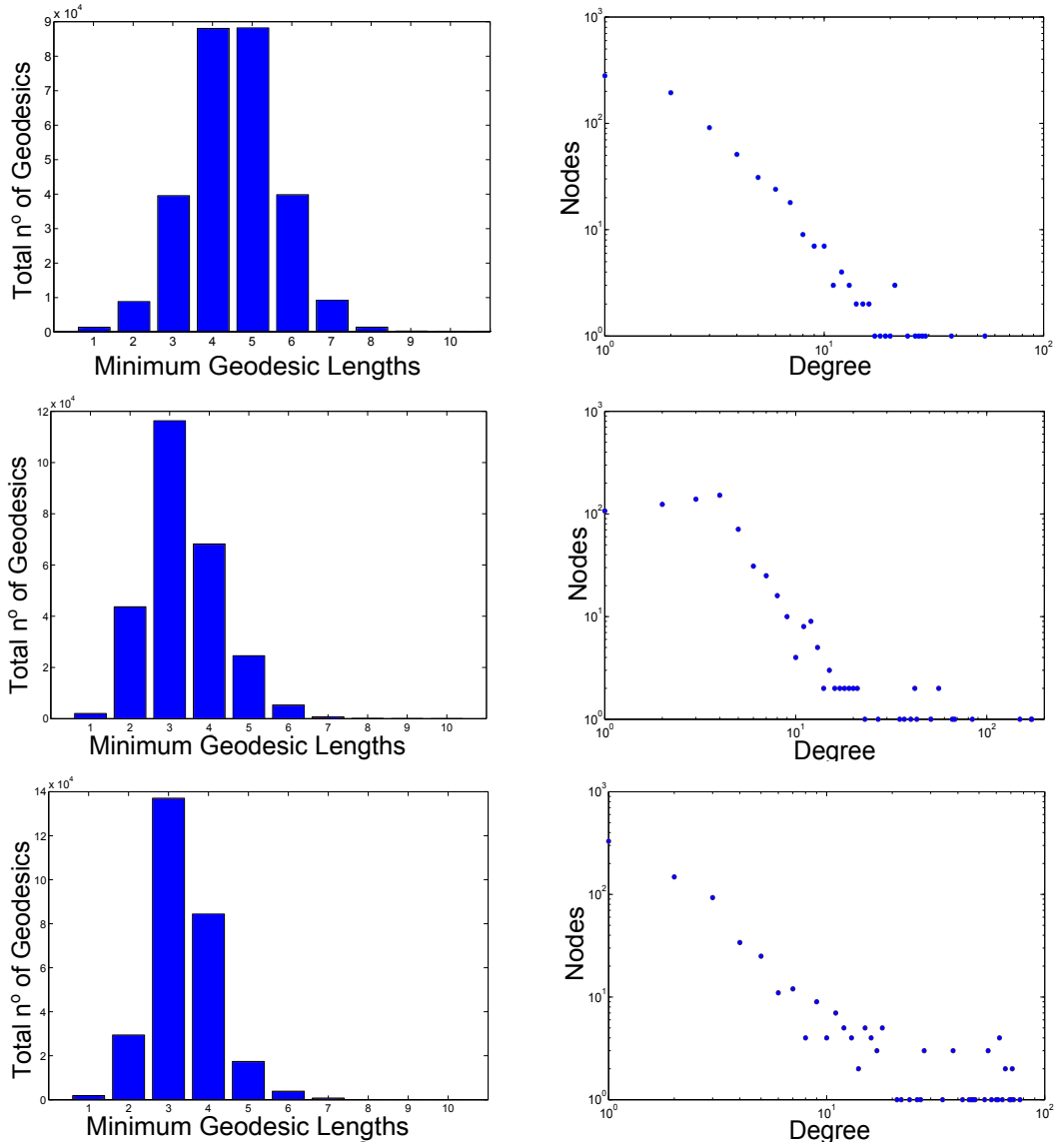

Figure 4.6: **Degree distributions and geodesic path length spectrum.** *Geodesic lengths (left) and degree distribution plots (right) for* S Aureus *(middle), a scale-free network (top) and size dependent network (bottom) all of* 741 *nodes. Note the geodesic distribution for the size dependent network matches closely the metabolic network. The degree distribution for the metabolic network is not entirely scale-free due to fewer nodes of degree* 1 *and* 2 *than would be expected. The size dependent growth model demonstrates a degree distribution which is in many respects closer that of the metabolic pathways.*

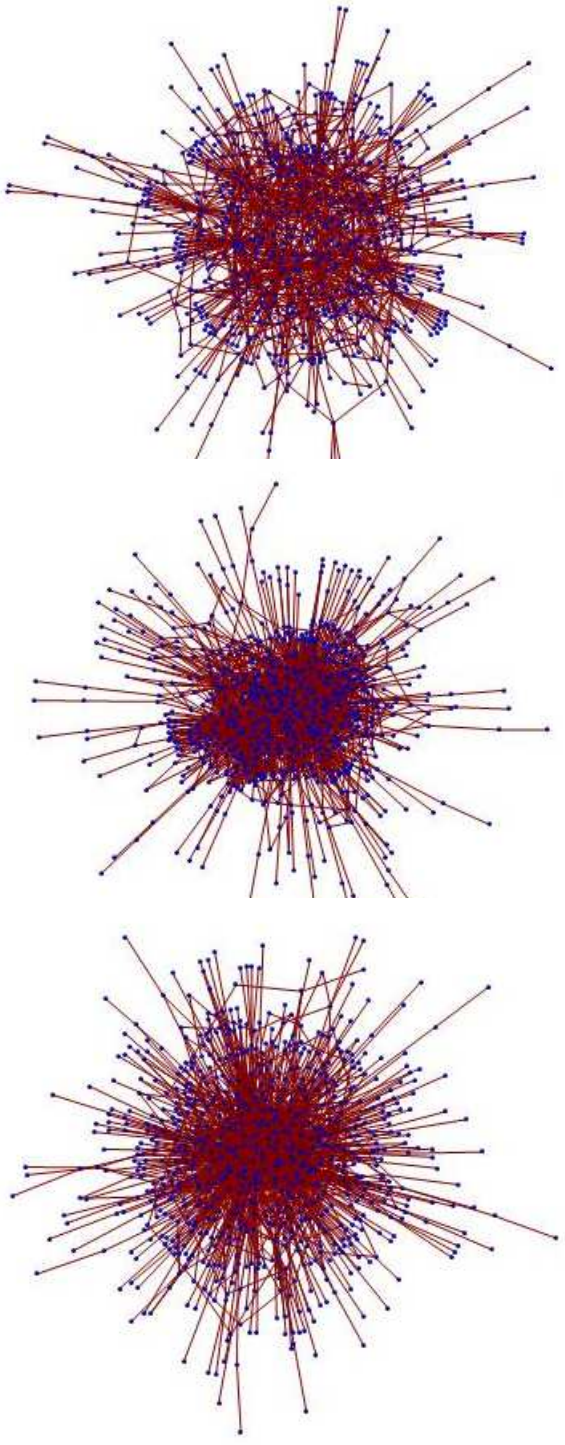

Figure 4.7: **Graphical representations of various networks** *Graphical representations of artificially generated Barabási Albert scale-free network (741 nodes, top), S Aureus (741 nodes, middle) and artificially generated size dependent network (741 nodes, bottom). All Figures are generated using the Mathematica spring embedding algorithm.*

4.2 Mass Transit Network Analysis

The final real world network to be examined is that of transportation, specifically mass transit networks, or underground railway networks. Underground railway networks are a feature of many developed cities. Understanding the structure of these networks can demonstrate important properties that these systems exhibit, from resistance to attacks, to providing the simplest way to increase connectivity. In this work the underground railway networks of New York and London will be examined, due to the differences in their geography. To analyse these networks an adjacency matrix first needs to be produced from the maps (see Figure 2.14). In the adjacency matrix two stations are considered connected if there exists a direct train connecting the two of them, and are given a value of one, and zero otherwise. From the two adjacency matrices measures can be taken on the network, as well as producing network images with all the geographical features removed (Figure 4.8 shows representations using spring algorithms in Mathematica). In the original maps some of the underlying geography has been removed, which allows the user to differentiate stations at the compacted centre, but the rough position of stations remains in the map.

The first step in analysing these networks is to calculate simple metrics which are given in Table 4.2.

The degree distributions (Figure 4.9) show that both these networks have a similar structure. With their low clustering and high average geodesic path length (Table 4.2) these networks would not seem to fall into any of the usual network

Table 4.2: **Various measures for two underground railway networks**

Measure	London Underground	New York Underground
Nodes	353	422
Maximum Geodesic Length	34	49
Mean Geodesic Length	11.8	16.7
Algebraic Connectivity	0.01	0.004
Clustering coefficient	0.024	0.02

categories. Having higher average paths than a random network of a similar connectance and higher average geodesic path lengths than the BA scale-free model. Similarly, they would not fit with the size dependent networks described here. A more in depth analysis will demonstrate certain topological properties that a standard network analysis can not detect, which will be explored in Chapter 5.

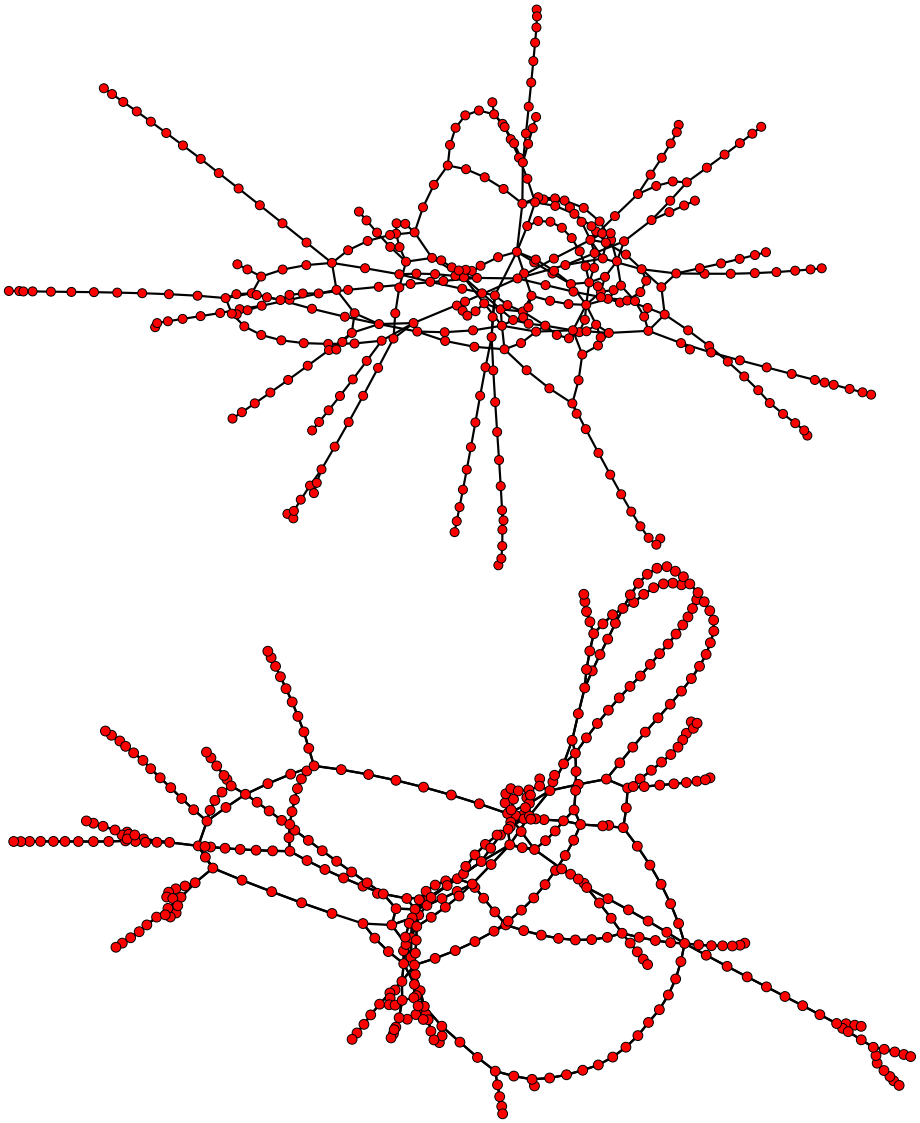

Figure 4.8: **Graphical representations of two underground railway networks.** *The London Underground (top) and New York Metro (bottom) networks, presented as non spatially-arranged graphs. Note the presence of two central clusters in the New York graph, which represent the concentration of stations in South Manhattan and Brooklyn. Both of these networks representations were generated using the Mathematica spring algorithm.*

