Introduction to correlation networks: interdisciplinary approaches beyond thresholding

Masuda, N.; Boyd, Z.M.; Garlaschelli, D.; Mucha, P.J.

Physics Reports 1136 (2025) 1–39

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

Review article

Introduction to correlation networks: Interdisciplinary approaches beyond thresholding

Naoki Masuda [a,b,*], Zachary M. Boyd [c], Diego Garlaschelli [d,e], Peter J. Mucha [f]

[a] Department of Mathematics, State University of New York at Buffalo, United States of America
[b] Institute for Artificial Intelligence and Data Science, State University of New York at Buffalo, United States of America
[c] Department of Mathematics, Brigham Young University, United States of America
[d] Lorentz Institute for Theoretical Physics, Leiden University, The Netherlands
[e] IMT School of Advanced Studies, Lucca, Italy
[f] Department of Mathematics, Dartmouth College, United States of America

ARTICLE INFO

Article history:
Received 8 April 2025
Received in revised form 17 June 2025
Accepted 25 June 2025
Available online 30 June 2025
Editor: H. Orland

Keywords:
Network
Pearson correlation
Partial correlation
Covariance selection
Graphical lasso

ABSTRACT

Many empirical networks originate from correlational data, arising in domains as diverse as psychology, neuroscience, genomics, microbiology, finance, and climate science. Specialized algorithms and theory have been developed in different application domains for working with such networks, as well as in statistics, network science, and computer science, often with limited communication between practitioners in different fields. This leaves significant room for cross-pollination across disciplines. A central challenge is that it is not always clear how to best transform correlation matrix data into networks for the application at hand, and probably the most widespread method, i.e., thresholding on the correlation value to create either unweighted or weighted networks, suffers from multiple problems. In this article, we review various methods of constructing and analyzing correlation networks, ranging from thresholding and its improvements to weighted networks, regularization, dynamic correlation networks, threshold-free approaches, comparison with null models, and more. Finally, we propose and discuss recommended practices and a variety of key open questions currently confronting this field.

Contents

* Correspondence to: Department of Mathematics, State University of New York at Buffalo, North Campus, Buffalo, NY 14260-2900, USA.
E-mail address: naokimas@gmail.com (N. Masuda).

https://doi.org/10.1016/j.physrep.2025.06.002

1. Introduction

Correlation matrices capture pairwise similarity of multiple, often temporally evolving signals, and are used to describe system interactions in various diverse disciplines of science and society, from financial economics to psychology, bioinformatics, neuroscience, and climate science, to name a few. Correlation analysis is often a first step in trying to understand complex systems data [1]. Existing methods for analyzing correlation matrix data are abundant. Very well established methods include principal component analysis (PCA) [2] and factor analysis (FA) [3,4], which can yield a small number of interpretable components from correlation matrices, such as a global market trend when applied to stock market data, or spatio-temporal patterns of air pressure when applied to atmospheric data. Another major method for analyzing correlation matrix data is the Markowitz's portfolio theory in mathematical finance, which aims to minimize the variance of financial returns while keeping the expected return above a given threshold [5,6]. The model takes as input a correlation matrix among different assets that an investor can invest in. In a related vein, random matrix theory (RMT) [7–9] has been a key theoretical tool for estimating and analyzing economic and other correlation matrix data for a couple of decades [6]. Various new methods for analyzing correlation matrix data have also been proposed. Examples include detrended cross-correlation analysis [10–12], correlation dependency, defined as the difference between the partial correlation coefficient and the Pearson correlation coefficient given three nodes [13,14], determination of optimal paths between distant locations in correlation matrix data [15], early warning signals for anticipating abrupt changes in multidimensional dynamical systems including the case of networked systems [16–18], and energy landscape analysis for multivariate time series data particularly employed in neuroscience [19,20].

The last two decades have also seen successful applications of tools from network science and graph theory to correlational data. A correlation matrix can be mapped onto a network, which we refer to here as a *correlation network*, where nodes represent elements and edges are informed by the strength of the correlation between pairs of elements. Correlation network analysis generally intends to extract useful information from data, such as the patterns of interactions among nodes or a ranking of nodes. We show a typical workflow of correlation network analysis in Fig. 1. With multivariate data with multiple (and not too few) samples as input, the analysis entails calculation of correlation matrices, construction of correlation networks from the correlation matrices, and downstream network analysis on the resulting correlation networks. The network analysis often includes discussion of the implication of the network analysis results in application domains in question. An ideal correlation network analysis appropriately adapts concepts and methods developed in network science to the case of correlation networks, generating knowledge that standard methods for correlation matrices (such as PCA) do not produce. Although correlation does not necessarily reflect a physical connection or direct interaction between two nodes, correlation matrices are conventionally used as a relatively inexpensive substitute of such direct connections, whose data are often less available than correlation matrix data. Correlation networks are also useful for visualization [21]. Correlation network analysis has been used in various research disciplines, typically not much behind wherever correlation matrix analysis is used, as we will review in Section 2. In our survey here, we focus on correlation

