UNIVERSITAT DE BARCELONA

From Dynamics to Structure of Complex Networks: Exploiting Heterogeneity in the Sakaguchi-Kuramoto Model

Gemma Rosell Tarragó

PhD Thesis

From Dynamics to Structure of Complex Networks: Exploiting Heterogeneity in the Sakaguchi-Kuramoto Model

Author

Gemma Rosell Tarragó

Advisor

Dr. Albert Díaz Guilera

Universitat de Barcelona

From Dynamics to Structure of Complex Networks: Exploiting Heterogeneity in the Sakaguchi-Kuramoto Model

MEMÒRIA PRESENTADA PER OPTAR AL GRAU DE DOCTOR PER LA UNIVERSITAT DE BARCELONA

PROGRAMA DE DOCTORAT EN FÍSICA

Autor

GEMMA ROSELL TARRAGÓ

Director

DR. ALBERT DÍAZ GUILERA

Tutor

DR. GIANCARLO FRANZESE

UNIVERSITAT DE BARCELONA

A Dani

Agraïments

El camí recorregut en aquesta tesi no hauria estat possible sense

l'ajuda i la presència de Déu,

les oportunitats i orientacions rebudes de l'Albert,

la confiança i l'amor del Dani,

l'alegria i la tendresa de les meves filles

i dels que vindran,

el pilar que han sigut sempre els meus pares,

l'intercanvi d'idees i formació amb companys i professors,

els motius recordats per tants amics,

els germans,

els padrins,

i tota la família...

A tots i totes – i a molts que no sabré mai – moltes gràcies per haver format part d'aquesta etapa.

Contents

CHAPTER 1

Introduction

The whole is more than the sum of its parts

Aristotle

1.1 The Complex System Approach to Nature

Already Aristotle realized that considering the entities of many systems as a whole could not be explained by a linear aggregation of their individual rules or properties. Conversely, non-trivial phenomena and unexpected properties may arise when the interactions between units are considered. The ensemble of neurons producing consciousness and intelligence, a large population of starlings showing complex flocking patterns, a crowd clapping spontaneously in unison or a systemic financial shock are examples of such emergent behaviours [71].

Figure 1.1: **Flock of birds.** Drawing of a flock of birds peforming complex patterns. *(Source: Designed by Neus Rosell)*

A complex system is a large collection of components that interact with each other at the local scale and still is able to spontaneously self-organize and show non-trivial global patterns and conducts involving the whole system, often without external guidance. Complex system science benefits from and contributes to the required novel mathematical framework and methodologies to study this type of systems [51].

The interactions between the constituents of the system may occur in multiple ways, at different time and size scales, and potentially with the environment as well. The society is a good case in point: citizens living and moving within a city, but also commuters and visitors, relating with each other at the level of households, neighbourhoods, workplaces, leisure and academic places, not to mention the sizeable communication activity through online platforms and information technology.

One cannot infer the behaviour of the whole system only by observing single individuals. The phenomena of unexpected and non-trivial global patterns coming out and unpredicted behaviours occurring at the large scale is known as 'emergence' [75]. Consider, for instance, the massive amount of particles composing the surface of the Earth and that, eventually, experience the rupture of geological faults, namely an earthquake, resulting from the lithosphere's seismic waves.

Complex systems typically evolve over time. In other words, the states or variables that characterize their constituents, but also the system as a whole, experience changes that depend linearly or non-linearly with time, its current state or external inputs. Small shifts in the intrinsic parameters or the environment may lead to dramatic changes in the behaviour of the system. Certain dynamic processes can lead a system to a chaotic state, i.e., being extremely sensitive to initial conditions or perturbations and becoming unpredictable in the long run [29, 166]. Consider, for instance, the difficulty of accurate weather forecasting, that results from the unpredictability of the climate-weather complex system[1].

We can also look at the synchronization of flashing among large groups of fireflies during twilight, or the spontaneous formation of discrete spatial patterns present in the clustering of some shrubs or the spots of jaguars [4, 161]. Complex systems have the capability of self organizing, that is, the monitoring of the whole system is distributed among their units without external intervention or guidance.

[1] Syukuro Manabe, Klaus Hasselmann and Giorgio Parisi were awarded the 2021 Nobel Prize in Physics for their contributions to our understanding of complex systems. In particular, Syukuro Manabe and Klaus Hasselmann shared one half of the prize for their mathematical modelization of the Earth weather-climate system and their predictions on global warming. Giorgio Parisi was awarded with the other half of the prize for his study of the interactions between disorder and fluctuations in physical systems at very different scales [60].

Finally, complex systems are present in different domains and disciplines, including biology and medicine, economy and finance, physics and engineering, information technology, politics, sociology, and more. All complex systems share properties and are made up by similar building blocks that make them exhibit common phenomena and features, a fact known as 'universality'. Consequently, the same mathematical and computational framework, methods and models are applicable and worthwhile to all of them [170].

One of the most useful representation of a complex system is considering the network of interactions between its units. Most of the real-world complex systems are best described as complex networks: the nodes and edges of the network corresponding to the constituents and interactions of the system, respectively. Importantly, real-world complex networks, such as biological, socio-economical or technological networks, exhibit common features which are neither purely regular nor purely random. Section 1.2 goes through the fundamentals of network theory, including the mathematics of networks and the common properties of real networks.

1.2 Fundamentals of Network Theory

1.2.1 Mathematics of Networks

1.2.1.1 Definitions and Notation

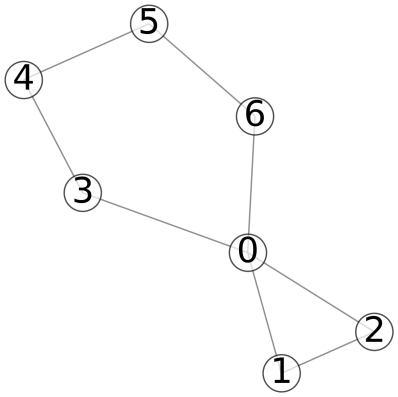
Formally, a complex network can be mathematically described as a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ consisting of a collection of N vertices (or nodes), the set $\mathcal{V}(\mathcal{G})$, and m edges (or links) connecting them, the set $\mathcal{E}(\mathcal{G})$. Both $\mathcal{V}(\mathcal{G})$ and $\mathcal{E}(\mathcal{G})$ can be given attributes [122, 20, 176].

In an undirected graph or network, each edge is defined by a pair of adjacent nodes, (i, j), without a preferred direction. Conversely, in a directed graph, the order of the nodes is important, and the pair (i, j) stands for an edge going from node i to node j. In a weighted graph, each edge (i, j) is characterized by a real number, w_{ij}, that represents the strength or intensity of the connection.

In addition, a graph can contain *self-loops*, i.e., edges from a node to itself, and *multiple edges*, nodes that are connected by more than one edge. Graphs that contain either of these last features, are called *multigraphs*. Otherwise, a *simple graph* is an unweighted, undirected graph containing no self-loops or multiple edges.

A graph can be represented by an $N \times N$ square matrix called the *adjacency matrix* of $\mathcal{G}$, $\mathcal{A}(\mathcal{G})$, which elements indicate the connectivity structure of the graph. In the case of simple graphs, the element A_{ij} equals 1 if node i is linked to node j, and 0 otherwise. Undirected graphs are represented by symmetric

matrices, with zeros on its diagonal. An example of a simple graph and its corresponding adjacency matrix is shown in Figure 3.1.



$$\mathscr{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Figure 1.2: **Example of a simple graph and the corresponding adjacency matrix.** Simple graph made of 7 nodes and 8 edges, i.e., $N = 7$ and $m = 8$, and its corresponding symmetric adjacency matrix, $\mathscr{A}$. The network is defined in Ref.[124].

1.2.1.2 Graph spectra

The spectrum of a graph is the set of eigenvalues of its adjacency matrix. $\mathscr{A}$ has N eigenvalues, $\{\lambda_i\}$, and N associated eigenvectors, $\{\vec{v}_i\}$. The adjacency matrix of an undirected simple graph is a real symmetric matrix and hence, can be orthogonally diagonalizable. Moreover, all its eigenvalues are real numbers. However, the eigenvalues of a directed graph can be complex numbers.

The Perron-Frobenius theorem [138, 65] states that a real square matrix with positive entries has a unique largest real eigenvalue and which corresponding eigenvector can be chosen so that all its entries are positive. The extension for non-negative matrices (with non-negative elements) states that the leading eigenvector will be non-negative and greater than or equal, in absolute value, to the remaining eigenvalues. The corresponding eigenvector can be chosen to have non-negative components.

Consequently, considering such matrix to be the adjacency matrix $\mathscr{A}$ of a graph, the theorem shows that $\mathscr{A}$ has a real non-negative eigenvalue λ_1 such that $|\lambda_i| \leq \lambda_1$, for all other eigenvalues. If the graph is connected, the multiplicity of λ_1 is 1 and $|\lambda_i| < \lambda_1$. For an undirected graph, all entries of the corresponding eigenvector $\vec{v}_1$ are non-negative. The same theorem tells that if the graph is connected the following inequalities hold: $\max\{\langle k \rangle, \sqrt{k_{max}}\} \leq \lambda_1 \leq k_{max}$.

1.2.1.3 Node degree and degree distribution

The degree k_i of node i is the number of edges connected to it. For an undirected graph, the degree can be computed from the adjacency matrix as[2]

$$k_i = \sum_{j \in \mathcal{V}} A_{ij} \tag{1.1}$$

The list of node degrees is called the *degree sequence*.

If the graph is directed, the degree of a node is composed by the number of outgoing links (known as the *out-degree*) and the number of ingoing links (referred to as the *in-degree*). The total degree consists in the sum of both components.

One of the most fundamental properties of networks is the heterogeneity in its degrees, a feature which can be captured by plotting the histogram of node degrees, also known as the degree distribution $P(k)$. The quantity p_k is defined as the probability that a randomly chosen node has a degree k or, alternatively, the fraction of nodes that have a degree k.

Relevant information can be obtained by plotting the degree distribution of a graph, as we will show in Section 1.2.2. However, further knowledge on how the degree is distributed among the nodes can be obtained by computing the n-moment of $P(k)$, defined as

$$\langle k^n \rangle = \sum_k k^n P(k) \tag{1.2}$$

The first moment of the distribution, $\langle k \rangle$, informs us about the mean degree of the graph. The second moment $\langle k^2 \rangle$ is, however, more decisive for determining the behaviour of a dynamical process being held on the corresponding graph.

1.2.1.4 Assortativity

Assortative mixing – also called *homophily* in the specific case of social networks – is the tendency that the nodes of a network have to link to other nodes that are similar to them in some way. Different similarity measures can be considered, although node degree is the most frequently used when studying complex networks.

In particular, the tendency that the nodes of a network with a given degree are attached to other with a similar degree value is called *assortativity*, while the converse situation – high degree nodes having a bias towards linking to low degree nodes – is called *disassortativity*. This property is often measured as a correlation between two nodes. The two most notable measures are the *assortativity coefficient* and the *neighbour connectivity*.

[2] If self-edges are present in the graph, the corresponding diagonal element has to be set to $A_{ii} = 2$.

The assortativity coefficient is given by the Pearson correlation coefficient of the degree between pairs of neighbouring nodes [117]. A network is considered to be assortative if the values of the correlation coefficient are positive, and conversely.

Nevertheless, one can study the correlations of a network using the neighbour connectivity [117, 126, 160]. When the degree k' of the neighbouring nodes to which a node of degree k is connected depends on k, another quantity shall be considered: the average degree of the nearest neighbours of nodes with degree k, $\langle k_{nn} \rangle (k)$ is defined as [3]

$$\langle k_{nn} \rangle (k) = \sum_{k'} k' P(k'|k) \tag{1.3}$$

This function can be plotted and inform us whether the network is assortative (positive slope) or disassortative (negative slope).

Eq.(1.3) can be alternatively calculated by taking the average degree of the nearest neighbours of nodes with degree k once

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} k_j = \frac{1}{k_j} \sum_{j=1}^{N} A_{ij} k_j \tag{1.4}$$

is computed for each node. $\mathcal{N}_i$ stands for the set of neighbouring nodes of node i.

1.2.1.5 Walks

A *walk* in a graph (network) is any sequence of consecutive vertices (nodes) that are linked through the corresponding edges. The preferred direction of edges has to be preserved in directed graphs. A *path* is a walk in which neither vertices nor edges are repeated. A *cycle* is a closed path, that is, the starting and ending nodes are the same.

The *length* of a walk is the number of edges that are spanned, taking into account that they can be traversed more than once.

The number of walks of a given length r can be easily computed from the elements of the adjacency matrix as follows: for a simple graph, the element A_{ij} is 1 if an edge exists between nodes i and node j. Similarly, the product $A_{ik}A_{kj}$ is 1 if a walk of length 2 exists between nodes i and j through node k, and 0 otherwise. Hence, the sum for all possible k results in the total number of walks of length 2 as follows

$$\sum_{k} A_{ik} A_{kj} = [A^2]_{ij} \tag{1.5}$$

[3] Besides the correlations between k and k' are formally characterized by the conditional probability $P(k'|k)$, the quantity $\langle k_{nn} \rangle (k)$ is preferred in order to avoid noisy results.

We note that the right-hand side of Eq.(1.5) corresponds to the element ijth of the squared adjacency matrix.

In the same way, the number of walks of length r from node i to node j is given by the element ijth of the matrix A^r.

1.2.1.6 Clustering

The nodes of many real-world networks are naturally divided into groups that share certain local features. There are several ways of measuring and detecting such clusters. We present some of the constructs that are typically used.

A very important property, specially in social networks, is *transitivity* [156, 119]. Mathematically, the relation "$\circ$" is said to be transitive if $a \circ b$ and $b \circ c$ implies $b \circ c$. For a network, the 'the friend of my friend is also my friend' relation is represented by the following triangle: if node u is connected with node v by an edge and node v is connected to node w, then nodes u and w are also connected. A network is a perfect transitive one only if each component is a fully connected subgraph or clique[4]. However, if nodes u and v are connected and nodes v and w are connected, they form a path of length 2. This, however does not imply that nodes u and w are also connected. For this reason, we are interested in assessing the level of transitivity of a network. Mathematically, we define the *global clustering coefficient* [102, 130] as

$$C = 3 \times \frac{\text{number of triangles}}{\text{number of connected triples}} \tag{1.6}$$

where connected triple means three nodes uvw with, at least, edges (u, v) and (v, w).

We can also define a *local clustering coefficient*, that is, a value for each node, as follows [174]

$$C_i = \frac{\text{number of pairs of neighbours of } i \text{ that are connected}}{\text{number of pairs of neighbours of } i} \tag{1.7}$$

Using Eq.(1.7) we can define an alternative measure of the global clustering coefficient as the mean of the local clustering coefficient of all nodes.

1.2.1.7 Community Structure

Another property of many real-world networks is the presence of groups of nodes that are more densely connected internally than with the remaining nodes. The *community structure* of the network refers to such frequent division of the network into – potentially overlapping – communities or clusters [74, 121, 120, 28, 105, 63, 57].

[4] a subgraph in which all vertices are connected to all others

Each community is made up of a set of nodes such that each node belongs to one community[5]. Most of the algorithms designed to find communities within a network rely on this definition of community. A network division into such separate communities or groups is called a *partition*. Despite the difficulties of finding the best network partition into communities[6], several approaches can be considered and are applied with different degrees of success and adequacy [140, 62, 63]. We name a few of them: spectral methods, methods based on statistical inference, methods based on optimisation and methods based on dynamics, among others. We should also consider other issues such as the decision on how many clusters are more suitable to some algorithms that require such pre-detection input, how to deal with stochastic techniques that do not deliver a unique answer or the problem of soft clustering, that is, communities that may overlap.

One of the most frequently used methodology relies on the maximization of a function that captures the goodness of a given network partition. The most noted quality function was defined by Michelle Girvan and Mark Newman and is known as *modularity* [121].

Modularity is a measure of the extent to which a network is naturally divided into modules or communities, i.e., it captures whether the density of edges within clusters is significantly larger than the quantity expected in a random network. It is mathematically expressed as the fraction of edges that belong to the different groups minus the fraction of edges we would expected to find if they were distributed randomly.

On the one hand, the total number of edges that connect nodes of the same type (that is, belonging to the same module or community) is given by

$$\frac{1}{2}\sum_{i,j} A_{ij}\delta(c_i, c_j) \tag{1.8}$$

where c_i is the class or community identifier to which node i belongs. Hence $\delta(c_i, c_j) = 1$ only if nodes i and j belong to the same community. The factor $1/2$ accounts for the fact that edges are counted twice.

[5]This definition of *non-overlapping* communities, however, can be relaxed so that nodes can belong to more than one community [7].

[6]Why is the partition (fixed groups' size) of a graph into groups such a difficult task? Because of its prohibitive scaling on the network size. In general the *Bell number* $B(\mathcal{G})$ of a graph is the number of partition of its vertex set (For example, $B_5 = 52$ corresponds to the 52 possible partitions of a set with 5 elements). For instance, if we were interested in dividing the nodes only into two parts (*graph bisection*), an exhaustive search, that is, trying all possible partitions in order to find the best one, would require $n!/n_1!n_2!$ prohibitive possibilities to test, where n_1 and n_2 are the sizes of both groups. Actually, the generating function of the Bell numbers shows that they grow more rapidly than an exponential function with respect to its number of components: $B(x) = \exp^{\exp^x - 1}$

On the other hand, the expected number of edges between nodes that belong to the same community if edges were placed randomly is given by[7]

$$\frac{1}{2}\sum_{i,j}\frac{k_i k_j}{2m}\delta(c_i,c_j) \tag{1.9}$$

Finally, if we want to calculate the fraction between the current and expected number of edges within groups, we have to divide the subtraction by m, i.e., the number of edges in the network. Therefore, modularity is defined as [121, 46]

$$Q=\frac{1}{2m}\sum_{i,j}\left(A_{ij}-\frac{k_i k_j}{2m}\right)\delta(c_i,c_j) \tag{1.10}$$

The values of Q fall in the range $[-1/2,1]$ [37].

There are, however, some limitations on the definition of modularity [61, 91]. Firstly, the maximum value of Q depends on the sizes of the groups and the degrees of the nodes and, hence, it can be considerably less than 1, even for a perfectly mixed network[8]. Secondly, standard methodologies to maximize modularity in large networks fail to distinguish small communities, as they tend to merge them. For these reasons, several studies suggest alternative methodologies to find the communities of a network. For example, in Ref.[14] the authors considered the dynamics of a system of coupled oscillators and its path to synchronization in order to obtain modular structures that emerge in different time scales and which correspond to the well-defined communities of a network.

1.2.1.8 Centrality measures

From the topology or structure of the network, enclosed in the adjacency matrix, several measures can be considered in order to characterize it. Triggered by modern sociologists, much effort has been devoted to the notion of *centrality*, which addresses the question 'Which are the most important or central nodes in the network?' [32, 34, 115]. Several *centrality measures* came up, emerging from social network analysis, each of them with different considerations regarding node importance. Hereafter, we revise some of the most well-known centralities [25].

- *Degree Centrality*

  The most simple measure of importance that one could define is the degree of the node, i.e, the number of edges that are connected to it. Although it is a straightforward meter, degree centrality is widely used in the

[7] There are $2m$ ends of edges in the network, where m is the total edges. If the network is built randomly, the probability that the free end of the edge attached to node i is connected to node j is given by $k_j/2m$.

[8] A network where each node is connected only to nodes of the same class

social sciences as a measure of (local) influence, and in many other fields as a first approach to node importance. Mathematically, degree centrality can be written as

$$x_i = \sum_j A_{ij} = k_i \tag{1.11}$$

Note that degree centrality can be normalized by dividing by the maximum possible degree in a simple graph, i.e., $N-1$.

In directed networks, both the in-degree and the out-degree of nodes can be considered as separate centrality measures.

- *Eigenvector Centrality*

  Unlike degree centrality, eigenvector centrality considers the fact that a node is important if it is connected to other important nodes. In other words, the centrality of a node is proportional to the sum of scores of its neighbours. Mathematically, the idea behind eigenvector centrality can be written as

$$x_i \propto \sum_j A_{ij} x_j \tag{1.12}$$

  Starting from the guess $\vec{x}(0) = \vec{1}$, we can iterate to the next steps as

$$x_i(t+1) = \sum_j A_{ij} x_j(t) \tag{1.13}$$

  and in matrix notation, $\vec{x}(t+1) = A\vec{x}(t)$. Therefore, at iteration t,

$$\vec{x}(t) = A^t \vec{x}(0) \tag{1.14}$$

  But, if $\vec{x}(0)$ is expressed in terms of the eigenvectors, $\vec{v}_i$, of A matrix, $\vec{x}(0) = \sum_i c_i \vec{v}_i$, Eq(1.14) turns to

$$\vec{x}(t) = A^t \sum_i c_i \vec{v}_i = \sum_i c_i \lambda_i^t \vec{v}_i, \tag{1.15}$$

  where λ_i is the eigenvalue corresponding to the eigenvector $\vec{v}_i$. Considering the Perron-Frobenius theorem, we can use the largest eigenvalue λ_1 and its corresponding eigenvector $\vec{v}_1$ (with non-negative entries) and write

$$\vec{x}(t) = \lambda_1^t \sum_i c_i \left[\frac{\lambda_i}{\lambda_1}\right]^t \vec{v}_i \to c_1 \lambda_1 \vec{v}_1 \tag{1.16}$$

  when $t \to \infty$, since $\lambda_i/\lambda_1 < 1$ for $i > 1$. Finally,

$$\vec{x}(t) = c_1 \lambda_1 \vec{v}_1 \tag{1.17}$$

Therefore, eigenvector centrality is proportional to the leading eigenvector of A.

Alternatively, eigenvector centrality satisfies the matrix equation

$$\vec{x}A = \lambda_1 \vec{x} \tag{1.18}$$

Eigenvector centrality is also known as *Bonacich's centrality*, who first defined it in 1987 [31, 32, 33].

Eigenvector centrality can also be defined for directed networks using either the leading left eigenvector or the leading right eigenvector, accounting for the out-degree or the in-degree of nodes, respectively, as the adjacency matrix is typically asymmetric. However, eigenvector centrality is not a suitable measure for directed acyclic graphs[9] as if the progression of nodes being pointed by others ends up at a vertex set that has in-degree zero, the resulting centrality will be zero.

- *Katz Centrality*

  The definition of Katz centrality, first introduced by Leo Katz in 1953 [87, 83], is similar to that of the eigenvector centrality, but it tackles the issue of zero in-degree nodes contributing zero to the centrality of the neighbouring nodes. To this end, each node is provided with a small amount of centrality, irrespective of the value of its degree. Mathematically,

  $$x_i = \alpha \sum_j A_{ij} x_j + \beta \tag{1.19}$$

  where α and β are positive parameters that tune the importance of both contributions. In matrix terms, Eq.(1.19) is written as

  $$\vec{x} = \alpha A\vec{x} + \beta\vec{1} \tag{1.20}$$

  If $\beta = 1$, the solution of Eq(1.20) for the centrality $\vec{x}$ is

  $$\vec{x} = (\mathbb{I} - \alpha A)^{-1}\vec{1}, \tag{1.21}$$

  where $\mathbb{I}$ is the identity matrix.

  A proper choice of α is $0 < \alpha < 1/\lambda_1$, where λ_1 is the largest eigenvalue of A matrix[10].

[9] Directed graphs with no directed cycles

[10] The matrix identity $(\mathbb{I} - M)^{-1} = \sum_n^\infty M^n$ is only true if the series on the right-hand side converges. This last condition is true as long as the absolute value of all eigenvalues of M matrix are smaller than 1, i.e., $|\lambda_i| < 1$.

The free parameter α determines the balance between the constant term and the eigenvector term. When $\alpha \to 0$ then the eigenvector term vanishes and Katz centrality approaches degree centrality. Conversely, as α increases, the values of Eq(1.20) become large and on the limit when $det(\mathbb{I} - \alpha A) = 0$, they diverge. This value is achieved when $\alpha = 1/\lambda_1$

- *Closeness Centrality*

  Closeness centrality provides a measure of how close is a node from all others on average [149, 24]. The length of the shortest path between node i and node j [11] is denoted as d_{ij}. Then, the mean geodesic distance for node i is

  $$l_i = \frac{1}{n}\sum_j d_{ij} \tag{1.22}$$

  Nodes with low l_i are those which are located close to others on average. In other words, they have better access to all other nodes. However, in order to define a centrality measure, more central nodes are required to get higher centrality value. For this reason, closeness centrality is defined as

  $$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \tag{1.23}$$

  Despite closeness centrality is widely used in the social sciences it has some problems: firstly, it is difficult to distinguish between nodes with a high and a low centrality score as the values tend to be piled up around a narrow range. Moreover, the values easily fluctuate with small changes in the network. Finally, alternative definitions of closeness centrality, such as efficiency [111], are proposed in order to overcome the issue of infinite l_i values found in networks with several components.

- *Betweenness Centrality*

  Betweenness centrality measures the extent to which a node lies on the paths between other nodes [64]. That is to say, if information or messages are passed through the nodes of a network such that they follow the shortest or geodesic paths, how many of such paths cross over the different nodes? Hence, the betweenness centrality of a nodes is proportional to the number of geodesic paths that nodes lies on. The removal of nodes with a high betweenness value produces a larger disruption of the communication in the network. Formally and for a general network, betweenness

[11] The shortest path between two nodes is also called the geodesic path, and corresponds to the path with less required edges between the given pair.

centrality is defined as

$$x_i = \sum_{s,t} \frac{n^i_{s,t}}{g_{s,t}} \tag{1.24}$$

where $n^i_{s,t}$ is the number of geodesic paths from node s and node t that pass through node i and $g_{s,t}$ is the total number of geodesic paths going through node i. If both quantities are 0 the fraction $n^i_{s,t}/g_{s,t}$ is set to 0.

The values of betweenness centrality are typically distributed over a wide range. Moreover, they inform of how a node lies between other nodes and hence, need not be correlated with the degree of that node.

1.2.2 Properties of Real Networks

In the previous section we have reviewed some of the most important concepts of networks (and, consequently, graphs as its more adequate mathematical representation). In this section we discuss how networks look like in the real world (real systems that are represented as networks, for example technological networks, social networks or biological networks). The analysis of systems that are rightly described as networks has uncovered the surprising fact that networks of very different nature share common patterns, a concept known as *universality* in the field of Statistical Physics. In this chapter we will enumerate and explain the most relevant of such common features.

- *Small-world property*

  One of the most well-known and unexpected properties of real networks is the *small-world* effect [174], the finding that in most networks, and despite having thousands and millions of nodes, the distance between nodes remains small. This was first evidenced by Stanley Milgram's letter passing experiment in the 1960s [107, 171][12]. Mathematically, the small-world effect means that the mean distance, l, between nodes in a network is smaller than the expected in a lattice. For a network of one component l is defined as

  $$l = \frac{1}{n}\sum_i l_l \tag{1.25}$$

  where l_i is the average geodesic distance of node i defined in Eq.(1.23). In other words, l represents the average number of steps that a node needs to reach another node [8].

[12] People were asked to send a received letter to a distant person passing it from acquaintance to acquaintance. The letters that reached the final destination, made it in a small number of steps, around six.

- *Power-law degree distribution*

  In Section 1.2.1 we define the degree distribution of a network. In order to obtain more information of the network it is often very informative to plot $P(k)$ as a function of the degrees k (for instance, using a histogram). Looking at real-world networks, it turns out that most of the systems display a very particular shape: most of the nodes have a very small degree, i.e., are poorly connected, while there is a few of them, known as *hubs*, that have a large degree. These right-skewed degree distributions are commonly well described by a power law distribution [52, 141, 19]. Mathematically,

  $$P(k) = Ck^{-\alpha} \tag{1.26}$$

  where the constant α is the exponent of the power law. Typically, for real networks the values of the exponent fall in the range $2 \leq \alpha \leq 3$.

  Nevertheless, despite the tail of the degree distribution can be generally described by Eq.(1.26), the full degree distribution is more complex than that. Moreover, there is often a cut-off that limits the maximum degree that a node can have[13]. Networks that can be described by a power-law distribution are known as *scale-free* networks because this density distribution is a function which is invariant under rescaling, i.e., $P(c \cdot k) = c^{\alpha} \cdot P(k)$, and thus free of a natural scale. This last property is very familiar to statistical physicists and that is why the interest and contributions from this field.

- *Clustering coefficient*

  As defined in Section 1.2.1, the clustering coefficient is a measure of the probability that two neighbours of a node are themselves neighbours. For many real-world networks, the number of this closed triadic relations, i.e., triangles, is significantly larger than that expected if connections were chosen at random. Importantly, the clustering coefficient of many social networks are typically large, between 0.1 and 0.6 [122]. Nevertheless, other systems such as technological or biological networks present values which are not different to that of the equivalent random networks.

  If we look at the local clustering coefficient, defined in Eq.(1.7), it turns out that, on average, nodes with a higher degree tend to have a small clustering coefficient [97, 122].

[13] Although it is generally claimed that most real-world networks follow this pattern, several empirical studies are more reluctant to accept such general behavior[39, 129, 167, 47, 82].

1.2.3 Network Models

One of the aims of network science is generating synthetic networks with specific features, such that they resemble real-world networks. For this reason, the characterization of the latter is of particular importance (see Section 1.2.2).

In addition, network models are used to have a better understanding of network behaviour and are often used as null models to reject or accept more complex structures.

Hereafter, we briefly present three well-known *random network models*[14] that are widely used in network science: the Erdös-Rényi model, the Barabási-Albert model and the Newman modular model.

1.2.3.1 Erdös-Rényi network model

The Erdös-Rényi model (ER) generates a graph $G(N,m)$ of N nodes that are linked through m randomly placed edges [55, 73, 59]. Alternatively, the graph $G(N,p)$ considers that each pair of nodes is connected with probability p. In the following chapters we consider the second definition of the model and thus we constrain ourselves to showing some of its basic results.

In the $G(N,p)$ model, the network is built as follows: starting from N disconnected nodes, each of the $N(N-1)/2$ possible pairs is connected with probability p (generate a random number in the range $[0,1]$ from a uniform distribution. If this value is smaller than the considered p, then connect the corresponding pair of nodes.).

Despite each realization of the model looks slightly different, the network class $G(N,p)$ is characterized by the same features.

The degree distribution of $G(N,p)$ follows a Binomial distribution [116]. Therefore, the probability that a randomly chosen node has degree k is

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{1.27}$$

Also, the peak is located at

$$\langle k \rangle = p(N-1) \tag{1.28}$$

and the width of the distribution is

$$\sigma_k = p(1-p)(N-1) \tag{1.29}$$

[14] Generally, complex networks are not described as lattices or complete networks, but are characterized by an apparently random connectivity pattern, i.e., links between nodes are placed randomly, but following certain rules that are precisely prescribed by different random network models.

However, when $N \gg \langle k \rangle$, the Binomial distribution can be approximated with a Poisson distribution [173]. This approximation enables easier analytical calculations, because the latter has only one parameter, $\langle k \rangle$:

$$p_k = \frac{\langle k \rangle}{k!} e^{-\langle k \rangle} \tag{1.30}$$

The peak of the Poisson distribution is located at $\langle k \rangle$ and the width of the distribution is $\sqrt{\langle k \rangle}$.

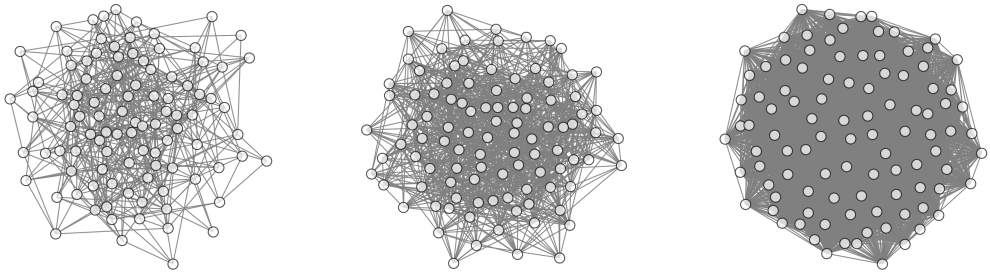


Figure 1.3: **Erdös-Rényi networks of different densities.** Erdös-Rényi networks of size $N = 100$ with three different values of connection probabilities p. From left to right, $p = 0.1$, $p = 0.2$ and $p = 0.7$.

1.2.3.2 Barabási-Albert network model

Section 1.2.2 highlights some of the most relevant features that are shared among many real-world networks. Scale-free networks are therefore not rightly described by the Erdös-Rényi random network model, as the degree distribution is not right-skewed and other properties, such as the small-world effect or the presence of hubs are not well represented.

The fact that many real networks share a common architecture, namely, they are rightly described by power law degree distributions, could be due to simple rules or mechanisms that are common for all such systems.

The following two processes are present in the evolution of real-world networks: they grow through the addition of new nodes and some of their constituent nodes are preferred to others.

Albert-László Barabási and Réka Albert proposed a model, as a special case of a more general model called *Price*'s model [141, 17], having both of the characteristics and for this reason it is known as the Barabási-Albert preferential attachment model (BA).

It turns out that the resulting networks could be rightly described by power law degree distributions and hence the scale-free property emerges from the two

simple formation mechanisms: growth of the network through the addition of new nodes and linear preferential attachment, i.e., nodes are preferred proportionally to its degree.

The networks are built as follows: starting from m_0 initially connected nodes, at each time step t we add a new node with m free stubs that are linked randomly to existing nodes with a probability given by $k_i / \sum_j k_j$ [8]. In other words, nodes that have a higher degree have more chances to receive new connections when new nodes are added.

For large k and t the degree distribution of the BA model is given by [20]

$$p_k \approx 2m^{1/\beta} k^{-\gamma} \tag{1.31}$$

where β is called the dynamical exponent and has a value of $\beta = 1/2$, $\gamma = 1/\beta + 1 = 3$. Hence, the exponent of the power law is 3.

Despite having some limitations, the BA model is widely used to generate synthetic scale-free networks.

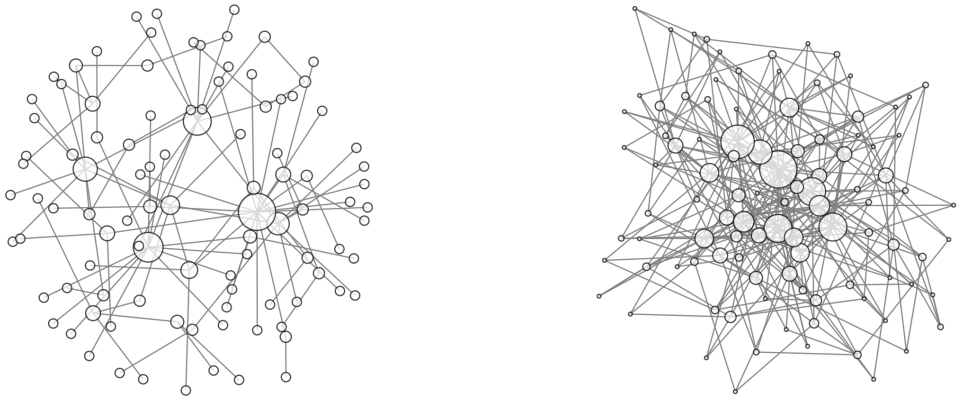


Figure 1.4: **Barabási-Albert networks of different densities.** Barabási-Albert networks of size $N = 100$ with two different values of new nodes' free stubs m. From left to right, $m = 1$ and $m = 3$. The radius of nodes is proportional to the degree.

1.2.3.3 Girvan-Newman modular network model

Michelle Girvan and Mark Newman, in order to test their algorithms of community detection, defined a model of random networks with known community structure [74]. The nodes of the network are distributed in a number $N_{modules}$ of modules or communities. Similarly to ER networks, edges are placed independently at random between pairs of nodes belonging to the same community with probability p_{in}, whereas pairs belonging to different communities are

linked with probability p_{out}. They considered that the expected degree of each node $\langle k \rangle$ and the average degree within a community $\langle k_{in} \rangle$ were fixed parameters. p_{in} and p_{out} are thus given by

$$p_{in} = \frac{\langle k_{in} \rangle}{n_{in} - 1} \qquad p_{out} = \frac{\langle k \rangle - \langle k_{in} \rangle}{n_{out}} \tag{1.32}$$

where $n_{in} = N/N_{modules}$ is the number of nodes in each module, and $n_{out} = N - n_{in}$ is the number of nodes that are not in a particular module.

As $\langle k_{in} \rangle$ increases the modules are easier to be identify. In other words, the modularity Q (defined in Eq.(1.10)) of the network increases too.

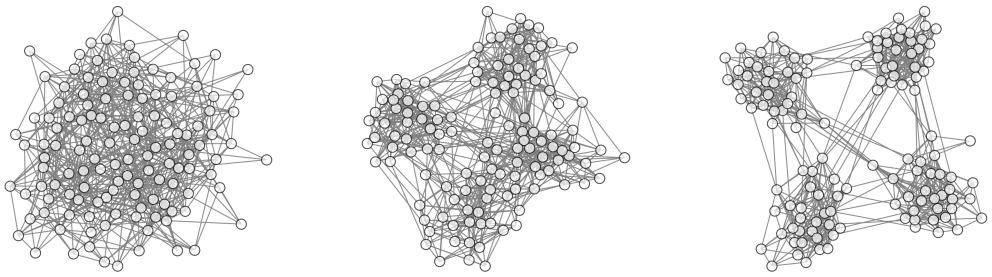


Figure 1.5: **Girvan-Newman modular networks with different modularity** Q**.** Girvan-Newman modular networks of size $N = 100$ and average degree $\langle k \rangle = 10$ with three different values of average degree within communities $\langle k_{in} \rangle$. From left to right, $\langle k_{in} \rangle = 6$, $\langle k_{in} \rangle = 8$ and $\langle k_{in} \rangle = 9$.

1.3 Processes on Complex Networks

Section 1.2 looks into the fundamentals of the structure of complex networks. In this section, we present some of the most relevant results of dynamical systems on complex networks [43, 122, 21].

1.3.1 Dynamical Systems on Networks

Dynamical systems theory is a mathematical field that studies the behaviour of dynamical systems, i.e., systems that, at any given time, are represented by a vector of real numbers, which evolution is usually defined by differential (continuous time) or difference (discrete time) equations which can be either deterministic or stochastic.

In particular, complex systems theory studies the common and emergent properties of dynamical systems which are complex in nature because they are

made of many components that interact with each other or, equivalently, they have a large number of degrees of freedom (see Section 1.1 for a detailed explanation). Researchers in this field study the global properties of systems, rather than the behaviour of their isolated constituents. The synchronization in a system of interacting oscillators [15, 139, 131] or the epidemic spreading in a network of connected individuals [118] are examples of such global properties.

The ideas of dynamical systems can thus be applied to dynamical systems on networks. Such systems are typically characterized by independent dynamical variables, $x_i, y_i, ...$ on each node that are coupled together only through the edges of the network. In other words, the time evolution equation of an individual variable x_i on node i can involve x_i, other variables on node i, or one or more variables of a node which is adjacent to i in the network, i.e., variables of neighbouring nodes.

For a system with a single variable on each node and considering pairwise interactions, the general first-order equation describing the dynamics of node i can be written as

$$\frac{dx_i}{dt} = f_i(x_i) + \sum_j A_{ij} g_{ij}(x_i, x_j) \tag{1.33}$$

where f_i is a function of the intrinsic dynamics of node i and g_{ij} describes the contribution of the connection $i - j$[15]. Note that the sum over j of A_{ij} ensures that only neighbouring nodes contribute to the evolution of a given node.

In general, nodes are representations of the same element - for example cells, people or neurons - and hence, the dynamics may be identical. For this reason, the functions f_i and g_{ij} are the same for all nodes and Eq.(1.33) turns to

$$\frac{dx_i}{dt} = f(x_i) + \sum_j A_{ij} g(x_i, x_j) \tag{1.34}$$

Suppose we are interested in finding the fixed points $\{x_i^*\}$ of the system[16]. The latter are obtained by solving the simultaneous equations

$$f(x_i^*) + \sum_j A_{ij} g(x_i^*, x_j^*) = 0 \ \forall i \tag{1.35}$$

Note that the position of the fixed point depends on both the dynamics of the system and the topology of the network.

[15] Higher-order interactions is the natural extension of pairwise interactions. As a proper framework for these type of interactions, the connection between simplicial complexes – a notion based on the set of possible generalized triangles – together with hypergraphs – interactions between an arbitrary number of nodes through hyperedges – and complex systems is being recently explored[151, 23].

[16] A fixed point is a steady state of the system, i.e., any set of values of the variables which do not change over time.

In order to analyse the linear stability of the fixed point we expand Eq.(1.34) about $\{x_i^*\}$ considering $x_i = x_i^* + \varepsilon_i$, where ε_i is a small quantity and hence keeping only linear terms. Finally,

$$\frac{d\varepsilon_i}{dt} = \left[\alpha_i + \sum_j \beta_{ij} A_{ij}\right] \varepsilon_i + \sum_j \gamma_{ij} A_{ij} \varepsilon_j \tag{1.36}$$

where $\alpha_i = \frac{\partial f}{\partial x}\Big|_{x=x_i^*}$, $\beta_{ij} = \frac{\partial g(u,v)}{\partial u}\Big|_{\substack{u=x_i^* \\ v=x_j^*}}$ and $\gamma_{ij} = \frac{\partial g(u,v)}{\partial v}\Big|_{\substack{u=x_i^* \\ v=x_j^*}}$.

In matrix form, Eq.(1.36) can be written as

$$\frac{d\vec{\varepsilon}}{dt} = M\vec{\varepsilon} \tag{1.37}$$

where

$$M_{ij} = \left[\alpha_i + \sum_j \beta_{ij} A_{ij}\right] \delta_{ij} + \gamma_{ij} A_{ij} \tag{1.38}$$

Eq.(1.37)can be solved by writing $\vec{\varepsilon}$ as a linear combination of the right eigenvectors[17] of the matrix M:

$$\vec{\varepsilon}(t) = \sum_k c_k(t)\, \vec{v}_k, \tag{1.39}$$

where $c_k(t)$ is the coefficient corresponding to the contribution of the kth eigenvector $\vec{v}_k$. Using Eq.(1.39), Eq.(1.37) becomes

$$\sum_k \frac{dc_k}{dt} \vec{v}_k = \sum_k \lambda_k c_k(t)\, \vec{v}_k, \tag{1.40}$$

where λ_k is the eigenvalue corresponding to $\vec{v}_k$ eigenvector. Finally, for each mode, i.e., in terms of each eigenvector we have

$$\frac{dc_k}{dt} = \lambda_k c_k(t), \tag{1.41}$$

which solution is given by

$$c_k(t) = c_k(0) e^{\lambda_k t}. \tag{1.42}$$

From Eq.(1.42) we conclude that if the real part of all of the eigenvalues λ_k is negative, all modes $c_k(t)$ decay in time, as well as $\vec{\varepsilon}$. In other words, the corresponding fixed point is a stable one[18]. Conversely, if all of the $\Re(\lambda_k)$ are positive, the associated fixed point is unstable. A combination of negative and positive

[17] In general, M is not a symmetric matrix

[18] Note that $c_k(t)$ is the contribution of mode k to $\vec{\varepsilon}$. But ε_i is the deviation of variable x_i from the fixed point element x_i^*.

values of $\Re(\lambda_k)$ leads to a saddle fixed point, that is, a position which at least has one repelling direction[19].

As a final remark of this very general framework for the stability analysis of the fixed points of a dynamical system, we highlight the fact that, very importantly, both the dynamical process characteristics and the topology of the network are enclosed in the spectra of M matrix. Therefore, a particular analysis is required for each pair of dynamical process and network.

1.3.2 Coupled Phase Oscillator Models

Many real systems of interest can be mathematically described as oscillatory systems, that is, an ensemble of units that are individually modelled as oscillators of one sort or another, but that they are coupled with the neighbours through the connections of the network. The flashing of fireflies, the neuronal brain signals or the energy flow through the power grid are examples of oscillatory systems [4, 15]. Many biological, technological and even socio-economical systems are rightly described as networks of couple phase oscillators. Very often, the dynamics of the oscillators considers that neighbouring nodes are coupled through its phases and regardless of the amplitude of the oscillations.

Within this framework, researchers have drawn particular attention to the study of *synchronization*, that is, the whole set (or a fraction) of network oscillators being locked at the same frequency [143, 15, 139, 131].

Despite results on several oscillatory ensembles have been obtained, such as pulse-couple models or coupled map systems, we focus our attention on the *limit cycle oscillators*.

1.3.2.1 Limit Cycle Oscillators: the Kuramoto Model

Already in 1665 Christiaan Huygens realized that two pendulum clocks, its own invention, suspended side by side on a wall could synchronize in frequency, swinging in opposite directions. He called the phenomena 'an odd kind of sympathy' in his letter to the *Royal Society* and considered that it was due to the 'imperceptible movements' of the common supporting structure [180]. This effect for a few number of oscillators is already well studied. However, when examining the analogue effect for large population of oscillators the mathematical treatment requires a different approach.

[19] Additionally to fixed points, some dynamical systems present *limit cycles*, which implies that the system remains indefinitely trapped on a loop of the dynamics. Physically, they represent stable oscillatory dynamics of systems. Nonetheless, limit cycles can also be repelling or attracting, namely, a small displacement from it may tend to the limit cycle or move away from it.

In 1948 Norbert Wiener published the book *Cybernetics* where he asked 'How is it that thousands of neurons or fireflies or crickets can suddenly fall into step with one another, all firing or flashing or chirping at the same time, without any leader or signal from the environment?' [178] The first attempt to mathematically model this non-linear collective dynamics was done by Arthur Taylor Winfree in 1967 [181]. He considered all-to-all interacting[20] weakly coupled limit-cycle oscillators characterized by intrinsic frequencies from an uni-modal probability distribution. His main contribution was considering only the phase of the oscillators and neglecting their amplitude. Finally, although not proved analytically, he observed that, even if different, the oscillators could display a relatively collective coherent behaviour when the frequencies were close enough.

After the work of Winfree and following his approach, in 1975 Yoshiki Kuramoto [92, 93] came up with a tractable mathematical model that could capture the phenomenology of collective synchronization even if several simplifications were assumed. He suggested that oscillators were coupled by a sinusoidal function of their phase differences as follows

$$\frac{d\theta_i}{dt} = \dot{\theta}_i = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i) \qquad \forall i \in [1,...,N], \tag{1.43}$$

where θ_i denotes the phase of the ith oscillator, K is the coupling strength of the interactions and ω_i the intrinsic frequency of node i. The factor $1/N$ ensures that the systems behaves correctly in the thermodynamic limit, $N \to \infty$. We note that in his original work, Kuramoto considered a mean field approach, i.e., all oscillators interacting simultaneously with all other units. The frequencies are obtained from a given probability distribution $g(\omega)$, which is usually assumed to be uni-modal and symmetric about its mean value Ω. Additionally, due to the rotational symmetry of the model, a rotating frame $\theta_i \to \theta_i - \Omega t$ is normally assumed[21]. After this shift, $g(\omega) = g(-\omega)$ and thus, ω_i corresponds to deviations from the mean frequency [15].

The collective behaviour of the oscillators can be measured by a complex-valued order parameter defined as

$$Z = r(t)e^{i\phi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j(t)}, \tag{1.44}$$

where the modulus $r(t)$ captures the macroscopic coherence of all oscillators and $\phi(t)$ represents the average phase. $r(t)$ is usually considered as the effective order parameter of the system, ranging $0 \le r(t) \le 1$. The limits $r(t) \approx 0$ and $r(t) \approx$

[20] Also known as mean field model

[21] This assumption is essentially based on the fact that the system defined in Eq.(1.43) remains invariant by a uniform rotation $\theta_i \to \theta_i + \theta_0$.

1[22] correspond to all oscillators moving incoherently and being phase locked, i.e., frequency synchronized, respectively.

By means of Eq.(1.43) and despite its simplicity, the Kuramoto model (KM) is able to capture the phase transition between the purely chaotic state, where all oscillators move independently, to a coherent state, where more and more oscillators reach the frequency synchronized state and eventually end with all units swinging in unison. When the variance of the natural frequencies distribution $g(\omega)$ is too wide compared to the strength of the coupling parameter K, the oscillators keep moving incoherently from each other. Conversely, once the coupling strength is able to balance the width of the distribution, a sub-population of oscillators starts locking their phases with respect to the others and hence become synchronized at an average frequency. This threshold represents the onset of synchronization, which is usually represented by the critical coupling K_c. The oscillators which natural frequency is too separate from that of the locked population remain moving at a different velocity. Therefore, the population is described by a partially synchronized state. As the coupling strength grows, more and more oscillators join in the coherent group, swinging around a mean frequency. Eventually, the entire population ends up locked in phase oscillating accordingly to a collective average rhythm (See Figure 1.6).

Mathematically, two oscillators i and j are frequency synchronized when

$$\dot{\theta}_i(t) = \dot{\theta}_j(t) \qquad \forall\, t > t_s, \tag{1.45}$$

where t_s is the time step after which both oscillators become synchronized. The whole system becomes synchronized when Eq.(1.45) is true for all pair of oscillators.

Additionally to frequency synchronized, two oscillators may become phase synchronized, a more restrictive state, when their phases have the same value after a certain point in time. Mathematically,

$$\theta_i(t) = \theta_j(t) \qquad \forall\, t > t_s. \tag{1.46}$$

Notice that two oscillators that are phase synchronized are consequently frequency synchronized. When Eq.(1.46) is true for all pair of oscillators, the corresponding state is known as fully synchronized state and the order parameter of the system takes exactly its maximum value, i.e., $r = 1$.