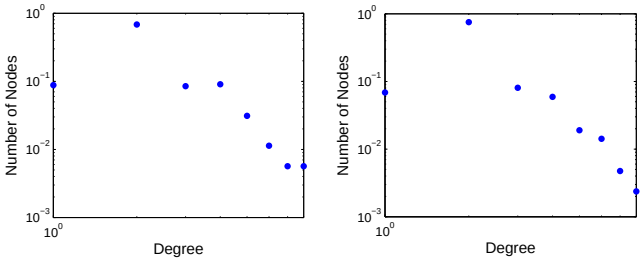

Figure 4.9: **Degree distribution for transport networks.** *Degree distributions for London (left) and New York (right) subway networks.*

4.3 Conclusion

In this analysis the structure of metabolic pathway networks have been investigated to give a network growth model which agrees with their topological properties. This investigation has demonstrated that metabolic networks of microorganisms are more accurately modelled with a network growth model in which network size is a modifying factor. Due to the size dependent networks having a densely connected cluster, these networks will have an increased resilience to targeted perturbations than that of the BA model, which are known to be drastically effected by this type of network attack [116]. This suggests that the structure of metabolic pathways gives them a greater resilience to node removal than if they were examples of scale-free networks, modelled by preferential attachment. Other representations of the metabolic pathways can also be modelled using an adjusted version of the growth model. The compartmentalised versions of the pathways could be modelled via smaller size dependent networks, joined together

to form highly clustered networks, representing the sub-cellular compartmentalisation seen in these formulations.

This network growth model can allow for man made networks to be produce which share the properties of these biological networks, for example the low average geodesic path length found in these networks is known to allow for the rapid transmission of information across the network [116]. This is also useful in testing how information will spread across the network, and in deducing how the topological properties are related to the actual properties of the networks [140]. This network growth model allows for various graphical measures to be estimated for other metabolic pathway networks. Comparing this to the measures taken on the actual network can then help in discovering similarities between them or in detecting networks which are actual very different from a topological perspective.

Analysing the structure of these real world complex networks has yielded an insight in to their properties and how they may function. Although an important step in understanding these networks, further insight can be gained by adding greater complexity to these representations by allowing the nodes to perform some function. This is addressed in this next chapter and it is this extension in the analysis will help to facilitate a greater understanding as to how these networks function in a more global sense.

Chapter 5

Detecting Topological Hierarchy Using Coupled Oscillator Models

In this chapter a novel method of hierarchy detection is derived in which the nodes are replaced by *oscillators* which interact with their neighbours. This has the affect of the network structure influencing the behaviour of the ensemble of oscillators, which allows for topological properties to be deduced [25, 103, 117, 153].

5.1 Introduction

The problem of relating the structure of a network to the dynamical behaviour it supports is of significant interest in many domains. Many different systems may

be represented as networks of connected entities, from friends communicating via social media [21, 108], to groups of neurons [155] and chemical reactions [77]. The fundamental issue addressed here is how to link the observed dynamics of a system to certain properties of its underlying network structure. The hope is that, by deepening our understanding of how particular types of network behave (in a global sense) over time, the structure of a network with a similar structure may be predicted. In addition, studying the recurring features of complex networks from a number of different disciplines, may help to gain a deeper, more over-arching theoretical understanding of network dynamics.

Early work in this area focused on the development of model systems, which were used to analytically study the onset of certain behaviours (such as oscillations) [87, 110] (see also [4, 142] for reviews). These model systems have been successfully applied in a number of different disciplines, including chemistry [101], ecology [113] and sociology [21]. Of particular interest are networks which possess some form of *community structure* [63, 66, 73, 92, 136]; These are generally characterised as having groups of nodes that are tightly knit (i.e. highly connected) with less dense connections existing between these groups [116]. Such structures are interesting because many 'real world' networks (e.g. social, biological, technological) are naturally partitioned into sets of loosely-connected 'communities', or 'modules' [102, 115, 124, 137, 163].

Moreover, this analysis is not restricted to networks which are static (i.e. the possibility that connections are added and removed and nodes update their state is also considered) since such structures capture the fact that links between individ-

ual nodes - and the properties of nodes - may change over time. Recent work [30] on community structure in dynamic networks has shown that allowing nodes to influence the state of other nodes facilitates the spontaneous emergence of dynamic equilibrium (that is, the community structure of the network remains stable, even as group composition changes over time) [96].

The idea of nodes influencing one another leads naturally to the notion of synchronisation. The ability of connected dynamic elements to synchronise their behaviour through interaction is ubiquitous (see [141] for a general introduction) and has profound implications for a wide variety of systems. For the purpose of this analysis, the situation where the connected elements are oscillators [110] is of particular interest, as their synchrony is observed in many settings, from the human heart [71] and brain [55], to insect locomotion [40] and novel forms of computation [10]. Previous work [9] has established a strong correlation between the connectivity of groups of nodes and the time required for oscillators to synchronise. However, given that full synchronisation does not (and, indeed, *should* not) occur in many networks (for example, the abnormal synchronisation in neurones is known to be a feature of epilepsy [96]) the possible relationship between structure and dynamical behaviour for oscillator networks where the coupling between oscillators is weak enough and the connectivity in the graph is sparse enough, such that synchronisation does not occur is considered. This chapter precisely addresses this question.

Network topology has a strong effect on the observed dynamics of oscillator networks [25, 103, 117, 153]. Previous work has mainly focused on whether a net-

work will synchronise, relating this to graphical measures such as the eigenvalues of the laplacian [117] or clustering coefficients [105]. This work suggests that the ability of an oscillator network to synchronise is enhanced by homogeneity in the distribution of connections [117].

Many complex networks have been shown to demonstrate periodic dynamics. Neural systems, for example, display modes of oscillation at particular frequencies and this has in turn been linked to the hierarchical organisation of the brain network itself [32].

In coupled oscillator networks with all-to-all coupling, oscillating waves of synchronization have been observed in systems with bimodal and trimodal frequencies [3, 5] and in systems of interacting populations of oscillators [112]. Such oscillations may also be observed in globally coupled oscillators, where there is both an excitatory and inhibitory component to the interactions, as observed in [11, 12]. However, in each of these cases the global oscillations are in some way attributable to the individual nodes in the network and not to the network structure itself.

In this chapter we demonstrate how the community structure of a complex network may actively drive periodic dynamics and that such periodic dynamics occur in real world networks. Section 5.2 demonstrates a new method of detecting hierarchy in networks using oscillator dynamics. This method is then applied to artificial networks with both two and three levels of hierarchy. The effects of decaying this hierarchical structure on the observed dynamics are also investigated, demonstrated the suitability of this method in detecting underlying hierarchy. The

final part of this analysis applies the method to the networks of metabolism and mass transit networks, with two examples from each, one with an underlying hierarchy and one without, demonstrating that this method can be used to determine pseudo-hierarchy in networks. This chapter is concluded with a discussion and suggestions for further work.

5.2 Methodology

In order to investigate the relationship between network structure and dynamics, a model system that is broadly applicable, but which supports a wide range of dynamical behaviours is required. This model also needs to be able to measure the global network dynamics in a way that readily admits analysis. The well-established Kuramoto model [85, 87, 88] meets all of these requirements and is widely used in related work [13, 47, 52, 72]. Please see Section 3.2 for an outline of this model.

The model describes a system of coupled oscillators described by ordinary differential equations (ODEs) where interaction terms between oscillators are connected according to the specific network topology. As the model is iterated over, the phase of each node is altered according to Equation 3.2.8 (Section 3.2).

Motivated, in part, by the realisation that many naturally-occurring networks have complex topologies, recent studies have been extended to systems where the pattern of connections is local but not necessarily regular [109]. Due to the complexity of the analysis, further assumptions have generally been introduced. For

example, it is usually assumed that the oscillators are identical. Obviously, therefore, in the absence of disorder, (i.e. if $\omega_i = \omega, \forall i$) there is only one attractor of the dynamics: the fully synchronised regime, where $\theta_i = \theta, \forall i$. This scenario suggests that, starting from random initial conditions, a complex network with a non-trivial connectivity pattern will exhibit the following behaviour: first, the highly interconnected units that form local clusters will synchronise; second, in a sequential process, increasingly large synchronised spatial structures will emerge, until, finally, the whole population is synchronised [9]. However, for many dynamical complex networks, synchronisation is neither realised nor desirable. In these instances, weakly coupled oscillators may display partial synchronisation or clustering, but not full synchronisation. More formally, Equation 3.2.8 can give rise to a variety of dynamical behaviours. For strongly coupled networks (those with high connectivity and coupling strength k) the phases of all oscillators quickly synchronise. With weak coupling, the oscillators appear to move randomly. Between these regimes, *partial* synchronisation is observed, where some oscillators are synchronised and others form clusters, but no global synchronisation is evident. For this analysis the Kuramoto order parameter is used as a measure of coherence over the network as a whole, Equation 3.2.9 Section 3.2 [85].

Section 5.2 demonstrates a novel method of detecting hierarchy in networks by observing the behaviour of oscillators coupled according to the adjacency matrix of said network. This will detail the methodology and how it is implemented, which can be outlined as follows:

- The adjacency matrix of the network under analysis is taken with each node being considered an oscillator.

- The natural frequency, ω, of the oscillator are normally distributed with a standard deviation of $\sigma(\omega)$.

- For a range of incremental values of $\sigma(\omega)$, for example 0.001 to 0.1 in increments of 0.001, a simulation is ran.

- On each simulation, the initial phase of each oscillator is randomly distributed in the range of 0 to 2π.

- In each of these simulations the range of coupling strengths to be searched through are chosen, for example 0.001 to 1 as well as the number of iterations for which the simulation will run, typically 20000+.

- On each of these iterations, the coupling strength is increased incrementally within this range. The change in phase of each oscillator calculated according to Equation 3.2.8 as well as the Kuramoto order parameter (Equation 3.2.9).

- From each simulation the Kuramoto order parameter, χ, can then be plotted. The range of coupling strengths, K, for which long range oscillations can be found for each $\sigma(\omega)$ value, if they are present.

- Using the particular value of K and $\sigma(\omega)$ a single simulation is ran, but with a fixed coupling strength through the iterations. The behaviour of the nodes over time can be seen by plotting the Kuramoto order parameter.

- The Fourier transform can then be performed on this signal to determine the modes of oscillation.

Results

The results of the experimental investigations are now presented. The over-arching aim is to show how global oscillatory behaviour may be related directly to the community structure of the underlying complex network.

Artificial networks

The first networks to be analysed are from two classes of graph; those with and those without any community structure. For example, consider the typical community structured graph in Figure 5.1. Given weak coupling, the dynamics of such a graph allow for the possibility of synchronisation within the smaller globally connected clusters, while the entire graph remains only partially synchronised.

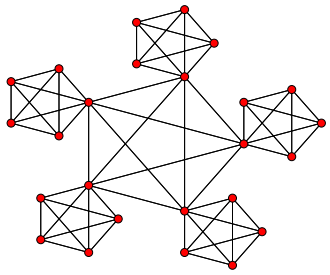

Figure 5.1: **Example graph with community structure (one level of hierarchy).** *For such a network there exist parameter regimes where the smaller, globally connected sub-graphs may synchronise but the network as a whole does not (partial synchronisation or clustering).*

As such, any global measure of synchronisation appears to oscillate (Figure 5.3) the oscillation being dependent upon the differences in the frequencies of oscillations between each of the clusters. These oscillations are present irrespective of the sampling interval used or even if the samples are irregularly spaced, demonstrating that the oscillations are not an artefact of the sampling procedure.