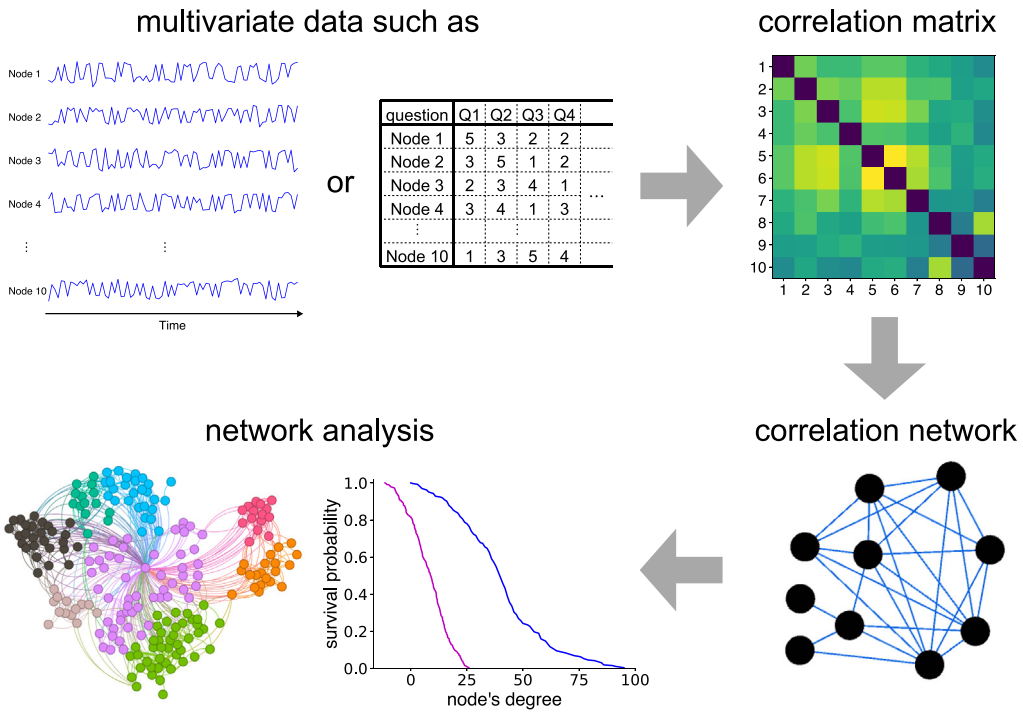

Fig. 1. Typical workflow of correlation network analysis. First, we are given multivariate data with L samples. A sample may correspond to a time point in the case of time series data or a respondent in the case of psychological questionnaires. Second, we compute the correlation matrix. In the present article, we focus on the case of Pearson correlation matrix although the subsequent steps of analysis equally apply to the case of other types of correlation or similarity matrices. Third, we construct a correlation network from the correlation matrix by, for example, thresholding on the pairwise correlation value. Fourth, we carry out network analysis on the generated correlation network. Network analysis typically includes calculation of some quantities or network features, such as communities (also called modules) and node centrality values, for the network and assess their implications, such as the difference between a disease group of samples and healthy control group of samples. See e.g. [21,24,26] for similar diagrams.

networks, with an emphasis on identifying different methods used to transform correlation matrices into correlation networks. See [21–25] for shorter reviews.

The validity of correlation network analysis remains an outstanding question, especially because the decisions about how to best construct network representations from correlation matrices is far from straightforward. One of the simplest methods is to threshold on the correlation value measured for each pair of nodes (see Section 3.2). However, while such a simple thresholding is widely used, it introduces various problems. These problems have led to proposals of alternative methods for generating correlation networks, which we will cover in sections 3.4–3.7.

Before proceeding, we raise some important clarifications. First, correlation networks as we consider here are different from network architectures that exploit correlation in data [27–29].[1] These "networks" are in the sense of neural network architecture in artificial intelligence and machine learning, whereas here we consider "networks" to denote graphs in network science.

Second, we focus on correlation networks based on the Pearson correlation coefficient or its close variants such as the partial correlation coefficient. In fact, there are numerous other definitions for quantifying the similarity between data obtained from node pairs [23,24,30–34]. Examples include similarity networks whose edges are determined using the rank correlation coefficient [35,36], the mutual information [37–39], and partial mutual information [40,41]. Co-occurrence of two nodes over samples, such as two authors co-authoring a paper (a paper is a sample in this example), also gives us an unnormalized variant of correlation. See sections 2.5 and 2.8 for examples of co-occurrence networks. However, a majority of concepts and techniques explained in our main technical sections 3 and 4, such as the detrimental effect of thresholding and dichotomizing the edge weight, use of weighted networks, graphical lasso, and importance of null models, also hold true when one constructs correlation networks using these or other alternative methods.

Third, we do not discuss causal inference in the present paper. A plethora of methods are available for inferring causality between nodes and associated directed networks. For example, a Bayesian network is a directed acyclic graph

[1] For example, the Progressive Spatio-Temporal Correlation Network (PSCC-net) is an algorithm to detect and localize manipulations in the input image data by taking advantage of spatial correlation structure in images [27]. The superpixel group-correlation network (SGCN) [28] and the deep correlation network (DCNet) [29] are encoder–decoder and deep-learning network architectures, respectively, for salient object detection in images.

that fully represents the joint probability distribution of the N variables. The edge of the directed acyclic graph represents directed and pairwise conditional dependency of one random variable (corresponding to a node) on another variable. The absence of the edge represents conditional independence between the two nodes. For Bayesian networks, see, e.g., [42,43]. For other techniques, see, e.g., [21,34,44]. While these methods reveal potential causal links, even from cross-sectional data, we do not consider them further here in our discussion of correlation matrices and related general similarity matrices, whose inherently symmetric natures mean that these matrices or networks do not in principle inform us of causality or directionality between nodes (at least not in a straightforward manner [45–47]). In a related vein, we do not discuss time-lagged correlation of multivariate time series in this paper, since these are also asymmetric in general, although many of the same considerations we raise here also apply to lagged correlations.

2. Data types leading to correlation networks

Correlation network analysis is common in many research fields. In this section, we survey typical correlation networks and their analysis in representative research fields.

2.1. Psychological networks

There are various multivariate psychological data, from which one can construct networks [21,25]. For example, in personality research, researchers construct personality networks in which each node can be a personal trait or goal such as being organized, being lazy, and wanting to stay safe. Edges between a pair of nodes typically represent a type of conditional association between the two nodes. Samples are frequently participants in the research responding to various questionnaires on a numeric scale (e.g., 5-point scale ranging from 1: strongly disagree to 5: strongly agree) corresponding to nodes. From a cross-sectional data set, one can calculate (conditional) correlation between pairs of nodes. Researchers are also increasingly combining surveys with alternative data collection modalities, for example, sensor data for daily movement or neural markers of stress [48,49]. It is reasonable to use correlation between signals from different modalities (e.g., smartphones and brain scanners) to construct a correlation network [49].

Another major type of psychological network is symptom networks employed in mental health research. Symptoms of a psychological, including psychopathological, condition, such as major depression and schizophrenia, are interrelated. Furthermore, causality between symptoms such as fatigue, headaches, concentration problems, and insomnia, and a psychopathological disorder, is often unclear. It has been suggested that a disorder does not originate from a single root cause, which motivates the study of symptom networks [21,50–52]. Nodes in a symptom network are symptoms, and one can use association between pairs of symptoms calculated from the participants' responses to define edges. Analysis of symptom networks may help us to predict how an individual develops psychopathology in the future, understand comorbidity as strong connection between symptoms of two mental disorders, and propose central nodes as possible targets of intervention [52]. Health-promoting behaviors can also be treated as nodes in these networks to suggest key behavioral intervention points [53].

Panel data, i.e., longitudinal measurements of variables from samples, are increasingly common for network approaches [21]. In this case, one obtains correlation networks at multiple time points. Then, one can construct time-varying correlation networks (see Section 3.9) or within-person correlation networks [54] that reflect temporal symptom patterns and ideally expose individual differences and possible causal pathways in mental health patterns related to disorders [55,56]. However, the validity of psychological network approach should be further studied. Research has shown that symptom networks have poor reproducibility across samples, likely due to measurement error in assessing symptoms among other reasons [51,57].

