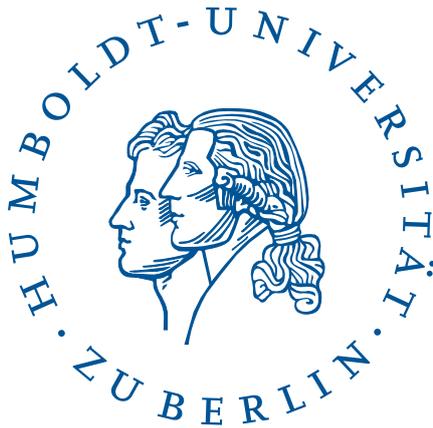


FLORIAN KLIMM

CHARACTERISATION OF
INDIVIDUAL NODES IN
THE MESOSCALE OF
COMPLEX NETWORKS

MASTER'S THESIS



SUPERVISED BY
PROF. DR. DR. H. C. JÜRGEN KURTHS
DR. GORKA ZAMORA-LÓPEZ

Masterarbeit zur Erlangung des akademischen Grades Master of Science im Fach Physik
Vorgelegt der Mathematisch-Naturwissenschaftliche Fakultät der Humboldt-Universität zu Berlin
Angefertigt am Potsdam-Institut für Klimafolgenforschung und in der Arbeitsgruppe Nichtlineare Dy-
namik und Kardiovaskuläre Physik am Institut für Physik der Humboldt-Universität zu Berlin
Eingereicht am 26.8.2014 von Florian Klimm, geboren am 3.5.1989 in Leipzig.

Acknowledgements

I thank Gorka Zamora-López for his outstanding supervision during this thesis, despite the long distance between Barcelona and Berlin. I thank you especially for sharing this interesting topic with me and all the fruitful discussions and advices in Skype sessions, sometimes on daily basis. I thank Jürgen Kurths for supporting me with this thesis and my future goals. I am also grateful to Niels Wessel and Javier Borge-Holthoefer for their input on the manuscript that was the foundation for this thesis. I appreciate the whole research group at the HU and at the PIK being such a friendly and supporting environment. A huge thanks goes to Veronika Stolbova for providing me with the data for the climate networks, discussing all findings in great detail and giving input on the final thesis. I thank Marc Wiedermann, Jakob Kolb, and Benjamin Maier for proofreading parts of this thesis and giving me helpful suggestions and corrections. Last but not least I thank my parents for their enduring support during my studies and my brother Max for being a continuous source of inspiration and motivation.

Abstract

In recent years the analysis of their community structure gave important insights into the functionality of many complex systems. Originating in the Social Sciences this approach has been successfully applied to neuronal networks, human traffic data, and many other fields. However, while analysing this mesoscale structure only minor attention was given to the position, or role, individual nodes present in it. I propose a multidimensional framework that captures this and therefore identifies important nodes. Those tools are then applied to well-known synthetic networks as well as neuronal connectivity data, extreme precipitation measurements, and a traffic network.

To be more specific, in this thesis, firstly former attempts to measure such mesoscale contribution are reviewed. We then define local and global hubness indices to parameterise the relevance of nodes locally within their community and globally on the whole network. Another crucial attribute is the contribution of nodes to the different modules in the network. For analysing that participation vectors are introduced, representing the likelihood of nodes to belong to each community; these account for inhomogenous relative sizes of the communities. Information in the participation vectors is reduced to two scalar indices. The dispersion index characterises how difficult it is to classify a node in one and only one community and the participation index indicates how uniformly the links of a node are distributed among all the communities. Combining those measures leads to a four-dimensional mapping of the role individual nodes have in the mesoscale of networks.

An exemplary graph that illustrates the different contributions a node might have in a network's mesoscale is analysed. The behaviour of well-known network models as regular, random, and scale-free graphs gives further insights. It can be shown that synthetic graphs with intrinsic modular structure do inherit very different behaviour. To model a combination of modular organisation and hub structure I then introduce a new model that shows similar behaviour as the neuronal networks analysed afterwards. Those networks from the *Caenorhabditis elegans*, cat, macaque, and human neuronal systems show striking similarities in terms of the introduced measures, despite being of different nature and size. All four networks show an interplay of segregation and integration, features that are known to be necessary for effective information processing.

I apply those measures to networks constructed from rainfall data of the Indian Summer Monsoon. The community structure analysis shows that modules tend to be shaped by natural topological boundaries as mountain chains in interplay with the onshore movement of moist air masses originating from the Indian Ocean. The further investigation of the role individual sites have in this structure show that North Pakistan is the most dominant region in terms of extreme precipitation synchronisation. This matches

with earlier analysis with other graph theoretical tools and this insight might be used by climatologists for the estimation of monsoon onsets.

Finally I generalise those tools in order to measure the participation of nodes in different layers of a multilayer network. Its utility is demonstrated by applying it to the European Air Transport Network that connects airports with direct flights of different airlines. The analysis shows that airlines are most strongly competing for airports in the holiday regions of southern Europe. Interestingly, major airlines and low cost airlines have striking structural differences that are directly based on the need of the latter to minimise expenses in order to provide a profit while offering affordable tickets to their customers.

To conclude, this thesis introduces new measures for the analysis of individual node's position in the mesoscopic formation of networks. Their behaviour on synthetic graphs is illustrated, followed by the use on networks arising from very different sub-fields of network science. I hope that those tools will be useful for fellow researchers in order to decode the interplay between structure and function in networks.

This thesis is in parts based on the following manuscript:
Florian Klimm, Javier Borge-Holthoefer, Niels Wessel, Jürgen Kurths, and Gorka Zamora-López. Individual node's contribution to the mesoscale of complex networks. *New Journal of Physics* (submitted), 2014

Zusammenfassung

In vielen Zusammenhängen ist es sinnvoll komplexe Systeme als Netzwerke zu repräsentieren und deren Gemeinschaftsstruktur zu untersuchen. Dieser, ursprünglich aus den Sozialwissenschaften stammende, Ansatz wurde in den letzten Jahren erfolgreich auf neuronale Netzwerke, Daten menschlicher Bewegungen und viele weiterer Gebiete angewendet. Jedoch wurde bei der Analyse der Position, oder Rolle, einzelner Knoten in der mesoskalen Struktur nur geringe Aufmerksamkeit geschenkt. Hier führen wir eine multidimensionale Methode ein, die dies untersucht und damit wichtige Knoten identifiziert. Diese Methode wird auf bekannte Netzwerkmodelle, neuronale Netzwerke, Netzwerke extremer Niederschlagsmessung und ein Verkehrsnetzwerk angewandt.

In dieser Arbeit befassen wir uns zunächst mit zwei bereits bekannte Konzepten um die mesoskale Struktur zu analysieren. Dann schlagen wir die folgenden vier Maße vor: *local hubness* misst die Wichtigkeit eines Knoten in seinem eigenen Modul und *global hubness* im gesamten Netzwerk. *Participation* ('Teilnahme') und *Dispersion* ('Verteilung') sind Maße für die Verteilung der Nachbarn eines Knoten in den Modulen des Netzwerkes. Hierfür definieren wir *participation vectors* ('Teilnahmevektoren') deren Elemente die Zugehörigkeit der Knoten zu den verschiedenen Gemeinschaften repräsentieren. Hierbei berücksichtigen wir, anders als bisherige Ansätze, die unterschiedliche Größe der Module. Um die Interpretation zu erleichtern fassen wir die Information der Vektoren in zwei skalaren Größen zusammen. *Dispersion* quantifiziert die Zugehörigkeit eines Knotens zu seinem eigenen Modul und *participation* misst die Verteilung der Nachbarn eines Knotens zwischen allen Modulen des Netzwerkes.

Wir erläutern die Funktion der verschiedenen Indizes auf einem Beispielgraphen. Dann untersuchen wir das Verhalten verschiedener künstlicher Netzwerke wie Zufallsgraphen und skalenfreie Netzwerke. Es kann gezeigt werden, dass sich Graphen mit intrinsischer modularer Struktur deutlich im Verhalten unterscheiden. Um eine Kombination aus modularer und skalenfreier Struktur zu simulieren führen wir ein neues Modell ein. Dies zeigt ähnliches Verhalten wie die neuronalen Netzwerke die wir danach untersuchen. Diese stammen von *Caenorhabditis elegans*, Katze, Makakeaffe und Mensch und haben alle ausgeprägte Gemeinsamkeiten, obwohl sie sich in ihrer Größe und Art unterscheiden. Sie alle zeigen Anzeichen von Segregation und Integration, zwei Konzepte die bekanntermaßen wichtig für die Informationsbearbeitung in neuronalen Systemen sind.

Anschließend wenden wir die entwickelte Methodik auf Klimanetzwerke an und untersuchen damit den extremen Niederschlag während des indischen Sommermonsuns. Die aufgedeckte Gemeinschaftsstruktur ist stark durch natürliche Erhebungen und deren Einfluss auf die Luftbewegungen geprägt. Die genauere Un-

tersuchung einzelner Orte in dieser Struktur offenbart, dass Gebiete in Nordpakistan besonders hohen Einfluss auf die Synchronisation des extremen Regenfalls haben. Dies ist in Übereinstimmung mit bisherigen Erkenntnissen die anderen Methoden der Netzwerkanalyse nutzen. Diese Einsichten sind wichtig für die Untersuchung und Vorhersage des Monsuns.

Im letzten Kapitel werden die vorgestellten Indexe angepasst, um sie auf Netzwerke mit verschiedenen Arten von Kanten anzuwenden (*multilayer networks*). Dies wenden wir auf ein Netzwerk der europäischen Flughäfen an, die durch Flüge verschiedener Airlines verbunden sind. Die Untersuchung zeigt, dass Airlines Flughäfen in den Urlaubsregionen in Südeuropa besonders stark nutzen. Des weiteren zeigen große Fluggesellschaften und Billigflieger strukturelle Unterschiede, da letztere ihre Ausgaben minimieren müssen.

Der zentrale Punkt der Arbeit ist die Definition der neuen Indizes um den Einfluss einzelner Knoten in der modularen Struktur eines Netzwerkes zu quantifizieren. Die Funktion dieser Indices wird zunächst anhand künstlicher Netzwerke verdeutlicht und dann auf reale Netzwerke verschiedenster Bereiche angewendet. Ich hoffe, dass diese Methoden von anderen Wissenschaftler angenommen werden, um Struktur und Funktion in Netzwerken besser zu verstehen.

Contents

| | |
|---|----|
| <i>Introduction</i> | 11 |
| <i>Materials & Methods</i> | 17 |
| <i>Application to synthetic networks</i> | 29 |
| <i>Deciphering segregation and integration in neuronal networks</i> | 43 |
| <i>Climate networks – Investigating the Indian Summer Monsoon</i> | 51 |
| <i>The multilayer Airport Transportation Network</i> | 63 |
| <i>Summary and discussion</i> | 75 |
| <i>Appendix</i> | 79 |
| <i>Bibliography</i> | 83 |
| <i>Selbständigkeitserklärung</i> | 89 |

Introduction

The representation of real systems as complex networks has become a successful practice in the literature across different scientific disciplines: biology, technology, sociology, climatology, etc [2] [3] [4] [5]. Graph analysis allows to describe the topological organisation of the constituents of a multi-component system and uncover their functional implications. There exist a wide variety of graph measures to explore the structural organisation of networks across different scales, from local node properties to global organisation. The intermediate scale, the *mesoscopic* scale, has also received significant attention through investigation of modular structure present in the vast majority of real networks. Most efforts have focused on the determination of the modules, a computational problem commonly referred as the *community detection problem* [6] [7] [8] [9], rather than on describing the mesoscale itself in an informative manner. Besides community detection, in many real networks the nodes are classified into categories based on meta-information about their real-world allocation. Geographically embedded networks such as power-grids, the internet at the autonomous system level or the airport transportation networks can be subdivided into countries or continents. Agents of social networks may be classified according to gender, race, age or any other categories found in social structures. The remaining challenge is then, once a classification is known that segregates nodes into groups, how to characterise the interrelations between them and the function every node takes. A very special type of mesoscale structure is the grouping into different types of interactions, organised in layers. The deciphering of the contribution of individual nodes in such an organisation just recently started to be discussed [10].

The first sound investigation about the analysis of the roles of nodes in relation to the community structure happened in the framework of the social sciences [11] [12] [13]. Using block-models they characterised the roles of cliques in the social structure, e.g. identifying the leaders of two opposing groups in a monastery.

- [2] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010
- [3] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009
- [4] Björn H Junker and Falk Schreiber. *Analysis of biological networks*, volume 2. John Wiley & Sons, 2008
- [5] Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. Complex networks in climate dynamics. *The European Physical Journal-Special Topics*, 174(1):157–179, 2009
- [6] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002
- [7] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010
- [8] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005
- [9] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009
- [10] Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3):032804, 2014
- [11] C W Harrison, S A Boorman, and R L Breiger. Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81:730–780, 1976
- [12] S A Boorman and C W Harrison. Social structure from multiple networks. II. Role structures. *American Journal of Sociology*, 81:1384–1446, 1976
- [13] Francois Lorrain and Harrison C White. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80, 1971

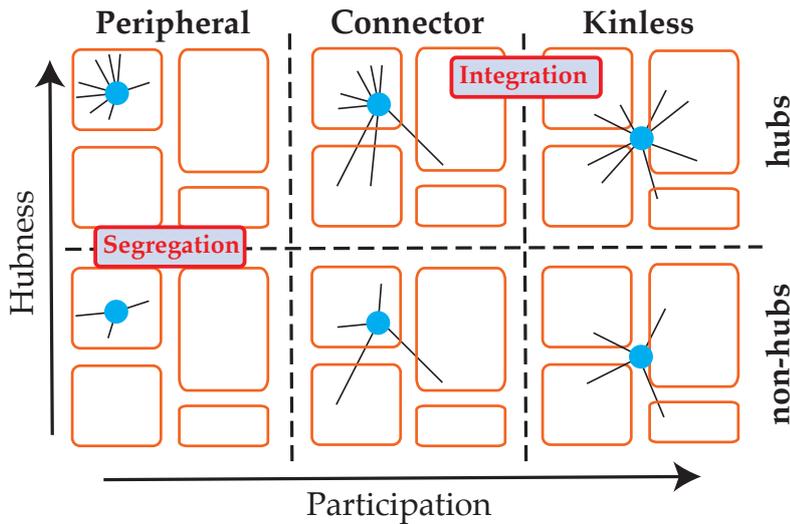


Figure 1: Roles of nodes in the mesoscale can simplified be mapped in a 2-dimensional space: A measure of *participation* (horizontal axis) validates whether a node has neighbours only in its own modules (peripheral), mostly in its own but also in others (connector), or evenly distributed amongst all (kinless). On the vertical measure the importance of a node is measures as *hubness*, the majority of nodes will be non-hubs and only a small amount of them has an exclusive position, making them hubs. In terms of information processing the low participation nodes will allow the segregated handling of information and the connector and kinless hubs will integrate the information from different modules.

- [14] Roger Guimera and Luís A. Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005
- [15] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005
- [16] Alex Arenas, Javier Borge-Holthoefer, Sergio Gomez, and Gorka Zamora-Lopez. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12(5):053009, 2010

- [17] Danielle S. Bassett and ED Bullmore. Small-world brain networks. *The Neuroscientist*, 12(6):512–523, 2006

- [18] Florian Klimm, Danielle S. Bassett, Jean M. Carlson, and Peter J. Mucha. Resolving structural variability in network models and the brain. *PLoS Comput Biol*, 10(3):e1003491, 03 2014

- [19] O. Sporns and G. M. Tononi. Classes of network connectivity and dynamics. *Complexity*, 7(1):28–38, 2001

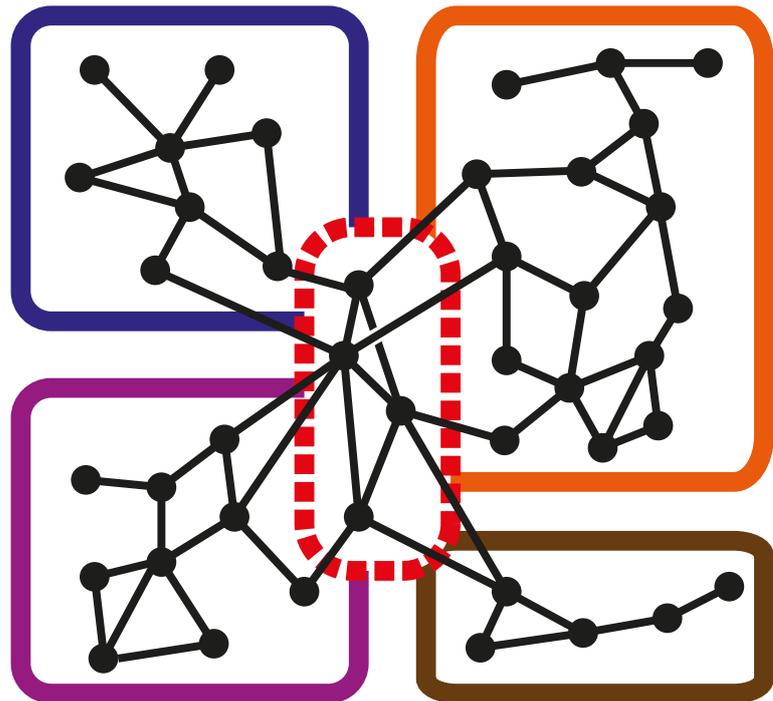
More recently, Guimera & Amaral characterised the roles of nodes according to two parameters: an indicator of the hubness and a participation index assessing how distributed are the links of a node across communities [14] [15]. In another effort, Arenas and coauthors introduced a formalism to study the mesoscale of complex networks based on a dimensionality reduction of the community structure [16].

In this thesis a set of graph descriptors is introduced in order to characterise the position every node takes within the modular and hierarchical architecture of complex networks overcoming limitations of previous approaches. This characterisation is a multidimensional problem and define parameters with a direct interpretation in terms of basic graph properties. Contrary to previous efforts, it is considered that communities are often inhomogeneous and account for the likelihood of nodes to belong to each community depending on their sizes.

Segregation and integration in neuronal networks

For illustration we apply our framework to a set of physiological networks. First, we investigate the neuronal network of the nematode *Caenorhabditis elegans* and the anatomical connectivity between cortical regions in cats, and humans. These are known to inherit a highly complex architecture due to the variety of information processes they host at multiple scales [18]. Therefore graph theory is an appropriate tool for analysing this highly complex structure. A modular and hierarchical structure is known to be important in neuronal networks[17]. We use the representation of node's roles in the modular structure in order to show that the structure of neuronal networks is optimised to process information in a way that combines specialisation and integration [19]. Those two features are known to coexist thanks to the combination of modular differentia-

Figure 2: Segregation and integration are two coexisting features in neuronal networks. Segregation refers to the clustering into distinct modules (solid boxes). These information of different kinds and sensory inputs can be processed in a parallel manner. The different modules can be very different in size and their intrinsic organisation and density since the tasks are diverse as well. However, in order to enable an efficient information processing the information of the different modules has to be combined and filtered in a higher organisational structure. Those are nodes that connect to all modules and serves as integrational centre (dashed red box). Since those nodes connect to different parts of the network they will not be part of a module but show in intermediate position, resulting in a high participation p . Such a modular architecture with an integrating central hub structure optimises short communication ways with specialisation of the modules, and are therefore *small-world* networks. But note that not all small-world networks implicate a modular or even hierarchical organisation. [17]



tion and highly interconnected hubs [21] [22] [23] [20]. Anatomical and functional segregation refers to the existence of specialised neurons, groups of neurons or brain area that have minimised interaction with areas. Those areas have often very specialised tasks and respond to only very distinct input features. Those inputs can be of sensory nature or might already be pre-processed by other preceding modules. The existence of such functionally specialised modules induce also a structural modularity that can be detected by community detection algorithms as indicated in Fig. 2. However, those specialised modules do not operate isolated and therefore a communication between them is necessary. This combination of information takes place in a superordinate structure of hubs and is known as integration. Those two features, integration and segregation, coexist in neuronal systems at a wide range of scales [24]. For example, the primary visual cortex is highly specialised for pattern recognition and consists of a six-layered structure. Those layers are formed by ocular columns which integrate basic neuronal operations as orientation of visual inputs. This indicates that modular and hierarchical architectures are emerging in combination.

Since segregation and integration are contrary features a complex structure is necessary to achieve both. Both being directly connected with the modular structure the community analysis is a suited tool for deciphering both. By measuring the contribution of individual nodes in this organisation we can show that a simple modular organisation does not fulfil the need to integrate information. Although this discussion is inspired by questions arising

[21] Olaf Sporns. Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology*, 23(2):162–171, 2013

[22] G. Zamora-López, C. S. Zhou, and J. Kurths. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.*, 4:1, 2010

[23] Olaf Sporns, Christopher J Honey, and Rolf Kötter. Identification and classification of hubs in brain networks. *PLoS one*, 2(10):e1049, 2007

[20] Gorka Zamora-López, Changsong Zhou, and Jürgen Kurths. Exploring brain function from anatomical connectivity. *Frontiers in neuroscience*, 5, 2011

[24] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013

from neuroscience, the tools are applicable to networks of all types of backgrounds. To illustrate this the mesoscale structure of climate networks and the human made flight traffic network are also analysed.

Earlier attempts

Before presenting our formalism to characterise the roles of nodes in modular networks, we first review two previous efforts.

Mapping of functional roles:

The framework proposed in [14] [15] consists of a mapping of the nodes based on two parameters; one parameter evaluates the internal importance of the node within its module and the other one evaluates its external connectivity. The internal parameter, named as *within-module degree*, is defined as the z-score of the internal degree of the node:

$$z_i = \frac{k'_i - \langle k' \rangle}{\sigma'} \quad (1)$$

where k'_i is the number of links the node makes with members of its own community, $\langle k' \rangle$ denotes the average internal degree of the nodes in this community and σ' their standard deviation. The external parameter, the *participation coefficient*, quantifies how distributed are the links of a node among all the communities:

$$P_i = 1 - \sum_{m=1}^M \left(\frac{k_{im}}{k_i} \right)^2 \quad (2)$$

$P_i = 0$ if all the links of a node fall in the same community and $P_i = 1$ if its links are uniformly distributed among all the modules. Following these two parameters, the nodes of a network can be mapped into a z - P plane in which nodes of different roles occupy distinguished regions, similar to those illustrated in Fig. 1. This framework suffers from a few shortcomings that limit its application and the interpretation of the results obtained: (i) Nodes are classified as hubs or non-hubs based exclusively on the information of their internal degrees when in the literature the term hub is associated with the degree a node, which comprises the whole network. (ii) All communities are assumed to be characterised by identical statistical properties. For example, if two modules had different internal degree distributions the z_i of nodes in them is evaluated over different statistical baselines. On the other hand, the formulation of the participation index assumes that all communities are of the same size. (iii) The largest value of P_i depends on the number of communities. In a partition with M communities the largest participation a node can take is $P_i = 1 - \frac{1}{M} < 1$ when according to [14], it should be $P_i = 1$. As a consequence P_i values across networks are only comparable if both networks contain the same number of communities.

[14] Roger Guimera and Luís A. Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005

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

[14] Roger Guimera and Luís A. Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005

Algebraic approach to characterise the modular skeleton:

The approach introduced in [16] is based on the linear decomposition of the modular organisation of a network. It accounts for the multidimensional nature of the problem to classify nodes. Given a partition \mathcal{L} of the network into M communities, the *contribution matrix* \mathbf{C} is defined as the number of links each node devotes to every community, $C_{is} = k_{is}$. The contribution matrix contains all the information of the modular structure and the challenge is to extract this information in a structured manner. This is achieved by application of singular value decomposition (SVD) to the \mathbf{C} matrix and investigating its principal directions. In this representation every module has its own intrinsic direction $\tilde{\mathbf{e}}_s$ in the M -dimensional space corresponding to the direction of those nodes that are only internally connected and make no external connections with other modules. The deviation of a node's projection $\tilde{\mathbf{n}}_i$ from $\tilde{\mathbf{e}}_s$ represents its tendency to connect with other modules. Equivalently, the deviation of the sum of a module's nodes projections $\tilde{\mathbf{m}}_s$ represents the tendency of the module C_s to connect to other modules; $\tilde{\mathbf{m}}_s = \tilde{\mathbf{e}}_s$ only when the module is disconnected from the others. On the other hand, the scalar product of the modular projections $\tilde{\mathbf{m}}_s$ reflects the relationships between modules allowing to investigate how modules are interrelated. Despite its elegance and its success to map the skeleton formed by the communities, the information for individual nodes is more difficult to extract and differentiate. As it usually happens with linear decomposition methods the interpretation of the optimal dimensions and their projections in terms of the natural parameters of the system is not trivial due to the mixture of information.

Altogether, we find that the functional mapping formalism lacks of universal criteria and the parameters given by the algebraic approach are difficult to interpret node-wise. In the following we introduce a formalism based on four parameters, local and global, whose combination leads to a rich understanding of the roles that nodes play within modular networks. Our approach aims to be universal such that all networks are comparable within the same criteria. Additionally, it recognises the fact that the probability of a node to connect to a community depends on the relative sizes of the communities.

[16] Alex Arenas, Javier Borge-Holthoefer, Sergio Gomez, and Gorka Zamora-Lopez. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12(5):053009, 2010

Materials & Methods

Networks and graphs

Foundations

In this work we use a framework from the discrete mathematics that is called *graph* or *network*. This field was introduced by Leonard Euler in 1753 in his infamous work on the *Seven Bridges of Königsberg* [25] and reached further attention at the end of the 20th century from a wide range of disciplines. This includes the Physical Sciences, Sociology, Biology, Climatology and many others. Due to this broad applicability also a wide range of network measures and metrics were introduced and successfully utilised in the last years. Here we will introduce only the necessary ones and we refer to [2] for a comprehensive and to [26] for a mathematically stricter introduction.

In the simplest sense a *graph* is an ordered pair $G = (V, E)$ composed of a set V of vertices (nodes) that are connected via edges (links) from the set E . The size of the network is determined by the number of nodes $n = |V|$ and the number of edges $m = |E|$. Nodes that are connected to each other via an edge are called *neighbours*, the union of all neighbours of a node u is its *neighbourhood* $N(u)$. A graph without nodes connected with themselves (self loops) and multiple connections between a pair of nodes (multiedges) is called a *simple graph* to which we will restrict our discussion if not otherwise stated. The most common representation of a graph is the *adjacency matrix* which is an $n \times n$ matrix with the elements

$$A_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ connected} \\ 0 & \text{else} \end{cases} \quad (3)$$

Accordingly this is a binary matrix. In more general frameworks the network's edges might be given an individual weight $w_i \in \mathbb{R}$. Furthermore since an edge connects two vertices the matrix is symmetric $A_{ij} = A_{ji}$. In *directed* network edges have a starting and end vertex, which gives rise to a non-symmetric adjacency matrix.

The maximal number of edges in a simple graph is $\binom{n}{2} = \frac{1}{2}n(n-1)$ and a network with this number of edges is called the *complete graph* on n K_n . We then define the density ρ of a graph to be the

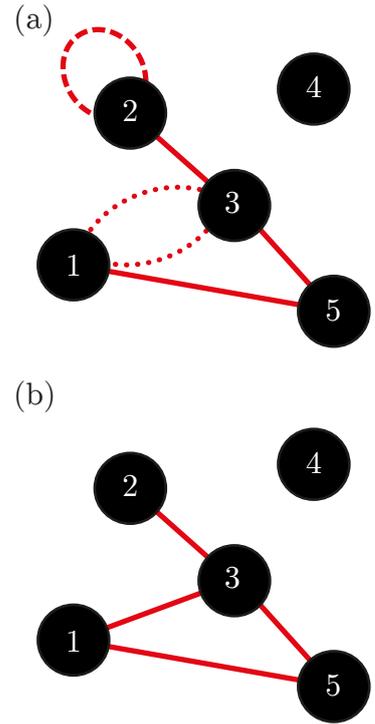


Figure 3: Panel (a): A graph consisting of $n = 5$ nodes (black dots) that are connected with $m' = 6$ edges (red lines). Panel (b) shows the simple graph where self-loop (dashed line) is erased and the multiedges (dotted lines) collapsed into a single one and the number of edges becomes $m = 4$. The neighbourhood of node 1 is given by $N(1) = \{3, 5\}$.

[25] Leonhard Euler. *Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741

[2] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010

[26] Geir Agnarsson and Raymond Greenlaw. *Graph Theory: Modeling, Applications, and Algorithms*. Prentice-Hall, Inc., 2006

The adjacency matrix for the simple graph shown in (3b) is given by the 5×5 binary matrix

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

The edge list for the same network is

$$E = \{(1,3), (1,5), (2,3), (3,5)\}.$$

fraction of edges that are present of all possible

$$\rho = \frac{2m}{n(n-1)}. \quad (4)$$

Graphs with a low density are called *sparse* and graphs with a high density are referred to as *dense*¹. Another graph representation that is especially useful for such sparse graphs is an *edge list* and is widely used to increase computational performance, in particular in terms of memory use.

Many properties of graphs must be formulated in terms of numerical values. The first such attribute we define is the *degree* k of a node and is the number of edge ends that connect to an edge. It can easily be calculated from the adjacency matrix

$$k_i = \sum_{j=1}^n A_{ij}. \quad (5)$$

When summing up the degrees of all nodes we receive twice the number of edges since each edge is connected to exactly two nodes

$$\sum_{i \in V} k_i = 2m, \quad (6)$$

what is known as the *Hand-Shaking-Theorem*. The mean degree is then

$$\langle k \rangle = \frac{\sum_{i \in V} k_i}{n} = \frac{2m}{n}. \quad (7)$$

A *path* in a network is a route from one node in a network to another along the edges. More strictly speaking it is any sequence of vertices such that consecutive pairs of vertices are neighbours. In (4) two of such paths are shown. Both start at vertex 2 and end at vertex 5. They differ in the amount of edges crossed, what defines its *length* l . Therefore the blue one has a $l = 2$ and the longer one of $l = 3$.