Figure 5.3 shows the Kuramoto order parameter oscillating between relatively low levels of synchronisation and almost full synchronisation. It should be emphasised, though, that the internal frequencies of the oscillators, ω, have been specifically selected in order to demonstrate such dynamics and that this will not occur in all cases. These values are chosen from the oscillating region found by varying the coupling strength as the model is iterated through, shown in Figure 5.2.

In graphs without any community structure, no discernible oscillation above that of the natural frequency of the oscillators are observed. Figure 5.4 shows the

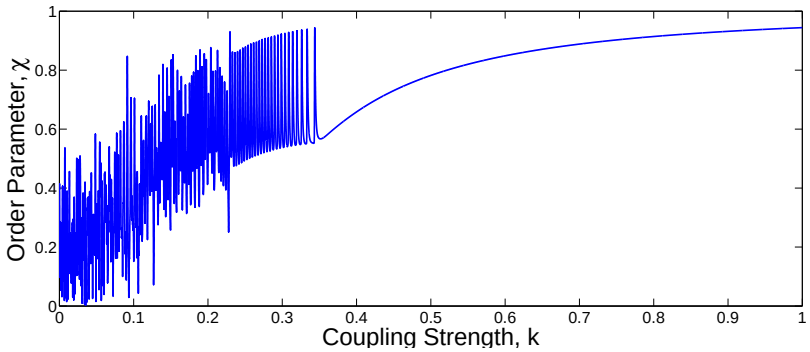

Figure 5.2: **Change in qualitative behaviour as measured by the Kuramoto order parameter, χ, for varying coupling strength (k).** *For the hierarchical network shown in Figure 5.1 with $\sigma(\omega)$=0.013.*

change on behaviour of the Kuramoto order parameter for varying coupling strengths for a complete (fully connected) and an Erdös and Rényi random network [56].

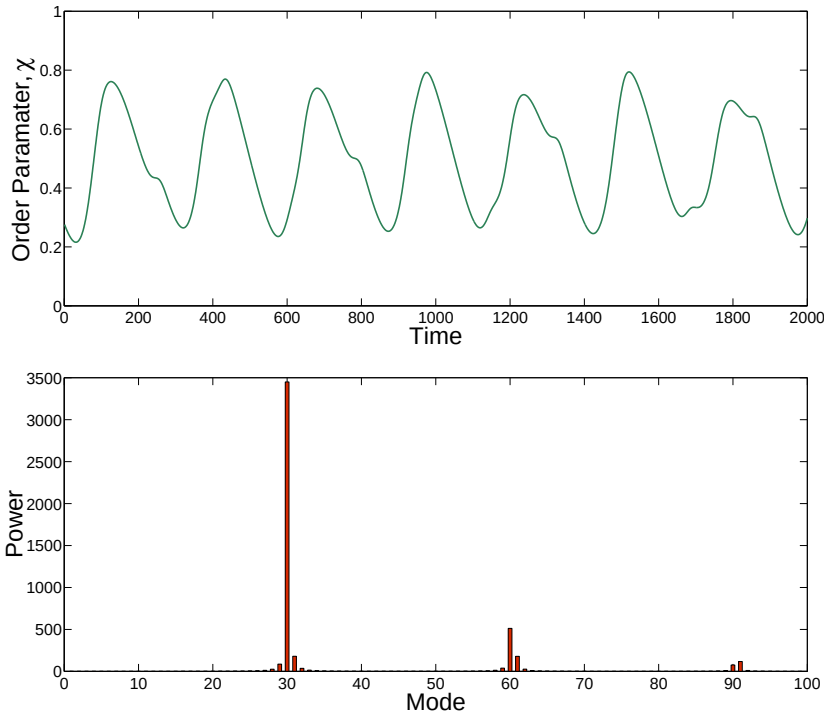

Figure 5.3: **Simulation and Fourier transform for the network in Figure 5.1.** *Time series for Kuramoto order parameter, χ (top), showing oscillatory dynamics for a network of Kuramoto oscillators coupled as in Figure 5.1 along with its corresponding Fourier transform (bottom). The coupling strength $k = 0.2$ and the frequencies are normally distributed with standard deviation $\sigma(\omega) = 0.02$. Note, for such parameter values it is possible to observe full synchronisation or oscillating dynamics as shown above depending on the individual frequencies of the oscillators. The example demonstrated here, although fairly typical, is not the only observable dynamics for such a network.*

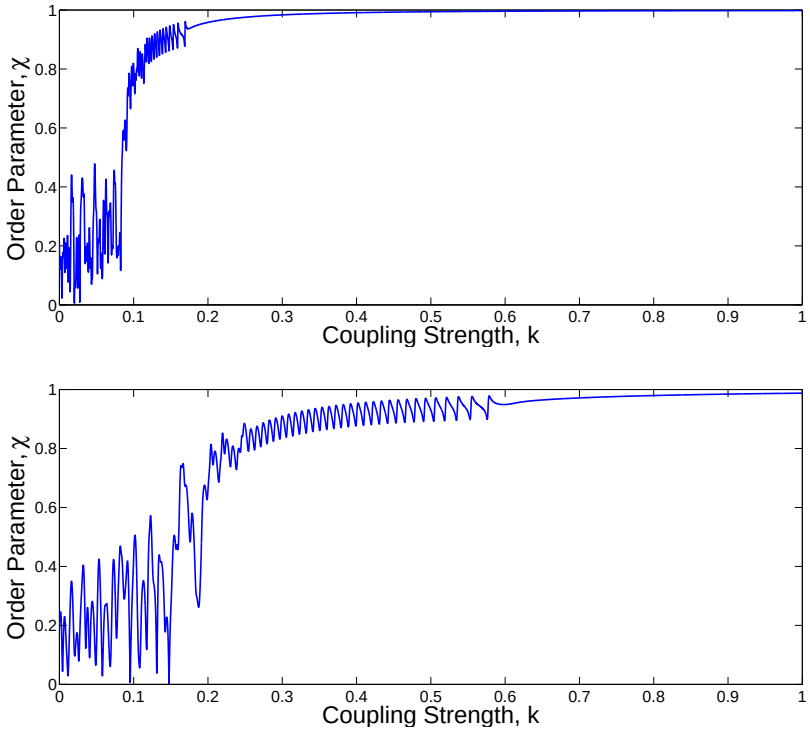

Figure 5.4: **Change in qualitative behaviour as measured by the Kuramoto order parameter, χ, for varying coupling strength (k).** *For a complete network with $\sigma(\omega)$=0.066 (top) and of a random network, $\sigma(\omega)$=0.036 (bottom).*

Figure 5.4(a) (top) shows a chaotic region where k is between 0 and 0.12 and then a region of short range oscillation before the network synchronises. This short range oscillation is cause by a single node not being synchronised with the rest of the system and oscillating its phase past the synchronised population, due to its initial frequency. Figure 5.4(b) (bottom) shows the change on behaviour of the Kuramoto order parameter for varying coupling strengths for a random graph, the same size as Figure 5.4(a) which shows similar behaviour. This demonstrates that oscillating behaviour occurs in networks of various classes, caused by a single node oscillating freely past the synchronised cluster. This short range oscillatory behaviour is not typical for random and globally coupled networks which mostly just move from chaos to synchronisation, however it is important to demonstrate that oscillating behaviour can occur in most types of graph, the type of oscillation produced is a feature that can be used to determine hierarchy.

The oscillatory behaviour in Figure 5.3 is caused by clusters synchronising due to the hierarchical nature of the network. This in contrast to the much studied Chimera states in which a part of the network synchronises with the remaining nodes remaining incoherent [2, 90, 91]. This clustered synchronisation can be demonstrated by observing the phase of each oscillator as the model is iterated, shown in Figure 5.5. It can be seen that most of the nodes have synchronized into a large cluster in which the phases of the nodes are all synchronised relative to each other, with a second cluster which also have their phases synchronised relative to one another. The oscillating behaviour can be seen to be caused by the phase of one cluster oscillating in relation to the phase of the other. This behaviour is determined by the topology of the network, in which the connectivity determines

whether clustered synchronisation will occur. This demonstrates that this large-range oscillation, caused by sub-populations forming, can be used to determine the presence of hierarchy in a network.

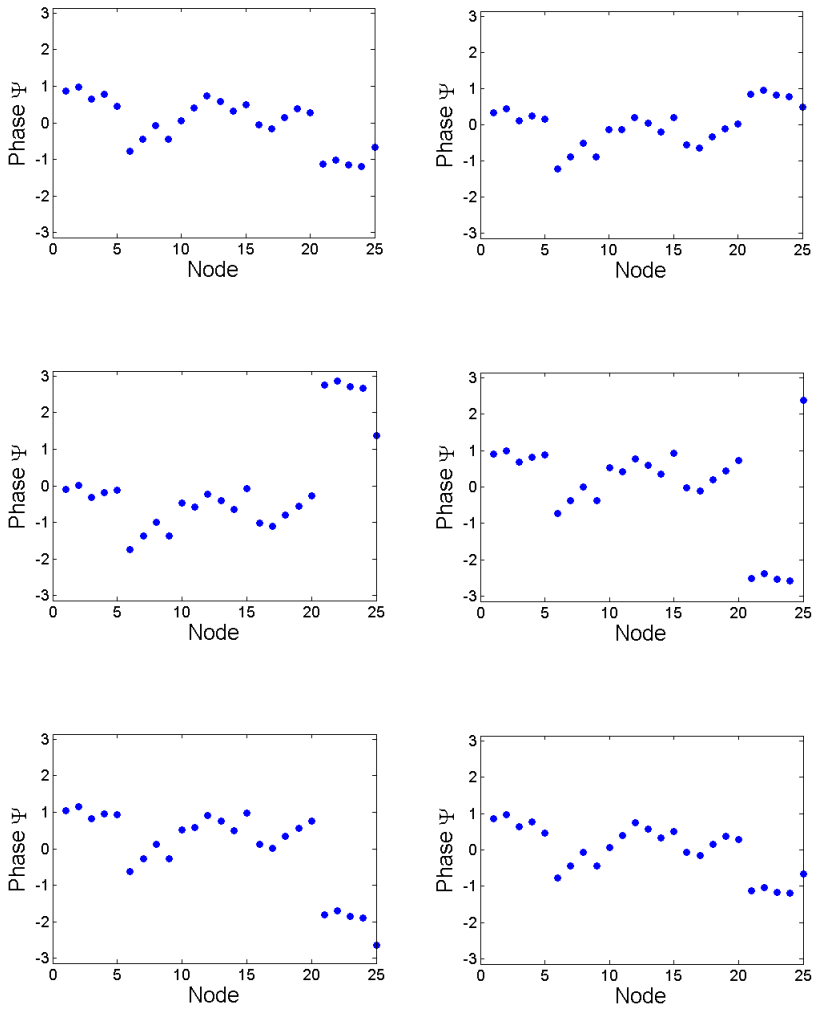

Figure 5.5: **Oscillators phase as the Kuramoto model is iterated through.** *Phase of nodes during Kuramoto model. Two clusters have formed, the twenty nodes on the right have synchronised with each other, as have the 5 nodes on the right, but the two clusters remain unsynchronised to each other, causing the regular oscillation seen in Figure 5.3.*

In order to demonstrate that the oscillating dynamics shown above are not simply an artefact of network symmetry, the original network is perturbed by repeatedly adding random connections. Figure 5.6 demonstrates the structural stability of the modal dynamics when the network structure is no longer symmetric, but the community structure is retained.