2.2. Brain networks

Various notions of brain connectivity have been essential to better understanding different neural functions. Studies of such brain networks constitute a major part of a research field that is often referred to as *network neuroscience* [58,59]. (See also the related material about network representations in [60].) Multivariate time series of neuronal signals recorded from the brain are a major source of data used in network neuroscience research. Such data may be recorded in a resting state or when participants are performing some task. Functional networks or functional connectivity refer to correlation-based networks constructed from multivariate neuronal time series data, obtained through, e.g., neuroimaging or electroencephalography, where the term "functional" in this setting effectively means correlational. A typical node in the brain networks is either a voxel (i.e., cube of side length of, e.g., 1 mm) or a spherical region of interest (ROI), which is a sphere in the brain. In the case of multivariate time series data, there are various other methods to infer directed brain networks, which is referred to as effective connectivity, but we do not cover directed networks in this article. Brain networks based on anatomical connectivity between brain regions are referred to as structural networks. Functional connectivity, or a correlation-based edge between two nodes in the brain does not imply the presence of an edge between the same pair of nodes in the structural network. Indeed, one does not expect a one-to-one correspondence between functional and structural brain networks because several brain states and functions continuously arise from the same anatomical structure [61]. Still, the possibility of studying structural networks in combination with functional networks

on the same set of nodes is a distinguishing feature of brain networks, which can be used for an additional comparison when validating the outcome of correlation-based networks [62]. See [24,32,58,62–65] for reviews and comparisons of techniques for the estimation and validation of brain networks from the (partial) correlations. Examples of the use of these methods for (functional) network analysis are discussed later in this review.

The most typical functional neuronal networks come from neuroimaging data, in particular functional magnetic resonance imaging (fMRI) data, which are measured using blood-oxygeneration-level-dependent (BOLD) imaging techniques [66]. Functional connectivity between voxels or between spherical ROIs, or other types of nodes, is calculated by a correlation between fMRI time signals at the two nodes after one has bandpassed the fMRI time series at each node to remove artifacts, with a frequency band of, e.g., 0.01–0.1 Hz. Functional MRI improves on electroencephalogram (EEG) and magnetoencephalography (MEG) in spatial resolution at the expense of temporal resolution, but functional EEG and MEG networks are not uncommon. We also note that EEG and MEG signals are oscillatory, so one has to calculate the functional connectivity between each pair of nodes using methods that are aware of the oscillatory nature of the signal, such as using phase lag index or amplitude envelope correlation [67] rather than conventional correlation coefficients or mutual information.

Structural covariance networks are another type of correlation brain network where the edges are defined as the correlation/covariance of the corrected gray matter volume or cortical thickness between brain regions i and j, where the samples are participants [68,69]. Morphometric similarity networks are a variant of structural covariance networks. In morphometric similarity networks, one uses various morphometric variables, not just a single one such as cortical thickness, for each node (i.e., ROI) [70]. One calculates the correlation between two nodes by regarding each morphometric variable as a sample. Therefore, differently from structural covariance networks based on cortical thickness, one can calculate a correlation network for each individual.

In neuroreceptor similarity networks, an edge between two nodes, or ROIs, is the correlation in terms of receptor density [71]. Specifically, one first calculates a vector of neurotransmitter density for each ROI, with each entry of the vector corresponding to one type of receptor. Then, one computes the correlation between each pair of ROIs, called receptor similarity.

2.3. Gene co-expression networks

Genes do not work in isolation. Gene co-expression networks have been useful for figuring out webs of interaction among genes using network analysis methods [34,72–79]. They are a type of data in a subfield of network science often referred to as *network biology* or *network medicine*. Gene co-expression networks are correlation networks in the generalized sense considered here, including the case of other measures of similarity. A typical measurement is the amount of gene expression for different genes and samples, where a sample most commonly corresponds to a human or animal individual. If one measures the expression of various genes for the same set of samples, we can calculate the co-expression between each pair of genes by calculating the sample correlation, yielding a correlation matrix. Depending on the questions being asked in the study, it may be important to calculate the underlying correlations with different factors to account for the effects of heterogeneous gene frequencies [80,81]. It is common to transform a correlation matrix into a network and then apply various network analysis methods, for example community detection with the aim of estimating the group of genes that are associated with the same phenotype[2] such as a disease. In this manner, correlation network analysis has been a useful tool for gene screening, which can lead to identification of biomarkers and therapeutic targets. In addition to community detection, identifying hub genes in co-expression networks helps finding key genes, for example, for cancer [82].

Different ways of defining co-expression matrices and networks from gene expression data include tissue-to-tissue co-expression (TTC) networks [83] (also see [84,85]). A TTC network proposed in [83] is a bipartite network, and its node is a gene–tissue pair. An edge between two nodes, denoted by $(\tilde{g}_i, \tilde{t}_i)$ and $(\tilde{g}_j, \tilde{t}_j)$, where $\tilde{g}_i$ and $\tilde{g}_j$ are genes, and $\tilde{t}_i$ and $\tilde{t}_j$ are tissues, represents the sample correlation as in conventional co-expression networks. However, by definition, the correlation is calculated only between node pairs belonging to different tissues, i.e., only for $\tilde{t}_i \neq \tilde{t}_j$. Therefore, TTC networks characterize co-expression of genes across different tissues.

Co-expression of genes i and j implies that i and j are both expressed at a high level in some samples (usually individuals) and both expressed at a low level in other individuals. Co-expressed genes tend to be involved in common biological functions. There are in fact multiple biophysical and non-biophysical reasons for gene co-expression [76]. For example, a transcription factor, a protein that binds to DNA, may regulate different genes i and j that are physically close on a chromosome. If this is the case, differential levels of regulation by the transcription factor across individuals can create co-expression of i and j. Another mechanism of co-expression is that the expression of genes i and j, which may be located far from each other on the chromosome or on different chromosomes, may depend on the temperature. Then, i and j would be co-expressed if different individuals are sampled from living environments with different temperatures. Variation in ages of the individuals can similarly create co-expression among age-related genes. Alternatively, co-expression may originate from non-biological sources, such as technical or laboratory ones, whose exact origins are often unknown.

[2] A *phenotype* is a set of observable traits of an organism and is usually contrasted with the underlying *genotype* that causes (or influences) the phenotype.

One is often interested in looking for differential co-expression, which refers to the different levels of gene co-expression between two phenotypically different sets of samples, such as a disease set versus a control set, or in two types of tissues [76,78]. Differential co-expression often reveals information that one cannot obtain by examining differential expression (as opposed to co-expression), i.e., different levels of gene expression between the two sets of samples [86].