The calculation of the number of paths between a pair of nodes in a network is received by iterative multiplication of the previously introduced adjacency matrix. The entry $A_{ij}^{(k)}$ in A^k is the number of distinct walks from node i to node j of length k . Note that this holds for $k = 1$ since the adjacency matrix has non-zero entries for those nodes that are neighbours. The *distance matrix* D is the $n \times n$ matrix where each entry d_{ij} gives the geodesic distance between node i to j . In an undirected graph $d_{ij} = d_{ji}$ and the distance matrix therefore symmetric $D = D^T$.

A graph is *connected* if every pair of distinct vertices can be connected via a path. If this is not the case the graph can be divided into disjoint subgraphs that are connected themselves, we call those subgraphs *components*. Components that consist of only a single node are termed *isolated nodes*.

The *geodesic distance* d between a pair of nodes (i, j) defined as the length of the shortest path between those nodes. Applying the

¹ Strictly spoken sparsity is only distinguishable from dense graph in the limit $n \rightarrow \infty$ and is given if $\rho \rightarrow 0$, in the case of a finite limit the network is considered dense.

| node | 1 | 2 | 3 | 4 | 5 | Σ | $\langle k \rangle$ |
|------------|---|---|---|---|---|----------|---------------------|
| degree k | 2 | 1 | 3 | 0 | 2 | 8 | $8/5$ |

Table 1: Node's degree for network shown in (3b). The sum over all node's degrees is equal to twice the number of edges of the graph according to the Hand-Shaking-Theorem and the mean degree $\langle k \rangle = 1.6$ since the network consists of five nodes.

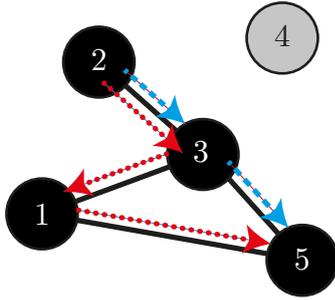


Figure 4: Two paths in the graph (3b), both from node 2 to node 5: The red (dotted) arrows indicate the path $(2 \rightarrow 3 \rightarrow 1 \rightarrow 5)$ of length three. The blue (dashed) arrows indicate the path $(2 \rightarrow 3 \rightarrow 5)$ of length 2 that is also the shortest path between nodes 2 and 3, their geodesic distance is therefore 2. Note that all nodes except node 4 can be connected via paths to each other, therefore the graph consists of two connected components $\{1,2,3,5\}$ (black nodes) and $\{4\}$ (grey node).

The number of distinct paths of length $l = 2$ of the network shown in (4) is given by

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

and paths with $l = 3$ by

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

and the paths between nodes 2 and 5 are highlighted.

iterative multiplication of the adjacency matrix we know that is given by the smallest value of r such that $A_{ij}^r > 0$. All nodes belonging to the same component have a finite distance to each other, to nodes in other components the distance is ∞ . The length of the longest geodesic path between any pair of nodes is the *diameter* of a graph. In the case of a multi-component graph often the components are analysed separately and the infinite distances ignored. This is also the case for calculating the *average path length* L that is defined as the mean distance between all pairs of nodes in a network

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}. \quad (8)$$

Closely connected is the measure of a networks *global efficiency* E [27], computed as the average of the inverse of the distance matrix

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (9)$$

The extreme cases are given by a graph of isolated nodes that has an efficiency of $E = 0$ since all nodes have infinite distances towards each other, a fully connected graph on the other hand leads to an efficiency of 1.

A node's *local clustering coefficient* C_i is a measurement of how densely its neighbours are connected with each other [28]. Suppose node i has k_i neighbours then there are at most $\frac{k_i(k_i-1)}{2}$ undirected edges between each other. The clustering coefficient is then defined as the fraction of existing links between neighbours to this maximal possible number

$$C_i = \frac{2 \cdot \# \text{ of edges between } i\text{'s neighbours}}{k_i(k_i - 1)}. \quad (10)$$

The *global clustering coefficient* C is then a property of the whole network and defined as the average over all node's individual local clustering

$$C = \langle C_i \rangle = \frac{1}{n} \sum_{i=1}^n C_i. \quad (11)$$

It is known that the global clustering is dominated by nodes with small degree's. To overcome this the global *transitivity* can be used instead. It is the ratio of connected triads to the number of potential ones in the whole graph.

Community detection – Network of networks

The above metrics can often be characterised in terms of the scale they investigate the graph Fig. 5: *local measures* as the degree k or clustering c take into account only the direct neighbourhood of each vertex. On the other side *global measures* as the betweenness g need information on the whole network (all shortest path lengths)

The pairwise distance between all nodes in the network shown in (4) is given by the distance matrix

$$D = \begin{pmatrix} 0 & \mathbf{2} & \mathbf{1} & \infty & \mathbf{1} \\ \mathbf{2} & 0 & \mathbf{1} & \infty & \mathbf{2} \\ \mathbf{1} & \mathbf{1} & 0 & \infty & \mathbf{1} \\ \infty & \infty & \infty & 0 & \infty \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & \infty & 0 \end{pmatrix}$$

thus the graph consists of two components, one of diameter 2 and size 4 and the other being an isolated node. The average path length of the larger component (mean of bold values) is $\frac{16}{43} = \frac{4}{3}$ and the efficiency of the whole graph $E = \frac{1}{2}$.

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

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

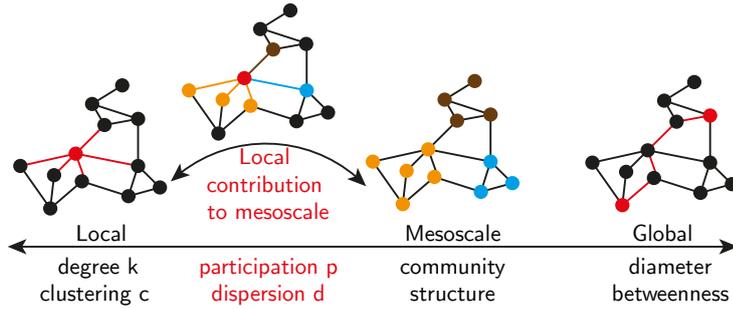


Figure 5: Metrics to analyse networks can access a wide range of scales: Ranging from local measures as the degree k of a node up to global metrics like the betweenness that need full information of the network. In the recent years the mesoscale or *community structure* has been analysed and shown to be of importance. Here we will introduce measures that bridge from the local measures to the mesoscale, e.g. how are a node's neighbours are distributed in the community structure.

- [6] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002
- [7] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010
- [29] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009
- [30] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977
- [31] Anna Lewis, Nick Jones, Mason Porter, and Charlotte Deane. The function of communities in protein interaction networks at multiple scales. *BMC systems biology*, 4(1):100, 2010
- [32] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008
- [33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008
- [34] Inderjit S. Jutla, Lucas G. S. Jeub, and Peter J. Mucha. A generalized louvain method for community detection implemented in matlab. <http://netwiki.amath.unc.edu/GenLouvain>, 2011-2012
- [35] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006

in order to validate each node's centrality. In the recent years the investigation of the networks organisation in between those extreme scales of observation has become an important tool. Such *mesoscale* analysis detects the *modular* or *community* structure of graphs. Such a community is a group of nodes that are relatively densely connected with each other but sparsely with other of such dense groups in the network [6] [7] [29].

Such grouping of single agents in groups originates from the Social Sciences and Fig. 6 shows a famous benchmark model [30]. In the recent years the detection of such a modular structure has been adopted by many different disciplines, from biological systems, e.g., protein interaction structure or collaboration [31] networks between scientists [32].

There are many different methods for detecting such a modular structure. Furthermore the measurement of the significance of different detected groupings is still an open discussion among experts. Here we will use the *Louvain*-method [33] [34] to extract the partitioning. As most community detection methods it tries to find a partition that optimises the modularity quality function

$$Q = \frac{1}{4M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) \delta_{c(i),c(j)}. \quad (12)$$

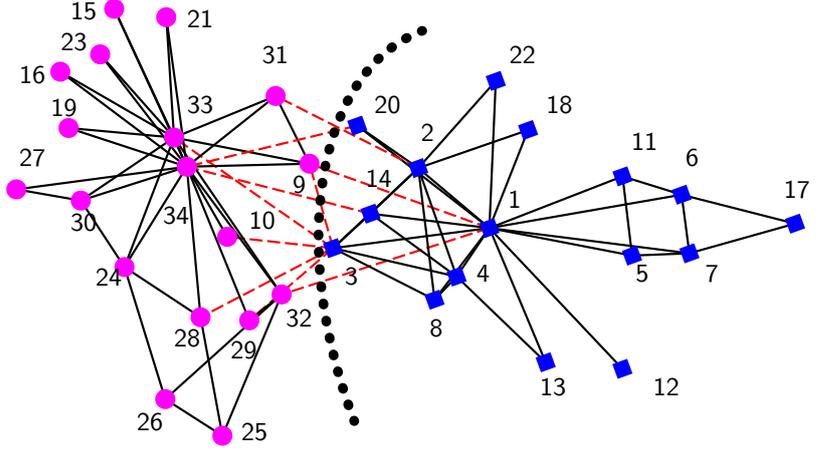
The modularity $Q \in [-1/2, 1)$ evaluates the goodness of a partition comparing the actual fraction of links falling within the modules to the expected fraction if links were distributed at random conserving the degrees of the nodes [35]. Positive values indicate stronger communities than expected by chance, with $Q \rightarrow 1$ for perfectly partitioned graph topologies.

Community detection algorithms are often nondeterministic. In order to compare modules from different runs the *Jaccard index* can be used. It is a statistic for comparing the similarity and diversity of two sets A and B

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (13)$$

In our case A and B are modules from different partitions of the same network.

Figure 6: The Zachary karate club network network shows $n = 34$ member of an University sports club. They are connected by one of the $m = 78$ edges if they are friends. The node's community affiliation is shown with pink circles and blue squares. After a disagreement the club split up into two competing clubs (indicated by dotted line) that are exactly predicted by the Louvain algorithm. Note that only 10 edges are connecting members of different groups (red dashed lines) and a majority of the friendships are inside the modules, leading to a relatively high modularity of $Q \approx 0.62$.



Participation – Bridging from community detection to local measures

Following we want to use the mesoscale structure in combination with local measures in order to get an in-depth picture of the networks modular organisation. As illustrated in Fig. 5 we are therefore interested in each node's local neighbourhood but do not simply count its magnitude as for the degree but take into account each neighbour's community association. We therefore will be able to measure the distribution of a node's links amongst all modules. In order to do so we introduce the concept of *participation vector* \mathbf{P}_i whose elements P_{im} represent the probability that node i belongs to community C_m , where $m = 1, 2, \dots, M$. This probability is given by $P_{im} = \frac{k_{im}}{N_m}$ where N_m is the size of the community. Since we are only interested in the relative differences participation vectors are normalised such that $\sum_{m=1}^M P_{im} = 1$. Otherwise the norm of \mathbf{P}_i would be proportional to the degree k_i but this information shall be disentangled and captured only by the hubness index. The vector of a node devoting all its links to the second community of a network with $M = 4$ communities is $\mathbf{P}_i = (0, 1, 0, 0)$ and for a node whose links are all equally likely distributed among the four communities, $\mathbf{P}_i = (1/4, 1/4, 1/4, 1/4)$.

Once the participation vectors have been computed for all nodes, we want to reduce that information into scalar values to quantify how distributed are the links of a node among all the communities. For consistency with previous definitions $p_i = 0$ if the node devotes all its links to a single community and $p_i = 1$ if its links are equally likely distributed among *all* the modules. Therefore we evaluate the standard deviation $\sigma(\mathbf{P}_i)$ of the elements of the participation vector \mathbf{P}_i and define the participation index as:

$$p_i = 1 - \frac{\sigma(\mathbf{P}_i)}{\sigma_{\max}(M)} = 1 - \frac{M}{\sqrt{M-1}} \sigma(\mathbf{P}_i). \quad (14)$$

The normalisation factor accounts for the fact that the standard deviation of an M dimensional vector with all elements equal to

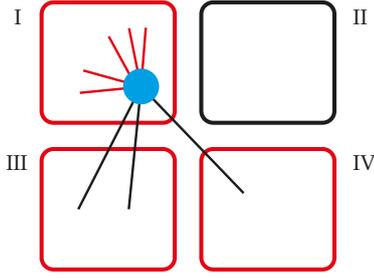


Figure 7: Schematic view of the node centric participation or *dispersion* d . The node itself is embedded in the module I and has furthermore connections to the modules III and IV but not to II. Thus for the node centric participation index the potential connection to sub-network II can be neglected. Furthermore for the intra-module degree of a node only the number of neighbours in the same module as the node is considered.

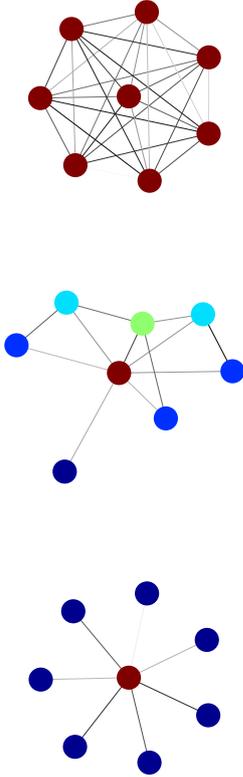


Figure 8: Three example networks of eight nodes with the central node having degree seven. In the uppermost one (full graph) all nodes have this degree, in the lowest one (star graph) all other nodes have degree one. In reality most networks will lie in between (middle panel). Color indicates degree (from 1 = dark blue to 7 = red). This shows that the degree of a node alone is not sufficient for the determination of its importance but as to be compared with those of the others (as with the standard score) or a null model (as the hubness).

zero but one is $\sigma_{max}(M) = \sqrt{M-1} / M$.

Dispersion – The node-centric participation

The investigation of a nodes connectivity to all modules in a network is a somewhat global perspective. It also can be interesting to analyse the distribution of neighbours from a node-centric perspective and take into account only modules it is connected with as indicated in Fig. 7.

The larger its p_i the more difficult it is to classify the node into one and only one community. But for a node to be difficult to classify it is not necessary that it connects to all the communities. If a node is equally likely connected with only two of the communities, it is also difficult to classify. Such a node has for example a participation vector $\mathbf{P}_i = (0, 1/2, 0, 1/2)$. To account for this information we define the dispersion of a node d_i equivalent to the participation index but considering only the non-zero entries of the participation vector:

$$d_i = 1 - \frac{\sigma(\mathbf{P}'_i)}{\sigma_{max}(M')} = 1 - \frac{M'}{\sqrt{M'-1}} \sigma(\mathbf{P}'_i), \quad (15)$$

where \mathbf{P}'_i is the subvector containing only the non-zero entries of \mathbf{P}_i , and M' its dimension. A node only connected within one community has $d_i = 0$ and a node equally likely connected among $M' \leq M$ has $d_i = 1$. We note that in general $d_i \geq p_i$ with the equality only happening when $M' = M$. The dispersion index shall be regarded as a measure of how difficult it is to classify a node into only one community and the participation index as the global reach of a node's links among all the communities. Furthermore we want to take into account that nodes are more likely to connect with larger modules. Imagine a network with only two communities, one of size $N_1 = 20$ and another of size $N_2 = 40$. A node devoting ten links to each of them is contributing more to the small community than to the large one and therefore, it is more likely to belong to the small community. We note also that not all combinations of dispersion and participation are possible. In fact for a node with dispersion d and number of connected communities M_c the participation is derived by the closed form

$$p(d, M, M_c) = 1 - \sqrt{\frac{M}{M-1} \left(\frac{1}{M_c} ((M_c - 1)(1 - d)^2 + 1) - \frac{1}{M} \right)}. \quad (16)$$

The derivation uses the normalisation of the participation vector \mathbf{P} and is shown in the appendix. The highest difference between a node's d and p is reached when $M_c = 2$ and follows:

$$p^+(d, M) = p(d, M, 2) = 1 - \sqrt{\frac{M}{M-1} \left(\frac{1}{2} ((1 - d)^2 + 1) - \frac{1}{M} \right)}. \quad (17)$$

Hubness

One of the most important features found in real networks is the presence of highly connected nodes or hubs. Despite their impor-

tance a formal definition of hubness is missing; they are colloquially defined as “those nodes with many more connections than others”. The hallmark of networks containing hubs, scale-free networks, possess a power-law degree distribution meaning that most nodes make only few links and a few nodes have many connections. The degrees of hubs is usually orders of magnitude larger than that of the sparsely connected nodes. The within-module degree in Equation (1) is intended to capture this notion of hubness by estimating the significance of a node’s degree compared to that of its neighbours.

Now, we note a paradoxical behaviour in the hubs of scale-free networks: despite having much larger degree than other nodes, hubs are only connected to a small fraction of the network. This happens because typically studied scale-free networks, both synthetic and real, are sparse. For example, in a scale-free network with $N = 100\,000$ nodes and exponent $\gamma = 3$ the most connected hubs are linked with only 2% – 5% of all nodes. In contrast, many real networks are dense and therefore the loose definition of hubness above does not comply. The same applies to the internal connectivity of the modules which are, by definition, densely connected parts of a network. In dense (sub-)networks the mean degree is of the same order of magnitude as its size, $\langle k \rangle \sim N$, and therefore the range of differences between the least and the most connected nodes is considerably reduced. Also, contrary to the hubs of scale-free networks, the hubs of dense networks are connected to a large fraction of the nodes. In dense networks it is usual to find hubs connected from 30 % up to 70 % of the network.

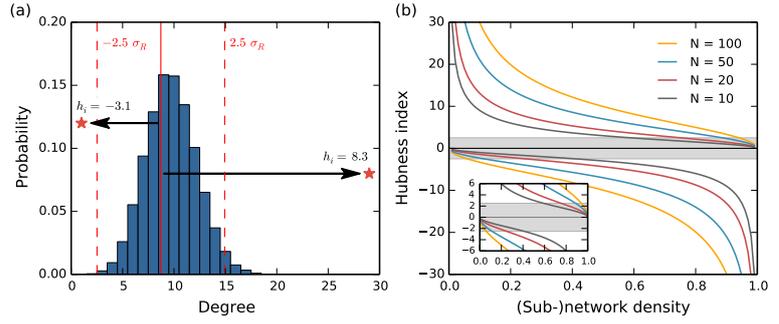
We find that for a measure of hubness to be universal it requires the assessment of the degrees under a common statistical baseline. Therefore, we define the hubness index h_i of a node as the comparison of its degree k_i with the degree distribution of an equivalent random graph of the same size N and density ρ . Since random graphs possess a narrow degree distribution around their mean degree $\langle k \rangle_R$, the more a node’s degree deviates from the degree distribution of an equivalent random graph, the more reasonable it is to be considered as a hub. It is well known that the mean degree of a random graph is $\langle k \rangle_R = (N - 1)\rho$ and the standard deviation of its degree distribution is $\sigma_R = \sqrt{(N - 1)\rho(1 - \rho)}$ [2]. Constraining to networks without self-loops, we define the hubness index of a node in a network of size N and density ρ as:

$$h_i = \frac{k_i - \langle k \rangle_R}{\sigma_R} = \frac{k_i - (N - 1)\rho}{\sqrt{(N - 1)\rho(1 - \rho)}}. \quad (18)$$

The hubness is negative for nodes with $k_i < \langle k \rangle_R$ allowing to identify also outliers that are significantly less connected than expected from randomness. This index differs from the one defined in Equation (1) in that all networks are compared to the same null-model – the random graph – instead of using the internal statistics of each network and of each community as a baseline for itself. This makes

[2] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010

Figure 9: Dependence of the boundaries of hubness index on the size N and density ρ of the network. (a) Illustrative case for networks of $N = 30$ and $\rho = 0.25$. Due to the position of the degree distribution of random graphs with N and ρ , nodes can achieve different extremal values of significance. A node with the smallest degree $k_i = 1$ will take much smaller hubness than a node with the largest degree $k_i = 29$. (b) Upper and lower boundaries of hubness for networks of different size with increasing density. Inset, shows the zoom near the thresholds $h_i \pm 2.5$.



the index universal and the results for different networks and modules comparable. Once the hubness index is defined we need to establish guidelines for interpreting the results and to identify its extremal values. In a random graph the degree of 99% of the nodes lies in the range bounded by $\langle k \rangle_R \pm 2.5 \sigma_R$. In any network, nodes within this range shall not be considered outliers. The larger the h_i index grows above 2.5σ the more likely is the node a hub. The largest and smallest values that h_i can take depend on the size and the density of the network. In sparse networks it is not possible to find outliers on the negative side because $0 < \langle k \rangle \ll N$ and even disconnected nodes lie in the range of statistically expected degrees. However, there is plenty of room for some nodes to have degree much larger than the mean and achieve high significance. Fig. 9(a) illustrates this for a network with $N = 30$ and $\rho = 0.25$. A node with degree $k_i = 1$ gives $h_i = -3.1$ while a hub with $k_i = 29$ achieves $h_i = 8.3$. As the density increases the situation is reversed because $\langle k \rangle \rightarrow N$. The room for hubs decreases limited by the size of the network while from the bottom, occasional nodes with few connections become very significant. The hubness of a node with degree $k_i = 0$ is:

$$h^-(N, \rho) = -\frac{\langle k \rangle_R}{\sigma_R} = -\sqrt{\frac{(N-1)\rho}{1-\rho}} \quad (19)$$

and for a node with degree $k_i = N - 1$:

$$h^+(N, \rho) = \sqrt{\frac{(N-1)(1-\rho)}{\rho}} \quad (20)$$

Equations (19) and (20) represent the lower and upper boundaries for the hubness index. In Figure 9(b) the boundaries are shown for networks of different size and varying density. When networks are sparse the lower limit $h^- > -2.5 \sigma$ regardless the size N and the upper bound tends to infinity. When networks are very dense, the opposite happens, the lower bound goes to $-\infty$ while there is no room for significant hubs because $h^+ < 2.5 \sigma$. As highlighted in the inset, the limitations are stronger for smaller networks. For example, in networks of $N = 10$ there is no regime of density in

which nodes with significant low and high degree can be found simultaneously.

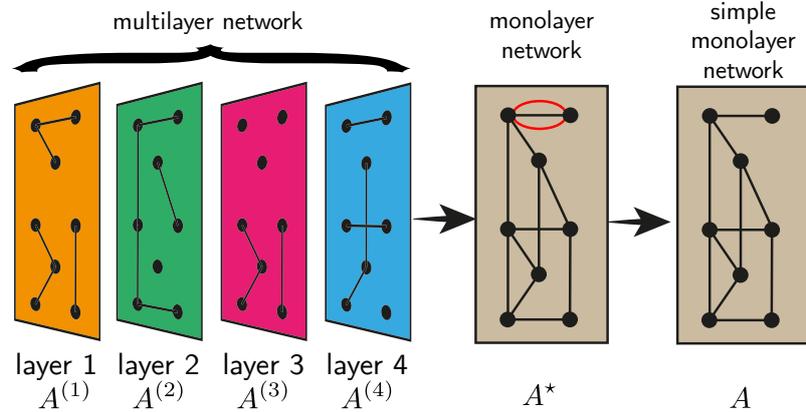
From now on we will characterise nodes by two hubness values: the *global hubness* $h^{(g)}$ is the hubness of the node in the network and the *local hubness* $h^{(l)}$ its hubness within the community it belongs to. In this case N and ρ are replaced by the size and density of the community N' and ρ' .

Role detection

We introduced four different indexes describing role of nodes: Participation p as a measure of the accessibility of all modules by a node. Dispersion d quantifies a nodes' affiliation with the modules it connects to. Local hubness validates the nodes importance inside its own module and global hubness in comparison with nodes of the whole network. Illustrating this four-dimensional space is not straightforward. Therefore we mostly discuss scatter or density plots that show the dependence and distribution of two out of those four indexes. For example, as illustrated in Fig. 1 the $(p, h^{(g)})$ allows us to investigate the role of nodes in terms of global importance, called hubs or non-hubs, in combination of their mesoscale position, referred to as ultra-peripheral, peripheral, connector, or kinless. The distinction between the different participation classifications is not always straightforward. Usually we divide the participation into three regimes of equal length $(0, 1/3]$ (peripheral), $(1/3, 2/3]$ (connector) and $(2/3, 3]$ (kinless), as well as the extreme case 0 (ultra-peripheral).

Multilayer networks

Figure 10: Multilayer networks consist of multiple layers of different types of edges that connect the nodes. Despite occurring frequently in real world data the multilayer networks are often merged to a monolayer adjacency matrix A^* for simplicity. Furthermore the multiple edges (red lines) are then removed to end up with a simple graph's adjacency matrix A . Such reduced representations do not distinguish between the different types of connectivity that the edges represent but can nevertheless give important insight in the network's structure. Note that the ordering of the layers is mostly arbitrary in multilayer networks, however in temporal networks that can be represented in a multilayer fashion the ordering is of fundamental importance.



Until now we exclusively explored networks in which all edges are given an equivalent footing. Some real-world networks consist of interconnected layers of different types of edges, and those are called *Multilayer networks* [36] [37]. We can think of those different links as different ways of transportation (e. g. buses, metro, and trains) between stations or different kind of social interactions (e. g. social media friendship, real-world friendship, and co-workership) between individuals. The interplay of such diverse structures is a current topic in the mathematical [36], as well as in the social sciences [38].

Until recently the different layers of a network were condensed into a single mono-layer in order to simplify the data analysis, as illustrated in Fig. 10. The monolayer can be further reduced to a simple graph by merging the multiple edges. Neuronal networks for example consist of two types of synapses between the neurons [39]: Chemical synapses as well as gap junctions. Both are junctions that enable neurons to exchange information so the reduction to a monolayer seems reasonable. Nevertheless they are fundamentally structurally and functionally different as chemical synapses can either be excitatory or inhibitory, whereas gap junctions are always excitatory since there is a direct pulse transfer from one neuron to the other.

To be more specific we represent a multilayer network by a sequence of graphs $\{G^\alpha\}_{\alpha=1}^b$, with each layer $G^\alpha = (V, E^\alpha)$ being indexed by $\alpha \in \{1, \dots, b\}$ and b being the number of layers [40]. Note that in our framework each layer has the same set of nodes V , however in more general situations as *multiplex networks* this can be replaced by an individual set V^α . Note that this formalism is similar to an *edge colouring* of a monolayer graph $G = (V, E)$ that allows multiple edges. The straightforward representation of such a multilayer graph is an adjacency tensor $A_{ij}^{(\alpha)}$ with $n \times n \times b$ elements [36].

We now can use the *a priori* information of each edges layer affil-

[36] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013

[37] Mikko Kivelä, Alexandre Arenas, Marc Barthélemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *arXiv preprint arXiv:1309.7233*, 2013

[36] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013

[38] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. Measuring large-scale social networks with high resolution. *PloS one*, 9(4):e95978, 2014

[39] Dominik Traxl. C. elegans - Neural structure & dynamics. Master's thesis, Ludwig-Maximilians-Universität München, 2012

[40] Emanuele Cozzo, Mikko Kivelä, Manlio De Domenico, Albert Solé, Alex Arenas, Sergio Gómez, Mason A Porter, and Yamir Moreno. Clustering coefficients in multiplex networks. *arXiv preprint arXiv:1307.6780*, 2013

[36] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013

iation in order to define a multilayer participation $p^{(ml)}$ that measures how a node's neighbours are distributed amongst different layers. Receiving non-normalised participation vector elements $k^{(ml)}$ for each node i and layer α is straightforward as we simply count the number of edges of each layer attach to a node, similar to the degree

$$k_{i,\alpha}^{(ml)} = \sum_{j=1}^n A_{ij}^{(\alpha)}. \quad (21)$$

However, by definition all layers have the same number of nodes n and therefore do not allow normalisation in the same manner as introduced for the community sizes. To overcome this we quantify the size of each layer as the number of nodes that do have at least one neighbour in this layer $N_\alpha = |\{V : k_{i,\alpha}^{(ml)} > 0\}|$ and define the elements of the *multilayer participation vector* as

$$p_{i,\alpha}^{(ml)} = k_{i,\alpha}^{(ml)} / N_\alpha. \quad (22)$$

We now can treat the multilayer participation vector in same way as the monolayer version in order to get a multilayer participation $p^{(ml)}$ and dispersion $d^{(ml)}$. As indicated in Fig. 11 minimal value of multilayer participation and dispersion $p^{(ml)} = d^{(ml)} = 0$ show that a node is exclusively connected in a single layer. High values on the other hand indicated that a node's neighbours are equally distributed amongst all layers.

Last but not least we want to introduce the *multilayer degree* $k^{(ml)}$ as the number of neighbours a node does have amongst all layers

$$k_i^{(ml)} = \sum_{\alpha=1}^b k_{i,\alpha}^{(ml)}. \quad (23)$$

Note that this only equivalent to the degree of a node in the collapsed monolayer network if multiple edges were not merged into unweighted ones.

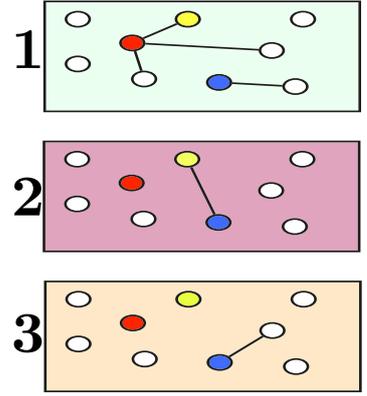


Figure 11: Example multilayer graph consisting of three layers (1,2, and 3) with sizes N_α of $N_1 = 6$, $N_2 = 2$, and $N_3 = 2$. The red node has only neighbours in the first layer and therefore minimal multilayer participation and dispersion $p^{(ml)} = d^{(ml)} = 0$. The blue node in contrast has neighbours in all layers, leading to a multilayer participation vector of $p^{(ml)} = (1/6, 1/2, 1/2)$ and therefore participation and dispersion of $p^{(ml)} = d^{(ml)} = 2/3$. The maximal value of 1 is not reached despite having one neighbour in each layer since the layer sizes differ. The yellow node has neighbours in only two layers and therefore $d^{(ml)} > p^{(ml)}$.

Application to synthetic networks

The omnipresence of networks that arise from data and real world applications raises the question whether they have underlying features in common and how those can be mimicked in artificial network models. Furthermore these allow us to tune and vary a certain network parameter, e.g. the density and see in what way other measures are effected. In this chapter we apply the introduced metrics global and local hubness, as well as participation and dispersion to a wide range of synthetic networks.