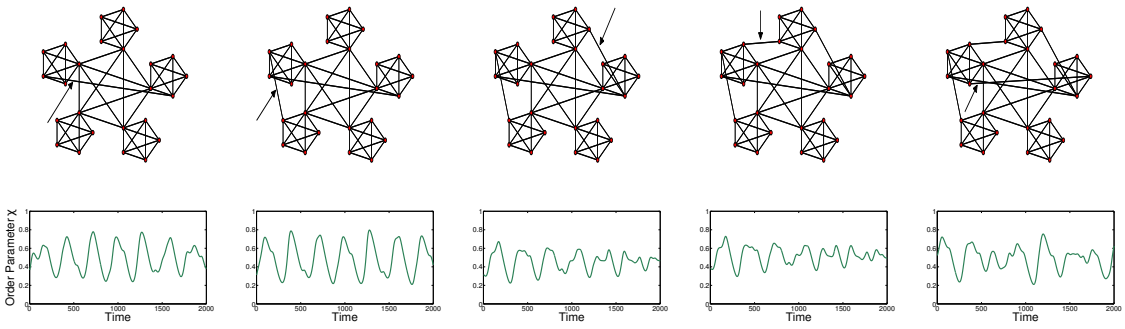

Figure 5.6: **Kuramoto simulation for increasingly perturbed hierarchical network.** *Time series of Kuramoto order parameter, χ, for increasingly perturbed hierarchical network of Kuramoto oscillators. New links are highlighted by arrows, demonstrating the robustness of dynamics to symmetry breaking, with $k = 0.2, \sigma(\omega) = 0.02$. Each simulation uses the same initial conditions and oscillator frequencies as in Figure 5.3. Similar observations occur whether the simulations are conducted as individual runs (as shown here) or with the network structure being perturbed as the simulation is performed. Note: Although the time series for the third and forth network (left to right) appear very similar they are simulations from their respective graphs.*

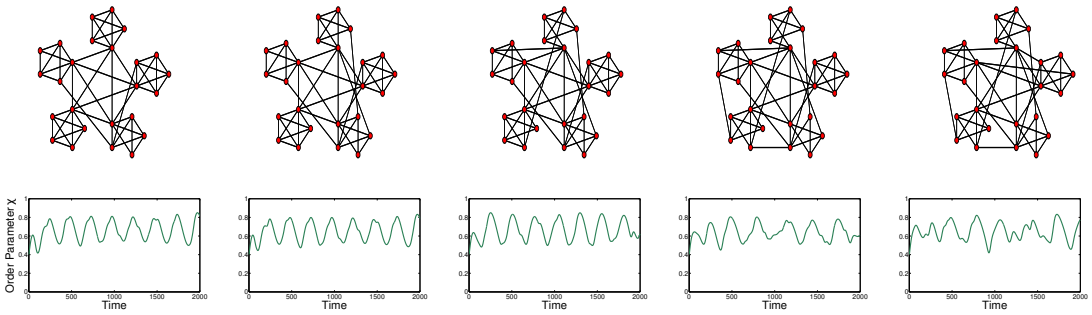

Figure 5.7: **Kuramoto simulations for rewired hierarchical networks.** *Time series for Kuramoto order parameter, χ, for various networks of coupled Kuramoto oscillators. Each network has been rewired using the Xswap rewiring algorithm which maintains the degree of each node. Two pairs of edges have been rewired from one graph to the next from left to right and it can be seen that the oscillating behaviour begins to break down as the hierarchical structure is decreased. $k = 0.2, \sigma(\omega) = 0.02$. Again the same initial conditions and internal oscillator frequencies are used as in Figure 5.3.*

Figure 5.6 demonstrates that the global oscillations observed are not due to symmetry of the graph structure. Although the asymmetric graphs no longer produce strong regular oscillations, the dynamics are not significantly affected by symmetry-breaking through the addition of connections. For this particular graph, it is possible to add a further 5 connections before the onset of global synchronisation.

Figure 5.7 shows another example of network rewiring, in this case using the Xswap algorithm [68], in which the network is randomised, with the degree of each node remaining constant. This is achieved by randomly selecting a pair of edges in the network, (i, j) and (k, l). If $(i, l) = (k, j) = 0$ then $(i, j) = (k, l) = 0$ and $(i, l) = (k, j) = 1$. It should be noted that in the unperturbed network the

nodes are self connected, so on some iterations these edges are swapped. The oscillations break down as the network is randomised demonstrating that it is the overall graphical structure that causes this behaviour.

To further develop the study of non-symmetric networks, an idealised network of oscillators arranged such that three highly coupled sub-networks of oscillators are connected via a sparse network of random connections are considered. For the following analysis, a network consisting of subgraphs of 45 oscillators with approximately 130 connections within each cluster is generated.

The effect of varying coupling strength, k, is investigated using standard techniques of increasing the coupling strengths as the model is iterated through. Figure 5.8 shows a typical example of the behaviour of the Kuramoto order parameter, χ as the coupling strength is increased from an initial value of $k = 0$ to $k = 1$. Here, the initial phases of the oscillators are drawn from a uniform distribution, $\theta_i \in [0, 2\pi]$. At each iteration of the simulation the value of k is increased in small increments, typically of around 2×10^{-6} for networks with varying random additional connections (see Figure 5.8).

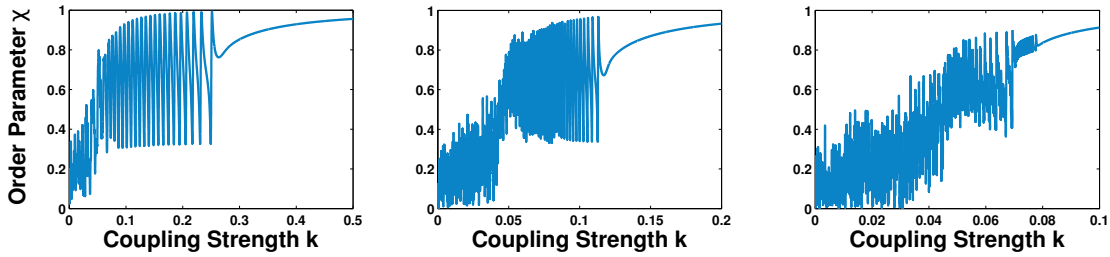

Figure 5.8: **Change in qualitative behaviour as measured by the Kuramoto order parameter, χ, for varying coupling strength (k) for networks composed of clustered random networks** *Each network contains* 3 *clusters of 45 randomly connected nodes with approximately* 130 *connections in each cluster. Here the frequencies are normally distributed with* $\sigma(\omega) = 0.01$. *(Left: 50 additional random connections over the whole network; Middle: 100 additional connections; Right: 150 additional connections. Note: The oscillatory regions indicate the parameter regimes where oscillatory behaviour will be observed..*

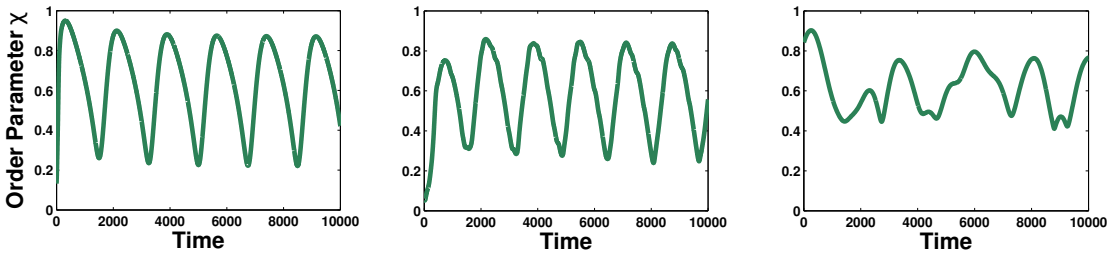

Figure 5.9: **Kuramoto simulations for networks of clustered random networks** *Time series for Kuramoto order parameter, with* $k = 0.4$ *(left),* 0.1 *(middle), and* 0.06 *(right) showing multi-modal dynamics. The simulations are for networks described in Figure 5.2 and the parameter values taken from the oscillating regions.*

In common with networks lacking community structure, these networks synchronise above a critical coupling strength and for small values of k, the oscillators are incoherent. In the first example there exists a specific region for $k \in [0.23, 0.6]$ for

which the Kuramoto order parameter, χ, appears to oscillate between the ordered and disordered state. Figure 5.9 shows the time series of the Kuramoto order parameter for the three networks described above, with a distribution of internal frequencies of 0.01 and respective coupling strengths of 0.4 (left), 0.1 (middle) and 0.06 (right).

A more complex network, which displays an additional level of hierarchy is now considered (Figure 5.10). For optimised parameter values of $\sigma(\omega) = 0.0012, k = 0.12$ multi-modal oscillations of the Kuramoto order parameter, χ, are observed, within a range of 0.1 to 0.5 (Figure 5.2). A Fourier spectrum of this time series demonstrates two modes of oscillation, at modes 22 and 46, with strong echoes at modes 66 and 93 (Figure 5.2). The relationship between these oscillating modes strongly mirrors the graphical structure of the network, in that the two levels of hierarchy cause a bimodal oscillation and therefore two peaks in the Fourier spectrum.

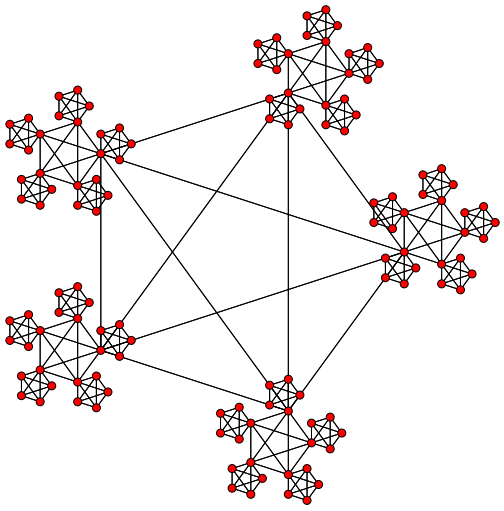

Figure 5.10: **Hierarchical network with two levels of hierarchy** *This network consists of five copies of the network in Figure 5.1 to produce two levels of hierarchy. For such a network there exist parameter regimes where the smaller, globally connected sub-graphs may synchronise but the network as a whole does not (partial synchronisation or clustering).*

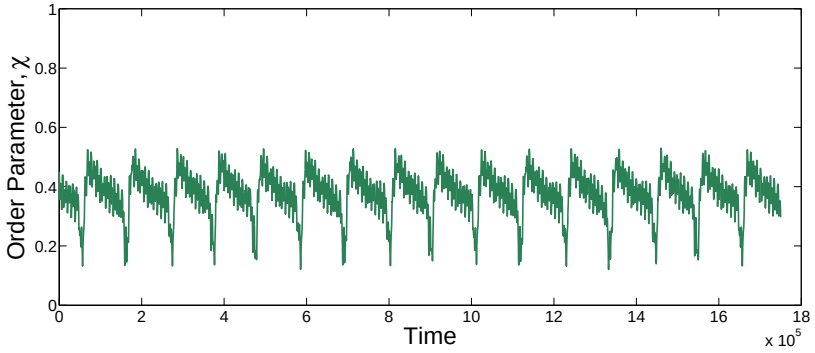

Figure 5.11: **Kuramoto simulation for a network with two levels of hierarchy** *With $\sigma(\omega)$=0.0012, k=0.12, showing bimodal oscillations.*

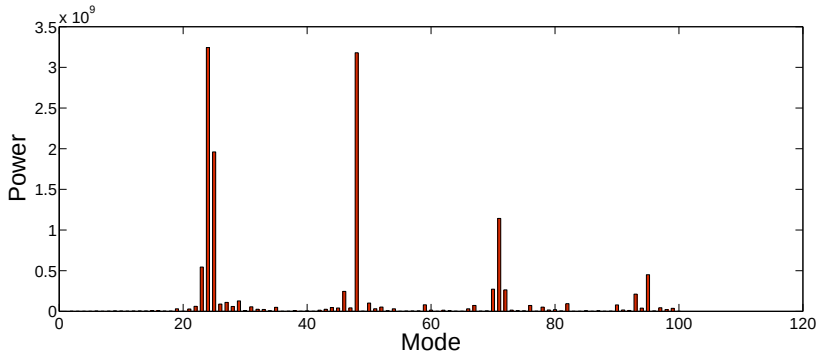

Figure 5.12: **Fourier transform of signal in Figure 5.2** *Two peaks can be seen at mode=23 and mode=47, followed by their respective echoes at mode=70 and mode=94 demonstrating a bi-modal oscillation*

'Real world' networks

In the previous section, the feasibility of using a global order measure to detect community structure in artificial networks has been established. This approach is now validated against two classes of 'real world' network, both of which present examples that may or may not possess community structure.