2.4. Metabolite networks

Metabolites are small molecules (e.g., amino acids, lipids, vitamins) that are intermediates or end products of metabolic reactions. One can also construct correlation networks from metabolomics data, or data of metabolites and their reactions [87,88]. To inform the edge, one measures pairwise correlation between the amounts of two metabolites given the samples. Like gene co-expression, correlation between metabolites can occur for multiple reasons, including knock-out of a gene coding an enzyme that is involved in a chemical reaction consuming or producing two metabolites, different temperatures or other environmental conditions under which different samples are obtained, or intrinsic variability owing to cellular metabolism [87]. Note that mass conservation within a moiety-conserved cycle produces negative correlation between at least one pair of metabolites involved in the reaction [89]. That said, in some cases one may consider correlation or other similarity between only a subset of metabolites that are not necessarily associated to one another by direct chemical processes but instead draw from a set of alternative biochemical processes (see, e.g., [90]).

2.5. Microbiome networks

Microbes interact with other microbe species as well as with their environments. Understanding of microbial composition and interaction in the human gut is expected to inform multiple diseases. Similarly, understanding soil microbial communities may contribute to enhancing plant productivity. Network analysis is adept at revealing, e.g., ecological community assembly and keystone taxa, and has been increasingly contributing to these fields.

In microbiome network analysis, one collects samples from, e.g., soil, at various time points or locations. Each sample from an environment (e.g., soil, gut, animal corpus, or water) contains various microorganisms with different quantities. Co-occurrence network analysis is increasingly common in this field, aided by an increasing amount and accuracy of data [33,91,92]. In a microbiome co-occurrence network, nodes are microorganisms (e.g., bacteria, archaea, or viruses), specified at the taxa level, for example, and an edge is defined to exist if two nodes co-occur across the samples. Therefore, microbiome co-occurrence networks are essentially microbiome correlation networks, and the usual correlation measures, such as Pearson correlation, can be used to determine edge data, but more sophisticated methods to define edges are more often used. (See [33,91] for various co-occurrence network construction methods.) Positively weighted edges result because of, e.g., cooperation between two taxa, sharing of niche requirements, or co-colonization. Negatively edges result because of, e.g., competition for space or resources, prey–predator relationships, or niche partitioning. A historically famous example of negative co-occurrence in ecological community assembly study is the checkerboard-like presence–absence patterns of two bird species inhabiting an island, discussed by Jared Diamond [93]. (Also see [33] for a historical account.) Regardless, one should keep in mind that correlation, or co-occurrence, does not immediately imply physical interaction between two taxa.

2.6. Disease networks

A node in a disease network is a disease phenotype. Correlation between two diseases defines an edge, and there are various definitions of edges as we introduce in this section. Each definition of edge creates a different type of disease network.

Comorbidity, also called multimorbidity [94], is the simultaneous occurrence of multiple diseases within an individual. One cause of comorbidity is that the same gene or disease-associated protein can trigger multiple diseases. Other causes, such as environmental factors or behaviors, such as smoking, can also result in comorbidity. A collection of potentially comorbid diseases can be modeled as the nodes of a network, and the edges, which are based on comorbidity or other similarity index between diseases [88], are correlational in nature.

The authors of [95] constructed phenotypic disease networks (PDNs) in which nodes are disease phenotypes. The edges are sample correlation coefficients or a variant, and the samples are patients in a hospital claim record (i.e., Medicare claims in the US). Note that here one uses correlation for binary variables because each sample (i.e., patient) is either affected or not affected by any disease i. The authors found that, for example, patients tend to develop illness along the edges of the PDN [95].

Similarly, prior work constructed a human disease network when two diseases share at least one associated gene, which is similar in principle to the phenotypic disease network despite that the edge of the human disease network is not a conventional correlation coefficient [96] (also see [97]). Similarly, an edge in a metabolic disease network is defined to exist when two diseases are either associated with the same metabolic reaction or their metabolic reactions are adjacent to each other in the sense that they share a compound that is not too common [98]. (H_2O and ATP, for example, are excluded because they are too common.) Alternatively, in a human symptom disease network [99], the edge between a pair of diseases is a correlation measure in which each sample is a symptom. In other words, roughly speaking, the edge weight is large when two diseases share many symptoms.

2.7. Financial correlation networks

Stocks of different companies are interrelated, and the prices of some of them tend to change similarly over time. A common transformation of such financial time series before constructing correlation matrices and networks is into the time series of logarithmic return, i.e., the successive differences of the logarithm of the price, given by

$$x_i(t) = \ln \frac{z_i(t+1)}{z_i(t)}, \tag{1}$$

where $z_i(t)$ is the price of the ith financial asset at time t, such as the closure price of the ith stock on day t, and $t \in \{0, \ldots, T-1\}$. An advantage of this method is that $x_i(t)$ is not susceptible to changes in the scale of $z_i(t)$ over time [100]. Then, one constructs the correlation matrix for N time series $\{x_i(1), \ldots, x_i(T-1)\}$, where $i \in \{1, \ldots, N\}$.

Financial correlation matrices have been analyzed for decades. For example, Markowitz's portfolio theory provides an optimal investment strategy as vector $\boldsymbol{w} = (w_1, \ldots, w_N)^\top$, where w_i represents the fraction of investment in the ith financial asset, and $^\top$ represents the transposition [5,6]. The theory formulates the minimizer of the variance of the return, $\boldsymbol{w}^\top C \boldsymbol{w}$, where C is the covariance matrix, as the solution of a quadratic optimization problem with the constraint that the expected return, $\boldsymbol{w}^\top \boldsymbol{g}$, where $\boldsymbol{g} = (g_1, \ldots, g_N)^\top$, and g_i is the expected return for the ith asset, is larger than a prescribed threshold.

Financial correlation matrices have also been extensively studied in econophysics research since the 1990s, with successful uses of RMT [6,100–106] and network methods such as maximum spanning trees [107,108], community detection [104–106,109,110], and more advanced methods (see [22,23] for reviews). One usually employs RMT in this context to verify that most eigenvalues of the empirical financial correlation matrices lie in the bulk part of the distribution of eigenvalues for random matrices. Such results imply that most eigenvalues of the empirical correlation matrices can be regarded as noise, and one is primarily interested in other dominant eigenvalues of the empirical correlation matrices whose values are not explained by random matrix theory [101–106]. The largest eigenvalue is usually not explained by RMT and is often called the market mode because it represents the movement of the market as a whole; moreover, other deviating eigenvalues are also present, encoding for the presence of groups of stocks that move coherently, as we discuss in Section 4.3.

