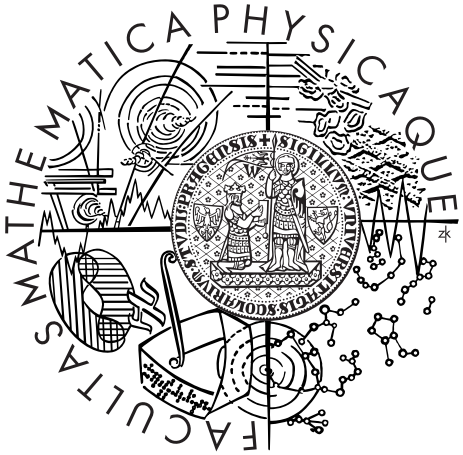
Charles University
Faculty of Mathematics and Physics

HABILITATION THESIS



David Hartman

Properties of networks characterizing dynamical systems

Prague 2022

Acknowledgments

First, I would like to thank Jaroslav Hlinka for the long-term scientific cooperation on complex network topics in various real-world domains. His ideas about nonlinearity in time series, presented in paper [61], initiated our subsequent colaborative papers about nonlinearity in brain networks, see [49] or its reprint P-17 , and stock networks, see [48] or its reprint P-6 , as well as another paper about nonlinearity in climate, see [59] or its reprint P-16 , where he was the first author and I was a second coauthor. Jaroslav Hlinka also initiated and lead our studies about small-world motivated by complex networks in brain and represented by next two results, see [58] or its reprint P-18 and see [56] or its reprint P-8 . Finally, I was a coauthor of two paper mainly authored by Jaroslav Hlinka about causality, see [60] or its reprint P-9 , and about evolving network, see [57] or its reprint P-14 .

I also want to thank Milan Paluš, my mentor at postdoc, for many great ideas and general support and travel opportunities when achieving selected results of Sections 1.2 and 1.3. I also thank him for a particular collaboration on paper dealing with dynamical memory effects, see [91] or its reprint P-19 . I also want to thank Martin Vejmelka for cooperation on the paper built on his idea for a network decomposition method, see [113] or its reprint P-13 . Altogether thank the whole group of my coauthors from the Institute of Computer Science of the Czech Academy of Sciences, including other colleagues from the Cobra group.

I also want to thank a great group of researchers at Potsdam Institute for Climate Impact Research, including Reik Donner, Jakob Runge, Jonathan Donges, Norbert Marwan, and Jürgen Kurths, for excellent cooperation on research topics and warm welcome when visiting their institute. Mainly thanks to Jakob Runge for the opportunity to work with him on a paper about complex causal effect assessment, see [99] or its reprint P-12 .

Let me also thank my coauthors on topics of homogeneity of structures, namely Dragan Mašulovic, Jan Hubička, Jaroslav Nešetřil, and Andres Aranda for great discussions and generally intensive collaboration on the results of Section 1.4. It was a pleasure to spend time together constructing claims and re-proving the theorems. I also thank Manuel Bodirsky for his support and hospitality during my visits to Dresden. Special thanks to Jaroslav Nešetřil for the fantastic support and opportunity to travel to achieve desired collaboration. Also, I appreciated many great days spent with Andres Aranda while drinking coffee and discussing the results of homomorphism-homogeneity in Dresden or Prague.

Finally, thank Milan Hladík for introducing interval analysis to me and for great intensive collaboration on various results of interval optimization and algebra as well as general optimization presented in Section 1.5. Thanks also to other members of the Group of interval methods at Charles University for their support and friendship.

Many thanks to my wife and all three children, who are my inspiration and the reason for my work.

Preface

This thesis consists of selected papers coauthored by David Hartman. The common topic is understanding complex networks and their processing. These networks are models of dynamic complex systems, and their utilization requires working with data, minimizing uncertainty, as well as studying the theoretical properties of graphs and other structures. For that reason, such a topic combines combinatorics, algebraic graph theory, optimization, interval algebra, statistics, data processing, and others.

To use a complex network for analysis of a real-world system, what we need to apply is a pipeline starting with data preprocessing and ending with network analysis itself. We study most of the steps in this pipeline. The presented results are subdivided into groups according to their position in this pipeline. The first group studies reliability of analyses for the systems defined by *behavioral connectivity*. Special attention is given to constructing networks from data accounting for properties of the dataset when using various types of connectivity measures [6, 7, 15, 16, 17]. The next group contains results about *global structures* of these networks such as graph decompositions or small-world character [12, 13, 14, 18, 19]. The third group deals with potential *symmetry of complex network* studying various symmetries applicable to large networks such as homomorphism-homogeneity or regularity of centralities [1, 8, 9, 10]. The last group deals with *uncertainty in the data* discussing the adoption of the interval algebra on matrices as potential representative of complex networks [2, 3, 4, 5, 11].

I also want to thank all my colleagues for their collaboration, see my kind acknowledgments for their contributions above in a separate section.

[1] A. Aranda and D. Hartman. The independence number of HH-homogeneous graphs and a classification of MB-homogeneous graphs. *European Journal of Combinatorics*, 85:103063, 2020. doi:10.1016/j.ejc.2019.103063.

[2] D. Hartman and M. Hladík. Tight bounds on the radius of nonsingularity. In M. Nehmeier, J. W. VonGudenberg, and W. Tucker, editors, *Scientific Computing, Computer Arithmetic, and Validated Numerics. SCAN 2015*, volume 9553 of *Lecture Notes in Computer Science*, pages 109–115. Springer, 2015. doi:10.1007/978-3-319-31769-4_9.

[3] D. Hartman and M. Hladik. Regularity radius: Properties, approximation and a not a priori exponential algorithm. *Electronic Journal Of Linear Algebra*, 33:122–136, 2018. doi:10.13001/1081-3810.3749.

[4] D. Hartman and M. Hladík. Complexity of computing interval matrix powers for special classes of matrices. *Applications of Mathematics*, 65(5):645–663, 2020. doi:10.21136/AM.2020.0379-19.

[5] D. Hartman, M. Hladík, and D. Říha. Computing the spectral decomposition of interval matrices and a study on interval matrix powers. *Applied Mathematics and Computation*, 403:126174, 2021. doi:10.1016/j.amc.2021.126174.

[6] D. Hartman and J. Hlinka. Nonlinearity in stock networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(8):083127, 2018. doi:10.1063/1.5023309.

[7] D. Hartman, J. Hlinka, M. Paluš, D. Mantini, and M. Corbetta. The role of nonlinearity in computing graph-theoretical properties of resting-state functional magnetic resonance imaging brain networks. *Chaos*, 21(1):art.num.013119, 2011. doi:10.1063/1.3553181.

[8] D. Hartman, J. Hubička, and J. Nešetřil. Complexities of relational structures. *Mathematica Slovaca*, 65(2):229–246, 2015. doi:10.1515/ms-2015-0019.

[9] D. Hartman, J. Hubička, and D. Mašulović. Homomorphism-homogeneous L-colored graphs. *European Journal of Combinatorics*, 35(0):313–323, 2014. doi:10.1016/j.ejc.2013.06.038.

[10] D. Hartman, A. Pokorná, and P. Valtr. On the connectivity and the diameter of betweenness-uniform graphs. In *Conference on Algorithms and Discrete Applied Mathematics*, pages 317–330. Springer, 2021. doi:10.1007/978-3-030-67899-9_26.

[11] M. Hladík, D. Hartman, and M. Zamani. Maximization of a PSD quadratic form and factorization. *Optimization Letters*, pages 1–14, 2020. doi:10.1007/s11590-020-01624-w.

[12] J. Hlinka, D. Hartman, N. Jajcay, D. Tomeček, J. Tintěra, and M. Paluš. Small-world bias of correlation networks: From brain to climate. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):035812, 2017. doi:10.1063/1.4977951.

[13] J. Hlinka, D. Hartman, N. Jajcay, M. Vejmelka, R. Donner, N. Marwan, J. Kurths, and M. Paluš. Regional and inter-regional effects in evolving climate networks. *Nonlinear Processes in Geophysics*, 21(2):451–462, 2014. doi:10.5194/npg-21-451-2014.

[14] J. Hlinka, D. Hartman, and M. Paluš. Small-world topology of functional connectivity in randomly connected dynamical systems. *Chaos*, 22(3):art.num.033107, 2012. doi:10.1063/1.4732541.

[15] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotná, and M. Paluš. Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity. *Climate dynamics*, 42(7-8):1873–1886, 2014. doi:10.1007/s00382-013-1780-2.

[16] J. Hlinka, N. Jajcay, D. Hartman, and M. Paluš. Smooth information flow in temperature climate network reflects mass transport. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):035811, 2017. doi:10.1063/1.4978028.

[17] M. Paluš, D. Hartman, J. Hlinka, and M. Vejmelka. Discerning connectivity from dynamics in climate networks. *Nonlinear Process Geophysics*, 18(5):751–763, 2011. doi:10.5194/npg-18-751-2011.

[18] J. Runge, C. Petoukhov, J.F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(1):1–10, 2015. doi:10.1038/ncomms9502.

[19] M. Vejmelka, L. Pokorná, J. Hlinka, D. Hartman, N. Jajcay, and M. Paluš. Non-random correlation structures and dimensionality reduction in multivariate climate data. *Climate Dynamics*, 44(9):2663–2682, 2015. doi:10.1007/s00382-014-2244-z.

Contents

1. Complex networks analysis

1.1 Introduction

Characterization, modeling, or even behavior prediction of complex dynamical systems are in some areas still far from being completely understood, see, e.g., the human brain [17] or the Earth's climate [75]. An important observation in science, see [12], is that many systems are made out of parts, and we can study the particular system via learning how these parts work. Authors of [12] also observe that there is a "new approach to science, studying how relationships between parts give rise to collective behaviors of a system, and how the system interacts and forms relationships with its environment". We can rephrase this observation into a simplified one, saying that many of these systems are composed of smaller subsystems interconnected via particular synchronization patterns [93]. One can imagine a simple model where a family of oscillators are connected via synchronization. We know that even such a simple system can produce quite complicated dynamics [93].

As a consequence of the above statements, we can see that even though we consider the representation power of particular models of subsystems as sufficient, we still might get into troubles when considering a specific interconnection structure. At the same time, it has been recently suggested that one can substantially contribute to the understanding of the systems mentioned above by exploting the so-called network character of the system [88, 2]. The system analysis is carried out by studying this "underlying network," assuming that crucial aspects of the system's behavior can be ascribed to specific connectivity patterns [88]. Such *complex networks* surprisingly often share structural properties across different real-world systems, and their characterization can benefit from using the same set of analytic tools [29].

There are a plethora of approaches when analyzing dynamical systems using complex networks. We can describe a relatively direct utilization within neuroscience. Consider the task of discriminating patients with a particular neurological disorders from healthy controls in various stages of the disease. This task represents a very complex problem that has been approached from multiple perspectives [9, 102, 24]. The tools used for the classification of various diseases range from standard classification techniques such as Support vector machine (SVM) [8] to more comprehensive ones using complex networks, e.g. Graph neural networks (GNN) [121, 118]. Except for the relatively direct application of the connectivity structure using GNN, we can also use specific network descriptors. Such an approach is always associated with a particular task, as we can see in a specific example of the characterization of Alzheimer's disease [53]. This study uses magnetic resonance imaging (MRI) data to construct a complex network representing the brain's functional structure. Alzheimer patients' networks exhibit longer distances indicating less efficient information exchange and a higher tendency to create clusters indicating higher local specializations. We can compute both properties for every subject using network characteristics and combine them into the well-known small-world coefficient [120]. Consequently, we can say that Alzheimer's disease modifies the small-world character of the patient's network.

The examples above show specific approaches. However, there is a typical analytical pipeline that is shared by many studies using complex networks. This pipeline is subject to many forms of uncertainty in various levels of processing, including problems in the accuracy of the data as well as the expressive power of the used graph characteristics. This thesis explores ways to minimize uncertainty in a complex network analytical pipeline. To better understand the motivation of this approach, we start with a short review of some applications of complex networks in section 1.1.1. In section 1.1.2, we identify research gaps and define general goals in this area. In the subsequent four sections (from 1.2 to 1.5), we describe results achieved when resolving the defined goals. The final section 1.6 provides some final remarks and describes several projects solved in this area having more applied outputs.

1.1.1 Real-world motivation

The areas of application of complex networks range from biological, natural, and socio-economic systems to characterizing purely technological ones.

Biological systems: Within biological systems, we can find tasks such as the analysis of the human brain [23, 94, 98] usually predicting the development of neurodegenerative diseases such as Alzheimer's [30, 117] or characterizing epileptic seizures [107, 41]. Other biological tasks are connected with protein-to-protein interactions [65] usually utilized in new drug design [64]. There are other commonly studied networks, such as metabolic networks [66] or genetic networks [33]. An example of a physical system handled via this approach is the Earth's climate [34, 108, 110], usually employed to explain complicated global phenomena such as El Niño/La Niña [109].