Firstly we introduce a handcrafted example graph that allows us to show the different roles nodes can play in the mesoscale. Then we investigate the lattice and ring-lattice networks that are mimicking spatial organisation of networks. This is followed by a discussion of Random and Scale-free graphs. We then conclude that networks with intrinsic modular structure show fundamentally different participation than the latter. Last but not least we introduce a new network model that shows an interplay between modular organisation and a scale-free hub structure in order to mimic biological information processing systems as the human brain.

Example graph

In order to demonstrate the usefulness of the introduced 4-dimensional mapping method we analyse a handcrafted example graph. The graph in Figure 13(a) consists of $n = 47$ nodes that are grouped into three modules. Each node's affiliations is represented by a red circle (community I), blue diamond (community II), or purple triangle (community III). The partitioning was delivered using the Louvain method for community detection [33] [34]. This method aims to maximise the modularity function Q , which evaluates the fraction of links falling into the communities compared to the expected fraction of links that would fall in the communities in a random graph with the same degree distribution. Community III is the largest one with 23 nodes and I and II are equally sized with 12 nodes each. Communities I has a density of 0.52, community II has 0.2 and III 0.29 as summarised in Table 2. Thus each community is distinct from the other two. We hand-picked seven nodes to illustrate the different roles that can be identified in the 4-dimensional space $(h^{(s)}, h^{(l)}, p, d)$. Instead of showing all the six possible relations we restrict to the three plots we find most informative: $(h^{(l)}, h^{(s)})$,

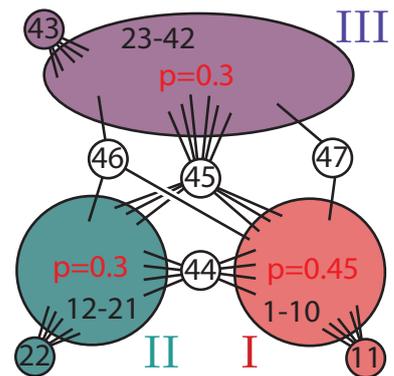


Figure 12: Schematic illustration of the constructing of the example graph: It consists of three modules (I, II, and III). I and II each consists of ten and III of twenty nodes. Those are internally randomly connected with probabilities $p = 0.3$ for II and III and with $p = 0.45$ for I. Each of those random graphs connects with one of the nodes 11, 22, and 43 via five edges. Furthermore 44, 45, 46, and 47 are introduced as different types of connector nodes between the modules. Although those nodes are not clearly belonging to one of the modules by contraction they will be partitioned into one with the Louvain algorithm, as illustrated in Fig. 13(a).

| community | size | edges | density |
|-----------|------|-------|---------|
| I | 12 | 34 | 0.52 |
| II | 12 | 14 | 0.20 |
| III | 23 | 73 | 0.29 |

Table 2: Sizes, number of edges and densities of communities I, II, and III of the example graph shown in Figure 13(a).

[33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008

[34] Inderjit S. Jutla, Lucas G. S. Jeub, and Peter J. Mucha. A generalized louvain method for community detection implemented in matlab. <http://netwiki.amath.unc.edu/GenLouvain>, 2011-2012

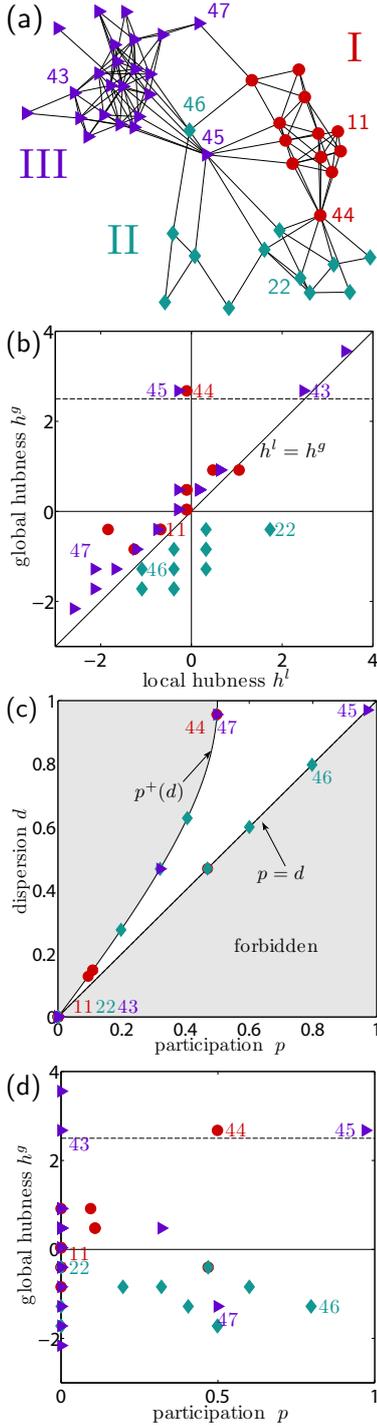


Figure 13: The exemplary graph consisting of three modules (red circles, blue diamonds, and purple triangles) of sizes 12, 12, and 23 for demonstrating different roles nodes can have in the mesoscale structure of a network. Panel (a) shows a force-layout of the graph, Panel (b) local hubness h^l against global hubness h^g , Panel (c) participation p against dispersion d , and Panel (d) participation p against global hubness h^g .

$(p_i, h^{(g)})$ and (p_i, d_i) corresponding to Figures 13(b), (c) and (d). The participation vectors representing the probability of each node to belong to the communities I, II and III for the selected nodes are:

| node | I | II | III |
|------|------|------|------|
| 11 | 1 | 0 | 0 |
| 22 | 0 | 1 | 0 |
| 43 | 0 | 0 | 1 |
| 44 | 0.52 | 0.48 | 0 |
| 45 | 0.32 | 0.32 | 0.35 |
| 46 | 0.38 | 0.41 | 0.2 |
| 47 | 0.48 | 0 | 0.52 |

– Nodes (11) and (22) are connected only within their modules and their participation and dispersion are therefore $p = d = 0$, we call them *peripheral* nodes. Both are connected to five nodes in their communities and have a small global hubness $h^{(g)} < 0$. They are representative of *peripheral non-hubs*, see Figure 1, which can functionally be identified with specialised or segregated function in the network. The examination of their local hubness shows a fundamental difference: (11) shows also $h^{(l)} < 0$ and does not fulfil a prominent position in its module since community I is dense but (22) is the node with highest internal degree in II, which is a sparse module and is therefore a local hub.

– Node (43) is also only connected with neighbours in its own module and its participation and dispersion are $p = d = 0$. Since it belongs to the largest community it is also one of the nodes with highest global hubness $h^{(g)}$. It is both a local and a global hub with no participation. It is characteristic of the rare *peripheral hub* category, see Figure 1.

– Node (44) is a member of community II but devotes the same number of links, six, to community I. It is therefore hard to determine to which of the two communities it belongs. This results in a dispersion index $d \approx 1$. However, it makes no links with community III and has a reduced participation index of $p = 0.5$. This node cannot be considered as a kinless hub but it is a typical *connector hub*. The reduced participation however restricts the node's ability to integrate information from all parts of the network.

– Node (45) instead has the same number of neighbours but those are spread out amongst all three modules. It is a *kinless hub* with the ability to integrate on a global scale. Since it belongs to community III, which is densely connected, its local hubness is $h^{(l)} < 0$.

– Node (46) has only three neighbours but they are each in one of the three communities. Therefore (46) is neither locally nor globally a hub but its high participation makes it a kinless node that is located between all communities.

– Node (47) has two neighbours in its own community III and one in the community II which is half the size. Therefore it has almost equal likelihood to belong to any of the two communities and achieves a large dispersion value close to one. Its lack of connectivity to the community III reduces its participation index to $p \approx 0.5$, therefore we can classify this node as a *non-hub connector*.

In Figure 13(e) the results for this graph are shown using the framework of the functional roles. Most of the nodes, including (11), (22) and (43) are classified as ultra-peripheral (R1) but none as hubs. Since only the internal degree is considered for the classification of nodes into hubs and non-hubs, node (22) is almost regarded as a hub but node (43) is assigned lower hubness despite being connected to more than twice the number of nodes as node (22). This misclassification is the result of assigning global hubness to local hubness and of a lack of normalisation that accounts for the size and density of each community. On the other hand, the dependence of P_i on the number of communities, the accessibility to the region (R3) of connector hubs is reduced and to the region of kinless nodes (R4), forbidden.

The participation vector's reduction onto the leading two dimensions with the SVD is able to detect mainly two insights: A larger radius R stands for a mixture between higher degree/hubness and participation. The intramodular projection vector for each module (dashed lines) \tilde{e} is close to the modular projection (solid lines) \tilde{m} if the module is mostly internally connected. The projection on this vector gives for each node a measure of the number of neighbours in this particular module. The angular distance then is a measure of how strongly a node participates in each module. Node 45 for example has a large angular distance from all modular vectors since it is a kinless node. However the SVD is strongly dominated by the largest module III and modules I and II are almost not distinguishable from each other. Therefore the method, in its 2-dimensional representation, is not suited for our purposes.

We have seen that our approach is capable of detecting a wide variety of node's roles in the mesoscale. However, the usage of four different metrics makes the illustration and interpretation a non-trivial task. Therefore we proceed with the illustration of the method applied to four artificial networks, followed by examples of biological networks. In the synthetic network examples we show results for a single realisation since the number of communities usually differs from realisation to realisation. Still, the results remain general. Averaging over many realisations only smoothens out the roles of outlier nodes, precisely those in which we are more interested.

Lattice & Ring Lattice graph

The *lattice* (see Fig. 15(a)) is widely used in Physics to shape networks that tend to be locally connected which can be a sign of embeddedness into a physical space. This is a reasonable first approach model for many real world interactions: In a social network people tend to befriend individuals that are close to them and in networks from climatological data sites that are close to each other show similar time series. Of course networks that arise from real data will show more complicated behaviour, e.g. the *small-world*-

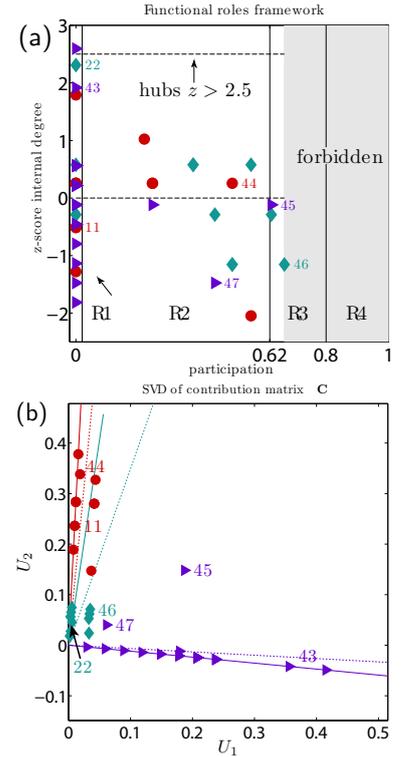


Figure 14: The exemplary graph consisting of three modules (red circles, blue diamonds, and purple triangles). Panel (a) shows the formerly introduced functional roles framework (participation and z-score internal degree). In panel (b) the two leading dimensions U_1 and U_2 from the SVD of the contribution matrix C are illustrated.

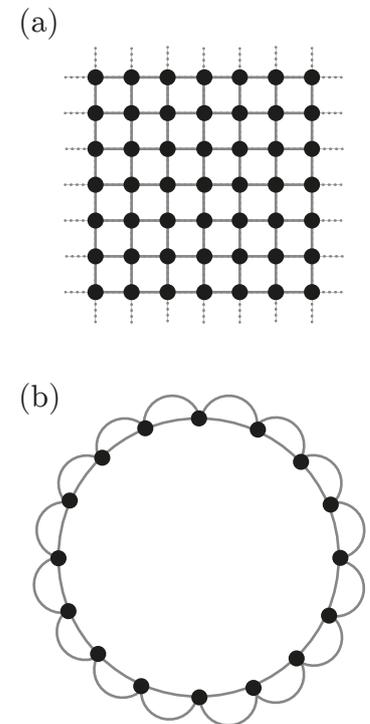


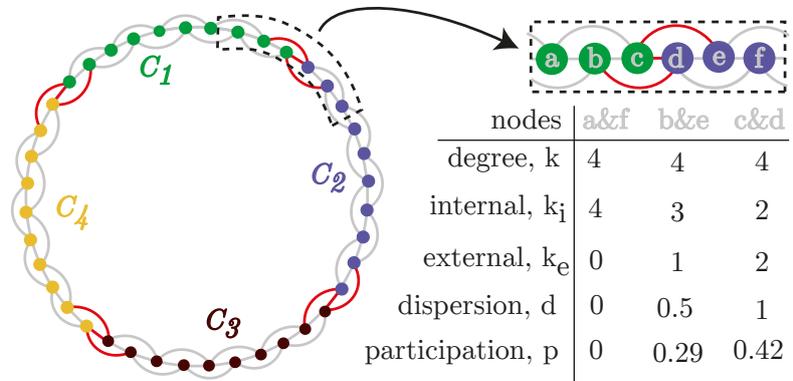
Figure 15: (a) Lattice network with each node having degree $k = 4$. Infinite size of the lattice is indicated with dotted lines. (b) Ring lattice network with 2nd next neighbour coupling and therefore each node having degree $k = 4$

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

effect as illustrated by Watts and Strogatz [28]. Thus, although it is a rough approximation such a regular networks is an interesting starting point and furthermore allows analytical predictions since it is a completely deterministic model.

The regular *ring lattice* (RL, see Fig. 15(b)) is a variant of the lattice with periodic boundary conditions in order to eliminate boundary effects. Each node has the same degree k and therefore we call the network k -regular. Since no node on this lattice is special, a proper clustering is not possible. Nevertheless as the Louvain method starts with random seed nodes we will be able to detect communities and we might use their very distinct shape to understand the role of the nodes and develop a theory that is applicable to other networks, too.

Figure 16: Participation of nodes along a regular ring lattice of $n = 40$ nodes that is partitioned into four communities (green, blue, brown, and yellow) of equal size ten. Each node is connected to the first and second next neighbour and has therefore degree $k = 4$. Links across communities are shown in red and are counted for the external degree k_e . Regarding this we can distinguish three types of nodes: Internal nodes with $k_e = 0$ (nodes a and f), border nodes with the same internal and external degree $k_e = k_i = 2$ (nodes c and d), and those in between (nodes b and e), named 2nd border nodes.



Before we approach the general case we want to investigate an exemplarily situation of a partitioning into four clusters of size 10. See Fig. 16 for an illustration of a RL with each node having degree $k = 4$. Unsurprisingly the modules are all spatially organised and each form a connected graph. Note that due to such a partitioning there are exactly three kind of nodes: nodes that are the direct border between two communities (border nodes), nodes that have only neighbours in their own community (internal nodes), and those in between. Since each node is at most connected with two communities their internal and external degrees completely determines the participation p and dispersion d , and those values are given in the adjacent table. We firstly want to focus on the derivation of a closed formula for the dispersion d since it is easier and it will be straightforward to expand this to the participation p later on. The reduced participation vector \mathbf{P}' has one or two elements for all nodes, because each node is connected to at maximum two modules. For a node connecting to only one module participation and dispersion are defined as $p = d = 0$. For a node with neighbours in two modules and a given external degree k_e and internal degree k_i

the dispersion reads

$$d = 1 - 2 \cdot \text{std}(\underbrace{\frac{k_i}{N_1 - 1}, \frac{k_e}{N_2}}_{\mathbf{p}'}) \quad (24)$$

where N_1 and N_2 are the community sizes of the communities connected to. For clusters of decent size we might drop the -1 accounting the forbidden self-looping of nodes. As stated before we want to assume that the network is partitioned into sub-networks of about even size. As investigated in an example of a RL with $n = 100$ nodes (see Fig. 17) this is not necessarily true but an appropriate assumption. Due to normalisation of the participation vector $\frac{k_i}{N_1} + \frac{k_e}{N_2} = 1$ it is useful to introduce the fraction of internal to external degree $\alpha = \frac{k_i}{k_e} = 1 - \frac{k_i}{N_2}$ and receive

$$d(\alpha) = 1 - 2 \cdot \text{std}(\alpha, 1 - \alpha) = \sqrt{\frac{1}{2} \left(\left(\alpha - \frac{1}{2} \right)^2 + \left(-\alpha + \frac{1}{2} \right)^2 \right)} \quad (25)$$

$$= 1 - 2 \left| \alpha - \frac{1}{2} \right| \quad (26)$$

Let us briefly analyse this result: We showed that each node's dispersion is exclusively depending on its internal and external degree. This holds for all nodes that connect to at maximum two modules, independent on the particular network topology. But on a regular ring lattice there will be very distinct values for the fraction of internal to external edges $\alpha \in [0, 0.5]$ since the degree is the same for all nodes: $\{0, \frac{1}{k}, \frac{2}{k}, \dots, \frac{k/2}{k} = \frac{1}{2}\}$ with the first one being internal nodes (nodes a & f in Fig. 16) and the last one complete border nodes (nodes c & d). The next highest value $\alpha = \frac{k-1}{k}$ are the nodes directly behind the border nodes and we want to call them *2nd border nodes* (nodes b & 2). Note that those nodes only exist since the degree of all nodes is $k = 4$, in a case for $k = 2$ they would not exist. Consequently for high degrees also *3rd border nodes* and so on will occur.

Now we tackle the behaviour of the participation p . It can be directly derived from the dispersion d but the number of modules M the RL is divided into is crucial for this. Firstly note that for the case of one global cluster $M = 1$ all nodes are internal nodes and therefore $p = d = 0$. For the case of two clusters we receive $p = d$ with d after the just derived equation (26). For the more general case we have to perform some algebra (see Appendix) to end at

$$p(\alpha, M) = 1 - \sqrt{\frac{M}{M-1} \left(\alpha^2 + (1-\alpha)^2 - \frac{1}{M} \right)} \quad (27)$$

with the limiting case of an infinite amount of communities

$$p_\infty(\alpha) = \lim_{M \rightarrow \infty} p(\alpha, M) = 1 - \sqrt{\alpha^2 + (1-\alpha)^2}. \quad (28)$$

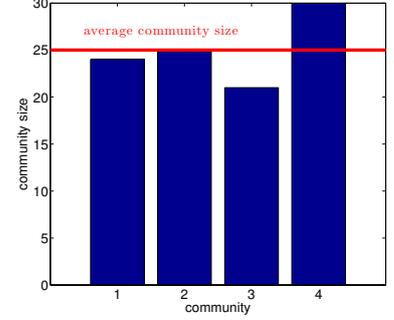


Figure 17: Communities sizes on the ring lattice with $n = 100$ nodes and each node degree two. Louvain algorithm detects $M = 4$ modules of approximately even size. Horizontal red line indicates the average community size $n/M = 25$.

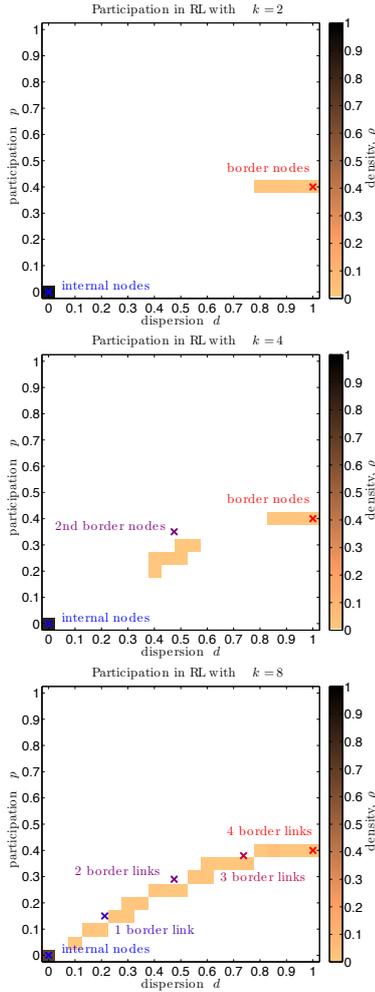


Figure 18: Participation p and dispersion d for three RL networks with varying degree k . Density plots in the background show numerical calculations and crosses are analytical predictions with eqn. 27 and 26. Upper panel show case $k = 2$ where a majority of nodes is internal nodes ($p = d = 0$) and some nodes are border nodes that have the same amount of internal and external neighbours and therefore $d = 1$. With increasing the degree to $k = 4$ (middle panel) 2nd border nodes emerge that have three neighbours in their module and one in another one. In the lower panel with $k = 8$ even more in-between cases occur and therefore the number of internal nodes is also increasing. Analytical predictions are not in perfect alignment with numerical results since the numerics assume equal sized modules and ignore the forbiddances of self-loops.

[41] M Puck Rombach, Mason A Porter, James H Fowler, and Peter J Mucha. Core-periphery structure in networks. *SIAM Journal on Applied mathematics*, 74(1):167–190, 2014

In Fig. 18 we compare the numerically calculated (p, d) pairs for RL of $k = 2, 4, 8$ with the just derived analytical expressions for the participation and dispersion. Note again the crucial approximation is that all communities are the same size, and ignoring the self-loop ban. The full analytical results for large networks with α in continuity approximation are shown in Fig. 19, including the two limiting cases of $M = 2$ and $M \rightarrow \infty$.

Calculating the global hubness $h^{(g)}$ is not possible for a regular network since the standard score is not defined for a sample with standard deviation of zero. However, calculating the local hubness $h^{(l)}$ gives interesting insights. As illustrated in Fig. 16 nodes in the centre of modules have the highest internal degree $k^{(i)} = k$, whereas nodes at the border between modules have same internal and external degree $k^{(i)} = k^{(e)} = k/2$. Thus the central nodes will have a local hubness $h^{(l)} > 0$ and the border nodes $h^{(l)} < 0$. Eventually existing second border nodes, third border nodes, etc. will show local hubness values in between those extrema.

In combination with the discussion of the participation we receive the following insight: Nodes in the centre of modules show minimal participation and dispersion $d = p = 0$ and maximal local hubness $h^{(l)}$. Towards the borders between modules dispersion increase up to $d = 1/2$ and the simultaneously the local hubness $h^{(l)}$ decreases (see Fig. 20). This can be described as a *core-periphery-structure* [41] since the central nodes are strongly connected with each other and form a dense *core*, whereas the border nodes are only loosely connected with the module and form the *periphery*. In the example of the RL the size of the core depends on two factors: the size of the module n' , as well as the degree k of each node. If the degree is large the adjacent modules connect with many nodes across the borders, decreasing the number of nodes that make up the core. Since the RL is regular the width of this periphery regions will be the same size for each module, exactly $k/2$ to each side.

In real world applications the networks are often spatially organised and therefore form networks not unlike the discussed RL and lattice but the graphs will unlikely be regular. In fact the degree of the nodes are a primer for how far-reaching the interactions are for each node. Therefore modules that consist of nodes with longer connections will show a wide periphery and modules with shorter influence range will show a narrow periphery. Thus we can use the local hubness, as well as dispersion and participation in order to analyse the borders between modules. Sharp border will show a narrow region of high participation, whereas fuzzy borders between modules are more likely to show a broad area of such high participation nodes. Those insights will be useful for the analysis of networks that are grounded on a strong spatial organisation, as for example climate networks.

Figure 19: Analytical calculations for participation p and dispersion d of nodes in a RL in dependence of the number of communities M . For $M = 1$ all nodes are internal nodes and therefore $d = p = 0$ (black dot). For $M = 2$ dispersion and participation are still identical $p = d$ but nodes range from internal nodes to border nodes $p = d = 1$ (dashed line). All in-between values can be reached for large networks and high degrees k . For larger number of modules M dispersion and participation will differ $d > p$ for non-internal nodes (coloured lines). The limiting case for an infinite number of modules is shown with dotted line.

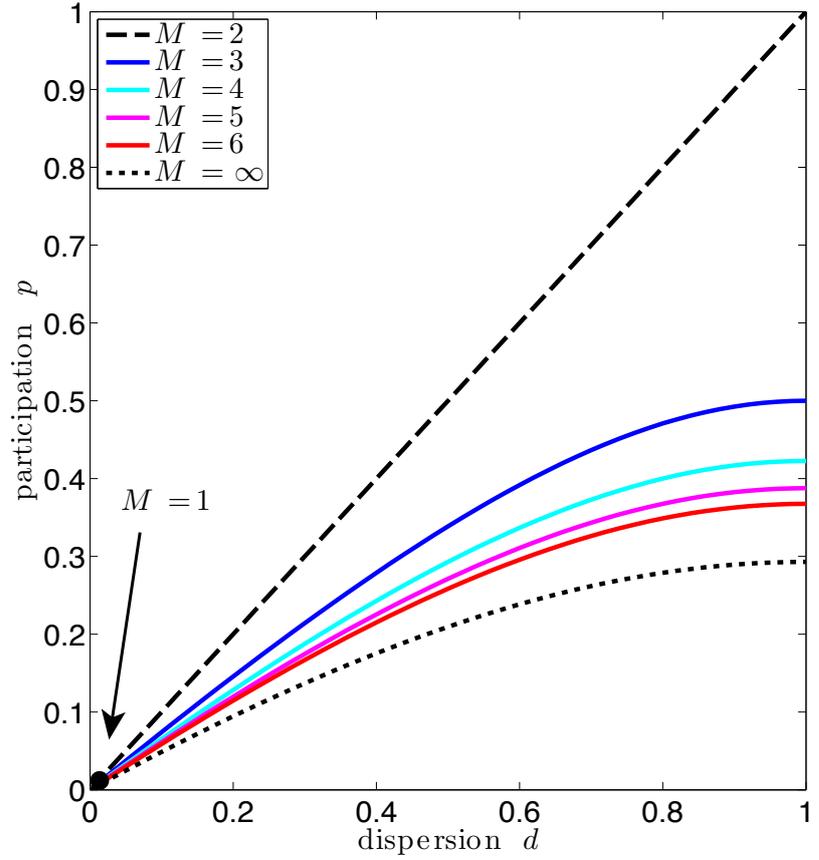
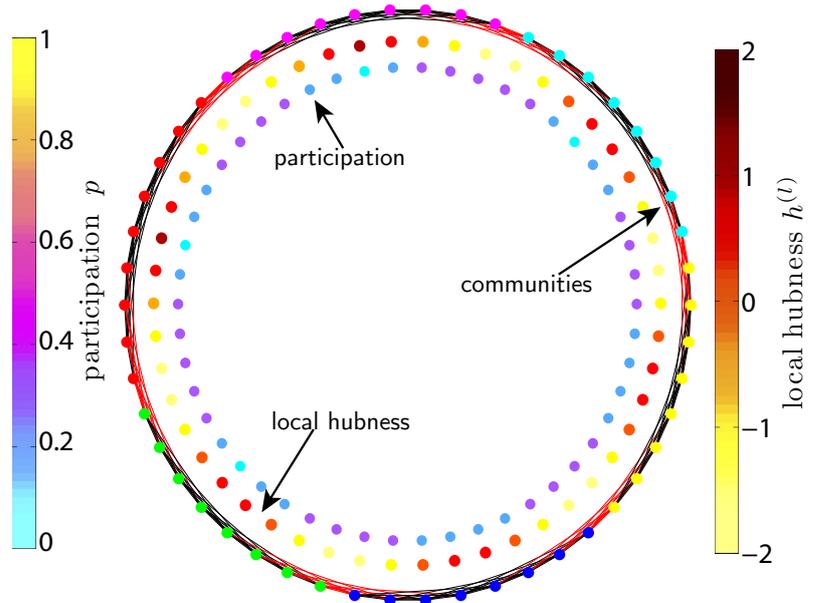


Figure 20: RL consisting of $n = 50$ nodes with degree $k = 8$. We illustrate affiliation to one of the $M = 6$ modules in the most outer ring. Edges are shown as lines that connect nodes, black lines are edges inside modules and in red lines are edges connecting nodes that are not in the same module. In the middle ring the local hubness $h^{(l)}$ is colour coded. The most inner ring shows the participation of the nodes. We see that in the centre, or core, of modules minimal participation and maximal local hubness is reached. Towards the border to other modules the participation increases since cross-module edges become more frequent. At the same time the local hubness decreases since the number of inner-modules edges decreases. The combination of both is a sign of a core-periphery structure.



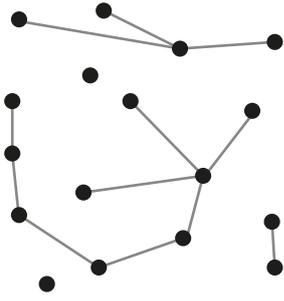


Figure 21: Typical ER random graph with $n = 17$ nodes and $m = 12$ edges, created with a connection probability of $p = 0.1$ and resulting in a density of $\rho \approx 0.09$.

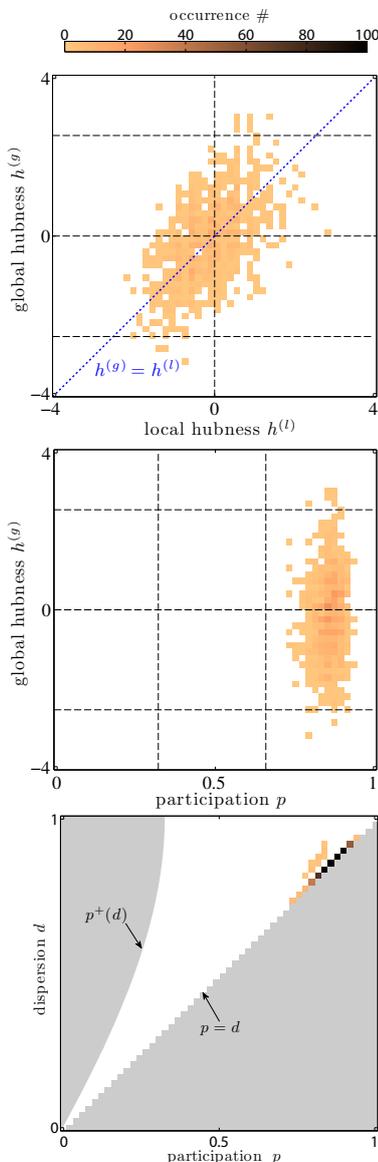


Figure 22: Mesoscale metrics density plots for a ER network of size $n = 1000$ with connection probability $p = 0.1$. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

Random graph

The *Erdős-Rényi* model (ER model) is the most famous one for generating random graphs. The widely used version has two parameters, the number of nodes n and the probability p to connect a pair of nodes with an edge. The creation of edges independently of other ones is a strong simplification and might be inappropriate for modelling real world networks.