Nonetheless, Eq.(1.43) does not include the topology of the interactions between oscillators. Therefore, in order to include the connectivity pattern in the dynamical model, Eq.(1.43) can be modified as follows:

$$\dot{\theta}_i = \omega_i + K \sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i) \qquad \forall\, i \in [1, ..., N], \tag{1.47}$$

[22] The approximate symbol corresponds to fluctuations of $\sim O(\sqrt{N})$

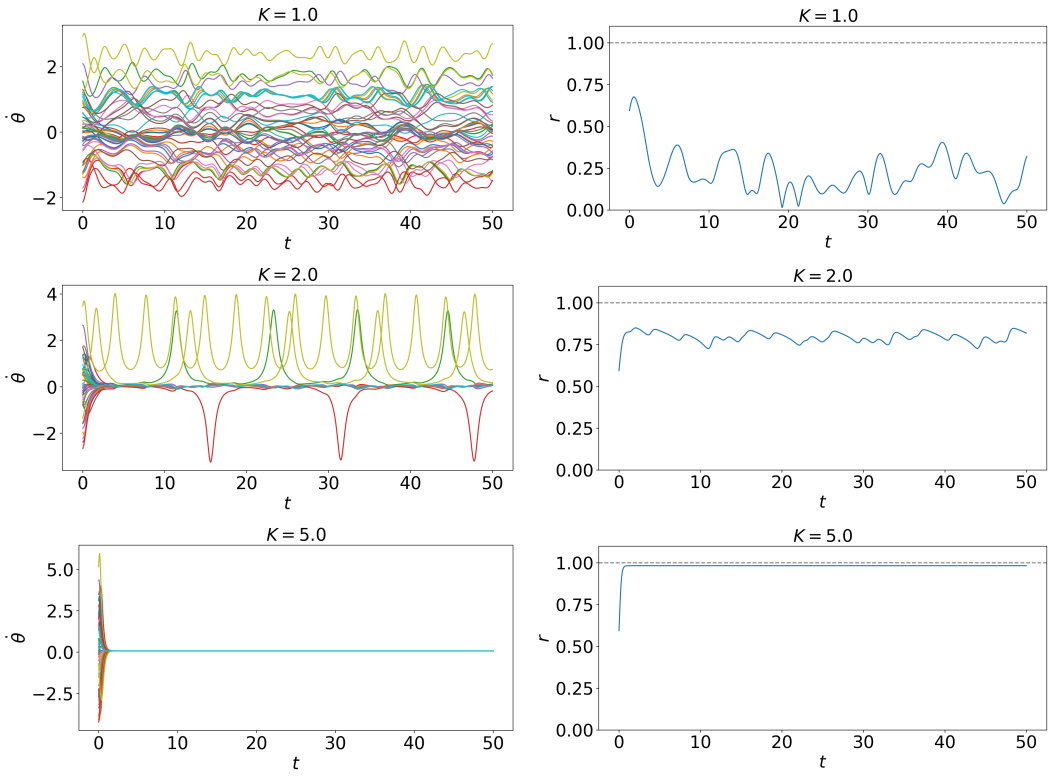


Figure 1.6: **Time evolution of the oscillators' instantaneous frequencies $\dot{\theta}$ and the corresponding order parameter r for a complete network of 50 nodes.** Time evolution of the oscillators' instantaneous frequencies $\dot{\theta}$ (first column) and the corresponding order parameter r (second column) for a complete network of 50 nodes, considering natural frequencies distributed according to a Gaussian distribution with unit variance $g(\omega) \sim \mathcal{N}(0,1)$ and three different values of the coupling strenght K. From top to bottom, $K = 1$, $K = 2$ and $K = 5$.

where A_{ij} is the element (i, j) of the adjacency matrix of the corresponding network. Several definitions of the coupling strength K can be considered to ensure a correct behaviour of the system in the thermodynamic limit. Usually, the effective value of the coupling is obtained by dividing it by the maximum degree of the network, k_{max}. Thereby, different topologies can be easily compared [137].

Much attention has been devoted to numerically exploring and analytically deriving (under certain assumptions) the onset of synchronization. In other words, researchers are interested in characterizing the value of the critical coupling strength beyond which clusters of nodes become frequency synchronized, given a particular network topology. Although this topic is not explicitly studied in the present work, the reader is referred to Refs.[165, 114, 15, 142, 50] for a deeper understanding of the topic.

1.3.2.2 The Kuramoto-Sakaguchi Model

In 1986, Yoshiki Kuramoto together with Hidetsugu Sakaguchi presented a generalization of the previous limit-cycle set of oscillators Kuramoto's model (see Eq.1.43) which incorporated a constant phase lag α between oscillators[23] and was originally written as [150]

$$\dot{\phi}_i = \omega_i - \sum_{j=1}^{N} K_{ij} \sin(\phi_i - \phi_j + \alpha), \qquad |\alpha| \leq \pi/2, \tag{1.48}$$

where ϕ_i represents the phase of the i-th oscillator [150]. They considered the case of uniform coupling, i.e., $K_{ij} = K/N$ and, for this particular case, derived analytical expressions such as the order parameter and the critical coupling for the onset of synchronization, as well as compared them with computer simulations.

Later studies of the Kuramoto-Sakaguchi model (KSM) [128, 124, 158, 38, 145, 146] included the network structure within the model together with the global shift α - or frustration - and considered identical oscillators, i.e., $\omega_i = \omega$ for all nodes. Therewith, the phase of each oscillator is governed by the equation

$$\dot{\theta}_i = \omega + K \sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i - \alpha) \qquad \forall i \in [1, ..., N], \tag{1.49}$$

where A_{ij} accounts for the connectivity between nodes i and j and the phase lag parameter α takes values in the range $[0, \pi/2]$.

For a wide range of $\alpha > 0$, the system becomes synchronized to a resulting frequency Ω, i.e., the dynamics reaches a stationary state. Nevertheless, α forces

[23] They suggested that because, empirically, two strongly coupled oscillators swing with a common frequency different from their average natural frequencies, a phase lag was needed in their functional interaction.

connected nodes to be locked in phase and hence breaks the phase synchronization. The magnitude of such locking is determined by both the network topology and the parameters of the dynamics. However, full synchronization is conserved for topological symmetric nodes; a phenomenon that has been called *remote synchronization* [124]. As the frustration increases, the asynchronous groups move away from each other until they reach the maximum separation, beyond which the system becomes chaotic. The threshold value α_C depends on the topology and the parameters of the dynamics.

In other words, when $\alpha = 0$ the system reduces to a network of identical Kuramoto oscillators. In this case, the fully synchronized state is globally stable[24].

Conversely, when $\alpha \neq 0$ the frustration parameter makes directly connected nodes to maintain a constant phase lag and hence forces the system to break the otherwise fully synchronized state. When the system reaches a stationary state, in which all oscillators are frequency synchronized, partial phase synchronization is conserved for nodes belonging to the same orbit in the network[25].

Figure 1.7 shows the temporal evolution of the Kuramoto-Sakaguchi dynamics for the network in Figure 3.1 considering three different values of the frustration parameter. In the cases the system achieves the stationary state (upper and middle panels) the nodes are divided into four equivalent clusters, as clearly reflected in the four separate dynamics. Differently, the lower panel shows a chaotic dynamic present when $\alpha > \alpha_C$, a threshold which depends on every system.

In order to quantify the level of synchronization between pairs of oscillators, we define a local order parameter between oscillators, based on Ref.[14]:

$$\rho_{ij}(t) = \cos(\theta_i(t) - \theta_j(t)) \tag{1.50}$$

Equation (1.50) measures the correlation between pairs of oscillators in the stationary regime, which is invariant under temporal translation.[26] We outline that the pairs of nodes that are structurally equivalent have a value of $\rho_{ij} = 1$.

Figure 1.8 shows the matrix representation of the local order parameter between all pair of nodes corresponding to the scenario described in the lower

[24] A proof of this statement is provided in Ref.[4], where the authors consider mean-field interactions in the limit of strong coupling $K \to \infty$ and $N \to \infty$, for the case of unimodal natural frequency distributions. The incoherent and partial synchronization states, as well as finite size effects are also analysed. Ref.[15] considers the more general case of complex interactions through network topologies and studies the stability of the completely synchronized state of populations of identical oscillators using the master stability function (MSF) formalism. Moreover, the fully synchronized state can coexist with other attractors of the system under certain conditions. The authors in Ref.[179] study the size of the synchronization basin.

[25] The vertices or nodes of the same orbit are structurally indistinguishable and play the same structural role in the network.

[26] The original definition of the measure is time dependent. We are concerned only with the stationary regime though.

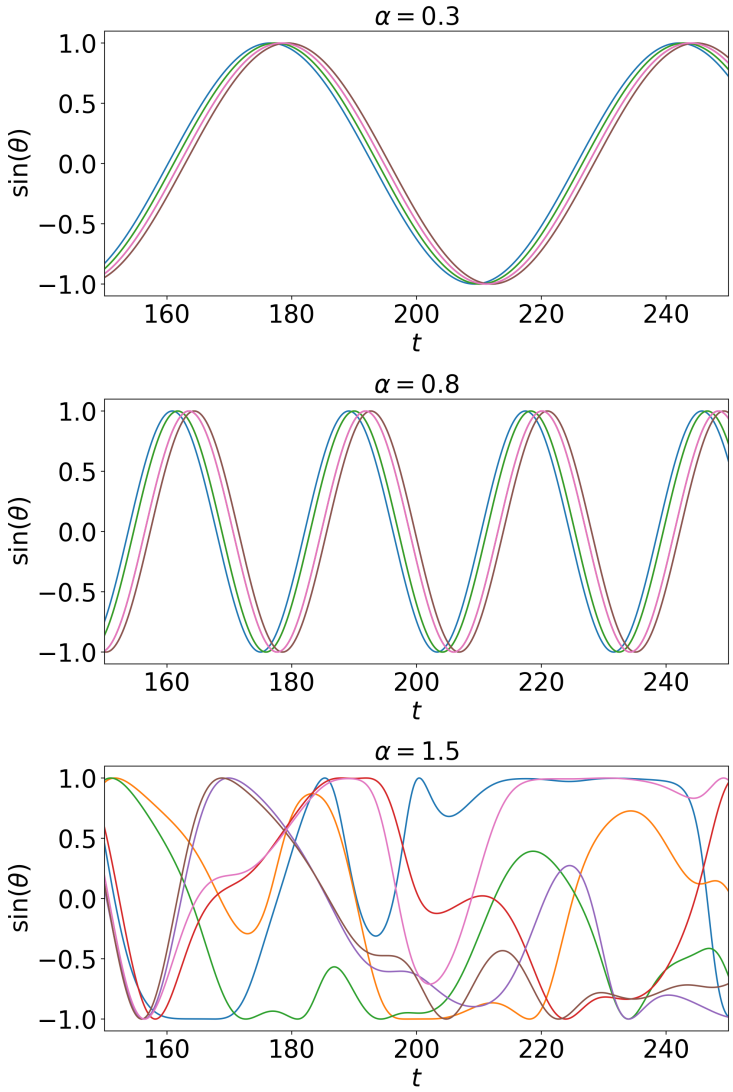


Figure 1.7: **Time evolution of the oscillators' phases θ for the network in Figure 3.1.** Time evolution of the oscillators' phases θ corresponding to the Kuramoto-Sakaguchi dynamics defined in Eq.(1.49) for the nodes of the network in Figure 3.1. The parameters of the model are set to $K = 1$ and $\omega = 0$ and the initial phases are distributed following a normal unimodal distribution $\mathcal{N}(0,1)$. The frustration parameter α is set to 0.3 (upper panel), 0.8 (middle panel) and $\alpha = 1.5$ (lower panel).

panel in Figure 1.7 (Kuramoto-Sakaguchi dynamics with $\alpha = 0.8$). Therein, four groups (orbits) are internally (phase) synchronized but remain asynchronous between them. The groups obtained capture the natural symmetries of the network: nodes 1 and 2, nodes 3 and 6, nodes 4 and 5 and node 0, which behaves uniquely. The Kuramoto-Sakaguchi model is extended to the General

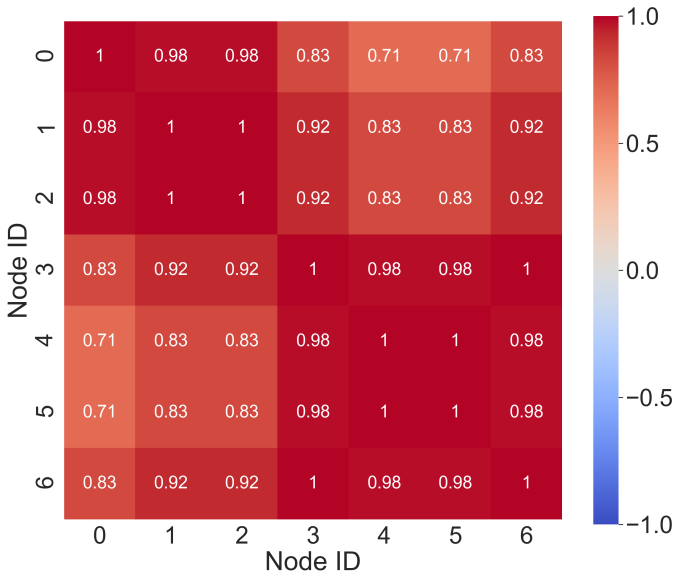


Figure 1.8: **Matrix representation of the local order parameter.** Matrix representation of the local order parameter defined in Eq.(1.50) between nodes corresponding to the scenario described in the lower panel in Figure 1.7.

Kuramoto-Sakaguchi model in Chapter 2.

Obtaining the analytical expression of the temporal evolution of the system of coupled non-linear differential equations defined in Eq.(1.49) for each oscillator $\theta_i(t)$ becomes an impossible task, even for small systems[27]. Nonetheless, the work presented throughout the following chapters relies on two basic assumptions or considerations: the system has reached a stationary state, i.e., it is not chaotic, and such scenario corresponds to the frequency synchronized state, i.e., all oscillators moving with the same instantaneous frequency, $\dot{\theta}_i$. Mathematically, $\dot{\theta}_i = \text{const.}$ and $\dot{\theta}_i = \dot{\theta}_j \ \forall(i,j)$. These two assumptions imply several restrictions on the parameters and initial conditions of the model that one should

[27] Solving numerically the dynamics of the system in order to obtain interesting insights may be, thus, a convenient approach.

consider when defining the initial configuration of the non-linear model (See Figure 1.7 for an example of stationary and chaotic behaviours of the system).

For small values of the frustration parameter and phases close to each other, i.e., $\alpha \approx 0$ and $\theta_j - \theta_i \approx 0$, which is the case in the frequency synchronized state, Eq.(1.49) can be linearized as follows:

$$\dot{\theta}_i \approx \omega + K \sum_{j=1}^{N} A_{ij}(\theta_j - \theta_i - \alpha), \tag{1.51}$$

where the sinus is replaced by its argument[28]. The linearized Kuramoto-Sakaguchi dynamics can be expanded as

$$\dot{\theta}_i = \omega - K \sum_{j=1}^{N} L_{ij}\theta_j - \alpha K k_i, \tag{1.52}$$

where L is the Laplacian matrix, defined as $L_{ij} \equiv \delta_{ij} k_i - A_{ij}$, and k_i is the degree of the ith node [122].

In the synchronized state all oscillators move coherently at the same constant collective rhythm, namely, $\dot{\theta}_i^* = \Omega$, where $\theta_i^*(t)$ corresponds to the evolution of the phases at the stationary state. We can derive the value of the common frequency oscillation Ω summing Eq.(1.52) over i index. Finally,

$$\Omega = \omega - \alpha K \langle k \rangle, \tag{1.53}$$

where we have used the Laplacian matrix property $\sum_i L_{ij} = 0$. Combining Eq.(1.52) and Eq.(1.53) we obtain the closed expression for the phases at the stationary state:

$$\sum_j L_{ij}\theta_j^* = \alpha(\langle k \rangle - k_i) \tag{1.54}$$

Eq.(1.54) corresponds to an undetermined system of linear equations due to the singular nature of the Laplacian matrix. In particular, there is one degree of freedom which corresponds to the phase of an arbitrary oscillator, which we should use as a reference value for the solution. In this way, we do not work directly with the functional form of phases because they are time dependent, i.e., $\theta_i^*(t) = f(t)$, but with the phase differences with respect to such reference node, once the stationary state is achieved,

$$\phi_i \equiv \theta_i - \theta_R, \tag{1.55}$$

where R corresponds to the label of the chosen reference node. By definition, $\phi_R = 0$.

To easily write the matrix expressions, we define the s*election matrix* $J_{(n,m)}$, which is, in general, an $(N-1) \times (N-1)$ identity matrix after the removal of the

[28] $sin(x) \approx x$ when $x \to 0$.

mth row and the nth column. Using the selection matrix, we define the *reduced Laplacian matrix* as

$$\tilde{L}(C,R) = J_{(,C)} \cdot L \cdot J_{(R,)} \equiv \tilde{L}. \tag{1.56}$$

In other words, $\tilde{L}$ corresponds to a (non-singular) square matrix of dimension $N-1$ after removing the Cth row and the Rth column from the original Laplacian matrix L. Notice that the index of the removed column corresponds to that of the reference node R, while the choice of the index C for the removal of the row is left free[29]. With these remarks and using the new set of variables $\{\phi_i\}$, Eq.(1.54) turns to

$$\sum_j \tilde{L}_{ij}\phi_j^* = \alpha(\langle k\rangle - k_i). \tag{1.57}$$

Finally, the solution for the phase differences $\{\phi_i\}$ in Eq.(1.57) in matrix form is given by

$$\vec{\tilde{\phi}}^* = \alpha\tilde{L}^{-1}(\langle k\rangle\,\vec{e} - \vec{\tilde{k}}), \tag{1.58}$$

where $\vec{e}$ is a vector of ones of dimension $N-1$ and the ith element of the vector $\vec{k}$ corresponds to k_i. $\vec{\tilde{k}}$ and $\vec{\tilde{\phi}}$ are defined as $J_{(,C)}\cdot\vec{k}$ and $J_{(,C)}\cdot\vec{\phi}$, respectively, i.e., the vector of node degrees and phase differences after the removal of the Cth entry [30].

Considering the topology defined in Figure 3.1, we derive the solution of the phase differences at the stationary regime (frequency synchronized state) to illustrate the procedure concluding at Eq.(1.58). The Laplacian matrix corresponding to the network in Figure 3.1 is given by

$$L = \begin{pmatrix} 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}. \tag{1.59}$$

The corresponding vector of node degrees is $\vec{k} = (4,2,2,2,2,2,2)$. Next, we select node 0 to be the reference node. Hence, $R = 0$. Note that, by definition, $\phi_0 = 0$. In order to compute the reduced Laplacian matrix we choose $C = 0$, an arbitrary

[29] Which equation to remove is left as a free choice in a underdetermined system of linear equations.

[30] An alternative approach to derive the phases at the stationary state is suggested in Ref.[158]

choice which does not affect the solution. Therefore, $\tilde{L} = J_{(,0)} \cdot L \cdot J_{(0,)}$. Particularly,

$$\tilde{L} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}. \tag{1.60}$$

Applying Eq.(1.58) we obtain the solution

$$\vec{\tilde{\phi}}^* = \frac{\alpha}{7}(2,2,4,6,6,4), \tag{1.61}$$

and $\phi_0 = 0$, by definition. Therefore, we obtain the four different dynamics according to the numerical simulation (see Figure 1.7): $\phi_1 = \phi_2$, $\phi_3 = \phi_6$ and $\phi_4 = \phi_5$, which numerical final value depends on the choice of the frustration parameter α and that captures the four different clusters of structurally equivalent nodes present in the network of Figure 3.1.

1.4 Outline of the thesis

After having reviewed the main concepts of network theory and revisited the principal results concerning limit-cycle oscillatory network models, this thesis explores the potential of considering a non-homogenous distribution of phase lag parameters among the population of oscillators, the main variation of the Kuramoto-Sakaguchi model with respect to the original Kuramoto model.

In Chapter 2 we consider the scenario of perturbing the otherwise fully synchronized state of the system by introducing a non-zero phase lag shift into the dynamics of a single node. In this way we obtain a rating of the nodes considering its potential to move the system away from the ground steady state. To this end, we define a novel centrality measure, which we call *funtionability*, that provides interesting insights into the network structure and node function.

In Chapter 3, we turn our attention to a second scenario by considering a more general setting in which the phase lag parameter is an intrinsic property of each node, not necessarily zero, and hence exploring the potential heterogeneity of the frustration among oscillators. In this second work, we bring forward a methodology to drive the system into any desired phase state, by means of a fine tuning of the phase lag distribution. In this way, the three intrinsic parameters of the nodes in the model, natural frequencies, frustration parameters and the phases in the steady state are coupled by an equation. In addition, we also address the question of finding, not only a plausible solution, but the optimal

solution such that the system requires the minimum cost to achieve a particular state.

Finally, in Chapter 4 the homogeneous scenario of phase lag distribution is revisited. We explore the distribution of obtained phase differences in this particular case and construct a framework that enables the detection of, not only the perfect symmetries of a network, but also approximate symmetries, defined as *quasi-symmetries*. We study the distributions of topological similarity among all pair of nodes and find a benchmark to determine whether a network has a more complex pattern to that of a random network concerning quasi-symmetries. Moreover, we define the 'dual network', a weighted network – and its corresponding binarized counterpart – that effectively encloses all the information of the quasi-symmetries in the original one. The dual network allows for the analysis of centrality measures and community detection with respect to approximate symmetries.

The overall conclusions are presented in Chapter 5, although a discussion of each work is included at the end of each chapter.

CHAPTER 2

Functionability in Complex Networks

Synchronization has become one of the most paradigmatic examples of emergent properties in complex systems [139, 131], since the degree of interaction between the oscillatory units of a discrete system makes that a variety of macroscopic states are available. Among the most studied such systems, because of its inherent simplicity, is the Kuramoto model, presented in Section 1.3.2, in which phase oscillators interact continuously with other units through a sine function of the phase difference [93, 150, 4]. In all-to-all models, there is a transition from an incoherent state to a coherent one that depends only on the relative strength of the two competing forces: the dispersion of intrinsic frequencies, and the intensity of the coupling between units.

Over the last four decades, the KM has been thoroughly studied in regular lattices, and, with the sudden interest in complex networks, its role in irregular connectivity patterns has been heavily exploited [15]. This has been achieved not only by analyzing synchronization properties (order parameters, control parameters, time to synchronize, etc.) but also through use of the path to synchronization of neighboring units in order to identify higher-order connectivity patterns, for instance in communities that form complex networks at different hierarchical levels [14, 13, 12]. As already stated, in the original KM the emphasis is on the relative strength in the two antagonistic contributions: frequency dispersion and coupling strength. However, when complex topologies are considered, it is important to disentangle these effects; for this reason special interest has been arisen concerning the evolution of identical oscillators. In particular, a simple change in the coupling function, by inserting a phase-lag or frustration parameter in the argument of the sine function, results in identical oscillators now being unable to synchronize and it generates complex patterns of phase differences, a fact which have been related to topological symmetries of the network [124, 136].

The introduction of an identical frustration parameter in all the coupling terms produces a global effect on the network. However, in complex network science there is a special interest in understanding what role the individual nodes play in the behavior of the overall network. Many centrality measures that are key to this field have been considered for many years in the social sciences,

and there are more recent proposals, such as Google Search PageRank as well as other measures related to the concept of controllability [100]. However, to the best of our knowledge, there is no measure of the effect that a given node can have on the range of states a network can achieve which, in terms of phase oscillators, is measured in the different phase-difference patterns a network can traverse. This is precisely the goal of the present chapter: to characterize the effect a node can have on the global asynchronous state.

Among the systems where this new concept that we name 'functionability' has an immediate application is the rapidly evolving field of brain networks [163]. The molecular and cellular mechanisms of synapse formation and plasticity shape and guide the development and change of synaptic connections in the long run [95, 40]. Therefore, brain networks, on a short timescale, are considered to be static. Such structural networks are the substrate on which different temporal co-activation patterns can occur, also known as functional networks [3, 108, 84, 41, 154, 27]. All the functionalities of the brain, at either the low or the high level, are captured by different networks which nonetheless occur within the same physical medium. How does this essential feature of the brain arise? What mechanisms are responsible for a static network undergoing many different states? Are there specialized brain regions that are better at this job? Are they easy to identify?

Higher functionability may be positive for the system, as it reflects the capacity of a node to be involved in different tasks, and can result in the network state shifting into one involving more complex temporal relationships between modules. However, highly functional nodes can also be potentially dangerous in systems where tiny perturbations can produce cascade-like effects which completely disrupt the network dynamics. An example of this is offered by the transfer networks of power grids, which have been widely studied in the field of complex networks, focusing on their structure to assess the damage of failures [16, 10]. However, power grids are highly sensitive to oscillatory dynamics and the synchronization of AC power [94], and hence to perturbations of the phase lags between individual agents. Other such examples include the synchronization in heartbeats [5], multiprocessors and multi-core processors and traffic signal synchronization [175].

We are specifically interested in detecting the nodes that have a major impact on the network by enabling a broader spectrum of states or a larger dispersion from the 'ground' state. Many centrality measures have been developed and defined over the years [29], some of them are even related to the dynamical properties of the nodes; but we specifically target a measure of variability or functionability that can be associated with the the physical phenomenon of synchronization in order to provide it with meaning [139]. Hence, we base a great

deal of our work on an oscillatory dynamical model; but we aim to arrive at a compact mathematical expression that emerges from it.

All the results presented in this chapter can be found in Ref.[144].

2.1 The Generalized Kuramoto-Sakaguchi Model

In Section 1.3.2.1 we show that a connected system described by the KM dynamics reaches the fully synchronized state when all the oscillators are identical. Nonetheless, besides tuning the natural frequency of each oscillator, is there an alternative way of breaking the natural synchrony of the system?

As explained in Section 1.3.2.2, later studies [26, 128, 124] based on the work of Kuramoto and Sakaguchi [150] have suggested the introduction of a homogeneous phase 'frustration' parameter α into the dynamics of the system. Equation 1.49 already includes the network structure and identical oscillators are considered.

Nonetheless, neither the Kuramoto model nor the Kuramoto-Sakaguchi model provide information on specific nodes, but on the network as a whole. In this chapter, we require that each oscillator is tagged by an intrinsic parameter capable of moving the system away from its natural fully synchronized state. What would the effect of a phase frustration parameter that characterizes distinctly each such oscillator be? Several studies have focused on the effect of different natural frequencies of the oscillators; but we may be concerned with other types of natural properties connected to the phase shift between oscillators.

We would like to identify the nodes that have the largest effect in leading the whole system away from full synchronization. To do so, we need to establish which nodes have the greatest capacity, with only a small perturbation, to produce a large dispersion in the phases of the whole population. In the next paragraphs we build a model that is capable of breaking the natural phase synchronization and search for central nodes that are best suited for doing this. To this end, we introduce a dynamic model based on Kuramoto-Sakaguchi model: the General Kuramoto-Sakaguchi model (GKSM), which enables us to individually characterize each node by means of an intrinsic frustration parameter.

In the present chapter we consider the natural generalization of the KSM by considering the frustration phase parameter to be intrinsic to each oscillator, α_i, rather than a homogeneous property of the population. This assumption may depict a more realistic scenario, in which oscillators represent real systems with individual properties that are determined by the nature of each oscillator. The GSKM is defined by the dynamics:

$$\dot{\theta}_i = \omega + K \sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i - \alpha_i) \qquad \forall i \in [1, ..., N], \tag{2.1}$$

where α_i is an intrinsic parameter of each oscillator. Other considerations regarding the model can be found in Refs. [85, 38].

2.1.1 Finding the most functional nodes

We may find or define very distinct (discrete) distributions of the frustration parameter among the population, each of them leading the system to a different behavior. As our main goal is to localize the nodes that are best able to move the system away from synchronization, we consider two possible distributions, although other interesting insights may be obtained from alternative possibilities:

- Homogeneous distribution: $\alpha_i = \alpha_h \ \forall i$.

  In this case, the GKSM reduces to the KSM, as previously explored, where all nodes share the same value of α (See Section 1.3.2.2).

- Delta distribution: $\alpha_i = \alpha \cdot \delta_{iC}$, where $\delta_{iC} = 0$ if $i \neq C$ and $\delta_{iC} = 1$ if $i = C$.

  In this case, only the node labelled as C has a frustration parameter value different from zero. In this way, we break the overall problem of moving the system away from the fully synchronized state into many individual problems.

Let us first suggest a simple way to measure phase dispersion between oscillators.

2.1.2 Measuring perturbations in oscillatory systems

Suppose a given dynamical system defined by a network of coupled oscillators experiences a shift in the original configuration of parameters produced by an external mechanism. The system moves from an initial configuration of parameters, p, to a new set of parameters, q. Since each oscillator is characterized by a phase $\theta_i(t)$, the phases corresponding to the configuration p, $\vec{\theta}(p)$, will transform to updated phases in the configuration q, $\vec{\theta}(q)$. In other words, the system experiences a change of state. We assume that configurations p and q lead the system to two possible frequency synchronized stationary states, characterized by a set of values for the phase locking between the oscillators, $\vec{\phi}$.

In this situation, we define the effect on nodes (i, j), ε_{ij}, generated by the configuration shift from p to q as

$$\varepsilon_{ij}(p \to q) \equiv \frac{1 - \cos(\Delta\phi_{ij})}{2} \tag{2.2}$$

where $\phi_{ij} \equiv \theta_j - \theta_i$ and $\Delta\phi_{ij} \equiv \phi_{ij}(q) - \phi_{ij}(p)$ [182]. Therefore, ε_{ij} is a measure of the change in phase difference between nodes i and j that, since it is defined in

the stationary regime, is time independent. Eq.(2.2) has the following properties: $\varepsilon_{ij}(p \to q) = \varepsilon_{ji}(p \to q)$, $\varepsilon_{ii}(p \to q) = 0$ and $\varepsilon_{ij}(p \to p) = 0$. Moreover, $\varepsilon_{ij}(p \to q) \in [0,1]$. In other words, if nodes (i,j) were initially in phase and they change to be in anti-phase, the effect or the change in phase difference would be the largest possible: $\varepsilon_{ij}(\Delta\phi = \pi) = 1$. When no changes are produced due to the change of configuration, that is, the phase difference between nodes i and j remains unchanged, the value of the effect is zero: $\varepsilon_{ij}(\Delta\phi = 0) = 0$.

Further considerations regarding the properties of ε_{ij} as a distance metric can be found in the Appendix 2.4.1.

2.2 Functionability: a new centrality measure

In order to assess the impact on the whole system of a node being perturbed, we make use of the GKSM model described in Eq.(2.1) and the effect measure defined in Eq.(2.2). Using the delta distribution defined in Section 2.1.1, we will consider that the change in the configuration is produced by just one single node C called the *control node* and we will compute the *functionability* $\mathscr{F}_C$ of such node. By performing the same procedure for each node, we will obtain a vector $\vec{\mathscr{F}}$ with the functionability of each of them in its different entries.

The Control Node C. The initial configuration of the system p is such that all the oscillators are completely synchronized at the stationary state. The system then switches to configuration q, which is determined by the control node. This change is enacted by setting the set of frustration parameters $\vec{\alpha}$ in Eq.(2.1) as follows:

$$\alpha_i(p) = 0 \ \forall \ i \qquad \alpha_i(q) = \begin{cases} 0 \text{ if } i \neq C \\ \alpha \text{ if } i = C \end{cases} \tag{2.3}$$

where C is the label of the control node, the effect of whose perturbation on the whole system we will assess.

Functionability $\mathscr{F}$. We define the *functionability* of node C as:

$$\mathscr{F}_C(\alpha) \equiv \sum_i \sum_j \varepsilon_{ij}(p \to q(\alpha)) \tag{2.4}$$

where p and q are defined in Eq.(2.3), ε_{ij} is defined in Eq.(2.2) and the state of the system is obtained from the dynamics described by Eq.(2.1) with the aforementioned configurations. As already seen, the initial configuration p corresponds to the fully synchronized state. Therefore, the functionability measures the total dispersion of the phases from this ground state owing to the perturbation of a single node. The larger the dispersion, the more functionability a node has.

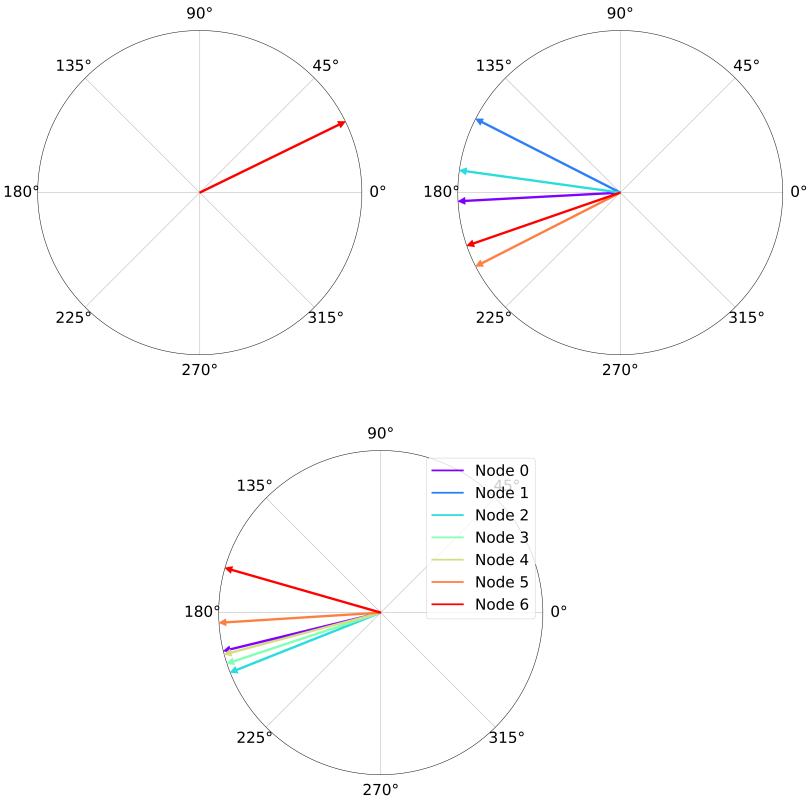


Figure 2.1: **Polar representation of the phases.** Polar representation of the final phases of the oscillators of the network in Figure 3.1 obtained from the dynamics described by Eq.(2.1) and conditions defined in Eq. (2.3). Upper left panel: The frustration parameter is set to $\alpha = 0$. The upper right and lower panels show the dispersion produced when nodes 1 and 6 are chosen as control nodes, respectively, when $\alpha = 0.5$.

In Figure 2.1 we observe that, regardless of the selected control node, if the frustration parameter is set to $\alpha = 0$, then the system is fully synchronized. Conversely, when $\alpha \neq 0$ – in the example considered, $\alpha = 0.5$ – the system becomes asynchronous, i.e., phases move apart from each other. We note that the effect of the frustration depends on the choice of the control node C and the quantification of this effect is captured by the corresponding value of the *functionability*.

2.2.1 Analytic expression of *functionability*

Equation (2.1) corresponds to a set of N coupled nonlinear differential equations whose solution, in general, cannot be derived analytically. However, if the system reaches a frequency synchronized state, that is, $\dot{\theta} = \Omega \; \forall \; (i, t > t_s)$, and the argument of the sine is small enough, we can linearize Eq.(2.1) similarly as it is done for the Kuramoto-Sakaguchi model in Section 1.3.2.2. Accordingly, Eq.(2.1) turns to

$$\dot{\theta}_i \approx \omega - K \left[\sum_{j=1}^{N} L_{ij}\theta_j + \alpha_i d_i \right], \tag{2.5}$$

where, for convenience, the symbol d_i corresponds to the degree of node i.

Without loss of generality, we can set $\omega = 0$ and $K = 1$[1]. With these remarks and equivalently to Eq.(1.53), the value of the common frequency is given by

$$\Omega = -\frac{1}{N}\sum_{i}^{N} d_i\alpha_i \equiv -\langle \alpha d \rangle \tag{2.6}$$

Finally,

$$\sum_{j=1}^{N} L_{ij}\theta_j = \langle \alpha d \rangle - \alpha_i d_i \tag{2.7}$$

In a connected graph, the Laplacian matrix has one null eigenvalue, which corresponds to the eigenvector $\vec{1}$, and hence the system of linear equations to solve $\vec{\theta}$ in Eq.(2.7) is singular. Intuitively, we are left with one free parameter which depends on the initial phase conditions, $\vec{\theta}(t = 0)$. Nonetheless, the phase differences between oscillators are well determined.

In accordance with the procedure used in Section 1.3.2.2, we need to work with the phase differences between oscillators $\theta_i - \theta_j$, instead of the corresponding absolute phases θ_i. To this end, we replicate Eq.(1.55),

$$\phi_i(R) = \phi_{iR} \equiv \theta_i - \theta_R, \tag{2.8}$$

[1] A non-zero shared natural frequency does not affect the synchronization of the system, and the coupling strength plays a role in the timescale of the path to synchronization.

where R corresponds to the label of a chosen reference node, with respect to which the set of phase differences are computed, once the stationary state is achieved. By definition, $\phi_R(R) = 0$.

Using the reduced Laplacian matrix $\tilde{L}(C,R)$ defined in Eq.(1.56) and considering q configuration defined in Eq.(2.3), Eq.(2.6) turns to:

$$\Omega = -\frac{1}{N}\sum_{i}^{N} d_i\alpha_i = -\frac{\alpha d_C}{N} \;\forall i \tag{2.9}$$

where C is the label of the control node. The GKSM in this particular case becomes

$$-\frac{\alpha d_C}{N} = -\left[\sum_{j=1}^{N} L_{ij}\theta_j + \alpha_i d_i\right] \;\forall i \tag{2.10}$$

The analytical solution of the set of phase differences for a given choice of control and reference nodes $\vec{\phi}(C,R)$ in matrix form is given by

$$\vec{\phi}(C,R) = \left[-\tilde{L}(C,R)\right]^{-1}\vec{\Omega} = \left[-\tilde{L}(C,R)\right]^{-1}\left(-\frac{\alpha d_C}{N}\right)\vec{1} \tag{2.11}$$

Hence,

$$\vec{\phi}(C,R) = \left[\tilde{L}(C,R)\right]^{-1}\left(\frac{\alpha d_C}{N}\right)\vec{1} \tag{2.12}$$

In order to calculate the matrix ε, whose elements are defined in Eq.(2.2)

$$\varepsilon_{ij}(p\to q) \equiv \frac{1-\cos(\Delta\phi_{ij})}{2}$$

we consider the fact that $\Delta\phi_{ij}(p\to q)\sim 0$ and thus we can linearize the cosine[2] and write ε_{ij} as:

$$\varepsilon_{ij}(p\to q) \approx \frac{\left(\Delta\phi_{ij}\right)^2}{4}. \tag{2.13}$$

To compute the *functionability* of a control node C as it is defined in Eq.(2.4), we use Eq.(2.12) and Eq.(2.13), where node R and node j are equivalent.

$$\varepsilon_{iR}(p\to q) \approx \left(\frac{\alpha d_C}{2N}\right)^2\left(\sum_{l}\left[\tilde{L}^{-1}(C,R)\right]_{il}\right)^2 \tag{2.14}$$

$\mathscr{F}_C \equiv \sum_i\sum_j \varepsilon_{ij}(p\to q)$, or $\mathscr{F}_C \equiv \sum_i\sum_R \varepsilon_{iR}(p\to q)$ can be expanded in order to obtain a more compact expression. Using Eq.(2.14) and rearranging summations:

$$\mathscr{F}_C = \left(\frac{\alpha d_C}{2N}\right)^2\sum_i\sum_R\left(\sum_l\left[\tilde{L}^{-1}(C,R)\right]_{il}\right)^2 = \left(\frac{\alpha d_C}{2N}\right)^2\sum_R\sum_i\left(\sum_l\left[\tilde{L}^{-1}(C,R)\right]_{il}\right)^2 \tag{2.15}$$

[2] $\frac{1-cos(x)}{2} = \sin^2\left(\frac{x}{2}\right) \approx \frac{x^4}{4}$ when $x\to 0$.

The last two summations can be expanded to

$$\sum_i \left(\sum_l \left[\tilde{L}^{-1}(C,R) \right]_{il} \right)^2 = \sum_i \sum_l \left[\tilde{L}^{-1}(C,R) \right]_{il} \sum_m \left[\tilde{L}^{-1}(C,R) \right]_{im} =$$
$$= \sum_l \sum_m \sum_i \left[\tilde{L}^{-1}(C,R) \right]_{il} \left[\tilde{L}^{-1}(C,R) \right]_{im} = \sum_l \sum_m \sum_i \left[\left(\tilde{L}^{-1}(C,R) \right)^T \right]_{li} \left[\tilde{L}^{-1}(C,R) \right]_{im} =$$
$$= \sum_l \sum_m \sum_i \left[\left(\tilde{L}^T(C,R) \right)^{-1} \right]_{li} \left[\tilde{L}^{-1}(C,R) \right]_{im} \tag{2.16}$$

Finally,

$$\mathscr{F}_C = \left(\frac{\alpha d_C}{2N} \right)^2 \sum_R^N \sum_{ij}^{N-1} \left[\left(\tilde{L}(C,R) \tilde{L}^T(C,R) \right)^{-1} \right]_{ij} \tag{2.17}$$

where we have used the matrix property: $A^{-1}B^{-1} = (BA)^{-1}$. If we are interested in normalizing Eq.(2.17) we can divide it by the prefactor $1/N^2$:

$$\hat{\mathscr{F}}_C \equiv \frac{1}{N^2} \left(\frac{\alpha d_C}{2N} \right)^2 \sum_R^N \sum_{ij}^{N-1} \left[\left(\tilde{L}(C,R) \tilde{L}^T(C,R) \right)^{-1} \right]_{ij} \tag{2.18}$$

The values of *functionability* for the network in Figure 3.1 taking $\alpha = 0.2$ are $\vec{\mathscr{F}} = \{0.34, \mathbf{0.43, 0.43}, 0.18, 0.36, 0.36, 0.18\}$. As we can see from Figure 2.2, nodes 1 and 2 obtain the highest scores, whilst nodes 3 and 6 have the lowest. The ranking of nodes is preserved regardless of the value of the frustration parameter, since all the dependence in α is a quadratic prefactor. A radial plot of the final phases at a particular point in time is shown in Figure 2.1 when nodes 1 and 6 are chosen as control nodes in the q configuration, with a value of the frustration parameter of $\alpha = 0.5$ (the linear model is considered).

2.2.2 Interpretation of *functionability*

At the beginning of the present Section, we define the *functionability* of node C, $\mathscr{F}_C$, as a measure of the effect that a phase-lag parameter introduced in the intrinsic properties of such node has on the original fully synchronized state of the network. In other words, $\mathscr{F}_C$ measures the potential that a single node has to move the network oscillators out of phase synchrony, as shown in Figure 2.1. Despite the fact that the general definition of *functionability* is built from the GKSM (see Eq.(2.1)), a non-linear dynamical model, we have derived a very compact analytical expression of the measure, given by Eq.(2.17).

Getting back to the origin point of the model and, therefore, the physical justification of the dynamics, the interpretation of the final analytical expression of the functionability is the following: high values of $\mathscr{F}_C$ inform of the large potential that the position where such nodes are located within the network have

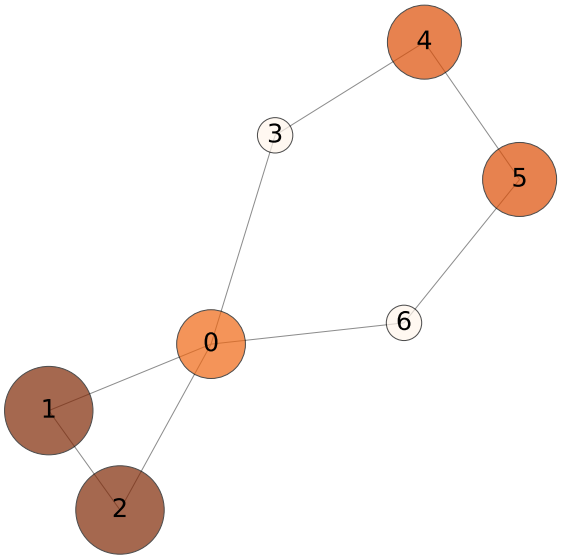


Figure 2.2: **Functionability of the nodes in a network.** Functionability values for the nodes of the network presented in Figure 3.1. Larger radius and darker colors correspond to higher functionability values. The frustration parameter is set to $\alpha = 0.2$.

in order to move the state of the whole system into more possible configurations, assuming the same perturbation is applied to individual nodes. As already pointed out, this effect may be either beneficial or disruptive for a given system, depending on its nature and functions.

To go one step further in the interpretation of functionability, we can take a closer look at its analytical expression, in Eq.(2.17), in order to understand the building blocks of it. Firstly, for fixed values of the network size N, and magnitude of the frustration parameter α, which can be included as a prefactor, we can identify two main contributions to $\mathscr{F}_C$:

1. The square of the degree of the node: d_C^2.

2. The contribution corresponding to the reduced Laplacian term, which we call *L-Periphery*:
$$\sum_{R}^{N} \sum_{ij}^{N-1} \left[\left(\tilde{L}(C,R) \tilde{L}^T(C,R) \right)^{-1} \right]_{ij}. \tag{2.19}$$

The first term stands for the importance of the degree of the node (see the first column in Figure 2.3). The more neighbors a node has (the more locally connected is), the more likely it is to be a more *functional* node. We note that this effect is further enhanced by the square of the degree.

Secondly, if we locate all the nodes using the Fruchterman-Reingold force-directed algorithm [67] available at the Networkx python library, which considers an attractive spring force between adjacent nodes and a repulsive electrical force between any pair of nodes, and use the second contribution of *functionability* as an attribute for size and color, we obtain an intuitive and qualitative meaning of it: nodes that have higher values of the *L-Periphery* are located at the periphery of the graphical layout of the network (see the second column in Figure 2.3). Hence, higher values of *functionability* correspond to nodes that are both well connected and also peripheral. Therefore, *functionability* provides us with more information than other classic measures of node importance or centrality (see the third column in Figure 2.3). *Functionability* and its two contributions are shown in Figure 2.3. We highlight that the product of the squared degree and the *L-periphery* is proportional to *functionability*.

2.2.3 New insights from *functionability*: a real example

In Section 2.2.2 we carry out a thoughtful analysis of *functionability*, taking into account its physical meaning and its mathematical expression. In the present section, we show that this new centrality measure provides us with unique information about the network when we compare it with other centrality measures, specially those used in the analysis of brain networks [90].

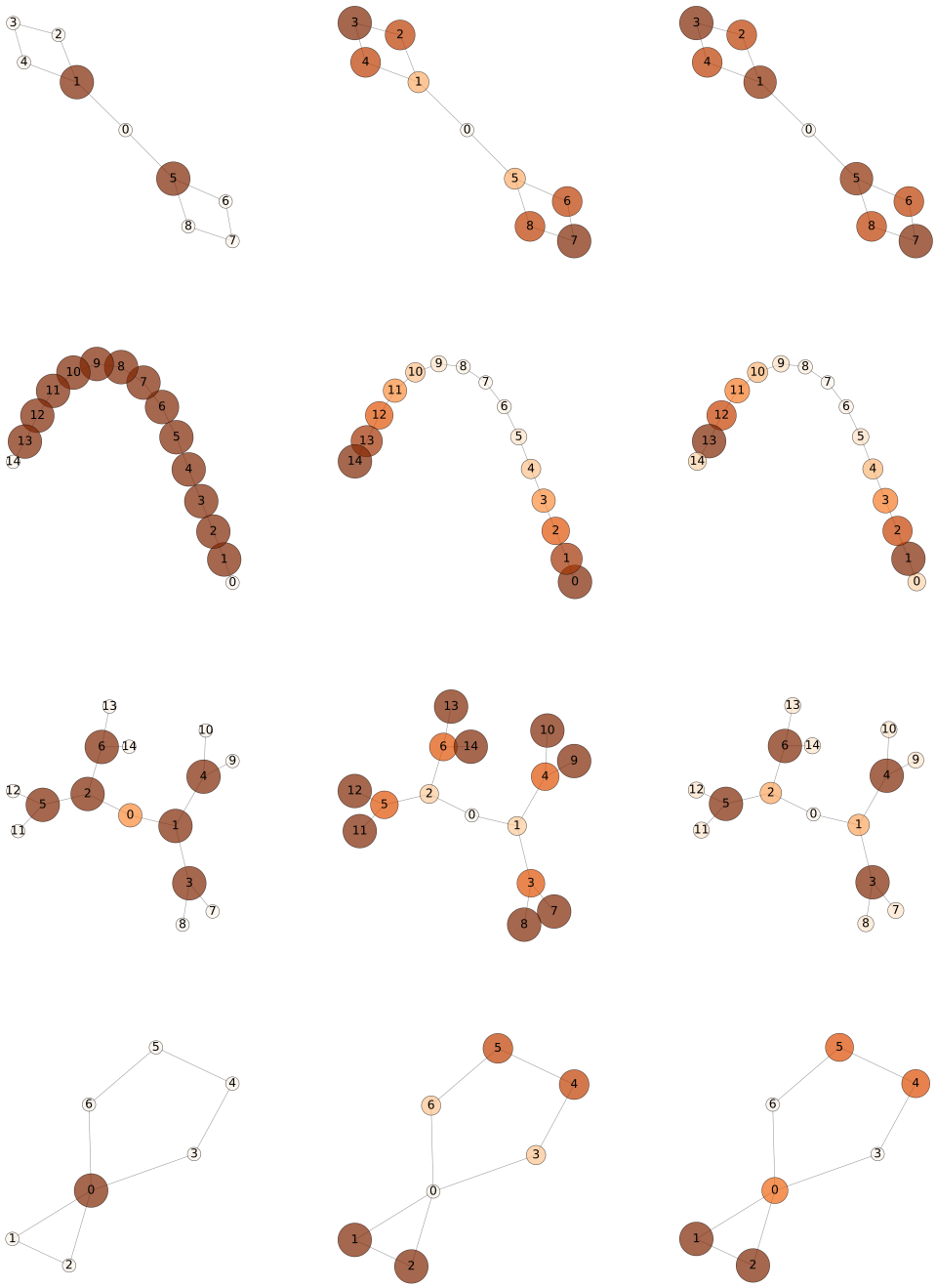


Figure 2.3: **Degree, *L-Periphery* and functionability of the nodes in a network.** Degree, *L-Periphery* and functionability values for the nodes corresponding to four synthetic network topologies. Larger radius and darker colors correspond to higher degree (first column), *L-Periphery* (second column) and functionability (third column) values.The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67].