Other types of financial data are possible. For example, correlation networks were constructed from pairwise correlation between the daily time series of the investor's behavior (e.g., the net volume of Nokia stock traded or its normalized version) for two investors [111,112]. One can also renormalize the covariance matrix using other indices, such as momentum [113].[3]

2.8. Bibliometric networks

Apart from microbiome studies, bibliometric and scientometric studies are another research field in which co-occurrence networks are often used [115,116]. For example, in an academic co-authorship network, a node represents an author, and an edge represents co-occurrence (i.e., collaboration) of two authors in at least one paper. One can weigh the edge according to the number of coauthored papers or its normalized variant [117]. While keeping authors as nodes, one can also create other types of co-occurrence networks, such as co-citation networks in which an edge connects two authors whose papers are cited together by a later paper, and keyword-based co-occurrence networks in which an edge connects two authors sharing keywords associated with their papers. Nodes of co-occurrence bibliometric networks can also be journals, institutions, research areas, and so forth. These co-occurrence networks are mathematically close to correlation networks and have been useful for understanding research communities and specialities, communication among researchers, interdisciplinarities, and the structure and evolution of science, for example.

Various other web-based information has also been analyzed as co-occurrence networks. For example, tags annotating posts in social bookmarking systems can be used as nodes of co-occurrence networks [118]. Two tags are defined to form an edge if both tags appear on the same post at least once. One can also use the number of the posts in which the two tags co-appear as the edge weight. Another example is co-purchase networks in online marketplaces, in which a node represents an item, and an edge represents that customers frequently purchase the two items together [119].

2.9. Climate networks

Climate can be analyzed as a network of interconnected dynamical systems [120–123]. In most analyses, the nodes of the network are equal-angle latitude–longitude grid points on the globe. However, such angular partitions lead to grid cells with geometric areas that vary with latitude, which in particular might lead to spurious correlations in the measured quantities, especially near the poles; such biases might be addressed either by a node splitting scheme that aims

[3] Momentum in finance generally refers to the rate of change in price. If the prices of two assets are correlated, the momentum of one asset can be informative of the future price of the correlated asset [114]. In [113], momentum and price correlation are mixed in various ways to construct correlation-type networks that reflect collective price dynamics, and, for example, network centrality is predictive of large upcoming swings in asset prices.

to obtain consistent weights for the network parameters [124], or by choosing instead to work on a grid with (possibly only approximately) equal grid cell areas [125]. Each node has, for example, a time series measurement of the pressure level, which represents wind circulation of the atmosphere. The edge between a pair of nodes is based on the correlation between the two time series. An early study showed that all nodes in equatorial regions have large degree (i.e., the number of edges that a node has) regardless of the longitude, whereas only a small fraction of nodes in the mid-latitude regions had large degrees [120]. Climate networks have been further used for understanding mechanisms of climate dynamics and predicting extreme events. For example, early warning signals were constructed from the degree of the nodes and clustering coefficient for climate networks of the Atlantic temperature field [126]. The proposed early warning signals were effective at anticipating the collapse of Atlantic Meridional Overturning Circulation. See section 2.1 of [123] for more examples.

3. Methods for creating networks from correlation matrices

To apply network analysis to correlation matrix data, we need to generate a network from correlation data (usually in the form of a correlation matrix). We call such a network a *correlation network*. Whether correlation network analysis works or is justified depends on this process. Although there are various methods for constructing correlation networks from data, they have pros and cons. Furthermore, there are various unjustified practices around correlation network generation, which may yield serious limitations on the power of correlation network analysis. In this section, we review several major methods.

3.1. Estimation of covariance and correlation matrices

How to estimate covariance matrices from observed data when the matrix size is not small is a long-standing question in statistics and surrounding research fields. In particular, the sample covariance matrix, a most natural candidate, is known to be an unreliable estimate of the covariance matrix. See [6,127–129] for surveys on estimation of covariance matrices. Although the primary focus of this paper is estimation of correlation networks, not covariance or correlation matrices, it is of course important to realize that correlation networks created from unreliably estimated correlation matrices are themselves unreliable. Therefore, we briefly survey a few techniques of covariance and correlation matrix estimation in this section, including providing the notations and preliminaries used in the remainder of this paper. This exposition is important in particular because correlation network analysis in non-statistical research fields such as network science and also various applications often ignores statistical perspectives examined in the previous studies.

We denote by $(x_{i\ell})$ an $N \times L$ data matrix, where N is the number of nodes to observe the signal from, and L is the number of samples, which is typically the length of the time series or the number of participants in an experiment or questionnaire. The sample covariance matrix, $C^{\mathrm{sam}} = (C^{\mathrm{sam}}_{ij})$, is given by

$$C^{\mathrm{sam}}_{ij} = \frac{1}{L-1}\sum_{\ell=1}^{L}(x_{i\ell} - \bar{x}_i)(x_{j\ell} - \bar{x}_j), \tag{2}$$

where

$$\bar{x}_i = \sum_{\ell=1}^{L}\frac{x_{i\ell}}{L} \tag{3}$$

is the sample mean of the signal from the ith node.[4] Because Eq. (2) is a sum of L outer products, the rank of C^{sam} is at most L. One can understand this fact more easily by rewriting Eq. (2) as

$$C^{\mathrm{sam}} = \frac{1}{L-1}\sum_{\ell=1}^{L}\tilde{\boldsymbol{x}}_\ell\tilde{\boldsymbol{x}}_\ell^{\top}, \tag{4}$$

where $\tilde{\boldsymbol{x}}_\ell = (x_{1\ell} - \bar{x}_1, \ldots, x_{N\ell} - \bar{x}_N)^{\top}$ Therefore, C^{sam} is singular if $L < N$, while the converse does not hold true in general.