Nevertheless it inhabits interesting phenomena as the size of the largest component and is a good starting point for studying network approaches to various applications. Although each realisation will have a different number of edges m the expected density of the resulting graph is same as the probability p and this will be approached for infinite sized networks. In a way it is the exact opposite of the just discussed RL that showed strong localised connectivity and therefore a strong order. The ER in contrast does not show any localisation since all edges have same probability of existence.

In Fig. 30 the distribution, as density plots, of the four metrics for the ER graph partitioned into nine modules is shown. Firstly we note that the distribution of both local and global hubness is rather small as expected. There is a correlated trend between both since nodes with a large number of neighbours inside their module are, in general, those with high number of neighbours in the whole network. The high values of participation p and dispersion d mean that all nodes have a significant number of neighbours in other modules. Most of the nodes have neighbours in all nine modules since most nodes lie on the diagonal $p = d$. Nodes directly above the diagonal are linked to eight modules. There are no peripheral nodes having neighbours only in their own community ($p = d = 0$). These observations indicate that the encountered community structure is very poor, as it is expected for ER graphs. The $(p, h^{(g)})$ plot is often useful to detect the global roles and functionality of the nodes since it closely resembles the classification framework illustrated in Figure 1. Here we see that all nodes have a kinless contribution to all different modules due to the lack of structure at the mesoscale. A small number of nodes may be classified as hubs with $h^{(g)} > 2.5$. According to the definition of the hubness approximately 1% of all nodes will fall into this category.

The definition of the global hubness takes into account the density, as well as the size of the network. Therefore changing those parameters in the ER network model does not affect the hubness distribution (see Fig. 23) In panel (a) we fix the connection probability to $p = 0.1$ and vary the size of the network. Although there are fluctuations in the distributions, due to the finite network size, the overall distribution is barely affected by increasing the network's size. In panel (b) we fixed the number of nodes to $n = 1000$ and change the connection probability p . Overall we see that the hubness distribution is not strongly influenced increasing connec-

tion density. However, the minor trend of decreasing of the upper boundary of the global hubness $h^{(g)}$ is detectable for extremely dense networks, as predicted by 20.

Summarising we find that ER graphs with no intrinsic community structure do not show a diversity of roles when analysing their mesoscale. All nodes tend to be connected with all modules and have a high participation. A small number of global hubs is detectable, as we expect 1% of nodes to exceed the expected degree and thus have $h^{(g)} > 2.5$.

Scale-free graph

It is known that many real world networks show extremely broad and non-homogenous degree distributions. In particular if the degree distribution $P(k)$ follows a power-law $P(k) \propto k^{-\gamma}$ it is called *scale-free* (SF). Barabási and Albert introduced an evolving network model where new nodes are entering the network and connecting with already existing nodes following a so-called preferential attachment rule [42]. This rule gives an emphasis to nodes that are already highly connected (= have a high degree) and therefore a scale-free degree distribution with $\gamma = 3$ is achieved.

Goh *et al.* [43] introduced another scale-free network model we will use in this work since it is more universal: See Fig. 24 for an exemplary network. We start with an empty network consisting of all n isolated nodes, which are indexed with an index i ($i = 1, \dots, n$). The control parameter $\alpha \in [0, 1)$ is used in order to assign each node a weight $w_i = i^{-\alpha}$. Now we iteratively create edges between a pair of nodes (i, j) , whereas the probability to draw each node is proportional to the introduced weights w_i . This leads to a scale-free degree distribution with the exponent $\gamma = \frac{1+\alpha}{\alpha}$. With adjusting the control parameter α we are able to create networks with degree distributions that scale with any $\gamma \in (2, \infty)$. In Fig. 24 a small SF network is shown. It is observable that nodes with small indexes $i = 1, 2, 3$ show higher degrees as those with high indexes $i = 18, 19, 20$. Furthermore we detect in the force-directed placement that the hubs are in the centre of the network and form a rich-club since they are all connected to each other.

Here we analyse a SF graph with the same parameters as the ER graph, consisting of $n = 1000$ nodes with a density of $\rho = 0.1$. In order to receive a moderate power-law degree distribution we choose $\alpha = 1/2$ and thus $\gamma \approx 3$ is the power-law exponent. In Fig. 25 the density plots of the mesoscale parameters are shown. Firstly we note that global as well as local hubness are much wider distributed than in the ER graph. For the global hubness this is expected since the high degree nodes in the SF show much higher degree than in a compared ER graph. Such hub nodes tend also to be local hubs with $h^{(l)} > 2.5$ since many of their neighbours are in their module. In terms of participation p and dispersion d the results are very similar to the ones for the ER: All nodes show

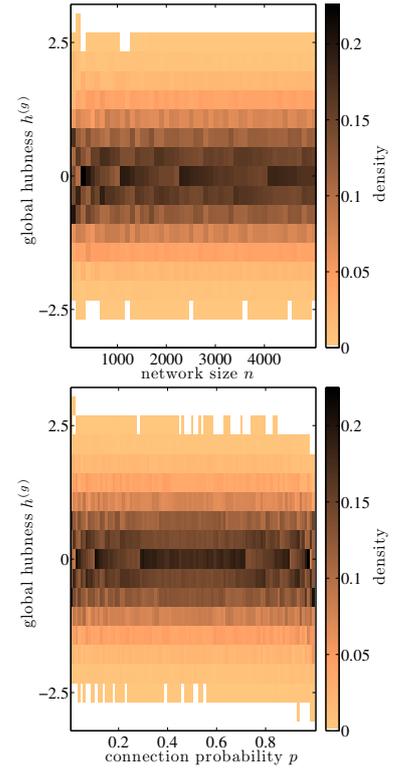


Figure 23: Global hubness $h^{(g)}$ distributions for ER networks with varying size n and connection probability p . Upper panel shows distribution of hubness for fixed $p = 0.1$ with varying size. Lower panel has fixed size to $n = 1000$ and varying probability p .

[42] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999

[43] K-I Goh, B Kahng, and D Kim. Universal behavior of load distribution in scale-free networks. *Physical Review Letters*, 87(27):278701, 2001

Figure 24: Goh model network with $n = 20$ nodes, connection probability $p = 0.18$, $\alpha = \frac{1}{2}$ and therefore $\gamma = 3$. Left Panel: Nodes are placed along the ring to show that nodes with small indexes (pink) have a high degree, whereas nodes with high indexes (cyan) will receive a small amount of edges. Right panel: Force-directed placement of the same network illustrates that the high degree nodes are also the central ones and strongly connected amongst each other and therefore form a rich-club.

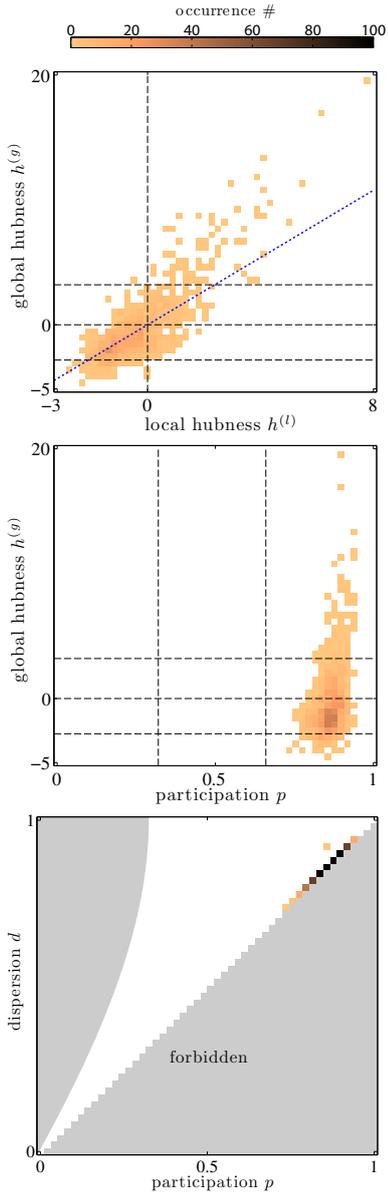
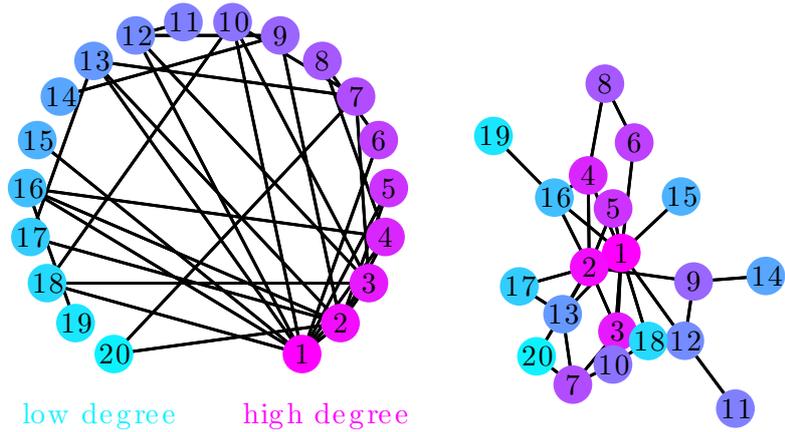


Figure 25: Mesoscale metrics density plots for a SF network of size $n = 1000$ with connection probability $p = 0.1$. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

extremely high values. This indicates that the community structure in such SF networks is very poor. Furthermore due to the relatively high density of connections almost all nodes connect to all modules and therefore lay on the diagonal $p = d$.

In terms of the role the nodes play in the mesoscale the poor community structure makes all nodes, independent of their degree kinless nodes. Note that the high degree nodes are therefore hubs that are able to access all modules, however since the modules are not very distinct a segregated processing of information can not be achieved there. In the following section we will discuss networks that show such an intrinsic modular structure.

Modular graph

Many real world networks show a modular structure. A simplistic way of creating networks with those characteristics is the *flat modular network* (FM, see Fig. 26). For this we firstly create M ER networks of size n' and choose a relatively high connection probability p_{in} . Now we cross-connect those created modules randomly with a connection probability p_{ex} . In order to create a detectable modular structure we have to choose $p_{ex} < p_{in}$. For $p = p_{ex} = p_{in}$ it would result in a normal ER graph with connection probability p and $n = M \cdot n'$ nodes.

We analyse such a FM graph consisting of $n = 1000$ nodes that are grouped into $M = 5$ equal sized modules with $n' = 200$ nodes. Inside the modules pairs of nodes are connected with probability $p_{in} = 0.3$ and externally with probability $p_{ex} = 0.05$.

We note that the local and the global hubness are strongly correlated, see Fig. 27, mainly because they reflect the internal degrees of the nodes. The small amount of external connections per node has a weak influence on the relation between local and global hubness.

Focusing on the measures of participation and dispersion we find that their values are much lower than those observed before in ER and SF networks. This happens because most nodes are strongly connected inside their module and only weakly connected to the other modules. The communities are well defined in this case. However, only a small amount of nodes have neighbours in only one module and thus $p = d = 0$. This shows that the strong community structure does not go in hand with a strong core of nodes inside each module. In contrast the small amount of external edges is evenly distributed amongst all nodes and therefore almost all of them show external affiliations. In the FM graph there is no dependence between $h^{(g)}$ and p . This occurs because external connections are created at random and therefore all nodes have the same probability of connecting to other modules.

In the density plot we investigate only a particular parameter set of internal and external connection probability. The chosen parameters result in low values for participation p , however we know that the ER network as a subset of such modular graphs with $p_{in} = p_{ex}$ show high participation $p \rightarrow 1$. Therefore we investigate in Fig. 28 five different FM networks with varying external connection probability $p_{ex} \in \{0, 0.025, 0.1, 0.2, 0.3\}$ while the internal probability is fixed to $p_{in} = 0.3$. A FM network with $p_{ex} = 0$ consists of separated modules that are not cross connected, therefore all nodes have exclusively neighbours in their own module and accordingly show all the minimal participation $p = 0$. With increasing the external connection probability p_{ex} nodes connect with other modules and therefore show increased participation p . Note that the random distribution of the edges shows a random distribution of the participation values around a mean value (as indicated with dashed vertical lines). The further increasing of the amount of external edges shifts

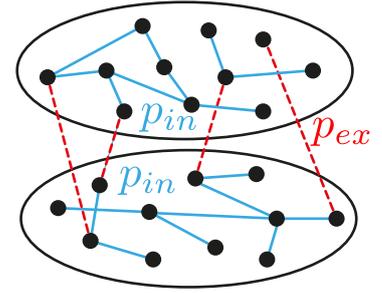


Figure 26: A flat modular network consisting of $M = 2$ modules (upper and lower ellipses) with each $n' = 11$ nodes. Inside each module the nodes are connected with a probability of p_{in} (cyan edges) and across modules with a probability of p_{ex} (red dashed edges).

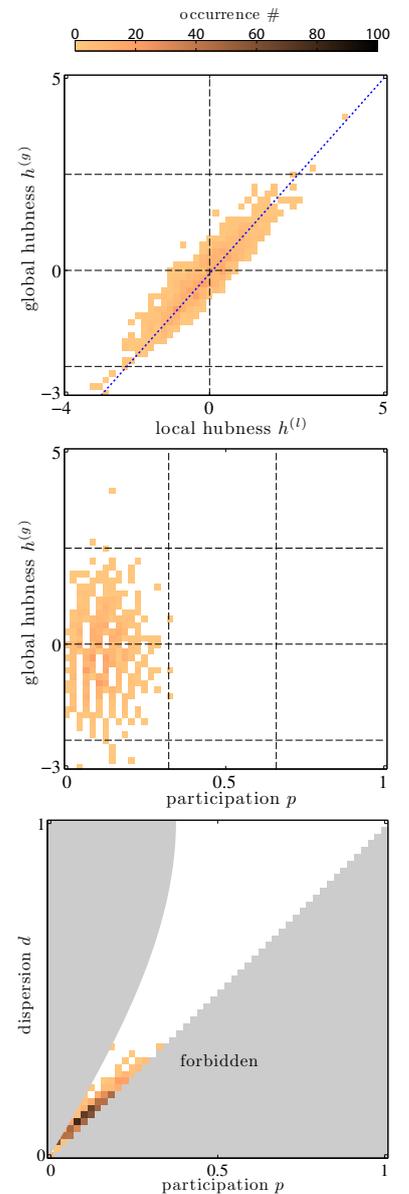


Figure 27: Mesoscale metrics density plots for a FM network of size $n = 1000$ consisting of $M = 5$ modules of even size $n' = 200$ with internal connection probability $p_{in} = 0.3$ and external $p_{ex} = 0.05$. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

the nodes to even higher participation values. In the limiting case of $p_{ex} = p_{in} = 0.3$ all nodes show very high participation values, same as an ER graph.

Since the participation of the nodes in a FM graph is very distinctively distributed around a mean value that increases with higher p_{ex} we analyse this behaviour in Fig. 29. Also we derive an analytical formula for the mean participation (see Appendix) of all nodes in a FM

$$\langle p \rangle = 1 - \sqrt{\frac{M}{M-1} \left(\left(\left(\frac{p_{in}}{\Sigma} \right)^2 + (M-1) \left(\frac{p_{ex}}{\Sigma} \right)^2 \right) - \frac{1}{M} \right)}. \quad (29)$$

This is achieved by calculating the participation of a node that has the mean value of neighbours to each of the modules. We see that this analytical expression is in good agreement with the numerical results (Fig. 29). For $p_{ex} \rightarrow 0$ the analytical expression fully matches $\langle p \rangle = 0$ for the unconnected modules. Some discrepancy is observable for the case were the ER networks are approached $p_{ex} \rightarrow p_{in}$: The prediction is $\langle p \rangle = 1$ but the numerical results are below. This happens since the participation is bounded in $[0, 1]$ and when averaging over all nodes some of them will show lower than the mean value and that can not be compensated by the bounded distribution ≤ 1 . The analytics hold better for dense networks since the chosen mean-field approach is suited for networks with a high number of edges, were the law of large numbers makes the mean approximation more eligible.

In real networks the community organisation happens as the consequence of functionally related nodes being gathered together. In this manner groups of specialised function are segregated, restricting the influence of nodes with different function on them. In the FM model the communication between the modules is limited due to the overall small dispersion.

Modular graph with scale-free attachment

Finally we want to introduce a network that have a scale-free degree distribution and thus hub structure, as well as a modular organisation of nodes. This *centralised modular* model (CM) is created in the first step exactly like the FM: We create M modules of each n' nodes and connect those internally with the probability p_{in} . The external links between modules are created following a preferential attachment rule. Similar to the SF network generation method as introduced before the nodes of each modules of size n' are indexed as i ($i = 1, 2, \dots, n'$) and are assigned a weight $w_i = i^{-\alpha}$, where α is a control parameter in $[0, 1)$. When connecting two different modules, nodes from each module are selected with a probability proportional to their weight. The degree distribution of network generated with this method follows a power law with $\gamma = (1 + \alpha)/\alpha$. In this case, the hubs from the different modules become interconnected and may form a so called *rich club*. This combination of character-

Figure 28: We investigate the participation vs. global hubness scatter $(p, h^{(g)})$ for five flat modular networks consisting of $M = 5$ modules of size $n' = 200$ with internal connection probability of $p_{in} = 0.3$ and varying $p_{ex} \in \{0, 0.025, 0.1, 0.2, 0.3\}$. For $p_{ex} = 0$ (orange dots) all nodes are only connected inside their modules and no cross connections exist. Therefore all nodes are internal nodes with $p = 0$. With introducing external connections with probability $p_{ex} = 0.025$ (red dots) nodes show increased participation due to those external connections. This increased participation is observable for further increasing of p_{ex} . For $p_{ex} = 0.3$ (blue dots) the network becomes an ER network since external edges have the same probability as external. Therefore we receive a plot similar to that case of a random network with almost maximal participation $p \rightarrow 1$. Vertical dashed lines indicated the mean participation $\langle p \rangle$ for each of the five networks. Note that we used the *a priori* partition into the five communities here.

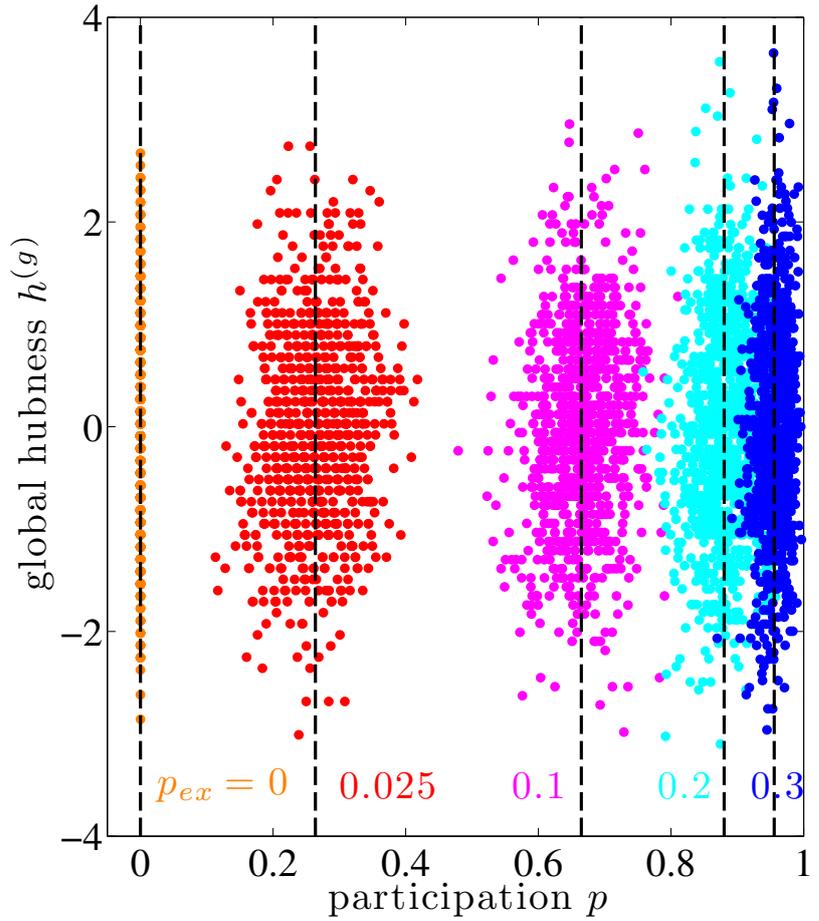
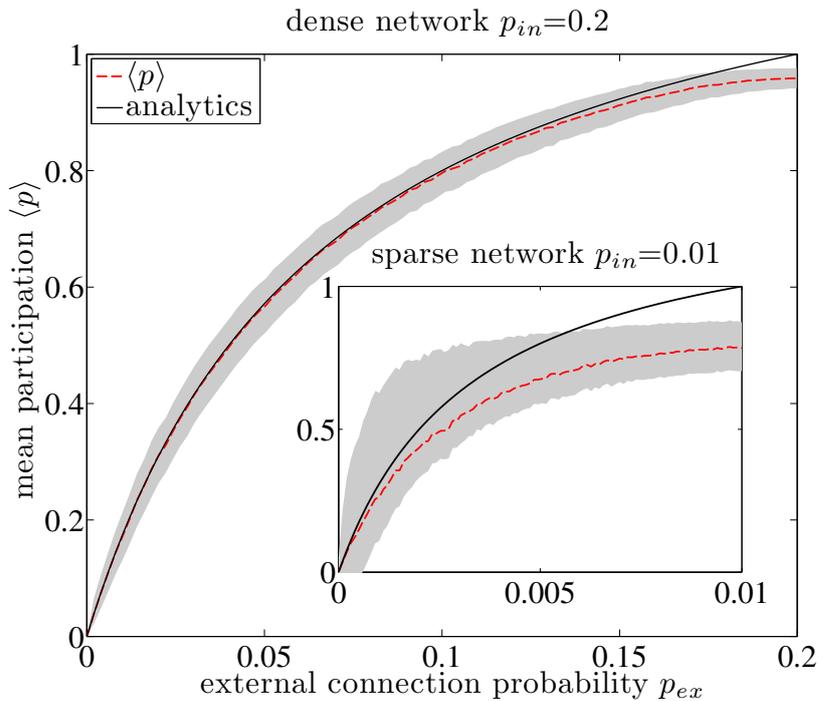


Figure 29: Mean participation $\langle p \rangle$ for FM networks of two different internal connection probabilities ($p_{in} = 0.2$ and $p_{in} = 0.01$ as inlay) with varying external connection probabilities from $p_{ex} = 0$ up to the ER graph case $p_{ex} = p_{in}$. Numerical means are shown with dashed red line and the standard deviation of the participation of all nodes as grey shaded area. The participation is minimal for separated modules ($p_{ex} = 0$) and increases up to ≈ 1 for the ER network cases. This is in good accordance to our analytics (solid black lines). Note that the analytics fit better in the case of a denser network since the law of large numbers makes our mean approximation more eligible for higher numbers of connections.



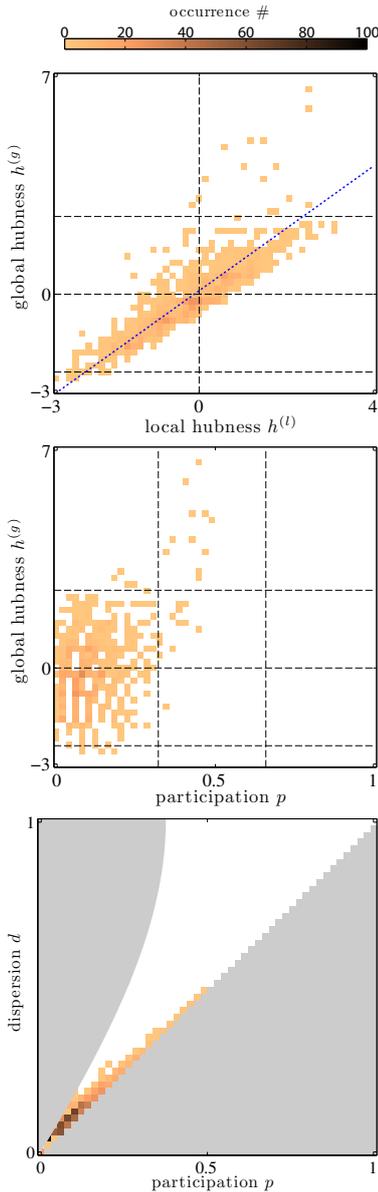


Figure 30: Mesoscale metrics density plots for a CM network of size $n = 1000$ consisting of $M = 5$ modules of even size $n' = 200$ with internal connection probability $p_{in} = 0.3$ and external $p_{ex} = 0.05$. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

istics, modular structure with centralised interconnectivity occur in neuronal networks and so we want to understand their influence on the mesoscale.

Similar to the FM the CM shows a correlation between local and global hubness. However, in the CM model the correlation is weakened since few nodes accumulate most of the links between modules. Some of the nodes that are hubs for the external connectivity happen to be sparsely connected locally. Overall, due to the preferential attachment, global hubness takes larger values in the CM than in the FM model.

Also the $(p, h^{(g)})$ plot for the CM model is very similar to the FM but the hubs clearly stand out – they have larger hubness and participation. Thus the hubs are connector hubs that interconnect the different modules. In the (p, d) -plane, the hubs lie on the $p = d$ line because they connect with the four modules. Due to this centralisation non-hub nodes receive less external links and see their participation lessened because they connect to less modules such that for them $d > p$. In the CM model the hubs densely connect with all the modules allowing them to integrate information from the distinct functional groups.

Conclusion

The investigation of synthetic graphs is very helpful for understanding the introduced mesoscale measures hubness, participation, and dispersion. As showed the participation and dispersion are strongly connected with the modularity of the network as only a strong modular structure will enable nodes to show low dispersion and participation. This is also connected with the discussion of the RL graph that is simplified network of localised connectivity. There we showed that participation and local hubness are suitable primers to detect and characterise the borders between spatially organised modules. Here a strong modular structure is associated with a narrow region of high hubness, whereas long ranging connections will lead to a blurred and wide border between the modules and therefore a overall low participation. Since SF and ER network do not show a strong community structure they also showed overall high participation.

As intrinsic to them SF network do show a hub structure. However, due to the high participation for all nodes those hubs do not receive an outstanding position in the mesoscale of the network. We then created a modular network with an on-top hub structure that interconnects the modules and enables the integration of the information of the segregated information from the individual modules. As we will show in the following chapter many real-world biological neuronal networks show such a structure in order to optimise the information processing.

Deciphering segregation and integration in neuronal networks

In this chapter we study the mesoscale structure of real neuronal networks. We start with the fully mapped neuronal network of the nematode *Caenorhaditis elegans* and finish with the corticocortical connectivities from the cat and human brains. Previous studies have shown that these networks combine the existence of modules of functionally related neurons or regions with the presence of highly connected hubs [19] [20] [22]. The segregation into modules allows the brain to handle information of different sensory systems in a temporal and spatial parallel manner. However, the brain needs to combine – integrate – the multisensory information in order to create a comprehensive understanding of the environment. It is believed that the neural hubs aid in the integration because of their capacity to reach information of different sensory systems.

C. Elegans

Caenorhaditis elegans (*C. elegans*) is a roundworm of about 1 mm length and is widely used as a model organism in neuroscience since its whole connectome has been mapped [44]. It has been mapped by electron micrographs of serial thin section in various studies [45]. The nervous system consists of a large somatic part along the body and a small pharyngeal protobrain in the head area. Despite the worm's small size (1 mm long) and simple nervous system it shows a broad range of non-trivial behaviour [46]. The animal is able to discriminate and move towards or away from chemicals, temperature, or food sources. Different strains of *C. elegans* are shown to either feed in groups (social) or alone (solitary) [47]. Each sex (male and hermaphrodite) show specific behaviour such as egg-laying or mating behaviour and also different number of neurons. Most of these behaviours are influencing each other, for example, in stressful environments egg-laying is suppressed.

The nervous system consists of neurons that connect with each other via synapses. Therefore we will represent the neurons by nodes and the synaptic connections as edges. Although different types of synapses are existing, electrical (or gap junctions) and chemical, we do not distinguish between them. Here we use the data from an adult hermaphrodite worm [48] in a binary form. The

- [19] O. Sporns and G. M. Tononi. Classes of network connectivity and dynamics. *Complexity*, 7(1):28–38, 2001
- [20] Gorka Zamora-López, Changsong Zhou, and Jürgen Kurths. Exploring brain function from anatomical connectivity. *Frontiers in neuroscience*, 5, 2011
- [22] G. Zamora-López, C. S. Zhou, and J. Kurths. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.*, 4:1, 2010

- [44] Lav R Varshney, Beth L Chen, Eric Paniagua, David H Hall, and Dmitri B Chklovskii. Structural properties of the *caenorhaditis elegans* neuronal network. *PLoS computational biology*, 7(2):e1001066, 2011
- [45] Samuel Ward, Nichol Thomson, John G White, and Sydney Brenner. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *caenorhaditis elegans*. *Journal of Comparative Neurology*, 160(3):313–337, 1975
- [46] William R Schafer. Deciphering the neural and molecular mechanisms of *C. elegans* behavior. *Current Biology*, 15(17):R723–R729, 2005
- [47] Mario de Bono. Molecular approaches to aggregation behavior and social attachment. *Journal of neurobiology*, 54(1):78–92, 2003

- [48] Beth L Chen, David H Hall, and Dmitri B Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4723–4728, 2006

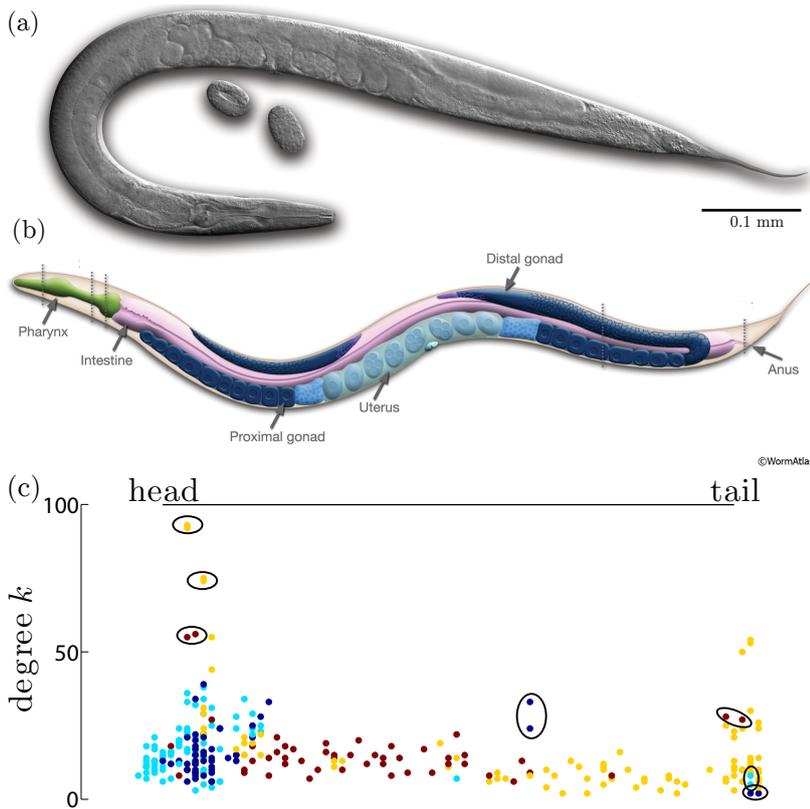


Figure 31: Anatomy of an adult *C. elegans* hermaphrodite. Panel (a) shows a microscopic image of the roundworm. Panel (b): Schematic drawing of anatomical strictures, left lateral side. *C. elegans* has an unsegmented, cylindrical body shape that is tapered at the ends. The cells of the nervous system are organised into ganglia in the head and the tail with a majority in the head around the pharynx (left side).