To this end, we compute the functionability centrality of the nodes of a well-known real network: the frontal cortex network (individual neurons) of the *Caenorhabditis Elegans* worm [86] (see Figure 2.4). Our aim is not to examine the details of the interpretation of the results, but rather to compare functionability with other well-known centrality measures. Nevertheless, we obtain that the ten nodes with the highest functionability score are ASKL, ASKR, ASER, OLQVR, AIAL, AINR, RIAR, OLQVL, AIAR, AIBR, RIH neurons, as standard naming, in descending order. Similarly, in Figure 2.5 we compute the values of several central-

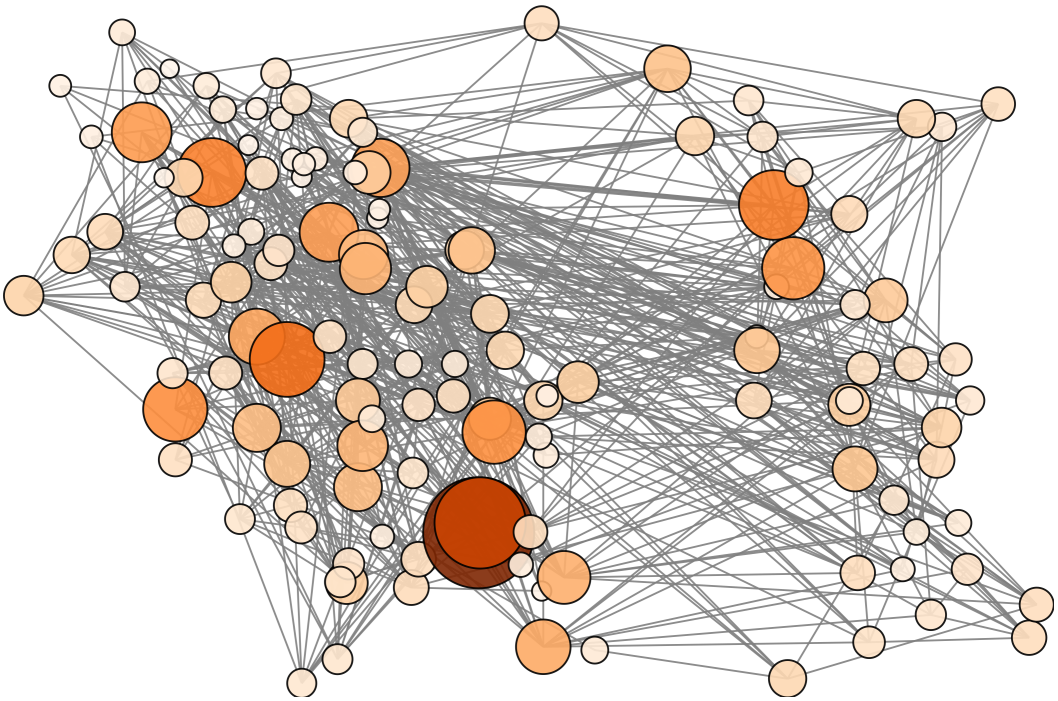


Figure 2.4: **Functionability of the 131 nodes of the *C. Elegans* frontal cortex network.** Functionability values for the nodes –representing individual neurons– of the *C. Elegans* frontal cortex network. Larger radius and darker colors correspond to higher functionability values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67].

ity measures and show the top ten neurons corresponding to the highest values. More classic centrality measures, such as node degree, betweenness, closeness, eigenvector and other spectral based centralities have different outcomes and rankings for nodes from those of *functionability*, even if they are similar. Other centrality measures also have different meanings [33, 25, 56, 34, 72, 147]. Figure 2.4 shows that the nodes that have the highest values of *functionability* correspond to neurons that do not usually appear as neurons with the highest degree, betweenness or eigenvector centralities; and hence, *functionability* gives us additional information concerning such nodes. As can be seen in Figure 2.6, *func-*

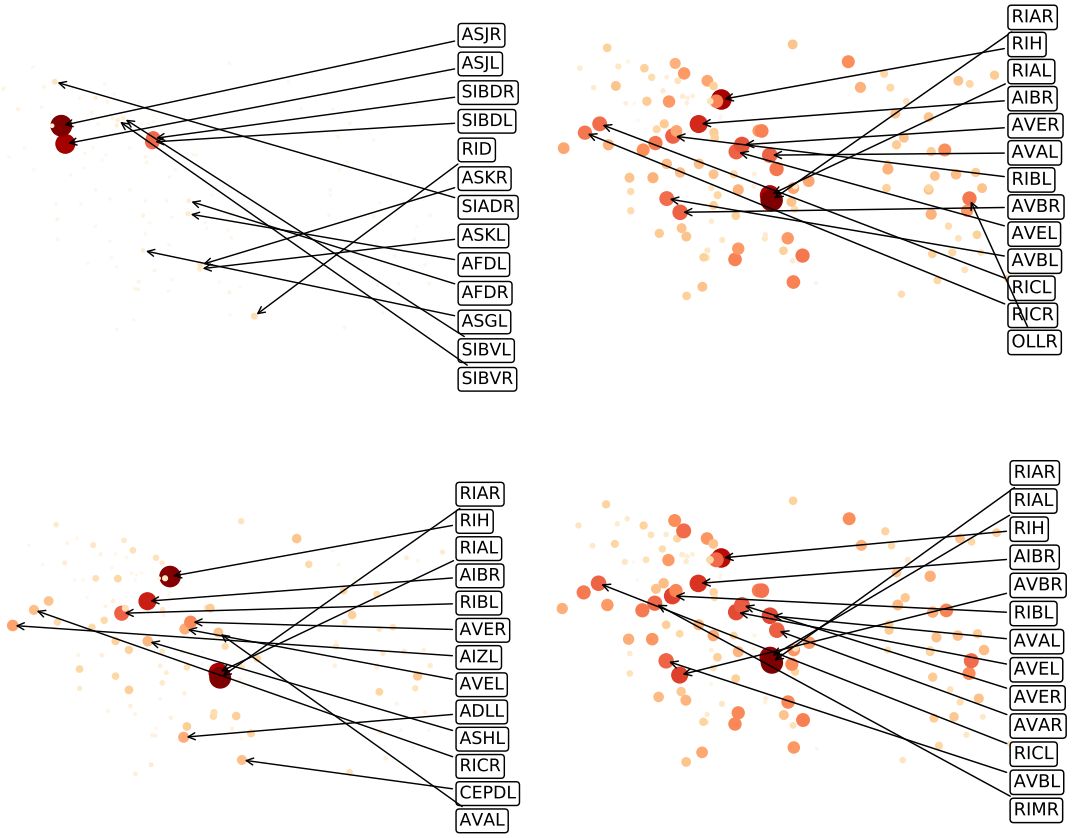


Figure 2.5: **Centrality scores of four different measures of the 131 nodes of the *C. Elegans* frontal cortex network.** *L-Periphery* (upper left panel), degree (upper right panel), betweenness (lower left panel) and eigenvector centrality (lower right panel) values for the nodes –representing individual neurons– of the *C. Elegans* frontal cortex network. Larger radius and darker colors correspond to higher centrality values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67].

tionability has a positive correlation with degree, betweenness and the eigenvector centralities, although all the Pearson coefficient values are below 0.5. We recall that nodes with higher values of *functionability* are more peripheral, as well as having higher degrees. We may also wish to consider whether *function-*

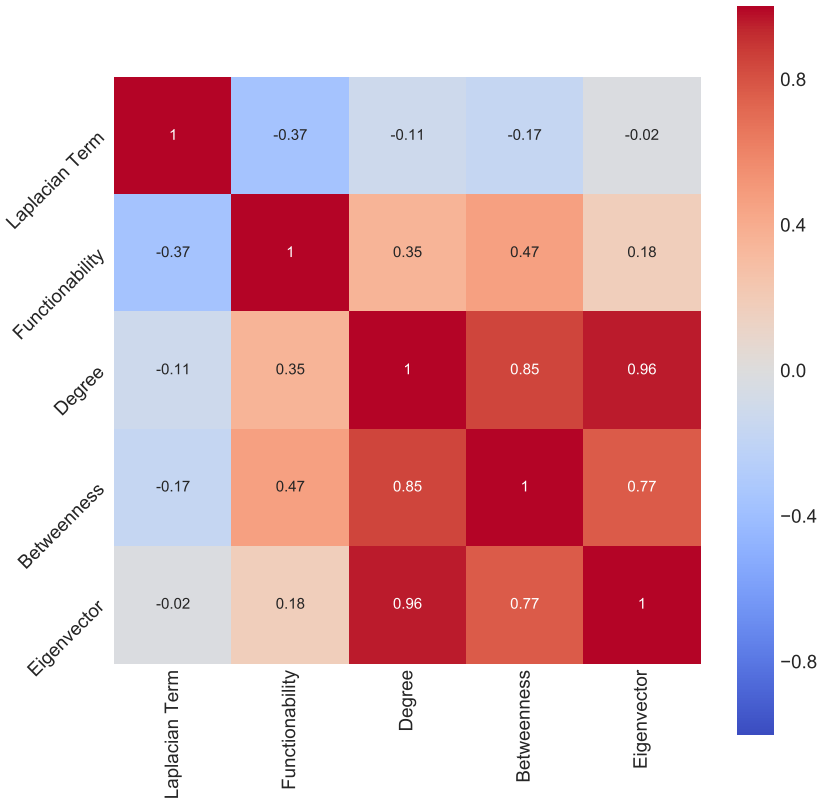


Figure 2.6: **Correlation matrix between five centrality measures corresponding to the *C. Elegans* frontal cortex network.** Correlation matrix, using the standard Pearson correlation coefficient, between five (classic) centrality measures: *functionability*, *L-periphery* (defined in Eq.(2.19)), degree, eigenvector and betweenness, from left to right. Correlations are computed from the centrality values of the nodes in the *C. Elegans* frontal cortex network.

ability would be equivalent to other alternative centrality measures that have been recently developed, such as controllability, core score or collective influence. These measures target specific properties of the network and have not yet been incorporated into standard network libraries and for this reason we do not provide a quantitative comparison with functionability. Nevertheless, we

will consider their definitions and analytical derivations in order to analyse its meaning:

Controllability, $\mathscr{C}$ Within the framework of control theory [96, 100, 134], special attention is paid to possible applications to complex networks, particularly brain networks [101, 77]. The term *control* refers to the ability of nodes to perturb the system in such a way that it reaches a desired state [68]. In order to assess *controllability*, several methods have been developed; all of them assume a linear response dynamical model [162]. We highlight two of them that were specifically designed to evaluate regional controllability, rather than as a global measure of the network.

- Average controllability: As defined in [77], the average controllability identifies the nodes that can steer the network to many easily reachable states. The result that we are most interested in concerns the mathematical expression of the particular case when the set of control nodes reduces to one node at a time, and hence, provides a measure of the average controllability of each node of the network. It can be proved that:

$$\mathscr{C}_a(\mathscr{K}) \equiv \mathrm{tr}(\mathscr{W}_{\mathscr{K}}^{-1}) = \sum_i \left[(\mathbb{I} - A^2)^{-1}\right]_{\mathscr{K} i} \tag{2.20}$$

  Eq.(2.20) resembles the well-known Katz centrality measure, considering only odd length walks. Moreover, if we expand it and we keep the second-order dependency on the adjacency matrix, we recover a measure that is proportional to the degree of the node $\mathscr{K}$.

- Modal controllability: A node that has large modal controllability is one that participates in most of the dynamical modes of the linear system. In other words, it is a node that is able to access states that are difficult to reach [80, 77]. Modal controllability is defined as:

$$\mathscr{C}_m(\mathscr{K}) \equiv \sum_{j=1}^{N} \left(1 - \lambda_j(A)^2\right) v_{ij}^2 \tag{2.21}$$

  where $\lambda(A)_j$ is the eigenvalue of the jth mode and v_{ij} is the contribution of node i to the eigenvector of the jth mode.

Other definitions can be similarly compared to *functionability*, but their meaning moves away from a measure of the effect on the states of a network, for example, boundary controllability. Eq.(2.20) and (2.21) are mathematical expressions which differ from Eq.(2.17) and have a different meaning [168]. However, by looking at them, we can also find similarities regarding dependence on the adjacency matrix. In general, many centrality measures tend to be partially correlated.

Core Score, CS. Despite much attention has been paid to community detection algorithms, another well-known mesoscopic property of a network is its core–periphery structure: which nodes are part of a more densely connected core and which are part of a sparsely connected periphery [22, 35, 78, 48, 103]. The authors in Ref.[22] proposed a continuous measure of a node's closeness to the core, called 'coreness'. The algorithm is based on an optimization procedure which considers cores with different sizes and boundaries, according to a transition function, and assesses to what extend a node matches this. In order to compute coreness, the authors define the *core quality* as:

$$R_\gamma = \sum_{i,j} A_{ij} C_{ij} \tag{2.22}$$

where γ is a vector that parametrizes the core quality. The elements C_{ij} are normally computed as $C_{ij} = C_i C_j$, where C_i are the elements of the local core values. The aim is to find a core vector, $\vec{C}$, that maximizes R_γ and is a normalized shuffle of the vector $\vec{C}^*$, which is determined using a transition function, providing a shuffled list of possible core vector values.

For a given set of parameters that determine the transition function of the core, $\gamma(\alpha, \beta)$, they define the aggregate *core score* of each node i as:

$$CS(i) = Z \sum_{\gamma} C_i(\gamma) \times R_\gamma \tag{2.23}$$

where Z is a normalization factor.

The aim of developing the core score measure differs from that behind *functionability* in many aspects. Nevertheless, the L-periphery centrality, which is one of the contributors to the former, may resemble the inverse core score outcome when the network is characterized by a clear core–periphery structure. Note that the aim of *functionability* is not related to finding communities or the core of a network.

Collective Influence, CI. A subset of measures aim to detect the most influential nodes in an adaptive way. Each method considers a different heuristics to rank all nodes, determines which node is ranked as the greatest spreader and removes it. Scores are recomputed and the procedure is repeated iteratively until no nodes are left in the network. The simplest approach is done by the highly degree adaptive (HDA) method [9].

Collective Influence (CI) tries to fill the gap left by the fact that the preceding set of methods does not optimize an objective global function. In

contrast, CI is defined in such a way that it potentially identifies the minimal set of nodes that, if removed, would cause the network to become disconnected, understood in the framework of network percolation theory. It does this, furthermore, by means of an energy cost function [112, 58, 113].

If $G(q)$ represents the probability of the existence of a giant component [9, 44, 116] in the limit of $N \to \infty$, then the problem reduces to finding the minimum fraction q_c such that:

$$q_c = \min\{q \in [0,1] : G(q) = 0\} \tag{2.24}$$

CI is computed by considering balls of different radius, l, whereby each size captures a different influence scale across the network. In Ref. [112], the authors show that the problem is equivalent to minimizing the cost function:

$$E_l(n) = \sum_{i=0}^{N} z_i \sum_{j \in \partial Ball(i,l)} \left(\prod_{k \in \mathscr{P}_l(i,j)} n_k \right) z_j \tag{2.25}$$

where $z_i \equiv d_i - 1$ and d_i stands for the degree of node i. The vector $\vec{n}$ represents whether a node, in the final configuration, belongs to the set of 'influencers' or not. Then the collective influence strength at level l of node i is:

$$CI_l(i) = z_i \sum_{j \in \partial Ball(i,l)} \left(\prod_{k \in \mathscr{P}_l(i,j)} n_k \right) z_j \tag{2.26}$$

and Eq.(2.25) becomes

$$E_l(n) = \sum_{i=1}^{N} CI_l(i) \tag{2.27}$$

Therefore, in order to minimize Eq.(2.27), we need to remove the node with the greatest CI_l value and iterate until a score is assigned to each node.

Functionability is not obtained from an optimization algorithm or any iterative procedure. However, the physical interpretation of the measure does also have a global scope in the following way: the mathematical definition of $\mathscr{F}_C$ for a node C is computed considering the effect that a local perturbation has on the whole network.

We have compared the proposed *functionability* centrality with two sets of measures of node importance. On the one hand, we compute the correlation matrix between the functionability, the built-in *L-periphery* centrality (defined in Eq.(2.19)) and three classic centrality measures: degree, eigenvector and betweenness centralities considering the frontal cortex network of the *C.Elegans* (see Figure 2.6). Also, the analytical expression of functionability, in Eq.(2.17), informs us that this centrality provides unique insights of the

considered network, since it cannot be replicated by other centralities. Note that the defined L-periphery centrality – which corresponds to the contribution of the reduced Laplacian term – is highly negatively correlated with betweenness centrality. Moreover, both *functionability* and L-periphery are algorithmic-free, parameter-free and deterministic centralities; all of these properties being highly beneficial for network analysis. On the other hand, three more recently developed measures, $\mathscr{C}, CS$ and CI have been explored in order to compare the definition of importance, and its computation as well as mathematical resemblance with *functionability*. We conclude that CS and CI aim to determine a more structural type of centrality, like revealing the participation of a node in the core or the set of nodes which would break the giant component apart. Hence, they may correlate in some ways with L-periphery. Conversely, controllability seeks the nodes which most enable the system to move towards a particular state, either those which are easy to access ($\mathscr{C}_a$) or more mode-like ones ($\mathscr{C}_m$). Average controllability resembles the intuitive motivation of *functionability*, although it is neither mathematically equivalent nor does it have a similar physical interpretation or building blocks (see Eq.(2.17) and (2.20)).

In addition, we should point out that *functionability* centrality does not rely on optimization procedures, nor is it bound to the values of the parameters. Actually, we have proved that Eq.(2.17) is a compact mathematical expression for the measure. The value of the frustration parameter α does not influence the ordering of the nodes (see Section 2.4.2 for more details).

Thereby, *functionability* is a unique measure of the effect of perturbing a node on the whole network by means of shifting the system to an asynchronous state (phase dispersion of the oscillators). The final expression is a deterministic, parameter-free and non-algorithmic measure of centrality, with an underlying physical model to support it and that enables an intuitive interpretation of it.

2.2.4 Weighted *functionability*

Both the dynamic model and the analytic expression of functionability can be easily extended to weighted networks. A weighted network is defined by the elements of the adjacency matrix W as follows:

$$\begin{cases} [W]_{ij} = w_{ij} & \text{if } i \leftrightarrow j \\ [W]_{ij} = 0 & \text{otherwise} \end{cases} \tag{2.28}$$

where $w_{ij} \in \mathbb{R}$.

In a general setting, the intensity of the connections in a network vary. Considering these weights, the GKSM defined in Eq.(2.1) can be rewritten as

$$\dot{\theta}_i = \omega + K \sum_{j=1}^{N} [W]_{ij} \sin(\theta_j - \theta_i - \alpha_i) \qquad \forall i \in [1, ..., N], \tag{2.29}$$

and the analytical expression of functionability for weighted networks is an extension of Eq.(2.17):

$$\mathscr{F}_C = \left(\frac{\alpha d_C}{2N}\right)^2 \sum_{R}^{N} \sum_{ij}^{N-1} \left[\left(\tilde{L}_W(C,R)\tilde{L}_W^T(C,R)\right)^{-1}\right]_{ij}, \tag{2.30}$$

where L_W corresponds to the weighted Laplacian matrix, defined as

$$[L_W]_{ij} \equiv s_i \delta_{ij} - [W]_{ij}. \tag{2.31}$$

The symbol s_i stands for the weighted degree of node i.

We show an example of a weighted version of the network topology in Figure 3.1. We highlight the difference in the color scale and size between Figure 2.2 and Figure 2.7.

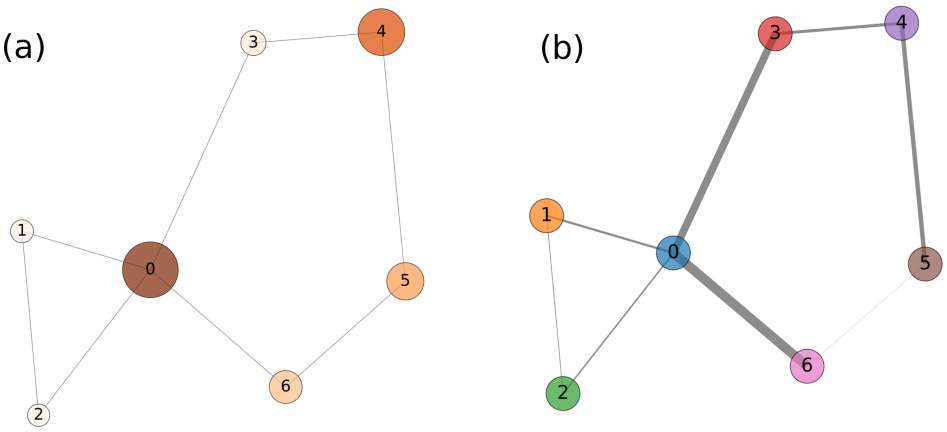


Figure 2.7: **Weighted functionability of the seven nodes of a weighted version of the network topology of Figure 3.1** Weighted functionability values for the nodes of the network presented in Figure 3.1 (right panel). Larger radius and darker colors correspond to higher weighted *functionability* $\mathscr{F}_W$ values (left panel). The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67].

2.3 Discussion

In the present chapter we have defined the *functionability*, a novel centrality measure that addresses the question of which are the nodes that, when individually perturbed, are best able to move the system away from the fully synchronized state. We aim to sort the nodes by their potential effect on the whole network when a change on their individual dynamics spreads over the entire oscillatory system and thereby disrupting the otherwise synchronized state. This issue may be relevant for the identification of critical nodes that are either beneficial – by enabling access to a broader spectrum of states – or harmful – by destroying the overall synchronization. Hence, depending on the system we are considering, the most *functional* nodes have to be considered when looking for a potential enhancement of the diversity of attainable states or the inhibition of risky instabilities in the system.

Our approach to this issue consists in defining a centrality measure called *functionability* $\mathscr{F}$ which characterizes each node and depends only on the connectivity of the network and the position or role of the corresponding node within it.

We consider a system to be represented as a network of connected phase oscillators, each of them corresponding to the nodes of the underlying graph. The dynamics that rules the evolution of individual nodes, and hence the system, is based on the popular Kuramoto phase oscillators [93]. The *functionability* of a node measures the dispersion of the phases at the stationary state produced by the insertion of a phase lag parameter in the dynamics of the considered node, as originally suggested by Kuramoto and Sakaguchi [150]. Despite $\mathscr{F}$ is defined in terms of the phase differences between nodes that are obtained from a dynamic model, in the present chapter we derive the analytical expression of the centrality. The corresponding ranking of node centralities is exclusively determined by the network structure. Therefore and importantly, *functionability* centrality is finally defined as a quantity which does not depend on the values of the model or the parameters of a numerical simulation. Moreover, the analytical expression is a compact and deterministic mathematical function of the network topology and, thus, is not based on optimization procedures. Additionally, the physical meaning of the measure, which is based on a dynamical model, continues to be worth and thus provides the centrality with an easy-to-handle interpretation of the obtained results.

In the present chapter we also compare the output given by functionability with that of other centrality measures, considering different network topologies. This analysis makes us conclude that, despite most centrality measures share certain common patterns, functionability delivers unique information about the network and the importance of its constituent nodes. From the analysis we

can also answer the question of which are the common features that are shared among the nodes with the largest values of $\mathscr{F}$: they are nodes that are locally well connected – have a high degree – but are also far from the central core of the network, that is, they are peripheral nodes. These two properties make functionability to provide non-trivial results regarding the importance of each node. In other words, the nodes that can potentially move the system further from the fully synchronized state are those which are both peripheral and are also locally well connected.

Many real systems are correctly represented as a network of coupled oscillators. For these systems, a perturbation that moves the system towards to or away from the fully synchronized state may represent a beneficial or harmful change to it. *Functionability* enables us to detect the nodes that are most central or relevant for moving the overall system away from synchronization. Epileptic attacks or power grid collapses may be derived from single nodes that, even if not located in the main core, change their intrinsic properties and spread asynchrony rapidly to the network, leading to potentially fatal states. It may be helpful to target such nodes in order to control both synchrony and asynchrony.

2.4 Additional Information

2.4.1 Is phase distance a proper metric?

In Eq.(2.2) we define a measure of the distance between two nodes or oscillators after a perturbation is made on the system. Our system consists of a set of phase oscillators, that is, oscillators whose main and only variable is its phase, and not amplitude. Hence, the distance between two nodes is a rather simpler function of an angular argument, symmetric with respect to π and bounded between 0 and 1. In this section we will comment on the mathematical implications of this definition. As a first consideration, the distance we are using it is not a proper metric, because it does not meet the triangle inequality, as we will see. Nevertheless, we will prove that it can be easily related to the Euclidean distance, which it is so. However, as we are not concatenating or adding different distances, this drawback will not be a problem. There are many optimization algorithms that use non-metric distances without modifying the expected results.

In order to make an intuitive description of the distance ε_{ij}, we will consider each oscillator to lay on a two dimensional plane. Each oscillator is characterized by a phase, which evolves in time, and a radius, which is equally set to one.

The Euclidean distance of two vectors, $\vec{a}$ and $\vec{b}$, in an n-dimensional space is defined as

$$d_E(\vec{a},\vec{b}) = \sqrt{||\vec{a}-\vec{b}||^2} = \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_N-b_N)^2}. \qquad (2.32)$$

The squared Euclidean distance can be expanded as follows

$$d_E^2(\vec{a},\vec{b}) = ||\vec{a}-\vec{b}||^2 = ||\vec{a}||^2 + ||\vec{b}||^2 - 2\vec{a}\cdot\vec{b} = 2\left(1-\cos(\vec{a},\vec{b})\right) \tag{2.33}$$

where we have considered that vectors $\vec{a}$ and $\vec{b}$ are normalized, that is, $||\vec{a}|| = 1$ and $||\vec{b}|| = 1$.

If we go back to the definition of the used distance between nodes, ε_{ij} in Eq.(2.2), we notice that the functional form is the same as the cosine distance between two vectors, albeit in one dimension. Looking at Eq.(2.33) we can rewrite cosine distance as

$$d_C(\vec{a},\vec{b}) \equiv \frac{1-\cos(\vec{a},\vec{b})}{2} = \frac{d_E^2(\vec{a},\vec{b})}{4}. \tag{2.34}$$

From the former equality, Eq.(2.34), we can state that, although, in general, cosine distance is not a proper metric, as it does not meet the triangle inequality, it can be easily interpret by means of the euclidean distance between both vectors.

The four properties a metric defined by a distance function $d(\vec{a},\vec{b})$ should satisfy are:

1. $d(\vec{a},\vec{b}) \geq 0$
2. $d(\vec{a},\vec{b}) = 0 \leftrightarrow \vec{a} = \vec{b}$
3. $d(\vec{a},\vec{b}) = d(\vec{b},\vec{a})$
4. $d(\vec{a},\vec{c}) \leq d(\vec{a},\vec{b}) + d(\vec{b},\vec{c})$

Considering cosine distance, $d_C(\vec{a},\vec{b})$, condition 4 can be written as

$$\begin{aligned} & d_C(\vec{a},\vec{c}) \leq d_C(\vec{a},\vec{b}) + d_C(\vec{b},\vec{c}) \Rightarrow \\ \Rightarrow \frac{1-\cos(\vec{a},\vec{c})}{2} \leq \frac{1-\cos(\vec{a},\vec{b})}{2} & + \frac{1-\cos(\vec{b},\vec{c})}{2} \Rightarrow \\ & \Rightarrow \cos(\vec{a},\vec{b}) + \cos(\vec{b},\vec{c}) - \cos(\vec{a},\vec{c}) \leq 1 \end{aligned} \tag{2.35}$$

Eq.(2.35) is, in general, not satisfied. Let us provide a counterexample by considering 3 2-dimensional normalized to unity vectors, $\vec{a} = (1,0)$, $\vec{b} = (\sqrt{2}/2, \sqrt{2}/2)$ and $\vec{c} = (0,1)$.

$$0.76 + 0.76 - 0 \approx 1.52 \not\leq 1 \tag{2.36}$$

Nevertheless, we are interested in small phase differences, and hence, small angular arguments. In this situation, if $\angle(\vec{a},\vec{b}) = \theta_1$, $\angle(\vec{b},\vec{c}) = \theta_2$ and $\angle(\vec{a},\vec{c}) = \theta_3 = \theta_1 + \theta_2$, Eq.(2.35) can be approximated to

$$1 - \frac{\theta_1^2}{2} + 1 - \frac{\theta_2^2}{2} - \left(1 - \frac{\theta_3^2}{2}\right) \leq 1 \Rightarrow -|\theta_1||\theta_2| \leq 0 \tag{2.37}$$

Condition in Eq.(2.37) is always true and, therefore, Eq.(2.2) is a proper distance for our purpose.

2.4.2 The linear model: assumptions and validity

The final definition and usage of *functionability* centrality, in Eq.(2.17), or Eq.(2.18) if we are interested in the normalized definition, is based on a non-linear oscillatory dynamics, as explained in Section 2.2. However, several assumptions have been made in order to obtain a more compact and useful expression of it. The more restrictive one is the linearization of the model. However, this assumption relies on the fact that we are concerned with the state where nodes are frequency synchronized and therefore, dispersion comes from the disruption of the phase synchronized state. For a set of coupled oscillators in a network which follow the conditions described in Section 2.1, the linearization assumption is not met when the frustration parameter α becomes too large (see lower panel in Figure 1.7).

In Figure (2.8) the value of functionability is computed in two different ways and for different values of α. On the one hand, the phase differences between oscillators are obtained from the numerical simulation of the non-linear model and $\mathscr{F}$ is determined applying its definition in Eq.(2.4). On the other hand, the analytical expression of functionability, derived in Eq.(2.17), is directly computed considering the same parameters.

Regardless of the type of network topology, there is a threshold value at which the two methods diverge. This value changes from one network to another, as well as from size to size.

Importantly, when the maximum simulation time increases, the values of the functionability delivered by the two methods become closer because the dynamics eventually reaches the stationary state (as the network size increases, the time needed to reach the stationary state becomes larger too). Nevertheless, our interest is to use the analytical expression of functionability, found in Eq.(2.17), rather than the values obtained from the results of the numerical simulation of a dynamical system, which results may be extremely sensible to the parameters of the simulation and the model (additionally, and depending on the values of the frustration parameter, the system may not achieve the stationary state, but continue trapped in a chaotic state). Nevertheless, the GKSM being the starting point of functionability enables a physical interpretation of it.

In addition, Eq.(2.17) is proportional to α, and therefore the ranking of nodes does not depend on this parameter, but only on the connectivity structure. From Figure 2.8 we can conclude that, if we set the value of the frustration parameter small enough, the two approaches – numerical simulation of the dynamical model and the linearized expression of functionability – are equivalent. However, as we move to larger values of α, the system following the GKSM dynamics results in a chaotic state and the corresponding values of functionability do not reflect the meaning of the centrality, whilst the analytical expression of func-

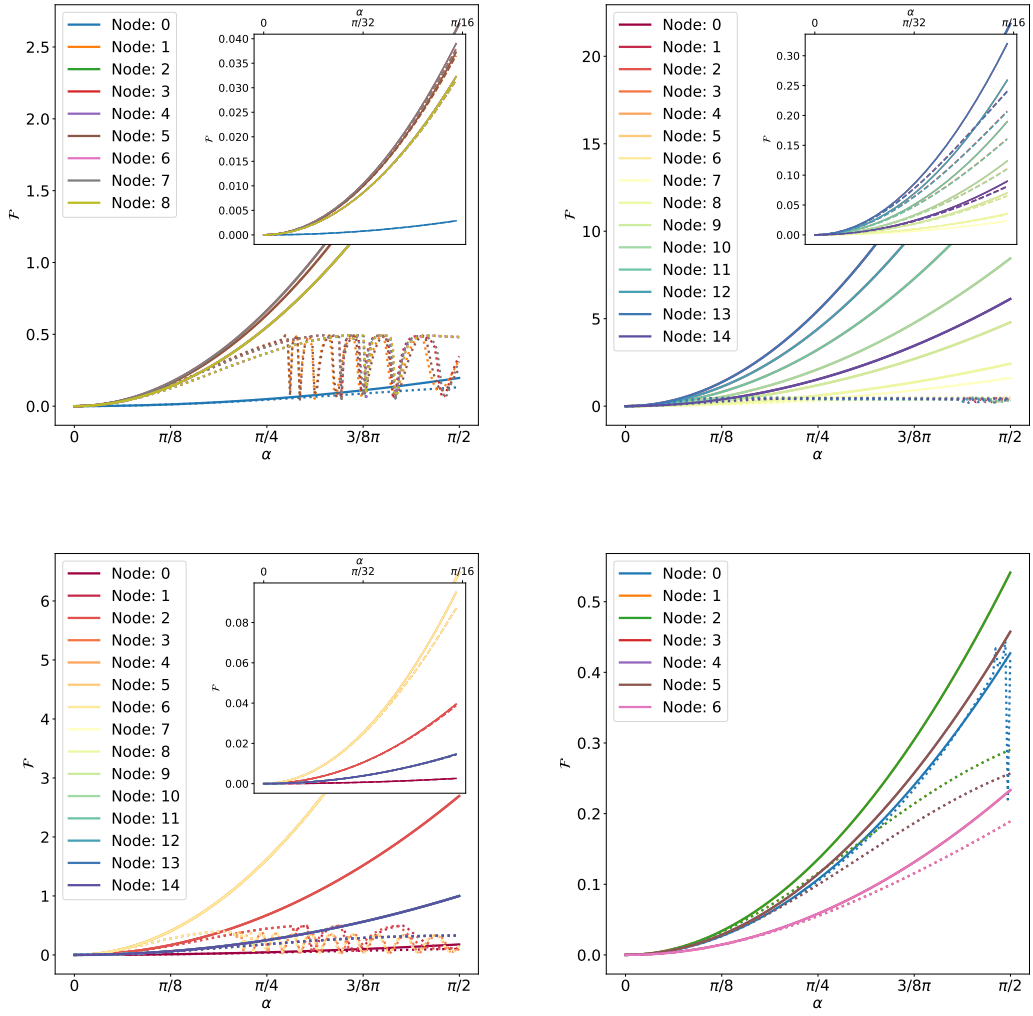


Figure 2.8: **Comparison of the functionability values obtained from the numerical simulation of the dynamical model and using the corresponding analytical expression for the four synthetic networks defined in Figure 2.3.** Comparison of the results obtained from the computation of functionability centrality $\mathscr{F}$ both considering the numerical simulation of the dynamical model defined in Eq.(2.1) and Eq.(2.4) (dotted lines), and the analytical expression derived in Eq.(2.17) (continuous lines). Functionability centrality is obtained considering different values of the frustration parameter α, in the range $[0, \pi/2]$, and all the nodes in the four synthetic networks defined in Figure 2.3 (maintaining the same ordering, from top to bottom, and from left to right).

tionability is always robust in the ranking of nodes. Moreover, the interpretation of the measure is still valid because it keeps rank ordering, that is, the ordering of the nodes according to its functionability score obtained from the analytical expression of the linear model is the same as that given by the numerical simulation of the dynamics described by the non linear one.

CHAPTER 3

Optimal Cost Tuning of Frustration

Emergence is one of the key concepts in the analysis of complex systems [123]. Collective properties emerge as a consequence of irregular interactions among its elemental constituents [21]. One of the most paradigmatic examples of emergence is synchronization [139, 131], because the interplay between populations of oscillatory units gives rise to a variety of global states, ranging from perfect synchronization or phase locked stationary configurations to chimera states [15, 53, 30]. Among the different models that have been used to understand such collective behavior, a lot of effort has been devoted to the Kuramoto model (KM), in which phase oscillators interact continuously with other units through a sine function of the phase difference [92, 150, 4].

In the past few years there has been a growing interest in the concept of controllability, which quantifies the feasibility to achieve a desired final state of a given dynamical system [99]. As stated above, the KM can give rise to a wide variety of stationary (phase or frequency synchronized) or not stationary states, being chimeras an unexpected mixture of both types of behaviors [66]. In this context, controllability can be understood as a tuning of the internal parameters of the oscillators to reach specific phase configurations. The most simple settings stand for a collection of identical oscillators interacting through a sinus function of the phase differences. In this case it is quite intuitive to see that the final state is a perfectly synchronized one in which all oscillators have exactly the same phase and frequency (the same frequency than the intrinsic one). It is the existence of a distribution of frequencies that gives rise to a transition, in terms of the strength of the coupling, from an incoherent state to a coherent one [93]. This transition is also present when we include a lag term – a constant phase added to the argument of the sinus function – for a wide range of such parameter [150, 124]. However, the introduction of this lag term for identical oscillators changes completely the structure of the, in principle, synchronized state. In Reference [124] it was shown that, for small and common values of the lag parameters, the synchronized state breaks into partially synchronized groups of oscillators, being symmetry the reason for the phase synchronization of the oscillators. When increasing this common lag parameter the system enters into a incoherent chaotic state. Actually, there has been an increasing interest in the last months on the role that symmetries plays in the synchronization of oscilla-

tory units and how the lack of homogeneity in some of the parameters can be compensated by other choices [125, 110, 184].

In the previous chapter we introduced the concept of “functionability" as a measure of the ability of a given node to change the state of the system by just tuning one internal variable, the node lag in the argument of the sinus function of the interaction [144]. Being an intrinsic property of the node, its change produces a global change in the phases of the system of oscillators that can be measured. Functionability stands for the reaction of the whole phase distribution to a small change in a node. The analytical expression of functionability reports its quadratic dependence on the node degree and the node lag value, but also a structural term, such that the most peripheral nodes in the network have also larger contributions to functionability centrality measure. The nodes with higher functionability values may represent positive actors for the network, because they enable more variability in the states of the system, but also potentially dangerous ones, as tiny perturbations can produce cascade like effects that completely changes the network dynamics.

As stated, the addition of a phase lag parameter enables a richer configuration state. However, it is clear that a tuning of a single parameter will not be enough to generate the wide variety of stationary states that a population of Kuramoto oscillators can achieve. Notwithstanding, the question that arises is whether a fine tuning of a set of individual parameters can make it possible. In this chapter, this is our proposal, we construct a general formalism that allows, within a linear approximation, to compute the set of lag parameters that may lead to any phase configuration for a fixed set of intrinsic frequencies. The problem can also be posed the other way around. Namely, given a set of frequencies, we may derive the configuration of phases that is produced by a set of lag parameters.

There are numerous examples of real-world systems that can be described as dynamical systems characterized by individual phases and which functioning are object of investigation. Some examples are the brain functional networks arising from temporal correlation patterns, ac power in power grids[94], heartbeats[6], multiprocessors and multicore processors, or traffic signalling. Not only the synchronization between their constituents may be intended or prevented, but also other particular configurations may be of relevant interest. For this reason, we propose a mechanism for tuning the intrinsic parameters of the system to achieve any desired phase configuration.

A previous work proposes a methodology to enhance frequency synchronization for the nonlinear Kuramoto-Sakaguchi model (extension of the Kuramoto model with a node phase lag parameter) [38]. Another work suggests that an unstable synchronized state becomes stable when, and only when, the oscillator parameters are tuned to nonidentical values [125]. We highlight the

work done in Reference [158], where the particular configuration of perfect synchronization is studied and the *synchrony alignment function* is defined in order to minimize the order parameter of the system considering different topologies and frequency scenarios. We address the most general question, following a similar path to that pursued by them, forcing the system to achieve any particular configuration for the linear case of the Kuramoto-Sakaguchi model by means of a fine tuning of the phase lag or frustration parameter set. Despite considering a linear approximation of the model, we show that the obtained tuned parameters for the case of full synchronization enhance frequency synchronization in the nonlinear model as well.

All the results presented in this chapter can be found in Ref.[145].

3.1 Analytic expression of the frustration parameters tuning

Similarly to the previous chapters, we consider the most general expression of the Kuramoto-Sakaguchi model, without any assumption about the distribution of natural frequencies and allowing the edges of the network to be weighted, a more realistic scenario for real-world networks, in which oscillators represent real systems with individual properties that are determined by the nature of each oscillator. The GSKM is thus defined by the dynamics:

$$\dot{\theta}_i = \omega_i + K\sum_{j=1}^{N} W_{ij}\sin(\theta_j - \theta_i - \alpha_i) \qquad \forall i \in [1,...,N], \tag{3.1}$$

where α_i is an intrinsic parameter of each oscillator and ω_i its natural frequency.

We next derive the general expression of the linearization of Eq.(3.1) following an equivalent procedure as Sections 1.3.2.2 and 2.2.1. For small values of the frustration parameters and phases close to each other, which is the case in frequency synchronization, we can linearize Eq.(3.1) as follows:

$$\dot{\theta}_i = \omega_i + K\sum_j W_{ij}(\theta_j - \theta_i - \alpha_i) = \omega_i - K\sum_j L_{ij}\theta_j - K\alpha_i s_i, \tag{3.2}$$

where W_{ij} is the value of the weight of the edge between node i and node j, $s_i \equiv \sum_j W_{ij}$ is the weighted degree of the ith node and L is the weighted Laplacian matrix defined as $L_{ij} \equiv \delta_{ij}s_i - W_{ij}$. In the stable regime, a synchronized frequency is achieved and, for all oscillators $\dot{\theta}_i = \Omega$. We can derive the value of the common frequency oscillation, Ω, summing Eq.(3.2) over i:

$$\sum_i \Omega = \sum_i \omega_i - K\sum_i\sum_j L_{ij}\theta_j^* - K\sum_i \alpha_i s_i \tag{3.3}$$

Taking into account the steady state $\dot{\theta}_i = \Omega \; \forall i$ and arranging summations:

$$N\Omega = \sum_i \omega_i - K \sum_j \theta_j^* \sum_i L_{ij} - K \sum_i \alpha_i s_i. \tag{3.4}$$

and finally,

$$\Omega = \langle \omega \rangle - K \langle \alpha s \rangle . \tag{3.5}$$

where we have used the Laplacian matrix property: $\sum_i L_{ij} = 0$ and defined the averages $\sum_i \alpha_i s_i / N = \langle \alpha s \rangle$ and $\sum_i \omega_i / N = \langle \omega \rangle$. Now we can plug expression Eq.(3.5) to Eq.(3.2) to get the stable phases of oscillators, θ_i^*:

$$\sum_j L_{ij} \theta_j^* = \frac{\omega_i}{K} - \frac{\langle \omega \rangle}{K} + \langle \alpha s \rangle - \alpha_i s_i \qquad \forall i \tag{3.6}$$

The solution of Eq.(3.6) regarding phases is undetermined due to the singular nature of the Laplacian matrix. Hence, Eq.(3.6) is, in general, an undetermined system of linear equations, that is, there is one free phase, which we should use as a reference value for the solution. Nonetheless, we do not work directly with the functional form of phases because they are time dependent $\{\theta_i^*\} = f_i(t)$, but with the phase differences with respect to a reference node, once the stationary state is achieved,

$$\phi_i \equiv \theta_i - \theta_R \tag{3.7}$$

In this way, we work with time independent values. In this situation, $\phi_R = 0$, by definition, as $\phi_R \equiv \theta_R - \theta_R = 0$.

On the other hand, the contribution $\langle \alpha s \rangle - \alpha_i s_i$ of the right-hand side of Eq.(3.6) can be written in matrix form as:

$$-\begin{pmatrix} \frac{N-1}{N} & -\frac{1}{N} & -\frac{1}{N} & \dots \\ -\frac{1}{N} & \frac{N-1}{N} & -\frac{1}{N} & \dots \\ \dots & \dots & \dots & \dots \\ -\frac{1}{N} & -\frac{1}{N}1 & \dots & \frac{N-1}{N} \end{pmatrix} \cdot \begin{pmatrix} s_0 & 0 & \dots & 0 \\ 0 & s_1 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & s_{N-1} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_{N-1} \end{pmatrix} = (-M \cdot D_s)\vec{\alpha} \tag{3.8}$$

where we have defined

$$M \equiv \begin{pmatrix} \frac{N-1}{N} & -\frac{1}{N} & -\frac{1}{N} & \dots \\ -\frac{1}{N} & \frac{N-1}{N} & -\frac{1}{N} & \dots \\ \dots & \dots & \dots & \dots \\ -\frac{1}{N} & -\frac{1}{N}1 & \dots & \frac{N-1}{N} \end{pmatrix} \tag{3.9}$$

and

$$D_s \equiv \begin{pmatrix} s_0 & 0 & \dots & 0 \\ 0 & s_1 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & s_{N-1} \end{pmatrix}. \tag{3.10}$$

We write Eq.(3.6) in matrix form as

$$L\vec{\theta}^* = \frac{1}{K}\vec{\Delta\omega} - M \cdot D_s\vec{\alpha} \tag{3.11}$$

where $\Delta\omega_i \equiv \omega_i - \langle\omega\rangle$. Finally, we obtain the set of unknowns $\{\alpha_i\}$:

$$M \cdot D_s\vec{\alpha} = \frac{1}{K}\vec{\Delta\omega} - L\vec{\theta}^* \tag{3.12}$$

Equation (3.12), however, does not have a solution, because of the singular nature of $M \cdot D_s$ matrix. M matrix is singular too, and hence, its inverse matrix does not exist. Mathematically, $det(M \cdot D_s) = det(M) \cdot det(D_s) = 0$

Similarly as we did for phases in Section 2.2.1, we solve the singularity problem by setting a reference node, which we call *control* node, regarding frustration parameters, i.e., we would not obtain the value for each of the parameters, but a relation between them:

$$\kappa_i \equiv \alpha_i - \alpha_C \tag{3.13}$$

where α_C is the value of the control node. In this situation, $\kappa_C = 0$, by definition, as $\kappa_C \equiv \alpha_C - \alpha_C = 0$.

To easily write the matrix expressions, we define the selection matrix $J_{(n,m)}$, which is, in general, an $(N-1)\times(N-1)$ identity matrix after the removal of the mth row and the nth column.

$L\vec{\theta}^*$ turns to $\tilde{L}(k,R)\vec{\phi}^*$, where we have removed the kth row and the Rth column. The result does not depend on which row we remove, hence we can choose any k. Using the selection matrix, $\tilde{L}(k,R) = J_{(,k)} \cdot L \cdot J_{(R,)} \equiv \tilde{L}$.

Similarly, $\vec{\tilde{\phi}}(k) = J_{(,k)} \cdot \vec{\phi} \equiv \vec{\tilde{\phi}}$, where we have removed the kth row.

In an equivalent way as the definition of the reduced Laplacian:

$$\tilde{MD_s}(k,C) = J_{(,k)} \cdot MD_s \cdot J_{(C,)} \equiv \tilde{MD_s}$$

where $\tilde{MD_s}$ is MD_s without the kth row and the Cth column.

Similarly, $\vec{\tilde{\kappa}}(k) = J_{(,k)} \cdot \vec{\kappa} \equiv \vec{\tilde{\kappa}}$ and $\vec{\tilde{\Delta\omega}}(k) = J_{(,k)} \cdot \vec{\Delta\omega} \equiv \vec{\tilde{\Delta\omega}}$, where we have removed the kth row.

Considering all the previous definitions and remarks, Eq.(3.11) can be rewritten as:

$$\tilde{MD_s}\vec{\tilde{\kappa}} = \frac{1}{K}\vec{\tilde{\Delta\omega}} - \tilde{L}\vec{\tilde{\phi}}^* - \alpha_C \cdot J_{(,k)}\vec{\sum_j}[MD_s]_{ij} \tag{3.14}$$

and finally,

$$\vec{\tilde{\kappa}} = \left(\tilde{MD_s}\right)^{-1}\left(\frac{1}{K}\vec{\tilde{\Delta\omega}} - \tilde{L}\vec{\tilde{\phi}}^* - \alpha_C \cdot \tilde{M}\vec{s}\right) \tag{3.15}$$

where we have used $J_{(,k)}\vec{\sum}_j[MD_s]_{ij} = \tilde{M}\vec{s}$. Notice that MD_s matrix is singular, but the row sum is not zero, although it is so for the column sum. Hence, we need to set $\alpha_C = 0$ if we want to avoid extra constant arrays in the final expression. In this particular case:

$$\vec{\tilde{\kappa}} = \left(\tilde{MD_s}\right)^{-1}\left(\frac{1}{K}\vec{\Delta\omega} - \tilde{L}\vec{\phi}^*\right) \tag{3.16}$$

and keep in mind that $\kappa_C = 0$.