In order to provide a metric for comparison, the standard measure of Q *modularity* [97] is used. The measure Q is in the range $[-1/2, 1]$ and gives a measure of community structure. It is defined as the proportion of the edges that fall within any cluster, minus the *expected* proportion if such edges were distributed at random, outlined in Section 2.2. Other metrics for determining such modularity have also been proposed (see [6], for example) however for this analysis the most well known is used (the MATLAB program to calculate Q modularity was downloaded from VisualConnectome [45]).

Human metabolic network

The metabolic pathwaynetwork of a cell or microorganism describes the connections between various cellular processes that are essential for sustaining function [77]. Metabolic networks often exhibit strong community structure [123, 76, 145] and existing examples are usually examples of *pseudo-hierarchical* networks, in that their structure is not fully hierarchical [147]. In this section this new method of hierarchy detection is used to correctly identify community structure in metabolic networks.

The metabolic pathway networks used are in SBML format [75], taken from the BiGG database [127]. These are imported to MATLAB using libSBML [27]. In this analysis, the *Homo Sapiens Recon 1* (human) metabolic network is used, as this is perhaps the most interesting example available. Similar results have been observed on other metabolic networks formulated in a similar manner.

In order to establish a relationship between community structure and dynamics, two versions of this network are considered. The first comprises the global connectivity matrix of all chemical reactants in the cell, a connection being present if two or more components are involved in a known reaction (water and ATP are excluded, as these occur in almost all reactions). The second formulation of the metabolic network partitions reactions into sub-cellular networks, each representing different regions of the cell (nucleus, golgi bodies, etc.) which are connected in turn by reactions. Graphical representations of these networks are shown in Figure 5.13.

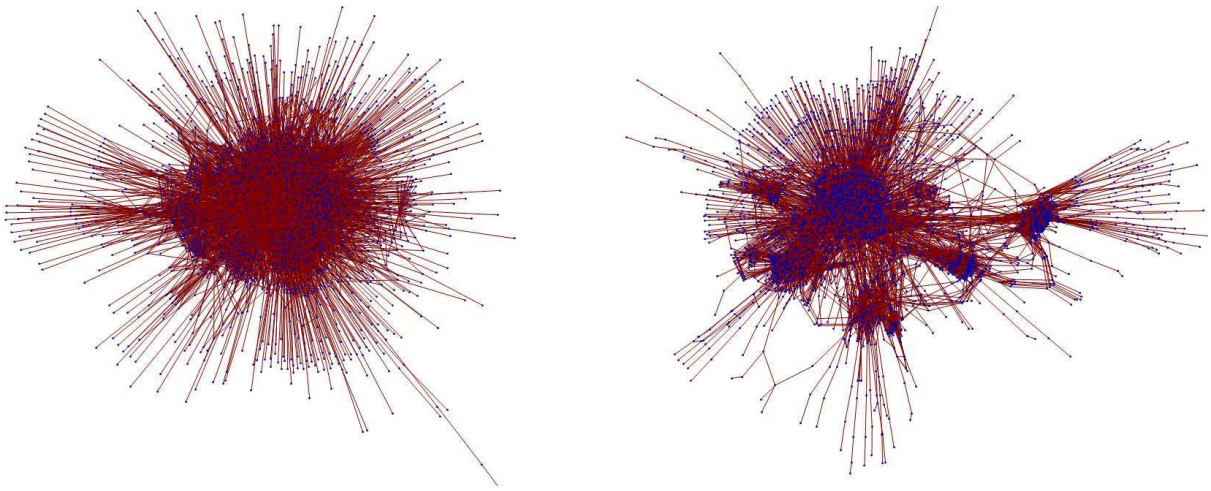

Figure 5.13: **Graphical representations of two versions of the same human metabolic network.** *Left: Non-partitioned representation of human metabolic network. Right: Partitioned representation of human metabolic network in which the network is partitioned into sub-cellular networks. The Mathematica spring algorithm is used to display the network structures, it can be seen that the two versions have a very different structure.*

From a graph theoretical perspective, these two networks are very similar. Standard graph metrics such as the clustering coefficient, mean and maximum geodesic path length do not distinguish between the two. Furthermore, the eigenvalue spectrum (as described in [9]) also shows no discernible difference.

The main difference between these two networks lies in the values for Q modularity, with the compartmentalised version having a value of $Q = 0.41$ and the non-compartmentalised version having a value of $Q = 0.29$. Due to the higher modularity of the compartmentalised version, regular oscillations would be expected when using this representation.

Simulations for optimised coupling strengths and frequency distributions are con-

ducted on both forms of the metabolic network. The coupling strength, as before, is taken from Figure 5.14.

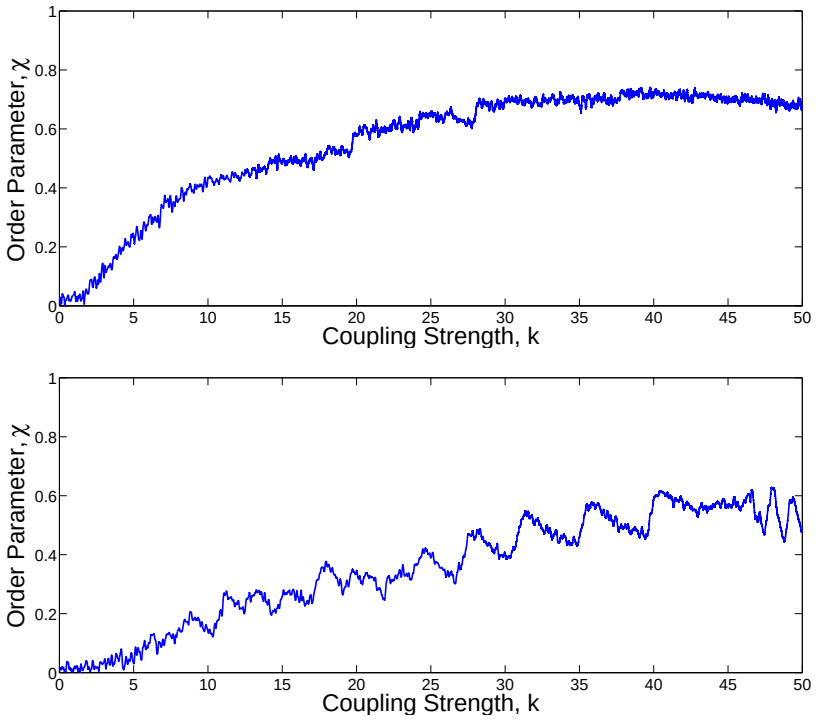

Figure 5.14: **Change in qualitative behaviour as measured by the Kuramoto order parameter, χ, for varying coupling strength (k).** *For the decompartmentalised (top) and compartmentalised (bottom) Homo Sapien metabolic pathway networks both with frequency deviation, $\sigma(\omega)$ =0.02*

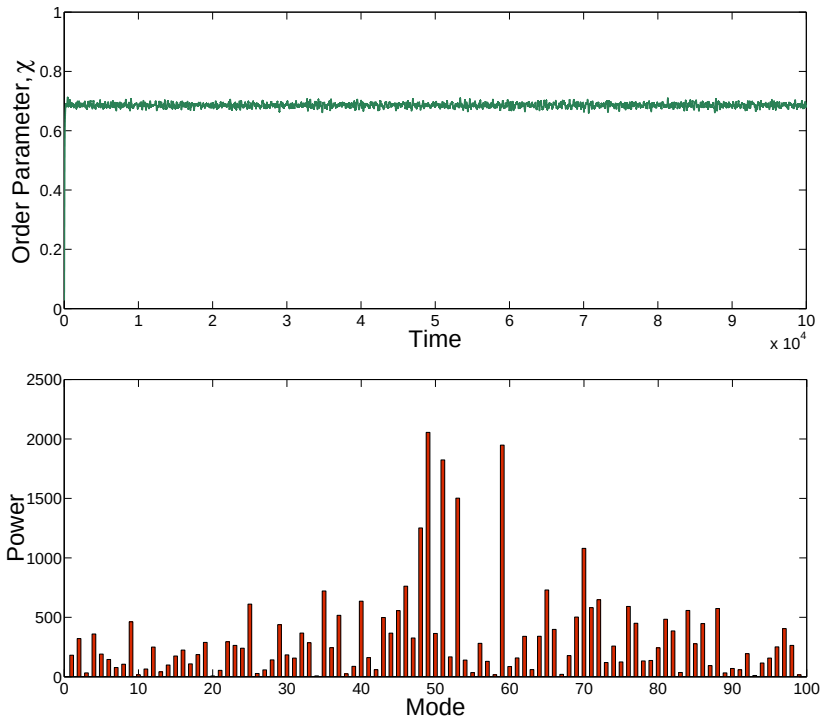

Figure 5.15: **Kuramoto simulation and corresponding Fourier transform for the network on the left in Figure 5.13.** *Kuramoto simulation (top) of decompartmentalised version of Homo Sapiens with* $\sigma(\omega)$*=0.02,* $k = 30.63$*. As no region of oscillation was found by increasing the coupling strength as the model is iterated through, the variables were chosen to be the same as that in Figure 5.16 for a comparison. The Fourier transform (bottom) shows no strong peaks, due to the irregularity of the signal*

For the non-partitioned network, multi-modal oscillations in the Kuramoto order parameter are not observed. However, for the partitioned network strong modal dynamics are observed (See Figures 5.15 and 5.16 for a comparison) which is consistent with the results for Q modularity. This demonstrates that this method of community detection is a viable method for use on complex real-world networks, where the underlying structure is not as regular as those formed using generative models.

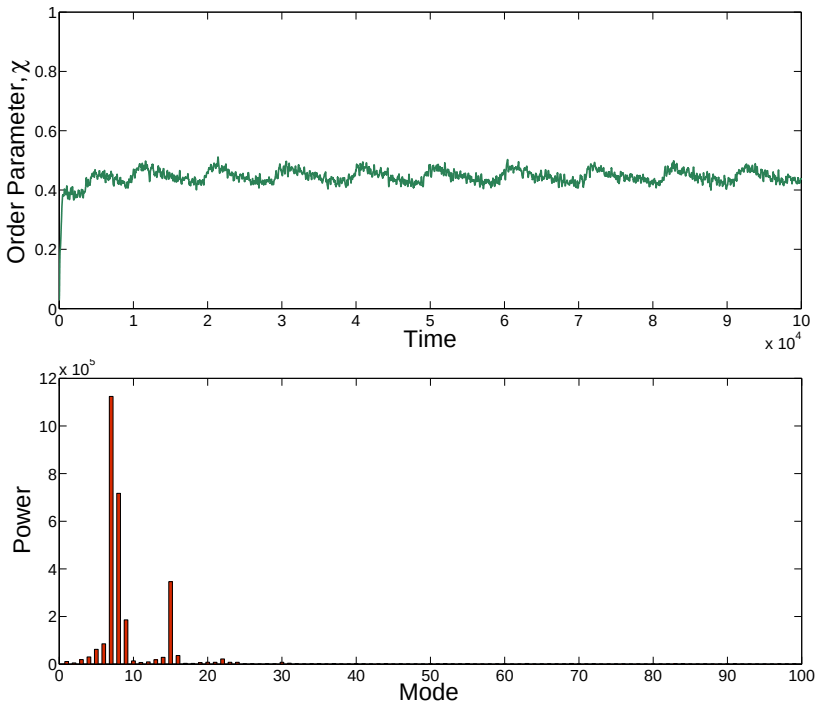

Figure 5.16: **Kuramoto simulation and corresponding Fourier transform for the network on the right in Figure 5.13.** *Kuramoto simulation (top) of compartmentalised version of Homo Sapiens with* $\sigma(\omega)$*=0.02,* $k = 30.63$*. These variables were chosen from the oscillating region shown in Figure 5.14. The Fourier transform (bottom) shows a strong peak at mode=7, followed by an echo at mode=15.*

Transport networks

Structurally, these networks are significantly different from the previous examples. Notably, there exist many long chains, the overall graph connectivity is low and there exists very few 'small-world' effects (see Section 4.2). As such, these networks present a novel challenge, over and above that offered by both the artificially-generated networks and the metabolic networks.