[39] Dominik Traxl. *C. elegans* - Neural structure & dynamics. Master's thesis, Ludwig-Maximilians-Universität München, 2012

network consists of 302 neurons but we restrict to the largest connected component of 279 nodes and a total of 2285 edges [39]. We use a Louvain-partition of the network in four communities of 50, 51, 88 and 90 neurons each. In Fig. 31(c) the position of all neurons along the body of the worm are shown, together with the number of neurons it connects to (degree k) and the community it belongs to. We see that the modules are spatially organised: In the proto-brain in the head consists of many densely connected neurons with high degree and is partitioned into two modules (light and dark blue). The large somatic nervous system is divided into two parts (red and yellow), too. The yellow modules stretches from the head to the tail, whereas the red one focusses on the frontal half. The nodes of highest degree are at the proto-brain and the region with the next highest degrees is the tail. The tail has long-range axons to the head since some of the nodes belong to the spatially very distant blue modules. Such node will likely be connector nodes in the mesoscale structure. Note that many of such distinct nodes occur pairwise, as indicated in some cases with ovals. This is due to the bilateral symmetry along the body.

When analysing the roles of nodes in the network, see results in Fig. 34, we find some characteristics similar to those observed in the synthetic networks and other distinctive properties. Similar to SF networks the *C. elegans* shows broad hubness distributions with local and global hubness being highly correlated. The range of

$h^{(g)}$ is larger than that of $h^{(l)}$ indicating that the hubs might form a rich club. On the other hand, many nodes have significant negative hubness. The participation and the global hubness are correlated in this case. Non-hubs have low participation and hubs have high participation. Altogether these observations clearly reflect the modular and hierarchical organisation of the network consisting of well defined communities (as shown by the large amount of nodes with negative hubness and very low participation) with the interconnections centralised by a set of global hubs forming a rich-club (hubs are both locally and globally hubs, and they have participation close to $p = 1$). In the (p, d) we appreciate that many nodes escape the $p = d$ line and take larger values of dispersion than of participation. This behaviour, not observed in the synthetic network models, is typical of connector nodes which have dispersed connectivity but do not connect with all the communities.

As discussed before the nervous system of the *C. elegans* is subdivided into two main organisatory parts. The small proto-brain' located in the pharynx and the rest of neurons are symmetrically arranged, left and right, along the ventral cord. We detect this bilateral symmetry in the pairwise occurrence of almost identical hubness and participation. Furthermore note that the high degree connectors are all located in the proto-brain the likely receives central organisatory tasks, similar to the human brain, that has higher integrator functions than the spinal cord (see Fig. 33). However, many nodes show a low participation p since they connect almost only in their own modules, those modules are the centres of segregated information processing.

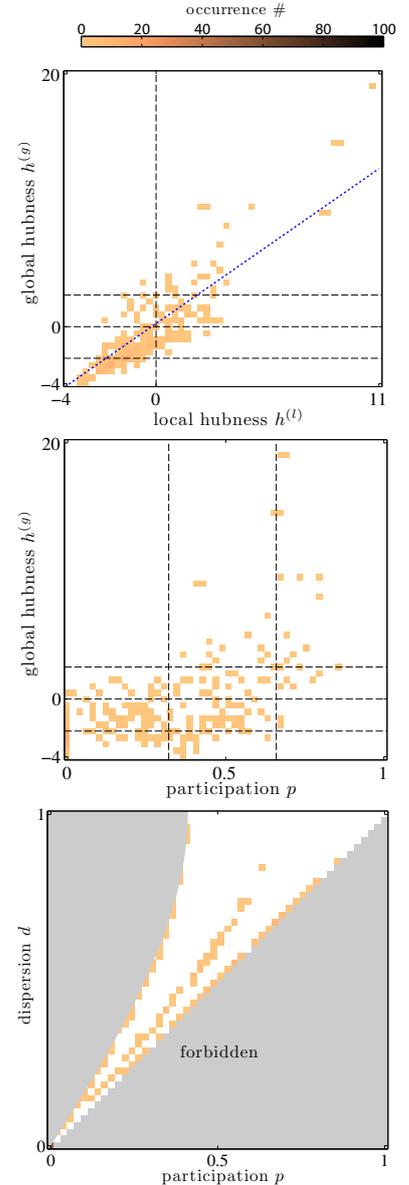


Figure 32: Mesoscale metrics density plots for *C. elegans* neuronal network. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

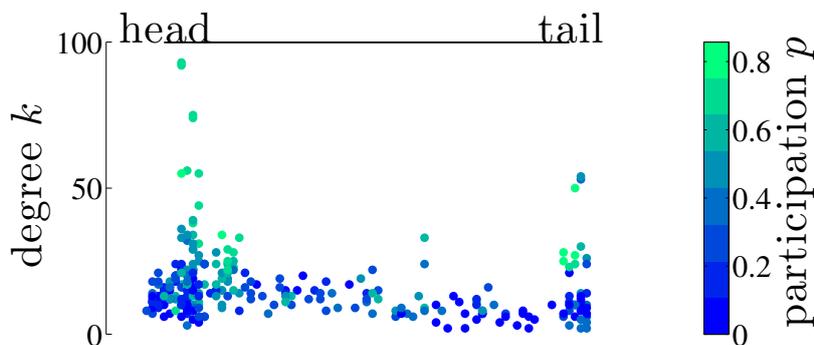


Figure 33: *C. elegans* neurons at their spatial location along the body from the head to the tail along the horizontal axis. Vertical axis indicates the degree k and colour the participation p of each node. The high degree nodes (hubs) are at the same time connectors and mostly located at the head and to smaller amount at the tail. The majority of nodes has low degree and is located along the whole length of the worms body.

[49] JW Scannell, GAPC Burns, CC Hilgetag, MA O'Neil, and Malcolm P Young. The connectional organization of the cortico-thalamic system of the cat. *Cerebral Cortex*, 9(3):277–299, 1999

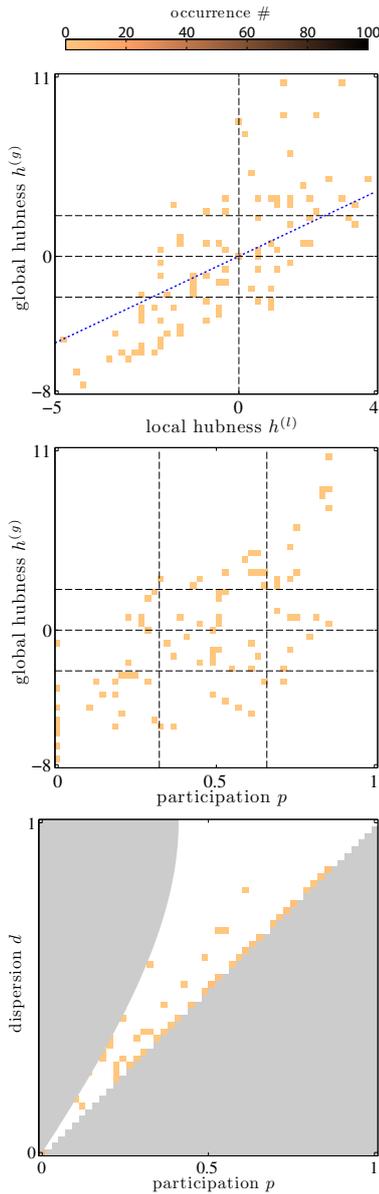


Figure 34: Mesoscale metrics density plots for cat neuronal network. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

Cat

The corticocortical connectivity of cats has been recently investigated to understand the relation between structure and function in the brains of higher organisms. The database used here is a combination of various published datasets [49]. We use a version with $n = 95$ cortical regions interconnected with $m = 1829$ white matter fibers. Here, we consider them undirected and unweighted. We use the Louvain partition into four modules of sizes 24, 21, 23 and 27. The analysis of the mesoscale reveals very similar to that found for the *C. elegans* despite being only a third of the size and much more dense; $\rho_{cat} \approx 20\%$ and $\rho_{eleg} \approx 6\%$. Global and local hubness are correlated with $h^{(g)}$ achieving again larger values than $h^{(l)}$. Global hubness and participation are also correlated. The non-hubs tend to have very low values of participation indicating that they are tightly connected inside their communities. The hubs, on the contrary, have large participation showing that they are well connected with all the four communities. In the (p, d) plot we see, in contrast to the *C. elegans* network, that the nodes are not likely to be at the $p_{max}(d)$ that is reached if nodes are connected to only two modules.

Altogether the two neuronal networks analysed so far show strong similarities in their structure. We now investigate the network of a primate to compare the information processing structures of a even higher developed animal.

Macaque monkey

Investigating the anatomical connectivity of the macaque monkey gives insights in the neuronal network organisation for a non-human primate. Macaques are investigated by neuroscientists as well as social scientists since they show intricate social structure. Here we use a data set that consists of $n = 71$ regions in the entire cerebral cortex [51] connected via $m = 438$ edges, resulting in a density of $\rho = 17\%$. The density lies in between the one of the two previously discussed networks and nevertheless the mesoscale analysis gives similar results, indicating that the found structural properties are universal in neuronal networks:

A partitioning into four modules of sizes 18, 14, 23, and 16 maximises the modularity, the same number of modules as for cat and *C. elegans*. Both hubness indexes are widely spread as in SF networks and strongly correlated: The global hubs are at the same time local hubs. Those hubs do also show the highest values of participation, they are kinless hubs and therefore able to connect the different modules in order to enable communication between the four modules. Those are made up of connector, as well as peripheral non-hubs. Those low-degree nodes with limited or no direct access to other modules enable the segregated information processing. In the (p, d) we see that nodes with a high participation p show the same dispersion d since they connect to all four modules. Nodes with smaller p in contrast are not on the diagonal $p = d$ since they connect to only three or two of the modules. By definition nodes that connect to only a single module have $p = d = 0$. The direct accessibility of all modules by the hubs further confirms their integrator function.



Figure 35: Self-portrait by the depicted macaque nigra female. [50]

[50] Macaque – Wikipedia, the free encyclopedia, 08 2014

[51] Jack W Scannell and Malcolm P Young. The connectional organization of neural systems in the cat cerebral cortex. *Current Biology*, 3(4):191–200, 1993

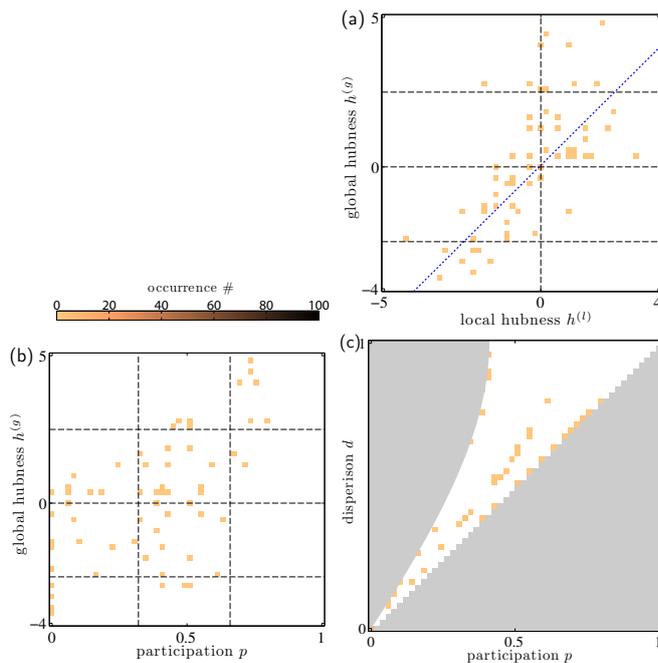
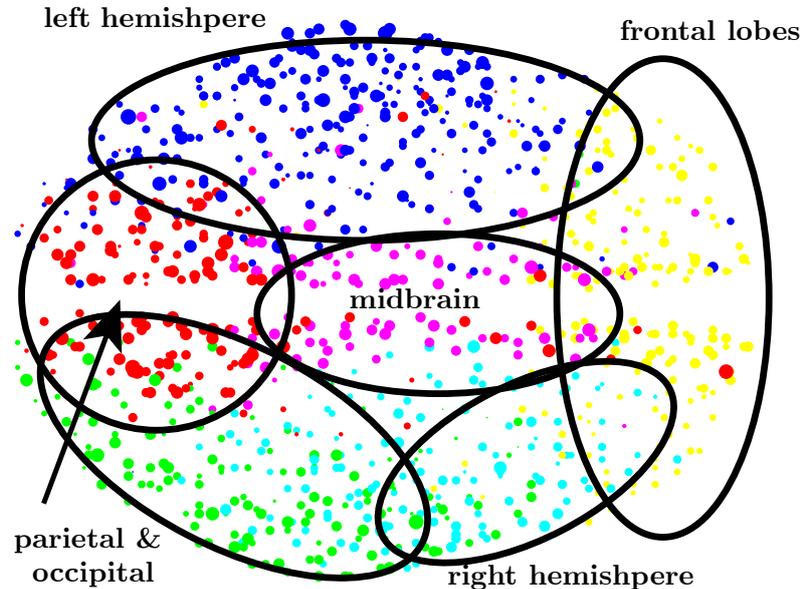


Figure 36: Mesoscale metrics density plots for macaque brain connectivity network. Panel (a) shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, panel (b) participation p against global hubness $h^{(g)}$, and panel (c) participation p against dispersion d .

Human

Figure 37: Dorsal view of the human brain nodes at the physical location. Dot's size indicates the degree of each node and the colour the community it belongs to as detected by the Louvain algorithm. Partitioning is spatially organised into modules that are also functionally grouped. However, some nodes are grouped into distant modules. For example a high degree node in the frontal lobe (yellow) is affiliated with the parietal lobe (red). Such nodes will have a connector function between the modules.



[52] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6(7):e159, 2008

For the human brain anatomical connectivity we used data first discussed in [52]. The connectivity is derived from non-invasive diffusion MRI able to detect white matter tracts that crisscross the human brain. We work with a binary adjacency matrix of a parcellation into $n = 994$ brain regions, after eliminating four isolated nodes, interconnected by $m = 13520$ white matter tracts. The network is partitioned into six modules of sizes 177, 143, 104, 156, 155 and 259. Those modules are illustrated with the node's dorsal position in Fig. 37. We see that the modules are strongly spatially organised and furthermore those modules show functional tasks: Firstly we detect that outer left and right hemispheres are forming separated modules, where the right hemisphere itself is subdivided into a rear and middle part. Furthermore the frontal lobes of both hemispheres form a joint module. Similarly the midbrain area with the limbic systems forms a central module. In the rear of the brain the parietal and occipital lobes form a joint module. Although most nodes are forming modules with spatially close nodes some of them belong to very distant modules, likely giving them a connector function between modules.

Now we want to investigate the functional roles of the different nodes in this mesoscale structure in more detail: In the corresponding plots of Fig. 38 we can clearly detect again the presence of hubs. In contrast to the observations of the two previous networks, in this case the hubs only show intermediate values of participation as observed in the $(p, h^{(g)})$ plot. The majority of nodes show a very low hubness and also very low participation indicating a well-defined modular organisation as in the networks of the *C. elegans* and of the cat. The lack of a high participation necessary for integration

is a very interesting discrepancy with the other networks. One possible explanation for this difference might be the known limitation of tractography methods to recover long-range, especially those crossing through the *corpus callosum* from one brain hemisphere to the other. This limitation could have reduced the number of connections in the data between the homologous hubs in both hemispheres and let their participation appear diminished.

Conclusion

Summarising we can state that the analysis of different topological roles that nodes can play inside the mesoscale structure of a network can give valuable insights in the structure and functionality of neuronal networks: we are able to link the functional task of information integration with the structural of connector and kinless hubs. On the other side the segregated information processing can only be achieved by nodes that are peripheral without strong connections to other modules. Both structural properties are present in the four neuronal networks we analysed. In particular we highlight the striking similarities in the structure of the *C. elegans* and the cat despite being two networks of very different scales. The first consists of single neurons that are connected with axons and the latter brain areas linked by white matter tracts, both consisting of millions of cells themselves. The omnipresence of integratory hubs and strong modular structure gives the neuronal networks excellent information processing abilities.

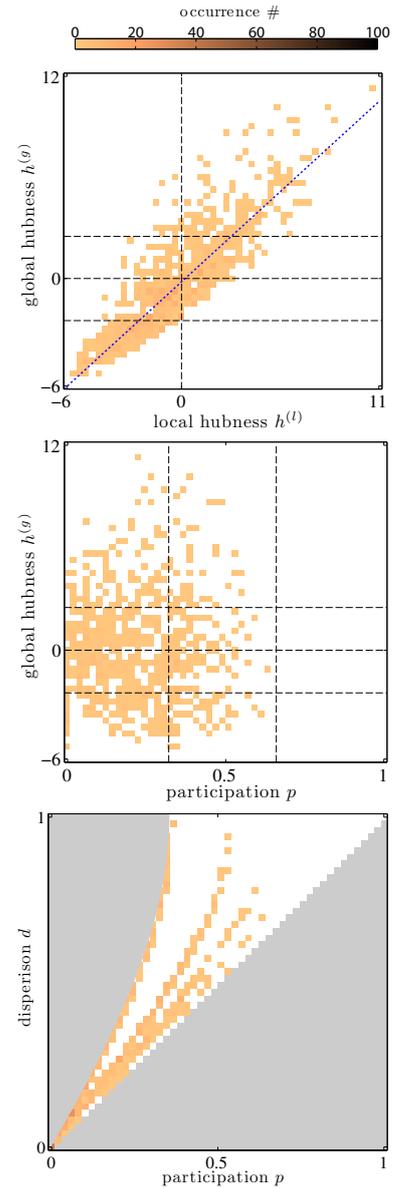


Figure 38: Mesoscale metrics density plots for human brain connectivity network. Upper panel shows local hubness $h^{(l)}$ against global hubness $h^{(g)}$, middle panel participation p against global hubness $h^{(g)}$, and lowest panel participation p against dispersion d .

Climate networks – Investigating the Indian Summer Monsoon

Although network theory has been successfully applied to many different technological and scientific problems it has only recently been adopted to climate research. This mainly focusses on interaction-networks, where links and their strength are based on a correlation measurement between chosen climate variables at different locations [53] [54]. Mutual information in combination with betweenness centrality, for example, has been used to identify ocean surface currents as important regions in the climate system [55]. More recently also the community structure in precipitation events [56] as well as sea surface temperature [57] has been analysed and shown to be strongly spatially organised due to physico-geographical factors. Here we use mesoscale analysis on data from precipitation during the Indian Summer Monsoon (ISM) and extend the analysis in order to extract the different roles sites play in this mesoscale structure. This chapter is organised as follows: Firstly we briefly introduce the phenomena ISM and the topology of the Indian subcontinent. Then we discuss the used precipitation data sets APHRODITE and TRMM. Thirdly we review event synchronisation as a method to measure similarity between time series. This is followed by an analysis of the mesoscale structure of the ISM extreme precipitation. Finally we utilise hubness and participation in order to detect high impact regions in the Indian subcontinent, completed by a discussion of our findings.

The Indian Summer Monsoon

The Indian Summer Monsoon (ISM) is among the most prominent phenomena of the climate system, influencing large parts of the tropics. It originates from the seasonal reversal of wind directions which themselves are caused by differential heating of land and surrounding water masses. This results in a strong seasonality of rainfall, with enormous socioecological impact on the Indian subcontinent [58]. The economical influence is given by the high population density on the Indian subcontinent and high dependency on agriculture which is mainly rain-fed. Any fluctuations from usual precipitation might also have cascading effects on other economical sectors, such as construction. Furthermore the ISM is

[53] Jonathan F Donges, H CH Schultz, Norbert Marwan, Yong Zou, and Juergen Kurths. Investigating the topology of interacting networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 84(4):635–651, 2011

[54] AA Tsonis and PJ Roebber. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 333:497–504, 2004

[55] Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. The backbone of the climate network. *EPL (Europhysics Letters)*, 87(4):48007, 2009

[56] N Malik, N Marwan, and J Kurths. Spatial structures and directionalities in monsoonal precipitation over south asia. *Nonlinear Processes in Geophysics*, 17(5):371–381, 2010

[57] A Tantet and HA Dijkstra. An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth System Dynamics*, 5:1–14, 2014

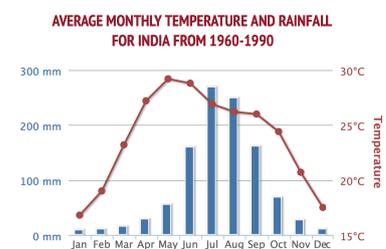


Figure 39: Average monthly rainfall and temperature for India from 1960–1990. 75 % of the annual precipitation occur during the summer monsoon months June, July, August, and September.

[58] Sulochana Gadgil. The Indian monsoon and its variability. *Annual Review of Earth and Planetary Sciences*, 31(1):429–467, 2003

[59] Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth's climate system. *Proceedings of the National Academy of Sciences*, 105(6):1786–1793, 2008

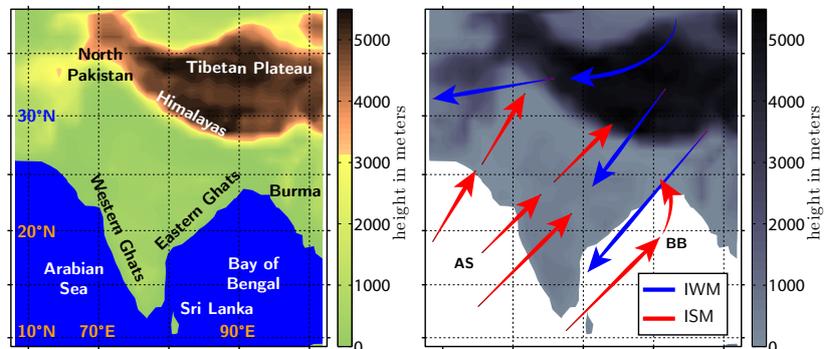
[60] Bodo Bookhagen and Douglas W Burbank. Toward a complete himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface* (2003–2012), 115(F3), 2010

considered a *tipping element* in the global climate system that might undergo drastic changes with minor changes of external parameters as planetary albedo [59]. Hence, understanding the monsoon is an important task for climatologist and also vital for the inhabitants of this region.

The Indian peninsula receives tremendous amounts of precipitation during the months of the ISM ranging from June to September (see Fig. 39) [60]. This is mainly triggered by onshore winds from the Southwest (see Fig. 40). Those are themselves caused by the different heating of land and sea and can be considered a large-scale sea breeze. We distinguish between two branches, a western one, originating from the Arabian Sea and spreading on the southern part of the peninsula. The moist air masses hit at first the Western Ghats and later encounter the high mountains of the Himalayas. Those heights lead to precipitation in the adjacent regions, known as *orographic precipitation*. The eastern branch moves northwards in the Bay of Bengal and breaks over Bangladesh and Assam. It is reflected westwards at the Himalayas and thus both branches merge later in North Pakistan.

During winter the wind directions are inverted, leading to the influx of dry air masses from the northeastern laying central asian regions. Given the offshore flow of air masses the winter monsoon brings less moisture. However, during its passage of the Bay of Bengal it absorbs water and therefore is responsible for rain in the most southern regions of India.

Figure 40: Indian subcontinent, colour coded the elevation above sea-level: Left panel shows furthermore important topographical features. Right panel indicates main wind directions during Indian summer monsoon (ISM) with red and during winter monsoon (WM) with blue arrows. In the summer the wind is coming from South West, and during winter from the North East. The ISM splits into the Arabian Sea branch (AS) and the Bay of Bengal branch (BB).



Data

For the investigation we use two types of observational data (see table 3). Firstly we use reanalysis gridded daily rainfall data for the years 1951–2007. This rain gauge measurement was provided by the *Asian Precipitation – Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources* project (APHRODITE [61] [62]). It is reliable long-term data, however the data was only collected for the land masses and therefore no measurements over the ocean are given. Secondly, we use satellite data of the years 1998–2012 from the *Tropical Rainfall Measuring Mission* (TRMM [63]). This data offers high spatial resolution as well as coverage of both, land and sea. However, it is only available for the most recent 15 years and, therefore, long-term effects are not observable.

Both data sets were analysed previously in terms of extreme rainfall events synchronisation using the climate network approach to extract spatial structures of extreme precipitation during ISM [64]. Here we add the community detection to the analysis of the extreme rainfall synchronisation. Using the measures introduced before (hubness, participation, and dispersion) we identify the roles of single nodes in the network of extreme precipitation. This analysis can help to give insights in the important spatial patterns of the extreme precipitation during the ISM.

- [61] Akiyo Yatagai, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh. Aphrodite: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9):1401–1415, 2012
- [62] Akiyo Yatagai, Osamu Arakawa, Kenji Kamiguchi, Haruko Kawamoto, Masato I Nodzu, and Atsushi Hamada. A 44-year daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Sola*, 5:137–140, 2009
- [63] George J Huffman, David T Bolvin, Eric J Nelkin, David B Wolff, Robert F Adler, Guojun Gu, Yang Hong, Kenneth P Bowman, and Erich F Stocker. The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55, 2007
- [64] Veronika Stolbova, Paige Martin, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlin. Processes Geophys. (in press)*, 2014

| Properties | APHRODITE | TRMM |
|-----------------------|------------------------------|--|
| Period | 1/1951 – 12/2007 | 1/1998 – 12/2012 |
| Geographical coverage | (62.5°– 97.5°E, 5°– 40°N) | (62.375°– 97.125°E, 5.125°– 39.875°N) |
| Spatial resolution | 0.5° | 0.25° |
| Temporal resolution | daily precipitation | 3-hourly, resampled to daily precipitation |
| Number of grid points | 70 × 70 = 4900 | 140 × 140 = 19600 |
| Type | interpolated rain-gauge data | satellite-derived |

Table 3: Properties of the used data sets *Asian Precipitation – Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources* (APHRODITE [61]) and *Tropical Rainfall Measuring Mission* (TRMM [63]). For both data sets we are interested only in the monsoon season consisting of June, July, August, and September, thus 122 days per year.

Event synchronisation

Both data sets provide us with time series of rainfall events during the whole year. Since we are especially interested in the ISM we will restrict the discussion to the data of the ISM period from the 1st of June through the 30th of September for each year, consisting

[56] N Malik, N Marwan, and J Kurths. Spatial structures and directionalities in monsoonal precipitation over south asia. *Nonlinear Processes in Geophysics*, 17(5):371–381, 2010

[65] R Quian Quiroga, T Kreuz, and P Grassberger. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical review E*, 66(4):041904, 2002

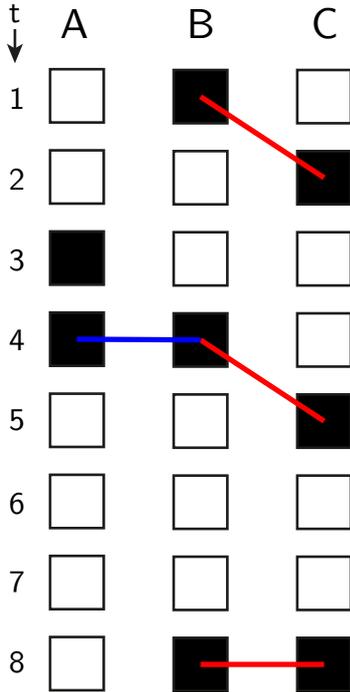


Figure 41: Pairwise event synchronization (ES) for three event time series A, B, and C of length 8: Filled boxes indicate occurrence of an event. Solid lines across time series shows that both events are inside the time lag τ_{ij} and therefore considered to be synchronised. Site A shows minor synchronisation with B as one pair of event occurs at the same time step 4. B and C are strongly synchronised (three pairs of events). Note that ES is not transitive since A does not show any synchronisation with C.

| $c(i,j)$ | A | B | C |
|----------|-----|-----|-----|
| A | | 1/2 | 0 |
| B | 1/2 | | 1/2 |
| C | 0 | 5/2 | |

Table 4: Event synchronization calculation between three example time series shown in Fig. 41. Diagonal is empty since we do not calculate self-synchronisation. In an adjacency matrix those entries would be empty since self-loops are not allowed.

of 122 days per year.