The obtained values of $\vec{\alpha}$ depend on both the chosen control node, C, and the value we set for its frustration parameter, α_C. Notice, therefore, that there is a continuous spectrum of values for the frustration parameter in order to achieve a particular phase configuration.

Moreover and more importantly, due to the non-row-sum equal to zero of MD_s matrix, the differences between the obtained values are dependent of the control node choice. Mathematically, $\alpha_i - \alpha_j(C = l) \neq \alpha_i - \alpha_j(C = k)$ if $l \neq k$. This property will lead us to the definition of a cost for the system to move to the final configuration, which will depend on both the control node and the value of its frustration parameter.

We provide an example of a synthetic network for the case of a homogeneous natural frequencies distribution, i.e., $\omega_i = \omega\ \forall i$. In this case, Eq.(3.15) turns to:

$$\vec{\tilde{\kappa}} = \left(\tilde{MD_s}\right)^{-1}\left(-\tilde{L}\vec{\tilde{\phi}}^* - \alpha_C \cdot \tilde{M}\vec{s}\right)$$

which in the case of the network depicted in Figure 3.1, leads to the solution

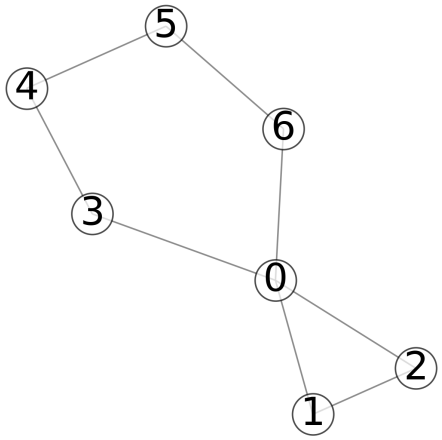


Figure 3.1: **Example of a simple graph.** Simple graph made of 7 nodes and 8 edges.

$$\begin{pmatrix} \kappa_0 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} = \begin{pmatrix} \frac{-2\alpha_1+3\phi_1+\phi_3+3\phi_6}{4} \\ \frac{3(\phi_1-\phi_2)}{2} \\ \frac{2\phi_1-\phi_2-2\phi_3+\phi_4}{2} \\ \frac{2\phi_1-\phi_2+\phi_3-2\phi_4+\phi_5}{2} \\ \frac{2\phi_1-\phi_2+\phi_4-2\phi_5-\phi_6}{2} \\ \frac{2\phi_1\phi_2+\phi_5-2\phi_6}{2} \end{pmatrix} \tag{3.17}$$

where we have chosen $\kappa_1 = 0$ and $\phi_0 = 0$. Hence, the results are written as a function of the value α_1 and ϕ_i $i \neq 0$. Therefore, we can achieve any phase configuration, given by the set $\{\phi_i\}$ by tuning the frustration parameters set $\{\alpha\}$, where $\alpha_i = \kappa_i + \alpha_C$.

To illustrate how we obtain the final values, let us consider the following phase configuration:

$$\vec{\tilde{\phi}}_{(R=0)} = (0.1, 0.2, 0.25, -0.2, -0.1, 0.0) \tag{3.18}$$

In the general case where $\alpha_C = \alpha_1 \neq 0$:

$$\vec{\tilde{\kappa}}_{(C=1)} = (0.1375 - \frac{\alpha_1}{2}, -0.15, -0.35, 0.275, 0.0, -0.05).$$

If we choose $\alpha_C = 0$, then $\alpha_i = \kappa_i$, we can include the value of the control node $C = 1$:

$$\vec{\tilde{\alpha}} = (0.1375, 0.0, -0.15, -0.35, 0.275, 0.0, -0.05).$$

Alternatively, we can choose whatever value we need regarding the control node. For instance, if $\alpha_C = \alpha_1 = 0.1$:

$$\vec{\tilde{\alpha}} = (0.1875, 0.1, -0.05, -0.25, 0.375, 0.1, 0.05)$$

and the phases configuration is the same. Importantly, we recover the same phase differences using the nonlinear model with the tuned α's, up to an error. For this last example and using the frustration parameters obtained by setting $\alpha_1 = 0.1$, the nonlinear model leads to final phases vector

$$\vec{\tilde{\phi}}_{(R=0)} = (0.09969, 0.19944, 0.25097, -0.19798, -0.09897, 0.00012) \tag{3.19}$$

which represents $\sim 0.3\%$ of relative error with respect to the initial Eq.(3.18). See the full derivation of the analytical solution in Section 3.4.1.

3.2 Optimal Cost tuning of frustration

As pointed out in Section 3.1, there is a continuous spectrum of values for the choice of the frustration parameters that enables the system access a particular phase configuration. The following question arises naturally: Among all the

possible solutions, which is the one that makes the system achieve a particular phase configuration with the minimum required cost?

This question is of particular relevance when we consider the plausible real nature of the system. If a real system needs to access a particular phase configuration, which may be associated with a precise function, it will tend to minimize the effort or cost to do so.

In order to quantify the required cost, we define it as follows:

$$e_T(C) \equiv \sum_i |\alpha_i(C)| \tag{3.20}$$

Henceforth, the cost associated to each node is given by the absolute value of the required frustration parameter. The absolute value operator allows for a sign-free contribution of each node, a very convenient choice in the case that the system is not beforehand specified, and a general definition is proposed instead. Furthermore, unlike other nonlinear cost functions such as the square sum of the parameters, no extra weight is given to larger values, besides the corresponding to a linear function.

As previously remarked, $e_T(C)$ will depend both on the chosen control node, C, as well as the particular choice of its frustration parameter, α_C.

The optimal configuration is given by the solution of the minimization problem

$$\min_{C,x} e_T(C,x) = \min_{C,x} \sum_i^N |\alpha_i(C,x)| \tag{3.21}$$

where the x variable is not yet specified. Depending on the problem we are interested in we would set it either to ω_i, s_i or any other combination of the parameters of the model. The minimal value of the cost will depend on the proper choice of the control node, C, in addition of the particular value of its frustration parameter, α_C, as the free parameter left to be set. In Sections 3.2.1 and 3.2.2 we provide a thorough analysis of it.

The cost required to achieve a particular phase configuration depends on that configuration, the control node and the chosen value of α_C. In Figure 3.2 we present an example, following with the network presented in Section 3.1 and choosing different values of α_C, we compute numerically the values of the required cost using Eq.(3.20) to achieve the phase configuration given in Eq.(3.18). Notice that the global minimum depends on the control node and its frustration parameter.

In Section 3.1 we have derived the general analytical solution of the frustration parameters as a function of a particular choice for the phase configuration. In this section we have defined a cost function in order to assess the optimal choice of such configuration.

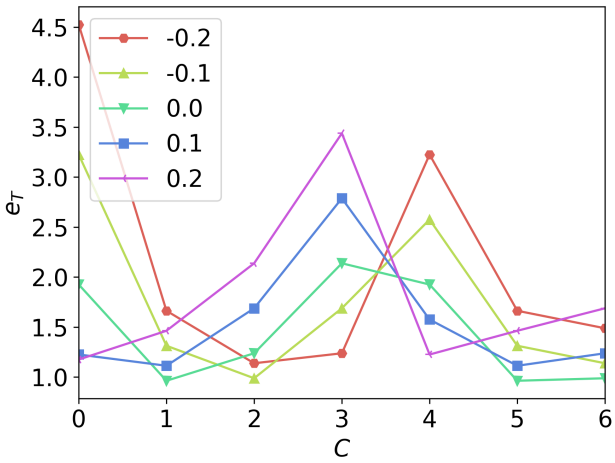


Figure 3.2: **Implied cost to achieve the phase configuration in Eq.(3.18) as a function of the chosen control node, C, for the network in Figure 3.1 and considering five different values of α_C** Implied cost to achieve the phase configuration in Eq.(3.18) as a function of the chosen control node, C, for the network in Figure 3.1 and considering five different values of α_C. Notice that the minimum cost is given, in this case, with the values $\alpha_C = 0$ and $C = 1$ or $C = 5$.

Depending on the phase configuration one is interested in achieving, results will vary and the analytical expressions will have different features.

In the following sections we will focus on two particular configurations, due to its intrinsic importance, in order to obtain and discuss the analytical solution of Eq.(3.21): The configuration given by the symmetries of the network [124] and the fully synchronized state.

3.2.1 Symmetric phase configuration

As explained in Section 1.3.2.2, a particular example of the Kuramoto-Sakaguchi model is the symmetric case, obtained by a homogeneous distribution of frustration parameters, i.e, $\alpha_i = \alpha_h \; \forall i$. For our purposes, we consider $\alpha_h > 0$. In this situation, the trivial solution of the frustration parameters, $\alpha_i = \alpha_h$, is another one of the values out of the continuous spectrum. That is, we can recover the landscape given by the symmetric configuration in many different ways. We are however, interested in computing the analytical expression of the cost function in order to select the one corresponding to the minimum cost.

3.2.1.1 Optimal cost tuning when $\alpha_C = 0$

Let us firstly consider the case where $\alpha_C = 0$ and homogeneous natural frequencies $\omega_i = \omega_h \ \forall i$. In the particular case of the symmetric configuration, that is, the phase configuration given by $\alpha_i = \alpha_h \ \forall i$ the solution of the frustration parameters is given by:

$$\vec{\tilde{\kappa}} = \left(\tilde{MD}_s\right)^{-1}\left(\frac{1}{K}\vec{\Delta\omega} - \tilde{L}\vec{\tilde{\phi}}^*\right) = -\left(\tilde{MD}_s\right)^{-1}\tilde{L}\vec{\tilde{\phi}}^* \tag{3.22}$$

But $\vec{\tilde{\phi}}^*$ corresponds to the symmetric case. Hence,

$$\vec{\tilde{\phi}}^* = \alpha_h \tilde{L}^{-1}\vec{\tilde{\Delta}s} \tag{3.23}$$

where $\vec{\tilde{\Delta}s}_i \equiv \langle s\rangle - s_i$ and the tilde touches on kth row removal.

Plugging Eq.(3.23) into Eq.(3.22):

$$\vec{\tilde{\kappa}} = -\alpha_h\left(\tilde{MD}_s\right)^{-1}\tilde{L}\tilde{L}^{-1}\vec{\tilde{\Delta}s} = -\alpha_h\left(\tilde{MD}_s\right)^{-1}\vec{\tilde{\Delta}s}$$

But $\vec{\tilde{\Delta}s}$ can be written as:

$$\vec{\tilde{\Delta}s} = -\tilde{M}\vec{s} \tag{3.24}$$

Putting it all together:

$$\vec{\tilde{\kappa}} = -\alpha_h\left(\tilde{MD}_s\right)^{-1}\tilde{L}\tilde{L}^{-1}\vec{\tilde{\Delta}s} = \alpha_h\left(\tilde{MD}_s\right)^{-1}\tilde{M}\vec{s}\,\vec{\tilde{\kappa}} \tag{3.25}$$

which in vector form is written as:

$$\vec{\tilde{\kappa}} = \alpha_h\left(\tilde{MD}_s\right)^{-1}\tilde{M}\vec{s}\,\vec{\tilde{\kappa}} = \alpha_h\begin{pmatrix} 1-\frac{s_C}{s_0} \\ 1-\frac{s_C}{s_1} \\ \cdots \\ 1-\frac{s_C}{s_{N-1}} \end{pmatrix} \tag{3.26}$$

And considering the relation between α and κ, in Eq.(3.13):

$$\vec{\alpha} = \alpha_h\begin{pmatrix} 1-\frac{s_C}{s_0} \\ 1-\frac{s_C}{s_1} \\ \cdots \\ 0\ (C\text{ node}) \\ \cdots \\ 1-\frac{s_C}{s_{N-1}} \end{pmatrix} \tag{3.27}$$

Equation (3.26) gives us the tuned values of the frustration parameters as a function of the chosen control node, C, when $\alpha_C = 0$. Notice that the result depends nonlinearly only on the ratio between the degree of each node and the control

node. This informs us that nodes with the same degree would be tuned to the same value or, in other words, the tuning depends only on the degree sequence of the network.

Once we have computed the analytical solution of the frustration parameters, we derive the expression of the required cost to achieve such state with the particular choice of C. Using the definition in Eq.(3.20):

$$e_T(C) = |\alpha_h| \sum_i^{N-1} \left| 1 - \frac{s_C}{s_i} \right| = |\alpha_h| \sum_i^{N-1} \left| \frac{s_i - s_C}{s_i} \right| \tag{3.28}$$

Before we provide the mathematical solution to the minimization problem defined in Eq.(3.21) for this particular case, let us gain an intuitive understanding of it. Looking at Eq.(3.28) we see that the contribution of the ith node to the cost increment depends on $|s_C - s_i|$ and, hence, if the chosen control node, C, has an extreme value, i.e, $s_C \ll s_i$ or $s_C \gg s_i$, the contribution will be larger. On the contrary, if the degree of the control node is similar to that of the remaining nodes, then the increase in cost will be smaller.

For example, the network in Figure 3.3(a), with $\vec{s} = (1,6,2,1,2,2,2,2)$ has the set of unique degrees $\vec{s}_{unique} = (1,2,6)$ and hence three possible values of the cost, shared by some nodes. If $C = \{0,3\}$, $s_C = 1$:

$$e_T(C) = |\alpha_h| \left(|1 - \frac{1}{1}| + 5|1 - \frac{1}{2}| + |1 - \frac{1}{6}| \right) = |\alpha_h| \left(\frac{5}{2} + \frac{5}{6} \right) = \frac{10}{3} |\alpha_h| \tag{3.29}$$

If $C = \{2,4,5,6,7\}$, $s_C = 2$:

$$e_T(C) = |\alpha_h| \left(2|1 - \frac{2}{1}| + 4|1 - \frac{2}{2}| + |1 - \frac{2}{6}| \right) = |\alpha_h| \left(2 + \frac{2}{3} \right) = \frac{8}{3} |\alpha_h| \tag{3.30}$$

And, finally, if $C = 1$, $s_C = 6$:

$$e_T(C) = |\alpha_h| \left(2|1 - \frac{6}{1}| + 5|1 - \frac{6}{2}| \right) = |\alpha_h| (10 + 10) = 20 |\alpha_h| \tag{3.31}$$

The minimum value of the energy is $\frac{8}{3}|\alpha_h|$, corresponding to the choice $C \in \{2,4,5,6,7\}$ with $s_C = 2$.

Notice that the optimal choice of the control node (or nodes) does not depend on the value of α_h in the symmetric configuration, but only on the degree sequence of the network. Moreover, this example illustrates that the degree of the control node corresponds to an intermediate value within the degree sequence of the network and not an extreme value. A more detailed inspection of Eq.(3.28) discloses that the proper choice of the control node (or control nodes) corresponds to the minimization of the relative error of degrees. In order to find

the particular value of the degree that the control node must have we should solve the minimization problem defined in Eq.(3.21):

$$\min_{C,x} e_T(C,x) = \min_{C,x} \sum_{i}^{N} \alpha_i(C,x)$$

which, when considering the symmetric configuration case, turns to

$$\min_{s_C} |\alpha_h| \sum_{i}^{N-1} \left| 1 - \frac{s_C}{s_i} \right| = |\alpha_h| \min_{s_C} \sum_{i}^{N} \left| \frac{s_C - s_i}{s_i} \right| \tag{3.32}$$

Equation (3.32) is equivalent to the minimization of the absolute value of the relative error of the degree:

$$|\alpha_h| \min_{s_C} \sum_{i}^{N} |\mathscr{E}_i| \tag{3.33}$$

where $\mathscr{E}_i = \left| \frac{s_C - s_i}{s_i} \right|$.

The most general minimization problem of the relative error of a variable [155] can be written as

$$\min_{d} \sum_{i=1}^{N} w_i |x_i - d| \; ; d > 0 \tag{3.34}$$

where d is the variable one is interested in and w_i is the weight corresponding to each x_i variable. The solution of Eq.(3.34) is given by

$$d = x_m \text{ , where } m \equiv \min\{i \mid \sum_{k=1}^{i} w_k \geq \sum_{k=i}^{n} w_k\} \qquad i \in \{1, ..., n\} \tag{3.35}$$

In other words, the value of d that minimizes Eq.(3.34) corresponds to the weighted median of the variable x or, equivalently, the 50% weighted percentile. The weighted median of a set n distinct ordered elements $x_1, x_2, ..., x_n$ with positive weights $w_1, w_2, ..., w_n$, is the element x_k satisfying $\min\{i | \sum_{k=1}^{i} w_k \geq \sum_{k=i}^{n} w_k\}$. In other words, the solution is given by x_k, the value such that the sum of the weights at each side of the pivot, k, are as even as possible.

The particular case defined in Eq.(3.32) can be mapped to the most general problem defined in Eq.(3.34), choosing $w_i = 1/s_i$, $x_i = s_i$ and $d = s_C$. Accordingly, the solution of s_C corresponds to the weighted median of the set $\{s_i\}$, with weights given by the inverse of the node degree.

Following the example of the network in Figure 3.3(a), with degree sequence $\vec{s} = (1,6,2,1,2,2,2,2)$, let us compute the optimal value of s_C by using Eq.(3.35):

$$\text{sorted}(\vec{s}) = (1,1,2,2,2,2,2,6) \qquad \vec{w} = \left(1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{6}\right) \tag{3.36}$$

To find the weighted median, we have to find the minimum value such that the sum of the weights at each side of the pivot are as even as possible.

$$1+1+\frac{1}{2}+\frac{1}{2}=3\geq 2.17=\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{6}$$

which corresponds to $s_C=2$, in agreement with the location of the minimum for $\alpha_C=0$ in Figure 3.3(b) corresponding to the network in Figure 3.3(a).

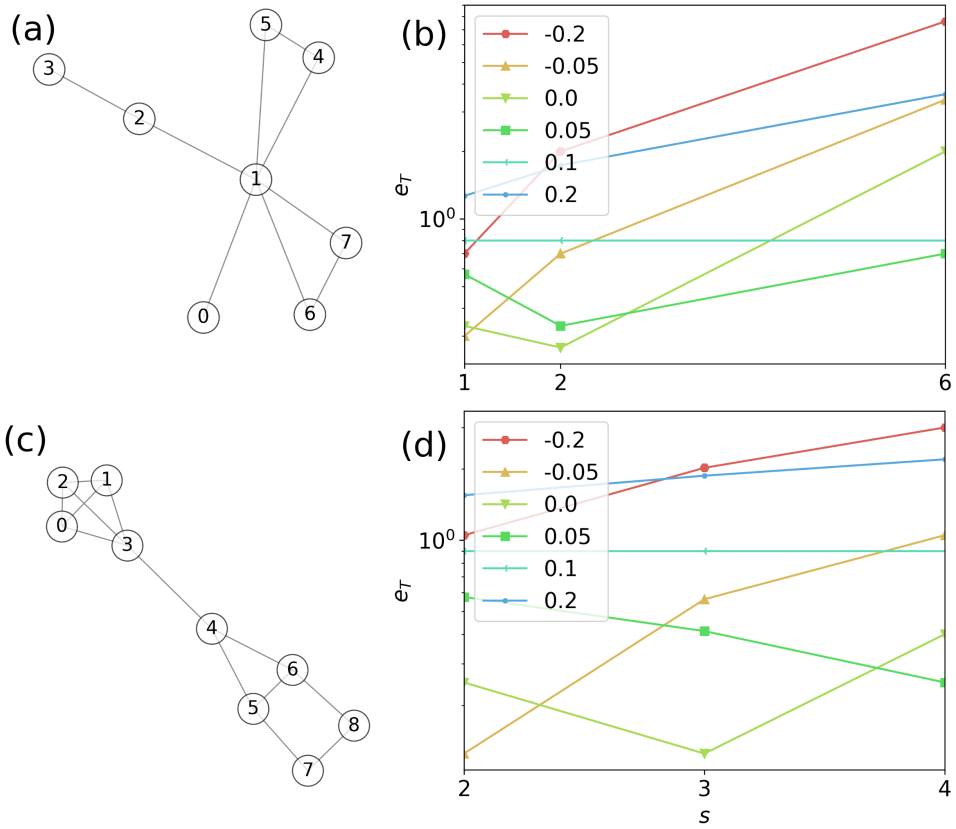


Figure 3.3: **Implied cost to achieve the symmetric configuration as a function of the degree corresponding to different choices of the control node for two different synthetic networks.** Implied cost to achieve the symmetric configuration as a function of the degree corresponding to different choices of the control node, s_C, for a network of 8 nodes (upper panels) and 9 nodes (lower panels). The distinct colors and markers correspond to different values of α_C. The symmetric configuration is generated by a value of $\alpha_h=0.1$.

3.2.1.2 Optimal cost tuning when $\alpha_C \neq 0$

We next ask which is the optimal choice of the control node in the case we let $\alpha_C \neq 0$ and $\omega_i = \omega_h \ \forall i$. In this case, we should look at Eq.(3.14) and set

$\vec{\tilde{\Delta}}\omega = 0$. Making use of the analytical solution of the symmetric configuration in Eq.(3.23):

$$\tilde{MD}_s\vec{\tilde{\kappa}} = -\alpha_h\tilde{L}\tilde{L}^{-1}\vec{\tilde{\Delta}}s - \alpha_C \cdot \vec{\sum_j}[MD_s]_{ij}$$

Using the properties $\tilde{L}\tilde{L}^{-1} = \mathbb{I}$ and $\vec{\tilde{\Delta}}s = \tilde{M}\vec{s}$,

$$\vec{\tilde{\kappa}} = \alpha_h\left(\tilde{MD}_s\right)^{-1}\left(\tilde{M}\vec{s} - \alpha_C\left(\tilde{MD}_s\right)^{-1}\vec{\sum_j}[MD_s]_{ij}\right)$$

Finally, in vector form,

$$\vec{\tilde{\kappa}} = \alpha_h\begin{pmatrix} 1-\frac{s_C}{s_0} \\ 1-\frac{s_C}{s_1} \\ \cdots \\ 1-\frac{s_C}{s_{N-1}} \end{pmatrix} - \alpha_C\begin{pmatrix} 1-\frac{s_C}{s_0} \\ 1-\frac{s_C}{s_1} \\ \cdots \\ 1-\frac{s_C}{s_{N-1}} \end{pmatrix} \tag{3.37}$$

Using Eq.(3.13),

$$\vec{\alpha} = (\alpha_h - \alpha_C)\begin{pmatrix} 1-\frac{s_C}{s_0} \\ 1-\frac{s_C}{s_1} \\ \cdots \\ 0\ (C\text{ node}) \\ \cdots \\ 1-\frac{s_C}{s_{N-1}} \end{pmatrix} + \alpha_C, \tag{3.38}$$

where we have used the result in Eq.(3.26) and the relation

$$\left(\tilde{MD}_s\right)^{-1}\vec{\sum_j}[MD_s]_{ij} = \left(\tilde{MD}_s\right)^{-1}\tilde{M}\vec{s} = = \begin{pmatrix} 1-\frac{\alpha_C}{\alpha_0} \\ 1-\frac{\alpha_C}{\alpha_1} \\ \cdots \\ 1-\frac{\alpha_C}{\alpha_{N-1}} \end{pmatrix}$$

In the particular case that $\alpha_C = \alpha_h$ we recover the trivial initial configuration $\alpha_i = \alpha_h\ \forall i$, as expected from the model.

Once we have computed the analytical solution of the frustration parameters, we derive the expression of the implied cost to achieve such state with the particular choice of C. Using the definition in Eq.(3.20):

$$e_T(C) = \sum_{i=0}^{N-1}\left|(\alpha_h - \alpha_C)\left(1-\frac{s_C}{s_i}\right) + \alpha_C\right| \tag{3.39}$$

We next derive the analytical solution of the optimal choice of the control node and finally proof that the global minimum corresponds to a value of $\alpha_C = 0$.

Equation (3.39) can be rearranged as

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{s_i - \left(1 - \frac{\alpha_C}{\alpha_h}\right) s_C}{s_i/\alpha_h} \right| \tag{3.40}$$

and thereby can be easily mapped to the solution of the minimization problem defined and solved in Eq.(3.34) and Eq.(3.35), respectively. Looking at Eq.(3.39), we should choose $x_i = s_i$, $w_i = \alpha_h/s_i$ and $d = (1 - \alpha_C/\alpha_h)\, s_C$. With this choice, the value of d that minimizes Eq.(3.39) corresponds to the weighted median of the set $\{s_i\}$ with weights α_h/s_i. Therefore, the value of d is the same as the solution of the case $\alpha_C = 0$, but $d \neq s_C$ and thus we must apply a transformation in order to obtain the optimal choice of s_C. We have to distinguish several cases, considering $\alpha_h > 0$:

- $\alpha_C > 0$. In this case we inspect Eq.(3.40) and distinguish two more cases:
  - o $\alpha_C > \alpha_h$: in this case, the prefactor of s_C is negative, and we can write:

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{s_i + \left|1 - \frac{\alpha_C}{\alpha_h}\right| s_C}{s_i/\alpha_h} \right| = \\ = \sum_{i=0}^{N-1} \left| \alpha_h + M\alpha_h \frac{s_C}{s_i} \right|$$

    where $M \equiv \left|1 - \frac{\alpha_C}{\alpha_h}\right| > 0$ is a positive number. Hence, as the cost function increases with increasing s_C, the minimum is achieved when $s_C = \min(s_i)$ (See Figure 3.3 at $\alpha_C = 0.2$).
  - o $\alpha_C < \alpha_h$: in this case, the prefactor of s_C is positive, and we can write:

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{s_i - \left|1 - \frac{\alpha_C}{\alpha_h}\right| s_C}{s_i/\alpha_h} \right|$$

    taking into account that $d = \left|1 - \frac{\alpha_C}{\alpha_h}\right| s_C$ and considering that, in this case, $0 < \alpha_C < \alpha_h$ and hence $0 \leq \left|1 - \frac{\alpha_C}{\alpha_h}\right| \leq 1$ and the weighted mean is bounded by $\min(s_i) \leq d \leq \max(s_i)$, the optimal value of s_C falls in the range $d \leq s_C \leq \max(s_i)$. Hence, the optimal value of s_C is always larger than the weighted median, d (see Figure 3.3 at $\alpha_C = 0.05$).
- $\alpha_C < 0$. In this case we can rewrite Eq.(3.40) as

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{s_i - \left(1 + \frac{|\alpha_C|}{\alpha_h}\right) s_C}{s_i/\alpha_h} \right|$$

  and distinguish two more cases:

- $|\alpha_C| > |\alpha_h|$: In this case, the prefactor of s_C is positive and bounded by $2 \leq \left(1 + \frac{|\alpha_C|}{\alpha_h}\right) \leq \infty$. In this case, $d = \left(1 + \frac{|\alpha_C|}{\alpha_h}\right) s_C$ and hence $0 \leq s_C \leq d/2$. Hence, the optimal value of s_C is always smaller than half the value of the weighted median, d (see Figure 3.3 at $\alpha_C = -0.2$).
- $|\alpha_C| < |\alpha_h|$:n this case, the prefactor of s_C is positive and bounded by $1 \leq \left(1 + \frac{|\alpha_C|}{\alpha_h}\right) \leq 2$. In this case, $d = \left(1 + \frac{|\alpha_C|}{\alpha_h}\right) s_C$ and hence $d/2 \leq s_C \leq d$. Hence, the optimal value of s_C is always smaller than the weighted median, d (see Figure 3.3 at $\alpha_C = -0.05$).

- $\alpha_C = 0$: This case is explored in Section 3.2.1.1. Equation (3.40) turns to

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{s_i - s_C}{s_i/\alpha_h} \right|$$

  . The optimal value of s_C is the same as the weighted median, d, without any further transformation (see Figure 3.3 at $\alpha_C = 0.0$).

- $\alpha_C = \alpha_h$: This case is discussed in the introduction of the present section. Eq.(3.40) turns to

$$e_T(C) = \sum_{i=0}^{N-1} \alpha_h = N\alpha_h$$

  and hence the value of the cost is the same constant value for all nodes (see Figure 3.3 at $\alpha_C = 0.1$).

Amid all the cases considered concerning the value of α_C, the global minimum cost is given by $\alpha_C = 0$, as shown in Figure 3.3. This result can be proved by considering a simplified version of Eq.(3.40), defined as

$$f(x) = \left| \frac{a - (1 - x/b)c}{a/b} \right| \tag{3.41}$$

The minimum value of Eq.(3.41) is achieved when $x = 0$, as long as $a > 0$, $b > 0$ and $c > 0$. This conditions are equivalent to $s_i > 0$, $\alpha_h > 0$ and $s_C > 0$, and are true for all the summation terms in Eq.(3.40). Therefore, the minimum value is given by setting $\alpha_C = 0$.

Summing up, in order to obtain the optimal $\{\alpha_i\}$ parameters' set in order to achieve the symmetric phase configuration with the minimum implied cost in the Kuramoto-Sakaguchi model, we should set $\alpha_C = 0$, independently of the value of α_h. The remaining parameters have to be tuned using Eq.(3.38). Moreover, the optimal choice of the control node (or nodes) corresponds to that with s_C located at the weighted median of $\{s_i\}$ (with weight equal to s_i^{-1}).

Notice also that nodes are grouped by degree regarding the tuned values of its frustration parameters. In other words, there may be different potential control nodes, as long as they share the same degree.

3.2.2 Fully synchronized phase configuration

Another particular phase configuration is given by the phase synchronization of nodes, that is, $\vec{\phi}^* = \vec{0}$. If we set, as in Section 3.2.1, $\omega_i = \omega_h \ \forall i$, we end up with the trivial solution $\alpha_i = 0 \ \forall i$. In the case of full synchronization we want to recover the completely in-phase state from a phase dispersion produced by a distribution of natural frequencies, which we consider to be positive. Hence, applying Eq.(3.14) to this case:

$$\vec{\tilde{\kappa}} = \left(\tilde{M}D_s\right)^{-1}\left(\frac{1}{K}\tilde{M}\vec{\omega} - \alpha_C \cdot \tilde{M}\vec{s}\right) \tag{3.42}$$

and in vector form,

$$\vec{\tilde{\kappa}} = \begin{pmatrix} \frac{\alpha_C(s_C - s_0) - (\omega_C - \omega_0)/K}{s_0} \\ \frac{\alpha_C(s_C - s_1) - (\omega_C - \omega_1)/K}{s_1} \\ \cdots \\ \frac{\alpha_C(s_C - s_{N-1}) - (\omega_C - \omega_{N-1})/K}{s_{N-1}} \end{pmatrix} \tag{3.43}$$

where we have used: $\vec{\tilde{\Delta}}\omega = \tilde{M}\vec{\omega}$. Finally, from the $\vec{\kappa}$ in Eq.(3.43) we can obtain $\vec{\alpha}$:

$$\vec{\alpha} = \begin{pmatrix} \frac{\alpha_C s_C - (\omega_C - \omega_0)/K}{s_0} \\ \frac{\alpha_C s_C - (\omega_C - \omega_1)/K}{s_1} \\ \cdots \\ \alpha_C \\ \cdots \\ \frac{\alpha_C s_C - (\omega_C - \omega_{N-1})/K}{s_{N-1}} \end{pmatrix} \tag{3.44}$$

Similarly as the result of the symmetric configuration, given in Eq.(3.38), the solution of the fully synchronized configuration concerning $\vec{\alpha}$ is a continuous spectrum of values, depending on the choice of the control node, C, the value of its frustration parameter α_C, which is a free parameter, and the natural frequencies of the oscillators. In Sections 3.2.2.1 and 3.2.2.2 we will make a in-depth analysis of the problem, as well as comment on the nonlinear expansion of the Kuramoto-Sakaguchi model and the validity of our approach in this case (Section 3.2.3).

3.2.2.1 Optimal cost tuning when $\alpha_C = 0$

Using the definition of cost in Eq.(3.20) and the general solution of the frustration parameters in Eq.(3.44) we get:

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{\alpha_C s_C - (\omega_C - \omega_i)/K}{s_i} \right| \tag{3.45}$$

In the particular choice $\alpha_C = 0$:

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{\omega_C - \omega_i}{K s_i} \right| \tag{3.46}$$

Equation (3.46) shows that the relevant piece of information regarding the control node is given by its natural frequency, ω_C. Similarly to the minimization problem posed in Section 3.2.1, and in order to find the optimal choice of the control node we need to solve Eq.(3.21) considering the solution of Eq.(3.46):

$$\min_{\omega_C} \left| \frac{\omega_C - \omega_i}{K s_i} \right| = \frac{1}{K} \min_{\omega_C} \left| \frac{\omega_C - \omega_i}{s_i} \right| \tag{3.47}$$

The optimization problem is equivalent to the most general problem, described in Eq.(3.34), with solution given by Eq.(3.35). In this case, $d = \omega_C$, $x_i = \omega_i$ and the weight $w_i = s_i^{-1}$. Accordingly, and in a similar way as in Section 3.2.1, the solution of ω_C corresponds to the weighted median of the set $\{\omega_i\}$, with weights given by the inverse of the node degree. Notice that the optimal choice of the control node is in general different to that given in Section 3.2.1.1). This is due to the fact that the weights of the weighted median have to be sorted according to descending order of natural frequencies instead of node degree.

Following with the example provided in Section 3.2.1.1, for the network in Figure 3.3(a), with degree sequence $\vec{s} = (1,6,2,1,2,2,2,2)$, let us compute the optimal value of ω_C by using Eq.(3.35). Consider the following natural frequencies:

$$\vec{\omega} = (0.1, 0.2, 0.05, 0.45, 0.3, 0.4, 0.25, 0.15), \tag{3.48}$$

which lead to

$$\text{sorted}(\vec{\omega}) = (0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.45) \tag{3.49}$$

and the corresponding weights

$$\vec{w} = \left(\frac{1}{2}, 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1 \right) \tag{3.50}$$

To find the weighted median, we have to find the minimum value such that the sum of the weights at each side of the pivot are as even as possible.

$$\frac{1}{2} + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{2} = 2.67 \geq 2.5 = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1$$

Therefore, the optimal value of natural frequency corresponds to the choice $C = 6$ [see $\alpha_C = 0$ line in Figure 3.4(a)], with $\omega_C = 0.25$ [see $\alpha_C = 0$ line in Figure 3.4(b)] and a degree of $s_C = 2$.

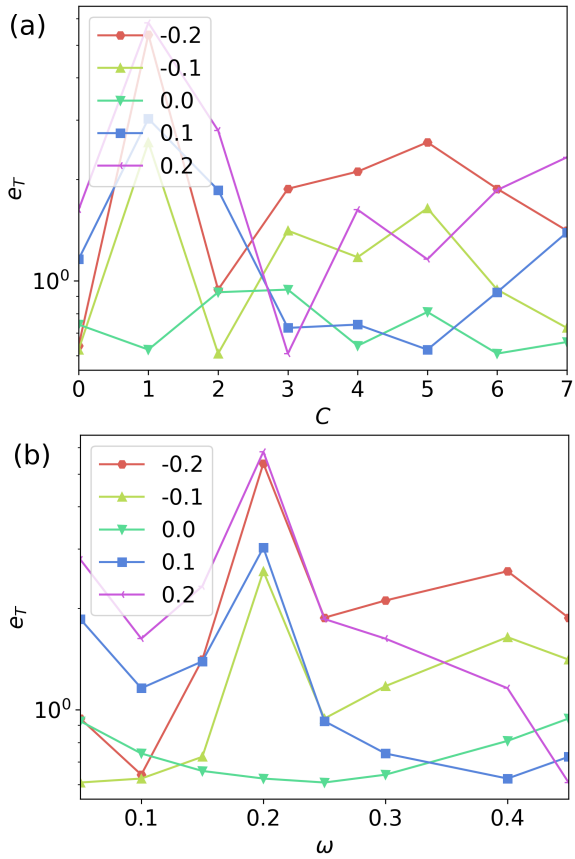


Figure 3.4: **Implied cost to achieve the fully synchronized configuration as a function of the chosen control node and natural frequencies of nodes for the network in Figure 3.3(a).** Implied cost to achieve the fully synchronized configuration as a function of the chosen control node, C (upper panel) and natural frequencies of nodes (bottom panel) for the network in Figure 3.3(a). Five different values of α_C are considered (marked colored lines). Natural frequencies are set as the example in Eq.(3.48).

3.2.2.2 Optimal cost tuning when $\alpha_C \neq 0$

The cost corresponding to the fully synchronized configuration case is given by Eq.(3.45). In the general case where $\alpha_C \neq 0$, we can minimize the cost with respect to ω_C or to s_C. If we minimize with respect to ω_i, we first have to rewrite Eq.(3.45) as

$$e_T(C) = \sum_{i=0}^{N-1} \left| \frac{\alpha_C s_C - (\omega_C - \omega_i)/K}{s_i} \right| = \frac{1}{K} \sum_{i=0}^{N-1} \left| \frac{\omega_i - (\omega_C - \alpha_C s_C K)}{s_i} \right| \tag{3.51}$$

Again, the problem and the solution of Eq.(3.51) can be taken from Eq.(3.34) and Eq.(3.35), choosing $d \equiv \omega_C - \alpha_C s_C K$, $w_i \equiv 1/Ks_i$ and $x_i \equiv \omega_i$.

Hence, the value d that minimizes the cost is the weighed median considering the same weight as in Section 3.2.2.1, $w_i = 1/s_i$ (notice, however, that the ordering is determined by natural frequencies and not degrees). Let us analyze the different possibilities regarding the values of α_C, maintaining ω_C and s_C constant:

- $\omega_C > \alpha_C s_C K$ or $\alpha_C < \frac{\omega_C}{K s_C}$: We can write

$$\sum_{i}^{N} \left| \frac{\omega_i - |\omega_C - \alpha_C s_C K|}{K s_i} \right|$$

  The value which minimizes cost is given by $d = \omega_k$, corresponding to the weighted median. However this is not directly the value of ω_C, as $d = |\omega_C - \alpha_C s_C K|$ in this case. The real values of the pair $\{\omega_C, s_C\}$ are given by $\min_C(\omega_k - (\omega_C - \alpha_C s_C K))$. Following with the example in Section 3.2.2.1, the value of the weighted median is $d = 0.25$. In the case we are considering, however, this is not the optimal choice of the parameters for the control node. We must shift the values considering the relation between d and the other parameters. If we choose $\alpha_C = 0.1$, for instance, we find that, $|\omega_C - 0.1 s_c| = 0.25$. In Figure 3.4 we see that the optimal choice is given by $\omega_C = 0.4$, which corresponds to $C = 5$ and $s_C = 2$.

- $\omega_C < \alpha_C s_C K$ or $\alpha_C > \frac{\omega_C}{K s_C}$: We can write

$$\sum_{i}^{N} \left| \frac{\omega_i + |\omega_C - \alpha_C s_C K|}{K s_i} \right|$$

  Hence, as the function increases with increasing $(\omega_i - \alpha_C s_i K)$, the minimum is achieved by $\min_C(\omega_C - \alpha_C s_C K)$.

3.2.3 Non-linear expansion of the Kuramoto-Sakaguchi model

The results obtained in Section 3.2.2 are based on a linear approximation of the Kuramoto-Sakaguchi model. We have derived the results based on the phase synchronization requirement, and assuming that frequency synchronization is already achieved in the steady state. Nevertheless, when measuring the order parameter with a large dispersion of natural frequencies or low coupling constant, we do not expect such steady state. However, we ask to which extend the proposed values of the obtained frustration parameters are also able to enhance frequency synchronization considering the original nonlinear Kuramoto-Sakaguchi model:

$$\dot{\theta}_i = \omega_i + K\sum_j W_{ij}\sin(\theta_j - \theta_i - \alpha_i) \tag{3.52}$$

We compare the results from Ref.[38] considering its *Type II* frustration parameters tuning for both the linear and the nonlinear Kuramoto model and we find that, despite our approach does not consider the enhancement of frequency synchronization on the nonlinear regime, it is able to improve the value of the order parameter, in a similar fashion as in Ref.[38]. This work considers the nonlinear Kuramoto-Sakaguchi model and seeks to improve the number of nodes that fall into the recruitment condition so as to achieve the same common oscillatory frequency. The considered network class is the same as the mentioned paper, as well as the statistics study.

We make use of the expression in Eq.(3.44) to tune the set of $\vec{\alpha}$ for a given configuration of random $\vec{\omega}$ and study the effect on the synchronization of the system for different values of the coupling strength.

We consider two cases: the linear and the nonlinear model with natural frequencies obtained from a uniform distribution $\omega_i \in [-1,1]$. From Figure 3.5(a), the linear case of the Kuramoto-Sakaguchi model [see Eq.(3.2)], our approach, derived from the analytic expression of the linear approximation, advances the analytic tuning of frustration parameters suggested by Ref.[38]. This is because they look for an enhancement in the number of nodes that are oscillating at the same frequency, Ω, but they do not worry about the exact values of the phases they achieve. On the contrary, we assume nodes are already synchronized (without setting the specific value of Ω, as they do) and we look for the full synchronization state.

In Figure 3.5(b), the linear tuning squared discontinuous purple line) approaches the type II (squared dashed green line) tuning in the case of the nonlinear Kuramoto-Sakaguchi model, even for small values of the coupling strength. Hence, despite the aim of our approach is not achieving frequency synchronization, the obtained tuning of the frustration parameters helps enhancing it as well. In principle this behavior is reminiscent of the so called explosive per-

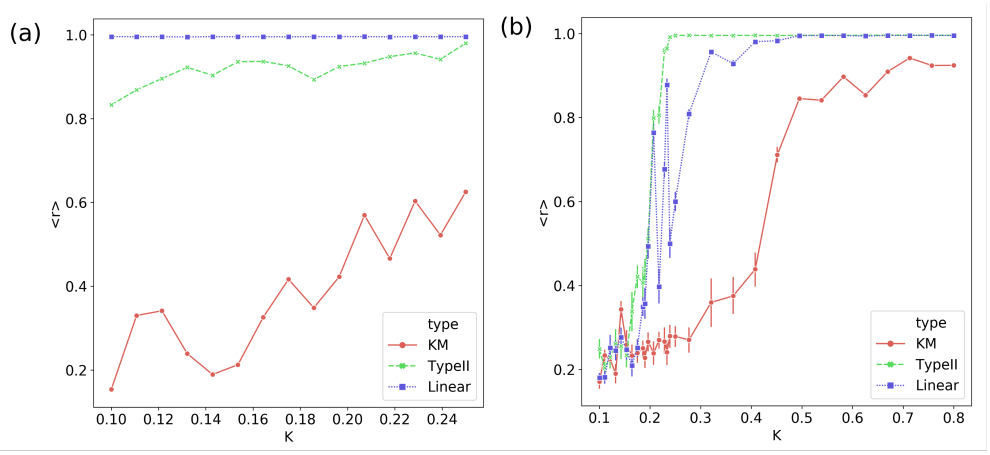


Figure 3.5: **Average order parameter as a function of the coupling strength for the linear and the nonlinear Kuramoto-Sakaguchi (KS) model on regular random graphs with homogeneous node degree.** Average order parameter, $\langle r \rangle$, as a function of the coupling strength, K, for the linear [in panel (a)] and the nonlinear [in panel (b)] Kuramoto-Sakaguchi (KS) model on regular random graphs with homogeneous node degree $s_i = 4$ and $N = 100$ nodes. Natural frequencies are obtained from a uniform random distribution in the range $\omega_i \in [-1, 1]$. Each data point represents an average over ten optimized configurations. We compare three types of tuning for the set of frustration parameters, $\{\alpha_i\}$: the original Kuramoto dynamics or $\alpha_i = 0 \ \forall i$ (spotted continuous red line); type II [38] KS dynamics with the frustration parameters set to $\sin(\alpha_i) = -\omega_i/(Ks_i)$ if $|\omega_i| < Ks_i$ and $\sin(\alpha_i) = \pm 1$ otherwise (squared dashed green line); and KS dynamics with the frustration parameters determined by the derived linear approximation in Eq.(3.44))(squared discontinuous purple line).

colation (see Ref.[54] and references therein), since the transition to the synchronized state is abrupt, as it happens in a first order phase transition. We are adjusting the phase-lag parameter as a response to the frequencies, and then in some sense it is similar to the original proposal in Ref.[76], the correlated degree-frequency framework.

3.3 Discussion

The Kuramoto-Sakaguchi model adds to the original Kuramoto model a homogeneous phase lag, α, between nodes which promotes a phase shift between oscillators. We consider a more general framework, in which the phase lag or the frustration parameter, α_i, is an intrinsic property of each node. A very relevant question in oscillatory models is finding the conditions of network synchronization. In the present chapter, we bring forward a methodology not only to obtain the desired synchronized state, but any convenient phase configuration in the steady state, by means of a fine tuning of the phase lag or frustration parameters, $\{\alpha_i\}$. We feature the analytical solution of frustration parameters so as to achieve any phase configuration, by linearizing the most general model. The three intrinsic parameters of the nodes in the model, natural frequencies, $\{\omega_i\}$, frustration parameters $\{\alpha_i\}$, and phases in the steady state ϕ_i^*, are coupled by an equation that allows to tune them for a desired configuration. While the set ϕ_i^* is uniquely determined, the set α_i has a continuous spectrum of solutions.

A main result is that a given phase configuration can be access via a continuous spectrum of frustration parameters, i.e, one phase and one frustration parameter are left as free parameters. The nodes we choose their values concerning phase and frustration parameter, are named reference and control nodes, respectively. Once the frustration parameters are tuned so as to obtain the desired state, we define a cost function to assess the overhead that the system requires to achieve such parameters' configuration. Among all possible tuning solutions of $\{\alpha_i\}$, we request those which minimize the cost to obtain them. We develop the analytical solution of the cost function for the cases of symmetric configuration and fully synchronized state and discuss them.

A key result is the solution to the minimization cost problem: For the case of symmetric configuration, the nodes which are to be set as control nodes are those whose degree is the weighed median of the sample, with a weight equal to the inverse of its degree. On the other hand, for the case of fully synchronized state, control nodes are those whose natural frequency is the weighted median of the sample, with a weight equal to the inverse of its degree. An extensive analysis of several cases is done in the text and a detailed example of a toy network is provided. We highlight the connection made with the nonlinear Kuramoto-

Sakaguchi model. Despite our analysis being based on the linear version of the model, we show that the proposed parameters' tuning is also able to enhance frequency synchronization, as done in Ref.[38]. We stress the fact that the question 'among all the possible solutions, which is the one that makes the system achieve a particular phase configuration with the minimum required cost?' is of particular relevance when we consider the plausible real nature of the system. If a real system needs to access a particular phase configuration, which may be associated with a singular function, then it will tend to minimize the effort or cost to do so. Further work can be done within this framework by doing real experiments on measuring the energy needed to access a particular configuration. Moreover, other nonlinear oscillatory models can be analyzed and compared with the Kuramoto-Sakaguchi model.

Other questions regarding the model are left open. We have considered the coupled trio of natural frequencies-frustration parameters-steady state phases. A natural extension to this would be to inspect the possibility to also tune the weights of the network edges in order to access a particular configuration. The higher dimension of the latter with respect to the vectors of parameters would require further assumptions about the model or the network structure, such as positive weights or particular distributions or topologies. Another research venue would be to consider the effect of removing a node of the network and the $\{\alpha_i\}$ set needed to minimize the effect on the removal on the whole network.

Despite we provide the analytical solution to the optimal choice of parameters in order to minimize the cost of achieving both the symmetric and the fully synchronized configurations, the access to all nodes' parameters requirement may not be feasible in real-world networks. Our methodology is quite general and the optimization procedure refers to a set of parameters to be tuned. In particular, a finite subset of nodes with accessible phase-lag parameter could be chosen (the choice could be restricted to any subset of nodes), holding all other nodes unaltered. This would provide a nonoptimal global condition but a restricted and approximated one that could deal with a subset of available nodes. A meaningful analysis would be to identify which subset of nodes is the one that enables to get a closest approximate solution, and relate those nodes with their topological properties, although this question is beyond the goal of this work.