Although covariance matrices are mathematically convenient, they are not normalized. In particular, if we multiply the data from the ith node by c (> 0), then C^{sam}_{ij} ($= C^{\mathrm{sam}}_{ji}$), where $j \neq i$, changes by a factor of c, and C^{sam}_{ii} changes by a factor of c^2, whereas all the other entries of C^{sam} remain the same. In practice, the data from different nodes may have different baseline fluctuation levels. For example, the price of the ith stock may fluctuate much more than that of the jth stock because the former has a larger average or the industry to which the ith company belongs may be subject to higher temporal variability. The correlation matrix, denoted by ρ, normalizes the covariance matrix such that ρ is not subject

[4] Note that the $L-1$ in the denominator of Eq. (2) is necessary to obtain an unbiased estimator.

to the effect of different overall amounts of fluctuations across different nodes. The sample Pearson correlation matrix, denoted by $\rho^{\rm sam} = (\rho_{ij}^{\rm sam})$, is defined by

$$\rho_{ij}^{\rm sam} = \frac{\sum_{\ell=1}^{L}(x_{i\ell} - \bar{x}_i)(x_{j\ell} - \bar{x}_j)}{\sqrt{\sum_{\ell=1}^{L}(x_{i\ell} - \bar{x}_i)^2}\sqrt{\sum_{\ell=1}^{L}(x_{j\ell} - \bar{x}_j)^2}} = \frac{C_{ij}^{\rm sam}}{\sqrt{C_{ii}^{\rm sam} C_{jj}^{\rm sam}}}. \tag{5}$$

Note that $\rho_{ii}^{\rm sam} = 1\ \forall i \in \{1, \ldots, N\}$. Also note that every sample correlation matrix is a sample covariance matrix of some data but not vice versa. A correlation matrix is characterized by positive semidefiniteness, symmetry, range of the entries only being $[-1, 1]$, and the diagonal being equal to 1 [130]. The set of full-rank correlation matrices for a fixed N is called the elliptope, which has its own geometric structure [131,132]. For standardized samples $y_{i\ell} = (x_{i\ell} - \bar{x}_i)/\sqrt{C_{ii}^{\rm sam}}$, the Euclidean distance between vectors $(y_{i1}, \ldots, y_{iL})$ and $(y_{j1}, \ldots, y_{jL})$ is given by

$$d_{ij}^2 = \sum_{\ell=1}^{L}\left(y_{i\ell} - y_{j\ell}\right)^2 = 2 - 2\rho_{ij}^{\rm sam}. \tag{6}$$

Therefore, given the sample correlation matrix, $d_{ij} = \sqrt{2(1 - \rho_{ij}^{\rm sam})}$ defines a Euclidean distance [100,107].

In the following, we refer to covariance matrices in some cases and correlation matrices in others, often interchangeably. This is because the correlation matrix, which is a normalized quantity, should be used in most data analyses, while the covariance matrix allows better mathematical analysis in most cases. This convention is not problematic because, if we standardize the given data first and then calculate the covariance matrix for the standardized data $(y_{i\ell})$, then the obtained covariance matrix is also a correlation matrix. Therefore, the mathematical results for covariance matrices also hold true for correlation matrices as long as we feed the pre-standardized data to the analysis pipeline.

With the above consideration in mind, we now stress that it is important to distinguish the *sample* covariance matrix, which is calculated from empirical data, from the theoretical or 'true' (also called *population*) covariance matrix. One may use the true covariance matrix to model the observed data mathematically in terms of a random process described by a (stationary) probability distribution. Let X_i denote a random variable for $i \in 1, \ldots, N$. The true covariance matrix $C = (C_{ij})$ is given by

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)], \tag{7}$$

where E represents the expectation, and $\mu_i = E[X_i]$ is the ensemble mean of X_i. Eq. (7) implies that a covariance matrix is a symmetric matrix. It is also a positive semidefinite matrix. Conversely, a symmetric positive semidefinite matrix C is always a covariance matrix for the following reason. Any positive semidefinite matrix M allows a positive semidefinite matrix, denoted by $M^{\frac{1}{2}}$, as square root. We set

$$\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = C^{\frac{1}{2}} \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix}, \tag{8}$$

where $U_1, \ldots, U_N$ are independent random variables, with each having mean 0 and variance 1. Then, it is straightforward to see that $(X_1, \ldots, X_N)^\top$ has mean 0 and covariance matrix C.

Having distinguished population and sample covariance matrices, we now look for a relationship between the two. If we regard X_i as a random variable, and the observed data as a possible realization of that variable, then we must also regard the sample covariance matrix $C^{\rm sam}$ as a random variable, and the empirical sample correlation matrix as a single realization of that variable. Then, the relevant question is the characterization of the probability distribution governing the sample covariance matrix, given the true covariance matrix C which acts a set of fixed parameters (to be estimated from the data) for the distribution. Clearly, the number L of observations, regarded as the number of independent draws for each random variable X_i, is another parameter (not to be estimated). For any finite L, the sample covariance matrix obeys the so-called Wishart distribution with L degrees of freedom, denoted by $W_N(L, C)$, under the assumption that the L samples are i.i.d. and obey the multivariate normal distribution whose covariance matrix is C [3,6,133,134]. We obtain $E[C^{\rm sam}] = C$. In other words, the sample covariance matrix is an unbiased estimator of the true covariance matrix, called the Pearson estimator in statistics. The variance of $C_{ij}^{\rm sam}$ is equal to $(C_{ij}^2 + C_{ii}C_{jj})/L$. In fact, $C^{\rm sam}$ is a problematic substitute of C, and the use of $C^{\rm sam}$ in applications in place of C tends to fail; see [6] for an example in portfolio optimization. An intuitive reason why $C^{\rm sam}$ is problematic is that, if L is not much larger than N, which is often the case in practice, one would need to estimate many parameters from a relatively few observations. Specifically, the covariance and correlation matrices have $N(N+1)/2$ and $N(N-1)/2$ unknowns to infer, respectively, whereas there are L samples of vector $(x_{1\ell}, \ldots, x_{N\ell})$ available [135]. If N/L is not vanishingly small (called the large dimension limit or the Kolmogorov regime [6]), then the estimation would fail. As an extreme example, if $L < N$, then $C^{\rm sam}$ is singular, but the true C may be nonsingular. Even if $L \geq N$, matrix $C^{\rm sam}$ may be ill-conditioned if L is not sufficiently greater than N, whereas C may be well-conditioned.

Therefore, covariance selection to reduce the number of parameters to be estimated is a recommended practice when L is not large relative to N [135]. One also says that we impose some structure on the estimator of the covariance matrix,

with the mere use of the sample covariance matrix as an estimate of the true covariance matrix corresponding to no assumed structure.