Rainfall time series inhabit often a high frequency component and are therefore less smooth and continuous as other meteorological quantities such as air temperature, or air pressure. In order to quantify the similarity of rainfall dynamics on different geographical sites, *event synchronisation* (ES) was proposed [56]. This method was introduced by Quiroga *et al.* [65] and first applied to rat and human electroencephalography (EEG) data but is suited to measure synchronisation and time delay patterns between all types of signals that inherit the occurrence of adequate *events*. In the framework of precipitation we define extreme events as days that receive rainfall amounts above the 90th percentile for all days at a given location.

The synchronisation between two grid points i and j is measured via the number of event pairs from both sites matching in time. The leading of each time series in those matching pairs can be considered. However, here we will restrict the discussion to an undirected network. An event l that occurs at grid site i at time t_l^i is considered to be synchronised with an event m occurring at site j at time t_m^j if it is within the maximal time lag of half the minimum time between two succeeding rainfall events $\tau_{lm}^{ij} = \min\{t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j\}$. Then $c(i|j)$ denotes the number of times an event appears in i shortly after it appears in j , i. e.

$$c(i,j) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij} \quad (30)$$

with

$$J_{ij} = \begin{cases} 1 & \text{if } 0 < t_l^i - t_m^j < \tau_{lm}^{ij} \\ 1/2 & \text{if } t_l^i = t_m^j \\ 0 & \text{else.} \end{cases} \quad (31)$$

Then we define the *strength of synchronisation*

$$Q_{ij} = \frac{c(i|j) + c(j|i)}{\sqrt{(s_i - 2)(s_j - 2)}} \quad (32)$$

between time series at different grid sites i and j and normalise it to be $0 \leq Q_{ij} \leq 1$. Hence $Q_{ij} = 1$ stands for complete synchronisation and $Q_{ij} = 0$ for the absence of any synchronisation. We conclude this analysis for all pairs of grid points and receive a square, symmetric matrix which represents the strength of synchronisation of extreme rainfall events between each pair of grid points. Finally we apply a threshold θ and yield a binary adjacency matrix

$$A_{ij} = \begin{cases} 1 & , \text{if } Q_{ij} > \theta, \\ 0 & , \text{else} \end{cases} \quad (33)$$

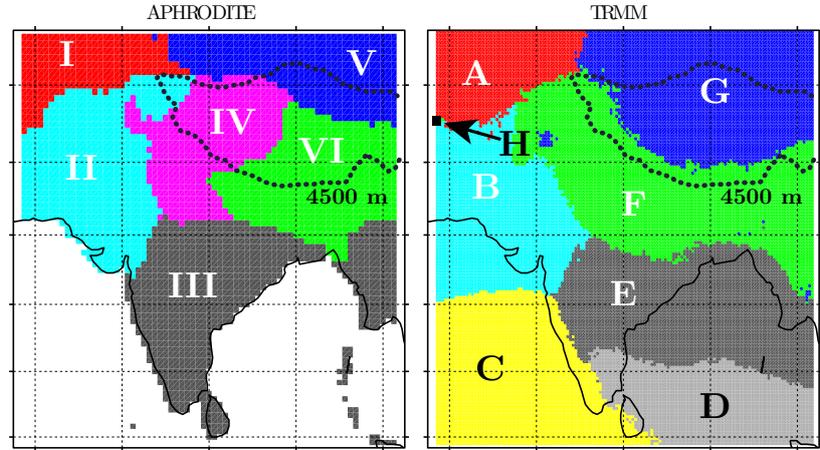
that will be used as adjacency matrix to construct an undirected, and unweighted graph. Here the threshold θ is chosen for each

data set so that 5% of the strongest links are kept. Note that we did not calculate the self-synchronisation $Q_{ii} = \frac{1}{2} \frac{s_i + s_i}{\sqrt{(s_i - 2)(s_i - 2)}} \approx 1$, consequently forbidding self-loops in order to receive a simple graph.

Having arrived with this simplified but meaningful graph representation of the extreme rain event synchronisation on the Indian subcontinent we apply the formerly introduced mesoscale graph measures in order to understand the mesoscale organisation of modules of similar precipitation.

Community detection

Figure 42: Partitions for the extreme rainfall synchronisation derived from APHRODITE and TRMM precipitation data. White spaces indicate missing data. APHRODITE is grouped in six (I–VI) and TRMM data into seven modules A–G (plus a single isolated node in its own community H, highlighted with a black square and arrow). Contour line of 4500 m is indicated with dotted to show Tibetan Plateau and sea-level with solid black line. Modules with the same colour in both data sets have a strong similarity ($J > 0.3$, see Tab. 5) and can be considered representing the same region. Note that module F is falling apart into the two modules IV and VI in the APHRODITE data set. Most communities are spatially closed sets of nodes, with some fringe at the border between modules. But community III connects the Indian peninsula with Burma. G consist of nodes mainly from the Tibetan Plateau and furthermore a group of nodes in North Pakistan, as well as a few scattered nodes in Burma. In the TRMM data set the border between modules F and G matches the location of the Himalaya. Furthermore the module C terminates at the Western Ghats.



Firstly, we analyse the community structure of the two extreme rainfall climate networks. As shown in Fig. 42 both partitions consists of modules that are strongly spatially organised. This is as expected since precipitation is mostly occurring in a local area, e.g. in moving weather fronts. Before analysing the location of the modules and their interplay with the topological and meteorological circumstances, we focus on differences and similarities of the partitions of the two data sets. The usual way for comparing two modules is the *Jaccard index*. However, it can only be applied for modules originating from the same network but the two networks we consider differ in the number of nodes they consist of. To overcome this, we pick only every second node (column- and row-wise) of the TRMM network and match it with the APHRODITE set as illustrated in Fig. 43. Since oceanic precipitation data is missing for APHRODITE we have to exclude such data points as well. Consequently, with this procedure we compare networks of the same size. Furthermore, nodes with the same index are approximately at the same geographical position, making it reasonable to compare them as if they were from the same network. Note that the reduced APHRODITE network was only used for the Jaccard index calculation and are otherwise fully considered. In Tab. 5 we show the pairwise Jaccard indexes for all combination of modules A–H from TRMM and I–VI from APHRODITE. The results are in good agreement with the observations. We can conclude that from looking at the spatial distribution of the partitions:

The modules and I and A are in very high accordance ($J(I, A) \approx 0.8$) and both represent the very northwestern region of the considered region belonging to the territories of Afghanistan and Tajikistan. Attaching southernly the communities II and B are matching well with $J(II, B) \approx 0.5$, we note that there is a minor agreement between the modules II and F ($J(II, F) \approx 0.1$) due to some joint nodes in northern Pakistan. The module F, located in the plains in front of

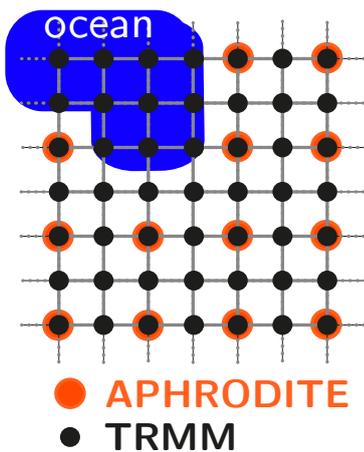


Figure 43: Jaccard index calculation between the modules in the TRMM and the APHRODITE data is undertaken for a subset of the TRMM data set: By taking into account only every second column's and row's nodes we create a subset of the TRMM nodes that matches the location of the APHRODITE set. Since this data misses oceanic precipitation data we have to exclude such data points as well, as indicated by the blue area.

the Himalayas is an interesting case since it is split up into different regions in the APHRODITE network, indicated by Jaccard indexes of 0.1, 0.2, and 0.4 with the modules III, IV, and VI, respectively. Module III is a spatially segregated module with nodes located on the Indian peninsula as well as Burma. The matching community in the TRMM set is E ($J(III, E) \approx 0.3$) and connects those parts via the Bay of Bengal, whose data is non-existent in the APHRODITE set. The modules C and D are both mostly oceanic nodes and show therefore almost no overlap with any module from APHRODITE, where those sites do not exist. Interestingly the module C ends abruptly at the west coast of the Indian subcontinent.

There the Western Ghats are located and act as an orographic barrier for the rainfall propagation. A similar situation occurs at the Himalayas (4500 m contour line given in Fig. 42), it clearly separates the Tibetan plateau (modules, V and G, respectively) from the southern regions. This shows that the high mountains segregate the extreme rainfall events of both sides. However, in the TRMM network community G does not only consists of nodes in the Tibetan Plateau but also has a cluster of nodes in North Pakistan as well as a group of few nodes in Burma. The isolated node module H in North Afghanistan is affiliated with community I in APHRODITE and therefore, shows minor similarity ($J(I, H) \approx 0.004$) with this module but no other.

| Jaccard index $J(i,j)$ | | TRMM modules | | | | | | | |
|---------------------------|-----|--------------|--------------|-------|-------|--------------|--------------|--------------|-------|
| | | A | B | C | D | E | F | G | H |
| APHRODITE modules | I | 0.795 | 0.003 | 0 | 0 | 0 | 0.003 | 0.007 | 0.004 |
| | II | 0.056 | 0.525 | 0 | 0 | 0 | 0.085 | 0.007 | 0 |
| | III | 0 | 0.030 | 0.069 | 0.057 | 0.344 | 0.112 | 0.002 | 0 |
| | IV | 0 | 0.002 | 0 | 0 | 0 | 0.229 | 0.157 | 0 |
| | V | 0.004 | 0 | 0 | 0 | 0 | 0 | 0.569 | 0 |
| | VI | 0 | 0 | 0 | 0 | 0.008 | 0.352 | 0.125 | 0 |

Table 5: Pairwise Jaccard Indexes $J(i,j)$ between the modules derived from the TRMM data set (A–H) and those from the APHRODITE data set (I–VI). High J indicates that both modules are at the same geographical position, $J = 0$ instead implies that there is no spatial overlap between the modules. Strong similarities $J > 0.3$ are indicated in bold. Note that the next highest index $J = 0.229$ is given for modules IV and F. But F is already matched with VI and therefore it would not be an injective mapping, as wanted in our case of colouring two communities from different networks with the same colour.

Concluding and summarising results from both networks we see a clear layered structure in the regions of similar rainfall events, perpendicular to the main wind direction Southwest: C and D form the oceanic regions, B (II) and E (III), the coast regions where the moist air first goes onshore, and finally the Indian plains F (IV & VI) where extreme rainfall occurs when the moist air masses hit the Himalayas. The division of those layers in western and eastern modules indicates the areas of influence of the Arabian Sea branch and Bay of Bengal branch, respectively. The Tibetan Plateau G (V) is a separate module since it is separated by the Himalayas as orographic barrier from the North Indian River Plain. Continental air masses mainly influence Northeast Pakistan (A, I) and thus a separated module is formed there.

This mesoscale analysis is an appropriate tool for comparing networks from data sets that gives a simplified but informative diagnostic on the similarities and differences between them. Here, the main difference lies in the unconsidered oceans for the

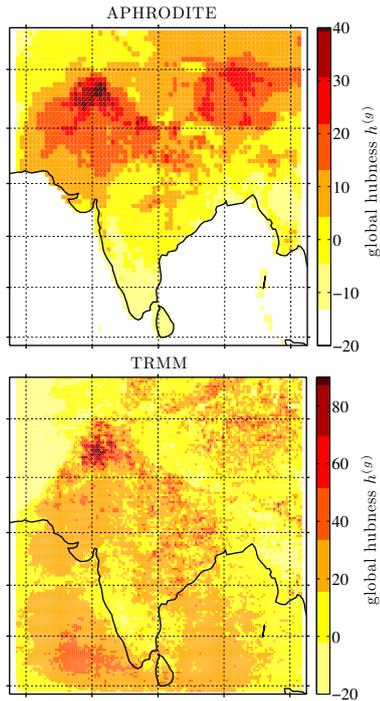


Figure 44: Global hubness for APHRODITE (upper panel) and TRMM (lower panel). North Pakistan, the Tibetan Plateau, and Eastern Ghats are regions of high hubness in both data sets. In TRMM the oceanic sites at the Arabian Sea and Bay of Bengal show high hubness as well.

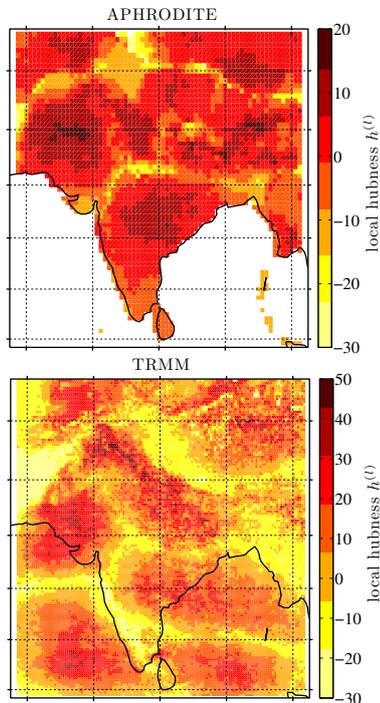


Figure 45: Local hubness for APHRODITE (upper panel) and TRMM (lower panel). In both data sets the communities show a spatial core-periphery structure: Sites in the centre of modules have high hubness whereas those close to other modules show lower values. Sites of particularly low hubness are located at the Himalayas and the Western Ghats.

APHRODITE, accompanied with a splitting of the North Indian River Plain module F in a western (IV) and eastern (VI) part.

Hubness

Here we analyse the graph measure global hubness. Nodes with a high hubness indicate that the corresponding site is strongly synchronised with other sites. Since it is a pure local measure it counts all neighbours and does not take into account the community structure. We furthermore calculate the local hubness, where we only count internal edges and compare nodes exclusively with nodes in the same module.

In the Fig. 44 we show the hubness for both precipitation networks. We see that the overall hubness is low $h < 20$ with some distinct regions of high hubness: In both data sets Northwestern Pakistan and the Tibetan Plateau are regions with particular high hubness nodes. In the TRMM network where oceanic precipitation is available, we see that Arabian Sea shows higher hubness as well, although not as high as the indicated continental regions. Intermediate hubness occurs in the southern part of the Bay of Bengal. Interestingly the Eastern Ghats shows high hubness in the TRMM network but not in the APRODITE. This indicates that Eastern Ghats' neighbours in APHRODITE are mainly oceanic sites that are missing in TRMM. The formation of the module E, consisting of the Eastern Ghats and the northern parts of the Bay of Bengal, in APHRODITE corroborates this hypothesis.

The local hubness (see Fig. 45) is an interesting measure of the precipitation synchronisation since it provides information on nodes regarding to the importance inside their module. Especially in the TRMM data it is observable that most modules show a clear *core-periphery structure*: Modules consist of a the spatial centre of nodes with high internal hubness, those nodes are densely connected with each other and referred to as core. Towards the borders to the other modules the internal hubness decreases since those spatially peripheral nodes connect not only with the own module but also to the adjacent module. This indicates a high spatial organisation of the extreme precipitation events. Compare here with the results for the regular lattice, a common model to mimic strong spatial organisation, that shows similar behaviour. Consequently, regions that are at the borders between two or more modules have particularly low local hubness. This is especially true for the Himalayas, the Western Ghats, and Southern Pakistan. The later is at the border of the available data and also effected by the Westerlies rather than the ISM and therefore only weakly synchronised with sites in India. The other regions are both known to be orographic barriers and therefore stop rainfall propagation.

In the TRMM local hubness we see that especially the oceanic modules C and D show a smooth and clear core-periphery structure. Continental modules in contrast, especially F and G, located

adjacent to the Himalayas, show a rougher structure. This implies that continental air masses are highly influenced by mountain ranges on a much smaller spatial and temporal scale than the oceanic counterpart. This is especially visible for the TRMM based network that has twice the spatial resolution.

Participation and Dispersion

We now use the detected modular structure of the synchronisation of extreme rainfall events to spot regions that are strongly interconnected to regions in other modules and such sites that are only locally connected. The measures of participation and dispersion are suited for this analysis and allow us to quantify the influence of nodes, whether it is exclusively on sites in their own module or towards other modules.

In Fig. 46 we show the participation p and in Fig. 47 the dispersion d for all sites of the two networks. We observe a very strong core-periphery structure due to localised links: Nodes in the centre of the modules tend to have minimal participation and dispersion $p = d = 0$, whereas nodes at the borders between modules interconnect with all adjacent modules and have a high participation p and up to maximal dispersion $d \rightarrow 1$. Such border regions are: Himalayas (as border between the Indian plains), the Tibetan Plateau, North Pakistan, as well as Burma. However, there are some regions with high participation that are not located at the border between modules: The Western Ghats are peripheral for region III of the APHRODITE data set but only the most northern part is adjacent to the Pakistanian region II. Nevertheless also the most southern part of the Western Ghats show almost maximal dispersion. This indicates that the extreme rainfall there is not only synchronised with the locally close sites in India but with sites in Pakistan, too. The picture is completed when the oceanic data of TRMM is included: The Western Ghats do not form a module with the continental Indian nodes but with the Arabic Sea. The Arabic Sea itself is closely connected to the Pakistan coastal regions and therefore a synchronisation of extreme rainfall events is detectable.

Overall we conclude that high participation and dispersion coincide strongly with borders between modules. As discussed in the preceding section they are themselves shaped by orographic barriers in interplay with the movement of air masses. Measuring participation provides further information on how sharp those borders are: Between modules E and F, for example, the region of nodes with high participation is quite broad, indicating that no sharp separation occurs. At the Western Ghats in contrast the region of high participation nodes is narrow, indicating a clear distinction of extreme rainfall events synchronisation on the eastern and western side of the mountain ranges.

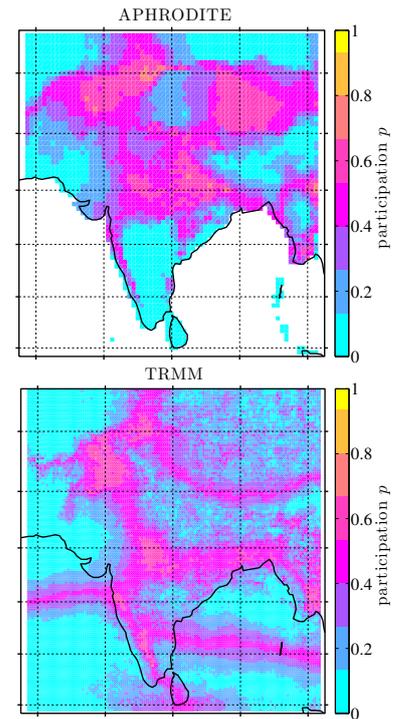


Figure 46: Participation for APHRODITE (upper panel) and TRMM (lower panel). Participation is low in the centre of modules and high at the border of those. Note that natural barriers as the Western Ghats and the Himalayas show distinct and narrow borders, unlike low-elevation areas.

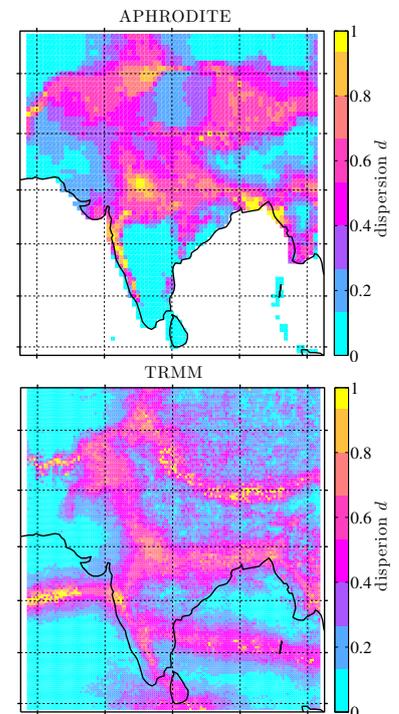


Figure 47: Dispersion for APHRODITE (upper panel) and TRMM (lower panel). Similar characteristics as for the participation. Some sites at the borders show almost maximal dispersion $d = 1$, indicating that they are exactly between two modules. The influential region in North Pakistan in contrast does not show increased dispersion in comparison to participation since it connects to all modules.

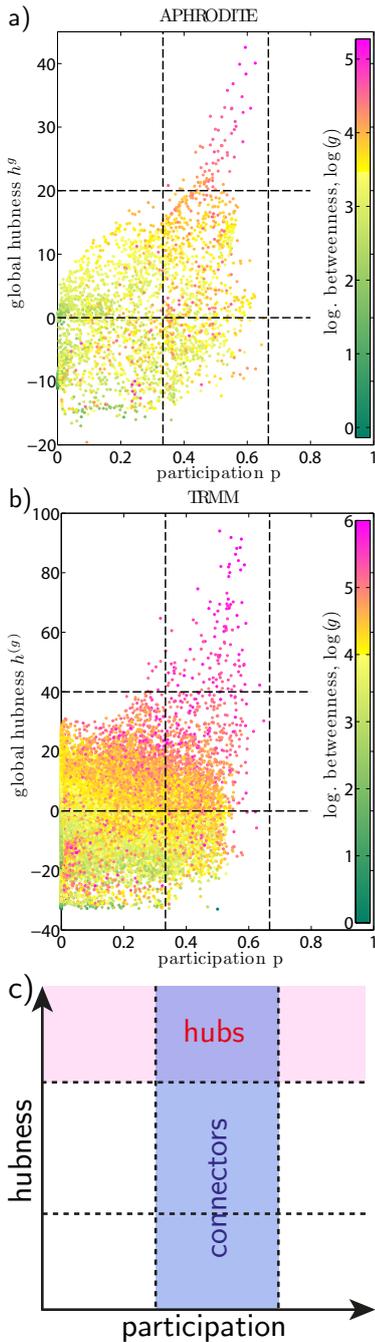


Figure 48: Participation p and global hubness h^g are used for the 2-dimensional detection of roles of nodes in the precipitation network. Colour of nodes indicate their betweenness centrality g . Panel a: Scatter plot for the APHRODITE data set. Panel b: Scatter plot for the TRMM data set. Panel c: Schematic representation of the roles that have distinguished position in those climate networks: Hubs are nodes that are strongly synchronised with many other sites and connectors are those that synchronised not only with nodes in their own module but also with sites from other modules. Especially important are regions with nodes that fall in both categories and hence connector hubs.

Role detection - Unifying the measures

Finally, we analyse the individual nodes contribution to the mesoscale structure of the in those two extreme rainfall networks. Firstly we investigate the (h, p) -scatter plots (see Fig. 48). It allows us to classify nodes into different structural roles and investigate their position in the region (see Fig. 50). Additionally we compare our results with the betweenness centrality measure g .

Scatter plots for both networks show that all nodes are in the range from connector to peripheral nodes, with the majority being peripheral. Those nodes with very low participation also tend to have small to intermediate hubness. The connector nodes in contrast are in the range from low to very high hubness. The nodes with hubness ≈ 0 are nodes that have the mean degree, that indicates that they do not have an outstanding position in terms of their local influence. Therefore, they must be located at the borders between modules (see Fig. 49). The connector hubs do not only show connections to different modules but have also many of those. Therefore, these nodes are at sites that are strongly influential as extreme rainfall events at those sites are synchronised with many other sites from different areas of the region. Interestingly, there are almost no peripheral hubs detectable. Those would be peripheral regions that are highly influential in their own module but not in others.

Investigating the betweenness g , another global measure of influence, we see that this analysis matches our results: The peripheral non-hubs tend to show low levels of betweenness since they are neither locally well connected nor to other modules. Nodes with higher degree show also increased betweenness due to the better local connectivity. Connector nodes with intermediate hubness show quite high betweenness since the effects of increased local connectivity is combined with the cross module links; letting them lay on the shortest path between a lot of pairs of nodes. Those are nodes at the borders between modules (see Fig. 49). Finally the connector hubs show the highest observed values of betweenness since they have many long-range connections to other modules that serve as shortcuts for the shortest paths between nodes that are spatially distant.

Nevertheless in the TRMM network some nodes with high betweenness do not fall in the class of connector hubs. Those are nodes that are detected as important by the path based measure betweenness but not in the mesoscale organisation. We now analyse their position on the Indian subcontinent (see Fig. 50). The red marked sites are the connector hubs, they are almost exclusively in North Pakistan. The non-hub connectors are detectable at the borders between the modules: Of special meteorological importance are here the Himalayas and the Western and Eastern Ghats. As discussed above we expect only the connector nodes (red as hubs, blue non-hubs) to show high betweenness. To test this we plot all

nodes with high betweenness that do not fall into this category as purple. We note that there are no of such undetected high betweenness nodes in the APHRODITE data set. Though there are some in the TRMM set. Those sites are scattered all over the Indian subcontinent and do not show a certain geographical area that we might have missed with our tools. Therefore the mesoscale analysis is an appropriate tool for detecting influential regions in such a interaction network from precipitation data.

Discussion

In this chapter we analysed the mesoscale structure of the synchronisation of high precipitation events during the ISM. We accomplished this in two steps: Firstly a community detection analysis and afterwards we analysed the roles that different sites play in this modular organisation.

We find that the communities are strongly spatially structured since precipitation occurs mostly in the form of moving rain clouds. We also showed that the communities detected from two different data sets show striking similarities, indicating that the encountered organisation represents features of the actual climatological phenomenon. The location of the modules is mainly driven by three factors: (1) Modules are organised in a layered structure from the Southwest, where the monsoon air masses originate towards the Tibetan Plateau. (2) Inside this layered structure a division in western and eastern parts occurs, as the monsoon divides into two branches, one from the Arabian Sea and the other from the Bay of Bengal. (3) Borders between modules are furthermore strongly affected by natural orographic barriers as the Himalayas and the Western Ghats.

The sites play diverse roles inside this mesoscale organisation: Most of the nodes do show a low hubness, are located in the core of modules, and have therefore low participation. Nodes at the borders between modules show higher participation, making them connectors. The width of the region of connector nodes enables us to distinguish sharp and imprecise borders between modules. There we detect that borders originating from natural barriers as mountain ranges are sharper than the ones where the two ISM branches meet.

Furthermore the role detection is used to detect regions that are very influential during the ISM. Our tools show, in coincidence with former attempts, that North Pakistan is the area of most impact during the Indian summer since both branches of the monsoon merge there. Therefore it is synchronised with sites from all over the subcontinent, resulting in high hubness and betweenness.

Finally the role detection is a primer of centrality in terms of betweenness. The connector hubs show the highest centrality values as their long-range connections serve as shortcuts to other modules. Border nodes between modules show high betweenness as well as

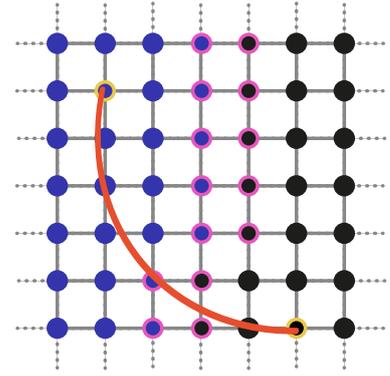


Figure 49: Role of connectors illustrated on a regular lattice, partitioned into two modules (blue and black). Nodes with pink border are at the border between two modules, they will have increased participation and are connector non-hubs. Long range connections, as illustrated with red arc will lead to increased hubness and participation for the connected nodes (yellow border) and therefore they are connector hubs. Border connector nodes will show higher betweenness than normal nodes, the connector hubs will have highest betweenness since the long-range connections are shortcuts to other modules.

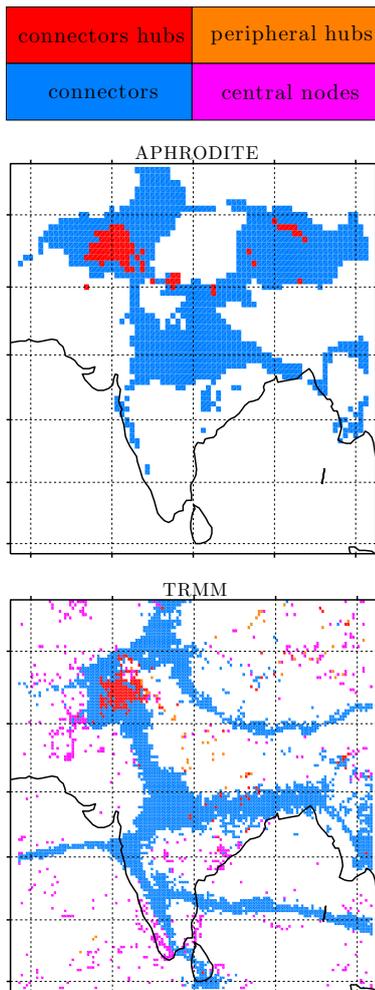


Figure 50: Functional roles of nodes: Red nodes indicate connector hubs, blue nodes are connectors, orange hubs without strong participation, the rare group of peripheral hubs. Purple sites are nodes with high betweenness that were not detected by the roles classification as connector hubs.

shortest paths between the modules lead through them.

This is an interesting feature as the detection of all shortest paths is a computationally very demanding task. The role detection instead is extremely fast, especially when the organisation into modules is already known.

We conclude that the here presented mesoscale analysis is an appropriate tool for detecting areas of similar behaviour in climate data, as well as the characterisation of their borders and the identification of particularly influential regions.

The multilayer Airport Transportation Network

Introduction

Understanding the structure of human mobility networks, especially traffic infrastructure is of high importance for socio-economic reasons, including the prediction of pandemics [66]. While a considerable effort has recently been devoted to the characterisation of its structure, their multiplex nature has not been considered appropriately. The air transportation network is furthermore an interesting study object since it shows scale-free, as well as small-world characteristics [67]. It inherits a modular structure that is shaped by geographical proximity as well as geopolitical factors. This chapter is organised as follows: Firstly we introduce the data used in this work. Afterwards we utilise a multilayer variant of the participation index together with the degree to distinguish the roles of different airports in the multilayer network. Finally we use this to show that low cost airlines inherit striking structural differences in comparison to major airlines.

Multilayer data

The Airport Transportation Network (ATN) is as many traffic networks a good example for a network consisting of multiple layers: Nodes represent airports and links stand for direct flights between two airports. We distinguish between the different commercial airlines and therefore each layer consists of flights by the same carrier. The ATN is an interesting study object since it has undergone significant growth in the last decades and is shaped by a variety of socio-economical and geographical factors.