Social systems: Social systems represent a vast area of applications giving birth to many notions such as centrality [39], homophily [81] or the small-world property [83, 120]. These systems range from classical social systems to online communities. Social networks can represent, e.g., personal relations such as friendship or trust [119, 19]. These models can also be defined for online communities or online social activities represented, e.g., by Facebook [72] or Email communications [43]. These networks give rise to the area of Social Network Analysis (SNA) [119]. This type of analysis can help in several application domains such as target marketing [68] or recommender design [18]. Similar to a computer virus that can spread through a particular structure of email contacts, see [122], physical contacts are responsible for transmitting diseases. This, in fact, gave birth to a whole area of research called *epidemic spreading* [92] that studies the spread dynamics of various diseases on social networks [85]. There are particular examples of diseases whose analysis can be improved using a social network viewpoint, such as sexually transmitted diseases [73]. Another critical topic is the analysis of rumors and fake-news [124] that has been applied to several social networks such as Twitter [67, 70]. There are other types of social networks based on scientific collaboration, criminal networks, etc., see survey [29].

Social-economic systems: Another type of real-world system that is quite complex and has a network character is a social-economic system [42]. An economy is usually considered as a system of complex interactions between players

of various types. This phenomenon is further inherited into various subsystems of the economy. A nice example is that of financial markets. A financial market exhibits several properties that characterize general complex systems in which large numbers of complex units interact together [79, 80]. Due to the usual lack of data, information about the interconnection structure can be helpful [80]. There are various economic subsystems handled via the complex network approach, including world-trade networks [103], stocks networks [78] or currency market networks [84]. Applications include the characterization of markets, such as the task of designing strategies for market indices [106]. Another example is portfolio design and trading strategies providing a tool to support portfolio optimization, e.g., a study of topological properties of Markowitz portfolio assets [35]. As the last example, we can name systematic risk management, more specifically characterizing risk emerging on markets, e.g., characterization and prediction of financial crises using network properties [16, 116].

Technological systems: Another group of systems suitable for complex network analysis are the technological systems. The internet [4] has a prominent position among these systems. This network is fascinating due to its enormous size and growing character, giving rise to the problem of efficient approximation of the whole structure without direct access to it, e.g., via sampling [74]. One reason to obtain information about the structural properties of the internet is to analyze its resilience to random breakdowns [27]. Resilience to random attacks is related to the degree distribution of the internet graph. However, on the internet, we can also expect directed attacks to highly interconnected systems called hubs. This leads to the research of both types of resiliences [3]. The internet has been found to possess a specific degree distribution [36]. Following these findings, random graph models were suggested to represent this phenomenon, such as the Barabási-Albert model [13] or Linearized Chord Diagram [20]. Apart from the whole network structure, we can also study networks constituted by particular malware such as a botnet. The botnet is a type of malware formed by a botmaster, a commanding program installed on the command-and-conquer server, and an army of bots, malicious microprograms installed on multiple devices over the internet obeying commands from the botmaster. Such a botnet structure can implement a Distributed Denial-of-service (DoS) attack capable of shutting down large and well-protected world servers. Characterizing models of potential botnet structures is helpful in the detection of existing botnets on the internet [31].

Web graph: Another complex network that is often studied is a network representing the interconnections of the World-Wide-Web network abbreviated as the *Web graph* [15, 21]. This network represents directed references between web pages. This graph is already massive, and it is still growing quite rapidly. Its network analysis is crucial to optimize search engines and facilitate access to information; note algorithms for searching the web such as PageRank introduced by Brin and Page [22] or HITS introduced by Kleinberg [69]. Apart from immediate application areas, this network is fascinating by itself. It is vast and sparse in edges, self-organizing, efficient in communication, and resistant to various attacks [21]. This extensive network also motivated new lines of studies in the area of infinite random models, including the area of countable random graphs [21] as well as the area of uncountable graph limits [74].

The overview of potential application areas continues to other fields such as logistic networks, linguistics, or even software architecture. For their review, see the comprehensive study [29].

1.1.2 Common tasks and corresponding gaps

As shown above, complex networks are used to understand some complex features of the underlying dynamical system. From a general perspective, the primary task in this field is to *characterize the system's behavior*, such as

- changes of brain structure and information processing emerging from a potential neurological disorders, or
- critical features of the epidemic spreading process depending on social network structure, or
- the reactions of financial markets to economic crises based on the development of mutual relationships between financial elements, or
- principles of internet threats such as Distributed Denial-of-service attacks using botnet networks.

The second, more specific, task using complex networks is *classifying or predicting the behavior* of the underlying interconnected system. This task can highly benefit from the general characterization task defined above. Examples include

- classification of subjects having Alzheimer's disease, or
- prediction of the rate of epidemic spreading in society, or
- prediction of an incoming economic crisis, or
- identification of a botnet structure or prediction of Distributed Denial-of-Service attacks.

Even though both tasks are connected, the corresponding analytical pipelines can differ significantly. However, there are some typical steps in the analytical pipeline using complex networks. We can depict a simplified pipeline as in Figure 1.1.

The output of this complex network analysis might contribute to a machine learning process or enable a better understanding of the underlying system. However, we need to apply every step carefully, considering the overall goal. A typical example that substantially impacts the accuracy of complex network analysis is omitting some preprocessing steps. Examples include

1. removing head movement artifacts in magnetic resonance brain data,
2. minimizing the bias introduced by periodic changes in the solar input via removing the mean annual cycle in climate data, or
3. using logarithmic return instead of prices in stock market analysis.

This leads to a general goal, or rather a motivation, of the research presented in this thesis.

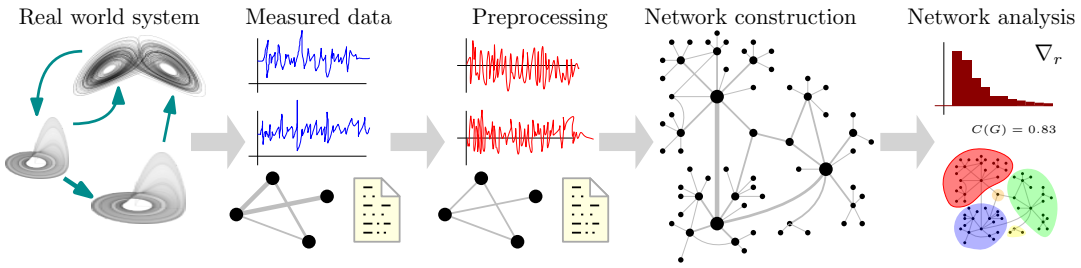


Figure 1.1: General simplified complex network analysis pipeline. The pipeline starts with the measurement of observables of the dynamical system (as usual we assume the system to be composed of many smaller subsystems). The next step is preprocessing. In this step, the data are adjusted with respect to their properties, the nature of the constructed model and the target task. Having the data prepared properly, we can construct the network in the next step. In the final step, we analyze the network and infer the target property of the underlying system.

Main Goal 1.1.1. *Considering a particular dynamical system and a corresponding task, design a complex network analytical pipeline solving the task that minimizes uncertainty.*

As mentioned above, many tasks solved via complex networks share a large part of the analytical pipelines. For that reason, we can approach this goal in a general form improving complex network utilization in general. Complex networks built from various underlying systems share some properties such as degree distribution or tendency to exhibit the small-world character. Therefore, an approach to solving issues of an analytical pipeline for a particular system can be, after some modifications, reused for other systems. Still, the task given above is quite general, and we can further subdivide it into several subtasks. The idea is to explore the effect of model-building steps on the robustness of the final characteristics utilized in the complex network analysis. Considering this, we can rephrase and concretize the above given main goal 1.1.1 as

Main Goal 1.1.2. *Considering a particular dynamical system, minimize the effect of network model creation on the robustness of the chosen graph characteristics.*

For many systems mentioned above, we have their overall behavior encoded in the form of multiple time series (see references in Section 1.1.1), e.g.

- brain networks, or
- Earth's climate networks, or
- financial system networks, or
- malware activity networks.

We create a corresponding network for a system of this type by applying some dependency measures, such as correlation coefficients or various versions of mutual information, on this multivariate time series. This immediately brings

questions about the effect of statistical properties of the data, such as nonlinearity, on the computation of the dependency measure and consequently on the network structure. We can define our first goal solved in section 1.2 as follows.

Goal 1.1.1. *Enable complex network analysis of a particular complex system via proper computation of dependency measures accounting for particular data properties.*

The main object in our analyses is a network. We need to understand the building of the global network structure to represent the phenomenon of our task. From one perspective, such understanding includes the proper space decomposition and consequent determination of the vertices. From another perspective, it also includes computation of characteristics of interest, expressing the studied phenomenon correctly and minimizing the effect of the pipeline steps.

Considering the first perspective, note that complex networks are often defined on various levels of detail. Different details may provide other information about the studied phenomenon. Examples are various decompositions of climatological data of the Earth using different globe grids or even using a set of major dynamical components such as El Niño Southern Oscillation or North Atlantic Oscillation. Other examples can be various Internet decompositions using single computers as nodes or some well-defined subnetworks. Usually, some near-optimal decomposition of the original space offers the best expressive power of the network characteristic. For that reason, the search for proper decomposition is one of the most critical steps in complex network analysis. It stands on the boundary of preprocessing and analysis. One of the critical aspects of the successful application of this step is the interpretability of the results. This leads to the following goal.

Goal 1.1.2. *Determine the best approach for decomposition for a particular dynamical system that would enable to explain observed phenomena.*

An example of the second perspective is the small-world phenomenon. This global characteristic of the network is frequently used in analyses. The question is to what extent the observed small-world character is due to the system's structure or due to the applied steps of the pipeline. The following goal represents this perspective.

Goal 1.1.3. *Determine the sensitivity of the global network characteristics considering the commonly applied network-building steps.*

Both perspectives are discussed in section 1.3.

Motivated by the global network structure studies such as small-world, we consider another property. Note that one important property of complex networks is their potential growing character that might result in large networks. Studies of such enormous networks are complicated. For that reason, some substitute notions as graph limit [74] are often used. One of the prominent examples, in this case, is the Web graph [15]. Among various fascinating properties of the Web graph, researchers were also attracted by its symmetric properties, see [21]. The symmetry of a general complex network was also studied from the perspective of the corresponding automorphism group, see [76, 101]. It appears, however, that such a definition of symmetry might not be the best for real-world networks due

to its strict definition. Motivated by approach of [21] and properties of motif counting, we also study the following goal.

Goal 1.1.4. *Considering a large network, what are existing notions of symmetries, and what are their properties?*

As mentioned above, one of the critical issues of complex networks built from data is uncertainty. As suggested in the sections referenced above, we can handle uncertainty using proper preprocessing or improving the representative model. Still, there can be remaining uncertainty due to irremovable inaccuracies in data. We can represent inaccuracies in different ways. A fruitful way to do this is the utilization of interval arithmetic, see [86], which accounts for all possible measured values.

Many problems in graph theory can be translated into matrix questions. Since interval algebra is mostly developed for matrices, we have also focused on that perspective. A good understanding of handling uncertainty in this area can help significantly in consequent complex network applications, see [6, 5].

Another type of uncertainty can be hidden in an optimization task that is not solvable in a reasonable time, and we need to apply an approximation method. Altogether, we can formulate the following goal.

Goal 1.1.5. *Understand the various forms of uncertainty when representing network tasks as matrix problems with particular attention to the interval algebra approach.*

1.2 Behavioral connectivity

One specific type of complex network is designed from a multivariate image of the dynamic system's behavior, often expressed as time series [123]. Examples of systems analyzed in this way include brain networks [23], Earth's climate networks [34], or econo-financial networks such as stock market networks [78]. The crucial step in corresponding analyses is the design of the network, i.e., the computation of the weight for each edge.

A general way to construct the weight of an edge connecting two subsystems is to calculate the mutual relationship between time developments of the corresponding subsystems. One of the most straightforward cases, still surprisingly popular, is Pearson's correlation coefficient. This choice can face problems in many datasets representing measured states of dynamical systems. We can use other more elaborate methods to evaluate dependence in some applications. This choice, however, can bring other problems with estimation.

Let us show the analysis of such a system on the example of brain complex network analysis. We can redraw Figure 1.1 for complex network analysis of brain activity measurement as in Figure 1.2, see also [23].