A major method of covariance selection is to impose sparsity on the covariance matrix or the so-called precision matrix (also called the concentration matrix), which is the inverse of the covariance matrix (with entries representing un-normalized partial correlation coefficients; see Section 3.6). Note that a sparse precision matrix does not imply that the corresponding covariance matrix (i.e., its inverse) is sparse. Graphical lasso (see Section 3.7) is a popular method to estimate a sparse precision matrix. Another major method to estimate a sparse correlation matrix is to threshold on the value of the correlation to discard node pairs with correlation values close to 0 (see Section 3.2). Another common method of covariance selection, apart from estimating a sparse covariance matrix, is covariance shrinkage (see [136] for a review). With covariance shrinkage, the estimated covariance matrix is a linear weighted sum of the sample covariance matrix, C^{sam}, and a much simpler matrix, called the shrinkage target, such as the identity matrix [137–139] or the so-called single-index model (which is a one-factor model in factor analysis terminology and is an approximation of C^{sam} by a rank-one matrix plus residuals) [137]. Note that the shrinkage target is a biased estimator of C. These and other covariance selection methods balance between the estimation biases and variances.

An advanced estimation method for the entire correlation matrix, based on RMT, is the so-called optimal rotationally invariant estimator (RIE) [6,65]. Roughly speaking, the RIE is the closest (in some spectral sense) matrix, among all those having the same eigenvectors as the sample correlation matrix, to the 'true' correlation matrix. It uses a certain self-averaging property to infer the spectrum of the true matrix from that of the sample matrix, and then to compute the spectral distance from the true matrix to be minimized [6]. A specific cross-validated version of the RIE has been recently shown to outperform several other estimators [65]. Since the RIE requires some notion of RMT to be properly defined, we discuss it in Section 4.3.

3.2. Dichotomization

In this and the following subsections, we present several methods to generate undirected networks from correlation matrices.

A simple method to generate a network from the given correlation matrix is thresholding, which in its simplest form entails setting a threshold θ, and placing an unweighted edge (i, j) if and only if the Pearson correlation $\rho_{ij} \geq \theta$; otherwise, we do not include an edge (i, j). It is also often the case that one thresholds on $\left|\rho_{ij}\right|$. There are mainly two choices after the thresholding. First, we may discard the weight of the surviving edges to force it to 1, creating an unweighted network. Second, we may keep the weight of the surviving edge to create a weighted network. See Fig. 2 for these two cases. The literature often use the term thresholding in one of the two meanings without clarification. In the remainder of this paper, we call the first case *dichotomizing* (which can be also called binarizing), which is, precisely speaking, a shorthand for "thresholding followed by dichotomizing". We discuss dichotomized networks in this section and threshold networks without dichotomization (yielding weighted networks) in sections 3.4 and 3.7.

Dichotomizing has commonly been used across research areas. However, researchers have repeatedly pointed out that dichotomizing is not recommended for multiple reasons.

First, no consensus exists regarding the method for choosing the threshold value [24,140–144] despite that results of correlation network analysis are often sensitive to the threshold value [141–147]. In a related vein, if a single threshold value is applied to the correlation matrix obtained from different participants in an experiment, which is typical in neuroimaging data analysis and referred to as an absolute threshold [142,148], the edge density can vary greatly across participants. Since edge density is heavily correlated with many network measures, this can be seen as introducing a confound into subsequent analyses and casts doubt on consequent conclusions, e.g., that sick participants tend to have less small-world brain networks than healthy controls. (In this example, a network with a large edge density would in general yield a small average path length and large clustering coefficient, leading to the small-world property, so that density differences alone could have driven the observed effect.) An alternative method for setting the threshold is the so-called proportional thresholding, with which one keeps a fixed fraction of the strongest (i.e., most correlated) edges to create a network, separately for each participant [142,148]; also see [149–151] for early studies. In this manner, the thresholded networks for different participants have the same density of edges. However, while the proportional thresholding may sound reasonable, it has its own problems [148]. First, because different participants have different magnitudes of overall correlation coefficient values, the proportional threshold implies that one includes relatively weakly correlated node pairs as edges for participants with an overall low correlation coefficients. This procedure increases the probability of including relatively spurious node pairs, which can be regarded as type I errors (i.e., false positives), increasing noise in the resulting network. (Also see [152,153] for discussion on this matter.) Second, the overall correlation strength is often predictive of, for example, a disease in question. The proportional threshold enforces the same edge density for the different participants' networks. Therefore, it gives up the possibility of using the edge density, which is a simplest network index, to account for the group difference. If one uses the absolute threshold, the edge density is different among participants, and one can use it to characterize participants. The edge density in the proportional thresholding is also an arbitrary parameter.

Second, apart from false positives due to keeping small-correlation node pairs as edges, correlation networks at least in its original form suffer from false positives because pairwise correlation does not differentiate between direct effects (i.e., nodes i and j are correlated because they directly interact) and indirect effects (i.e., nodes i and j are correlated

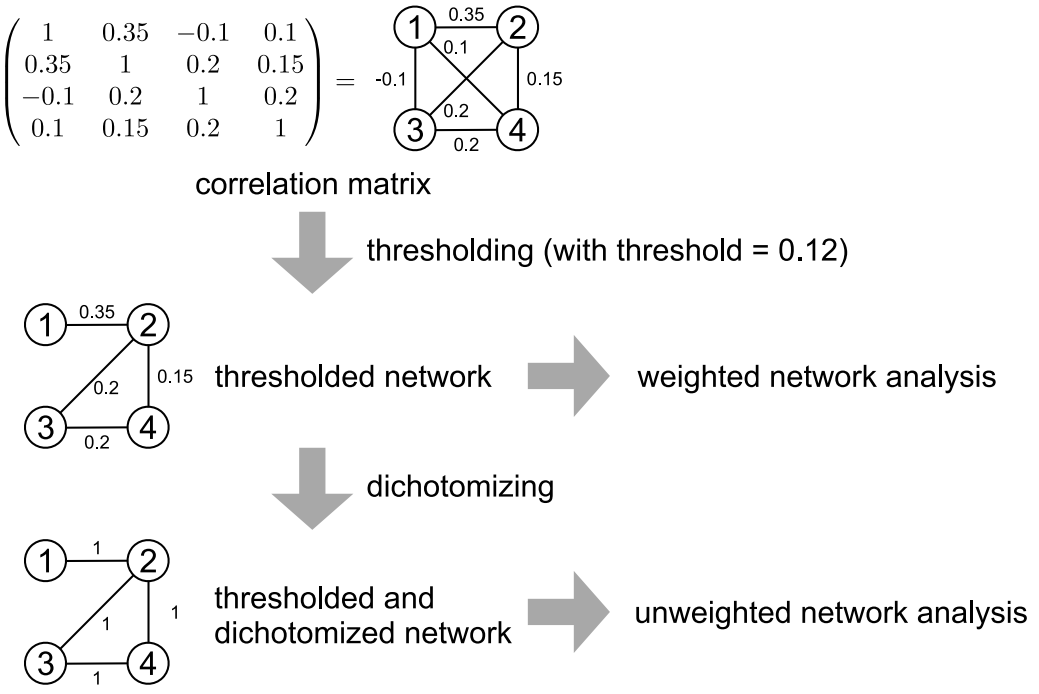

Fig. 2. Thresholding a correlation matrix. We set the threshold at $\theta = 0.12$. If we only threshold the correlation matrix, we obtain a weighted network. If we further dichotomize the thresholded matrix, we obtain an unweighted network. A different θ yields a different network in general.