As before, numerical simulations are run in order to optimise model parameters,

in an attempt to maximise any oscillatory dynamics. On the London network, a small amount of oscillatory behaviour is observed, although the amplitude of such oscillation is small - the maximum observed oscillation has an amplitude of 0.05. The resulting Fourier spectrum has a peak strength of 2.8×10^6 (Figure 5.17).

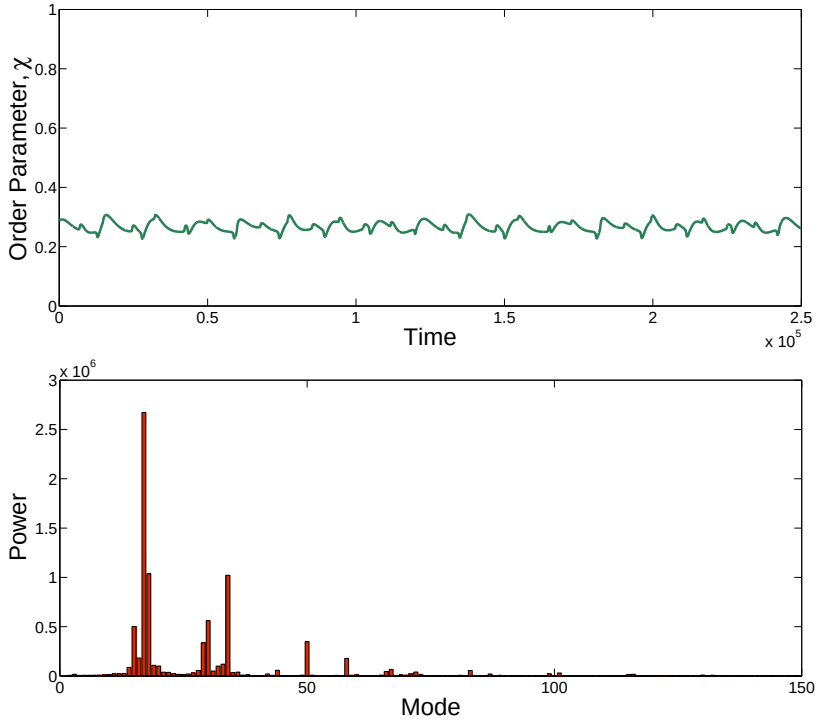

Figure 5.17: **Kuramoto simulation (top) and Fourier transform (bottom) for the London underground nework.** *With $\sigma(\omega)$=0.0021, K=3.6. These values were chosen to maximise the oscillatory behaviour.*

On the other hand, experiments on the New York network yield a *significantly* more pronounced oscillation, which displays very strong periodicity. The primary oscillatory mode has a strength of 8.5×10^7 - and a strong echo. A second oscillatory mode is also observed (Figure 5.18). In order to demonstrate that this oscillating behaviour is indeed caused by the underlying hierarchy of the network,

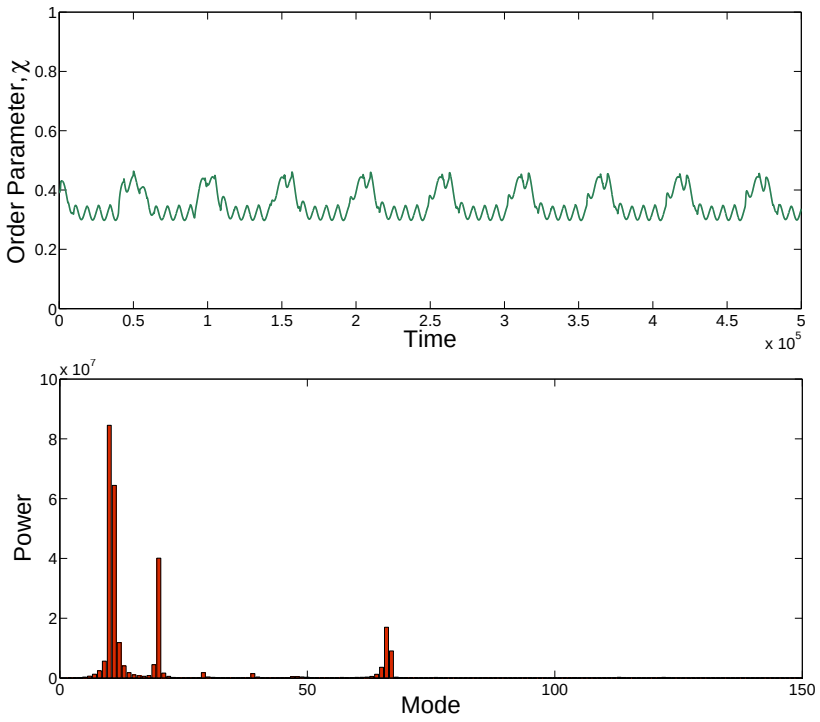

Figure 5.18: **Kuramoto simulation (top) and Fourier transform (bottom) for the New York subway network.** *With $\sigma(\omega)$=0.0028, K=6.5. These values were chosen to maximise the oscillatory behaviour.*

the New York subway network was rewired using the Xswap algorithm previously described. We observe that as the network is rewired and the modularity reduced to below $Q \approx 0.4$, oscillations no longer occur. As the Xswap algorithm maintains the degree distribution of the network but reduces modularity, this precisely demonstrates that modularity directly causes the oscillations in the Kuramoto order parameter of the phase.

Conclusion

This chapter has demonstrated a robust and structurally stable relationship between form and function in complex networks whereby global oscillations are shown to be a factor of network topology. Modal oscillations in a measure of global synchronisation are observed which can be directly related to the structure of the network itself. As this method relies on the overall behaviour of the network it allows for the presence of an underlying hierarchical structure to be detected in complex networks.

By applying the method to two types of real world networks - whereby examples exist with significantly different community structures but with similar underlying topology, this method is also shown to work on realistic, more irregular structures. The breakdown in oscillatory behaviour when networks are rewired (with the degree of each node remaining constant) has also been demonstrated. This confirms that network modularity drives oscillations, as reducing the degree of modularity causes these oscillations to break down. It should be noted, however, that for the real world examples given, the underlying dynamics of the nodes on the network (chemical reactions and subway trains) are considerably more complex than the simple Kuramoto oscillators used to demonstrate the principle. As such, it is not possible to directly attribute any observed oscillatory dynamics in such systems to the network structure alone.

Many real world networks (e.g. Transport, the brain) are examples of pseudo-

hierarchical networks, in that their structure is not fully hierarchical [147]. In the particular example of the brain, for instance, multi-modal oscillations (observed as Gamma ($> 30(Hz)$ Beta $(13-30Hz)$ and Alpha $(8-12Hz)$ waves etc. in EEG measurements) may attribute to structural hierarchies in the neural connectivity. As such, *for systems where the dynamics of the individual elements of a complex network are known to be unimodal and the interactions between them are likewise,* global observations of oscillatory behaviour may give some indication as to underlying structures and network connectivity, yielding novel methods of community detection.

Chapter 6

Discussion

In this thesis an analysis on real world networks has been presented which has given significant insight into their structure. The main contributions of this thesis are:

- A graphical analysis has been conducted on the real world networks of metabolism and underground rail networks, demonstrating common features of each type of network.
- A novel network growth model has been described which closely matches the features seen in the metabolic networks investigated.
- A novel method of hierarchy detection has been derived which can be used to perform a more in depth analysis on networks.
- This novel method has then been applied to the networks of metabolism and

underground railway networks, demonstrating the durability of this method in detecting hierarchy in real world networks, as well as giving an insight into their specific structure.

The graphical analysis has demonstrated that the metabolic networks investigated are more accurately modelled as networks which are restricted in their growth by the number of nodes in the network. This size dependent model has the advantage over other growth model, such as the BA scale-free model, in that networks can be grown to the same size which have different topological properties, a feature seen in many of the real world networks analysed. This allows for a network model which can be tuned to fit with specific real world networks using a variable which in this work has been arbitrary but could be expanded on to be related to some property of the network. In the case of the metabolic pathways the type of network representation can indicate which value this variable should take, producing a more accurate network model. For example, if the particular representation used is known to increase the clustering observed in the network (such as in the case of compartmentalised representations) then a higher value for this constant can be used to give a more appropriate network model.

Further work in this area would be in expanding the model to also produce weighted and directed networks which would allow for a more complex network representation to be modelled. As recent work has demonstrated that many networks thought to be scale-free are not of this class it would also be worthwhile to test if they have the characteristics of the size dependent model described here.

One of the advantage of having a network growth model which fits closely with the graphical measures found in certain networks is that it allows for man made networks to be designed to have similar features. As structure directly effects the behaviour of networks we can emulate a particular desirable behaviour. For example, the high clustering and low mean geodesic length allows for information to travel efficiently through the network, which can now be emulated in man made networks by using this network growth model. The artificial network model can also be used to test how the networks behave under certain conditions, such as targeted removal of nodes. Rather than just testing the actual networks, we can also generate artificial networks of the same size with a similar structure to allow for more comprehensive analysis.

The second part of this work was in the field of oscillator dynamics, in which the Kuramoto model was applied to networks to reveal hierarchical structures. This method of hierarchy detection utilises the dynamic behaviour observed in these networks, which is caused by the underlying structure. This method was first applied to artificial networks to demonstrate the concept in networks with known hierarchy. By using networks with a known structure the dynamical effects observed on the network can be related to the structure. In order to test the robustness of this method it was applied to networks which are constructed from clusters of randomly generated networks. The importance of this step was due to the fact that real world networks do not tend to be regular, but have an inherit randomness to them so any method of detecting hierarchy in real world networks needs to be capable of detecting an underlying hierarchy (or pseudo-hierarchy). The method was also applied to increasingly perturbed networks which as well as demonstrat-

ing that the method is suitable for detecting pseudo-hierarchy, also shows how the dynamic behaviour breaks down as the structure breaks down. In contrast to previous research, this method has also been applied to real world networks which demonstrates its ability to detect underlying structural properties of complex networks, allowing for a wider range of applications. By using examples from both biological and man made networks where there is some indication as to possible differences in their structure this dynamical behaviour can be seen to be a direct cause of topology.

An important advantage of this method of hierarchy detection is the fact it can be used to determine topological properties without knowing the network structure. In effect, if a network has an oscillating output this could be attributed to an underlying hierarchical structure. One example where this method may be of particular importance is in neurology. As the brain is known to exhibit regular waves of oscillation these may be attributed to the structure of the brain itself. This would allow for abnormalities in these oscillating waves to be directly linked to the physical structure of the brain, allowing for the detection of certain illnesses from the brain waves exhibited. For example, in the case of epilepsy where the seizures are known to be caused by abnormal synchronisation of neurons [96], it may be possible to attribute these to a specific problem in the topological structure of the brain. Many real world network have been classified as scale-free networks, however they may exhibit the hierarchical properties seen in the network examined here. One example where this may be the case is in social network, although preferential attachment may seem a viable method for creating networks with similar properties it is probable that they will have an underlying hierarchy, due to the

way these network develop, when someone joins a social network they tend to be joining their own social network with that of the new acquiescence, creating a highly modular network. This could be easily tested for by taking a large sample of a social media network and applying the method described here, which would allow for the structure to be detected. This would not only be helpful in our understanding of how these networks are structured but could also have applications in epidemiology. By having a better understanding of the structure of a social network we can gain a deeper understanding as to how many diseases spread across these networks and how best to control this.