The first step is the measurement itself, resulting in a multivariate time series dataset. The next step is preprocessing, which resolves problems such as head movement or known measurement artifacts. This step can also deal with statistical issues of the data. Having well-preprocessed data, we can compute for every pair of subsystems the interdependence of a corresponding pair of time series resulting in the weight of an edge. In this way, we design the whole weighted

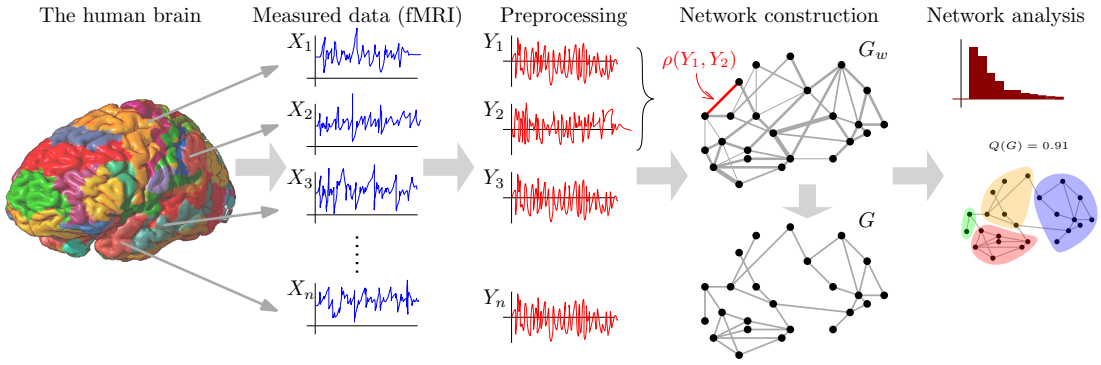


Figure 1.2: The figure shows a simplified depiction of the complex network analysis of the brain using the brain's activity measurement, such as functional magnetic resonance imaging (fMRI).

network G_w. The next step often applies proper thresholding resulting in the unweighted network G. We run the final analyses on this network. Examples might include direct utilization of simple measures such as small-world [53, 105], centralities [32, 1], or community detection [37, 82]. These networks can also be used as an input to further machine learning techniques such as Graph neural networks, see, e.g., a recent review of machine learning methods for the classification of schizophrenia [24].

Utilizing network characteristics in brain studies is highly dependent on the robustness of this pipeline. One of the core issues in this process is the computation of dependence measures for pairs of time series. The most popular dependence measure is (Pearson's) correlation coefficient. For random variables X, Y we can simply define it as

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}, \tag{1.1}$$

where $\mathrm{Cov}(X, Y)$ denotes covariance and σ_X denotes standard deviation of X. Several problems may arise during such calculation, namely the presence of nonlinearity discussed in Section 1.2.1 and the effect of the dynamical memory discussed in section 1.2.2. Resolving these is given as the Goal 1.1.1. We studied several issues of these computations as described in the subsections below.

1.2.1 Nonlinearity in connections

The most common way to determine the weight of an edge is via some dependency measure such as Pearson's correlation coefficient, which has been developed to assess the linear dependence. While Pearson's correlation detects dependence reliably in the case of multivariate Gaussian probability distributions, it may be suboptimal in the case of complex non-Gaussian dependence patterns. We use the term "linear" to denote the "Gaussianity" and the term "nonlinear" to say that the distribution is not Gaussian. These terms are motivated by the distinction between respective linear and nonlinear methods. Apart from linear correlation coefficients, we can use other measures sensitive to nonlinearities, such as mutual information. Mutual information is defined using a statistical distance measure between two probability distributions. For two discrete probability distributions

P and Q defined on the same probability space $\mathcal{X}$. We define Kullback–Leibler divergence as

$$D_{KL}(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \tag{1.2}$$

We define the divergence using the integral instead of the sum for continuous probability distributions. The mutual information is then defined as Kullback-Leiber divergence of the bivariate distribution from a hypothetical bivariate distribution of two independent variables with the same marginal distributions as the original ones, i.e.

$$I(X,Y) = D_{KL}(P(X,Y)\|P(X)P(Y)), \tag{1.3}$$

where $P(X,Y)$ is joint probability distribution and $P(X)$ is marginal distribution. This measure can account for nonlinearity, but it has problems with related estimators, such as dependence on space discretization, computational bias, or computational demands. For these reasons, their utilization can bring more problems than solutions.

We suggest applying a careful process accounting for potential nonlinearity in data when performing complex network analysis. This effort includes assessing the effect of nonlinearity and possible steps to mitigate it. To determine the effect of nonlinearity, we suggest using multivariate Fourier transform (FT) surrogate datasets, i.e., realizations of multivariate linear Gaussian stochastic process that mimics individual spectra of the original time series and their cross-spectrum, see e.g. [95, 90, 89]. The multivariate FT surrogates are obtained by computing the Fourier transform of the series, keeping unchanged the magnitudes of the Fourier coefficients (the spectrum), but adding the same random number to the phases of coefficients of the same frequency bin; the inverse FT into the time domain is then performed. The multivariate FT surrogates (partially) preserve synchronization, if present in the original data, which can be explained by a multivariate linear Gaussian stochastic process. This step creates a linear counterpart for the original dataset and enables comparison with the original complex network results.

Nonlinearity in brain complex networks: For brain magnetic resonance datasets, we suggest a testing pipeline for nonlinearity based on a particular utilization of Fourier surrogate datasets, see [49] or its reprint P-17. To enable the testing, we make use of the lower bound for mutual information given as $I(X,Y) \geq -\frac{1}{2}\log(1-r(X,Y)^2)$ if some conditions are met such as univariate Gaussianity. The tests are performed for network measures of various types inspired by their application in neuroscience. We consider local measures such as betweenness centrality or clustering coefficient and global measures such as efficiency or assortative coefficient measures. To perform the test themselves, we introduce several dominance indicators and functions.

We considered resting-state fMRI (no particular subject activity) data from healthy volunteers measured by the cooperating institute from Italy with standard neuroimaging preprocessing; see [49] for details. We used a box-counting algorithm based on the marginal equiquantization method to compute mutual information. Considering global network characteristics computed for the whole network, we found only a negligible effect of nonlinearity. We have found statistically significant deviation for local parameters computed for vertices of the

network. Nevertheless, a quantitative comparison showed that the nonlinearity effect is practically negligible when compared to the intersession variability.

Nonlinearity in climate complex networks: The major contribution of the study mentioned above is delivering a particular pipeline to assess the nonlinearity. The advantage of brain studies is the possession of multiple subjects and potential measurement sessions. We have also explored a different type of dataset describing the evolution of the Earth's climate, see [59] or its reprint P-16. In the case of climate systems, the data availability is significantly lower compared to brain studies. We develop a multi-step approach that allows the detailed assessment of the nonlinear contribution to dependence patterns in a dataset, including quantification, localization, and analysis of sources of this contribution. This pipeline enables researchers to identify parts of the data responsible for nonlinear contribution and assess the effect on the network construction. Apart from nonlinearity given by rescaling of variables that can be controlled by Spearman's rank correlation or marginal normalization, we also elaborate working with "extra-normal" information. It is well known that for a bivariate Gaussian distribution using variables X and Y, the mutual information is directly defined by correlation $\rho(X, Y)$ as $I_G(X, Y) = -\frac{1}{2}\log\left(1 - \rho^2(X, Y)\right)$. For a non-Gaussian distribution this is not generally true and thus, based on [62], we define "extra-normal" information as $I_E(X, Y) = I(X, Y) - I_G(X, Y)$. Since some of this nonlinearity remains in the dataset, we suggest a localization method based on the corresponding complex network.

To have robust results, we consider two major datasets in climatology, the NCEP/NCAR reanalysis dataset as well as the ERA dataset. The main variable was surface air temperature. Similar to the brain nonlinearity study shown above, we have used box-counting marginal equiquantization method. Due to some properties of climate data, we have also used an estimator based on k-nearest neighbor obtaining similar results.

Nonlinearity in financial complex networks: Based on experience with network constructions for the brain and the climate datasets, we studied financial data that are well known to possess nonlinear character, see [48] or its reprint P-6. We have investigated various complex networks created for the different stock markets expressing synchronization of the strongest stocks in the chosen region, namely stock markets given by indices New York Stock Exchange 100 (NYSE100), Financial Times Stock Exchange 100 (FTSE100), and Standard & Poor 500 (SP500). The selection of the stock markets covers more countries and dimensions. For the construction of the networks, we consider the logarithmic return of the daily closing prices.

In this case, the source of nonlinearity hides quite often in the time development of a particular stock. We propose a pipeline to identify and remove several effects of nonlinearity using various commonly used thresholding techniques for financial market networks. There can be remaining components for which we suggest a method of localization. These remaining nonlinearities have been found to have a negligible impact on the complex network level, especially if we account for the behavior around the crisis of 2008.

1.2.2 Dynamics of the time series

Even if we resolve the problem of nonlinearity in the dataset, we can still end up with different issues causing uncertainty in the resulting connectivity. The time series dynamics can influence resulting dependence measures, including nonlinear mutual information. Note that for stochastic processes $X_i(t)$, the potential uncertainty is measured by entropy rate $h = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$. In particular, processes with lower entropy rates (i.e., processes with higher regularity or dynamical memory) tend to have their non-negative dependence measures biased upward.

In our study, see [91] or its reprint P-19, we have demonstrated such a bias for simple process defined by an autoregressive process. Apart from observing this phenomenon, we suggest a method to correct this bias. The bias correction is done by replacing the absolute correlations with their Z-scores based on independent Fourier transform surrogate data used above for nonlinearity handling. These surrogate datasets are realizations of processes preserving the original spectra of the studied process and their entropy rates in the Gaussian approximation. As a result of this operation, the cross-correlations of the independent FT surrogates have a approxamately normal distribution, and the Z-score is a statistical quantity suitable for the thresholding used to construct binary networks. On the network level, we have considered area weighted connectivity and betweenness centrality (BC) to test our approach.

We have observed the phenomenon described above for the surface air temperature data obtained from NCEP/NCAR reanalysis data for tropical areas. After correcting for this bias, the most connected areas of the climate networks move from the tropical areas influenced by the El Niño Southern Oscillation to the Northern Hemisphere areas dominated by the North Atlantic Oscillation (NAO). Using this corrected connectivity measure we are able to observe the influence of NAO and solar variability on the connectivity in the climate networks.

1.2.3 Causal network discovery

For some complex systems, we can also be interested in the directionality of the dependence. Undirected measures are interesting to characterize phenomena such as teleconnections, see [115]. Directed dependency measures can be helpful to reconstruct information flow being in close relation to the air-mass flow, see [60] or its reprint P-9. The directional dependence is delivered by causality computed using the Granger causality analysis. The space of the globe has been subdivided into an equidistant geodesic grid to suppress the effect of unequal distances between grid points of pairs located in different latitudes. We need to apply appropriate thresholding to obtain the final unweighted graph suitable for interpretations. We found common approaches, such as using (extensions of) minimal spanning trees or a particular threshold, not ideal in this case. For that reason, we constructed a new winner-takes-all approach considering the situation in each node of the network. For the case of the air temperature field, our methodology clearly uncovers a smooth flow structure; evident both qualitatively and in quantitative comparison with appropriate random graphs. The climatological relevance is shown by the close relation to the air-mass flow.

1.3 Global network structure

To use a network for analytical purposes, we need to believe that its global structure represents the studied phenomenon reliably. We have already commented on the problem of edge determination in Section 1.2. As described in the introduction, there are other problems with the global structure of the network.

1.3.1 Network decompositions

One of the critical problems when building complex networks is decomposing the space into subsystems, i.e., vertices of the constructed network, representing the studied phenomenon well. Solving this task relates to the Goal 1.1.2.

Decomposition of complex networks: For some dynamical systems, determining subsystems represented by nodes can be complicated. In the case of climate networks, a simple globe grid can be insufficient for some applications. The reason is that instead of the situation at particular positions on the globe, we are interested in global events such as El Niño or North Atlantic Oscillation. These localized phenomena are well-known modes of variability and play a crucial role in Earth's climate. Principal component analysis (PCA) has a long tradition in the climatological community. A common approach is performing a regional analysis or focusing on a smaller selection of the highest-variance components. The goal of our method, see [113] or its reprint P-13, is to handle a meaningful portion of the variability in the entire globe and thus automatically reduce the dimensionality of large climate datasets into spatially localized components suitable to represent nodes of a complex network that allows interpretable utilizations.

The method mentioned above identifies non-random modes of variability, where non-random means with a high probability, i.e., such a mode could not have been generated by a stochastic model in which the processes at each grid point are independent. This method statistically separates the modes that cannot be explained by random fluctuations of independent processes (the non-random modes) from those that can (and are therefore discarded before further analysis). A fully automatic method proposed here is essential if large bodies of data are being analyzed as they are likely to result in many modes of variability, among which it is not easy to choose. In this way, we can reduce the dimensionality of the data prior to applying complex network approaches and support a more straightforward interpretation of the results. Some of the applications, such as using FFT surrogates described in Section 1.2.2, would not be computationally possible for the whole network, and dimensionality reduction is necessary.