Here we will restrict the discussion to the European airport network which consists of $n = 450$ airports that are approached by $b = 37$ different airlines, each forming an individual layer [68]. The data was collected for the year 2011. It consists of quite different types of airlines; ranging from major airlines as Lufthansa, over low cost airlines as Ryanair, up to regional airlines as Air Nostrum. Furthermore it includes Netjets, a business jet rental airline, as well as TNT and European Air Transport, two cargo airlines. We note that this data set does not cover the whole worldwide airport trans-

[66] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, 2004

[67] Roger Guimerà, Stefano Mossa, Adrian Turtchi, and LA Nunes Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005

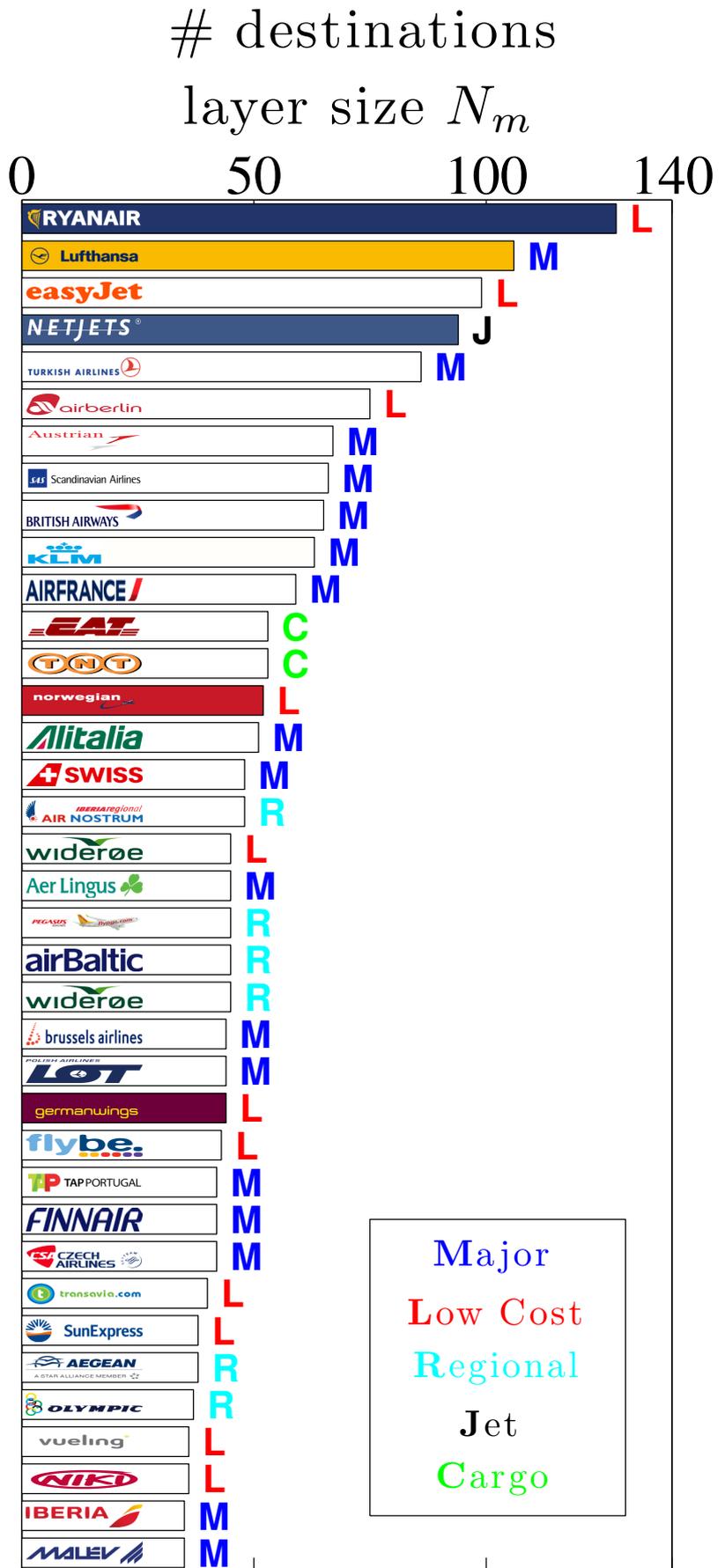
[68] Alessio Cardillo, Jesús Gómez-Gardeñes, Massimiliano Zanin, Miguel Romance, David Papo, Francisco del Pozo, and Stefano Boccaletti. Emergence of network features from multiplexity. *Scientific reports*, 3, 2013

portation network. Since it is an excerpt some of the results have to be lighted in this way. Turkey for example, is geographically peripheral from a Eurocentric perspective, although considering flights to Asia would change the picture.

First of all we want to investigate the separate layers/airlines that make up the multilayer ATN graph. As previously defined, the layer size N_m is given as the number of nodes that have a non-zero degree for each airline m . Therefore it is identical with the number of destinations an airline flies to. As shown in Fig. 51 the number of destinations N_m is widely heterogeneously distributed, ranging from only 35 destinations up to 128 destinations. Taking into account the different types of airlines; major, low cost, regional, jet rental, and cargo carrier we detect the following: Major airlines (flag carriers, or former national airlines) are ranging from a small number of destinations (Malev Airlines and Iberian, 35) up to very large (Lufthansa, 106). Same is true for low cost airlines (from Niki, 36 up to Ryanair, 128). Ryanair is also overall the airline with the highest number of destinations. Regional airlines tend to have small (Olympic Air, 37) up to intermediate numbers (Norwegian Air Shuttle, 52) of connected airports since they operate in a spatially limited region. Netjets is able to provide a very high number of 94 destinations since it offers fractional ownership and rental of business jets. The two cargo carriers TNT and European Air Transport both operate from 53 airports whereof 28 are approached by both airlines.

In Fig. 53 we show the ATN of six different airlines with the physical location of their destinations in Europe. Those include major airlines (Lufthansa, Scandinavian Airlines, and Turkish Airlines), as well as low cost airlines (Ryanair and Pegasus Airlines), and a regional airline (Wideroe). All airlines are spatially organised and serve more strongly particular parts of Europe. Scandinavian Airlines for example offers many direct flights between airports in northern Europe but a smaller amount of destinations in the west, and almost none to the Balkan or Turkey. This can be formulated as a *core-periphery* structure similar to the climate network in the previous chapter or for the regular lattice model: Layers tend to connect nodes that are close to each other and are less likely to form edges between pairs of nodes that are far away. However, airlines do not connect nodes that are extremely close to each other, e.g. two airports of the same city like London Heathrow and London Luton. In the climate network spatially adjacent nodes will show similar rainfall with high probability since weather is a localised phenomenon and therefore edges are formed between them. This localisation of the layers indicates that the different layers give rise to a modular organisation of the whole multilayer network. In contrast to a normal modular structure here nodes can be affiliated with multiple layers, similar to an overlapping community structure, indicating that those airlines are competing against each other at those airports.

Figure 51: The European airline network consists of $b = 37$ layers, each one being a single airline connecting different airports. The size of each layer N_m is defined as the number of destination for each airline, thus nodes with degree $k_m > 0$. Airlines are grouped into four different types: Major airlines (blue M), low cost airlines (red L), regional airlines (cyan R), jet rentals (black J), and cargo transporters (green C). Airlines are ordered in decreasing order of sizes N_m . We note that major airlines (flag carriers, or former national airlines) are ranging from a small number of destinations (Malev Airlines and Iberian, 35) up to very large (Lufthansa, 106). Same is true for low cost airlines (from Niki, 36 up to Ryanair, 128). Ryanair is also the airline with the highest number of destinations. Regional airlines tend to have small (Olympic Air, 37) up to intermediate numbers (Norwegian Air Shuttle, 52) of connected airports since they operate in a spatially limited region. Netjets is able to provide a very high number of 94 destinations since it offers fractional ownership and rental of business jets. The two cargo carriers TNT and European Air Transport both operate from 53 airports whereof 28 are approached by both airlines.



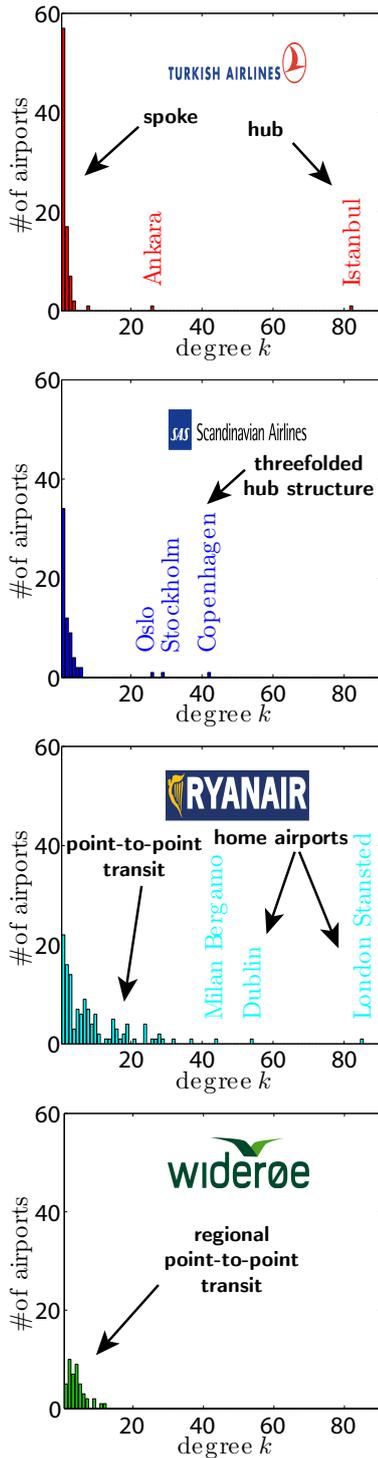


Figure 52: Degree distributions for four different airlines: Turkish Airlines shows a clear *hub-and-spoke* structure with Istanbul as main, Ankara as secondary hub and many regional airports with degree 1. Scandinavian Airlines is a multinational flag carrier and shows a less distinct three-folded hub structure. Ryanair's home airports are London Stansted and Dublin, however the airline operates in a point-to-point transit as indicated by many airports with intermediate number of connections. Widerøe shows as a regional airline point-to-point transit with only a small number of destinations.

A particularly interesting modular structure in interplay with a scale-free degree distribution is known as *hub and spoke*: Airlines maintain highly frequented airports that serve as hubs. This central airport is then connected to many smaller local airports. Overall this structure enables a high efficiency in terms of mean shortest path-lengths. In addition the smaller number of flights in comparison with a full graph allows more efficient use of resources. For example, planes are more likely to fly at full capacity from the central airports and can fly multiple times a day.

This organisation is especially observable with airlines from Turkey: For both, Turkish Airlines and Pegasus Airlines, Istanbul serves as the main hub that connects to many domestic and international locations (see Fig. 52). The domestic airports are mostly peripheral airports with only a small number of airlines serving them. In contrast the international destinations are mostly hubs in other countries in order to enable a high number and fast connection flights. Here, the two Turkish Airlines differentiate in their strategy, too: The flag carrier Turkish Airlines offers flights to many European cities, especially capitals as Madrid or Dublin, whereas the low cost airline Pegasus focuses on the central European region, where Turkish emigrants generate a high demand for flights.

The example of the two Turkish carriers shows that the interplay of the different layers is truly a complex one that is driven by a variety of processes: On the one hand airlines are competing with each other for passengers on the same flights (e.g. the direct flight Istanbul – Ankara). On the other hand the infrastructure provided by other airlines can be used to allow passengers to catch connecting flights. Therefore both airlines connect to European hubs in order to allow passengers to change there to flights to all other European destinations. This creates symbiotic effects since even peripheral regions are connected to each other with only a small number of steps as indicated by the small average path-length of $L = 2.7$.

Scandinavian Airlines is the flag carrier of Denmark, Norway, and Sweden. Therefore it shows a three-folded hub structure with airports in all three capitals Copenhagen, Oslo, and Stockholm. Lufthansa is the formerly owned by the German government and therefore focusses on the German market: The hub structure is less distinct with major airports in industrial strong regions in South Germany (Munich), West Germany (Frankfurt), and Belgium (Brussels).

Ryanair is the most successful low cost airline in Europe. It covers destinations in all parts of Europe with an emphasis on central Europe. In contrast to the other airlines it does not show the hub organisation but rather a *point-to-point* transit: Most destinations are connected by direct flights and changing the flight at a central hub is not necessary. For example many direct connections are offered from the British Isles to Spain and Germany, and also between those two regions. In a hub and spoke structure as employed by

Lufthansa there are no direct flights between Spain and Britain. Despite not using a central hub Ryanair also uses London Stansted and Dublin as home airports with an increased number of connections. However, Stansted connects only with 50% of Ryanair's destinations, Istanbul instead with 95% of Turkish Airlines' and 80% of Pegasus Airlines' airports. This makes Istanbul truly a hub with a centralised position for those two airlines.

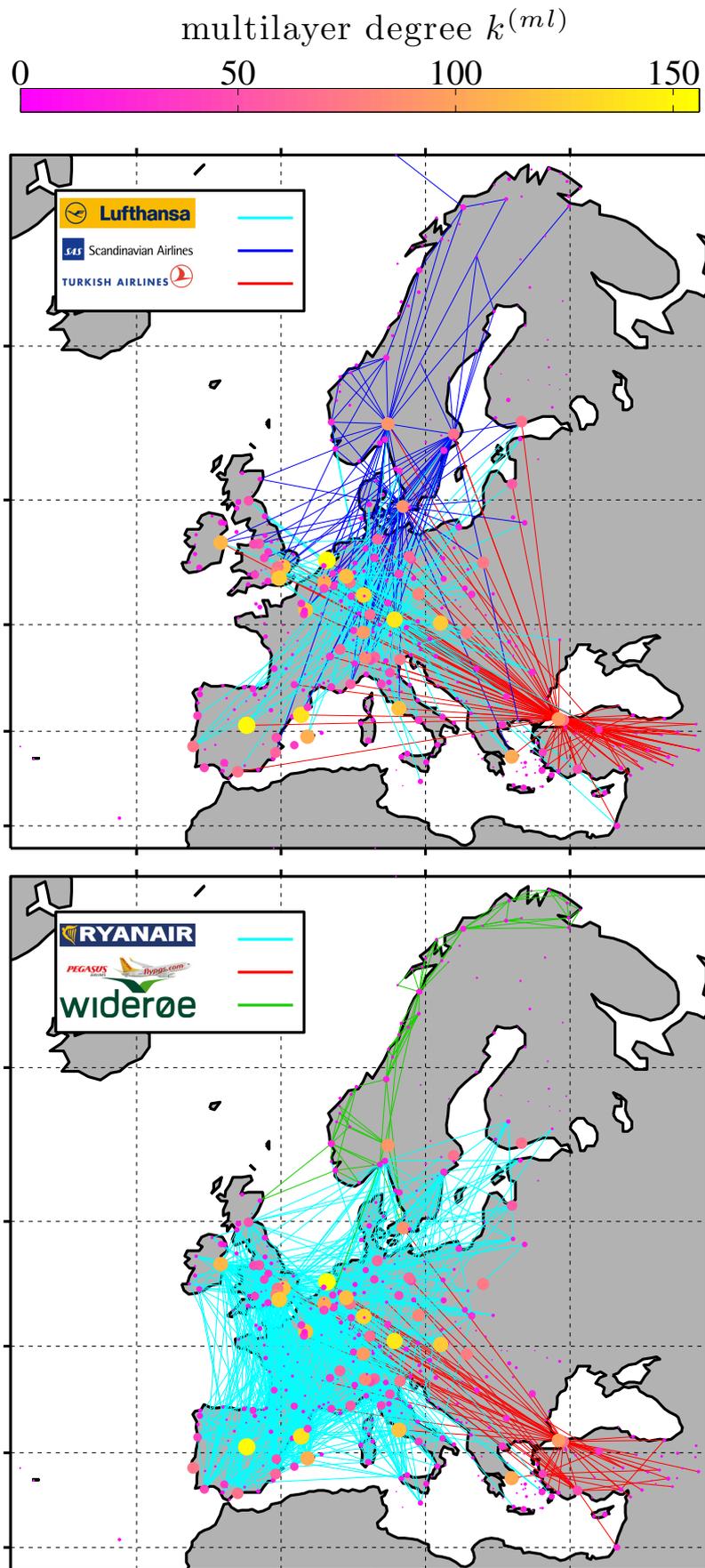


Figure 53: Six layers (airlines) of the multilayer ATN are shown with the physical location of the airport nodes in Europe. Node's size and colour indicate their multilayer degree $k^{(ml)}$. Nodes are connected if there is a direct flight between the two airports. Upper panel shows three major airlines: Cyan lines indicate flights by Lufthansa, blue lines flights by Scandinavian Airlines, and red lines flights by Turkish Airlines. All three airlines are spatially organised and show a core-periphery structure known as *hub and spoke*: For Turkish Airlines Istanbul is the hub that connects with other Turkish airports as well as international destinations. Scandinavian Airlines shows a less distinct hub structure since it is a multinational cooperation and therefore has hubs in the capitals Copenhagen, Oslo, and Stockholm. Lufthansa focuses on the German market. The hub structure is less distinct with major airports in industrial strong regions in South Germany (Munich), West Germany (Frankfurt), and Belgium (Brussels). Lower Panel shows two low cost airlines and a regional airline: Cyan lines indicate flights by Ryanair and red lines flights by Pegasus airlines, both low cost airlines. Ryanair is the airline with the highest number of connected airports (128) and connects all parts of central and western Europe. Interestingly they do not fly to airports in the peripheral regions of Scandinavia as well as Turkey. Pegasus Airline is a Turkish low cost airline and has as Turkish Airlines Istanbul as main hub. From there it connects to domestic airports as well as cities in central Europe. Due to the same regional hub it is for many flights a direct competitor with the major airline Turkish Airlines. Widerøe, a Norwegian regional airline (indicated with green lines) connects peripheral airports in Norway and a small number of international destinations in the proximity as Scotland or Sweden. Partially Widerøe and Scandinavian Airlines are competitors for Norwegian flights, however the major airline has way more international flights and the regional airline instead provides better coverage of inter-norwegian flights.

Participation and degree of individual airports

Now we analyse the position each airport plays in the multilayer structure of the ATN. For this we use a two-dimensional mapping similar to the ones used for the mesoscale organisation in communities: On the one side we take the multilayer degree $k^{(ml)}$ and on the other side the multilayer participation $p^{(ml)}$. The first measures the number of airports a city is connected to with multiple counts for flights by different airlines. The participation instead measures how distributed the airports flights are among the different airlines: $p^{(ml)} = 0$ means an airport is approached by only a single airline, whereas $p^{(ml)} = 1$ indicates that a variety of airlines is flying to this location.

In Fig. 54 this mapping is shown in the upper panel in a scatter plot as well as with the airports at their spatial location in the lower panel. First we notice that the participation ranges from ultra-peripheral nodes ($p = 0$) to kinless nodes ($p > 2/3$). This indicates a wide structural variability the nodes play in the ATN. The ultra-peripheral nodes tend to have low multilayer degrees $k^{(ml)}$. Those are mostly small, regional airports that are only connected by one airline in order to provide accessibility of this region by plane. They are particularly observable in geographically periphery regions of Europe as Scandinavia or Turkey. However, the two ultra-peripheral airports with the highest degree are interesting special cases: Frankfurt-Hahn Airport (EDFH) is a major airport for Ryanair, offering 28 different destinations, but none for other airlines, since those use the close Frankfurt Airport (EDDF). Liège Airport (EBLG) offers 35 different destinations. Since it is a cargo airport no passenger airline is flying here, but only TNT Airlines, leading to minimal participation $p^{(ml)} = 0$.

The category of peripheral nodes ($p \in (0, 2/3)$) is the largest with 185 airports (see Tab.6. Those are connected by different airlines but still show a stronger affiliation with some airlines than others. An example is Stockholm Skavsta Airport (ESKN), a low cost airport 100 km outside of Stockholm. It is mainly used by Ryanair but furthermore by Wizz Air, a Hungarian low cost airline that specialises in using secondary airports all over Europe. But no other airlines flies there, leading to a small participation. Istanbul Airport (LTBA) is the only example of an a peripheral hub: It is used strongly by the two Turkish air carriers and to a much less extend by the other European airlines.

165 nodes fall in the category of connector nodes that are used by different airlines to similar amounts. An example of non-hubs in this category is Sarajevo International Airport (LQSA). It is used by a variety of different airlines but none of them offer more than three destinations. Therefore the degree $k^{(ml)}$ stays low whereas the participation $p^{(ml)}$ is fairly high. Many of the well known airports are in this category but show a high degree. Those include London Heathrow (EGLL), Frankfurt Airport (EDFF), and Madrid Airport

| | ultra-peripheral | peripheral | connector | kinless |
|---------|------------------|--------------------|------------------------------|--------------------|
| $p \in$ | $\{0\}$ | $(0, \frac{1}{3})$ | $(\frac{1}{3}, \frac{2}{3})$ | $(\frac{2}{3}, 1)$ |
| # | 53 | 185 | 165 | 17 |

Table 6: Participation of the airports grouped into four distinct groups, ultra-peripheral nodes, peripheral nodes, connector nodes, and kinless nodes.

(LEMD). Those airports are frequented by a high number of different airlines with a high number of flights. The Amsterdam Airport (EHAM) is the airport with the highest degree and falls into this category, too.

The fourth category of kinless nodes is a small (only 17 airports) but interesting one, since those airports are used by an extremely high number of different airlines. The node with the highest participation is the Bucharest Airport (LROP) and the second largest Tel Aviv Airport (LLBG). The latter one is used by a high number of European airlines since it provides access to the middle eastern region and is known for its high level of security. However, the long distance to Europe does not make multiple destinations for each airline necessary, leading to a low degree. Nodes with higher degree in this category are frequented by many different airlines from a high number of destinations. Those include Milan (LIMC), Venice (LIPZ), and Malaga (LEMG). All of those airports are so called *sun and beach* locations – Passengers from all over Europe fly there for beach holidays. The demand to fly there is so high that many airlines fly to this destinations from a variety of airports. The node with the highest degree in this category is Barcelona (LEBL), one of the world’s leading touristic centres.

Altogether we can conclude that the airports fulfil a large variety of roles in the multilayer ATN: Nodes with low degree and low participation are small, regional airports. Nodes with higher degree tend to be connected by different airlines as connectors. However some nodes can be classified as peripheral hubs since they are largely used by a particular airline. Most interestingly the nodes with high degree and high participation are holiday destinations in southern Europe and many different airlines are competing against each other for customers.

Characterising airlines via mesoscale measures

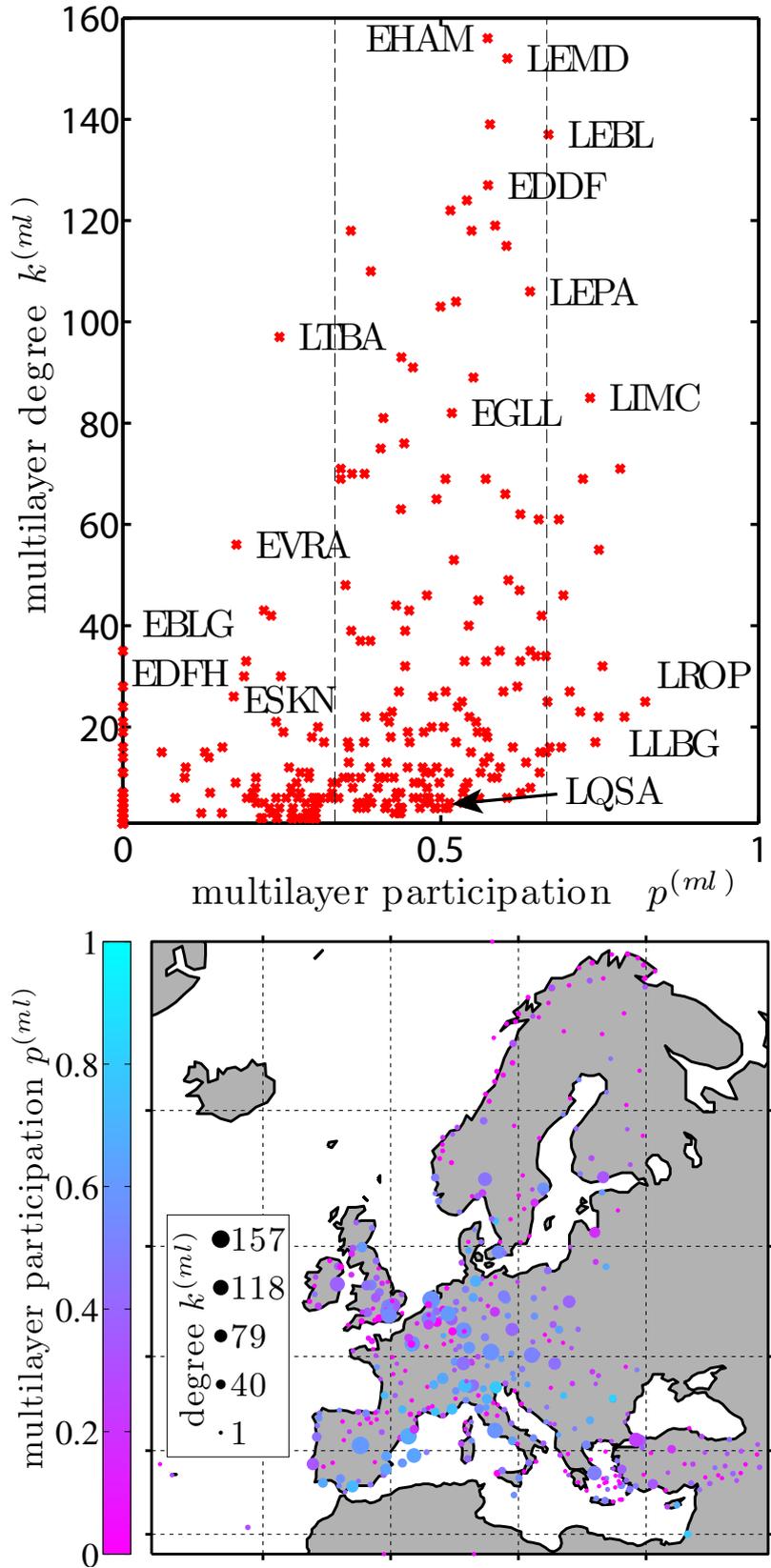
After analysing the different role each airport plays in the ATN we use this information to get an insight in the organisation of the different airlines. For this purpose we calculate the mean multilayer participation $\langle p^{(ml)} \rangle^\alpha$ and mean multilayer degree $\langle k^{(ml)} \rangle^\alpha$ for each layer α as the mean values for each node a layer connects to

$$\langle p^{(ml)} \rangle^\alpha = \left(\sum_{i|k_{i,\alpha}>0} p_i^{(ml)} \right) / N_\alpha \quad (34)$$

$$\langle k^{(ml)} \rangle^\alpha = \left(\sum_{i|k_{i,\alpha}>0} k_i^{(ml)} \right) / N_\alpha. \quad (35)$$

Thus we receive summarising information to what kind of airports the different airlines tend to connect to. In Fig. 55 we show those for all 37 different airlines and distinguish between the five types

Figure 54: Multilayer participation $p^{(ml)}$ and degree $k^{(ml)}$ for all airports in the ATN. Upper Panel: Shown as a scatter plot. Participation ranges from ultra-peripheral nodes (EBLG, Liège) to kinless nodes (LROP, Bucharest). Majority of nodes show intermediate participation. Airports with high degree tend to have intermediate participation and are for example Amsterdam (EHAM), Madrid (LEMD), and Barcelona (LEBL). Palma de Mallorca (LEPA) and Milan (LIMC) can be classified as kinless hubs since they show high degree with also high participation. Istanbul (LTBA) is the only example for a peripheral hub due to the peripheral location in Europe. Lower Panel: Participation and degree of each airport illustrated at their physical location. Size of dots indicate degree and colour the participation from zero (magenta) to 1 (cyan). Clearly we notice core-periphery structure in Europe: Airports in central Europe tend to have higher participation values than nodes in the periphery regions as Turkey or Scandinavia. But note that small airports and only a small number of connections are spread out all over Europe (little red dots). Furthermore we detect that especially the airports in southern Europe (Spanish coast, Italy) show high participation. This indicates that many different airlines are flying there and competing for customers.



of airlines. First of all we note that most airlines show intermediate values for the mean participation $\langle p^{(ml)} \rangle \approx 0.5$. This is due to the fact that this is the largest group, and furthermore the high degree nodes are mostly in this, as well. However three airlines show lower values. The regional Norwegian airline Wideroe has the lowest mean participation and at the same time also the minimal mean degree. This is caused by the peripheral location of Norway in Europe: Most destinations are small regional airports only used by Wideroe and therefore having $p^{(ml)} = 0$ with also a small degree $k^{(ml)}$ since it is not economically feasible to connect such a sparsely inhabited region with many flight routes. Same is true for Olympic Air, a regional airline from Greece. They fly to many regional airports on islands in the Aegean Sea that are not used by other European Airlines, resulting in low mean values for both indexes.

The third airline that is clearly distinguishable from the main field is Ryanair. Despite flying to many destinations in central Europe it shows low participation and degree. Ryanair is known to use many secondary airports in order to minimise flight costs that are much higher for major airports. Often they are the only (Frankfurt Hahn) or almost single (Stockholm Skasta) airline at those airports in the outskirts of the cities. Therefore the degree and participation of those locations is particularly low.

The highest values of participation and degree are achieved by the major airlines Iberia and Swiss. That means that those two connect airports that are heavily used by many different airlines. Other major airlines tend to show high mean index values in comparison to the regional and low cost airlines. This indicates that flag carriers can afford to fly to more expensive destinations since they charge more than low-cost airlines. The latter ones have to reduce their operating costs and therefore switch to smaller, less frequented airports that are cheaper to fly from.