The results from both sides of the investigation have demonstrated that some of the networks investigated, namely the compartmentalised versions of the metabolic pathways, showed properties of being both size dependent networks and having an underlying hierarchical structure. This would suggest that a more accurate network model for these networks would be small size dependent networks growing and then coalescing with other networks to form networks with a hierarchical structure as well as having size dependent features. Another important finding was that the network representation format used for the metabolic pathways has a dramatic effect on the structure they posses. This can lead to the same metabolic pathways being classified as different topological structures which can drastically alter our understanding of how they behave. This demonstrates the importance of using the appropriate representations when analysing these biological networks. The underground railway networks were not found to be examples of size dependent networks, or any other standard networks models. This is not a surprising result as these networks began as separate single lines which have since been

joined up, very different from standard network models. A dynamical analysis did reveal that the two different underground railway systems investigated have a very different topological structure, probably down to differences in the underlying geography of their location.

The motivation for this work was to derive a method of hierarchy detection which could be used for the pseudo-hierarchical real world networks which exists in many fields. Further work in this area should focus on deriving a measure for this type of hierarchy, building on the method described for detection. If a general measure of hierarchy could be determined, the differences in how these networks function could be inferred, which may help to explain differences in their functionality. Deriving more information from these oscillatory patterns may also be possible, such as the size of clusters and the strength of their connection to one another which could help to develop this method into a more useful tool. In some real world networks too much complexity is lost when using unweighted undirected representations. Further work would be in developing this method to also be applicable to weighted directed networks which could be achieved by having individual coupling strengths for each connection.

This analysis has demonstrated a new network growth model which shares features of metabolic pathway networks, and may be applicable to others. This gives a great insight into how these networks develop and how their growth may be constrained. Deriving a novel method of hierarchy detection has not only given more insight into the networks examined here but also allows for the presence of hierarchy to be detected in any real world networks, allowing for a more concrete

view on the structure of many complex networks previously examined in more standard topological ways.

Bibliography

[1] J Abello, PM Pardalos, and MGC Resende. On maximum clique problems in very large graphs. *External Memory Algorithms and Visualization*, 50:199–130, 1998.

[2] D.M. Abrams, R. Mirollo, S.H. Strogatz, and D.A. Wiley. Solvable model for chimera states of coupled oscillators. *Physical Review Letters*, 101(8):84103, 2008.

[3] JA Acebrón, LL Bonilla, S de Leo, and R Spigler. Breaking the symmetry in bimodal frequency distributions of globally coupled oscillators. *Physical Review E*, 57:5287–5290, 1998.

[4] J.A. Acebrón, L.L. Bonilla, C.J.P. Vicente, F. Ritort, and R. Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137, 2005.

[5] JA Acebrón, A Perales, and R Spigler. Bifurcations and global stability of synchronized stationary states in the kuramoto model for oscillator populations. *Physical Review E*, 64(1):016218, 2001.

[6] Rodrigo Aldecoa and Ignacio Marín. Deciphering network community structure by surprise. *PloS One*, 6(9):e24195, 2011.

[7] Gerald Alexanderson. About the cover: Euler and Königsberg's bridges: A historical view. *Bulletin of the American Mathematical Society*, 43(4):567–573, 2006.

[8] Panagiotis Angeloudis and David Fisk. Large subway systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, 367:553–558, 2006.

[9] A. Arenas, A. Diaz-Guilera, and C.J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11):114102, 2006.

[10] P. Ashwin and J. Borresen. Discrete computation using a perturbed heteroclinic network. *Physics Letters A*, 347(4-6):208–214, 2005.

[11] Peter Ashwin and Jon Borresen. Encoding via conjugate symmetries of slow oscillations for globally coupled oscillators. *Physical Review E*, 70(2):026203, 2004.

[12] Peter Ashwin, Gábor Orosz, and Jon Borresen. Heteroclinic switching in coupled oscillator networks: Dynamics on odd graphs. In *Nonlinear Dynamics and Chaos: Advances and Perspectives*, pages 31–50. Springer, 2010.

[13] Salvatore Assenza, Ricardo Gutiérrez, Jesús Gómez-Gardeñes, Vito Latora, and Stefano Boccaletti. Emergence of structural patterns out of synchro-

nization in networks with competitive interactions. *Scientific reports*, 1, 2011.

[14] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[15] A.L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1-4):69–77, 2000.

[16] A.L. Barabási and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[17] A.L. Barabási, E. Ravasz, and T. Vicsek. Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 299(3-4):559–564, 2001.

[18] LÁSZLÓ Barabási, A.L and Eric Bonabeau. Scale-free. *Scientific American*, 2003.

[19] R.J. Barsanti and J. Gilmore. Comparing noise removal in the wavelet and fourier domains. In *System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium on*, pages 163–167, March 2011.

[20] D.A. Beard, S. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83(1):79–86, 2002.

[21] P.S. Bearman, J. Moody, and K. Stovel. Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1):44–91, 2004.

[22] M.M. Bianchi. Collective behavior in gene regulation: Metabolic clocks and cross-talking. *FEBS Journal*, 275(10):2356–2363, 2008.

[23] Norman L Bigg, Edward Keith Lloyd, and Robin James Wilson. *Graph Theory: 1736-1936*. Oxford University Press, 1976.

[24] Peter Bloomfield. *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.

[25] S Boccaletti, M Ivanchenko, V Latora, A Pluchino, and A Rapisarda. Detecting complex network modularity by dynamical clustering. *Physical Review E*, 75(4):045102, 2007.

[26] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[27] B. J. Bornstein, S. M. Keating, A. Jouraku, and M. Hucka. LibSBML: an API library for SBML. *Bioinformatics*, 24:880–881, Mar 2008.

[28] R. Boulet, B. Jouve, F. Rossi, and N. Villa. Batch kernel SOM and related Laplacian methods for social network analysis. *Neurocomputing*, 71(7-9):1257–1273, 2008.

[29] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1-6):309–320, 2000.

[30] John Bryden, Sebastian Funk, Nicholas Geard, Seth Bullock, and Vincent AA Jansen. Stability in flux: community structure in dynamic networks. *Journal of The Royal Society Interface*, 8(60):1031–1040, 2011.

[31] F. Buckley and M. Lewinter. *A Friendly Introduction to Graph Theory*. Prentice Hall, 2003.

[32] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[33] Kenneth L Calvert, Matthew B Doar, and Ellen W Zegura. Modeling internet topology. *Communications Magazine, IEEE*, 35(6):160–163, 1997.

[34] Andrea Capocci, Vito DP Servedio, Guido Caldarelli, and Francesca Colaiori. Detecting communities in large networks. *Physica A: Statistical Mechanics and its Applications*, 352(2):669–676, 2005.

[35] Stephan C. Carlson. The Encyclopedia Britannica: Königsberg bridge problem @ONLINE, 2013. URL: `http://www.britannica.com/EBchecked/topic/321794/Konigsberg-bridge-problem`.

[36] José Manuel Casado. Synchronization of two Hodgkin–Huxley neurons due to internal noise. *Physics Letters A*, 310(5):400–406, 2003.

[37] Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

[38] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[39] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Breakdown of the internet under intentional attack. *Physical review letters*, 86(16):3682, 2001.

[40] James Collins and Ian Stewart. Hexapodal gaits and coupled nonlinear oscillator models. *Biological Cybernetics*, 68(4):287–298, 1993.

[41] James J Collins and Carson C Chow. It's a small world. *Nature*, 393(6684):409–410, 1998.

[42] E. Mary Cooke. *Escherichia Coli and Man*. Edinburgh : Churchill Livingstone, 1974.

[43] Gábor Csányi and Balázs Szendroi. Fractal-small-world dichotomy in real-world networks. *Physical Review E*, 70(1):16122, 2004.

[44] CS Cummins and H Harris. The chemical composition of the cell wall in some gram-positive bacteria and its possible value as a taxonomic character. *Journal of General Microbiology*, 14(3):583–600, 1956.

[45] Huiguang He Dai Dai. Visualconnectome: Toolbox for brain network visualization and analysis (abstract), human brain mapping 2011 @ONLINE, 2013. URL: `http://code.google.com/p/visualconnectome/`.

[46] D.J. de Solla Price. Networks of scientific papers. *Science*, 149(3683):510, 1965.

[47] Pietro DeLellis, Mario DiBernardo, and Francesco Garofalo. Novel decentralized adaptive strategies for the synchronization of complex networks. *Automatica*, 45(5):1312–1318, 2009.

[48] Sybil Derrible and Christopher Kennedy. Network analysis of world subway systems using updated graph theory. *Transportation Research Record: Journal of the Transportation Research Board*, 2112(1):17–25, 2009.

[49] Jane F Desforges and Walter L Peterson. Helicobacter pylori and peptic ulcer disease. *New England Journal of Medicine*, 324(15):1043–1048, 1991.

[50] T. Dickmeis and N.S. Foulkes. Glucocorticoids and circadian clock control of cell proliferation: At the interface between three dynamic systems. *Molecular and Cellular Endocrinology*, 2010.

[51] Reinhard Diestel. *Graph Theory*. Springer-Verlag, 2005.

[52] Florian Dorfler and Francesco Bullo. Synchronization and transient stability in power networks and nonuniform kuramoto oscillators. *SIAM Journal on Control and Optimization*, 50(3):1616–1642, 2012.

[53] Henry Dorrian, Jon Borresen, and Martyn Amos. Community structure and multi-modal oscillations in complex networks. *PLOS ONE*, 8(10):e75569, 2013.

[54] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002.

[55] James T Enright et al. Temporal precision in circadian systems: a reliable neuronal clock from unreliable components? *Science*, 209(4464):1542, 1980.

[56] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist*, 38(4):343–347, 1961.

[57] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 251–262. ACM, 1999.

[58] D.A. Fell and A. Wagner. The small world of metabolism. *Nature Biotechnology*, 18(11):1121–1122, 2000.

[59] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[60] M. Field. Combinatorial dynamics. *Dynamical Systems*, 19(3):217–244, 2004.

[61] Richard FitzHugh. Mathematical models of threshold phenomena in the nerve membrane. *The Bulletin of Mathematical Biophysics*, 17(4):257–278, 1955.

[62] Niloy Ganguly, Andreas Deutsch, and Animesh Mukherjee. *Dynamics on and of complex networks: applications to biology, computer science, and the social sciences*. Springer, 2009.

[63] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[64] Yubing Gong, Maosheng Wang, Zhonghuai Hou, and Houwen Xin. Optimal spike coherence and synchronization on complex Hodgkin–Huxley neuron networks. *Chemphyschem*, 6(6):1042–1047, 2005.

[65] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

[66] N. Gulbahce and S. Lehmann. The art of community detection. *Bioessays*, 30(10):934–938, 2008.

[67] J. Hadlaw. The London underground map: Imagining modern time and space. *Design Issues*, 19(1):25–35, 2003.

[68] Sami Hanhijärvi, Gemma C. Garriga, and Kai Puolamäki. Randomization techniques for graphs. In *Proceedings of the 9th SIAM International Conference on Data Mining (SDM '09)*, pages 780–791, 2009.

[69] D. Hansel, G. Mato, and C. Meunier. Clustering and slow switching in globally coupled phase oscillators. *Physical Review E*, 48(5):3470, 1993.

[70] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500, 1952.

[71] J Honerkamp. The heart as a system of coupled nonlinear oscillators. *Journal of Mathematical Biology*, 18(1):69–88, 1983.

[72] H. Hong and S.H. Strogatz. Kuramoto model of coupled oscillators with positive and negative coupling parameters: An example of conformist and contrarian oscillators. *Physical Review Letters*, 106(5):54102, 2011.

[73] C. Hu, J. Yu, and H. Jiang. Synchronization of complex community networks with nonidentical nodes and adaptive coupling strength. *Physics Letters A*, 2010.

[74] Wenyu Huang, Kathryn Moynihan Ramsey, Biliana Marcheva, and Joseph Bass. Circadian rhythms, sleep, and metabolism. *The Journal of Clinical Investigation*, 121(6):2133, 2011.

[75] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, Cuellar A. A. Cornish-Bowden, A., S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19:524–531, 2003.

[76] A.A. Ioannides. Dynamic functional connectivity. *Current Opinion in Neurobiology*, 17(2):161–170, 2007.