The PCA method identifies a predefined number of principal components serving as a new basis for data vectors. The problem is selecting the correct number of components. We proposed to identify non-random components by comparing the eigenvalues of the sample covariance matrix of the original data with their distributions arising from a surrogate stochastic model. Due to some climate data properties, including autocorrelation, we cannot directly apply results from Random Matrix Theory. We showed that the identified non-random modes characterize most global dynamics in two variables of interest, the monthly surface level pressure (SLP) and the monthly surface air temperature (SAT) using NCEP/NCAR reanalysis data.

Decomposition in evolving networks: Time series covering long time intervals are convenient to obtain robust estimates of a dependency measure. For non-stationary processes, structural information can be, and for long time intervals usually is, time-dependent. For that reason, evolving networks are often applied, e.g., they can be used to disentangle different types of El Niño episodes. The evolving network is a dynamic network whose states are determined using sliding window approaches on time series. The question is how to apply dimensionality reduction in such a setting. Assume we apply the PCA-based method to identify climate system components as described above. Key questions studied in [57] or its reprint P-14 are as follows: Firstly, how and to what extent the temporal evolution of the grid-based climate network is already reflected in the dynamics of the higher level inter-component network and vice versa; and secondly, what is the role of local (within-component) and distant (between-component) links in the global network evolution.

The applied dimensionality reduction (based on PCA) provides clusters of the grid-based network. For each grid vertex, we determine the major cluster that this vertex belongs to. We compare differences in connectivity for local grid-based discretization and more crisp PCA cluster-based decompositions. Comparisons are made for several coarsening equidistant grids. We have also commented on the effect of several climatological phenomena. Apart from the important role of the ENSO region that was suggested earlier, our detailed analysis provided evidence for the additional role of other tropical regions. For these reasons, we concluded that the component network reflects other climate phenomena compared to coarsened grid-based network and might be valuable for climatological analyses.

Causality reconstruction with dimensionality reduction: Putting together experiences with dependency measures and decompositions of the datasets, we have proposed a methodology to identify regions important for spreading and mediating perturbations in dynamical systems, see [99] or its reprint P-12. This complex network methodology combines dimensionality reduction, causality reconstruction, and causal effect assessment. We have applied this approach to a dataset of atmospheric dynamics, which results in identification of several strongly uplifting regions acting as major gateways of perturbations spreading in the atmosphere. Additionally, the method provides a stricter statistical approach to pathways of atmospheric teleconnections, yielding insights into the Pacific-Indian Ocean interaction relevant to monsoonal dynamics. This methodology is not limited to climate. It has been suggested to be used in a broader context of neuroscience or power grid analyses. Note that, in general, it can be helpful in the potential increase of resilience of a dynamical system to shocks or extreme events. Examples of perturbations studied by complex networks are volcanic eruptions and geoengineering in climate systems, epileptic seizures in the brain, epidemic spreading, blackouts in power grids, or failure of banks in the financial system.

To identify processes responsible for spreading and mediating perturbations in spatio-temporal systems, we need to reconstruct both variables of interest from the potentially invalid grid dataset and their causal interaction distinguishing direct from indirect interactions. The final pipeline consists of the above-studied dimensionality reduction based on PCA, iterative causality discovery avoiding spurious links, and the construction of suitable causal network measures to ana-

lyze the underlying system. We use a causal discovery algorithm based on the PC algorithm [104]. This approach removes spurious links due to common drivers (influencing subsystems) and transitivity effects which has been shown to have an effect also in undirected networks when studying the small-world property, see Section 1.3.2. We also provided arguments supporting the power of the suggested method for handling perturbations compared to the classical correlation approach.

1.3.2 Small-world phenomenon

Assume we have already determined the network structure, and we have the network ready for analysis. We can decide to study some of its properties and use them to characterize the underlying system. An example of such approach can be utilizing the small-world character; see an example at the beginning of Section 1.1. Roughly speaking, a graph is small-world if it tends to create clusters and its average shortest path is not long. To quantify this, we can make use of the following characteristics. For a graph $G = (V, E)$ of size $|V| = n$ define *average shortest path* as

$$L(G) = \frac{1}{n(n-1)} \sum_{u \neq v, u, v \in V} d(u, v), \tag{1.4}$$

where $d(u, v)$ is the distance between vertices u and v. Let us denote the degree of a vertex as $k_v = \deg(v)$. For any vertex $v \in V$ define *clustering coefficient* of v as $C(v) = \frac{|N_G(v)|}{k_v(k_v-1)}$, where $N_G(v)$ is the neighborhood of the vertex v. We define the clustering coefficient of the graph as

$$C(G) = \frac{1}{n} \sum_{v \in V} C(v). \tag{1.5}$$

Note that this represents the density of triangles. We want to compare quantities of these characteristics to a specific null model. We can use standard Erdős-Renyi random graph $G_{n,p}$ defined on vertex set $\{1, \ldots, n\}$ with each edge included independently with probability p. Let us denote $L_{n,p}$ and $C_{n,p}$ the expected values of average shortest path and clustering coefficient of the random graph $G(n, p)$. For a given graph G of size n and density given by p introduce the following indices $\lambda(G) = \frac{L(G)}{L_{n,p}}$ and $\gamma(G) = \frac{C(G)}{C_{n,p}}$. The small-world condition can be formulated as follows

$$\lambda(G) \gtrsim 1 \qquad \text{and} \qquad \gamma(G) \gg 1, \tag{1.6}$$

or we can formulate the whole property as $\sigma(G) = \frac{\lambda(G)}{\gamma(G)} \gg 1$.

Bias of the small-world for functional networks: The small-world characteristic, defined above, is a popular graph descriptor in neurological studies used to characterize several neurological diseases. Networks in brain studies are built using data from various measurements, including magnetic resonance (MRI) or electroencephalography (EEG). Apart from *structural connectivity* expressing connections of brain regions via white matter, we also measure so-called *functional connectivity* describing synchronization of brain regions during the predefined task. We use the name functional MRI (fMRI) when using MRI for this type of measurement. Since the results of these measurements are time series, the resulting network is computed using the methods discussed above. We argue in [58]

or its reprint P-18 that using dependence measures having partial transitivity, such as Pearson's correlation coefficient, leads to upwardly biased estimates of small-world characteristics.

A simple model for the relation between structural and functional connectivity is the autoregressive (AR) process defined as

$$X_t = c + AX_{t-1} + e_t, \tag{1.7}$$

where c is a vector of constants, A is the association matrix, and e_t is the vector of error terms. The matrix A corresponds to structural connectivity. We set up this matrix in a way corresponding to random structure; more precisely, we use Erdős-Renyi adjacency with several other parameters controlling properties of the AR process. The values of small-world indices indicate increased clustering and approximately conserved average path length with respect to a corresponding random graph. This signifies a small-world like structure of the functional connectivity matrix, even though the coupling structure of the generating system is completely random. This remains true even if we use some degree preserving random generations such as the Maslov-Sneppen procedure.

Evaluation of the small-world bias in real-world networks: The study mentioned above provides arguments for the existence of potential bias of the small-world characteristics when working with functional networks of the dynamical systems. However, the strength of this bias depends heavily on a range of parameters. The relevance of this bias for real-world data is further studied in a subsequent study for brain and climate datasets, see [56] or its reprint P-8. The small-world properties of fMRI functional connectivity graph obtained using standard methods have been shown to be largely reproduced or even exceeded by a matching randomly connected multivariate autoregressive process. This result shows for the first time that the small-world properties of functional connectivity real-world graphs can be attributed to the correlation coefficients' transitive properties. For climate data, we have observed even lower small-world index than for a matching randomly connected multivariate autoregressive process, suggesting a "large-world property" of global climate.

1.4 Symmetry of networks

As mentioned in the introduction, there is growing attention given to large network symmetries. This is not limited to usual symmetries, such as the size of the automorphism group, but we also consider various notions of homogeneity generalizing vertex transitivity. Apart from such a perspective, we can also study symmetry connected with network characteristics, namely regularities in their distribution motivated by similar degree distribution studies.

1.4.1 Notions of homogeneity

A countable graph $G = (V, E)$ is ultrahomogeneous if for any finite $A, B \in V$ and for any isomorphism $f : A \to B$ we can find an automorphism $g \in \mathrm{Aut}(G)$ such that $g|_A = f$, i.e., any local isomorphism is extendable to an automorphism. The classification of all countable undirected graphs is a famous result in this

field [71]. This classification provides interesting results such as ultrahomogeneity of the Rado graph, a countably infinite random graph, or utilization of the Fraïssé limit theory [38].

Relational complexity of graphs: Consider a finite graph G that is not ultrahomogeneous. Such a graph has some local isomorphisms that are not extendable by an automorphisms, think of a cycle C_6 with vertex set $\{1, 2, \ldots, 6\}$ and the local isomorphism $\{1, 3\} \to \{1, 4\}$. We can, however, introduce new colored edges for pairs of distant vertices with colors corresponding to their distance. If we add all such possible edges, the resulting graph is ultrahomogeneous, see Figure 1.3. In this case, edges are sufficient, but for the Petersen graph, we need relations of arity 3.

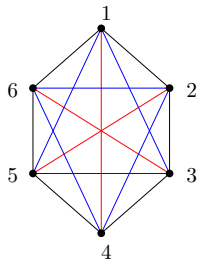


Figure 1.3: Homogenization of C6: Introducing new edges colored by distance create an edge colored graph that is ultrahomogeneous.

We can introduce a notion of homogenization to a relational structure as generalizations of hypergraphs. A *relational structure* $\mathbf{A}$ is a pair $(A, (R^i_{\mathbf{A}} : i \in I))$, where $R^i_{\mathbf{A}} \subseteq A^{\delta_i}$ is δ_i-ary relation on A. The family $(\delta^i : i \in I)$ is called a *type* Δ. The class of all (countable) relational structures of type Δ is denoted as $\mathrm{Rel}(\Delta)$. Classical graphs (V, E) can be understood as relational structure with one binary symmetric irreflexive relation with $\Delta_G = (2)$. Let $\Delta' = (\delta^i : i \in I')$ be a subtype of Δ. Then any structure $\mathbf{X} \in \mathrm{Rel}(\Delta')$ may be viewed as a structure $\mathbf{A}$ with additional relations $X^i_{\mathbf{X}}, i \in I' \setminus I$. We call such $\mathbf{X}$ a *lift*.

Let $\mathbf{A}$ be a relational structure and let $\mathrm{Aut}(\mathbf{A})$ be the automorphism group of $\mathbf{A}$. A k-ary relation $\rho \subseteq A^k$ is an *invariant* of $\mathrm{Aut}(\mathbf{A})$ if $(\alpha(x_1), \ldots, \alpha(x_k)) \in \rho$ for all $\alpha \in \mathrm{Aut}(\mathbf{A})$ and all $(x_1, \ldots, x_k) \in \rho$. Let $\mathrm{Inv}_k(\mathbf{A})$ denote the set of all k-ary invariants of $\mathrm{Aut}(\mathbf{A})$ and let $\mathrm{Inv}(\mathbf{A}) = \bigcup_{k \geq 1} \mathrm{Inv}_k(\mathbf{A})$, $\mathrm{Inv}_{\leq k}(\mathbf{A}) = \bigcup_{1 \leq k' \leq k} \mathrm{Inv}_{k'}(\mathbf{A})$. It easily follows that a lift $(A, (R^i_{\mathbf{A}} : i \in I), \mathrm{Inv}(\mathbf{A}))$ (possibly of infinite type) is an ultrahomogeneous structure for every structure $\mathbf{A} = (A, (R^i_{\mathbf{A}} : i \in I))$. For a structure $\mathbf{A}$ the *relational complexity* $\mathrm{rc}(\mathbf{A})$ of $\mathbf{A}$ is the least k such that $(A, (R^i_{\mathbf{A}} : i \in I), \mathrm{Inv}_{\leq k}(\mathbf{A}))$ is ultrahomogeneous, if such a k exist. If no such k exists, we say that the relational complexity of $\mathbf{A}$ is not finite and write $\mathrm{rc}(\mathbf{A}) = \infty$. To define the *lift complexity*, $\mathrm{lc}(\mathbf{A})$, of a relational structure just omit the condition about invariancy of the added reations.

We study relational complexity for graphs in [50], see its reprint P-11. The paper's major contribution is represented by steps towards the classification of finite graphs having fixed relational complexity, namely 1 and 2, and providing a bound for relational complexity for the specific countable class of graphs. Relational complexity 1 decomposes the graph into sets colored by unary relations of the corresponding lift. We call them *parts*. In the following, c-coloured k-graph is a relational structure on V partitioned into $V = V_1 \cup V_2 \cup \ldots \cup V_k$ such that

structures induced on V_i are graphs and there are c coloured edges existing between parts [26]. Note that colored edge and non-edge play the same role for isomorphism. We provided the following classification result.