3.4 Additional Information

3.4.1 Step-by-step derivation of the frustration parameters

In the present Section we illustrate the procedure and result equivalent to the compact expression derived in Eq.(3.16) with a simple example of a synthetic

network. Consider the network in Figure 3.1, with its Laplacian matrix:

$$L = \begin{pmatrix} 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & 1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

We expand

$$\sum_j L_{ij}\theta_j^* = \frac{\omega_i}{K} - \frac{\langle\omega\rangle}{K} + \langle\alpha s\rangle - \alpha_i s_i \;\forall i$$

(Eq.(3.6) in previous sections) step by step:

$$\begin{pmatrix} 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & 1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \theta_0^* \\ \theta_1^* \\ \theta_2^* \\ \theta_3^* \\ \theta_4^* \\ \theta_5^* \\ \theta_6^* \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\omega_0-\langle\omega\rangle}{K} \\ \frac{\omega_1-\langle\omega\rangle}{K} \\ \frac{\omega_2-\langle\omega\rangle}{K} \\ \frac{\omega_3-\langle\omega\rangle}{K} \\ \frac{\omega_4-\langle\omega\rangle}{K} \\ \frac{\omega_5-\langle\omega\rangle}{K} \\ \frac{\omega_6-\langle\omega\rangle}{K} \end{pmatrix} - \begin{pmatrix} \frac{6}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & \frac{6}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & \frac{6}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & \frac{6}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & \frac{6}{7} & -\frac{1}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & \frac{6}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & \frac{6}{7} \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\omega_0-\langle\omega\rangle}{K} \\ \frac{\omega_1-\langle\omega\rangle}{K} \\ \frac{\omega_2-\langle\omega\rangle}{K} \\ \frac{\omega_3-\langle\omega\rangle}{K} \\ \frac{\omega_4-\langle\omega\rangle}{K} \\ \frac{\omega_5-\langle\omega\rangle}{K} \\ \frac{\omega_6-\langle\omega\rangle}{K} \end{pmatrix} + \begin{pmatrix} -\frac{24}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix}$$

If we set all natural frequencies to the same value: $\omega_i = \omega \; \forall i$:

$$\begin{pmatrix} 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & 1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \theta_0^* \\ \theta_1^* \\ \theta_2^* \\ \theta_3^* \\ \theta_4^* \\ \theta_5^* \\ \theta_6^* \end{pmatrix} =$$

$$= \begin{pmatrix} -\frac{24}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{pmatrix}$$

Now we choose $R = 0$ and $C = 1$, i.e., all $\phi_i = \theta_i - \theta_0$ and $\kappa_i = \alpha_i - \alpha_1$ (arbitrary choices). Let us write explicitly the change of variables (the red and the blue columns are ones we can remove due to the change of variables, as they do not affect to the system of equations anymore):

$$\begin{pmatrix} 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & 1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \phi_0^* = 0 \\ \phi_1^* \\ \phi_2^* \\ \phi_3^* \\ \phi_4^* \\ \phi_5^* \\ \phi_6^* \end{pmatrix} =$$

$$= \begin{pmatrix} -\frac{24}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} \end{pmatrix} \cdot \begin{pmatrix} \kappa_0 \\ \kappa_1 = 0 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} + \begin{pmatrix} \frac{-12}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \end{pmatrix}$$

If we look carefully at Eq.(3.53), we see that although the left-hand side and the right hand-side matrices are both singular, the first one has both column and row sums equal to zero, while the second one has only column sum equal to zero. This is reflected in the additional constant term that appears when doing

the change of variables regarding α_i, which can be written as:

$$b_i = \sum_j [M \cdot D_s]_{ij} \neq 0, \qquad \text{in general} \tag{3.53}$$

We can choose whatever row to remove from either sides. We choose row 0:

$$\begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \phi_1^* \\ \phi_2^* \\ \phi_3^* \\ \phi_4^* \\ \phi_5^* \\ \phi_6^* \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} \end{pmatrix} \cdot \begin{pmatrix} \kappa_0 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} + \begin{pmatrix} \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \end{pmatrix}$$

In this situation, we can solve for the set $\vec{\vec{\kappa}}$:

$$\begin{pmatrix} \kappa_0 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} = \begin{pmatrix} \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{12}{7} \end{pmatrix}^{-1} \cdot$$

$$\cdot \left[\begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \phi_1^* \\ \phi_2^* \\ \phi_3^* \\ \phi_4^* \\ \phi_5^* \\ \phi_6^* \end{pmatrix} - \begin{pmatrix} \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \\ \frac{2}{7}\alpha_1 \end{pmatrix} \right],$$

which leads to the result:

$$\begin{pmatrix} \kappa_0 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} = \begin{pmatrix} \frac{-2\alpha_1+3\phi_1+\phi_3+3\phi_6}{4} \\ \frac{3(\phi_1-\phi_2)}{2} \\ \frac{2\phi_1-\phi_2-2\phi_3+\phi_4}{2} \\ \frac{2\phi_1-\phi_2+\phi_3-2\phi_4+\phi_5}{2} \\ \frac{2\phi_1-\phi_2+\phi_4-2\phi_5-\phi_6}{2} \\ \frac{2\phi_1\phi_2+\phi_5-2\phi_6}{2} \end{pmatrix} \cdot \tag{3.54}$$

We recall that $\kappa_1 = 0$ and $\phi_0 = 0$.

Let us consider the following arbitrary phase configuration, as an example:

$$\vec{\phi}_{(R=0)} = (0.1, 0.2, 0.25, -0.2, -0.1, 0.0). \tag{3.55}$$

In this case and using Eq.(3.54) we obtain the values of the frustration parameters:

$$\vec{\kappa}_{(C=1)} = (0.1375 - \alpha_1/2, -0.15, -0.35, 0.275, 0.0, -0.05)$$

Bear in mind the definition $\kappa_i \equiv \alpha_i - \alpha_C \Rightarrow \alpha_i = \kappa_i + \alpha_C$ to obtain the final values. If we choose $\alpha_C = 0 \Rightarrow \alpha_i = \kappa_i$, then we can include explicitly the value of the control node $C = 1$:

$$\vec{\alpha} = (0.1375, 0.0, -0.15, -0.35, 0.275, 0.0, -0.05)$$

Alternatively, we can choose whatever value we need regarding the control node. For instance, if $\alpha_C = \alpha_1 = 0.1$ and the phases configuration is also defined by Eq.(3.55),. the solution of the frustration parameter tuning is

$$\vec{\alpha} = (0.1875, 0.1, -0.05, -0.25, 0.375, 0.1, 0.05).$$

Figure 3.6 shows an example of possible phase configuration (the one provided by the symmetric configuration, i.e., four distinct cluster of nodes)

3.4.2 Solution of the cost optimization problem

In order to gain a more intuitive understating of the analytical expression and solution of the considered cost function, we consider the analysis of the continuous case.

Symmetric configuration case

Considering the symmetric configuration case and choosing $\alpha_C = 0$, the continuous optimization problem can be written as

$$\frac{\partial e_T(C)}{\partial s_C} = \frac{\partial}{\partial s_C}|\alpha_h| \sum_i^{N-1} \left|1 - \frac{s_C}{s_i}\right| = |\alpha_h| \sum_i^{N-1} \frac{\mathrm{sgn}(s_C - s_i)}{s_i} \tag{3.56}$$

Equation (3.56) is based on the function

$$f(x) = \left|\frac{a-x}{a}\right| \; a, x > 0 \tag{3.57}$$

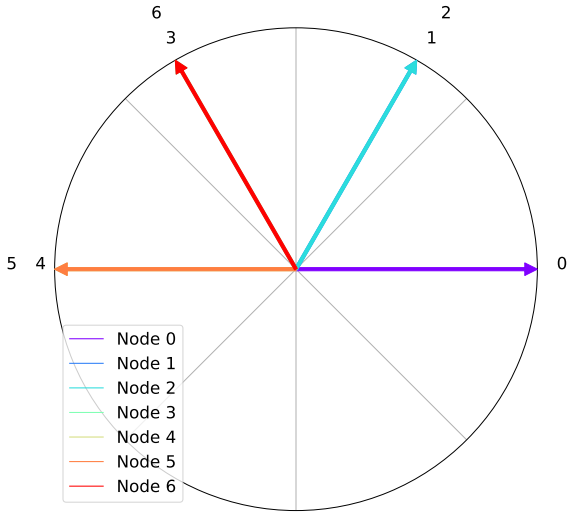


Figure 3.6: **Polar plot of the scaled phases after tuning the set of frustration parameters to the symmetric configuration.** For the network in Fig 3.1, phases obtained after tuning the set of frustration parameters to the symmetric configuration (four distinct symmetries).

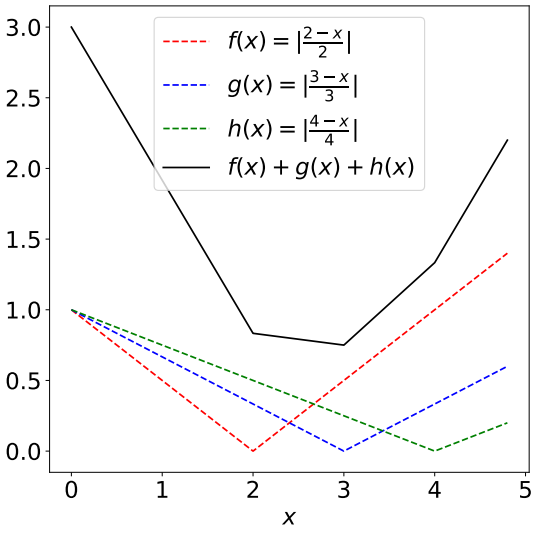


Figure 3.7: **Examples of the function defined in Eq.(3.57).** Three examples of the general function $f(x) = |\frac{a-x}{a}|$, with $a = 2$, $a = 3$ and $a = 4$, and the resulting sum of them.

which is depicted in Figure 3.7 for different values of a and the sum of all of them. Regardless of the set of a_i values, the sum function $\sum_i f(x,a_i)$ (see the example in the black line in Figure 3.7), is a concave function and has a unique minimum, which corresponds to one of the a_i values.

In order to assess the value of a_i where the minimum is located, we compute the derivative of Eq.(3.57):

$$\frac{df(x)}{dx} = \frac{\text{sgn}(x-a)}{a} \tag{3.58}$$

and hence, $d\sum_i f(x,a_i)/dx = \sum_i \text{sgn}(x-a_i)/a_i$, which is depicted in Figure 3.8. Notice that, despite the derivative of the function is not defined at the values

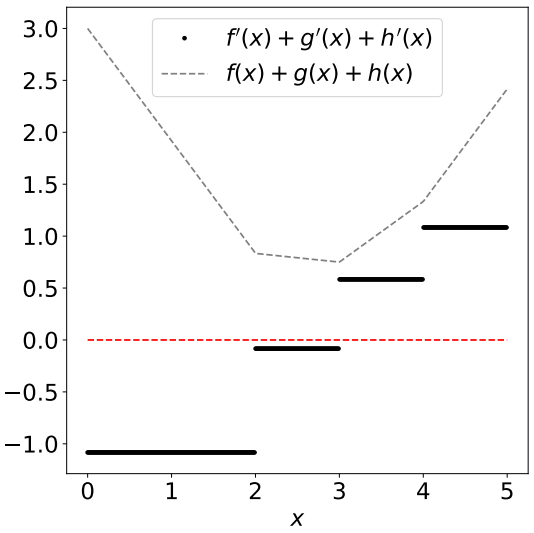


Figure 3.8: **Derivative of a composed function based on Eq.(3.57).** Derivative of the function $f(x) = |\frac{2-x}{2}| + |\frac{3-x}{3}| + |\frac{4-x}{4}|$ defined in Figure 3.7. Red dashed line at $y = 0$.

$x = a_i$, the derivative changes its sign when moving from $x < 3$ to $x > 3$ and hence, the minimum is located at this value of a_i.

To conclude, Eq(3.56) behaves equivalently as the function defined in Eq.(3.57) and hence, displays only one minimum, which is achieved at the s_i where there is a change of sign in the derivative.

Alternatively and as explained in the main text, we can understand the minimization problem as part of a general framework. The minimization of Eq.(3.56) is equivalent to the minimization of the absolute value of the relative error:

$$\sum_i^{N-1} \left|1 - \frac{s_C}{s_i}\right| = \sum_i^{N} \left|\frac{s_C - s_i}{s_i}\right| = \sum_i^{N} |\mathscr{E}_i| \tag{3.59}$$

The general problem can be written as [155]:

$$\min_{d} \sum_{i=1}^{N} w_i |x_i - d| \; ; d > 0$$

with the solution:

$$d = x_m \text{ where } m \equiv \min\{i \mid \sum_{k=1}^{i} w_k \geq \sum_{k=i}^{n} w_k\} \quad i \in \{1, ..., n\} \tag{3.60}$$

In other words, the d value that minimizes Eq.(3.60) corresponds to the weighted median of the variable x, or the 50% weighted percentile.

Weighted median: For n distinct ordered elements $x_1, x_2, ..., x_n$ with positive weights $w_1, w_2, ..., w_n$, the weighted median is the element x_k satisfying $\min\{i | \sum_{k=1}^{i} w_k \geq \sum_{k=i}^{n} w_k\}$

Therefore, the solution is given by x_k, the value such that the sum of the weights at each side of the pivot, k, are as even as possible.

Our problem is a special case of the discrete weighted medians with weights $1/s_i$, which are a special case of the medians of a measure.

Following the example provided in Figure 3.7, $\{x\} = \{2, 3, 4\}$ and $\{w\} = \{1/2, 1/3, 1/4\}$.

The weighted median is achieved for $k = 2$, corresponding to $x_2 = 3$ and weight $w_2 = 1/3$ as $1/2 + 1/3 = 5/6 > 1/3 + 1/4 = 7/12$. Conversely, if we let $k = 1$, and hence $x_1 = 2$ and $w_1 = 1/2$, the condition on the weights will not be true: $1/2 \not> 1/2 + 1/3 + 1/4$.

CHAPTER 4

Quasi-Symmetries in Complex Networks

Complex networks – from biological networks such as the brain connectome or regulatory networks to social and technological networks, like scientific collaboration networks or the Internet [42, 18, 43, 135, 116] – are widely used to model the structure and behaviour of complex systems. Despite these apparently diverse networks are unique in its nature, many studies have shown that they share a number of properties, which distinguish them from other mathematical graphs of interest. Such common features include the heterogeneity in its node degrees, captured by a power-law distribution, high clustering coefficients, and the 'small-world' property, among others [174, 8, 29, 166, 74]. Additionally, a certain degree of symmetry is also an attribute of real-world networks [159, 152]. The study of the symmetries of a network is of great relevance for several reasons: it may help us to have a better understanding of the formation of certain real-world networks, they can also provide information about node function, and have an effect on network redundancy and robustness. Moreover, symmetries are known to influence the outcome of network dynamics, such as synchronization or controllability [100, 98, 177, 124, 136, 85].

The notion of 'symmetry' or 'invariance' includes several specifications depending on the field it is applied [183]. Mathematically, a symmetric transformation, or a symmetry is the set of transformations that leaves an object invariant or unmodified [49]. Differently than continuous transformations, such as a translation or a rotation applied to a geometric shape, symmetries in complex networks are necessarily discrete transformations applied to graphs, which are defined as discrete entities. Importantly, graphs are topological objects and generally, their properties are independent of the positions of vertices or lengths of the links. For this reason, a geometric transformation of their components has no effect on the topology, but to the visualization of the graph . In a different way, a topological transformation of a graph maps each vertex to another one as a permutation. Finally, the set of permutations of a graph that leaves the topology invariant are the automorphisms of the graph (in Section 4.1 the notion of symmetries in complex networks is explained in depth). Other types of symmetries that may be present in graphs are scale invariance or translational symmetries, which are not considered in the present chapter [69].

Built on the standard notion of graph symmetry that we have reported, i.e, topological or structural symmetry, other weaker or approximate symmetries may be present in real-world networks. Despite they are not included in the finite number of automorphisms of graphs, they indeed play an important part in determining the network behaviour [69, 164, 127]. Alternatives for approximate symmetries in graphs include 'near' symmetries and 'stochastic symmetries' [159, 81]. A 'near' symmetry is described in terms of properties of the network that are left unchanged when some other transformation is applied on the network. Examples include whether two nodes have the same degree, and/or the same number of second neighbours, and/or the same local clustering coefficient. A more relaxed condition consists in whether two nodes are 'statistically' equivalent, that is, whether these topological properties are the same in an average sense. The permutation of statistically equivalent nodes are called stochastic symmetries and they result in a family of statistically equivalent networks with the same statistical properties [69].

The given alternatives to perfect or standard topological symmetries in graphs are of great interest as small fluctuations or errors may be present when constructing the graphs, as well as additional and/or missing links could be included/removed. The resulting graphs or networks may lead to very significant changes in the analysis of topological symmetries, as many of them will remain hidden due to its approximate nature.

In the present chapter and in line with the analysis of approximate symmetries, we propose a different extension of the latter, which we call 'Quasi-Symmetries'. This alternative definition of weaker symmetries remains free to impose any invariance of a particular topological property. Quasi-Symmetries are obtained from the network as an extension to structural equivalence; structural or topological similarity is derived for all pair of nodes from an oscillatory dynamical model: the Kuramoto-Sakaguchi model [150]. According to this model, all nodes are considered as individual phase-oscillators that are coupled with its neighbours by a sinus function of its phase difference. The phase differences between them at the steady-state configuration determine the degree of structural similarity, as shown in Section 4.1.2. The analysis of quasi-symmetries provides insights to otherwise hidden properties of real-world networks. Firstly, we explore the distributions of structural similarity among all pairs of nodes and we find a benchmark to determine whether a network has a more complex pattern than that of a random network concerning quasi-symmetries . Secondly, we define the 'dual network', a weighted network (and its corresponding binarized counterpart) that effectively encodes all the information of quasi-symmetries in the original one. The dual network allows for the analysis of centrality measures and community detection. The first informs us about the nodes that play a unique role in the network or those that behave similarly to many other nodes.

The latter results to a classification of nodes into quasi-symmetric communities, the natural extension of the automorphism group orbits (structurally symmetric nodes) of a network.

All the results presented in this chapter can be found in Ref.[146].

4.1 Symmetries in complex networks

A network or, mathematically, a simple graph, $\mathcal{G}(\mathcal{V},\mathcal{E})$, consists of a set of nodes, $\mathcal{V}(\mathcal{G})$, linked by a set of edges $\mathcal{E}(\mathcal{G})$. A network of n nodes, labelled from 0 to $n-1$, can be represented by its adjacency matrix, A, a $n \times n$ matrix with $a_{ij} = 1$ if there is a link between nodes i and j and $a_{ij} = 0$ otherwise. A permutation, or relabelling, of the nodes of a network can be written as $\pi(\mathcal{V}) : \{0,1,...,n-1\} \rightarrow \{\pi(0),\pi(1),...,\pi(n-1)\}$ where, for instance, node 0 changes to $\pi(0)$. Equivalently, a permutation can be represented in a two-line form as follows,

$$\pi(\mathcal{V}) : \begin{pmatrix} 0 & 1 & ... & n-1 \\ \pi(0) & \pi(1) & ... & \pi(n-1) \end{pmatrix} \tag{4.1}$$

P_π is a square matrix that corresponds to the permutation $\pi(\mathcal{V})$ and is obtained by permuting the columns of the identity matrix, i.e., the element $p_{ij} = 1$ if $\pi(i) = j$ and 0 otherwise.

The concept of network symmetry is akin to the mathematical definition of a graph automorphism, which is a permutation of the network nodes but preserving adjacency. In other words, neighbouring nodes still remain neighbours after the permutation is applied. Namely, a graph automorphism $\sigma(\mathcal{V})$ is a permutation of the vertices $\sigma(\mathcal{V})$ such that $(\sigma(i),\sigma(j))$ is an edge only if (i,j) is an edge: the set of edges is preserved. Consequently, the permutation matrix corresponding to a graph automorphism or a symmetry, P_σ, commutes with the adjacency matrix of the network.

$$AP_\sigma = P_\sigma A \tag{4.2}$$

The set of all the symmetries of a graph form the automorphism group of the graph, $Aut(\mathcal{G})$. In Reference [55], a graph is defined as symmetric when there exits at least a non-identical permutation of its vertices that leaves the graph invariant or, equivalently, the group of its automorphisms has a degree greater than 1.

The set of vertices can be split into the core of fixed points, V_0, that is, vertices which are moved by none of the automorphisms of $Aut(\mathcal{G})$, and the vertex set of symmetric motifs, M_i. This partition is called the geometric decomposition of the network and can be written as

$$V = V_0 \cup M_1 \cup ... \cup M_m \tag{4.3}$$

being m the number of symmetric motifs. Each symmetric motif can be further partitioned into clusters. Two nodes, v_i and v_j, belong to the same cluster if $\sigma(v_i) = v_j$ and conversely, where $\sigma \in Aut(\mathcal{G})$. Clusters are alternatively called orbits induced by $Aut(\mathcal{G})$. The vertices or nodes of the same orbit are structurally indistinguishable and play the same structural role in the network (nodes are colored by orbit in Fig. 4.1). We can classify symmetric motifs into two types:

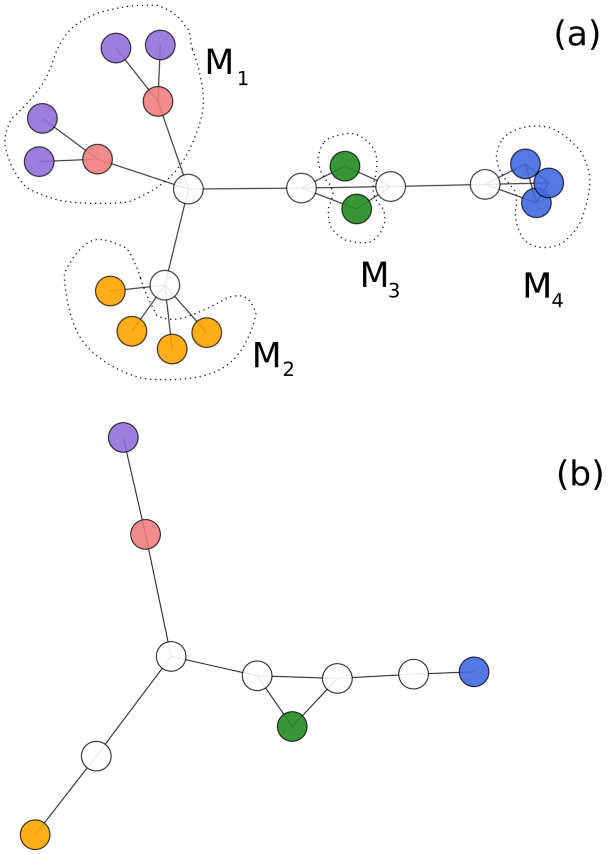


Figure 4.1: **Geometric decomposition into the asymmetric core and four symmetric motifs of a network.** Geometric decomposition into the asymmetric core and four symmetric motifs of a network, in panels (b) and (a), respectively. Nodes are colored by orbit and fixed points are in white color. Motifs M_2, M_3 and M_4 correspond to BSMs and motif M_1 is non-basic.

basic and complex. Basic symmetric motifs (BSMs) are made of one or more orbits of the same number of vertices (motifs M_2, M_3 and M_4 in Fig. 4.1) and complex symmetric motifs are hardly found in real-world networks, and they are typically branched trees (motif M_1 in Fig. 4.1)[104, 70, 152]. The detection of graph automorphisms and the corresponding geometric decomposition of a network is vastly used to simplify the topology of the network by compressing re-

dundant information. Moreover, the basic structural properties of the network can be derived only from the geometric decomposition of the graph or the so called quotient graph. Network eigenvalues are an example of it.

In the present chapter we are interested in detecting the nodes that are structurally equivalent, that is, nodes that play the same role in a network and therefore, we will be detecting the orbits generated by the automorphism group of a network, $Aut(\mathcal{G})$. Notice that a symmetric motif may be subdivided into several orbits and that the isolated permutation of two nodes belonging to the same orbit needs not correspond to an automorphism of the network.

The notion of 'structural equivalence' or a pair of nodes being structurally equivalent is alternatively defined in the social sciences as: if two nodes have exactly the same set of neighbours, regardless of whether they are neighbours of each other, then a permutation between them exists such that the network remains unchanged. Notice, however, that this definition is more restrictive that two (or more) nodes being structurally equivalent as long as they belong to the same orbit, which may not share the same neighbours, however.

4.1.1 Generation of symmetric networks

By examining the automorphism group of real-world networks, several studies show that real networks, unlike random graphs, contain a large amount of symmetries, namely, network motifs[11]. This is partly due to the fact that symmetry can arise from growth processes present in nature. However, the availability of real network datasets is often scarce, especially, when looking for enough variability regarding symmetry. Alternately, we can use random graphs generating models, such as Erdös-Rényi, Watts-Strogatz or Barabási-Albert, but these models do not generate graphs with symmetries, and hence we should turn to regular graphs in order to work with symmetries. Such motifs are however trivial and easy-to-identify by visual inspection.

In the present chapter we will use an algorithm that is able to generate graphs with any desired symmetry pattern [89]. Hereafter, we provide a schematic explanation of the algorithm and the main required concepts.

An *equitable partition*(EP) of the nodes divides the graph into non-overlapping clusters of nodes, $\{C_i\}$, such that the number of connections to C_j from any node $v \in C_i$ only depends on i and j, that is, their corresponding clusters [153].

The automorphism group, $Aut(\mathcal{G})$, of a graph induces an equitable partition of nodes, where the clusters of the EP are the orbits generated by $Aut(\mathcal{G})$.

An equitable partition of a graph can be represented by its *quotient graph*, $\mathscr{Q}$. The quotient graph of an EP consists of five components:

$$\mathscr{Q} = \{\mathscr{C}, \mathscr{F}, \vec{n}, \vec{s}, \vec{f}\} \tag{4.4}$$

$\mathscr{Q}$ is made of p quotient nodes and q quotient edges. $\mathscr{C}$ represents the set of clusters or quotient nodes and $\mathscr{F}$ represents the set of quotient edges that link the clusters of the EP. The integer vector $\vec{n}$ of length p contains the size of each cluster or quotient nodes, while the integer vector $\vec{s}$ of length p represents the intra-cluster degree of each cluster, that is, the number of edges of a node with all the others within the same cluster (which is a shared number for all nodes in the cluster). The integer vector $\vec{f}$ of length $2q$ consists of pairs of quotient edge weights assigned to each quotient edge $(C_i, C_j) \in \mathscr{F}$ as (f_{jk}, f_{kj}) defined as

$$\begin{aligned} f_{jk} &= \sum_{v_a \in C_k} A_{ia},\ v_i \in C_j \\ f_{kj} &= \sum_{v_a \in C_j} A_{ia},\ v_i \in C_k \end{aligned} \tag{4.5}$$

In Fig. 4.2 we show the quotient graph corresponding to the network in Fig. 4.1(a). However, not all quotient graphs are feasible, that is, not all combination

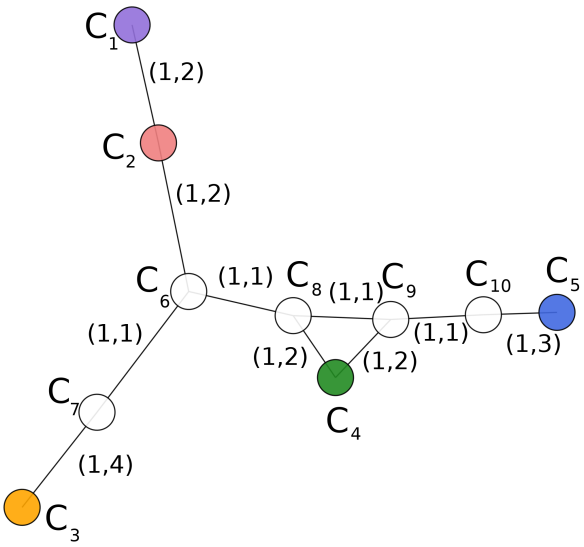


Figure 4.2: **Quotient graph $\mathscr{Q}$ of a network.** Quotient graph $\mathscr{Q}$ corresponding to the network in Fig. 4.1(a). C_i correspond to the different clusters or quotient nodes of $\mathscr{Q}$. Using the components described in Eq.(4.4), $\vec{n} = (4,2,4,2,3,1,1,1,1,1)$, $\vec{s} = (0,0,0,0,2,0,0,0,0,0)$ and the weights of the quotient edges are included in the figure using the notation (f_{jk}, f_{kj}) with $j < k$.

of the components of $\mathscr{Q}$ described in Eq.(4.4) represents some original graph.

The authors of the algorithm for generating symmetric graphs [89] take into account several constraints that must be considered: firstly and in order to satisfy that the number of nodes of each cluster, n_i, can satisfy the connectivity requirements implied by s_i, the following restrictions have to be met:

$$\mathrm{mod}\,(n_i s_i, 2) = 0,\ n_i \geq s_i + 1 \tag{4.6}$$

In addition, the number of edges, m_{jk}, going through two linked clusters, C_j and C_k, must be consistent:

$$n_j f_{jk} = n_k f_{kj} = m_{jk} \tag{4.7}$$

which also imply that there must be enough nodes in cluster C_k to satisfy the demands of each node in cluster C_j and the other way round:

$$n_k \geq f_{jk},\ n_j \geq f_{kj} \tag{4.8}$$

The constraints defined in Eqs.(4.6)-(4.8) gathers the conditions that a quotient graph, $\mathscr{Q}$, must meet to be feasible. In addition, one could construct a representative graph $\mathscr{G}$ from a given quotient graph. This last implication is of particular relevance as the authors suggest a methodology to obtain samples of symmetric graphs that fulfill the requirements of a particular quotient graph. We will briefly present the steps of the algorithm, but we encourage the reader to find all the details in Ref.[89].

The required input consists of the sets $\vec{n}$, $\vec{s}$ and $\mathscr{F}$, together with the number of quotient nodes and quotient edges, p and q, respectively. The resulting graph, $\mathscr{G}$, has $\sum_{i=1}^{p} n_i$ nodes and $\sum_{i=1}^{p} \frac{n_i s_i}{s} + \sum_{(C_j,C_k)} \in \mathscr{F}\, n_i f_{jk}$ edges. They next propose a method to provide a proper choice of the quotient edges weights, without having first made sure that the constraints defined in Eqs.(4.6)-(4.8) are met. They divide the set of $\mathscr{G}$ edges into the intra-cluster and the inter-cluster edge sets and suggest a wiring scheme for the edges, based on mathematical proofs. The equitable partition induced by the created $Aut(\mathscr{G})$ is verified by using software `Nauty`[106].

4.1.2 Detection of symmetries: a dynamic model approach

There are many discrete algebra software that is able to determine the automorphism group, that is, the symmetries, of a graph as well as to extract the orbits that locate the nodes in each cluster. `Saucy3`[2], `GAP`[1] or `Nauty`[106] are some examples. We are however interested in constructing a framework that enables the detection of, not only perfect symmetries, but what we will call *Quasi-Symmetries* (See Section 4.2).

To this end, we present an alternative method to detect the orbits of a network by using the steady state of the Kuramoto-Sakaguchi dynamic model with

homogeneous phase lag, introduced in Chapter 1 and analysed throughout the previous chapters. We recall that this model considers the dynamics of N identical phase oscillators θ_i, for $i = \{1, ..., N\}$, coupled in a network whose evolution is governed by

$$\dot{\theta}_i = \omega + K \sum_j A_{ij} \sin(\theta_j - \theta_i - \alpha), \ j \in \Gamma_i \tag{4.9}$$

Eq.(4.9) corresponds to the Kuramoto-Sakaguchi model (1986) [150], which adds to the original Kuramoto model (1975) [92, 4, 15] a homogeneous phase lag α between nodes that promotes a phase shift between oscillators. Each unit is influenced directly by the set of its nearest neighbours via the adjacency matrix of the network corresponding to the system, $\mathcal{G}(\mathcal{V}, \mathcal{E})$. The coupling strength, $K > 0$, adjusts the intensity of such interactions, Γ_i is the set of neighbours of node i and ω is the natural frequency of each unit, which we consider to be homogeneous among oscillators.

If the system becomes synchronized to a resulting frequency, α forces the system to break the otherwise original fully synchronized state, that is, phase synchronization. However, partial synchronization is conserved for nodes belonging to the same orbit in the network [124, 125]. We hereafter provide a proof of this last statement. To this end, we refer to sections 1.3.2.2, 2.2.1 and 3.1 of the previous chapters for a full explanation of the steps and further clarification of the definitions.

Let us first derive the analytical solution of the phases in the stationary state: if the system reaches the synchronized state and α is small enough, Eq.(4.9) can be linearized and the values of the phases at any time in the stationary state are given by

$$\sum_j L_{ij} \theta_j = \alpha(\langle d \rangle - d_i), \tag{4.10}$$

where d_i is the degree of the ith node and L corresponds to the Laplacian matrix of the network $\mathcal{G}$, in matrix form defined as

$$L \equiv D - A, \tag{4.11}$$

where A is the adjacency matrix of the network and D is the diagonal matrix $[D]_{ij} = d_i \delta_{ij}$ and d_i is the degree of the ith node. Equivalently, $L_{ij} = d_i \delta_{ij} - A_{ij}$. In matrix notation,

$$L\vec{\theta} = \alpha(\langle d \rangle \vec{e}_n - \vec{d}) \tag{4.12}$$

where $[\vec{d}]_i = d_i$. In a connected network, L has one null eigenvalue. Consequently, Eq.(4.12) is singular. Nonetheless, we can solve it by computing the phase difference between each node and a node which we choose as reference. Hence,

$$\phi_i \equiv \theta_i - \theta_R \tag{4.13}$$

where R is the index of the reference node and its corresponding θ_R is left as a free variable. Obviously, $\theta_R = 0$. The new system can be written as

$$\tilde{L}\tilde{\vec{\phi}} = \alpha(\langle d\rangle \vec{e}_{n-1} - \tilde{\vec{d}}) \tag{4.14}$$

where $\tilde{L}$, the so called reduced Laplacian [124, 144], is obtained by removing the Rth row and column of L, although the result does not depend on which row we remove. Similarly, the vector $\tilde{\vec{d}}$ is obtained by removing the Rth element of $\vec{d}$. Finally, the phases with respect to a reference node in the frequency synchronized steady state of the Kuramoto-Sakaguchi model are given by

$$\tilde{\vec{\phi}} = \alpha\tilde{L}^{-1}(\langle d\rangle \vec{e}_{n-1} - \tilde{\vec{d}}) \tag{4.15}$$

We next show that the phases of nodes belonging to the same orbit will be equal at any time.

If P corresponds to the permutation matrix of an automorphism $\sigma \in Aut(\mathcal{G})$, then Eq.(4.2) is true. The Laplacian matrix of the network L also commutes with the permutation matrix, as

$$PL = P(D - A) = PD - PA$$

We already know that $P = P_\sigma$ commutes with A, as $\sigma \in Aut(\mathcal{G})$. P also commutes with D on account of the general statement that any diagonal matrix with equal values for all elements corresponding to the same orbit of the automorphism permutes with the corresponding permutation matrix (See the Appendix section for a detailed proof and Ref.[152] for a generalization of this result). All nodes belonging to the same orbit have the same degree, and hence, D meets the required conditions so as to permute with P. Hence,

$$PL = PD - PA = DP - AP = (D - A)P = LP \tag{4.16}$$

If we left-multiply Eq.(4.12) by P we get

$$PL\vec{\theta} = \alpha(\langle d\rangle P\vec{e}_n - P\vec{d}) = \alpha(\langle d\rangle \vec{e}_n - \vec{d})$$

as symmetric nodes have the same degree ($P\vec{d} = \vec{d}$). In addition, $PL = LP$, as derived in Eq.(4.16). Consequently,

$$LP\vec{\theta} = L\vec{\theta} \tag{4.17}$$

Similarly as done in Eq.(4.14), we define $\tilde{P}$ as P with the removal of the Rth row and column and Eq.(4.17) turns to

$$\tilde{L}\tilde{P}\tilde{\vec{\phi}} = \tilde{L}\tilde{\vec{\phi}} \tag{4.18}$$

Now, the inverse of $\tilde{L}$ exists and we can left-multiply Eq.(4.18) by $\tilde{L}^{-1}$, leading to

$$\tilde{P}\tilde{\vec{\phi}} = \tilde{\vec{\phi}} \tag{4.19}$$

Since $\tilde{P}\tilde{\vec{\phi}}$ corresponds to the permutation of the phases of symmetric nodes, Eq.(4.19) implies that the phases of nodes belonging to the same orbit (those permuted within an automorphism) are equal at any time.

The reverse conditional statement is always true with the exception of a very unlikely case. Only when two nodes i and j that have different degrees, i.e., $d_i \neq d_j$, verify this very restrictive condition (see Appendix 4.5.2)

$$\frac{\sum_k [\tilde{L}^{-1}]_{ik}}{\sum_k [\tilde{L}^{-1}]_{jk}} = \frac{\langle d \rangle - d_i}{\langle d \rangle - d_j} \tag{4.20}$$

and, additionally the degrees of both nodes meet the inequality

$$d_i \geq \langle d \rangle \text{ and } d_j \geq \langle d \rangle \text{ or } 0 < d_i \leq \langle d \rangle \text{ and } 0 < d_j \leq \langle d \rangle \tag{4.21}$$

then the two considered nodes can have the same phases despite not belonging to the same orbit.

Nevertheless, we note that the condition expressed in Eq.(4.20) represents a highly unlikely event and hence would require a very fine tuning of the degree sequence of the corresponding (weighted) network. Moreover, from a probabilistic perspective, the probability that a continuos random variable takes a specific value is zero and so is the chance that the quotient of weighted degrees in Eq.(4.20), resulting from a non-linear transformation, takes a particular value. Henceforth we will assume that the bi-conditional stated as 'Nodes that have the same phases $\Longleftrightarrow$ Nodes that belong to the same orbit' is effectively true.

In this section we have proved that the phases at the steady state of the Kuramoto-Sakaguchi model with homogeneous natural frequencies and phase lag parameters capture the clusters of nodes corresponding to the orbits of the network. Therefore, a straightforward method to detect the orbits of a network is computing the phases analytically as in Eq.(4.15) and classify nodes into clusters according to their values. Nodes with equal values of ϕ belong to the same orbit.

As α behaves as a scaling factor in Eq.(4.15) one could always normalize the results such that $\phi_i \in [0, \pi]$. As an example, the values of $\vec{\phi}$, choosing $R = 0$, for the network in Fig. 4.3(a) are

$$\begin{aligned}\vec{\phi} = (&0.0, 0.0, 0.0, 0.11, 0.76, 1.02, 1.02, 1.20, 2.30, 2.72,\\ &2.99, 2.99, 2.72, 2.99, 2.99, 2.88, 3.14, 3.14, 3.14, 3.14)\end{aligned}$$

The corresponding clusters or orbits of the scaled values can be easily identified in the polar plot shown in Fig. 4.3(b). Notice that the obtained groups are the same as the orbits coloured in Fig. 4.1(a), as expected.

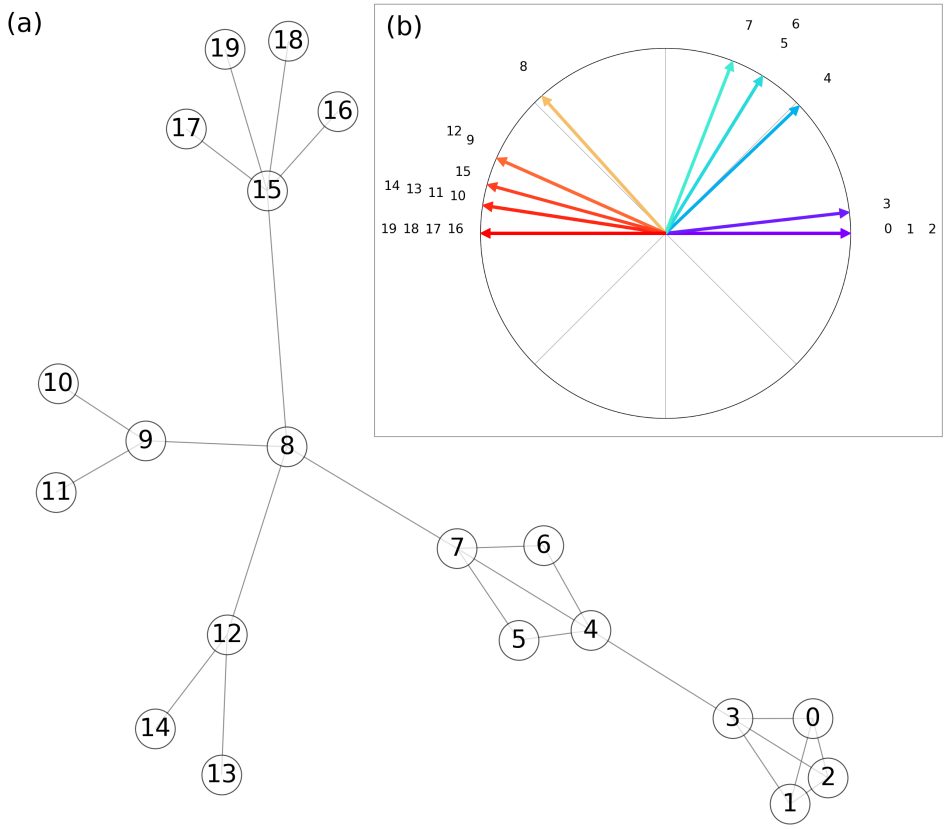


Figure 4.3: **Polar plot of the scaled phases corresponding to the orbits of a network.** (a) Labels' choice for the network topology of 20 nodes further examined in Figs 4.1 and 4.2 and (b) its corresponding polar plot of the $\{\phi_i\}$ scaled phases (in the range $[0,\pi]$) obtained according to Eq.(4.15). We note that the obtained number of groups are the same as the orbits of the network.

4.2 Quasi-Symmetries in complex networks

The concept of approximate symmetry is not new. Approximate symmetry detection for 3D geometry [109] or approximate symmetry methods for solving differential equations [132] are some examples. We address the question of what do we understand by approximate symmetries, or what we call *Quasi-Symmetries*, in complex networks and how do they emerge. For that purpose, we will establish a simile with a circle, a geometric shape consisting of all points in a plane that are a constant distance, the radius, from the center. The circle is highly symmetric as every line that passes through the center generates a reflection and every angle represents a rotational symmetry around the center. However, one could obtain slightly different shapes if the points are obtained experimentally. Despite the underlying true shape being a circle, owing to missing data or experimental errors, the derived shape may lead to a deformed circle or *quasi-symmetric* circle. Similarly, besides synthetic regular networks, real-world networks represent samples of processes that generate them and they are gathered by data collecting methods, either computational or experimentally. Ultimately, researchers deal with networks with missing or additional edges or nodes, as well as with noisy weighted networks. Hence, despite a group of nodes being structurally indistinguishable up to an error, that is, belonging to the same orbit, they may remain as separate independent units by applying traditional symmetry detection methods.

As defined in Section 4.1, the extent of symmetry of a symmetric graph can be measured by the number of possible symmetric permutations of its group of automorphisms [55]. We are concerned, however, by symmetry as a node-wise attribute. *Symmetry,* as a mathematical concept, is a binary attribute of a node with respect to another, either true or false, depending on whether they belong to the same orbit or not. We however introduce the concept of *Quasi-Symmetry* as a continuous variable that characterizes the degree of structural similarity of a pair of nodes. Obviously, a pair or a group of nodes that belong to the same orbit will be perfectly symmetric and therefore, have the largest possible value of quasi-symmetry. This new attribute enables us to characterize the degree of symmetry of all pair of nodes and provides richer information of the network. Notice that the concept of *quasi-symmetry* can be applied not only to networks which have been perturbed, but also to networks of which we want to obtain the degree of symmetry between its nodes, even if we know, beforehand, that they do not belong to the same orbit.

Other authors have defined the notion of 'near symmetry', a more restrictive definition of approximate symmetry, present in complex networks when certain properties remain invariant under some other network transformation, for example, node degree. Accordingly, notions of 'stochastic symmetry' have

also been established in order to characterize near symmetries in real networks [69, 159].

4.2.1 Building synthetic networks with Quasi-Symmetries

Real-world networks, both weighted and unweighted, are potential quasi-symmetric networks. In order to provide a general framework, we need to work with synthetic samples. As exposed in Section 4.1.1, random networks hardly present symmetric patterns and the latter are difficult to control. For this reason, we use the algorithm presented in Section 4.1.1 in order to generate networks with any desired symmetry pattern. This networks are considered to be the underlying perfectly symmetric networks. On top of them, we build the quasi-symmetric networks by either swapping a given number of edges randomly or by modifying the weight of its edges. These mechanisms can be applied in very different ways. We present two particular implementations that can be applied in order to perturb the original networks. The first class of synthetic (unweighted) quasi-symmetric networks is constructed by swapping a random pair of edges, $\{(x,y),(u,v)\}$, that become $\{(x,u),(y,v)\}$ such that degree is preserved and the new edges do not already exist (See Fig. 4.4(b) for an example). The second class of (weighed) synthetic quasi-symmetric networks is constructed by adding a uniform random real number $w \in \mathscr{U}(-w_{max}, w_{max})$ to the otherwise binary edge (See Fig. 4.4(c) for an example). The random transformations that result to a negative weight are ignored.

4.2.2 Characterization of Quasi-Symmetries

In Section 4.1.2 we propose an alternative methodology to those based on discrete algebra for detecting the clusters of equivalent nodes or orbits of the network by bundling the nodes that have the same value of ϕ_i computed analytically from Eq.(4.15). Using the same result, we extend the notion of symmetry into that of quasi-symmetry to characterize the degree of structural equivalence of all pair of nodes.

We first compute the steady state phases of the nodes with respect to any reference node (we note that results do not depend on this choice) using Eq.(4.15). The α parameter in Eq.(4.15) acts as a scaling factor and hence one could always re-scale the set of phases such that they fall in the range $[0,\pi]$. In this way, the most distant nodes are separated by, at most, π (See Fig. 4.3(b) as an example) and results are independent of the network size and the number of edges.

Next, the phase difference is computed between all pair of nodes as

$$\Delta\phi_{ij} = |\phi_i - \phi_j| = |\phi_j - \phi_i| \tag{4.22}$$

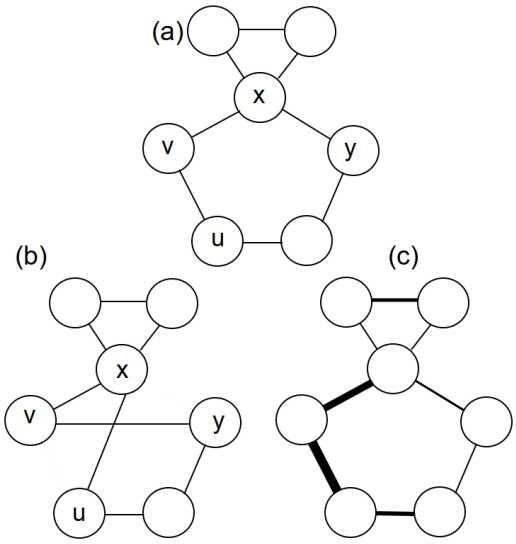


Figure 4.4: **Two possible mechanisms for the generation of quasi-symmetric networks.** Example of a toy network with four orbits or clusters of equivalent nodes in panel (a), and two feasible quasi-symmetric networks drawing from it. Panel (b) shows one possible mechanism leading to the creation of quasi-symmetries by double-edge swapping. Panel (c) exemplifies the perturbation of perfect symmetries by adding random weights to the edges.

Notice that $\Delta\phi_{ij} = 0$ if nodes i and j are completely symmetric.

4.2.2.1 Distribution of Quasi-Symmetries

One could easily count the number of distinct orbits of a network with perfect symmetries either using a discrete algebra software or following the steps described in Section 4.1.2. But besides quantifying perfect symmetries, we may be interested in characterizing the topology of a network, regarding the structural similarity between the nodes, or quasi-symmetries. The first measure that we propose corresponds to the distribution of the scaled phases and phase differences.

In order to obtain more information about a network and distinguish whether it presents more structurally equivalent or similar nodes (quasi-symmetries) than those expected by a random network, we study two baseline types of networks and their distributions of quasi-symmetries: regular networks and random networks.

- Regular Networks

1. Complete Networks K_N: all nodes are structurally equivalent, that is, they belong to the same orbit and, accordingly, they have the same value of ϕ_i. Hence, $\Delta\phi_{ij} = 0$ for all pairs of nodes.
2. Circulant Networks $G(k_1, ..., k_m)$: in a circulant network, each node i is connected to the nodes with indexes $i + k_s$ and $i - k_s$, for all the set of m numbers. Many well-known graph families are subfamilies of the circulant networks. For example if $m = 1$ and $k_1 = 1$, the resulting network is a circular network. The resulting distributions are delta-like, as for a complete network, as all nodes are structurally equivalent.
3. Balanced Tree Networks $G(r, h)$: A tree with a branching factor of r and a height of h has $\sum_{k=0}^{h} r^k = \frac{r^{h+1}-1}{r-1}$ nodes. The number of perfect symmetries or distinct orbits of the network is $h+1$, with a size given by r^k, where k is the current height of the leaf. Therefore, there are $h(h+1)/2+1$ different values of $\Delta\phi_{ij}$, each one having $r^{k_1} r^{k_2}$ repetitions, where k_1 and k_2 are the height of the two leaves which we are considering. The frequency of $\Delta_{ij} = 0$ corresponds to the count of all possible pairs of nodes in the same leaf, i.e., $\sum_{k=1}^{h} r^k(r^k-1)/2 = \frac{r(r^h-1)(r^{h+1}-1)}{2(r^2-1)}$. Figure 4.5 shows the distributions of scaled phases and the corresponding phase differences of a balanced tree network with a height of $h = 3$ and two values of the branching factor, $r \in \{2,3\}$. Notice that there are four distinct values of phases, according to $h+1 = 3+1$, with frequency given by r^k: $\{1,2,4,8\}$ and $\{1,3,9,27\}$, for $r = 2$ and $r = 3$, respectively (see Fig. 4.5(a,b). There are seven distinct values of phase differences, according to $h(h+1)/2+1 = 3(3+1)/2+1$ (see Fig. 4.5(c,d)).

- Random Networks

1. Erdös-Rényi (ER) Network $G(N, p)$: in this model, each of the $\binom{N}{2}$ possible edges is included with probability p, independently from every other edge. Figure 4.6 shows the distributions (relative frequencies) of the scaled phases, $\{\phi_i\}$ and the phase differences between nodes, $\{\Delta\phi_{ij}\}$ for an ER network of 500 nodes and three different values of the p. As the probability of connection approaches 1, the network becomes closer to a complete network and therefore, there are more nodes that are structurally similar. Consequently, the distribution of scaled phases and phase differences is discrete (see the bottom panels in Fig. 4.6(a-b)). Intermediate values of p lead to a continuous distribution of scaled phases which average approaches $\pi/2$ as p in-

creases (see the middle panels in Fig. 4.6(a-b)). The final shape of the distribution is a reflection of the degree distribution of the original network.