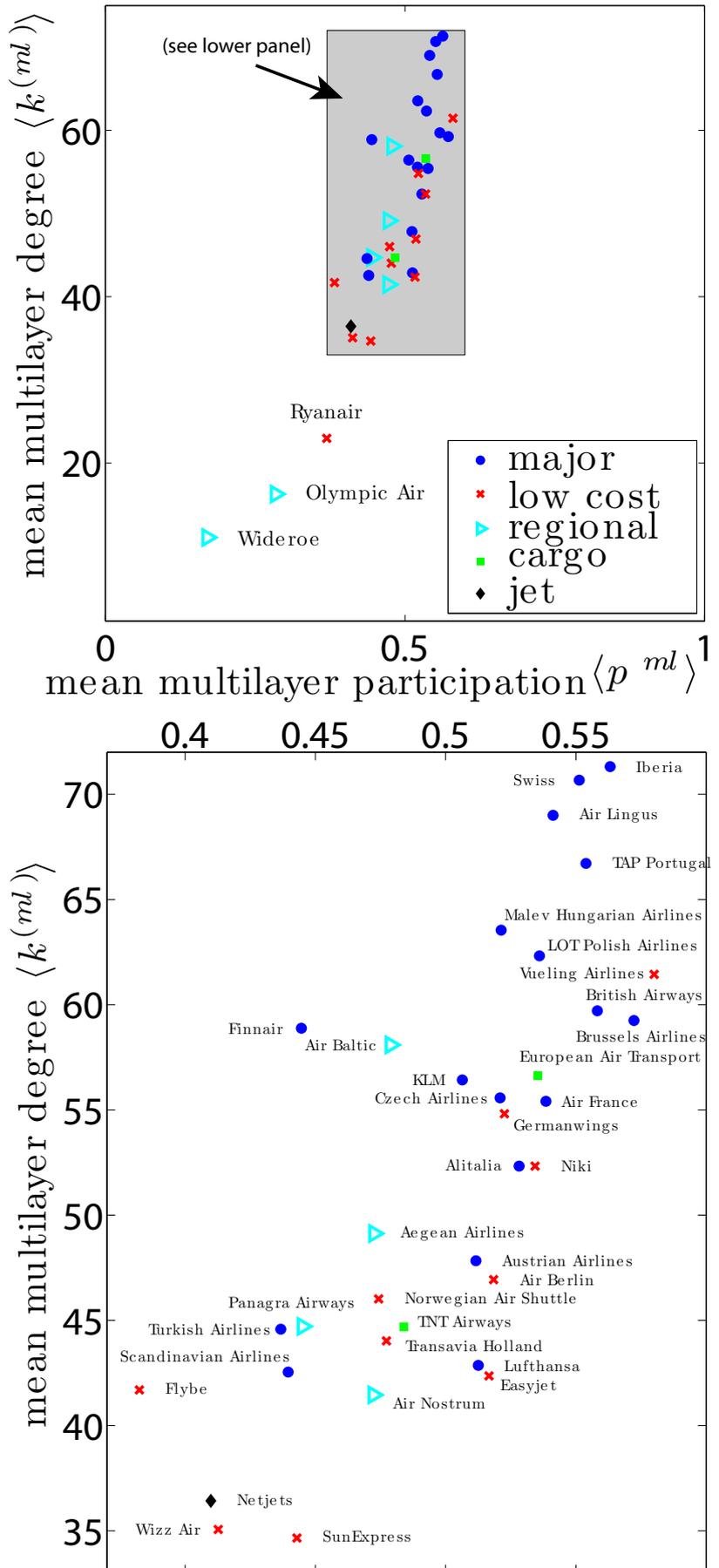
Additionally we detect the influence of the attractiveness of the Spanish holiday regions at the Mediterranean Sea. For both groups, major and low cost airlines, the airlines with highest indexes are Spanish airlines, the flag carrier Iberia and the low cost Vueling Airlines. Both airlines connect many Spanish airports and those are highly used by other airlines, as shown in the foregoing section. Therefore the Spanish airlines show high mean values.

In contrast the major airlines with the lowest participation are Turkish Airlines, Finnair, and Scandinavian Airlines. Those mainly operate in regions that are spatially peripheral in Europe and show therefore particularly low participation.

Discussion

Analysing the multilayer participation and degree of the nodes we see that airports inherit a wide diversity of roles in the European ATN. Those range from peripheral hubs over connector nodes to

Figure 55: For each layer (airline) we show the mean value of multilayer participation $\langle p^{(ml)} \rangle$ and multilayer degree $\langle k^{(ml)} \rangle$. Upper panel shows for all 37 airlines and lower panel an excerpt of the region where most airlines fall. Major airlines are indicated with blue dots, low cost airlines with red crosses, regional airlines with cyan triangles, cargo airlines with green squares, and the jet rental service with a black diamond. Most airlines show intermediate mean participation values ≈ 0.5 . Lowest mean values for both degree, and participation are reached by Wideroe and Olympic Air, two regional airlines, that are focusing on Norway and Greece respectively, two peripheral regions in Europe. Furthermore Ryanair shows low mean participation values since they often fly to small cheap airports that are not used by other airlines. Overall it is detectable that regional and low cost airlines show lower mean values for participation and degree for this reason. Major airlines in contrast use frequently used, big airports that are contacted by many other airlines. Those show high multilayer participation. Additionally the geographical position in Europe is important: The major airline with lowest participations are Turkish Airlines, Finnair, and Scandinavian Airlines, all three originate from peripheral regions in Europe and therefore have small airports as destinations that other airlines do not connect to. In contrast the major airline with highest participation and degree is Iberia, the flag carrier of Spain. It therefore connects all the airports in the Spanish holiday regions that have high participation values themselves. This effect also influences the regional and low cost airlines: the low cost airline with highest degree and participation is Vueling Airline, a Spanish low cost airline. Cargo carriers and the Jet rental service show intermediate values for both participation and degree.



kinless non-hubs. A variety of social-economical and geographical circumstances triggers those: Firstly the spatial location in Europe is influencing the amount of airlines that are flying to a destination, with central regions being more used than peripheral as Norway. Secondly the southern holiday regions in Spain and Italy are frequently used by many different airlines since customers from all over Europe want to spend their vacations there. Furthermore we detect that low cost airlines as Ryanair tend to use secondary airports that are not used by major airlines. This reduces the operating costs for the airlines, allowing them to make inexpensive tickets available to their customers.

We conclude that the participation approach is suited for the analysis of the multilayer nature of networks. It allows to distinguish between nodes that are exclusively contributing to a single layer from those that are meaningful in many or all layers. This type of analysis is a novel approach that is extendable to many different fields of networks science.

Summary and discussion

In this thesis new measures to uncover the contribution that individual nodes make in modular networks or in networks with an *a priori* classification due to meta-information of the system have been described. This approach uses four descriptors that help mapping every node's contribution. On the one hand, local and global hubness indices parametrise the relevance of nodes locally within their community and globally on the whole network. On the other hand, we have proposed measures to evaluate how distributed are the links of a node among the communities. Participation vectors representing the likelihood of nodes to belong to each community have been introduced; these account for inhomogenous relative sizes of the communities. Information in the participation vectors is reduced into two scalar indices. The dispersion index characterises how difficult it is to classify a node in one and only one community and the participation index indicates how uniformly are the links of a node distributed among *all* the communities.

We have illustrated the use of the measures applying them to both synthetic and empirical networks. The example graph in Fig. 13 has been designed to contain nodes playing many different roles. In comparison with the results from previously defined frameworks, we show that only ours is able to distinguish the richness of roles that nodes take. Results in random and scale-free networks on the one hand, and in modular networks on the other hand, show fundamental differences between them. These examples demonstrate how the outcome of the participation and dispersion indices differ when the community structure is well-defined or not.

Analysis of empirical neuronal data has shown that our method is able to detect mesoscale features that are important to the information processing of those structures: on the one side a strong community structure is indicated by the presence of many nodes with very low participation values. These are the sites of segregated information processing. On the other side the hub structure is similar to a SF network organisation. Importantly those hubs are connector or kinless hubs, showing that they are able to access information from all the different modules in order to combine and integrate it. The coexistence of integration and segregation is fundamental for the successful handling of information from multisensory input and both principles occur at multiple scales in

- [20] Gorka Zamora-López, Changsong Zhou, and Jürgen Kurths. Exploring brain function from anatomical connectivity. *Frontiers in neuroscience*, 5, 2011
- [69] Steven L Bressler and JA Kelso. Cortical coordination dynamics and cognition. *Trends in cognitive sciences*, 5(1):26–36, 2001
- [70] Joaquín M Fuster. *Cortex and mind: Unifying cognition*. Oxford university press, 2003
- [71] Murray Shanahan. A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, 17:288–303, 2007
- [14] Roger Guimera and Luís A. Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005
- [15] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005
- [16] Alex Arenas, Javier Borge-Holthoefer, Sergio Gomez, and Gorka Zamora-Lopez. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12(5):053009, 2010

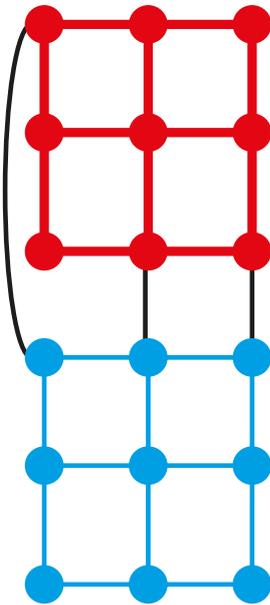


Figure 56: Network consisting of two separated modules (red and blue) that are cross-connected with three shortcuts (black edges). The network does not show a standard *small-world* effect since no node has clustering $C > 0$. Nevertheless the shortcuts lead to an effect similar to the small-world. This can be detected by an increased participation p for the nodes with shortcuts.

neuronal networks. [20] [69] [70] [71].

Comparison with previous approaches

The framework here proposed builds upon previous efforts to characterise the roles that nodes take in modular networks: the functional roles framework introduced in Guimerá & Amaral [14] [15] and the SVD approach by Arenas *et al.* [16]. A common limitation of the two frameworks is that both assume all communities to be of the same size. While this is a reasonable approximation for large networks containing several large and near-to-homogeneous communities, many real networks are small or contain inhomogeneous communities.

Here, we have taken a probabilistic approach and evaluated the likelihood of a node to belong to a community that depends not only on the degree of the node but also on the size of the community. Stacking our participation vectors would lead to a matrix similar to the contribution matrix which would additionally account for the inhomogeneity of communities. We have also proposed a measure of hubness that is consistent with the common and original understanding of hubs, say, that the degree of hubs in scale-free networks largely deviates from the expected, narrow degree distribution of random graphs. We have defined the hubness of a node as the difference between its degree and the typical degree distribution in equivalent random graphs. This definition allows to compare the results out of different networks under a common statistical baseline. Additionally, we have shown that the hubness is bounded by finite-size effects depending on the size and on the density of the (sub-)networks.

Comparison with small world effect

One of the most famous discoveries in modern graph theory was the *small-world* effect. Roughly speaking Watts and Strogatz showed that RL networks with high clustering are likely to show shortest path lengths similar to random graphs with randomisation of only a small number of edges.

The neuronal networks discussed in this work also show this small-world effect. The segregated modules show high clustering and the integration hubs are the shortcuts that enable short average path lengths. However, since we are tracking the contribution of the nodes in the mesoscale we do not compare the clustering (a local graph measure) with the average path length (a global measure). Nodes that have shortcuts are likely to show high participation p , since they connect to far-away modules. Nodes with only local neighbours instead show minimal participation $p = 0$ since all neighbours are in the same module. Therefore a wide distribution of participation p is a sign of small-world networks with nodes that range from locally organised to long-range connectors. Fur-

thermore participation is able to detect small-world like effects for networks that have a global transitivity of 0, in case no triads exist, see Fig. 56 for an example of a graph with no triangles. Assume this network is created from two separated modules (red and blue) and afterwards the black edges are introduced as shortcuts. The participation of the linked nodes will be increased, indicating that they connect to the other module. However, a standard small-world detection is not possible since all nodes have clustering $C_i = 0$.

Outlook

Future challenges of the formalism are its application to weighted or signed networks. From a technical point of view, we note the potential of participation vectors to improve community detection methods because they are ideal tools to identify misclassified nodes. This information can be used to switch nodes accordingly until the partition minimising the number of misclassified nodes is reached. Interestingly the framework of multilayer networks is also applicable to time dependent networks. For this the layers are arranged in temporal order, each indicating a discrete time window and edges are present if there is a connection between a pair of nodes at this time. The multilayer participation vector approach is then able to distinguish nodes that are active only at distinct times from those that are active at all times.

Conclusion

The tools presented in this thesis give additional insights on the mesoscale of networks. This reaches from the measurement of the contribution of single nodes' role in the community structure up to the core-periphery structure of spatially embedded networks. Furthermore the application to multilayer networks can give insights in the structural differences of single nodes or whole layers. Those tools are applicable to all kinds of data sets and hopefully are able to decipher structure and function in various kinds of networks.

Appendix

Louvain Method

There is a wide range of community detection algorithms available. Here we used the *Louvain*-method [33] in a Matlab implementation [34]. As most current community detection methods it optimises the modularity quality function

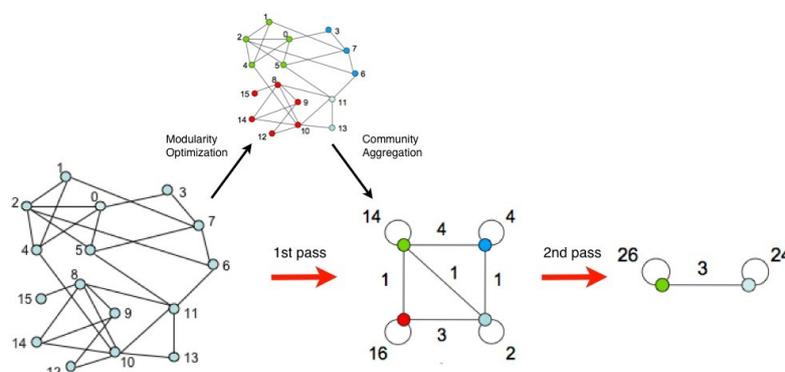
$$Q = \frac{1}{4M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) \delta_{c(i),c(j)}, \quad (36)$$

as introduced before.

The Louvain method is extremely fast, interestingly its limitations are given by memory availability and not computational time, as most other algorithms. For a benchmark network with 118 million nodes finding the partitioning took only 152 min. For this 24 GM of memory were sufficient. This method is also suited to identify hierarchical organisation of modules, however this was not used in this work. It also is able to treat weighted networks.

In Fig. 57 the steps of the algorithm are illustrated and we explain them now in more detail: The algorithm is divided into two phases that are repeated iteratively. Each repetition of those two phases is called a *pass*. Starting with a network of n nodes we firstly assign each node a separate community. Thus in this initial partition there are exactly $M = n$ modules.

In the first phase (modularity optimisation) for each node i we consider the neighbours j of i and we evaluate the gain of modularity ΔQ from removing i from its community and by placing it in



- [33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008
- [34] Inderjit S. Jutla, Lucas G. S. Jeub, and Peter J. Mucha. A generalized louvain method for community detection implemented in matlab. <http://netwiki.amath.unc.edu/GenLouvain>, 2011-2012

Figure 57: Visualisation of the steps of the *Louvain*-algorithm. Each pass is made to two phases: One where modularity is optimised by allowing only local changes of communities; one where the communities found are aggregated in order to build a new network of communities. The passes are repeated iteratively until no increase of modularity is possible. See original description of this algorithm in [34].

the community of j . The node i is then placed in the community for which this gain is maximum, but only if this gain is positive. This process is applied repeatedly until no further improvement can be achieved. Note that nodes are usually considered more than once for the swapping until the local maximum of modularity is reached. The order in which the nodes are considered for changing the module might change the resulting community structure and therefore the algorithm is non-deterministic. Running the algorithm multiple times for our networks resulted in only minor changes in the final partitions.

In the second phase of the algorithm (community aggregation) a new network whose nodes are the communities found during the first phase is created. Weights of the links between the nodes are given by the sum of the weight of the links between nodes in the corresponding two communities. This includes self-loops. Now a pass is completed and the first phase might be reapplied in order to receive a hierarchical clustering. Note that the first pass takes almost all of the time since the network size decreases with each pass.

Since we are only interested in a single partition with maximum modularity we use the partition received after the first pass.

Analytical expressions for participation and dispersion

We want to derive analytical expressions for the participation p and dispersion d for a given participation vector P of a single node (for simplicity we dropped the node index i). As introduced before this is a vector with M elements that are normalised to $\sum_{i=1}^M P_i = 1$. M is the number of modules in the network and the participation vector P will have exactly M_c non-zero elements, where M_c is the number of modules the node is connected to. Thus P will have $M - M_c = K$ zeros. For the dispersion d only the non-zero elements are considered, whereas for the participation p all elements are taken into account, indicated by the reduced participation vector P' . The formulas for participation p and dispersion d then read

$$p = 1 - \frac{\sigma(\mathbf{P})}{\sigma_{max}(M)} = 1 - \frac{M}{\sqrt{M-1}} \sigma(\mathbf{P}) \quad (37)$$

$$d = 1 - \frac{\sigma(\mathbf{P}')}{\sigma_{max}(M_c)} = 1 - \frac{M_c}{\sqrt{M_c-1}} \sigma(\mathbf{P}'). \quad (38)$$

Where σ is the standard deviation in the definition as the second moment of a set of n values $X = \{x_1, x_2, \dots, x_n\}$ about their mean μ

$$\sigma(X) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}. \quad (39)$$

Now let us investigate the behaviour of the standard deviation and come back to the initial dispersion and participation later on. Since the participation vector is normalised the mean of its elements is

given by

$$\mu(\mathbf{P}) = \frac{\sum P_i}{M} = \frac{1}{M} \quad (40)$$

and for the reduced participation vector

$$\mu(\mathbf{P}') = \frac{\sum P'_i}{M_c} = \frac{1}{M_c}. \quad (41)$$

Let us now focus on the standard deviation of the reduced participation vector \mathbf{P}' since the analysis of the full one will be a straightforward extension. Plugging its elements and the mean into eq. 39 leads to

$$\sigma(\mathbf{P}') = \sqrt{\frac{1}{M_c} \sum_{i=1}^{M_c} \left(P'_i - \frac{1}{M_c} \right)^2} \quad (42)$$

$$= \sqrt{\frac{1}{M_c} \left(-\frac{2}{M_c} \underbrace{\sum_{i=1}^{M_c} (P'_i)}_{=1} + \sum_{i=1}^{M_c} (P'^2_i) + \frac{M_c}{M_c^2} \right)} \quad (43)$$

$$= \sqrt{\frac{1}{M_c} \left(-\frac{1}{M_c} + \sum_{i=1}^{M_c} (P'^2_i) \right)} = \sqrt{-\frac{1}{M_c^2} + \frac{1}{M_c} \underbrace{\sum_{i=1}^{M_c} (P'^2_i)}_{=\chi}}. \quad (44)$$

We note that the standard deviation of such a normalised vector is fully defined by the sum over its squared elements χ , where we may sum not only over the elements of the reduced participation vector but over all of the full vector since the additional elements are 0. The standard deviation for the full participation vector is similarly arrived but we have to take into account the changed mean, as well as the zero elements

$$\sigma(\mathbf{P}) = \sqrt{\frac{1}{M} \left(\sum_{i=1}^{M_c} \left(P'_i - \frac{1}{M} \right)^2 + (M - M_c) \left(0 - \frac{1}{M} \right)^2 \right)} \quad (45)$$

$$= \sqrt{\frac{1}{M} \left(\chi - \frac{1}{M} \right)}. \quad (46)$$

Again, the standard deviation will depend only on the sum of the squared elements. Using this information we derive simple formulas for the participation and dispersion

$$p = 1 - \sqrt{\frac{M}{M-1} \left(\chi - \frac{1}{M} \right)} \quad (47)$$

$$d = 1 - \sqrt{\frac{M_c}{M_c-1} \left(\chi - \frac{1}{M_c} \right)}. \quad (48)$$

We now may express $\chi(d)$, substitute it into $p(\chi)$ to receive the final form

$$p(d, M, M_c) = 1 - \sqrt{\frac{M}{M-1} \left(\frac{1}{M_c} ((M_c - 1)(1 - d)^2 + 1) - \frac{1}{M} \right)}. \quad (49)$$

The maximum participation is then reached for $M_c = 2$

$$p^+(d, M) = p(d, M, 2) = 1 - \sqrt{\frac{M}{M-1} \left(\frac{1}{2} ((1-d)^2 + 1) - \frac{1}{M} \right)}. \quad (50)$$

Mean participation for flat modular graphs

We calculate the mean participation for flat modular graphs consisting of M modules of equal size n' . Internal connections occur with a probability p_{in} and external to nodes in other modules with p_{ex} . Similar to a random ER network the number of neighbours each node has in each module are distributed with binomials, where the internal degree is maximal $n' - 1$ and the number of nodes in each other module at most n' . From the distribution of the number of neighbours in each module the distribution of the participation vectors \mathbf{P} follows directly. For simplicity we restrict the discussion to the behaviour of the average node, in a mean field like approach. The mean number of neighbours a node has in its own module is given by $(n' - 1)p_{in}$ and for other modules $n'p_{ex}$. The unnormalised participation vector is then given simply by

$$\langle \mathbf{P} \rangle = \left(p_{in}, \underbrace{p_{ex}, \dots, p_{ex}}_{(M-1) \text{ times}} \right), \quad (51)$$

since we divide the number of neighbours by each modules size. The normalisation factor is then given by $1/\Sigma = p_{in} + (M - 1)p_{ex}$ and according to eq. (48) the mean participation is then exactly

$$\langle p \rangle = 1 - \sqrt{\frac{M}{M-1} \left(\left(\frac{p_{in}}{\Sigma} \right)^2 + (M-1) \left(\frac{p_{ex}}{\Sigma} \right)^2 \right) - \frac{1}{M}}. \quad (52)$$

Now we want to look at the two extreme cases: (1) A network of separated modules $p_{ex} = 0$ and (2) a random network $p_{in} = p_{ex} > 0$. For (1) the participation vector consists of only a single non-zero element p_{in} and due to normalisation $\Sigma = p_{in}$ we end at $\langle p \rangle = 0$ as we expect from nodes that are not able to connect with other modules.

The ER case is more interesting and the participation vector consists of identical elements $\langle \mathbf{P} \rangle = \underbrace{\left(\frac{1}{M} \right)}_{M \text{ times}}$. Thus the sum of the square elements becomes $\chi = M \cdot \frac{1}{M^2} = 1/M$ and the mean participation reaches the maximal value of $\langle p \rangle = 1$ and indicates a vanished modular structure.

Bibliography

- [1] Florian Klimm, Javier Borge-Holthoefer, Niels Wessel, Jürgen Kurths, and Gorka Zamora-López. Individual node's contribution to the mesoscale of complex networks. *New Journal of Physics (submitted)*, 2014.
- [2] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [3] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
- [4] Björn H Junker and Falk Schreiber. *Analysis of biological networks*, volume 2. John Wiley & Sons, 2008.
- [5] Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. Complex networks in climate dynamics. *The European Physical Journal-Special Topics*, 174(1):157–179, 2009.
- [6] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- [7] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.
- [8] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.
- [9] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.
- [10] Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3):032804, 2014.
- [11] C W Harrison, S A Boorman, and R L Breiger. Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81:730–780, 1976.

- [12] S A Boorman and C W Harrison. Social structure from multiple networks. II. Role structures. *American Journal of Sociology*, 81:1384–1446, 1976.
- [13] Francois Lorrain and Harrison C White. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80, 1971.
- [14] Roger Guimera and Luís A. Nunes Amaral. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02):P02001, 2005.
- [15] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.
- [16] Alex Arenas, Javier Borge-Holthoefer, Sergio Gomez, and Gorka Zamora-Lopez. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12(5):053009, 2010.
- [17] Danielle S. Bassett and ED Bullmore. Small-world brain networks. *The Neuroscientist*, 12(6):512–523, 2006.
- [18] Florian Klimm, Danielle S. Bassett, Jean M. Carlson, and Peter J. Mucha. Resolving structural variability in network models and the brain. *PLoS Comput Biol*, 10(3):e1003491, 03 2014.
- [19] O. Sporns and G. M. Tononi. Classes of network connectivity and dynamics. *Complexity*, 7(1):28–38, 2001.
- [20] Gorka Zamora-López, Changsong Zhou, and Jürgen Kurths. Exploring brain function from anatomical connectivity. *Frontiers in neuroscience*, 5, 2011.
- [21] Olaf Sporns. Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology*, 23(2):162–171, 2013.
- [22] G. Zamora-López, C. S. Zhou, and J. Kurths. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.*, 4:1, 2010.
- [23] Olaf Sporns, Christopher J Honey, and Rolf Kötter. Identification and classification of hubs in brain networks. *PloS one*, 2(10):e1049, 2007.
- [24] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013.
- [25] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.

- [26] Geir Agnarsson and Raymond Greenlaw. *Graph Theory: Modeling, Applications, and Algorithms*. Prentice-Hall, Inc., 2006.
- [27] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.
- [28] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [29] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009.
- [30] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.
- [31] Anna Lewis, Nick Jones, Mason Porter, and Charlotte Deane. The function of communities in protein interaction networks at multiple scales. *BMC systems biology*, 4(1):100, 2010.
- [32] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [34] Inderjit S. Jutla, Lucas G. S. Jeub, and Peter J. Mucha. A generalized louvain method for community detection implemented in matlab. <http://netwiki.amath.unc.edu/GenLouvain>, 2011-2012.
- [35] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
- [36] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013.
- [37] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *arXiv preprint arXiv:1309.7233*, 2013.
- [38] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. Measuring large-scale social networks with high resolution. *PloS one*, 9(4):e95978, 2014.

- [39] Dominik Traxl. *C. elegans - Neural structure & dynamics*. Master's thesis, Ludwig-Maximilians-Universität München, 2012.
- [40] Emanuele Cozzo, Mikko Kivelä, Manlio De Domenico, Albert Solé, Alex Arenas, Sergio Gómez, Mason A Porter, and Yamir Moreno. Clustering coefficients in multiplex networks. *arXiv preprint arXiv:1307.6780*, 2013.
- [41] M Puck Rombach, Mason A Porter, James H Fowler, and Peter J Mucha. Core-periphery structure in networks. *SIAM Journal on Applied mathematics*, 74(1):167–190, 2014.
- [42] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [43] K-I Goh, B Kahng, and D Kim. Universal behavior of load distribution in scale-free networks. *Physical Review Letters*, 87(27):278701, 2001.
- [44] Lav R Varshney, Beth L Chen, Eric Paniagua, David H Hall, and Dmitri B Chklovskii. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS computational biology*, 7(2):e1001066, 2011.
- [45] Samuel Ward, Nichol Thomson, John G White, and Sydney Brenner. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *Journal of Comparative Neurology*, 160(3):313–337, 1975.
- [46] William R Schafer. Deciphering the neural and molecular mechanisms of *C. elegans* behavior. *Current Biology*, 15(17):R723–R729, 2005.
- [47] Mario de Bono. Molecular approaches to aggregation behavior and social attachment. *Journal of neurobiology*, 54(1):78–92, 2003.
- [48] Beth L Chen, David H Hall, and Dmitri B Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4723–4728, 2006.
- [49] JW Scannell, GAPC Burns, CC Hilgetag, MA O'Neil, and Malcolm P Young. The connective organization of the corticothalamic system of the cat. *Cerebral Cortex*, 9(3):277–299, 1999.
- [50] Macaque – Wikipedia, the free encyclopedia, 08 2014.
- [51] Jack W Scannell and Malcolm P Young. The connective organization of neural systems in the cat cerebral cortex. *Current Biology*, 3(4):191–200, 1993.
- [52] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biology*, 6(7):e159, 2008.

- [53] Jonathan F Donges, H CH Schultz, Norbert Marwan, Yong Zou, and Juergen Kurths. Investigating the topology of interacting networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 84(4):635–651, 2011.
- [54] AA Tsonis and PJ Roebber. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 333:497–504, 2004.
- [55] Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. The backbone of the climate network. *EPL (Europhysics Letters)*, 87(4):48007, 2009.
- [56] N Malik, N Marwan, and J Kurths. Spatial structures and directionalities in monsoonal precipitation over south asia. *Nonlinear Processes in Geophysics*, 17(5):371–381, 2010.
- [57] A Tantet and HA Dijkstra. An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth System Dynamics*, 5:1–14, 2014.
- [58] Sulochana Gadgil. The Indian monsoon and its variability. *Annual Review of Earth and Planetary Sciences*, 31(1):429–467, 2003.
- [59] Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth’s climate system. *Proceedings of the National Academy of Sciences*, 105(6):1786–1793, 2008.
- [60] Bodo Bookhagen and Douglas W Burbank. Toward a complete himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface (2003–2012)*, 115(F3), 2010.
- [61] Akiyo Yatagai, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh. Aphrodite: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9):1401–1415, 2012.
- [62] Akiyo Yatagai, Osamu Arakawa, Kenji Kamiguchi, Haruko Kawamoto, Masato I Nodzu, and Atsushi Hamada. A 44-year daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Sola*, 5:137–140, 2009.
- [63] George J Huffman, David T Bolvin, Eric J Nelkin, David B Wolff, Robert F Adler, Guojun Gu, Yang Hong, Kenneth P Bowman, and Erich F Stocker. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55, 2007.

- [64] Veronika Stolbova, Paige Martin, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlin. Processes Geophys. (in press)*, 2014.
- [65] R Quian Quiroga, T Kreuz, and P Grassberger. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical review E*, 66(4):041904, 2002.
- [66] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, 2004.
- [67] Roger Guimerà, Stefano Mossa, Adrian Turtchi, and LA Nunes Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005.
- [68] Alessio Cardillo, Jesús Gómez-Gardeñes, Massimiliano Zanin, Miguel Romance, David Papo, Francisco del Pozo, and Stefano Boccaletti. Emergence of network features from multiplexity. *Scientific reports*, 3, 2013.
- [69] Steven L Bressler and JA Kelso. Cortical coordination dynamics and cognition. *Trends in cognitive sciences*, 5(1):26–36, 2001.
- [70] Joaquín M Fuster. *Cortex and mind: Unifying cognition*. Oxford university press, 2003.
- [71] Murray Shanahan. A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, 17:288–303, 2007.

Selbständigkeitserklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit mit dem Titel *Characterisation of individual nodes in the mesoscale of complex networks* selbständig und ausschließlich unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Florian Klimm Berlin, 26.8.2014