Proposition 1.4.1 (Proposition 3.5 in [50]). *Let* $\mathbf{G}$ *be a graph with* $\mathrm{rc}(\mathbf{G}) = 1$ *or* $\mathrm{lc}(\mathbf{G}) = 1$ *and let* $V_1, V_2, \ldots, V_k$ *be its parts. Then the following holds.*

1. *The subgraph induced by each part is an ultrahomogeneous graph.*
2. $\mathbf{G}$ *corresponds to an ultrahomogeneous* 2*-colored* k*-graph with partitions* $V_1, V_2, \ldots, V_k$*.*
3. *The subgraph induced by each pair of parts corresponds to an ultrahomogeneous* 2*-colored* 2*-graph.*

This result relies on the classification of c-colored n-graphs that are only partially classified, see [97]. Therefore the complete classification is still open.

Consider finite graphs with relational complexity 2. Apart from some obvious cases such as metrically homogeneous graphs or cographs, we also provided classification results for trees as follows.

Proposition 1.4.2 (Proposition 3.9 in [50]). *Finite (graph) trees have relational complexity at most* 2*.*

For the countably infinite case, we need to adopt approaches from Fraïssé theory. Let $\mathrm{Age}(\mathbf{A})$ be the class of all finite structures isomorphic to an (induced) substructure of $\mathbf{A}$. Call this class the *age* of $\mathbf{A}$. Studying such classes can help you to find ultrahomogeneous countable graphs via Fraïssé's theorem [38, 63]. Apart from some other class properties, the core property for ultrahomogeneity is an amalgamation property. Class $\mathcal{K}$ has *amalgamation property* if for $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{K}$ and α, β embeddings $\alpha : \mathbf{C} \to \mathbf{A}$ and $\beta : \mathbf{C} \to \mathbf{B}$, there exists $\mathbf{D} \in \mathcal{K}$ and embeddings γ, δ embeddings $\gamma : \mathbf{A} \to \mathbf{D}$ and $\delta : \mathbf{B} \to \mathbf{D}$ such that $\gamma \circ \alpha = \delta \circ \beta$. A structure $\mathbf{U}$ is *universal* for $\mathcal{K}$ if all $\mathbf{A} \in \mathcal{K}$ embedds to $\mathbf{U}$. Roughly speaking, Fraïssé theorem says that for a class $\mathcal{K}$ having amalgamation property there exists a unique universal ultrahomogeneous structure $\mathbf{U}$. This results can be used to seach for ultrahomogenous relational structures.

We are interested in countably infinite graphs that are not ultrahomogeneous but can be extended to an ultrahomogeneous lift (note that it might not be possible – if infinite arities are needed). We consider a class $\mathrm{Forb}_{\mathrm{h}}(\mathcal{F})$ where $\mathcal{F}$ is a family of connected structures. $\mathrm{Forb}_{\mathrm{h}}(\mathcal{F})$ denotes the class of all structures $\mathbf{A}$ for which there is no homomorphism $\mathbf{F} \to \mathbf{A}$, $\mathbf{F} \in \mathcal{F}$. The main result for infinite structures shows that we can find (bounds of) relational complexity via simple exploration of the family of forbidden graphs.

Theorem 1.4.1 (Theorem 5.2 in [50]). *Let* $\mathcal{F}$ *be a finite minimal family of finite connected relational structures and* $\mathbf{U}$ *an* ω*-categorical universal structure for* $\mathrm{Forb}_{\mathrm{h}}(\mathcal{F})$*. Denote by* n *the size of the largest minimal g-separating g-cut in* $\mathcal{F}$*. Then*

(a) $\mathrm{rc}(\mathbf{U}) \geq n$*;*

(b) if $\mathbf{U}$ *is the canonical universal structure for* $\mathrm{Forb}_{\mathrm{h}}(\mathcal{F})$*, then* $\mathrm{rc}(\mathbf{U}) = n$*.*

The g-separating g-cut is a specific edge cut of the structure. For other notions, such as ω-categoricity and canonical universal structure, see [50].

Homomorphism-homogeneity of colored graphs: The notion of ultrahomogeneity was generalized in multiple ways, see [77]. One of the generalizations was due to Cameron and Nešetřil [25]. They use homomorphisms instead of isomorphisms in the definition of homogeneity. Let X and Y are two characters such that $X \in \{H, M, I\}$ that stands for prefixes *homo, mono*, and *iso* and $Y \in \{H, A, B, M\}$ that stands for prefixes *homo, auto, bi*, and *mono*. We can call relational structure XY-homogeneous if any local X-morphism can be extended to Y-morphism over the whole graph. In this way, IA-homogeneity stands for the original ultrahomogeneity. Other interesting classes are defined by HH-homogeneity and MH-homogeneity. These extend either a local homomorphism or a local monomorphism to an endomorphism. Since the original paper of Cameron and Nešetřil there exists a question about potential equality of classes HH and MH. For graphs, this equality has been proven by Rusinov and Schweitzer [100].

In our work about homomorphism-homogeneity, see [51] or its reprint P-15, we have made a step towards generalization of such results. We investigated the equality mentioned above for graphs with sets of colors assigned to vertices and edges and homomorphisms preserving these colors. We can see on the example from Figure 1.4 that for a general graph of this type, the equality does not hold.

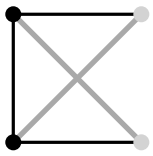


Figure 1.4: An example of a colored graph that is MH-homogeneous but not HH-homogeneous.

For that reason, we have introduced a notion of L-colored graphs, where L is a partially ordered set with the ordering relation $\prec$, with the least element 0 and the greatest element 1 (the idea is that L contains sets of colors ordered by containment). For some partially ordered sets L, we can prove the equality. For instance, in the case of the linearly ordered set, that we call a *chain*, we have.

Theorem 1.4.3 (Theorem 3.3 in [51])**.** *Let G be a finite L-colored graph where L is a chain with the least element 0 and the greatest element 1. Then the following are equivalent:*

(1) G is HH-homogeneous,

(2) G is MH-homogeneous,

(3) G has the following structure:

- *every connected component of G is a uniform L-colored graph, and*
- *if $U(n_1, \alpha_1, \beta_1)$ and $U(n_2, \alpha_2, \beta_2)$ are connected components of G such that $\alpha_1 \preceq \alpha_2$, then $n_1 \leq n_2$ and $\beta_1 \preceq \beta_2$. Consequently, if $\alpha_1 = \alpha_2$, then $n_1 = n_2$ and $\beta_1 = \beta_2$.*

For definitions of L-colored graphs $U(\cdot,\cdot,\cdot)$, see [51]. We have obtained similar results for L, a set of incomparable elements enriched with minimal and maximal elements (we call it a *diamond*). This result can be found as Theorem 4.4. in [51].

Independence number of HH-homogeneous graphs: One of the problems that are still open is the classification of countable homomorphism-homogeneous (HH-homogeneous) graphs. According to the original paper of Cameron and Nešetřil [25] there are uncountably many countably infinite HH-homogeneous graphs – namely, graphs having the Rado graph as a spanning subgraph. Rusinov and Schweitzer gave the first example of a graph with a different structure [100].

Let $\alpha(G)$ denote an *independence number*, i.e., the largest set of vertices that span no edges in a graph G. Define the *star number* of G as $\sigma(G) := \sup\{\alpha(N(v)) : v \in V(G)\}$, where $N(v)$ denote neighborhood of a vertex v. To complete the classification, we need to search for such examples that are not images of the Rado graph. These graphs are characterized by having bounded star numbers. Exploring possible values of star numbers can help the classification. We elaborate, see [7] or its reprint P-5, a structural argument to show a bound for an independence number of HH homogeneous graph.

Theorem 1.4.2 (Theorem 20 in [7]). *If G is a countably infinite connected HH-homogeneous graph with finite star number $\sigma(G)$, then $\alpha(G) \leq 2\sigma(G)+\lceil\frac{\sigma(G)}{2}\rceil-1$.*

This theorem shows that any connected HH-homogeneous graph with infinite independence number always has the Rado graph as a spanning subgraph. Using these results, we can narrow the space for potential future classification via limiting to particular combinations of $\alpha(G)$ and $\sigma(G)$.

Using the property for connected infinite HH-homogeneous graphs with infinite independence mentioned above, we also finished the classification of MB-homogeneous graphs nicely elaborated by Coleman-Evans-Gray [28]. We use the term *bimorphism* for bijective homomorphism. Any classification has to be specified up to a chosen equivalence. Graphs G and H are bimorphism-equivalent if there exist bijective homomorphisms $F : G \to H$ and $J : H \to G$. Note that for graphs, bimorphism-equivalence means that G is (isomorphic to) a spanning subgraph of H and H is (isomorphic to) a spanning subgraph of G. We have provided final classification.

Theorem 1.4.4 (Theorem 26 in [7]). *Let G be a countably infinite MB-homogeneous graph. Then G is bimorphism-equivalent to one of the following or its complement:*

1. *K_ω,*
2. *$I_\omega[K_\omega]$,*
3. *The Rado graph $\mathcal{R}$.*

For two graphs G and H, we use $G[H]$ to denote the lexicographic product of G and H.

1.4.2 Uniform distributions of graph characteristics

Apart from symmetry defined by extending local morphisms, we have also investigated a symmetry hidden in the regularity of a particular graph characteristic distribution – note a simple example of (degree) regular graphs. An interesting class of graph descriptors is the class of centralities, i.e., characteristics that measure the importance of a vertex in a graph. One of these characteristics is betweenness centrality defined as follows

$$B(x) = \sum_{\{u,v\} \in \binom{V(G)\setminus\{x\}}{2}} \frac{\sigma_{u,v}(x)}{\sigma_{u,v}}, \tag{1.8}$$

where $\sigma_{u,v}$ denotes the number of shortest paths between u and v and $\sigma_{u,v}(x)$ is the number of shortest paths between u and v passing through x.

Connectivity and the Diameter of Betweenness-Uniform Graphs: The graph G is called *betweenness uniform* (BUG) if all vertices from $V(G)$ have the same value of betweenness. There exist some trivial examples of BUGs, such as vertex-transitive graphs. However, there are betweenness-uniform graphs that are neither vertex- nor edge-transitive. Several authors attempt to classify these graphs. In one of the results, Gago, Hurajová-Coroničová, and Madaras [40] asked several open questions. Our results, see [52] or its reprint P-1, positively answer two of them.

Vertex connectivity of G, $\kappa(G)$, is minimal size of a vertex cut in G. We say that G is *k-connected* if $|V(G)| > k$ and G always remains connected after the removal of less than k vertices. We have shown the following.

Theorem 1.4.3 (Theorem 2 in [52])**.** *If G is a connected betweenness-uniform graph then it is a cycle or a 3-connected graph.*

This limits the space for the potential future search of BUGs. Another result deals with maximal distances in a graph. For two vertices x, y, the length of the shortest xy-path is their distance $d(x, y)$. Diameter $d(G)$ of a graph G is then $\max_{x,y \in V(G)} d(x, y)$. We get the following.

Theorem 1.4.4 (Theorem 3 in [52])**.** *If G is betweenness-uniform graph and $\Delta(G) = n - k$, then $d(G) \leq k$.*

We have also improved the bound given in last mentioned theorem to a the following: $d(G) \leq \lceil \frac{k+5}{2} \rceil$ (Theorem 11 in [52]).

1.5 Uncertainty in network design

One of the major issues of complex networks created from data is the uncertainty in the input. There are multiple ways to represent and handle uncertainty. One of the theoretically strong approaches is represented by interval methods [86]. Roughly speaking, this approach substitutes real numbers in a given problem with intervals representing lower and upper bound of potential values. Subsequent computations attempt to express the solution as an interval of possible outputs considering any combination of values from the input variables. This is

quite interesting for networks, as noticed recently [5, 6]. Networks are commonly represented by matrices such as the adjacency matrix. For that reason, we can explore the reliability of computation with interval matrices or related matrix perturbation theory aiming to contribute to the Goal 1.1.5.

1.5.1 Radius of regularity

One of the issues that might be problematic when working with dynamical systems is the singularity of corresponding matrices, e.g., when we care about the stability of linear time-invariant dynamical systems [96]. Assume that we know that the matrix A of our system is regular, and thus the system is stable. We are interested in the nearest singular matrix to keep the system stable even considering the small perturbation of its values. The distance can be measured using the radius of nonsingularity defined as follows. Given a matrix $A \in \mathbb{R}^{n\times n}$, the *radius of nonsingularity* (or *regularity radius*) is defined by

$$d(A) := \inf\{\varepsilon > 0;\ (\exists \text{ singular } B)(\forall i,j) : |a_{ij} - b_{ij}| \leq \varepsilon\}. \tag{1.9}$$

In other words, it is the minimum distance of A to a singular matrix in the maximum norm. We also consider a generalization of this radius as follows.

$$d(A,\Delta) := \inf\{\varepsilon > 0;\ (\exists \text{ singular } B)(\forall i,j) : |a_{ij} - b_{ij}| \leq \varepsilon\Delta_{ij}\}, \tag{1.10}$$

where $\Delta \in \mathbb{R}^{n\times n}$ is a given non-negative matrix. Thus, $d(A)$ is a special case of $d(A,\Delta)$ when putting $\Delta := E$, and the matrix E consists of all ones.