2. Barabási-Albert (BA) Network $G(N,m)$: in this model, called preferential attachment or Barabási-Albert network, nodes are added one at a time with m random edges which are linked to the existing nodes with a probability proportional to the degree of them. Figure 4.7 shows the distributions (relative frequencies) of the scaled phases, $\{\phi_i\}$ and the phase differences between nodes, $\{\Delta\phi_{ij}\}$ for a BA network of 500 nodes and three different values of the m. The resulting distributions are very similar to that of ER networks (see Fig. 4.6). Besides small values of m, resulting to star-like patterns, the distribution of phases is continuous. Again, the particular shape of the distributions is determined by the degree distribution of the original network.

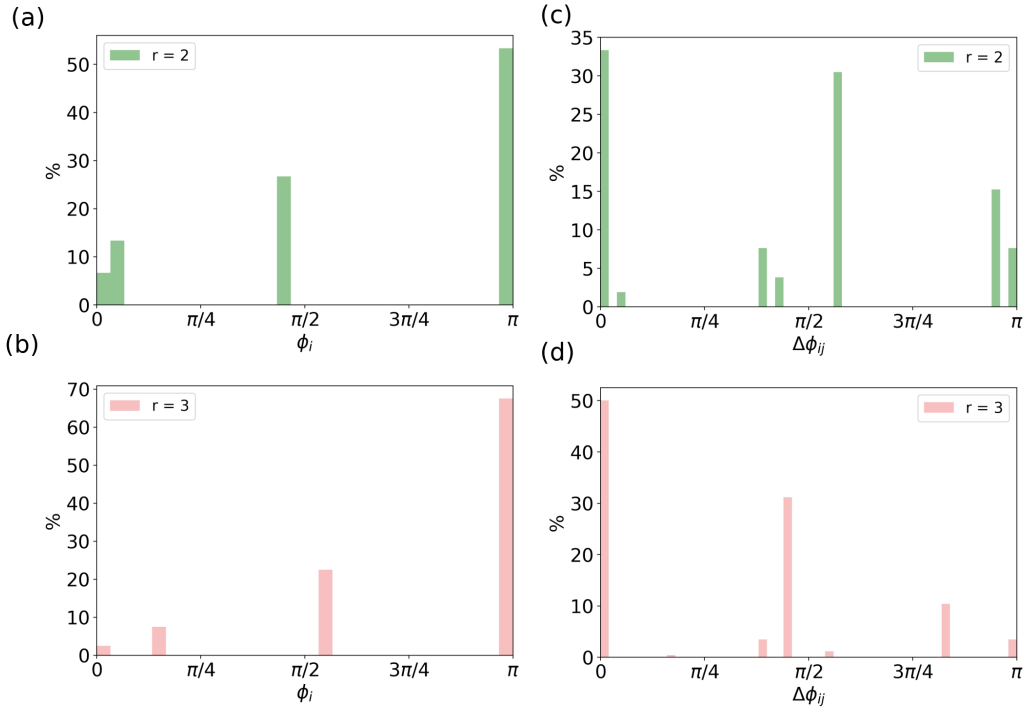


Figure 4.5: **Relative frequency of the scaled phases and phase differences of two balanced trees.** Relative frequency of the scaled phases, ϕ_i, obtained using Eq.(4.15), and phase differences, $\Delta\phi_{ij}$, of a balanced tree network of height, h, equal to 3 and branching factor, r, of 2 [panels (a) and (c)] and 3 [panels (b) and (d)].

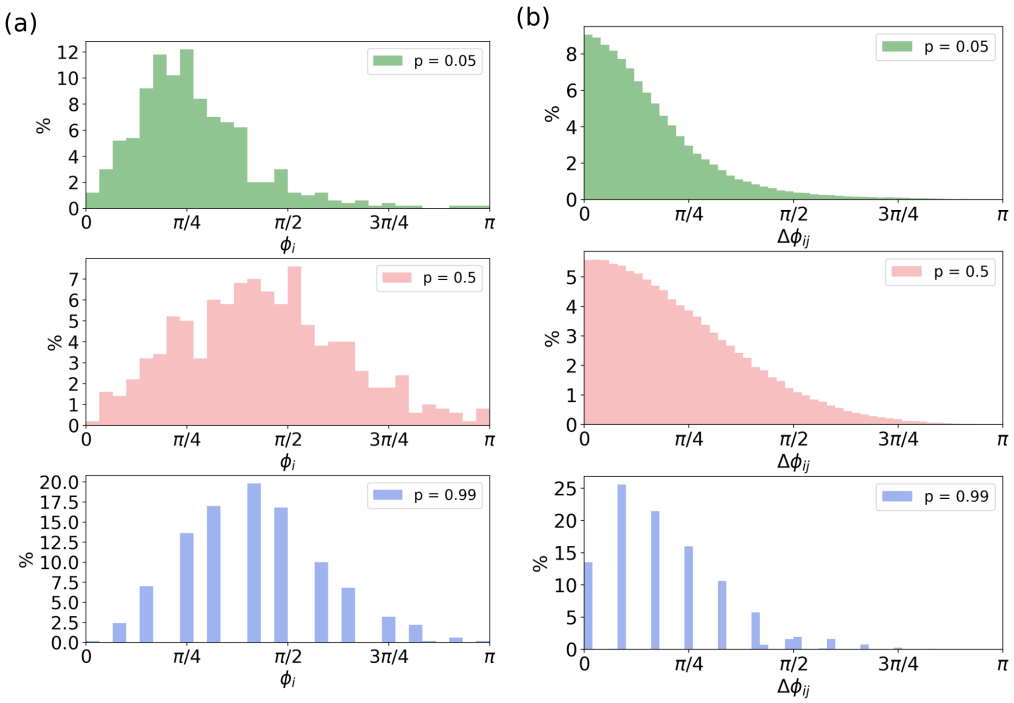


Figure 4.6: **Relative frequency of the scaled phases and phase differences of an Erdös-Rényi random network of 500 nodes and three different densities.** Relative frequency of the scaled phases, ϕ_i, obtained using Eq.(4.15), and phase differences, $\Delta\phi_{ij}$, of an Erdös-Rényi random network of 500 nodes and three different densities: $p = 0.05$, $p = 0.5$ and $p = 0.99$ (upper, middle and lower figures in panels (a) and (b), respectively).

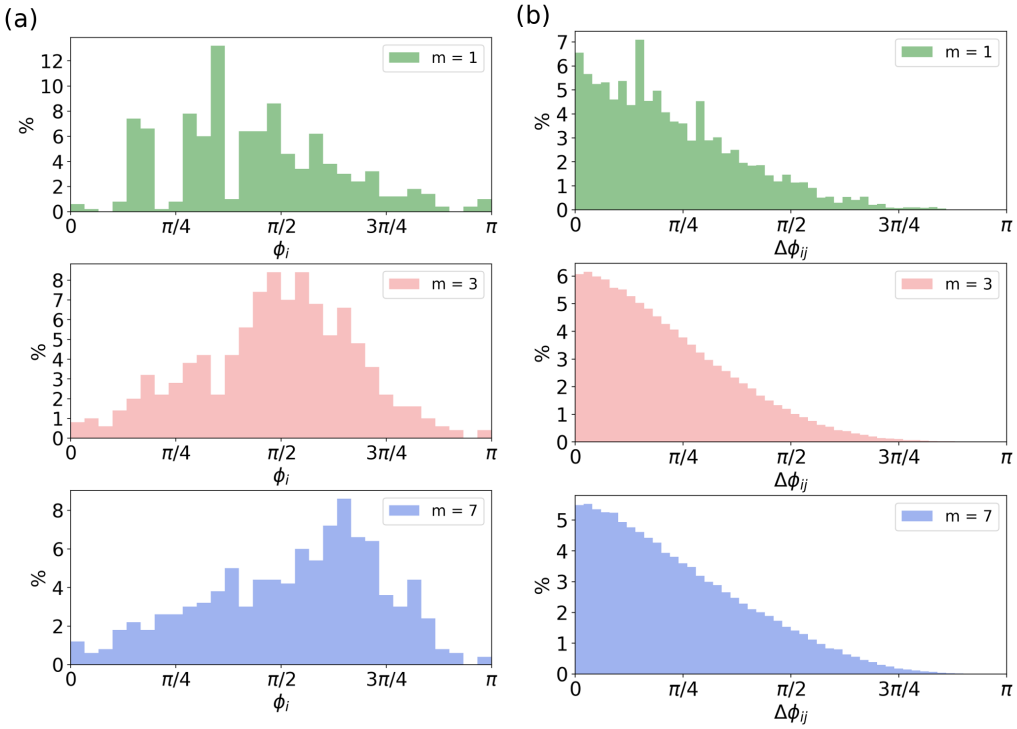


Figure 4.7: **Relative frequency of the scaled phases and phase differences of a Barabási-Albert random network of 500 nodes and three different densities.** Relative frequency of the scaled phases, ϕ_i, obtained using Eq.(4.15), and phase differences, $\Delta\phi_{ij}$, of a Barabási-Albert random network of 500 nodes and three different densities: $m = 1$, $m = 3$ and $m = 7$ (upper, middle and lower figures in panels (a) and (b), respectively).

From the analysis of random networks, namely, ER and BA models, we conclude that, despite both networks have distinct network topologies, i.e., different degree distributions, the level of structural similarity between nodes is very similar. We conclude that random networks display a uni-modal continuous distribution of phases, the shape of which is determined by the corresponding degree distribution. Extreme values of the parameters of the models, i.e., networks with very few connections[1] or very dense networks, conversely, lead to a discrete distribution of phases, resulting from most of the nodes being structurally similar.

- Networks with perturbed (quasi) symmetries

  Once regular and random networks have been analysed in terms of structure similarity, we conduct the equivalent analysis of networks of which we control the number of (perfect) symmetries, generated accordingly to the methodology presented in Section 4.1.1. In order to assess the effect of perturbing the originally perfect symmetries, we add a uniform random noise to each edge (see Section 4.2.1) and study the evolution of similarity, or presence of *quasi-symmetries*, of two networks with 5 and 12 symmetries in the non-perturbed network (originally perfect symmetries). Figs 4.8(a) and 4.9(a) show the relative frequency of the scaled phases, ϕ_i, obtained using Eq.(4.15) of a network of 264 nodes with originally 5 perfect symmetries or orbits and another of 209 nodes and originally 12 perfect symmetries, respectively. Six different perturbed networks are included for each one, by adding a uniform random noise in the range $[-w_{max}, w_{max}]$ to each edge. The upper panels in Figs 4.8(a) and 4.9(a) show that, in the non-perturbed case, i.e., $w_{max} = 0.0$, the distribution of scaled phases results to a discrete one that sets the group of nodes apart according to symmetries, 5 and 12, respectively. As the network becomes more noisy, i.e, the value of w_{max} increases, the symmetries are no longer perfect, but *quasi-symmetries*. In other words, equivalent nodes turn to similar nodes. When the perturbation applied to the networks is too large, the distributions of phases is similar to that of a random network (see lower panels in Figs 4.8(a) and 4.9(a)).

  Thereby, we have shown that networks which structure is enriched with quasi-symmetries, differently than random networks, present a very particular pattern regarding phases distribution, even if perfect symmetries have been removed by adding random noise: they are characterized by a multi-modal distribution of phases, rather than the uni-modal distribution that identify random networks.

[1] All networks are verified to be made of one single component

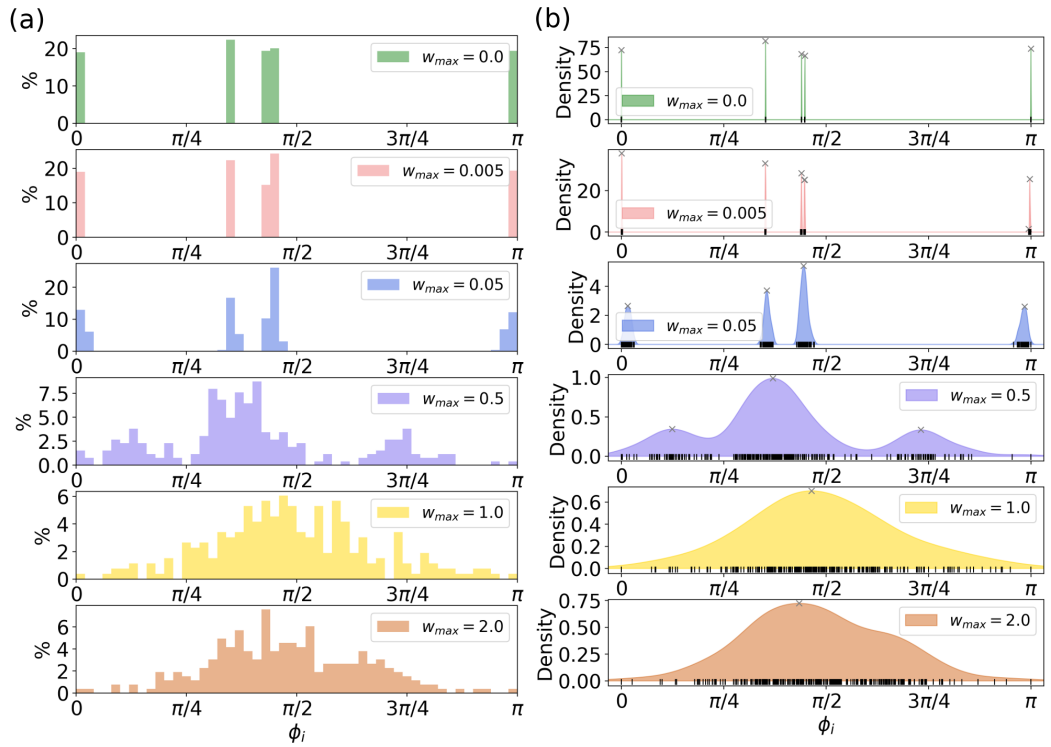


Figure 4.8: **Relative frequency of the scaled phases and the corresponding Gaussian Kernel Density distributions of a network of 264 nodes with originally 5 perfect symmetries or orbits.** Relative frequency of the scaled phases, ϕ_i [panel (a)], obtained using Eq.(4.15) and the corresponding Gaussian Kernel Density distributions with an optimal bandwidth choice according to cross-validation method of a network of 264 nodes with originally 5 perfect symmetries or orbits. A uniform random noise in the range $[-w_{max}, w_{max}]$ is added to each edge, avoiding negative values. Six different values of w_{max} are included, from upper to lower panels.

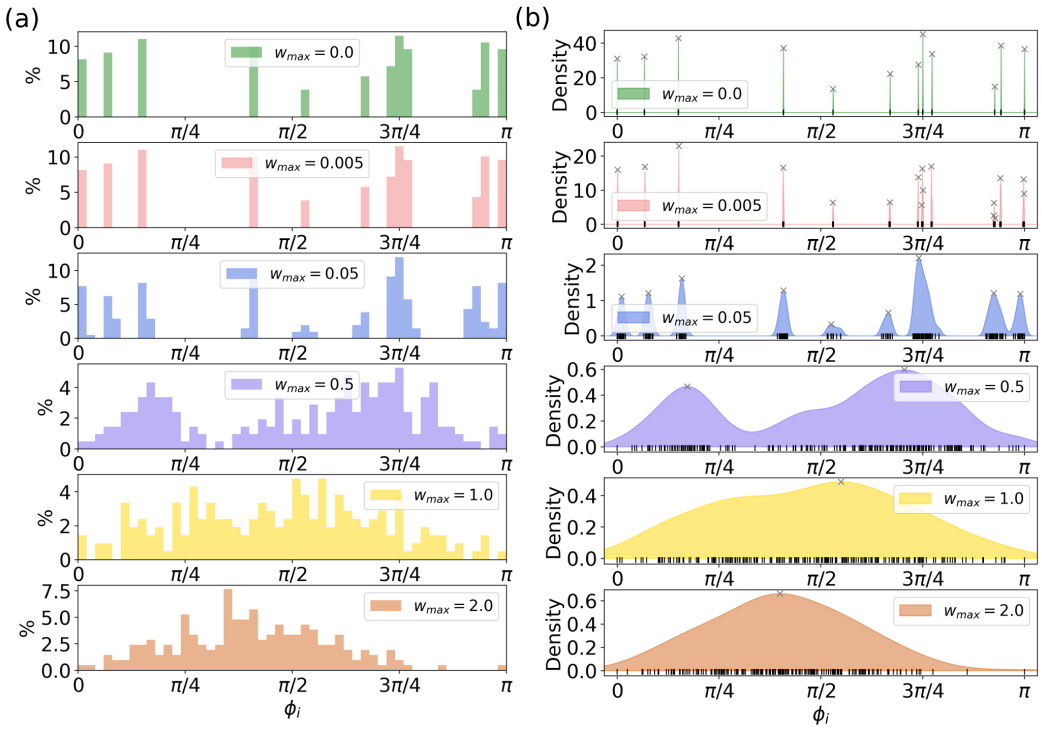


Figure 4.9: **Relative frequency of the scaled phases and the corresponding Gaussian Kernel Density distributions of a network of 209 nodes with originally 12 perfect symmetries or orbits.** Relative frequency of the scaled phases, ϕ_i [panel (a)], obtained using Eq.(4.15) and the corresponding Gaussian Kernel Density distributions with an optimal bandwidth choice according to cross-validation method of a network of 209 nodes with originally 12 perfect symmetries or orbits. A uniform random noise in the range $[-w_{max}, w_{max}]$ is added to each edge, avoiding negative values. Six different values of w_{max} are included, from upper to lower panels.

4.2.2.2 Counting Quasi-Symmetries

In order to assess the quality of the *quasi-symmetries* or structural similarity of real-world networks, we propose a methodology to reject the hypothesis that the network presents a structural similarity equivalent to that of a random network. To do so, we explore the modality of the scaled phases distribution. In other words, we detect the number of modes or peaks of the distribution of scaled phases. When the distribution of $\{\phi_i\}$ is uni-modal we can not say that the topology of the network is different to that of a random network, with respect to *quasi-symmetries* or structural similarity.

The methodology consists in fitting a Gaussian Kernel Density Estimator (KDE) distribution to the scaled phases (see the Appendix 4.33 for more details on KDE). The bandwidth of the Kernel, for each case, is selected using cross-validation, a non-parametric methodology. [79, 45]. The results for the networks of 5 and 12 symmetries are shown in Figs 4.8(b) and 4.9(b), respectively. Notice that the width of the distributions changes with increasing w_{max}, as expected. Once the optimal density is found for each w_{max}, we can detect the peaks for each case. Notice that the perfect symmetries are unequivocally detected (see upper panels in Figs 4.8(b) and 4.9(b)). The distribution becomes more broadened and the number of detected peaks diminishes. Finally, when the networks are completely perturbed, i.e., randomized, the distributions and corresponding number of peaks, or modes, are equivalent to the random networks presented in Figs 4.6 and 4.7, that is, uni-modal distributions.

From the analysis of the symmetric networks we can draw several conclusions, which will be applied to real-world networks: firstly and importantly, random networks present a uni-modal distribution of scaled phases. Secondly, networks that are made of groups of structurally similar nodes, present multi-modal distribution of scaled phases. Narrower peaks are a signal of more differentiated groups of nodes.

In Fig. 4.10 we show the KDE distribution, with the optimal choice of bandwidth, and the corresponding peaks, for a whole-cortex macaque structural connectome constructed from a combination of axonal tract-tracing and diffusion-weighted imaging data [157], which displays distinct modes and hence, informs us that the network is more richer than a random topology with regard to *quasi-symmetries*.

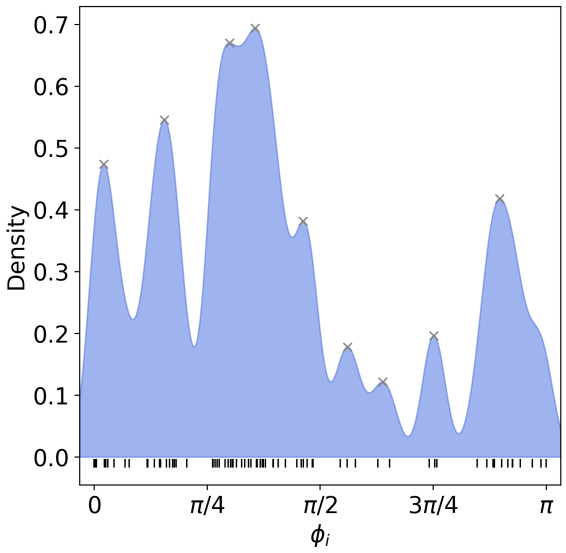


Figure 4.10: **Kernel Density Distribution of a Macaque brain network.** Kernel Density Distribution, using a Gaussian kernel, with an optimal choice of the bandwidth according to 10% cross-validation method of a Macaque brain network, where the nodes above the 95% percentile of phases are removed (outliers).

4.3 The Dual Network

The analysis of the distribution of scaled phases and the corresponding phase differences leaves much room for obtaining more in-depth insights of the structural similarity or *quasi-symmetries* in complex networks. Eq.(4.22) enables us to define the *dual network*, a mathematical object which gathers all the information regarding the *quasi-symmetries* of a network, as we will see.

We define the dual network, $\mathscr{H}(V,\mathscr{E}')$ of $\mathscr{G}(V,\mathscr{E})$, as a complete weighted network that inherits all nodes of the original one and which edges are given a weight according to

$$w_{ij} = \frac{\pi - \Delta\phi_{ij}}{\pi} \tag{4.23}$$

Hence, the weight of the edges is in the range $[0,1]$. An edge connecting two nodes which are completely symmetric has a weight of 1, while an edge connecting the most distant nodes has a weight of 0. Notice that weights are obtained from phase differences applying a linear transformation.

Figure 4.11(a) shows the dual network corresponding to the network in Fig. 4.3(a) with its phases distributed as shown in Fig. 4.3(b). Notice that the nodes which are structurally more similar are more strongly connected, i.e., the edges connecting them have a larger weight, and they are placed very close together

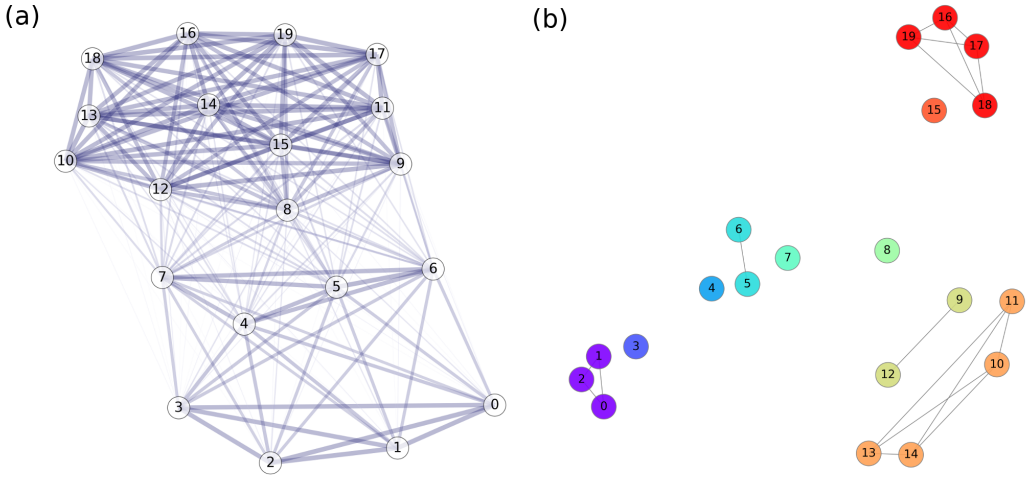


Figure 4.11: **Dual network and corresponding binarized dual network of the network topology defined in Fig. 4.3(a).** (a) Dual network, $\mathscr{H}$, corresponding to the network, $\mathscr{G}$, in Fig. 4.3(a) with its phases distributed as shown in Fig. 4.3(a). Edge width and color intensity scales with its weight, computed using Eq.(4.23). The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering $\mathscr{H}$. (b) Corresponding binarized dual network, $\mathscr{H}_{\mathscr{B}}$. The number of communities (in different colors) is set to 10, the known value of different orbits. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering $\mathscr{G}$.

when using the Fruchterman-Reingold force-directed algorithm [67] for assigning the position of the nodes in the layout of the network. In the network of Fig. 4.3(a), many nodes are part of tree-like motifs of different sizes. This structural similarity is reflected in them being tightly connected in the corresponding dual network.

The fact that the dual network, it is worth saying, is a complex network, entails that many measures developed in the field of network theory can be also applied to this particular network, unveiling interesting properties of the original one.

Before exploring the most informative measures on the dual network, we define the binarized dual network, $\mathscr{H}_{\mathscr{B}}$, as the network with the adjacency matrix given by

$$a_{ij} = \begin{cases} 1 & \text{if } w_{ij} \geq w_{threshold} \\ 0 & \text{otherwise} \end{cases} \tag{4.24}$$

In other words, a threshold value for the weight determines the sparsity of the binarized dual network. $\mathscr{H}_{\mathscr{B}}$ leads to more significant results of network measures, as explained in Section 4.3.1. Different values of $w_{threshold}$ enhance or weaken the presence of *quasi-symmetries*, ranging from a complete network to a completely disconnected one. Our approach consists in choosing a threshold such that the corresponding number of detected communities in $\mathscr{H}_{\mathscr{B}}$ is the same as the number of peaks obtained in the Gaussian Kernel density. Note that several values may verify the latter requirement, a fact that captures the probabilistic nature of a network with *quasi-symmetries*. As long as the main edges are conserved, one could obtain the same number of communities with slightly different connectivity patterns. Figure 4.11(b) shows the binarized dual network, $\mathscr{H}_{\mathscr{B}}$, corresponding to the network in Fig. 4.3(a). The number of communities, in this case, perfect symmetries, is 10, and $w_{threshold}$ is chosen to meet this requirement. Notice that only nodes that belong to the same orbit are connected by an edge, while all nodes remain connected in the original definition of (weighted) dual network (see Fig. 4.11(a)). Note that different values of $w_{threshold}$ may lead to the same number of communities. This behaviour becomes more clear when dealing with larger networks. If we consider the networks with originally 5 and 12 symmetries when we apply a perturbation on their edges with $w_{max} = 0.05$, the number of detected peaks is 4 and 9, respectively (see Figs 4.8(b) and 4.9(b)). Using Eq.(4.24), we obtain the corresponding $\mathscr{H}_{\mathscr{B}}$, by setting the number of communities to the number of detected peaks. A range of $w_{threshold}$ values leads to feasible networks and one can choose between more sparse networks (higher values of $w_{threshold}$) or more densely connected (lower values of $w_{threshold}$) realizations of $\mathscr{H}_{\mathscr{B}}$. Figure 4.12 shows two feasible $\mathscr{H}_{\mathscr{B}}$ for the network with originally 5 symmetries after apply-

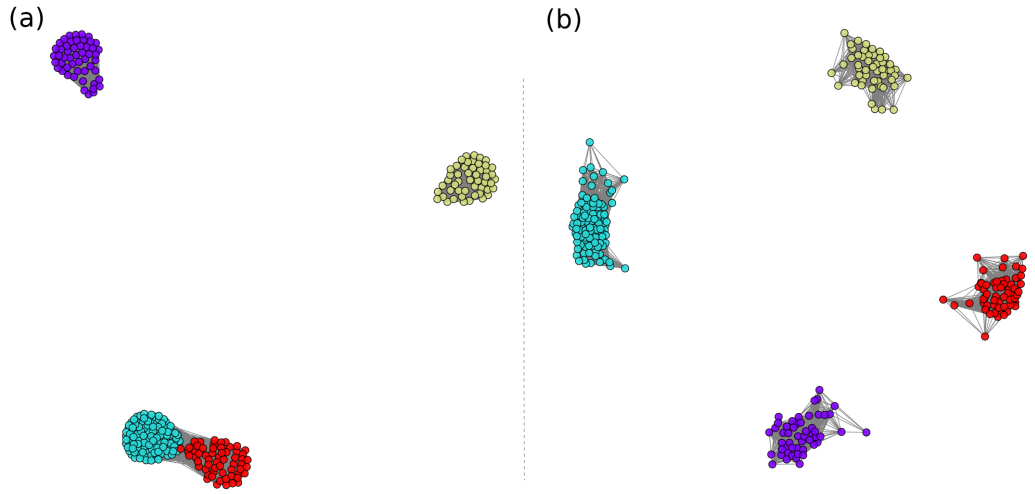


Figure 4.12: **Binarized network for the network with originally 5 symmetries considering two different weight thresholds.** $\mathscr{H}_{\mathscr{B}}$ for the network with originally 5 symmetries when we apply a perturbation with $w_{max} = 0.05$, corresponding to 4 peaks in the distribution of scaled phases (see Section 4.2.1). The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering $\mathscr{H}_{\mathscr{B}}$. The threshold values for the weight are $w_{threshold} = 0.92$ and $w_{threshold} = 0.99$, in panel (a) and (b), respectively.

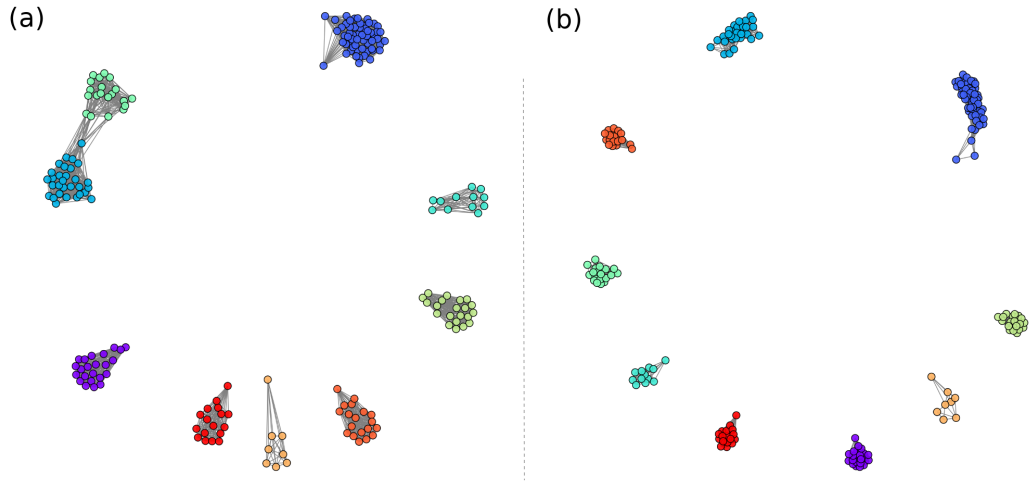


Figure 4.13: **Binarized network for the network with originally** 12 **symmetries considering two different weight thresholds.** $\mathscr{H}_{\mathscr{B}}$ for the network with originally 12 symmetries when we apply a perturbation with $w_{max} = 0.05$, corresponding to 9 peaks in the distribution of scaled phases (see Section 4.2.1). The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering $\mathscr{H}_{\mathscr{B}}$. The threshold values for the weight are $w_{threshold} = 0.96$ and $w_{threshold} = 0.98$, in panel (a) and (b), respectively.

ing a perturbation with $w_{max} = 0.05$ (see Section 4.2.1), corresponding to 4 detected peaks in the distribution of scaled phases. Notice that when $w_{threshold}$ is larger, the dual network becomes more sparse (Fig. 4.12(b)). However, the number of communities is preserved, as a requirement for the creation of $\mathscr{H}_{\mathscr{B}}$. The same applies in Fig. 4.13, for the network with originally 12 symmetries when we apply a perturbation with $w_{max} = 0.05$, corresponding to 9 detected peaks.

4.3.1 Centrality measures

The characterization of the nodes in a network includes the study of its centrality, a measure of its importance in the network, based on the application-specific context we are interested in. On this basis, centrality measures are classified into two types, depending on whether local information around the particular nodes or global information of the network is required.

At the beginning of this Section we have introduced the definition of the dual network, $\mathscr{H}$, and the corresponding binarized network, $\mathscr{H}_{\mathscr{B}}$, which represents a mapping of the structural similarity between nodes of the original networks, namely, its quasi-symmetries. In fact, the dual network is the more complete measure of the structural similarity or equivalence between nodes, relying on the distribution of quasi-symmetries, as exposed in section 4.2.2.1. Nonetheless, we may be interested in computing node-specific measures that inform us about the role that a particular node plays regarding the structural similarity of a network. To this end, we explore some well-known centrality measures on top of the dual network to obtain new insights about the nodes that are the most relevant concerning structural similarity. Although many centrality measures have been proposed, we focus on providing an analysis of one local and one global centrality measures, namely, degree and betweenness centralities, respectively.

We provide an example of the degree centrality values in $\mathscr{H}$ for the network defined in Fig. 4.3(a). The radius of the nodes in Fig. 4.14 is proportional to the values of the degree centrality of the dual network, and the color map is built such that darker values correspond to larger values of this centrality. Nodes that have the largest values are those that are more structurally similar to all others, while nodes with the smallest values are those whose position is more rare or unique. Table 4.1 shows the results of the degree centrality of the dual network corresponding to the network in Fig. 4.3(a) sorted in ascending order. The nodes that display a largest value of degree centrality in the dual network are the nodes 9 and 12, while those displaying the smallest values correspond to nodes 0, 1 and 2.

In Fig. 4.18 we show the results of degree centrality on $\mathscr{H}_{\mathscr{B}}$ for the Macaque brain network, which KDE distribution is presented in Fig. 4.10. Although we are not looking for the interpretation of the obtained results, as it is not our field

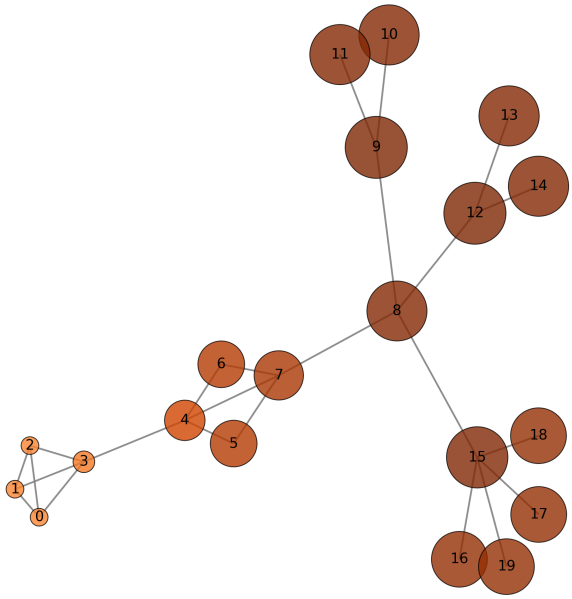


Figure 4.14: **Degree centrality corresponding to the dual network of the network defined in Fig. 4.3(a).** Degree centrality of the nodes in the dual network corresponding to the network presented in Fig. 4.3(a). Larger radius and darker colors correspond to higher centrality values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering the original network.

Node ID	Degree Centrality
0,1,2	0.342
3	0.369
4	0.499
5,6	0.542
7	0.560
16, 17, 18, 19	0.605
8	0.634
10, 11, 13, 14	0.636
15	0.643
9, 12	0.648

Table 4.1: **Degree centrality ranking of the nodes in Fig.4.3(a).** Values of the degree centrality of the dual network corresponding to the network in Fig. 4.3(a) sorted in ascending order.

of expertise, we highlight the fact that brain regions that display a larger value of degree centrality account for a larger similarity with many other nodes in the network (here we find dorsolatral premotor cortex, prefrontal polar cortex, superior parietal cortex, posterior insula and orbitolateral prefrontal cortex as the most central ones), while small values of degree centrality means that their role in the network is more unique, in the sense that no other nodes can play a similar structural role (here we find amygdala, inferior temporal cortex, primary visual cortex, anterior visual area, ventral temporal cortex and hippocampus as the less central ones).

Figure 4.15 shows for the Macaque brain network, the relation between the original network and its corresponding dual network regarding degree and eigenvector centralities, respectively. Note that the most central nodes of the original network and its dual counterpart are not the same. Therefore, the dual network provides new insights about the structure of the original one: nodes that play a role of being structurally similar to many others need not have specific requirements concerning its degree. Regarding eigenvector centrality, we can observe a non-linear tendency between both networks. The interpretation of the highest scores of eigenvector centrality in the dual network is the following: nodes that are, not only structurally similar to many other nodes, but whose neighbours are so (and the neighbours of their neighbours, and so on). Conversely, the nodes with low values of eigenvector centrality are those which are unique and which neighbours so are. Note that the values of the centralities in both the original and the dual network are positively correlated up to a threshold, from which a slightly negative correlation appears.

Despite the ranking of node importance obtained from centralities provides us with the more relevant information about structural similarity, the distribution of scores is rather homogeneous. In order to obtain a more clear pattern, we suggest using the binarized dual network, $\mathscr{H}_{\mathscr{B}}$, in order to compute network measures, such as centralities or community detection, because we get rid of non-significant low-weighted edges and the network becomes more sparse, a characteristic which leads to a better separation of the roles of nodes. On this basis, Figs 4.19 and 4.20 show the results of degree and eigenvector centralities using $\mathscr{H}_{\mathscr{B}}$, the binarized dual network. In this case, the ranking of nodes is similar to that of the weighted dual network but differences between nodes are more emphasized.

The brain regions with lower eigenvector centrality on the dual network are the amygdala, inferior temporal cortex, inferior temporal cortex, primary visual cortex, anterior visual area, ventral part and hippocampus. The regions with the highest eigenvector centrality score are the prefrontal polar cortex, primary auditory cortex, posterior cingulate cortex, posterior insula, orbitolateral prefrontal cortex and superior parietal cortex.

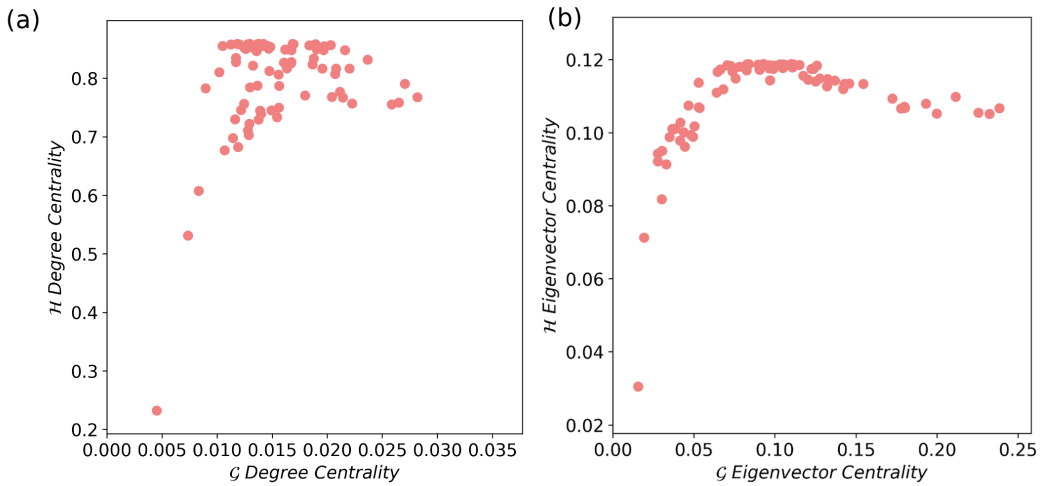


Figure 4.15: **Degree and eigenvector centralities computed for the nodes of the original and the corresponding dual network for eh Macaque brain network.** Scatter plot of the degree and eigenvector centralities, in panel (a) and (b), respectively, obtained for both the original and its corresponding dual network, $\mathscr{G}$ and $\mathscr{H}$, respectively, for the brain network of a Macaque.

4.3.2 Quasi-Symmetric communities

The classification of nodes into perfectly symmetric clusters or orbits has been already addressed and solved in the field of discrete algebra. In Section 4.1.2 we suggest an alternative approach to obtain these same results based on a dynamic model. Nonetheless, we are interested in classifying nodes into different communities based on the structural similarity, and not perfect equivalence, between them. This problem is a particular case of the more general topic of unsupervised classification algorithms, where no correctly classified samples are provided. In other words, we do not know the number of groups and the characteristics of the nodes belonging to each group. However and differently to classical classification problems, our main point is relying on the detection of the number of peaks of the Gaussian Kernel Estimator distribution fitted on the scaled phases (see Section 4.2.2.2). For large enough networks (those which considering a distribution makes sense), the number of peaks will be considered as the number of expected communities that the community detection algorithm should obtain. Hence, only the classification of nodes in each community is missing. To address this question, several approaches are proposed, although we do not reject alternative methodologies that may come up.

- Distance based approach: in section 4.2.2.1 we explore the distribution of phases by fitting a Kernel density distribution in order to decide whether

the structural similarity of a network has a richer pattern than a random network. Once the optimal bandwidth of the Gaussian kernel is numerically computed, we can easily count the number of peaks of the distribution (see, for example, Fig. 4.9(b)). Our method consists in considering this last value as the number of expected communities in the corresponding parameter of an unsupervised clustering algorithm, for example, k-means clustering or hierarchical clustering (following Reference [124]), and obtain the optimal partition into communities. Note that the algorithm delivers the cluster to which each node belongs to, but not in a network-like structure.

Figure 4.16(a,c) shows the dual network corresponding to the networks with originally 5 symmetries with no perturbation and $w_{max} = 0.05$, respectively. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering $\mathscr{H}$. Colors represent the distinct clusters obtained using k-means algorithm with the number of clusters given by the number of peaks of the Kernel density distribution, i.e., 5. (see the upper panel in Fig. 4.8(b)). Similarly, Fig. 4.17(a,c) shows the obtained communities for the networks with 12 perfect symmetries and after applying a random noise with $w_{max} = 0.05$, respectively. In order to verify whether all nodes are correctly classified into the different clusters (a number which is given by the detected peaks of the KDE distributions), we plot the obtained scaled phases of each node and its corresponding membership into the different communities. For the case of no perturbation, single points are expected, as nodes belonging to the same cluster collapse into a single phase value (see Figs 4.16(b) and 4.17(b)). For perturbed networks, a dispersion of phases around different centroids is expected (see Figs 4.16(d) and 4.17(d)).

Alternatively to k-means clustering algorithm we could apply hierarchical clustering in order to split the nodes into communities and obtain equivalent results.

The classification of items into clusters obtained by applying unsupervised algorithms, such as k-means or hierarchical clustering, depend on the choice of the number of clusters. The problem of the optimal choice has been widely studied and several approaches have been proposed in order to select the proper number of clusters or the cut height, for the cases of k-means and hierarchical clustering, respectively. We compare the results of the number of peaks obtained by using the Kernel density distribution approach with that obtained by choosing the optimal number of clusters with the most common method: the elbow curve [88]. The obtained optimal number of clusters does not coincide with our approach as, when

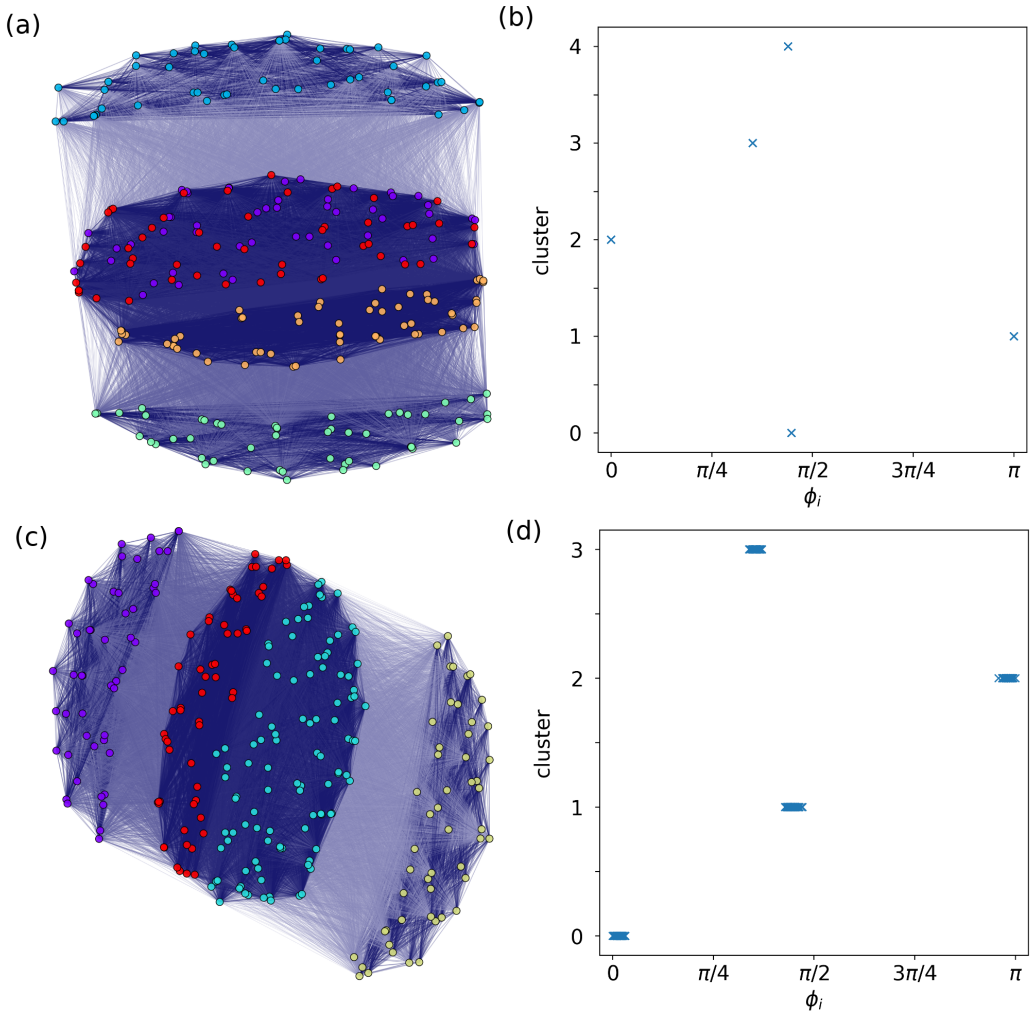


Figure 4.16: **Clusters delivered by k-means algorithm of the dual network of the network with originally 5 symmetries and the corresponding scatter plot of the scaled phases versus the cluster ID.** (a) Dual network corresponding to the network (264 nodes) with originally 5 symmetries and no perturbation applied ($w_{max} = 0.0$). Colors represent the distinct clusters obtained using k-means algorithm with the number of clusters given by the number of peaks of the Kernel density distribution, i.e., 5. (see the upper panel in Fig. 4.8(b)). (b) Corresponding scatter plot of the scaled phases versus the cluster ID. Nodes are correctly classified, as the number of distinct groups is 5, as expected for the case of a network with 5 symmetries and no perturbation applied. (c) Dual network obtained after applying a perturbation of $w_{max} = 0.05$ to the original network. The number of clusters is 4 (see the third panel Fig. 4.8(b)). (d) Corresponding scatter plot of the scaled phases versus the cluster ID. Notice that phases corresponding to nodes that belong to the same cluster have a dispersion different than zero.

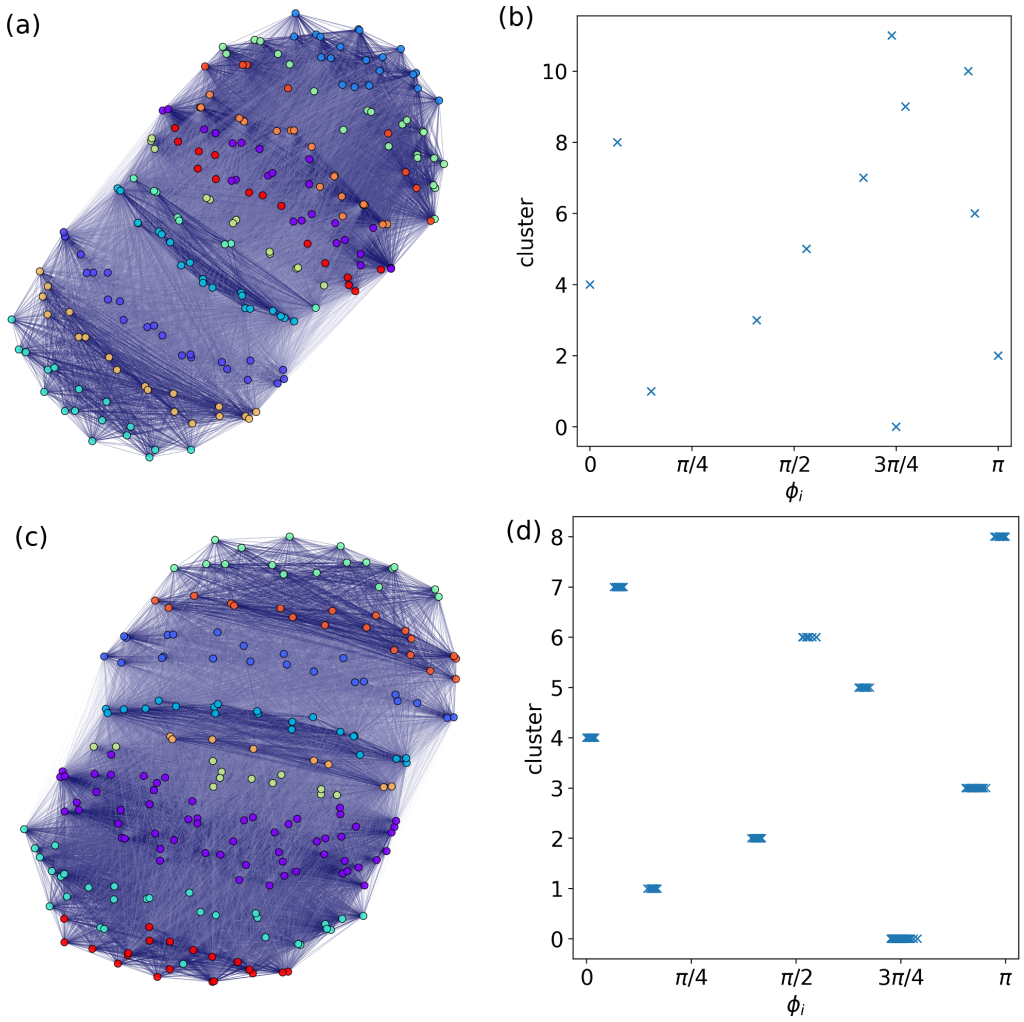


Figure 4.17: **Clusters delivered by k-means algorithm of the dual network of the network with originally 12 symmetries and the corresponding scatter plot of the scaled phases versus the cluster ID.** (a) Dual network corresponding to the network (209 nodes) with originally 12 symmetries and no perturbation applied ($w_{max} = 0.0$). Colors represent the distinct clusters obtained using k-means algorithm with the number of clusters given by the number of peaks of the Kernel density distribution, i.e., 12. (see the upper panel in Fig. 4.9(b)). (b) Corresponding scatter plot of the scaled phases versus the cluster ID. Nodes are correctly classified, as the number of distinct groups is 12, as expected for the case of a network with 12 symmetries and no perturbation applied. (c) Dual network obtained after applying a perturbation of $w_{max} = 0.05$ to the original network. The number of clusters is 9 (see the third panel Fig. 4.9(b)). (d) Corresponding scatter plot of the scaled phases versus the cluster ID. Notice that phases corresponding to nodes that belong to the same cluster have a dispersion different than zero.

the network is not perturbed, larger number of clusters are suggested. The Kernel density distribution approach automatically detects the optimal bandwidth and adapts to each distribution, from narrow peaks to broad and diffuse distributions.