[77] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[78] WILLIAM J Jones, DAVID P Nagle Jr, and WILLIAM B Whitman. Methanogens and the diversity of archaebacteria. *Microbiological Reviews*, 51(1):135, 1987.

[79] Ferenc Jordán. Predicting target selection by terrorists: a network analysis of the 2005 London underground attacks. *International Journal of Critical Infrastructures*, 4(1):206–214, 2008.

[80] K. Josic and J. Rubin. Deriving information about architecture from activity patterns in coupled cell systems. *SIAM Journal on Applied Dynamical Systems*, 4(1):53–77, 2005.

[81] Dieter Jungnickel. *Graphs, Networks and Algorithms*, volume 5. Springer, 2008.

[82] R. Khanin and E. Wit. How scale-free are biological networks? *Journal Of Computational Biology*, 13(3):810–818, 2006.

[83] T. Kitami and J.H. Nadeau. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genetics*, 32(1):191–194, 2002.

[84] Thomas William Körner. *Fourier Analysis*. Cambridge University Press, 1989.

[85] Y. Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics*, pages 420–422. Springer, 1975.

[86] Y. Kuramoto. Cooperative dynamics of oscillator community. *Prog. Theoret. Phys. Suppl*, 79:223–240, 1984.

[87] Y. Kuramoto and I. Nishikawa. Statistical macrodynamics of large dynamical systems. case of a phase transition in oscillator communities. *Journal of Statistical Physics*, 49(3):569–605, 1987.

[88] Yoshiki Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, 2003.

[89] V. Lacroix, L. Cottret, P Thébault, and M-F. Sagot. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Transactions On Computational Biology And Bioinformatics*, 5(4):594–617, 2008.

[90] Carlo R Laing. Chimera states in heterogeneous networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(1):013113–013113, 2009.

[91] C.R. Laing. The dynamics of chimera states in heterogeneous Kuramoto networks. *Physica D: Nonlinear Phenomena*, 238(16):1569–1588, 2009.

[92] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.

[93] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

[94] Vito Latora and Massimo Marchiori. Is the Boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, 314(1):109–113, 2002.

[95] Simon B Laughlin and Terrence J Sejnowski. Communication in neuronal networks. *Science*, 301(5641):1870–1874, 2003.

[96] Klaus Lehnertz, Stephan Bialonski, Marie-Therese Horstmann, Dieter Krug, Alexander Rothkegel, Matthäus Staniek, and Tobias Wagner. Synchronization phenomena in human epileptic brain networks. *Journal of Neuroscience Methods*, 183(1):42–48, 2009.

[97] Elizabeth A Leicht and Mark EJ Newman. Community structure in directed networks. *Physical Review Letters*, 100(11):118703, 2008.

[98] Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. Tastes, ties, and time: A new social network dataset using facebook. com. *Social Networks*, 30(4):330–342, 2008.

[99] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu. Weighted networks of scientific communication: the measurement and topological role of weight. *Physica A: Statistical Mechanics and its Applications*, 350(2):643–656, 2005.

[100] M. Li, J. Wu, D. , T. Zhou, Z. Di, and Y. Fan. Evolving model of weighted networks inspired by scientific collaboration networks. *Physica A: Statistical Mechanics and its Applications*, 375(1):355–364, 2007.

[101] Y.C. Lin, LT Fan, S. Shafie, B. Bertók, and F. Friedler. Graph-theoretic approach to the catalytic-pathway identification of methanol decomposition. *Computers & Chemical Engineering*, 34(5):821–824, 2010.

[102] O. Litvin, H.C. Causton, B.J. Chen, and D. Pe'Er. Modularity and interactions in the genetics of gene expression. *Proceedings of the National Academy of Sciences*, 106(16):6441, 2009.

[103] Xuyang Lou and Johan AK Suykens. Finding communities in weighted networks through synchronization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(4):043116–043116, 2011.

[104] A. Mazurie, D. Bonchev, B. Schwikowski, and G.A. Buck. Evolution of metabolic network organization. *BMC Systems Biology*, 4(1):59, 2010.

[105] Patrick N McGraw and Michael Menzinger. Clustering and the synchronization of oscillator networks. *Physical Review E*, 72(1):015101, 2005.

[106] Patrick N McGraw and Michael Menzinger. Laplacian spectra as a diagnostic tool for network structure and dynamics. *Physical Review E*, 77(3):031102, 2008.

[107] Metabolism and nutrition. *Metabolism and Nutrition*, volume 4. Dawson Books, 2012.

[108] S. Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

[109] R. Mirollo and S.H. Strogatz. The spectrum of the partially locked state for the Kuramoto model. *Journal of Nonlinear Science*, 17(4):309–347, 2007.

[110] R.E. Mirollo and S.H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics*, 50(6):1645–1662, 1990.

[111] B. Mohar. The Laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications*, 2:871–898, 1991.

[112] Ernest Montbrió, Jürgen Kurths, and Bernd Blasius. Synchronization of two interacting populations of oscillators. *Physical Review E*, 70(5):056125, 2004.

[113] Jose M Montoya and Ricard V Solé. Small world patterns in food webs. *Journal of theoretical biology*, 214(3):405–412, 2002.

[114] M.T.A. Metropolitan transportation authority (New York, NY, USA) @ONLINE, January 2013. URL: `http://www.mta.info/nyct/maps/submap.htm`.

[115] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[116] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[117] Takashi Nishikawa, Adilson E Motter, Ying-Cheng Lai, and Frank C Hoppensteadt. Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize? *Physical Review Letters*, 91(1):14101, 2003.

[118] Christos H Papadimitriou. The Euclidean Travelling Salesman Problem is NP-complete. *Theoretical Computer Science*, 4(3):237–244, 1977.

[119] Gaby E Pfyffer, PR Murray, EJ Baron, JH Jorgensen, ML Landry, MA Pfaller, et al. Mycobacterium: general characteristics, laboratory detection, and staining procedures. *Manual of Clinical Microbiology: Volume 1*, (Ed. 9):543–572, 2006.

[120] AS Pikovsky, MG Rosenblum, and J. Kurths. Synchronization in a population of globally coupled chaotic oscillators. *EPL (Europhysics Letters)*, 34:165, 1996.

[121] Victor M Preciado, Ali Jadbabaie, and George C Verghese. Structural analysis of laplacian spectral properties of large-scale networks. *Automatic Control, IEEE Transactions on*, 58(9):2338–2343, 2013.

[122] C. Priami and P. Quaglia. Modelling the dynamics of biosystems. *Briefings in Bioinformatics*, 5(3):259, 2004.

[123] E. Ravasz and A.L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

[124] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551, 2002.

[125] Michael A Romanos, Carol A Scorer, and Jeffrey J Clare. Foreign gene expression in yeast: a review. *Yeast*, 8(6):423–488, 1992.

[126] Otto E Rössler. An equation for continuous chaos. *Physics Letters A*, 57(5):397–398, 1976.

[127] J. Schellenberger, J.O. Park, T.M. Conrad, and B.Ø. Palsson. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213, 2010.

[128] Christophe H Schilling, Stefan Schuster, Bernhard O Palsson, and Reinhart Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*, 15(3):296–303, 1999.

[129] Christian M Schneider, Lucilla de Arcangelis, and Hans J Herrmann. Scale-free networks by preferential depletion. *EPL (Europhysics Letters)*, 95(1):16005, 2011.

[130] S. Schuster, D.A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332, 2000.

[131] Stefan Schuster and Björn H Junker. Topological analysis of metabolic and regulatory networks. *Modeling in Systems Biology*, 16:209–224, 2011.

[132] Katherine A Seaton and Lisa M Hackett. Stations, trains and small-world networks. *Physica A: Statistical Mechanics and its Applications*, 339(3):635–644, 2004.

[133] M.S. Shkarayev, G. Kovačič, and D. Cai. Topological effects on dynamics in complex pulse-coupled networks of integrate-and-fire type. *Physical Review E*, 85(3):036104, 2012.

[134] Julian Sienkiewicz and Janusz A Hołyst. Statistical analysis of 22 public transport networks in Poland. *Physical Review E*, 72(4):046127, 2005.

[135] Simon Singh. *Fermat's Last Theorem: The Story of a Riddle that Confounded the World's Greatest Minds for 358 Years*. HarperPerennial, 2007.

[136] Olaf Sporns, Dante R Chialvo, Marcus Kaiser, Claus C Hilgetag, et al. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425, 2004.

[137] Cornelis J Stam, Jaap C Reijneveld, et al. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed Phys*, 1(3), 2007.

[138] J. Stelling, S. Klamt, K. Bettenbrock, Schuster S., and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, November 2002.

[139] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.

[140] S.H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[141] S.H. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Hyperion, 2003.

[142] Steven H Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1):1–20, 2000.

[143] M. P. H. Stumpf, C. Wiuf, and R. M. May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences*, 102(12):4221–4224, 2005.

[144] Kazuhiro Takemoto. Metabolic network modularity arising from simple growth processes. *Physical Review E*, 86, Sep 2012.

[145] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs. structural. In *ACM SIGCOMM Computer Communication Review*, volume 32, pages 147–159. ACM, 2002.

[146] T.F.L. Transport For London @ONLINE, January 2013. URL: `http://www.tfl.gov.uk/`.

[147] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen. Hierarchy measures in complex networks. *Physical Review Letters*, 92(17):178702, 2004.

[148] A. Uçar, K.E. Lonngren, and E.W. Bai. Synchronization of chaotic behavior in nonlinear Bloch equations. *Physics Letters A*, 314(1-2):96–101, 2003.

[149] Vera Vasas, Christiane Lancelot, Véronique Rousseau, Ferenc Jordán, et al. Eutrophication and overfishing in temperate nearshore pelagic food webs: a network perspective. *Marine Ecology Progress Series*, 336(10):1–14, 2007.

[150] FH Verhoff and JE Spradlin. Mass and energy balance analysis of metabolic pathways applied to citric acid production by *aspergillus niger*. *Biotechnology and Bioengineering*, 18(3):425–432, 1976.

[151] Anders Vretblad, S Axler, FW Gehring, and KA Ribet. *Fourier analysis and its applications*, volume 223. Springer, 2003.

[152] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1478):1803, 2001.

[153] Xiao-Hua Wang, Li-Cheng Jiao, and Jian-She Wu. Extracting hierarchical organization of complex networks by dynamics towards synchronization. *Physica A: Statistical Mechanics and its Applications*, 388(14):2975–2986, 2009.

[154] Yuqing Wang, David TW Chik, and ZD Wang. Coherence resonance and noise-induced synchronization in globally coupled Hodgkin-Huxley neurons. *Physical Review E*, 61(1):740, 2000.

[155] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[156] J.H. Wilkinson, J.H. Wilkinson, and J.H. Wilkinson. *The Algebraic Eigenvalue Problem*, volume 155. Oxford Univ Press, 1965.

[157] Robin James Wilson. *Introduction to Graph Theory, 4/e*. Pearson Education India, 1985.

[158] J. Wolf and R. Heinrich. Dynamics of two-component biochemical systems in interacting cells; synchronization and desynchronization of oscillations and multiple steady states. *Biosystems*, 43(1):1–24, 1997.

[159] Y.I. Wolf, G. Karev, and E.V. Koonin. Scale-free networks in biology: new insights into the fundamentals of evolution? *Bioessays*, 24(2):105–109, 2002.

[160] C.W. Wu and L.O. Chua. Application of graph theory to the synchronization in an array of coupled nonlinear oscillators. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 42(8):494–497, 1995.

[161] S. Wuchty. Scale-free behavior in protein domain networks. *Molecular Biology and Evolution*, 18(9):1694, 2001.

[162] V. Zagursky, A. Riekstinch, and I. Zarumba. Testing the methods of sound signal compression. In *Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference focused on*, volume 2, pages 595–600 vol.2, July 2003.

[163] Changsong Zhou, Lucia Zemanová, Gorka Zamora, Claus C Hilgetag, and Jürgen Kurths. Hierarchical organization unveiled by functional connectivity in complex brain networks. *Physical Review Letters*, 97(23):238103, 2006.