Approximating regularity radius: To compute this regularity radius, we can make use of the following relationship.

$$d(A) = \frac{1}{\|A^{-1}\|_{\infty,1}}, \tag{1.11}$$

where $\|\cdot\|_{\infty,1}$ is a matrix norm defined as $\|M\|_{\infty,1} := \max\{\|Mx\|_1;\ \|x\|_\infty = 1\} = \max\{\|Mz\|_1;\ z \in \{\pm 1\}^n\}$. Computing this norm is known to be NP-hard problem. We develop, see [44] or its reprint P-10, a randomized approximation algorithm to calculate this norm. Note that computing this norm can be represented as an integer programming. We utilize the semidefinite relaxation of this problem motivated by the randomized approximation algorithm for the MaxCut problem.

The final randomized algorithm has an approximation ratio equal to 0.78343281. Besides the design of a new approximation algorithm for regularity radius, we also produced rigorous upper and lower bounds improving previously known results, see

$$0.78343281\gamma = \frac{\alpha}{2-\alpha}\gamma \leq \|M_{ij}\|_{\infty,1} \leq \gamma + \epsilon \tag{1.12}$$

where γ is an optimal solution for a norm and $\alpha = 0.87856723$ is the Goemans-Williamson value characterizing the approximation ratio of their approximation algorithm for MaxCut.

Regularity radius for various classes of matrices: We follow our interests in regularity radius in the subsequent work, see [45] or its reprint P-7. In this work, we show other properties of the regularity radius and explore its computation for

several classes of matrices, contributing thus to the overall understanding of this characteristic. At first, we show that checking the finiteness of this characteristic is polynomial.

Theorem 1.5.1 (Theorem 3.2 in [45]). *Checking whether* $\mathrm{r}(A, \Delta) = \infty$ *is a polynomial problem.*

We also provide a sharp upper bound on the number of non-zero elements of Δ to achieve infinite $\mathrm{r}(A, \Delta)$ (Proposition 3.4 in [45]). We use the relationship between the maximum (Chebyshev) norm and the spectral norm to construct new bounds for the radius of regularity. Let u and v be the left and right singular vectors, respectively, corresponding to minimum singular value $\sigma_{\min}(A)$ and set $y := \mathrm{sgn}(u)$ and $z := \mathrm{sgn}(v)$. Then we have

$$\mathrm{r}(A) \leq \frac{1}{\rho_0(A^{-1}yz^T)},$$

where ρ_0 is the real spectral radius providing maximum from absolute values of real eigenvalues of the matrix and equal to 0 if no such eigenvalue exists.

Several results are provided for a situation where an input matrix has a special form, such as exact formulas for several special classes of matrices, e.g., for totally positive or inverse non-negative. We also design approximation algorithms for special classes, e.g., rank-one radius matrices. For tridiagonal matrices, we design a polynomial algorithm to compute the radius of regularity.

Searching for the regularity radius can be understood as a problem from interval algebra. In this field, we consider matrices containing intervals instead of real numbers. More formally, the *interval matrix* is defined as

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \{A \in \mathbb{R}^{n\times n};\ \underline{A} \leq A \leq \overline{A}\},$$

We can read this as a set of all matrices having elements between defined bounds. Note that an interval matrix can also be understood as an interval given by the relation between *midpoint matrix* A_c and the *radius matrix* A_Δ as

$$\boldsymbol{A} = [A_c - A_\Delta, A_c + A_\Delta]$$

An interval matrix is called *regular* if it consists merely of nonsingular matrices; otherwise, it is called *singular* (or *irregular*). Let's consider the simple case that we have one parameter δ and define the corresponding interval matrix as $\boldsymbol{A}_\delta = [A_c - A_\Delta, A_c + A_\Delta]$, where $A_\Delta = \delta ee^T$ for e being the vector of ones. The search for the distance to a singular matrix can be rephrased as a search for the minimal δ such that the interval matrix $\boldsymbol{A}_\delta$ becomes singular, i.e.

$$\mathrm{r}(A) = \min\{\delta \geq 0 \mid \boldsymbol{A}_\delta \text{ is singular}\}. \tag{1.13}$$

Considering situations where the known bounds are not tight enough, a new method based on the Jansson-Rohn algorithm for testing the regularity of an interval matrix is presented, which is a priory not exponential.

1.5.2 Interval matrix powers

To control or analyze dynamical systems, you often need to use matrix powers, remind the stability of linear systems, or random walk algorithms. What if the matrix itself is again subject to uncertainty? Motivated by our results of the radius of nonsingularity, we consider interval uncertainty.

Before defining interval matrix power, let us define some auxiliary notions. Note that an interval matrix of dimension 1×1 is an *interval* $\boldsymbol{a} = [\underline{a}, \overline{a}]$. We denote the space of these intervals as $\mathbb{IR}$ and the space of interval $n \times m$ matrices as $\mathbb{IR}^{n\times m}$. For a bounded set $\mathcal{B} \subset \mathbb{R}^{n\times n}$ we define an *enclosure* of $\mathcal{B}$ as any $\boldsymbol{B} \in \mathbb{IR}^{n\times n}$ such that $\mathcal{B} \subseteq \boldsymbol{B}$. The *interval hull* of $\mathcal{B}$, denoted by $\square\mathcal{B}$, is the smallest enclosure of $\mathcal{B}$, that is, $\square\mathcal{B} := \bigcap_{\mathcal{B}\subseteq\boldsymbol{B}\in\mathbb{IR}^{n\times n}} \boldsymbol{B}$.

We define *k-th interval matrix power* or *k-th power of interval matrix* $\boldsymbol{A}$ as

$$\boldsymbol{A}^k := \{A^k;\ A \in \boldsymbol{A}\}.$$

Since this is not an interval matrix in general, we are content with its *interval hull* $\left[\boldsymbol{A}^k\right] := \square\{A^k;\ A \in \boldsymbol{A}\}$. We can imagine this hull as the tightest n-dimensional box aligned with the axes.

There exists arithmetic for intervals with many nice properties. However, some of the expected ones are missing, e.g., the absence of distributivity. For that reason, this arithmetic can lead to overestimation when enumerating expressions. Computing interval matrix powers can lead to n-dimensional boxes mentioned above that are not as tightest as possible, i.e., just enclosures. Overestimation within the enumeration of algebraic terms containing intervals is well characterized, and thus we can identify the case when it happens. Following this observation, it is well known that computing the square of an interval matrix is polynomial, but the cube is already an NP-hard problem. Note that matrix powers can result in irrational numbers. For that reason, here and in further text, when speaking about computational complexity, we use the term polynomial always for polynomial up to a given accuracy.

Computing matrix powers for special classes of matrices: Note that if all matrix elements are values and not intervals, the problem of computing matrix powers is easy. We can consider a real value as a special type of interval that we call *degenerate interval* and for which holds $\underline{a} = \overline{a}$. We can say that for any interval matrix having all elements degenerate, computation of powers is easy. What if some of the elements are not degenerate? We can consider a class of interval matrices, where some elements are degenerate. An example can be an interval matrix where diagonal elements are non-degenerate intervals, and all remaining elements are degenerate. We consider several such subclasses in our next work, see [46] or its reprint P-3. For a matrix with only diagonal elements non-degenerate, computing the cube is polynomial. The proof of the NP-hardness of the cube for interval matrices uses the class of matrices that only have interval components in one row and one column. Motivated by this structure, we have shown that computing the cube of matrices with only one interval column is polynomial (Theorem 2.5). However, the major result concerns a fixed power coefficient k and the following special class of matrices.

Definition 1.5.1 (Definition 3.1 in [46])**.** Let k and m_e be fixed. We say that a class of interval matrices has *constant expression of k-th power* if any component

of their k-th powers can be expressed as a polynomial in at most m_e interval variables.

This class is not bounded in size, as we can see on tridiagonal interval matrices having m_e fixed for a fixed k. We can show that the computation of kth powers for these matrices is easy.

Theorem 1.5.1 (Theorem 3.2 in [46]). *Let k be fixed and let $\boldsymbol{A}$ be an interval matrix having rational end-points with constant expression of k-th power. Then computation of k-th power of $\boldsymbol{A}$ up to any given accuracy ε is a polynomial problem with respect to input data and $\log(1/\varepsilon)$.*

This polynomiality is shown by transforming the problem into a quantifier elimination task and utilizing the Tarski elimination method. This result is used to show the polynomiality of computing the cube of matrices generalizing diagonal ones, namely *interval band matrices* (Proposition 3.5 in [46]).

Among other results for smaller classes, such as companion matrices, we consider a rather specific class: linear parametric interval matrices defined as

$$A(p) := \sum_{q=1}^{m} p_q A^{(q)}, \quad p_q \in \boldsymbol{p}_q, \tag{1.14}$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m \in \mathbb{IR}$. Note that this generalization of interval matrices represents potential dependencies between elements of a matrix. We can approximate the solution via direct evaluation of the expression $\boldsymbol{A} := \sum_{q=1}^{m} \boldsymbol{p}_q A^{(q)}$ using standard interval arithmetic. This approach leads to overestimation, and thus it is better to use the special structure of parametric matrices. The interval hull of the kth power can be defined as

$$\left[A(\boldsymbol{p})^k\right] := \square\{A(p)^k;\, p \in \boldsymbol{p}\}.$$

We show the following result about the hardness of the corresponding computations already for $k = 2$.

Theorem 1.5.2 (Theorem 4.1 in [46])). *Computation of $\left[A(\boldsymbol{p})^2\right]$ is an NP-hard problem.*

The area of parametric interval matrices contains many open problems. Moreover, many of the studied problems in this area are computationally hard. Sometimes, we can help ourselves by restricting the elements of matrices as above. In fact, the proof of the above-mentioned Tarski method can be rephrased and show the variant of the Theorem 1.5.1 for parametric matrices (Theorem 4.4 in [46]).

Spectral interval method and interval matrix powers: Computing powers of (real) matrices can be handled using spectral decomposition of the original matrix. We use this simple observation to handle powers of interval matrices in our next work, see [47] or its reprint P-2. Indeed, assume that a matrix A is diagonalizable and we have its spectral decomposition $A = V\Lambda V^{-1}$, where $\Lambda \in \mathbb{C}^{n\times n}$ is a diagonal matrix with the eigenvalues of A on the diagonal, and $V \in \mathbb{C}^{n\times n}$ is a matrix having the corresponding eigenvectors as its columns. We can compute the kth power of matrix A as $A^k = V\Lambda^k V^{-1}$. We can follow the same approach with the interval matrix. To enable this approach, we need to first handle the spectral decomposition for interval matrices. Given $\boldsymbol{A} \in \mathbb{IR}^{n\times n}$, our problem states.

Find a diagonal matrix $\mathbf{\Lambda} \in \mathbb{IC}^{n\times n}$ and a matrix $\boldsymbol{V} \in \mathbb{IC}^{n\times n}$ such that for each $A \in \boldsymbol{A}$ there are $\Lambda \in \mathbf{\Lambda}$ and $V \in \boldsymbol{V}$ such that $A = V\Lambda V^{-1}$.

Once we have an enclosure $\boldsymbol{V}$, we can compute an enclosure of the set of inverses $\{V^{-1};\, V \in \boldsymbol{V}\}$ by methods from interval computation (and use $\boldsymbol{V}^{-1}$ to denote an interval enclosure of the set of such inverses). The high-level algorithm for matrix powers is as follows

Algorithm 1: enclosure of the k-th power by **spectral method**

Input: Matrix $\mathbf{A}$, power coefficient k
Result: Enclosure $\boldsymbol{C}$ of power $\mathbf{A}^k$

1 $\boldsymbol{\lambda} \leftarrow$ eigenvalues of $\boldsymbol{A}$
2 $\mathbf{\Lambda} \leftarrow \mathrm{diag}(\boldsymbol{\lambda})$ // create a diagonal matrix with eigenvalues
3 $\boldsymbol{V} \leftarrow$ eigenvectors of $\boldsymbol{A}$ // stored in columns
4 estimate enclosure of $\boldsymbol{V}^{-1}$
5 $\boldsymbol{C} = \boldsymbol{V}\mathbf{\Lambda}^k\boldsymbol{V}^{-1}$

The resulting interval matrix $\boldsymbol{C}$ is an enclosure of the k-th power of the interval matrix $\boldsymbol{A}$. We need to resolve two tasks to make this algorithm work:

1. Compute (the enclosure of) eigenvalues of an interval matrix

2. Compute (enclosures of) eigenvectors of an interval matrix

The first task, computing enclosures of the eigenvalues, can be done using the well-known Bauer-Fike theorem. Roughly speaking, for two matrices A, B such that A is diagonalizable, this theorem shows for any eigenvalue $\lambda(A + B)$ that there is an eigenvalue $\lambda_i(A)$ that is relatively close. This closeness can determine an interval for eigenvalues. For that reason, we can use this theorem to estimate interval eigenvalues in some cases – see assumptions and principles in Section 2.1 of [47].