- Dual Network approach: using the definition of the binarized dual network, in Eq.(4.24), we choose $w_{threshold}$ such that the number of detected communities using the Louvain algorithm equals the number of detected peaks. In Figs 4.12 and 4.13, an example is provided for the networks with originally 5 and 12 symmetries, respectively.

  One key benefit of using $\mathscr{H}_{\mathscr{B}}$ is working with a sparse network and keeping only significant relations between similar nodes. On this basis, in Fig. 4.21 we present the result of applying community detection on the binarized dual network of the Macaque brain network, with the positions of the layout being determined by the original network. Notice that the left-right hemispheres separation is recovered from the communities and similar regions are gathered in the same *quasi-symmetric* community. Nodes belonging to the same community play a similar role or have a similar structural pattern.

4.4 Discussion

There have been a number of attempts to deal with approximate symmetries in networks. Beyond structural or topological symmetry, one should consider the fact that real-world networks are exposed to fluctuations or errors, as well as mistaken insertions or removals. Understanding network (approximate) symmetry is of great relevance for the analysis of real-world networks, as they have a significant effect on network dynamics and function. In the present chapter, we provide an alternative notion to approximate symmetries, which we call 'Quasi-Symmetries'. Differently from other definitions, quasi-symmetries remain free to impose any invariance of a particular network property and are obtained from an oscillatory dynamical model: the Kuramoto-Sakaguchi model.

A first main contribution is exploring the distributions of structural similarity among all pairs of nodes and finding a benchmark to determine whether a network has a more complex pattern to that of a random network concerning quasi-symmetries: the criteria consists in determining whether the number of quasi-symmetric groups is greater than one. The number of peaks is derived from the Gaussian Kernel Density Estimator (KDE) (a detailed explanation of the KDE is found in the appendix 4.5.3). Despite we have used this approach, we

are open to alternative methodologies to find a more accurate detection of the number of peaks. Moreover, other Kernels may be considered.

Secondly, we define the 'dual network', a weighted network (and its corresponding binarized counterpart) that effectively encloses all the information of quasi-symmetries in the original one. The dual network allows for the analysis of centrality measures and community detection. The first informs us about the nodes that play a unique role in the network and of those that behave similarly to many other nodes. The latter leads to a classification of nodes into quasi-symmetric communities, the natural extension of the automorphism group orbits (structurally symmetric nodes) of a network. The use of the binary dual network, $\mathscr{H}_{\mathscr{B}}$, is advantageous as it leads to more heterogeneous results in the ranking of node importance and it enables a more significant classification of nodes into quasi-symmetric communities. The number of detected peaks in the KDE distribution determines the family of feasible $\mathscr{H}_{\mathscr{B}}$. However, one could suggest other criteria, as well as threshold models, in order to create the binarized counterpart of the dual network.

Finally, we state that in the present chapter we bring out a general framework to deal with approximate symmetries in complex networks. The dual network has been presented as a useful tool to work with quasi-symmetries and a number of applications have been addressed. Nevertheless, there is a lot of room for obtaining other interesting insights. The analysis of network tolerance to attack to the quasi-symmetric structure or the analysis of quasi-symmetries in multilayer networks are some examples.

4.5 Additional Information

4.5.1 Condition on a diagonal matrix so that it commutes with an automorphism permutation matrix

We provide a proof of the condition that a diagonal matrix must meet in order to commute with the permutation matrix, P, corresponding to an automorphism $\sigma \in Aut(\mathscr{G})$. This last statement is needed to proof that the Laplacian matrix of a network also commutes with the permutation matrix, P:

Proof. Let P be a permutation matrix corresponding to an automorphism, $\sigma \in Aut(\mathscr{G})$. Hence,

$$Pe_i = e_{\sigma(i)} \ \forall i \in \{0, ..., n-1\} \tag{4.25}$$

where $\{e_0, ..., e_{n-1}\}$ denotes the standard basis of $\mathbf{R}^n$. For a $n \times n$ matrix M to commute with P we must have $MP = PM$, or equivalently, $P^{-1}MP = M$. If M is

the diagonal matrix

$$M = \begin{pmatrix} m_0 & & \\ & \ddots & \\ & & m_{n-1} \end{pmatrix}$$

for all $i \in \{0, ..., n-1\}$,

$$P^{-1}MPe_i = P^{-1}Me_{\sigma(i)} = P^{-1}m_{\sigma(i)}e_{\sigma(i)} = \\ = m_{\sigma(i)}P^{-1}e_{\sigma(i)} = m_{\sigma(i)}e_i \tag{4.26}$$

where we have used $e_i = P^{-1}e_{\sigma(i)}$ derived from Eq.(4.25) and the fact that $m_{\sigma(i)}$ is a scalar. From this we can write

$$P^{-1}MP = M_\sigma \tag{4.27}$$

where

$$M_\sigma \equiv \begin{pmatrix} m_{\sigma(0)} & & \\ & \ddots & \\ & & m_{\sigma(n-1)} \end{pmatrix}$$

So $M = diag(m_0, ..., m_{n-1})$ commutes with P if and only if $m_{\sigma(i)} = m_i$ for all i. But the condition in Eq.(4.27) holds as long as $m_i = m_j$ for all i, j that belong to the same orbit. □

4.5.2 The bi-conditional proof of the statement 'nodes with equal ϕ belong to the same orbit'

We derive the required conditions for the statement 'Nodes that have the same phases belong to the same orbit' to be true. As we will see, the implication of two nodes having the same phases is, in most cases, that these nodes belong to the same orbit. However, there might be some cases where the equality of phases is achieved by other conditions.

Suppose nodes i and j have the same phase, $\theta_i = \theta_j$ or $\phi_i = \phi_j$. Then, from Eq.(4.14) we can write the corresponding solutions using the reduced Laplacian

$$\phi_i = \alpha \sum_k [\tilde{L}^{-1}]_{ik}(\langle d \rangle - d_i) \text{ and } \phi_j = \alpha \sum_k [\tilde{L}^{-1}]_{jk}(\langle d \rangle - d_j) \tag{4.28}$$

The condition of identical phases leads to the equality

$$\sum_k [\tilde{L}^{-1}]_{ik}(\langle d \rangle - d_i) = \sum_k [\tilde{L}^{-1}]_{jk}(\langle d \rangle - d_j) \Rightarrow (\langle d \rangle - d_i)\sum_k [\tilde{L}^{-1}]_{ik} = (\langle d \rangle - d_j)\sum_k [\tilde{L}^{-1}]_{jk} \tag{4.29}$$

because d_i and d_j do not depend on k.

Condition (4.29) comes from assuming $\phi_i = \phi_j$.

We consider two different cases: the considered nodes having the same degree or different degree.

- $d_i = d_j$

  If nodes i and j have the same degree, from Eq.(4.29) we get $\sum_k [\tilde{L}^{-1}]_{ik} = \sum_k [\tilde{L}^{-1}]_{jk}$. This last equality is only true when i and j belong to the same orbit.

  Therefore, the straightforward implication is that i and j having the same phase implies that i and j belong to the same orbit.

- $d_i \neq d_j$

  This case is more tricky. We can write the relations between the sum along rows of the inverse of the reduced Laplacian as

  $$\frac{\sum_k [\tilde{L}^{-1}]_{ik}}{\sum_k [\tilde{L}^{-1}]_{jk}} = \frac{\langle d \rangle - d_i}{\langle d \rangle - d_j} \tag{4.30}$$

  Making use of the inequality $\sum_k [\tilde{L}^{-1}]_{ik} > 0 \ \forall i$, we get $\frac{\sum_k [\tilde{L}^{-1}]_{ik}}{\sum_k [\tilde{L}^{-1}]_{jk}} \geq 0$. Finally,

  $$\frac{\langle d \rangle - d_i}{\langle d \rangle - d_j} \geq 0 \tag{4.31}$$

  Considering that $d_i > 0 \ \forall i$, the inequality (4.31) has the following solution:

  $$d_i \geq \langle d \rangle \text{ and } d_j \geq \langle d \rangle \text{ or } 0 < d_i \leq \langle d \rangle \text{ and } 0 < d_j \leq \langle d \rangle \tag{4.32}$$

  where we have simplified by considering that the network is large enough.

  From this second case we can conclude that the equality (4.29) can be achieved when $d_i \neq d_j$ only if Eq.(4.30) and Eq.(4.32) are true. These last requirements represent very strong restrictions for a network. Firstly, the fine tuning (only feasible for weighted networks) of the degree sequence implied in Eq.(4.30) is hard to be achieved. Secondly, the inequality (4.32) adds further constrains on the first condition.

To conclude, we can say that the double implication 'Nodes that have the same phases $\Longleftrightarrow$ Nodes that belong to the same orbit' is always true except by the cases where a pair of nodes i and j that have different degrees, $d_i \neq d_j$ meet the conditions expressed in Eqs.(4.30) and (4.32). Note that restriction (4.30) requires that a quotient of degrees takes a specific value, resulting from a non-linear transformation of network parameters, and hence, it is highly unlikely. From a probabilistic perspective, the probability that a continuous random variable takes a specific value is zero.

4.5.3 Kernel Density Estimator

Kernel Density Estimator (KDE) is a non-parametric standard technique in explorative data analysis to estimate the probability density function of a random variable first introduced in References [148] and [133]. From a finite data sample the probability function of the whole population is inferred. The KDE method takes a kernel and a parameter, the bandwidth, that affects the level of smoothness the resulting curve has.

The problem can be posed as 'How can one estimate a probability density function $f(x)$ given a sequence of n independent identically distributed random variables $X_1, ..., X_n$ from this density f? [172] The estimator of f, $\hat{f}_h(x)$ is defined by

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \tag{4.33}$$

where h is the smoothing parameter called the bandwidth and K is the kernel, $K_h(u) = K(u/h)/h$. The Kernel is imposed to be symmetric and non-negative, and K itself being a probability density. Then, $\hat{f}_h(x)$ intuitively places at each observation point X_i a probability mass according to K_h and then averages. Some of the commonly kernels are uniform, triangle,quartic, triweight, Epanechnikov and Gaussian. It turns out that the choice of the bandwidth is much more important for the estimation of the density than the particular shape of the kernel. Small values of h result into an over-fitted density distribution, showing spurious features ot the latter, while to big values of h lead to an estimate which is too biased and may not reveal relevant features of the distribution.

The construction of a kernel density estimate finds and interpretation in thermodynamics, since the Gaussian KDE is the solution of the heat propagation model, i.e., the solution of the estimator is equivalent to the amount of heat generated when heat kernels are placed at each data point locations [36]. Note that the Gaussian kernel density estimator is the unique solution to the diffusion partial differential equation PDE

$$\frac{\partial}{\partial t}\hat{f}(x,t) = \frac{1}{2}\frac{\partial^2}{\partial^2 x}\hat{f}(x,t),\ t > 0 \tag{4.34}$$

with initial condition $\hat{f}(x,0) = 1/N\sum_i \delta(x - X_i)$ is the empirical density of the data and $t = \sqrt{h}$. Eq.(4.34) corresponds to the Fourier heat equation.

We however make a few comments on some drawbacks of the popular Gaussian Kernel density estimator: firstly, it lacks local adaptativity, which often leads to substantial sensitivity to outliers as well as a tendency to flatten the peaks and valleys of the distribution [169]. Secondly, just as most kernel estimators, it suffers from boundary bias, as most kernels do not take into account further information about the domain of the data, i.e., data being nonnegative. Some

of these issues have been addressed by introducing more complex higher-order kernels.

In this work, we pursue to hold the methodology as simple as possible. To this end and because it meets the objectives of the posed problem, we use the Gaussian kernel. Nonetheless, we suggest the reader to consider implementing a kernel based on diffussive processes, in Reference [36], as it solves the mentioned problems of standard kernels estimators.

4.5.4 Whole-cortex Macaque structural connectome: results

In several section we have applied our measures to the whole-cortex macaque structural connectome constructed from a combination of axonal tract-tracing and diffusion-weighted imaging data [157]. We present the corresponding figures of the results concerning centralities and community detection.

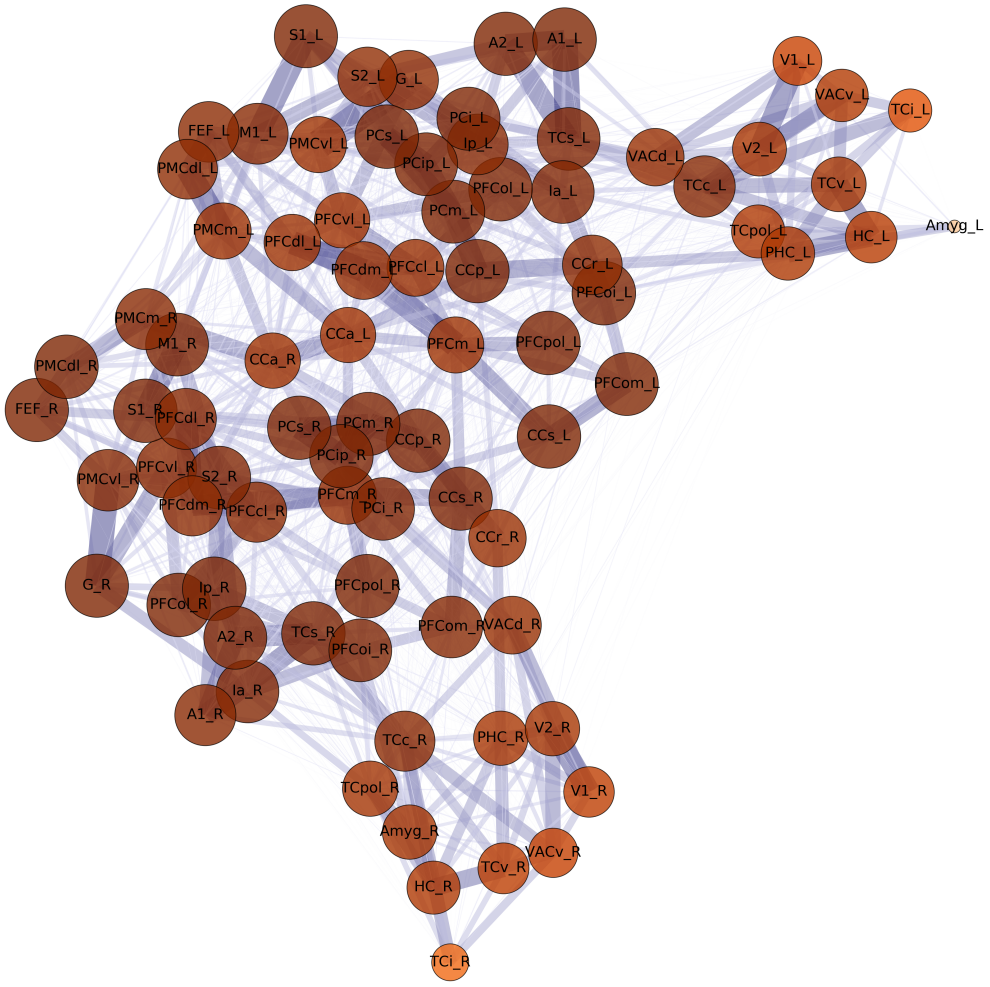


Figure 4.18: **Degree centrality of the nodes in the dual network corresponding to the Macaque brain network.** Degree centrality of the nodes in the dual network corresponding to the Macaque brain network. Larger radius and darker colors correspond to higher centrality values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering the original network.

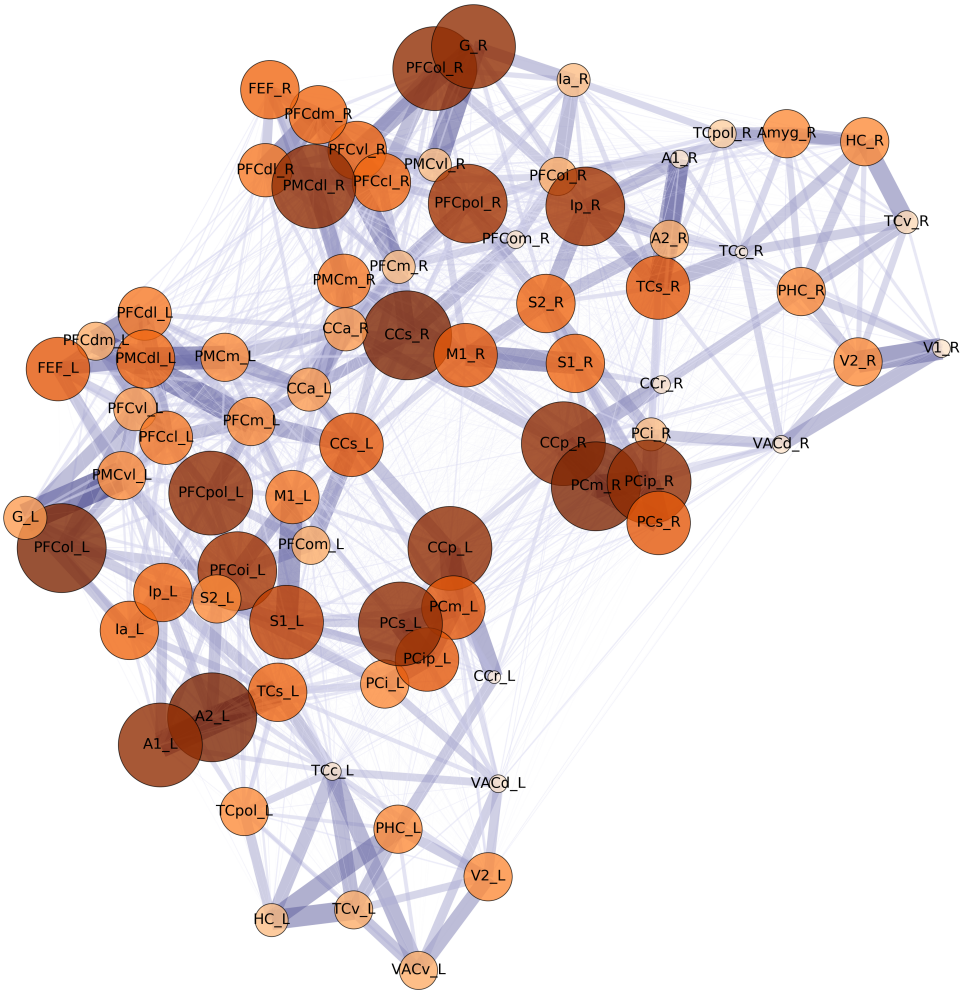


Figure 4.19: **Degree centrality of the nodes in the binarized dual network corresponding to the Macaque brain network.** Degree centrality of the nodes in the $\mathscr{H}_{\mathscr{B}}$ network of the Macaque brain network. Larger radius and darker colors correspond to higher centrality values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering the original network. The nodes above the 95% percentile of phases are removed (outliers).

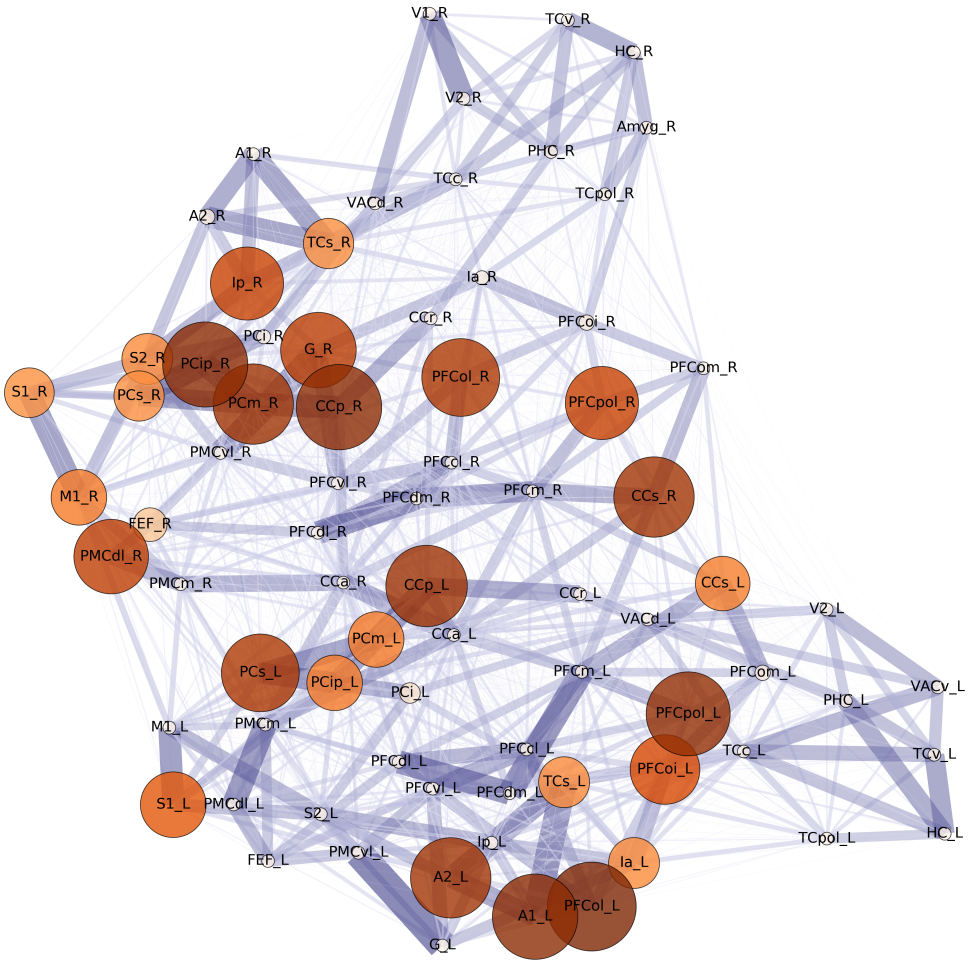


Figure 4.20: **Eigenvector centrality of the nodes in the binarized dual network corresponding to the Macaque brain network.** Eigenvector centrality of the nodes in the $\mathscr{H}_{\mathscr{B}}$ network of the Macaque brain network. Larger radius and darker colors correspond to higher centrality values. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering the original network. The nodes above the 95% percentile of phases are removed (outliers).

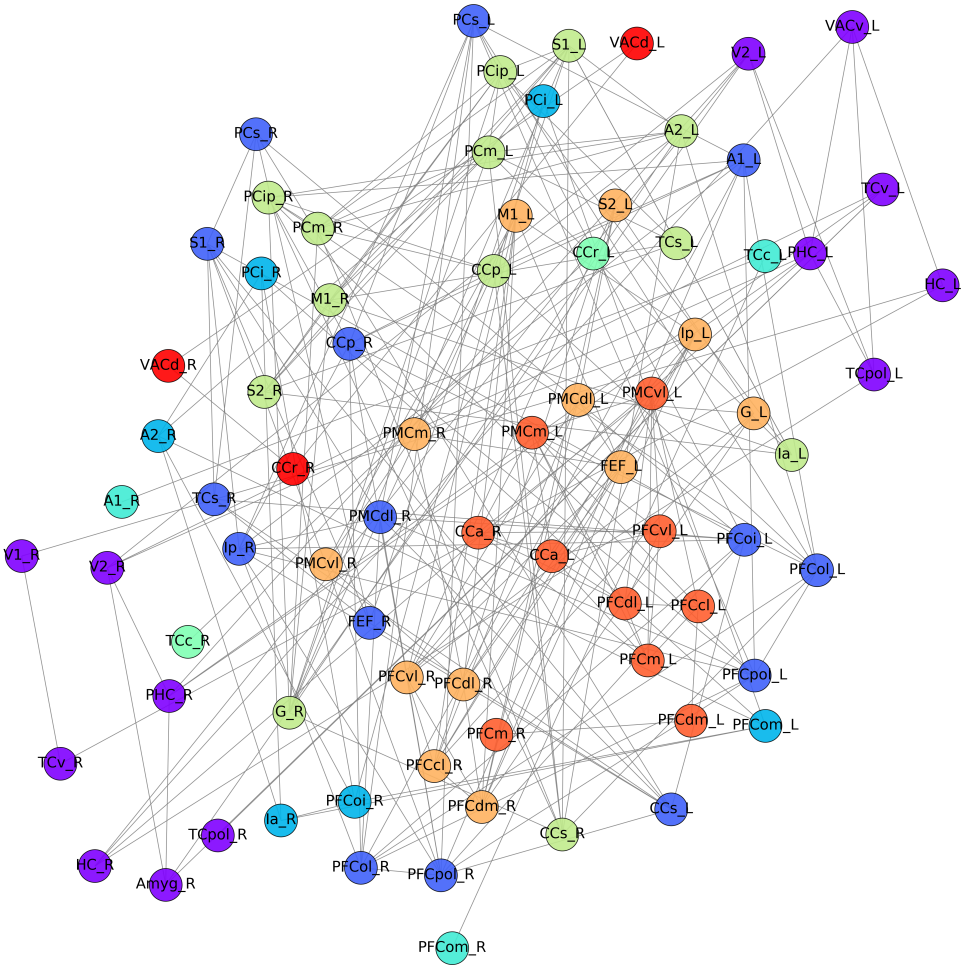


Figure 4.21: **Binarized dual network of the Macaque brain network with colors corresponding to the different communities obtained after applying Louvain community detection algorithm.** $\mathscr{H}_{\mathscr{B}}$ network of the macaque brain network with $w_{threshold} = 0.97$. Colors correspond to different communities. The position of nodes are computed using the Fruchterman-Reingold force-directed algorithm [67] considering the original network. We note that the left-right hemispheres separation is recovered from the communities and similar regions are gathered in the same *quasi-symmetric* community.

CHAPTER 5

Conclusions

Most of the real-world complex systems are best described as complex networks – biological, socio-economic or technological networks, among others. Notably, they exhibit common features which make them be neither purely regular nor purely random. Such of these well-known properties include the small-work effect, an heterogeneous distribution of node degrees and the presence of communities. The identification of the constituents and interactions of such systems as the nodes and edges of the corresponding network, respectively, allows to individually explore the importance and role of each element: centrality measures are the best example of node assessment. Nevertheless, the full description of a network might remain incomplete if its structure is considered detached from its dynamics.

Real networks are representations of evolving systems, which constituent parts and connexions may experience changes over time. Ongoing dynamical processes may be present even if the topology of a network is considered to be static. Such systems are typically characterized by independent dynamical variables on each node that are coupled together only through the edges of the network – acknowledge the transmembrane voltage spiking of individual neurons in a brain or the infectious disease status of an individual in the course of an epidemic spreading. Importantly, non-trivial patterns and phenomena emerge from the interactions existing in such dynamical networked systems.

Many real systems of interest can be mathematically described as oscillatory systems, that is, an ensemble of units that are individually modelled as oscillators of one sort or another, but that they are coupled with the neighbours through the connections of the network. The flashing of fireflies, the neuronal brain signals or the energy flow through the power grid are some examples. Many biological, technological and even socio-economic systems are rightly described as networks of couple phase oscillators. Very often, the dynamics of the oscillators considers that neighbouring nodes are coupled through its phases and regardless of the amplitude of the oscillations. Within this framework, researchers have drawn particular attention to the study of synchronization, that is, the whole set (or a fraction) of network oscillators being locked at the same frequency.

Within this framework, after Arthur Taylor Winfree had made the first attempt to mathematically model the latter non-linear collective dynamics [181]

and following his approach, Yoshiki Kuramoto came up with a tractable mathematical model that could capture the phenomenology of collective synchronization by suggesting that oscillators were coupled by a sinusoidal function of their phase differences [92, 93]. Despite its simplicity, the Kuramoto model is able to capture the phase transition between the purely chaotic state, where all oscillators move independently, to a coherent state, where more and more oscillators reach a frequency synchronized state and eventually end with all units swinging in unison.

Later, Yoshiki Kuramoto together with Hidetsugu Sakaguchi presented a generalization of the previous limit-cycle set of oscillators Kuramoto's model which incorporated a constant phase lag between oscillators [150]. Subsequent studies of the model included the network structure within the model together with the global shift – or frustration parameter – and considered identical oscillators for all nodes, i.e., sharing a common intrinsic frequency. For a wide range of the phase lag values, the system becomes synchronized to a resulting frequency, i.e., the dynamics reaches a stationary state. However, the frustration parameter forces connected nodes to be locked in phase and hence leading the system to break the phase synchronized state. The magnitude of such locking is determined by both the network topology and the parameters of the dynamics. However, full synchronization is conserved for topological symmetric nodes – a phenomenon that has been called remote synchronization [124].

In the original work of Kuramoto and Sakaguchi and in most of the consequent later studies, a uniform distribution of phase lag parameters is customarily assumed. However, just as the connectivity pattern of individual nodes is mostly heterogeneous, other intrinsic properties of nodes – that assuredly represent the constituents of real systems – do not need be identical but distributed non-homogeneously among the population. This thesis contributes to the understanding of the Kuramoto-Sakaguchi model with a generalization for non-homogeneous phase lag parameter distribution. Considering different scenarios concerning the distribution of the frustration parameter among the oscillators represents a major step towards the extension of the original model and provides significant novel insights into the structure and function of the considered network.

The first setting that the present thesis considers consists in perturbing the stationary state of the system by introducing a non-zero phase lag shift into the dynamics of a single node. The aim of this work is to sort the nodes by their potential effect on the whole network when a change on their individual dynamics spreads over the entire oscillatory system by disrupting the otherwise synchronized state. In particular, we define *functionability*, a novel centrality measure that addresses the question of which are the nodes that, when individually perturbed, are best able to move the system away from the fully synchronized state.

This issue may be relevant for the identification of critical nodes that are either beneficial – by enabling access to a broader spectrum of states – or harmful – by destroying the overall synchronization. Hence, depending on the system we are considering, the most functional nodes have to be considered when looking for a potential enhancement of the diversity of attainable states or the inhibition of risky instabilities in the system. In addition, despite functionability is obtained from a dynamical model, in the present thesis we derive the analytical expression of the centrality. It turns out that the corresponding ranking of node centralities is exclusively determined by the network structure. Therefore and importantly, functionability centrality is finally defined as a quantity which does not depend on the values of the model or the parameters of a numerical simulation. Moreover, the analytical expression is a compact and deterministic mathematical function of the network topology and, thus, is not based on optimization procedures.

After comparing functionability with other centrality measures we conclude that, although most centrality measures share certain common patterns, functionability delivers unique information about the network and the importance of its constituent nodes. Importantly, we obtain that the nodes with the largest values of functionability are both locally well connected – have a high degree – but are also far from the central core of the network – that is, they are peripheral nodes. To sum up, functionability enables us to detect the nodes that are most central or relevant for moving the overall system away from synchronization. For example, epileptic attacks or power grid collapses may be derived from single nodes that, even if not located in the main core, change their intrinsic properties and spread asynchrony rapidly to the network, leading to potentially fatal states. It may be helpful to target such nodes in order to control both synchrony and asynchrony.

The second scenario that the present thesis addresses considers a more general configuration in which the phase lag parameter is an intrinsic property of each node, not necessarily zero, and hence exploring the potential heterogeneity of the frustration among oscillators. A very relevant question in oscillatory models is finding the conditions of network synchronization. In this second work, we bring forward a methodology not only to obtain the desired synchronized state, but any convenient phase configuration in the steady state, by means of a fine tuning of the phase lag or frustration parameters. We obtain the analytical solution of frustration parameters so as to achieve any phase configuration, by linearizing the most general model. The three intrinsic parameters of the nodes in the model, natural frequencies, frustration parameters and the phases in the steady state are coupled by an equation that allows to tune them for a desired configuration. A main result is that a given phase configuration can be

access via a continuous spectrum of frustration parameters, i.e, one phase and one frustration parameter are left as free parameters.

We also address the fact that the question 'among all the possible solutions, which is the one that makes the system achieve a particular phase configuration with the minimum required cost?' is of particular relevance when we consider the plausible real nature of the system. If a real system needs to access a particular phase configuration, which may be associated with a singular function, then it will tend to minimize the effort or cost to do so. To this end, once the frustration parameters are tuned so as to obtain a particular state, we define a cost function to assess the overhead that the system requires to achieve such configuration. In this way, among all possible solutions of the tuning, we request those which minimize the cost to obtain them. Moreover, in the present thesis we obtain the analytical solution of the cost function for two particular configurations and we make a connection with the non-linear Kuramoto-Sakaguchi model: despite our analysis being based on the linear version of the model, we show that the proposed phase lag parameters tuning is also able to enhance frequency synchronization.

Finally, the homogenous distribution of phase lag parameters is revisited in the last scenario: the phase state obtained in the stationary state when all nodes share a common frustration parameter is tightly bounded to the topological symmetries of the network. As studied in the literature, a certain degree of symmetry is an attribute of real-world networks. The study of the symmetries of a network is of great relevance for several reasons: it may help us to have a better understanding of the formation of certain real-world networks, they can also provide information about node function, and have an effect on network redundancy and robustness. Moreover, symmetries are known to influence the outcome of network dynamics, such as synchronization or controllability. The notion of 'symmetry' or 'invariance' includes several specifications depending on the field it is applied: a topological transformation of a graph or a network maps each vertex to another one as a permutation, and the set of permutations of a graph that leaves the topology invariant are the automorphisms of the graph. Additionally, the set of vertices can be split into the core of fixed points, that is, vertices which are moved by none of the automorphisms of the graph, and the vertex set of symmetric motifs, corresponding to the different permutations. Each symmetric motif can be further partitioned into clusters, alternatively called orbits, and they represent nodes that are structurally indistinguishable and play the same structural role in the network.

Nevertheless, beyond structural or topological symmetry, one should consider the fact that real-world networks are exposed to fluctuations or errors, as well as mistaken insertions or removals. Hence, despite there are many discrete algebra software that is able to determine the automorphism group, that is, the

symmetries, of a graph as well as to extract the orbits that locate the nodes in each cluster, we are interested in constructing a framework that enables the detection of, not only perfect symmetries, but approximate symmetries. There have been a number of attempts to deal with approximate symmetries in networks. In the present thesis, we provide an alternative notion to approximate symmetries, which we call 'Quasi-Symmetries'. Differently from other definitions, quasi-symmetries remain free to impose any invariance of a particular network property and are obtained from the stationary state of the Kuramoto-Sakaguchi model with an homogeneous phase lag distribution.

A first contribution is exploring the distributions of structural similarity among all pairs of nodes and finding a benchmark to determine whether a network has a more complex pattern than that of a random network concerning quasi-symmetries: the criteria consists in determining whether the number of quasi-symmetric groups is greater than one.

Secondly, we define the 'dual network', a weighted network –and its corresponding binarized counterpart– that effectively encloses all the information of quasi-symmetries in the original one. The dual network allows for the analysis of centrality measures and community detection. The first informs us about the nodes that play a unique role in the network and of those that behave similarly to many other nodes. The latter leads to a classification of nodes into quasi-symmetric communities, the natural extension of the automorphism group orbits (structurally symmetric nodes) of a network. The use of the binary dual network is advantageous as it leads to more heterogeneous results in the ranking of node importance and it enables a more significant classification of nodes into quasi-symmetric communities.

Therefore, in this last work we bring out a general framework to deal with approximate symmetries in complex networks. The dual network is presented as a useful tool to work with quasi-symmetries and a number of applications are addressed.

Bringing all the pieces together, the general conclusion that we can draw from this thesis is that, despite the compact and simple mathematical expression of the original Kuramoto-Sakaguchi oscillatory model, interesting insights are obtained when we consider the versatility of the phase lag parameters distribution. Considering three different scenarios for the distribution of the frustration parameters among the population has lead us to obtain a more extensive understanding of the connection between the dynamics and the structure of complex networks. The definition and study of the functionability centrality or quasi-symmetries are examples of it.

Throughout this thesis, we have sought to obtain the analytic solution of several questions, wherever possible, because of the great benefit of having a more tractable and interpretable outcome of the problems set forth. In addi-

tion, we have illustrated most of the concepts and results by considering simple synthetic networks. In this way, the meaning and required procedures of each work become clearer to the reader. Moreover, special attention has been paid to exemplify the different results to real networks, particularly, real brain networks. However, we are aware that the interpretation by an expert of the field of such outcomes is missing and hence represents a limitation to the work. For this reason, a natural extension of the thesis is the application of the different measures and results to real world networks, together with a meaningful interpretation of the results, which would represent a significant step forward for the considered fields. Furthermore, there is a lot of room for obtaining other interesting results, such as the extension of the obtained results to multilayer or temporal networks.

APPENDIX A

List of Publications

Publications related to the thesis

Articles

- *Functionability in complex networks: Leading nodes for the transition from structural to functional networks through remote asynchronization*
  Gemma Rosell-Tarragó and Albert Díaz-Guilera
  Chaos **20**, 013105 (2020)
  https://doi.org/10.1063/1.5099621

- *Optimal cost tuning of frustration: Achieving desired states in the Kuramoto-Sakaguchi model*
  Gemma Rosell-Tarragó and Albert Díaz-Guilera
  Physical Review E **103**, 012216 (2021)
  https://doi.org/10.1103/physreve.103.012216

- *Quasi-symmetries in complex networks: a dynamical model approach*
  Gemma Rosell-Tarragó and Albert Díaz-Guilera
  Journal of Complex Networks **9**(3), cnab025 (2021)
  https://doi.org/10.1093/comnet/cnab025

Other publications

- *A Complex Network Framework to Model Cognition: Unveiling Correlation Structures from Connectivity*
  Gemma Rosell-Tarragó, Emanuele Cozzo and Albert Díaz-Guilera
  Complexity **2018**, 1918753 (2018). https://doi.org/10.1155/2018/1918753

- *Numerical simulations of energy deposition caused by 50 MeV-50 TeV proton beams in copper and graphite targets*
  Yuancun Nie, Rüdiger Schmidt, Vera Chetvertkova, Gemma Rosell-Tarragó, Florian Burkart and Daniel Wollmann
  Physical Review Accelerators and Beams **20**, 081001(2017)
  https://doi.org/10.1103/PhysRevAccelBeams.20.081001

APPENDIX B

Resum en Català

Les xarxes complexes són una bona representació matemàtica de la majoria dels sistemes complexos que es troben en la natura i en la societat: xarxes biològiques, socioeconòmiques o tecnològiques, entre d'altres. Aquestes xarxes, que representen sistemes reals, tenen característiques en comú que fan que no siguin ni purament regulars ni purament aleatòries. Aquestes propietats, ben conegudes, inclouen l'efecte de *small world*, la distribució heterogènia dels graus dels nodes i la presència de comunitats. La identificació dels components i les interaccions d'un sistema en els nodes i les arestes d'una xarxa, respectivament, permet explorar individualment la importància i el paper de cada element: les mesures de centralitat en són un bon exemple. No obstant això, la descripció completa d'una xarxa pot romandre incompleta si la seva estructura es considera desvinculada de la seva dinàmica.

Les xarxes complexes són representacions de sistemes en evolució, les parts constitutives i les connexions dels quals poden experimentar canvis al llarg del temps. Encara que la topologia d'una xarxa es consideri estàtica, hi poden estar tenint lloc processos dinàmics. Aquests sistemes es caracteritzen habitualment associant variables dinàmiques independents a cada node que s'acoblen només a través de les arestes de la xarxa: pensem en l'increment del voltatge transmembrana de les neurones en un cervell o l'evolució d'una malaltia infecciosa d'un individu en el curs d'una epidèmia. És important destacar que els patrons i fenòmens no trivials emergeixen degut a les interaccions existents en aquests sistemes dinàmics.

Molts sistemes reals d'interès es poden descriure matemàticament com sistemes oscil·latoris, és a dir, com un conjunt d'unitats que es modelitzen individualment com a oscil·ladors d'un tipus o un altre, però que s'acoblen amb els veïns a través de les connexions de la xarxa. El centelleig de les cuques de llum, els senyals neuronals del cervell o el flux d'energia a través de la xarxa elèctrica en són alguns exemples. Una descripció adequada de molts sistemes biològics, tecnològics i, fins i tot, socioeconòmics consisteix en considerar-los xarxes d'oscil·ladors de fase acoblats. Molt sovint, la dinàmica dels oscil·ladors considera que nodes veïns estan acoblats a través de les seves fases i independentment de l'amplitud de les seves oscil·lacions. Dins d'aquest marc, hi ha hagut especial interès en l'estudi de la sincronització, és a dir, quan tot el conjunt (o una fracció) dels oscil·ladors de la xarxa es mou a la mateixa freqüència.

Després que Arthur Taylor Winfree hagués fet el primer intent de modelitzar matemàticament aquesta dinàmica col·lectiva no lineal [181] i seguint el seu enfocament, Yoshiki Kuramoto va elaborar un model matemàtic senzill que podria capturar el fenòmen de la sincronització col·lectiva suggerint que els oscil·ladors estaven acoblats per una funció sinusoïdal de les seves diferències de fase [92, 93]. Tot i la seva simplicitat, el model de Kuramoto és capaç de capturar la transició de fase entre un estat purament caòtic, on tots els oscil·ladors es mouen de manera independent, a un estat coherent on, de manera progressiva, els oscil·ladors arriben a un estat de sincronització en freqüència i, finalment, totes les unitats acaben oscil·lant a l'uníson. .

Més tard, Yoshiki Kuramoto juntament amb Hidetsugu Sakaguchi van presentar una generalització del model d'oscil·ladors de cicle límit corresponent al model de Kuramoto, que incorporava un desfasament constant entre oscil·ladors [150]. Estudis posteriors del model han incorporat l'estructura de la xarxa dins del model juntament amb el desfasament constant (o paràmetre de frustració), tot considerant tots els oscil·ladors idèntics, és a dir, amb la mateixa freqüència natural. Per un rang ampli de valors del paràmetre de frustració el sistema se sincronitza a una freqüència resultant, o dit d'una altra manera, el sistema arriba a un estat estacionari. Tanmateix, el paràmetre de frustració obliga els nodes connectats a mantenir una diferència de fase constant i, per tant, el sistema deixa d'estar sincronitzat en fase. La magnitud d'aquest desfasament està determinada tant per la topologia de la xarxa com dels paràmetres de la dinàmica. Tanmateix, en els nodes topològicament simètrics la sincronització completa (en freqüència i en fase) es conserva , un fenomen que s'ha anomenat sincronització remota [124].

En el treball original de Kuramoto i Sakaguchi, així com en la majoria dels estudis conseqüents, és habitual considerar una distribució uniforme del paràmetre de desfasament. No obstant això, de la mateixa manera que el patró de connectivitat dels nodes és majoritàriament heterogeni, altres propietats que en són intrínseques no són necessàriament idèntiques en tots els componenets sinó que estan distribuïdes de manera no homogènia entre la població. Aquesta tesi representa una contribució a la comprensió del model de Kuramoto-Sakaguchi, especialment, en l'efecte de considerar una distribució heterogènia del paràmetre de frustració. El fet de tenir en compte diferents escenaris relacionats amb la distribució de la constant de desfasament entre parelles d'oscil·ladors representa un pas important en l'extensió del model original i proporciona informació única de l'estructura i la funció de la xarxa que s'està considerant.

El primer escenari que es considera en la tesi consisteix en pertorbar l'estat estacionari del sistema mitjançant la incorporació d'un paràmetre de desfasament diferent de zero a la dinàmica de cada node. L'objectiu d'aquest treball és proporcionar una ordenació dels nodes segons l'efecte potencial que tenen

sobre tota la xarxa quan un canvi en la seva dinàmica individual s'estén per tot el sistema oscil·latori, modificant així l'estat inicialment sincronitzat. En particular, definim la *funcionabilitat*, una nova mesura de centralitat que adreça la qüestió de quins són els nodes que, quan són pertorbats individualment, tenen més capacitat de moure el sistema fora de l'estat totalment sincronitzat.

Aquesta qüestió pot ser rellevant en la identificació dels nodes crítics del sistema, ja sigui pel fet de ser beneficiosos – permetent per exemple l'accés a un espectre més ampli d'estats – o bé perjudicials – destruint la sincronització general. Per tant, depenent del sistema que s'estigui considerant, caldrà identificar els nodes més funcionals, ja sigui per aconseguir una major diversitat d'estats accessibles com per inhibir inestabilitats que representin un risc pel sistema. A més, tot i que la *funcionabilitat* té el seu origen en un model dinàmic, en aquesta tesi obtenim l'expressió analítica d'aquesta centralitat. L'anterior expressió ens indica que l'ordenació dels nodes segons la seva *funcionabilitat* està determinada exclusivament per l'estructura de la xarxa. Cal destacar doncs que aquesta mesura de centralitat es defineix com una quantitat que no depèn dels paràmetres del model o d'una simulació numèrica. A més, l'expressió analítica ve donada per una funció matemàtica compacta i determinista de la topologia de xarxa i que, per tant, no es basa en mètodes numèrics d'optimització.

Després de comparar la *funcionabilitat* amb altres mesures de centralitat concloem que, tot i que la majoria de mesures de centralitat comparteixen certs patrons comuns, la *funcionabilitat* proporciona informació única sobre la xarxa i la importància dels seus nodes constitutius. És important destacar que obtenim que els nodes amb valors més grans de *funcionabilitat* estan ben connectats localment – tenen un grau alt – però també estan allunyats del nucli central de la xarxa – és a dir, són nodes perifèrics. Resumint, la *funcionabilitat* permet detectar els nodes que són més centrals o rellevants per tal de pertorbar el sistema de la sincronització total. Per exemple, els atacs epilèptics o els col·lapses de la xarxa elèctrica poden ser desencadenats per nodes que, encara que no estiguin situats al nucli principal, al modificar les seves propietats intrínseques, transmeten ràpidament l'asincronia a tota la xarxa, donant lloc a estats potencialment perillosos. Per tant, pot ser d'especial rellevància identificar aquests nodes per tal de controlar tant la sincronia com l'asincronia.

El segon escenari que aborda la present tesi considera una configuració més general en la qual el paràmetre de desfasament és una propietat intrínseca de cada node, no necessàriament zero, i n'explora la potencial heterogeneïtat. Una qüestió molt rellevant quan es consideren models oscil·latoris consisteix en trobar les condicions sota les quals la xarxa assoleix la sincronització. En aquest segon treball, proposem una metodologia, no només per obtenir l'estat de sincronització desitjat, sinó qualsevol altra configuració de les fases en l'estat estacionari mitjançant un ajust dels paràmetres de desfasament o frustració. Així,

obtenim la solució analítica dels paràmetres de frustració per aconseguir qualsevol configuració de les fases del conjunt d'oscil·ladors, linealitzant el model més general. D'aquesta manera, obtenim que els tres paràmetres intrínsecs dels nodes del model – les freqüències naturals, els paràmetres de frustració i les fases en estat l'estacionari – estan acoblats en una equació que permet ajustarlos a la configuració desitjada. Un resultat principal és que es pot accedir a una configuració de les fases determinada a través d'un espectre continu de paràmetres de frustració, és a dir, la fase d'un dels oscil·ladors, així com un dels paràmetre de frustració esdevenen paràmetres lliures degut als graus de llibertat.

En aquest mateix treball també es planteja la següent pregunta: 'D'entre totes les solucions possibles, quina és la que fa que el sistema assoleixi una determinada configuració de les fases amb el mínim cost requerit?'. Aquesta pregunta és d'especial rellevància si considerem la naturalesa de molts sistemes. Si un sistema real necessita accedir a una determinada configuració de les fases – que podria estar associada a una funció particular – aleshores tendirà a minimitzar l'esforç o el cost per fer-ho. Per això, un cop ajustats els paràmetres de frustració per obtenir un estat determinat, definim una funció de cost per avaluar la sobrecàrrega que li requereix al sistema arribar a aquesta configuració. D'aquesta manera, d'entre totes les possibles solucions de l'ajust dels paràmetres, ens quedem amb aquelles que minimitzin el cost d'obtenir-les. A més, en la present tesi obtenim la solució analítica de la funció de cost pel cas de dues configuracions particulars i fem una connexió amb el model no lineal de Kuramoto-Sakaguchi: tot i que l'anàlisi que realitzem es basa en la versió lineal del model, mostrem que l'ajust dels paràmetres de desfasament proposat també és capaç de potenciar la sincronització en freqüència del sistema.