The next step is to compute an interval eigenvector for a given interval eigenvalue. We enclose the null space of the interval matrix $\boldsymbol{A} - \boldsymbol{\lambda}I_n$. The idea is as follows. Let $A \in \boldsymbol{A}$ and λ be one of its (simple) eigenvalues, and x the corresponding eigenvector. This means that $Bx = 0$ for $B := A - \lambda I_n$ having rank $n - 1$. There has to be $x_j \neq 0$ and we can assume that x is normalized such that $x_j = 1$. The next step is to select the row and column indices carefully and rewrite the equation $Bx = 0$ into a solvable and not overdetermined system. Solving this new system can produce (after some adjustments) interval eigenvectors. This procedure includes the search for the best column and row indices. This can be computationally demanding. For that reason, we produce a heuristic search based on sufficient regularity conditions for interval matrices. This procedure reduces the complexity significantly.

The major contribution of the work mentioned above lies in providing previously absent spectral decomposition of interval matrices. This could help in various different applications. We utilize this approach relatively directly for the computation of interval matrix powers.

1.5.3 Quadratic programming approximations

It is well-known that many graph-theoretical problems can be transformed either to linear or quadratic programming. Even forgetting uncertainty in data, we can face problems with tractability, e.g., integer linear programming or maximization of a convex quadratic form. We can apply some approximation method, but we need to consider it as another source of uncertainty, this time connected with the precision of the computations. Considering approximation algorithms, further bounds for the optimized function are also helpful.

Bounds for maximization of a convex quadratic form: In our next work, see [55] or its reprint P-4, we consider maximization of a convex quadratic form on a convex polyhedral set, i.e.

$$f^* = \max\ x^T A x \ \text{ subject to } \ x \in \mathcal{M}, \tag{1.15}$$

where $\mathcal{M}$ is a convex polyhedral set and $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. This problem is known to be NP-hard even when $\mathcal{M}$ is restricted to be a box (hypercube). Several standard approaches include cutting plane methods, branch & bound methods, or polynomial-time approximation methods. Most of these methods exploit the structure of the feasible set and the objective function simultaneously. In contrast, we propose bounds based on a finer overestimation of the objective function in this work. This approach outperforms some of the other methods, see numerical tests in a separate paper [54].

The idea is as follows. Let A be a matrix with the properties given above. We can factorize this matrix into $A = G^T G$. We can express quadratic form as $x^T A x = x^T G^T G x = \|Gx\|_2^2$ and reformulate the problem as

$$\max\ \|Gx\|_2^2 \ \text{ subject to } \ x \in \mathcal{M}. \tag{1.16}$$

We focus on an upper bound of f^*. To derive such an upper bound we can estimate the Euclidean norm by some other norm. The main focus of this paper is to investigate the relation between the upper bounds and the factorization and the vector norm used. Denote by $\mathcal{H}$ the set of orthogonal matrices of size n and by

$$g^* \coloneqq \min_{R\in\mathbb{R}^{n\times n}: A = R^T R}\ \max_{x\in\mathcal{M}}\ \|Rx\|_\infty^2$$

the best upper bound obtained by a factorization of A. For this, we have the following.

Theorem 1.5.2 (Theorem 1 in [55])**.** *We have*

$$f^* = n \cdot \max_{x\in\mathcal{M}}\ \min_{H\in\mathcal{H}}\ \|HGx\|_\infty^2 \le n \cdot \min_{H\in\mathcal{H}}\ \max_{x\in\mathcal{M}}\ \|HGx\|_\infty^2 = g^*. \tag{1.17}$$

The theorem says that the upper bound g^* overestimates f^* the same way as max-min inequality. We provide an example (Example 1 in [55]) showing that the relation in the above inequality can be sharp. On the other hand, we show that the bound is not far away in general (Proposition 1 in [55]), namely $g^* \le n f^*$. We also provide arguments that this inequality holds as an equation in some instances (Proposition 2 in [55]).

We further consider one special case - the maximization on a box represented by an interval vector $\boldsymbol{x} = [\underline{x}, \overline{x}] = \{x \in \mathbb{R}^n;\ \underline{x} \le x \le \overline{x}\}$ that stands for $\mathcal{M}$.

We show that the bound given by the above theorem is not tight and provide an exemplary bound for f^* that stands in between f^* and g^* (Proposition 3 in [55]).

The original bound uses preconditioning with orthogonal matrices due to their zero effect on the quadratic form. We further consider a general class of matrices suitable for preconditioning when searching for upper bounds, i.e.

$$\begin{aligned}\mathcal{B} &\coloneqq \{B \in \mathbb{R}^{n\times n};\ \|x\|_2 \leq \sqrt{n}\|Bx\|_\infty\ \forall x \in \mathbb{R}^n\} \\ &= \{B \in \mathbb{R}^{n\times n};\ 1 \leq \sqrt{n}\|Bx\|_\infty\ \forall x \in \mathbb{R}^n : \|x\|_2 = 1\}.\end{aligned} \tag{1.18}$$

Consider that we have SVD decomposition of B. We show several properties of corresponding singular values (Propositions 6 - 8 in [55]) arriving at the end in the following claim.

Proposition 1.5.3 (Proposition 9 in [55])**.** *Checking $B \in \mathcal{B}$ is a co-NP-hard problem.*

To help in such unfortunate situation, we provide several sufficient conditions (Propositions 10 - 12 in [55]). We also consider a set similar to a set given in Equation 1.18 but for non-negative case of $Gx \geq 0$ (section 3.1 in [55]).

Finally, we were able to generalize Theorem 1.5.2 to general vector norm (Theorem 2 in [55]).

1.6 Closing remarks

We have provided a description of the research performed for complex networks covering many aspects of the analytical pipeline given in Figure 1.1. Section 1.2 covers the processing of data in the form of multivariate time series to build a complex network. This section discusses the handling of nonlinearity, dynamics of time series, or causality problems. In the following Section 1.3 we describe principles of building and analyzing the global structure of the network. The major topics are determining the vertices using dimensionality decomposition and the global network property of small-world. In the next Section 1.4 we consider one specific network global property: symmetry. First, we show research for various forms of homogeneity for large networks approximated by countable infinite graphs. This includes steps towards classifying homomorphism-homogeneous countable graphs and ranking homogenizations using relational complexity. Another presented topic dealing with symmetry was represented by research around the regularity of betweenness centrality. In the final Section 1.5, we describe research dealing with topics of numerical uncertainty in data used for network constructions. The first topic covers interval matrix tasks such as regularity radius and interval matrix powers. The second topic is about designing specific bounds for an NP-hard problem that can be used for approximate algorithms.

These research works cover various perspectives of complex networks. To solve related problems, we need to apply many scientific disciplines. Examples of domains are combinatorics, graph theory, model theory, optimization, linear algebra, interval algebra, statistics, data processing, theory of dynamical systems, data processing using neurological, climate, and financial data.

Notes to future work: There are many open questions in the above-described areas. Considering the dependency measures, we expect research around synchronization of dynamical systems given by specific differential equations, see [10]. As

an example, computing the Granger causality for diffusion processes, see [114]. Considering the global structure, we believe that there are many unresolved tasks. There are several interesting open problems around specific sparsity definitions, such as bounded expansion [87]. Another interesting generalization of a complex network is a hypergraph, see [14]. Considering the symmetry topics, we are still working on the classification of homomorphism-homogeneous graphs as well as the classification of betweenness uniform graphs. We expect to work with interval networks, see [6, 5], namely the community structure of interval graphs. Further research of parametric interval matrices properties represent another area.

Industrial project using complex networks: We have also used complex network approaches in several industrial projects, apart from scientific works. Let us provide a short note to several recent projects. The central topic was computer security. The first project was for Avast Software; see contract reference at [112]. The topic was the characterization of the structure of dependencies in a high-dimensional dynamic system. The next project was for Cisco Systems, Inc., see contract reference at [111]. The topic was the classification of malicious NetFlow communications, which will further serve to classify malicious hosts, with special attention to various time-series analysis methods. Another topic dealt with the automatic evaluation of crowd behavior and its modeling. The cooperation was with CertiCon a.s., see their annual report at [11].

References

[1] F. Agosta, S. Sala, A. Valsasina, P.and Meani, E. Canu, G. Magnani, S. F. Cappa, E. Scola, P. Quatto, M. A. Horsfield, et al. Brain network connectivity assessed using graph theory in frontotemporal dementia. *Neurology*, 81(2):134–143, 2013.

[2] R Albert and AL Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[4] R. Albert, H. Jeong, and A.L. Barabasi. Internet - Diameter of the World-Wide Web. *Nature*, 401(6749):130–131, 1999.

[5] Hélder Alves, Paula Brito, and Pedro Campos. Centrality measures in interval-weighted networks. *arXiv preprint arXiv:2106.10016*, 2021.

[6] Hélder Alves, Paula Brito, and Pedro Campos. Community detection in interval-weighted networks. *arXiv preprint arXiv:2106.10217*, 2021.

[7] A. Aranda and D. Hartman. The independence number of HH-homogeneous graphs and a classification of MB-homogeneous graphs. *European Journal of Combinatorics*, 85:103063, 2020.

[8] M. R. Arbabshirani, K. Kiehl, G. Pearlson, and V. D. Calhoun. Classification of schizophrenia patients based on resting-state functional network connectivity. *Frontiers in Neuroscience*, 7:133, 2013.

[9] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165, 2017.

[10] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Report - Review Section of Physics Letters*, 469(3):93–153, 2008.

[11] CertiCon a.s. Certicon Annual Report for 2020. `https://www.certicon.cz/about-us/?lang=en#media`, 2020. page 19, [Online; accessed 9-January-2022].

[12] Y. Bar-Yam. General features of complex systems. *Encyclopedia of Life Support Systems (EOLSS), UNESCO, EOLSS Publishers, Oxford, UK*, 1, 2002.

[13] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[14] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92, 2020.

[15] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner.

Creating a science of the Web. *Science*, 313(5788):769–771, 2006.

[16] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3, SI):535–559, 2012.

[17] B. B. et al. Biswal. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4734–4739, 2010.

[18] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez. Recommender systems survey. *Knowledge-based Systems*, 46:109–132, 2013.

[19] M. Boguná, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas. Models of social networks based on social distance attachment. *Physical review E*, 70(5):056122, 2004.

[20] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[21] A. Bonato. *A Course on the Web Graph*. Graduate studies in mathematics. American Mathematical Society, 2008.

[22] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

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

[24] X.-L. Cai, D.-J. Xie, K. H. Madsen, Y.-M. Wang, S. A. Bögemann, E. F. C. Cheung, A. Møller, and R. C. K. Chan. Generalizability of machine learning for classification of schizophrenia based on resting-state functional mri data. *Human Brain Mapping*, 41(1):172–184, 2020.

[25] P. J. Cameron and J. Nešetřil. Homomorphism-homogeneous relational structures. *Combinatorics Probability & Computing*, 15(1-2):91–103, 2006.

[26] G. I. Cherlin. The classification of countable homogeneous directed graphs and countable homogeneous n-tournaments. *Memoirs of American Mathematical Society*, 131(621):161, 1998.

[27] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628, 2000.

[28] T. D. H. Coleman, D. M. Evans, and R. D. Gray. Permutation monoids and MB-homogeneity for graphs and relational structures. *European Journal of Combinatorics*, 78:163–189, 2019.

[29] L. d. F. Costa, O. N. Oliveira, Jr., G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.

[30] N. A. Crossley, A. Mechelli, J. Scott, F. Carletti, P. T. Fox, P. McGuire, and E. T. Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

[31] D. Dagon, G. Gu, C. P. Lee, and W. Lee. A taxonomy of botnet structures. In *Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007)*, pages 325–339. IEEE, 2007.

[32] W. de Haan, W. M. van der Flier, H. Wang, P. F. A. Van Mieghem, P. Scheltens, and C. J. Stam. Disruption of functional brain networks in alzheimer's disease: what can we learn from graph spectral analysis of resting-state magnetoencephalography? *Brain Connectivity*, 2(2):45–55, 2012.

[33] L. Diambra and L. da F. Costa. Complex networks approach to gene expression driven phenotype imaging. *Bioinformatics*, 21(20):3846–3851, 2005.

[34] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. Complex networks in climate dynamics. *European Physical Journal - Special Topics*, 174:157–179, 2009.

[35] C. Eom and J. W. Park. Effects of common factors on stock correlation networks and portfolio diversification. *International Review of Financial Analysis*, 49:1–11, 2017.

[36] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In Guerin, R., editor, *ACM SIGCOMM'99 Conference: Applications, Technologies, Architectures, and Protocols for Computer Communications*, volume 29 of *Computer Communications Review*, pages 251–262, 1999.

[37] L. Ferrarini, I. M. Veer, E. Baerends, M.-J. van Tol, R. J. Renken, N. J. A. van der Wee, D. J. Veltman, A. Aleman, F. G. Zitman, B. W. J. H. Penninx, et al. Hierarchical functional modularity in the resting-state human brain. *Human Brain Mapping*, 30(7):2220–2231, 2009.

[38] R. Fraïssé. Sur certains relations qui généralisent l'ordre des nombres rationnels. *Comptes Rendus de l'Académie des Sciences*, 15:91–103, 2006.

[39] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

[40] Silvia Gago, Jana Coroničová Hurajová, and Tomáš Madaras. On betweenness-uniform graphs. *Czechoslovak Mathematical Journal*, 63(3):629–642, 2013.

[41] Z. Gao, W. Dang, X. Wang, X. Hong, L. Hou, K. Ma, and M. Perc. Complex networks and deep learning for eeg signal analysis. *Cognitive Neurodynamics*, pages 1–20, 2020.

[42] S. Goyal. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2012.

[43] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical*

review E, 68(6):065103, 2003.

[44] D. Hartman and M. Hladík. Tight bounds on the radius of nonsingularity. In M. Nehmeier, J. W. VonGudenberg, and W. Tucker, editors, *Scientific Computing, Computer Arithmetic, and Validated Numerics. SCAN 2015*, volume 9553 of *Lecture Notes in Computer Science*, pages 109–115. Springer, 2015.

[45] D. Hartman and M. Hladik. Regularity radius: Properties, approximation and a not a priori exponential algorithm. *Electronic Journal of Linear Algebra*, 33:122–136, 2018.

[46] D. Hartman and M. Hladík. Complexity of computing interval matrix powers for special classes of matrices. *Applications of Mathematics*, 65(5):645–663, 2020.

[47] D. Hartman, M. Hladík, and D. Říha. Computing the spectral decomposition of interval matrices and a study on interval matrix powers. *Applied Mathematics and Computation*, 403:126174, 2021.

[48] D. Hartman and J. Hlinka. Nonlinearity in stock networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(8):083127, 2018.

[49] D. Hartman, J. Hlinka, M. Paluš, D. Mantini, and M. Corbetta. The role of nonlinearity in computing graph-theoretical properties of resting-state functional magnetic resonance imaging brain networks. *Chaos*, 21(1):art.num.013119, 2011.

[50] D. Hartman, J. Hubička, and J. Nešetřil. Complexities of relational structures. *Mathematica Slovaca*, 65(2):229–246, 2015.

[51] D. Hartman, J. Hubička, and D. Mašulović. Homomorphism-homogeneous L-colored graphs. *European Journal of Combinatorics*, 35(0):313–323, 2014.

[52] D. Hartman, A. Pokorná, and P. Valtr. On the connectivity and the diameter of betweenness-uniform graphs. In *Conference on Algorithms and Discrete Applied Mathematics*, pages 317–330. Springer, 2021.

[53] Y. He, Z. Chen, and A. Evans. Structural insights into aberrant topological patterns of large-scale cortical networks in alzheimer's disease. *Journal of Neuroscience*, 28(18):4756–4766, 2008.

[54] M. Hladík and D. Hartman. Maximization of a convex quadratic form on a polytope: factorization and the chebyshev norm bounds. In *World Congress on Global Optimization*, pages 119–127. Springer, 2019.

[55] M. Hladík, D. Hartman, and M. Zamani. Maximization of a PSD quadratic form and factorization. *Optimization Letters*, pages 1–14, 2020.

[56] J. Hlinka, D. Hartman, N. Jajcay, D. Tomeček, J. Tintěra, and M. Paluš. Small-world bias of correlation networks: From brain to climate. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):035812, 2017.

[57] J. Hlinka, D. Hartman, N. Jajcay, M. Vejmelka, R. Donner, N. Marwan, J. Kurths, and M. Paluš. Regional and inter-regional effects in evolving

climate networks. *Nonlinear Processes in Geophysics*, 21(2):451–462, 2014.

[58] J. Hlinka, D. Hartman, and M. Paluš. Small-world topology of functional connectivity in randomly connected dynamical systems. *Chaos*, 22(3):art.num.033107, 2012.

[59] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotná, and M. Paluš. Nonlinear dependence and teleconnections in climate data: sources, relevance, nonstationarity. *Climate Dynamics*, 42(7-8):1873–1886, 2014.

[60] J. Hlinka, N. Jajcay, D. Hartman, and M. Paluš. Smooth information flow in temperature climate network reflects mass transport. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):035811, 2017.

[61] J. Hlinka, M. Paluš, M. Vejmelka, D. Mantini, and M. Corbetta. Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *Neuroimage*, 54(3):2218–2225, 2011.

[62] Jaroslav Hlinka, Milan Paluš, Martin Vejmelka, Dante Mantini, and Maurizio Corbetta. Functional connectivity in resting-state fmri: is linear correlation sufficient? *Neuroimage*, 54(3):2218–2225, 2011.

[63] W. Hodges. *A Shorter Model Theory*. Cambridge University Press, 1997.

[64] A. L. Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology*, 4(11):682–690, 2008.

[65] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

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

[67] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th workshop on social network mining and analysis*, pages 1–9, 2013.

[68] G. C. Kane, M. Alavi, G. Labianca, and S. P. Borgatti. What's different about social media networks? A framework and research agenda. *MIS Quarterly*, 38(1):275–304, 2014.

[69] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[70] S. Kwon, M. Cha, and K. Jung. Rumor Detection over Varying Time Windows. *PLoS ONE*, 12(1), 2017.

[71] A. H. Lachlan and R. E. Woodrow. Countable ultrahomogeneous graphs. *Transactions of the American Mathematical Society*, 262(1):51–94, 1980.

[72] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time: A new social network dataset using facebook.com. *Social Networks*, 30(4):330 – 342, 2008.

[73] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg.

The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[74] L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society colloquium publications. American Mathematical Society, 2012.

[75] S. Lovejoy and D. Schertzer. *The weather and climate: emergent laws and multifractal cascades*. Cambridge University Press, 2013.

[76] B. D. MacArthur, R. J. Sanchez-Garcia, and J. W. Anderson. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531, 2008.

[77] D. Macpherson. A survey of homogeneous structures. *Discrete Mathematics*, 311(15):1599 – 1634, 2011.

[78] R. N. Mantegna. Hierarchical structure in financial markets. *European Physical Journal B*, 11(1):193–197, 1999.

[79] R. N. Mantegna and H. E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.

[80] R. N. Mantegna and H. E. Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

[81] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[82] D. Meunier, S. Achard, A. Morcom, and E. Bullmore. Age-related changes in modular organization of human brain functional networks. *Neuroimage*, 44(3):715–723, 2009.

[83] S. Milgram. Small-world problem. *Phychology Today*, 1(1):61–67, 1967.

[84] T. Mizuno, H. Takayasu, and M. Takayasu. Correlation networks among currencies. *Physica A - Statistical Mechanics and its Applications*, 364:336–342, 2006.

[85] C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000.

[86] Ramon E. Moore, R. Baker Kearfott, and Michael J. Cloud. *Introduction to Interval Analysis*. SIAM, Philadelphia, PA, 2009.

[87] J. Nešetřil and P. Ossona De Mendez. Grad and classes with bounded expansion i. decompositions. *European Journal of Combinatorics*, 29(3):760–776, 2008.

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

[89] Milan Paluš. From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemporary physics*, 48(6):307–348, 2007.

[90] M. Paluš. Detecting phase synchronization in noisy systems. *Physics Letters A*, 235(4):341–351, 1997.

[91] M. Paluš, D. Hartman, J. Hlinka, and M. Vejmelka. Discerning connectivity from dynamics in climate networks. *Nonlinear Process Geophysics*, 18(5):751–763, 2011.

[92] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[93] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths. *Synchronization: a universal concept in nonlinear sciences*, volume 12. Cambridge university press, 2003.

[94] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen. Functional Network Organization of the Human Brain. *Neuron*, 72(4):665–678, 2011.

[95] D. Prichard and J. Theiler. Generating surrogate data for time-series with several simultaneously measured variables. *Physical Review Letters*, 73(7):951–954, 1994.

[96] L. Ravanbod, D. Noll, and P. Apkarian. Branch and bound algorithm with applications to robust stability. *Journal of Global Optimization*, 67(3):553–579, 2017.

[97] S. E. Rose. *Classification of Countable Homogeneous* 2*-Graphs*. PhD thesis, University of Leeds, Leeds, England, 2011. reprint available at `http://etheses.whiterose.ac.uk/1750/`.

[98] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[99] J. Runge, C. Petoukhov, J.F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(1):1–10, 2015.

[100] M. Rusinov and P. Schweitzer. Homomorphism–homogeneous graphs. *Journal of Graph Theory*, 65(3), 2010.

[101] Rubén J Sánchez-García. Exploiting symmetry in network analysis. *Communications Physics*, 3(1):1–15, 2020.

[102] A. Serra, P. Galdi, and R. Tagliaferri. Machine learning for bioinformatics and neuroimaging. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1248, 2018.

[103] M. A. Serrano and M. Boguna. Topology of the world trade web. *Physical Review E*, 68(1, Part 2):art.num.015101, 2003.

[104] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman. *Causation, Prediction, and Search*. MIT press, 2000.

[105] O. Sporns and J. D. Zwi. The small world of the cerebral cortex. *Neuroinformatics*, 2(2):145–162, 2004.

[106] C. K. Tse, J. Liu, and F. C. M. Lau. A network perspective of the stock

market. *Journal of Empirical Finance*, 17(4):659–667, 2010.

[107] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, Dimitrios D. Koutsouris, and I Fotiadis Dimitrios. A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals. *Computers in Biology and Medicine*, 99:24–37, 2018.

[108] A. A. Tsonis and P. J. Roebber. The architecture of the climate network. *Physica A*, 333(0):497 – 504, 2004.

[109] A. A. Tsonis and K. L. Swanson. Topology and predictability of El Nino and La Nina networks. *Physical Review Letters*, 100(22):art.num.228502, 2008.

[110] A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What do networks have to do with climate? *Bulletin of the American Meteorological Society*, 87(5):585–595, 2006.

[111] Ústav informatiky AV ČR v. v. i. and CISCO System Inc. Sponsored research agreement. `https://smlouvy.gov.cz/smlouva/12333592`, 2020. Czech National Register of Contracts, [Online; accessed 9-January-2022].

[112] Ústav informatiky AV ČR v. v. i. and Avast Software s.r.o. Research support agreement. `https://smlouvy.gov.cz/smlouva/12333592`, 2020. Czech National Register of Contracts, [Online; accessed 9-January-2022].

[113] M. Vejmelka, L. Pokorná, J. Hlinka, D. Hartman, N. Jajcay, and M. Paluš. Non-random correlation structures and dimensionality reduction in multivariate climate data. *Climate Dynamics*, 44(9):2663–2682, 2015.

[114] B. Wahl, U. Feudel, J. Hlinka, M. Wächter, J. Peinke, and J. A. Freund. Granger-causality maps of diffusion processes. *Physical Review E*, 93(2):022213, 2016.

[115] John M Wallace and David S Gutzler. Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly weather review*, 109(4):784–812, 1981.

[116] G.-J. Wang, C. Xie, K. He, and H. E. Stanley. Extreme risk spillover network: application to financial institutions. *Quantitative Finance*, 17(9):1417–1433, 2017.

[117] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao, J. Jia, Y. Han, and Y. He. Disrupted Functional Brain Connectome in Individuals at Risk for Alzheimer's Disease. *Biological Psychiatry*, 73(5):472–481, 2013.

[118] L. Wang, K. Li, and X. P. Hu. Graph convolutional network for fmri analysis based on connectivity neighborhood. *Network Neuroscience*, 5(1):83–95, 2021.

[119] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.

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

[121] D. Yao, J. Sui, E. Yang, P.-T. Yap, D. Shen, and M. Liu. Temporal-adaptive graph convolutional network for automated identification of major depressive disorder using resting-state fmri. In *International Workshop on Machine Learning in Medical Imaging*, pages 1–10. Springer, 2020.

[122] C. C. Zou, D. Towsley, and W. B. Gong. Email worm modeling and defense. In DaSilva, L. and Engbersen, T. and Park, E. K., editor, *ICCCN 2004: 13th International Conference on Computer Communications and Networks, Proceedings*, pages 409–414, 2004.

[123] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths. Complex network approaches to nonlinear time series analysis. *Physics Reports*, 787:1–97, 2019.

[124] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter. Detection and Resolution of Rumours in Social Media: A Survey. *ACM Computing Surveys*, 51(2), 2018.