Finalment, en l'últim escenari es torna a considerar una distribució dels paràmetres de desfasament homogènia: l'estat de les fases obtingut en l'estat estacionari quan tots els nodes comparteixen un paràmetre de frustració comú està estretament lligat a les simetries topològiques de la xarxa. En diversos estudis empírics es mostra que un atribut de les xarxes del món real és que presenten un cert grau de simetria. L'estudi de les simetries d'una xarxa és de gran rellevància per diversos motius: ens pot ajudar a entendre millor la formació de determinades xarxes del món real, també poden proporcionar informació sobre la funció dels nodes i tenir un efecte sobre la redundància i la robustesa de la xarxa. A més, se sap que les simetries influeixen en el resultat de la dinàmica de la xarxa, com ara en la sincronització o la controlabilitat. La noció de "simetria" o"invariància" inclou diverses especificacions segons el camp en què s'apliqui. Una transformació topològica d'un graf o d'una xarxa intercanvia cada vèrtex amb un altre en forma de permutació. El conjunt de permutacions d'un graf que deixen la topologia invariant són els automorfismes d'aquest graf. Tenint en compte els automorfismes d'un graf, el conjunt dels seus vèrtexs es pot di-

vidir en: un nucli format pels punts fixos – vèrtexs que es mantenen invariants en tots els automorfismes del graf – i el conjunt de vèrtexs que formen motius simètrics – corresponents a les diferents permutacions. Cada motiu simètric es pot dividir en grups, també anomenats òrbites, que representen nodes estructuralment indistinguibles, és a dir, que tenen el mateix paper estructural a la xarxa.

No obstant això, més enllà de la simetria estructural o topològica, cal tenir en compte el fet que les xarxes del món real estan exposades a fluctuacions o errors, així com insercions o eliminacions errònies. Per tant, malgrat que hi ha molts programaris d'àlgebra discreta que són capaços de determinar el grup d'automorfismes, és a dir, les simetries d'un graf, així com d'extreure les òrbites a les que pertanyen cada un dels nodes, agrupant-los en clusters, ens interessa construir una metodologia que permeti la detecció de, no només les simetries perfectes, sinó les simetries aproximades. Hi ha hagut una sèrie d'intents de tractar les simetries aproximades en xarxes complexes. En la present tesi, proporcionem una noció alternativa a les simetries aproximades, que anomenem "*Quasi-simetries*". A diferència d'altres definicions, les quasi-simetries romanen lliures d'imposar qualsevol invariància d'alguna de les propietats de la xarxa i s'obtenen a partir de l'estat estacionari del model de Kuramoto-Sakaguchi quan es considera una distribució homogènia del paràmetre de desfasament.

Una primera contribució consisteix en explorar les distribucions de similitud estructural entre tots els parells de nodes i establir un punt de referència per determinar si una xarxa té un patró més complex que el d'una xarxa aleatòria pel que fa a les quasi-simetries: aquest criteri consistirà en determinar si el nombre de quasi-simetries o grups simètrics és més gran que u.

En segon lloc, definim la "*xarxa dual*", una xarxa ponderada – i el seu corresponent homòleg binaritzat – que integra en definitiva tota la informació relativa a les quasi-simetries en la xarxa original. La definició de la xarxa dual permet aplicar els mètodes estàndards d'anàlisi de xarxes, com són les mesures de centralitat i la detecció de comunitats. El primer ens informa sobre els nodes que tenen un paper únic a la xarxa i dels que es comporten de manera semblant a molts altres nodes. El segon proporciona una classificació dels nodes en comunitats quasi-simètriques, l'extensió natural de les òrbites del grup d'automorfismes (nodes estructuralment simètrics) d'una xarxa. L'ús de la xarxa dual binària és avantatjós ja que condueix a resultats més heterogenis en la obtenció de la importància dels nodes i permet una classificació més significativa de nodes en comunitats quasi-simètriques.

Per tant, en aquest darrer treball presentem un marc general per tractar simetries aproximades en xarxes complexes. La xarxa dual es presenta com una eina útil per treballar amb quasi-simetries i s'especifiquen diverses aplicacions.

Els diferents treballs que integren la present tesi ens ajuden a concloure que, de l'expressió matemàtica compacta i senzilla del model oscil·latori original de Kuramoto-Sakaguchi se'n deriven nocions interessants quan considerem l'heterogeneïtat en la distribució dels paràmetres de desfasament. Considerar tres escenaris diferents per a la distribució dels paràmetres de frustració entre la població ens ha portat a obtenir una comprensió més profunda de la connexió entre la dinàmica i l'estructura de xarxes complexes. La definició i estudi de la *funcionabilitat* com a mesura de centralitat o les quasi-simetries en són exemples.

Al llarg d'aquesta tesi hem obtingut la solució analítica a diverses qüestions – sempre que ha estat possible – ja que tenir una expressió matemàtica com a resultat permet una interpretació i aprofundiment més significatius dels problemes plantejats. A més, hem il·lustrat la majoria dels conceptes i resultats considerant xarxes sintètiques simples. D'aquesta manera, el sentit i els procediments necessaris de cada treball es fan més clars per al lector. Addicionalment, s'ha prestat especial atenció a aplicar els diferents resultats a xarxes reals, en particular, xarxes reals del cervell. Tanmateix, som conscients que en manca una interpretació per part d'un expert en l'àmbit d'aquests resultats i, per tant, representa una limitació. Per aquest motiu, una extensió natural de la tesi és l'aplicació de les diferents mesures i resultats a xarxes del món real, juntament amb una interpretació significativa dels resultats, la qual cosa suposaria un important pas endavant per als camps considerats. A més, hi ha molt de marge per obtenir altres resultats interessants, com ara l'extensió dels resultats obtinguts a xarxes multicapa, temporals o considerant interaccions d'ordre superior, com són els hipergrafs.

References

[1] Gap - groups, algorithms, programming - a system for computational discrete algebra. https://www.gap-system.org/. 97

[2] saucy 3.0. http://vlsicad.eecs.umich.edu/BK/SAUCY/. 97

[3] Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, 15(3):247–262, September 2013. doi: 10.31887/dcns.2013.15.3/osporns. URL https://doi.org/10.31887/dcns.2013.15.3/osporns. 34

[4] Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., and Spigler, R. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, April 2005. doi: 10.1103/revmodphys.77.137. URL https://doi.org/10.1103/revmodphys.77.137. 2, 21, 26, 33, 59, 98

[5] Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. Heart rate variability: a review. *Medical & Biological Engineering & Computing*, 44(12):1031–1051, November 2006. doi: 10.1007/s11517-006-0119-0. URL https://doi.org/10.1007/s11517-006-0119-0. 34

[6] Acharya, U. R., Joseph, K. P., Kannathal, N., Min, L. C., and Suri, J. S. Heart rate variability. In *Advances in Cardiac Signal Processing*, pages 121–165. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-36675-1_5. URL https://doi.org/10.1007/978-3-540-36675-1_5. 60

[7] Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, June 2010. doi: 10.1038/nature09182. URL https://doi.org/10.1038/nature09182. 8

[8] Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002. doi: 10.1103/revmodphys.74.47. URL https://doi.org/10.1103/revmodphys.74.47. 13, 17, 91

[9] Albert, R., Jeong, H., and Barabasi, A.-L. Erratum: correction: Error and attack tolerance of complex networks. *Nature*, 409(6819):542–542, January 2001. doi: 10.1038/35054111. URL https://doi.org/10.1038/35054111. 49, 50

[10] Albert, R., Albert, I., and Nakarado, G. L. Structural vulnerability of the north american power grid. *Physical Review E*, 69(2), February 2004. doi: 10.1103/physreve.69.025103. URL https://doi.org/10.1103/physreve.69.025103. 34

[11] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, June 2007. doi: 10.1038/nrg2102. URL https://doi.org/10.1038/nrg2102. 95

[12] Arenas, A. and Díaz-Guilera, A. Synchronization and modularity in complex networks. *The European Physical Journal Special Topics*, 143(1):19–25, April 2007. doi: 10.1140/epjst/e2007-00066-2. URL https://doi.org/10.1140/epjst/e2007-00066-2. 33

[13] Arenas, A., Díaz-Guilera, A., and Pérez-Vicente, C. J. Synchronization processes in complex networks. *Physica D: Nonlinear Phenomena*, 224(1-2):27–34, December 2006. doi: 10.1016/j.physd.2006.09.029. URL https://doi.org/10.1016/j.physd.2006.09.029. 33

[14] Arenas, A., Díaz-Guilera, A., and Pérez-Vicente, C. J. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11), March 2006. doi: 10.1103/physrevlett.96.114102. URL https://doi.org/10.1103/physrevlett.96.114102. 9, 26, 33

[15] Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., and Zhou, C. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, December 2008. doi: 10.1016/j.physrep.2008.09.002. URL https://doi.org/10.1016/j.physrep.2008.09.002. 19, 21, 22, 25, 26, 33, 59, 98

[16] Arianos, S., Bompard, E., Carbone, A., and Xue, F. Power grid vulnerability: A complex network approach. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(1):013119, March 2009. doi: 10.1063/1.3077229. URL https://doi.org/10.1063/1.3077229. 34

[17] A.Simon, H. On a class of skew distribution functions. 42(3-4):425–440, 1955. doi: 10.1093/biomet/42.3-4.425. URL https://doi.org/10.1093/biomet/42.3-4.425. 16

[18] Barabási, A.-L. Scale-free networks: A decade and beyond. *Science*, 325 (5939):412–413, July 2009. doi: 10.1126/science.1173299. URL https://doi.org/10.1126/science.1173299. 91

[19] Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. 286(5439):509–512, October 1999. doi: 10.1126/science.286.5439.509. URL https://doi.org/10.1126/science.286.5439.509. 14

[20] Barabási, A.-L. and Pósfai, M. *Network science*. Cambridge University Press, Cambridge, 2016. ISBN 9781107076266 1107076269. URL http://barabasi.com/networksciencebook/. 3, 17

[21] Barrat, A., Barthelemy, M., and Vespignani, A. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008. doi: 10.1017/cbo9780511791383. URL https://doi.org/10.1017/cbo9780511791383. 18, 59

[22] Bassett, D. S., Wymbs, N. F., Rombach, M. P., Porter, M. A., Mucha, P. J., and Grafton, S. T. Task-based core-periphery organization of human brain dynamics. *PLoS Computational Biology*, 9(9):e1003171, September 2013. doi: 10.1371/journal.pcbi.1003171. URL https://doi.org/10.1371/journal.pcbi.1003171. 49

[23] Battiston, F., Amico, E., Barrat, A., Bianconi, G., de Arruda, G. F., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., Murray, M. M., Peixoto, T. P., Vaccarino, F., and Petri, G. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, October 2021. doi: 10.1038/s41567-021-01371-4. URL https://doi.org/10.1038/s41567-021-01371-4. 19

[24] Bavelas, A. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, November 1950. doi: 10.1121/1.1906679. URL https://doi.org/10.1121/1.1906679. 12

[25] Benzi, M. and Klymko, C. On the limiting behavior of parameter-dependent network centrality measures. *SIAM Journal on Matrix Analysis and Applications*, 36(2):686–706, January 2015. doi: 10.1137/130950550. URL https://doi.org/10.1137/130950550. 9, 45

[26] Bergner, A., Frasca, M., Sciuto, G., Buscarino, A., Ngamga, E. J., Fortuna, L., and Kurths, J. Remote synchronization in star networks. *Physical Review E*, 85(2), February 2012. doi: 10.1103/physreve.85.026208. URL https://doi.org/10.1103/physreve.85.026208. 35

[27] Betzel, R. F., Avena-Koenigsberger, A., Goñi, J., He, Y., de Reus, M. A., Griffa, A., Vértes, P. E., Mišic, B., Thiran, J.-P., Hagmann, P., van den Heuvel, M., Zuo, X.-N., Bullmore, E. T., and Sporns, O. Generative models of the human connectome. *NeuroImage*, 124:1054–1064, January 2016. doi: 10.1016/j.neuroimage.2015.09.041. URL https://doi.org/10.1016/j.neuroimage.2015.09.041. 34

[28] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008. doi: 10.1088/1742-5468/2008/10/p10008. URL https://doi.org/10.1088/1742-5468/2008/10/p10008. 7

[29] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006. doi: 10.1016/j.physrep.2005.10.009. URL https://doi.org/10.1016/j.physrep.2005.10.009. 2, 34, 91

[30] Boccaletti, S., Pisarchik, A. N., del Genio, C. I., and Amann, A. *Synchronization: From Coupled Systems to Complex Networks*. Cambridge University Press, March 2018. doi: 10.1017/9781107297111. URL https://doi.org/10.1017/9781107297111. 59

[31] Bonacich, P. Factoring and weighting approaches to status scores and clique identification. 2(1):113–120, January 1972. doi: 10.1080/0022250x.1972.9989806. URL https://doi.org/10.1080/0022250x.1972.9989806. 11

[32] Bonacich, P. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, March 1987. doi: 10.1086/228631. URL https://doi.org/10.1086/228631. 9, 11

[33] Bonacich, P. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, October 2007. doi: 10.1016/j.socnet.2007.04.002. URL https://doi.org/10.1016/j.socnet.2007.04.002. 11, 45

[34] Borgatti, S. P. Centrality and network flow. *Social Networks*, 27(1):55–71, January 2005. doi: 10.1016/j.socnet.2004.11.008. URL https://doi.org/10.1016/j.socnet.2004.11.008. 9, 45

[35] Borgatti, S. P. and Everett, M. G. Models of core/periphery structures. *Social Networks*, 21(4):375–395, October 2000. doi: 10.1016/s0378-8733(99)00019-2. URL https://doi.org/10.1016/s0378-8733(99)00019-2. 49

[36] Botev, Z. I., Grotowski, J. F., and Kroese, D. P. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, October 2010. doi: 10.1214/10-aos799. URL https://doi.org/10.1214/10-aos799. 129, 130

[37] Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188, February 2008. doi: 10.1109/tkde.2007.190689. URL https://doi.org/10.1109/tkde.2007.190689. 9

[38] Brede, M. and Kalloniatis, A. C. Frustration tuning and perfect phase synchronization in the kuramoto-sakaguchi model. *Physical Review E*, 93(6), June 2016. doi: 10.1103/physreve.93.062315. URL https://doi.org/10.1103/physreve.93.062315. 25, 36, 60, 79, 80, 82

[39] Broido, A. D. and Clauset, A. Scale-free networks are rare. *Nature Communications*, 10(1), March 2019. doi: 10.1038/s41467-019-08746-5. URL https://doi.org/10.1038/s41467-019-08746-5. 14

[40] Burke, S. N. and Barnes, C. A. Neural plasticity in the ageing brain. *Nature Reviews Neuroscience*, 7(1):30–40, January 2006. doi: 10.1038/nrn1809. URL https://doi.org/10.1038/nrn1809. 34

[41] Buzsaki, G. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, June 2004. doi: 10.1126/science.1099745. URL https://doi.org/10.1126/science.1099745. 34

[42] Caldarelli, G. *Scale-Free Networks*. Oxford University Press, May 2007. doi: 10.1093/acprof:oso/9780199211517.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199211517.001.0001. 91

[43] Caldarelli, G. and Vespignani, A. *Large Scale Structure and Dynamics of Complex Networks*. WORLD SCIENTIFIC, June 2007. doi: 10.1142/6455. URL https://doi.org/10.1142/6455. 18, 91

[44] Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471, December 2000. doi: 10.1103/physrevlett.85.5468. URL https://doi.org/10.1103/physrevlett.85.5468. 50

[45] Charles C, T. Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76(4):705–712, 1989. doi: 10.1093/biomet/76.4.705. URL https://doi.org/10.1093/biomet/76.4.705. 112

[46] Clauset, A., Newman, M. E. J., and Moore, C. Finding community structure in very large networks. *Physical Review E*, 70(6), December 2004. doi: 10.1103/physreve.70.066111. URL https://doi.org/10.1103/physreve.70.066111. 9

[47] Clauset, A., Shalizi, C. R., and Newman, M. E. J. Power-law distributions in empirical data. 51(4):661–703, November 2009. doi: 10.1137/070710111. URL https://doi.org/10.1137/070710111. 14

[48] Cucuringu, M., Rombach, P., Lee, S. H., and Porter, M. A. Detection of core–periphery structure in networks using spectral methods and geodesic paths. *European Journal of Applied Mathematics*, 27(6):846–887, August 2016. doi: 10.1017/s095679251600022x. URL https://doi.org/10.1017/s095679251600022x. 49

[49] Cundy, H. M. Geometric symmetry, by e. h. lockwood and r. h. macmillan. pp x, 228. £10·50. 1978. SBN 0 521 21685 0 (cambridge university press). *The Mathematical Gazette*, 63(425):212–214, October 1979. doi: 10.2307/3617910. URL https://doi.org/10.2307/3617910. 91

[50] da Fonseca, J. and Abud, C. The kuramoto model revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(10):103204, October 2018. doi: 10.1088/1742-5468/aadb05. URL https://doi.org/10.1088/1742-5468/aadb05. 25

[51] De Domenico, M. and Sayama, H. Complexity explained. 2019. doi: 10.17605/OSF.IO/TQGNW. URL https://osf.io/tqgnw/. 2

[52] de Solla Price, D. J. Networks of scientific papers. 149(3683):510–515, July 1965. doi: 10.1126/science.149.3683.510. URL https://doi.org/10.1126/science.149.3683.510. 14

[53] Dörfler, F. and Bullo, F. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50(6):1539–1564, June 2014. doi: 10.1016/j.automatica.2014.04.012. URL https://doi.org/10.1016/j.automatica.2014.04.012. 59

[54] D'Souza, R. M., Gómez-Gardeñes, J., Nagler, J., and Arenas, A. Explosive phenomena in complex networks. *Advances in Physics*, 68(3):123–223, July 2019. doi: 10.1080/00018732.2019.1650450. URL https://doi.org/10.1080/00018732.2019.1650450. 81

[55] Erdős, P. and Rényi, A. Asymmetric graphs. *Acta Mathematica Academiae Scientiarum Hungaricae*, 14(3-4):295–315, September 1963. doi: 10.1007/bf01895716. URL https://doi.org/10.1007/bf01895716. 15, 93, 102

[56] Estrada, E. and Higham, D. J. Network properties revealed through matrix functions. *SIAM Review*, 52(4):696–714, January 2010. doi: 10.1137/090761070. URL https://doi.org/10.1137/090761070. 45

[57] Fani, H. and Bagheri, E. Community detection in social networks. *Encyclopedia with Semantic Computing and Robotic Intelligence*, 01(01):1630001, March 2017. doi: 10.1142/s2425038416300019. URL https://doi.org/10.1142/s2425038416300019. 7

[58] Ferraro, G. D., Moreno, A., Min, B., Morone, F., Pérez-Ramírez, Ú., Pérez-Cervera, L., Parra, L. C., Holodny, A., Canals, S., and Makse, H. A. Finding influential nodes for integration in brain networks using optimal percolation theory. *Nature Communications*, 9(1), June 2018. doi: 10.1038/s41467-018-04718-3. URL https://doi.org/10.1038/s41467-018-04718-3. 50

[59] Fienberg, S. E. A brief history of statistical models for network analysis and open challenges. 21(4):825–839, October 2012. doi: 10.1080/10618600.2012.738106. URL https://doi.org/10.1080/10618600.2012.738106. 15

[60] for Physics, T. N. C. Scientific background on the nobel prize in physics 2021. URL https://www.nobelprize.org/uploads/2021/10/sciback_fy_en_21.pdf. 2

[61] Fortunato, S. and Barthelemy, M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, December 2006. doi: 10.1073/pnas.0605965104. URL https://doi.org/10.1073/pnas.0605965104. 9

[62] Fortunato, S. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, February 2010. doi: 10.1016/j.physrep.2009.11.002. URL https://doi.org/10.1016/j.physrep.2009.11.002. 8

[63] Fortunato, S. and Hric, D. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, November 2016. doi: 10.1016/j.physrep.2016.09.002. URL https://doi.org/10.1016/j.physrep.2016.09.002. 7, 8

[64] Freeman, L. C. A set of measures of centrality based on betweenness. 40(1):35, March 1977. doi: 10.2307/3033543. URL https://doi.org/10.2307/3033543. 12

[65] Frobenius, G. *Über Matrizen aus nicht negativen Elementen*. Preussische Akademie der Wissenschaften Berlin: Sitzungsberichte der Preußischen Akademie der Wissenschaften zu Berlin. Reichsdr., 1912. ISBN 9783111271903. 4

[66] Frolov, N., Maksimenko, V., Majhi, S., Rakshit, S., Ghosh, D., and Hramov, A. Chimera-like behavior in a heterogeneous kuramoto model: The interplay between attractive and repulsive coupling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(8):081102, August 2020. doi: 10.1063/5.0019200. URL https://doi.org/10.1063/5.0019200. 59

[67] Fruchterman, T. M. J. and Reingold, E. M. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, November 1991. doi: 10.1002/spe.4380211102. URL https://doi.org/10.1002/spe.4380211102. 43, 44, 45, 46, 52, 114, 115, 116, 117, 119, 122, 131, 132, 133, 134

[68] Galbiati, M., Delpini, D., and Battiston, S. The power to control. *Nature Physics*, 9(3):126–128, March 2013. doi: 10.1038/nphys2581. URL https://doi.org/10.1038/nphys2581. 48

[69] Garlaschelli, D., Ruzzenenti, F., and Basosi, R. Complex networks and symmetry i: A review. *Symmetry*, 2(3):1683–1709, September 2010. doi: 10.3390/sym2031683. URL https://doi.org/10.3390/sym2031683. 91, 92, 103

[70] Garrido, A. Symmetry in complex networks. *Symmetry*, 3(1):1–15, January 2011. doi: 10.3390/sym3010001. URL https://doi.org/10.3390/sym3010001. 94

[71] Gell-Mann, M. What is complexity? remarks on simplicity and complexity by the nobel prize-winning author of the quark and the jaguar. *Complexity*, 1(1):16–19, September 1995. doi: 10.1002/cplx.6130010105. URL https://doi.org/10.1002/cplx.6130010105. 1

[72] Ghosh, R., hua Teng, S., Lerman, K., and Yan, X. The interplay between dynamics and networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, August 2014. doi: 10.1145/2623330.2623738. URL https://doi.org/10.1145/2623330.2623738. 45

[73] Gilbert, E. N. Random graphs. 30(4):1141–1144, December 1959. doi: 10.1214/aoms/1177706098. URL https://doi.org/10.1214/aoms/1177706098. 15

[74] Girvan, M. and Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12):7821–7826, June 2002. doi: 10.1073/pnas.122653799. URL https://doi.org/10.1073/pnas.122653799. 7, 17, 91

[75] Goldstein, J. Emergence in complex systems. In Allen, P., Maguire, S., and McKelvey, B., editors, *The Sage Handbook of Complexity and Management*, pages 65–78. Sage Publications, 2011. 2

[76] Gómez-Gardeñes, J., Gómez, S., Arenas, A., and Moreno, Y. Explosive synchronization transitions in scale-free networks. *Phys. Rev. Lett.*, 106: 128701, Mar 2011. doi: 10.1103/PhysRevLett.106.128701. URL https://link.aps.org/doi/10.1103/PhysRevLett.106.128701. 81

[77] Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., Medaglia, J. D., Vettel, J. M., Miller, M. B., Grafton, S. T., and Bassett, D. S. Controllability of structural brain networks. *Nature Communications*, 6 (1), October 2015. doi: 10.1038/ncomms9414. URL https://doi.org/10.1038/ncomms9414. 48

[78] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6(7):e159, July 2008. doi: 10.1371/journal.pbio.0060159. URL https://doi.org/10.1371/journal.pbio.0060159. 49

[79] Hall, P., Sheater, S. J., Jones, M. C., and Marron, J. S. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269, 1991. doi: 10.1093/biomet/78.2.263. URL https://doi.org/10.1093/biomet/78.2.263. 112

[80] Hamdan, A. M. A. and Nayfeh, A. H. Measures of modal controllability and observability for first- and second-order linear systems. *Journal of Guidance, Control, and Dynamics*, 12(3):421–428, May 1989. doi: 10.2514/3.20424. URL https://doi.org/10.2514/3.20424. 48

[81] Holme, P. Detecting degree symmetries in networks. *Physical Review E*, 74(3), September 2006. doi: 10.1103/physreve.74.036107. URL https://doi.org/10.1103/physreve.74.036107. 92

[82] Holme, P. Rare and everywhere: Perspectives on scale-free networks. 10 (1), March 2019. doi: 10.1038/s41467-019-09038-8. URL https://doi.org/10.1038/s41467-019-09038-8. 14

[83] Hubbell, C. H. An input-output approach to clique identification. 28(4): 377, December 1965. doi: 10.2307/2785990. URL https://doi.org/10.2307/2785990. 11

[84] Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Penna, S. D., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A.,

de Pasquale, F., Sporns, O., Walter, M., and Chang, C. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 80:360–378, October 2013. doi: 10.1016/j.neuroimage.2013.05.079. URL https://doi.org/10.1016/j.neuroimage.2013.05.079. 34

[85] Jiang, X. and Abrams, D. M. Symmetry-broken states on networks of coupled oscillators. *Physical Review E*, 93(5), May 2016. doi: 10.1103/physreve.93.052202. URL https://doi.org/10.1103/physreve.93.052202. 36, 91

[86] Kaiser, M. and Hilgetag, C. C. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Computational Biology*, 2(7):e95, July 2006. doi: 10.1371/journal.pcbi.0020095. URL https://doi.org/10.1371/journal.pcbi.0020095. 45

[87] Katz, L. A new status index derived from sociometric analysis. 18(1):39–43, March 1953. doi: 10.1007/bf02289026. URL https://doi.org/10.1007/bf02289026. 11

[88] Ketchen, D. J. and Shook, C. L. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6):441–458, 1996. ISSN 01432095, 10970266. URL http://www.jstor.org/stable/2486927. 122

[89] Klickstein, I. and Sorrentino, F. Generating symmetric graphs. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(12):121102, December 2018. doi: 10.1063/1.5064375. URL https://doi.org/10.1063/1.5064375. 95, 97

[90] Kuhnert, M.-T., Geier, C., Elger, C. E., and Lehnertz, K. Identifying important nodes in weighted functional brain networks: A comparison of different centrality approaches. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):023142, June 2012. doi: 10.1063/1.4729185. URL https://doi.org/10.1063/1.4729185. 43

[91] Kumpula, J. M., Saramäki, J., Kaski, K., and Kertész, J. Limited resolution in complex network community detection with potts model approach. *The European Physical Journal B*, 56(1):41–45, March 2007. doi: 10.1140/epjb/e2007-00088-4. URL https://doi.org/10.1140/epjb/e2007-00088-4. 9

[92] Kuramoto, Y. Self-entrainment of a population of coupled non-linear oscillators. In *Lecture Notes in Physics*, volume 30, Berlín, Heidelberg, 1975.

Springer. ISBN 978-3-540-07174-7. doi: 10.1007/BFb0013365. 22, 59, 98, 136, 144

[93] Kuramoto, Y. *Chemical Oscillations, Waves, and Turbulence*. Springer Berlin Heidelberg, 1984. doi: 10.1007/978-3-642-69689-3. URL https://doi.org/10.1007/978-3-642-69689-3. 22, 33, 53, 59, 136, 144

[94] Lei, X., Li, X., and Povh, D. A nonlinear control for coordinating TCSC and generator excitation to enhance the transient stability of long transmission systems. *Electric Power Systems Research*, 59(2):103–109, September 2001. doi: 10.1016/s0378-7796(01)00139-0. URL https://doi.org/10.1016/s0378-7796(01)00139-0. 34, 60

[95] Lesch, K.-P. and Waider, J. Serotonin in the modulation of neural plasticity and networks: Implications for neurodevelopmental disorders. *Neuron*, 76(1):175–191, October 2012. doi: 10.1016/j.neuron.2012.09.013. URL https://doi.org/10.1016/j.neuron.2012.09.013. 34

[96] Levin, A. and Narendra, K. Control of nonlinear dynamical systems using neural networks: controllability and stabilization. *IEEE Transactions on Neural Networks*, 4(2):192–206, March 1993. doi: 10.1109/72.207608. URL https://doi.org/10.1109/72.207608. 48

[97] Li, Y., Shang, Y., and Yang, Y. Clustering coefficients of large networks. 382-383:350–358, March 2017. doi: 10.1016/j.ins.2016.12.027. URL https://doi.org/10.1016/j.ins.2016.12.027. 14

[98] Liu, Y.-Y., Slotine, J.-J., and Barabasi, A.-L. Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7):2460–2465, January 2013. doi: 10.1073/pnas.1215508110. URL https://doi.org/10.1073/pnas.1215508110. 91

[99] Liu, Y.-Y. and Barabási, A.-L. Control principles of complex systems. *Reviews of Modern Physics*, 88(3), September 2016. doi: 10.1103/revmodphys.88.035006. URL https://doi.org/10.1103/revmodphys.88.035006. 59

[100] Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. Controllability of complex networks. *Nature*, 473(7346):167–173, May 2011. doi: 10.1038/nature10011. URL https://doi.org/10.1038/nature10011. 34, 48, 91

[101] Lombardi, A. and Hörnquist, M. Controllability analysis of networks. *Physical Review E*, 75(5), May 2007. doi: 10.1103/physreve.75.056110. URL https://doi.org/10.1103/physreve.75.056110. 48

[102] Luce, R. D. and Perry, A. D. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, June 1949. doi: 10.1007/bf02289146. URL https://doi.org/10.1007/bf02289146. 7

[103] Ma, A. and Mondragón, R. J. Rich-cores in networks. *PLOS ONE*, 10(3): e0119678, March 2015. doi: 10.1371/journal.pone.0119678. URL https://doi.org/10.1371/journal.pone.0119678. 49

[104] MacArthur, B. D., Sánchez-García, R. J., and Anderson, J. W. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531, November 2008. doi: 10.1016/j.dam.2008.04.008. URL https://doi.org/10.1016/j.dam.2008.04.008. 94

[105] Malliaros, F. D. and Vazirgiannis, M. Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4):95–142, December 2013. doi: 10.1016/j.physrep.2013.08.002. URL https://doi.org/10.1016/j.physrep.2013.08.002. 7

[106] McKay, B. D. and Piperno, A. Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60:94–112, January 2014. doi: 10.1016/j.jsc.2013.09.003. URL https://doi.org/10.1016/j.jsc.2013.09.003. 97

[107] Milgram, S. The Small-World Problem. *Psychology Today*, 1(1):61–67, 1967. 13

[108] Mišić, B. and Sporns, O. From regions to connections and networks: new bridges between brain and behavior. *Current Opinion in Neurobiology*, 40: 1–7, October 2016. doi: 10.1016/j.conb.2016.05.003. URL https://doi.org/10.1016/j.conb.2016.05.003. 34

[109] Mitra, N. J., Guibas, L. J., and Pauly, M. Partial and approximate symmetry detection for 3d geometry. *ACM Transactions on Graphics*, 25(3):560–568, July 2006. doi: 10.1145/1141911.1141924. URL https://doi.org/10.1145/1141911.1141924. 102

[110] Molnar, F., Nishikawa, T., and Motter, A. E. Network experiment demonstrates converse symmetry breaking. *Nature Physics*, 16(3):351–356, January 2020. doi: 10.1038/s41567-019-0742-y. URL https://doi.org/10.1038/s41567-019-0742-y. 60

[111] Morer, I., Cardillo, A., Díaz-Guilera, A., Prignano, L., and Lozano, S. Comparing spatial networks: A one-size-fits-all efficiency-driven approach. *Physical Review E*, 101(4), April 2020. doi: 10.1103/physreve.101.042301. URL https://doi.org/10.1103/physreve.101.042301. 12

[112] Morone, F. and Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature*, 524(7563):65–68, July 2015. doi: 10.1038/nature14604. URL https://doi.org/10.1038/nature14604. 50

[113] Morone, F., Roth, K., Min, B., Stanley, H. E., and Makse, H. A. Model of brain activation predicts the neural collective influence map of the brain. *Proceedings of the National Academy of Sciences*, 114(15):3849–3854, March 2017. doi: 10.1073/pnas.1620808114. URL https://doi.org/10.1073/pnas.1620808114. 50

[114] Motter, A. E., Zhou, C., and Kurths, J. Network synchronization, diffusion, and the paradox of heterogeneity. 71(1), January 2005. doi: 10.1103/physreve.71.016116. URL https://doi.org/10.1103/physreve.71.016116. 25

[115] Negre, C. F. A., Morzan, U. N., Hendrickson, H. P., Pal, R., Lisi, G. P., Loria, J. P., Rivalta, I., Ho, J., and Batista, V. S. Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences*, 115(52):E12201–E12208, December 2018. doi: 10.1073/pnas.1810452115. URL https://doi.org/10.1073/pnas.1810452115. 9

[116] Newman, M. E. J. Scientific collaboration networks i. network construction and fundamental results. *Physical Review E*, 64(1), June 2001. doi: 10.1103/physreve.64.016131. URL https://doi.org/10.1103/physreve.64.016131. 15, 50, 91

[117] Newman, M. E. J. Assortative mixing in networks. *Physical Review Letters*, 89(20), October 2002. doi: 10.1103/physrevlett.89.208701. URL https://doi.org/10.1103/physrevlett.89.208701. 6

[118] Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002. doi: 10.1103/PhysRevE.66.016128. URL https://link.aps.org/doi/10.1103/PhysRevE.66.016128. 19

[119] Newman, M. E. J. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, January 2003. doi: 10.1137/s003614450342480. URL https://doi.org/10.1137/s003614450342480. 7

[120] Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), September 2006. doi: 10.1103/physreve.74.036104. URL https://doi.org/10.1103/physreve.74.036104. 7

[121] Newman, M. E. J. and Girvan, M. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), February 2004. doi: 10.1103/physreve.69.026113. URL https://doi.org/10.1103/physreve.69.026113. 7, 8, 9

[122] Newman, M. *Networks: an Introduction*. Oxford University Press, March 2010. doi: 10.1093/acprof:oso/9780199206650.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199206650.001.0001. 3, 14, 18, 29

[123] Nicolis, G. and Nicolis, C. *Foundations of Complex Systems*. WORLD SCIENTIFIC, September 2007. doi: 10.1142/6253. URL https://doi.org/10.1142/6253. 59

[124] Nicosia, V., Valencia, M., Chavez, M., Díaz-Guilera, A., and Latora, V. Remote synchronization reveals network symmetries and functional modules. *Physical Review Letters*, 110(17), April 2013. doi: 10.1103/physrevlett.110.174102. URL https://doi.org/10.1103/physrevlett.110.174102. 4, 25, 26, 33, 35, 59, 67, 91, 98, 99, 122, 136, 144

[125] Nishikawa, T. and Motter, A. E. Symmetric states requiring system asymmetry. *Physical Review Letters*, 117(11), September 2016. doi: 10.1103/physrevlett.117.114101. URL https://doi.org/10.1103/physrevlett.117.114101. 60, 98

[126] Noldus, R. and Mieghem, P. V. Assortativity in complex networks. *Journal of Complex Networks*, 3(4):507–542, March 2015. doi: 10.1093/comnet/cnv005. URL https://doi.org/10.1093/comnet/cnv005. 6

[127] Olver, P. The symmetry groupoid and weighted signature of a geometric object. *Journal of Lie Theory*, 26(1):235–267, January 2016. ISSN 0949-5932. 92

[128] Omel'chenko, O. E. and Wolfrum, M. Nonuniversal transitions to synchrony in the sakaguchi-kuramoto model. *Physical Review Letters*, 109 (16), October 2012. doi: 10.1103/physrevlett.109.164101. URL https://doi.org/10.1103/physrevlett.109.164101. 25, 35

[129] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. Structure and tie strengths in mobile communication networks. 104(18):7332–7336, April 2007. doi: 10.1073/pnas.0610245104. URL https://doi.org/10.1073/pnas.0610245104. 14

[130] Opsahl, T. and Panzarasa, P. Clustering in weighted networks. *Social Networks*, 31(2):155–163, May 2009. doi: 10.1016/j.socnet.2009.02.002. URL https://doi.org/10.1016/j.socnet.2009.02.002. 7

[131] Osipov, G. V., Kurths, J., and Zhou, C. *Synchronization in Oscillatory Networks*. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-71269-5. URL https://doi.org/10.1007/978-3-540-71269-5. 19, 21, 33, 59

[132] Pakdemirli, M., Yürüsoy, M., and Dolapçı, İ. T. Comparison of approximate symmetry methods for differential equations. *Acta Applicandae Mathematicae*, 80(3):243–271, February 2004. doi: 10.1023/b:acap.0000018792.87732.25. URL https://doi.org/10.1023/b:acap.0000018792.87732.25. 102

[133] Parzen, E. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962. doi: 10.1214/aoms/1177704472. URL https://doi.org/10.1214/aoms/1177704472. 129

[134] Pasqualetti, F., Zampieri, S., and Bullo, F. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, March 2014. doi: 10.1109/tcns.2014.2310254. URL https://doi.org/10.1109/tcns.2014.2310254. 48

[135] Pastor-Satorras, R. and Vespignani, A. *Evolution and Structure of the Internet*. Cambridge University Press, February 2004. doi: 10.1017/cbo9780511610905. URL https://doi.org/10.1017/cbo9780511610905. 91

[136] Pecora, L. M., Sorrentino, F., Hagerstrom, A. M., Murphy, T. E., and Roy, R. Cluster synchronization and isolated desynchronization in complex networks with symmetries. *Nature Communications*, 5(1), June 2014. doi: 10.1038/ncomms5079. URL https://doi.org/10.1038/ncomms5079. 33, 91

[137] Peron, T., Messias F. de Resende, B., Mata, A. S., Rodrigues, F. A., and Moreno, Y. Onset of synchronization of kuramoto oscillators in scale-free networks. *Phys. Rev. E*, 100:042302, Oct 2019. doi: 10.1103/PhysRevE.100.042302. URL https://link.aps.org/doi/10.1103/PhysRevE.100.042302. 25

[138] Perron, O. Zur theorie der matrices. *Mathematische Annalen*, 64:248–263, 1907. URL http://eudml.org/doc/158317. 4

[139] Pikovsky, A., Rosenblum, M., Kurths, J., and Hilborn, R. C. Synchronization: A universal concept in nonlinear science. *American Journal of Physics*, 70(6):655–655, June 2002. doi: 10.1119/1.1475332. URL https://doi.org/10.1119/1.1475332. 19, 21, 33, 34, 59

[140] Porter, M. A., Onnela, J.-P., and Mucha, P. J. Communities in networks, 2009. 8

[141] Price, D. D. S. A general theory of bibliometric and other cumulative advantage processes. 27(5):292–306, September 1976. doi: 10.1002/asi.4630270505. URL https://doi.org/10.1002/asi.4630270505. 14, 16

[142] Restrepo, J. G., Ott, E., and Hunt, B. R. Onset of synchronization in large networks of coupled oscillators. 71(3), March 2005. doi: 10.1103/physreve.71.036151. URL https://doi.org/10.1103/physreve.71.036151. 25

[143] Rodrigues, F. A., Peron, T. K. D., Ji, P., and Kurths, J. The kuramoto model in complex networks. *Physics Reports*, 610:1–98, January 2016. doi: 10.1016/j.physrep.2015.10.008. URL https://doi.org/10.1016/j.physrep.2015.10.008. 21

[144] Rosell-Tarragó, G. and Díaz-Guilera, A. Functionability in complex networks: Leading nodes for the transition from structural to functional networks through remote asynchronization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(1):013105, January 2020. doi: 10.1063/1.5099621. URL https://doi.org/10.1063/1.5099621. 35, 60, 99

[145] Rosell-Tarragó, G. and Díaz-Guilera, A. Optimal cost tuning of frustration: Achieving desired states in the kuramoto-sakaguchi model. *Physical Review E*, 103(1), January 2021. doi: 10.1103/physreve.103.012216. URL https://doi.org/10.1103/physreve.103.012216. 25, 61

[146] Rosell-Tarragó, G. and Díaz-Guilera, A. Quasi-symmetries in complex networks: a dynamical model approach. *Journal of Complex Networks*, 9(3), June 2021. doi: 10.1093/comnet/cnab025. URL https://doi.org/10.1093/comnet/cnab025. 25, 93

[147] Rosell-Tarragó, G., Cozzo, E., and Díaz-Guilera, A. A complex network framework to model cognition: Unveiling correlation structures from connectivity. *Complexity*, 2018:1–19, July 2018. doi: 10.1155/2018/1918753. URL https://doi.org/10.1155/2018/1918753. 45

[148] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, September

1956. doi: 10.1214/aoms/1177728190. URL https://doi.org/10.1214/aoms/1177728190. 129

[149] Sabidussi, G. The centrality index of a graph. *Psychometrika*, 31(4):581–603, December 1966. doi: 10.1007/bf02289527. URL https://doi.org/10.1007/bf02289527. 12

[150] Sakaguchi, H. and Kuramoto, Y. A soluble active rotater model showing phase transitions via mutual entertainment. *Progress of Theoretical Physics*, 76(3):576–581, September 1986. doi: 10.1143/ptp.76.576. URL https://doi.org/10.1143/ptp.76.576. 25, 33, 35, 53, 59, 92, 98, 136, 144

[151] Salnikov, V., Cassese, D., and Lambiotte, R. Simplicial complexes and complex systems. *European Journal of Physics*, 40(1):014001, nov 2018. doi: 10.1088/1361-6404/aae790. URL https://doi.org/10.1088/1361-6404/aae790. 19

[152] Sánchez-García, R. J. Exploiting symmetry in network analysis. *Communications Physics*, 3(1), May 2020. doi: 10.1038/s42005-020-0345-z. URL https://doi.org/10.1038/s42005-020-0345-z. 91, 94, 99

[153] Schaub, M. T., O'Clery, N., Billeh, Y. N., Delvenne, J.-C., Lambiotte, R., and Barahona, M. Graph partitions and cluster synchronization in networks of oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26 (9):094821, September 2016. doi: 10.1063/1.4961065. URL https://doi.org/10.1063/1.4961065. 95

[154] Schmidt, R., LaFleur, K. J. R., de Reus, M. A., van den Berg, L. H., and van den Heuvel, M. P. Kuramoto model simulation of neural hubs and dynamic synchrony in the human cerebral connectome. *BMC Neuroscience*, 16(1), September 2015. doi: 10.1186/s12868-015-0193-z. URL https://doi.org/10.1186/s12868-015-0193-z. 34

[155] Semsar-Kazerooni, E. and Khorasani, K. Multi-agent team cooperation: A game theory approach. *Automatica*, 45(10):2205–2213, October 2009. doi: 10.1016/j.automatica.2009.06.006. URL https://doi.org/10.1016/j.automatica.2009.06.006. 70, 89

[156] Serrano, M. Á. and Boguñá, M. Clustering in complex networks. i. general formalism. *Physical Review E*, 74(5), November 2006. doi: 10.1103/physreve.74.056114. URL https://doi.org/10.1103/physreve.74.056114. 7

[157] Shen, K., Bezgin, G., Everling, S., and McIntosh, A. R. The virtual macaque brain: A macaque connectome for large-scale network simulations in the-virtualbrain, 2018. URL https://zenodo.org/record/1471588. 112, 130

[158] Skardal, P. S., Taylor, D., and Sun, J. Optimal synchronization of complex networks. *Physical Review Letters*, 113(14), September 2014. doi: 10.1103/physrevlett.113.144101. URL https://doi.org/10.1103/physrevlett.113.144101. 25, 30, 61

[159] Smith, D. and Webb, B. Hidden symmetries in real and theoretical networks. *Physica A: Statistical Mechanics and its Applications*, 514:855–867, January 2019. doi: 10.1016/j.physa.2018.09.131. URL https://doi.org/10.1016/j.physa.2018.09.131. 91, 92, 103

[160] Smith, K. M. On neighbourhood degree sequences of complex networks. *Scientific Reports*, 9(1), June 2019. doi: 10.1038/s41598-019-44907-8. URL https://doi.org/10.1038/s41598-019-44907-8. 6

[161] Solé, R. V. and Bascompte, J. *Self-Organization in Complex Ecosystems. (MPB-42)*. Princeton University Press, December 2006. doi: 10.1515/9781400842933. URL https://doi.org/10.1515/9781400842933. 2

[162] Sorrentino, F. Effects of the network structural properties on its controllability. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(3): 033101, September 2007. doi: 10.1063/1.2743098. URL https://doi.org/10.1063/1.2743098. 48

[163] Sporns, O. *Networks of the Brain*. The MIT Press, 2010. doi: 10.7551/mitpress/8476.001.0001. URL https://doi.org/10.7551/mitpress/8476.001.0001. 34

[164] Stewart, I. Networking opportunity. *Nature*, 427(6975):601–604, February 2004. doi: 10.1038/427601a. URL https://doi.org/10.1038/427601a. 92

[165] Strogatz, S. H. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, September 2000. doi: 10.1016/s0167-2789(00)00094-4. URL https://doi.org/10.1016/s0167-2789(00)00094-4. 25

[166] Strogatz, S. H. Exploring complex networks. *Nature*, 410(6825):268–276, mar 2001. ISSN 0028-0836. doi: 10.1038/35065725. URL http://www.nature.com/doifinder/10.1038/35065725. 2, 91

[167] Stumpf, M. P. H. and Porter, M. A. Critical truths about power laws. 335(6069):665–666, February 2012. doi: 10.1126/science.1216142. URL https://doi.org/10.1126/science.1216142. 14

[168] Tang, E., Giusti, C., Baum, G. L., Gu, S., Pollock, E., Kahn, A. E., Roalf, D. R., Moore, T. M., Ruparel, K., Gur, R. C., Gur, R. E., Satterthwaite, T. D., and Bassett, D. S. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications*, 8(1), November 2017. doi: 10.1038/s41467-017-01254-4. URL https://doi.org/10.1038/s41467-017-01254-4. 48

[169] Terrell, G. R. and Scott, D. W. Variable kernel density estimation. *The Annals of Statistics*, 20(3):1236–1265, 1992. ISSN 00905364. URL http://www.jstor.org/stable/2242011. 129

[170] Thurner, S., Klimek, P., and Hanel, R. *Introduction to the Theory of Complex Systems*. Oxford University Press, November 2018. doi: 10.1093/oso/9780198821939.001.0001. URL https://doi.org/10.1093/oso/9780198821939.001.0001. 3

[171] Travers, J. and Milgram, S. An experimental study of the small world problem. pages 179–197. Elsevier, 1977. doi: 10.1016/b978-0-12-442450-0.50018-3. URL https://doi.org/10.1016/b978-0-12-442450-0.50018-3. 13

[172] Turlach, B. A. Bandwidth selection in kernel density estimation: A review. In *CORE and Institut de Statistique*, 1993. 129

[173] Viti, A., Terzi, A., and Bertolaccini, L. A practical overview on probability distributions. *Journal of Thoracic Disease*, 7:E03–7, 03 2015. doi: 10.3978/j.issn.2072-1439.2015.01.37. 16

[174] Watts, D. J. and Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998. doi: 10.1038/30918. URL https://doi.org/10.1038/30918. 7, 13, 91

[175] wei Huang, D. and neng Huang, W. Traffic signal synchronization. *Physical Review E*, 67(5), May 2003. doi: 10.1103/physreve.67.056124. URL https://doi.org/10.1103/physreve.67.056124. 34

[176] West, D. B. *Introduction to Graph Theory*. Featured Titles for Graph Theory. Prentice Hall, 2001. ISBN 9780130144003. 3

[177] Whalen, A. J., Brennan, S. N., Sauer, T. D., and Schiff, S. J. Observability and controllability of nonlinear networks: The role of symmetry. *Physical Review X*, 5(1), January 2015. doi: 10.1103/physrevx.5.011005. URL https://doi.org/10.1103/physrevx.5.011005. 91

[178] Wiener, N. *Cybernetics: or Control and Communication in the Animal and the Machine*. MIT Press, 2 edition, 1948. 22

[179] Wiley, D. A., Strogatz, S. H., and Girvan, M. The size of the sync basin. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1):015103, March 2006. doi: 10.1063/1.2165594. URL https://doi.org/10.1063/1.2165594. 26

[180] Willms, A. R., Kitanov, P. M., and Langford, W. F. Huygens' clocks revisited. 4(9):170777, September 2017. doi: 10.1098/rsos.170777. URL https://doi.org/10.1098/rsos.170777. 21

[181] Winfree, A. T. Biological rhythms and the behavior of populations of coupled oscillators. 16(1):15–42, July 1967. doi: 10.1016/0022-5193(67)90051-3. URL https://doi.org/10.1016/0022-5193(67)90051-3. 22, 135, 144

[182] Ye, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling*, 53(1-2):91–97, January 2011. doi: 10.1016/j.mcm.2010.07.022. URL https://doi.org/10.1016/j.mcm.2010.07.022. 36

[183] Zee, A. *Fearful Symmetry*. Princeton University Press, December 2007. doi: 10.1515/9781400874507. URL https://doi.org/10.1515/9781400874507. 91

[184] Zhang, Y. and Motter, A. E. Symmetry-independent stability analysis of synchronization patterns. *SIAM Review*, 62(4):817–836, January 2020. doi: 10.1137/19m127358x. URL https://doi.org/10.1137/19m127358x. 60