because i and k interact and j and k interact). In other words, correlations are transitive. The correlation coefficient is lower-bounded by [154]

$$\rho_{ij} \geq \rho_{ik}\rho_{jk} - \sqrt{(1-\rho_{ik})^2(1-\rho_{jk})^2}. \tag{9}$$

Eq. (9) implies that if ρ_{ik} and ρ_{jk} are large, i.e., sufficiently close to 1, then ρ_{ij} is positive. Furthermore, this lower bound of ρ_{ij} is usually not tight, suggesting that ρ_{ij} tends to be more positive than what Eq. (9) suggests when $\rho_{ik}, \rho_{jk} > 0$ [155,156]. This false positive problem is the main motivation behind the definition of the partial correlation networks and related methods with which to remove such a third-party effect, i.e., influence of node k in Eq. (9). (See Section 3.6.) Instead, one may want to suppress false positives by carefully choosing a threshold value. Let us consider the absolute thresholding. For example, if i and j do not directly interact, i and k do, j and k also do, yielding $\rho_{ij} = 0.4$, $\rho_{ik} = 0.7$ and $\rho_{jk} = 0.6$, then setting $\theta = 0.5$ enables us to remove the indirect effect by k. However, it may be the case that i' and j' do not directly interact, i' and k' do, j' and k' also do, yielding $\rho_{i'j'} = 0.2$, $\rho_{i'k'} = 0.3$, and $\rho_{j'k'} = 0.4$. Then, thresholding with $\theta = 0.5$ dismisses direct as well as indirect interactions (that is, it introduces false negatives). A related artifact introduced by the combination of thresholding and indirect effects is that thresholding tends to inflate the abundance of triangles, as measured by the clustering coefficient for dichotomized (and therefore unweighted) networks, and other short cycles [87,156]; even correlation networks generated by dichotomizing randomly and independently generated data $\{x_{i\ell}\}$ have high clustering coefficients [156]. This phenomenon resembles the fact that spatial networks tend to have high clustering just because the network is spatially embedded [157,158].

Third, whereas thresholding has been suggested to be able to mitigate uncertainty on weak links (including the case of the proportional thresholding to some extent) and enhance interpretability of the graph-theoretical results (e.g., [24]), thresholding in fact discards the information contained in the values of the correlation coefficient. For example, in Fig. 2, thresholding turns a correlation of -0.1 and 0.1 into the absence of an edge. Furthermore, if we dichotomize the edges that have survived thresholding, a correlation of 0.2 and 0.35 are both turned into the presence of an edge.

There are various methods to try to mitigate some of these problems. In the remainder of this section, we cover methods related to dichotomizing.

One family of solutions is to integrate network analysis results obtained with different threshold values [24] (but see [159] for a critical discussion). For example, one can calculate a network index, such as the node's degree, denoted by α, as a function of the threshold value, θ, and fit a functional form to the obtained function $\alpha(\theta)$ to characterize the node [146]. Similarly, one can calculate α for a range of θ values and take an average of α [26,160]. In the case of group-to-group comparison, an option along this line of idea is the functional data analysis (FDA), with which one looks at α as a function of θ across a range of θ values and statistically test the difference between the obtained function for different

groups by a nonparametric permutation test [161,162]. In these methods, how to choose the range of θ is a nontrivial question.

A different strategy is to determine the threshold value according to an optimization criterion. For example, a method was proposed [143] for determining the threshold value as a solution of the optimization of the trade-off between the efficiency of the network [163] and the density of edges. Another method to set θ is to use the highest possible threshold that guarantees all or most (e.g., 99%) of nodes are connected [164].

The so-called maximal spanning tree is an easy and classical method to automatically set the threshold by guaranteeing that the network is connected [100,107], while at the same time avoiding the creation of edges that would form loops (and are therefore unnecessary for connectedness). One adds the largest correlation node pairs as edges one by one under the condition that the generated network is the tree. In the end, the maximal spanning tree contains all the N nodes, and the number of edges is $N-1$. Thanks to the mapping from (large) sample correlation coefficients to (small) Euclidean distances established by Eq. (6), the maximal (in the sense of correlation) spanning tree is sometimes called the *minimum* (in the sense of distance) spanning tree [100,107] (MST). The MST can be viewed as the graph achieving the overall minimum length among all graphs that make the N nodes reachable from one another. Here, the minimum length is defined as the sum of the lengths of its realized edges, and the length of an edge is the metric distance between its endpoints. The maximal spanning tree also allows a hierarchical tree representation, which facilitates interpretation [107,165,166]. However, the generated network is extreme in the sense that it is a most sparse network among all the connected networks on N nodes, without any triangles. A variant of the maximum spanning tree is to sequentially add edges with the largest correlation value under the constraint that the generated network can be embedded on a surface of a prescribed genus value (roughly speaking, the given number of holes) without edge intersection [167]. If the genus is constrained to be zero, the resulting network is a planar graph, called the planar maximally filtered graph (PMFG). The PMFG contains $3N-6$ edges. The PMFG contains more information than the maximum spanning tree (which is in any case contained in the PMFG), such as some cycles and their statistics. Note that these methods effectively produce a threshold for the correlation value to be retained, but on the other hand preserve only some of the edges that exceed the threshold. Indeed, they are (designed to be) irreducible to the mere identification of an overall threshold value, with their merit residing in the introduction of higher-order geometric constraints guiding the dichotomization procedure.

The maximal spanning forest (MSF) is similar to the maximal spanning tree but extracts different information from correlation matrix data [168]. In the MSF, each node has out-degree 1 with the out-going edge from node v pointing to the node that is the most strongly correlated with v. A node may then have no in-coming edge (i.e., in-degree equal to 0) or in-degree larger than 1. The MSF is a directed network with N (directed) edges and without cycles of length greater than two (reciprocated edges may exist where two nodes are most strongly correlated with each other). The MSF is in general not (even weakly) connected, and the different "chain-like" components may be of practical interest [168]. In contrast, the MST is composed of a single connected component. Indeed, if one adds edges to connect the components of the MSF into a single component by considering all node pairs between different components and sequentially adding the highest-correlation edges only when the added edge connects two components into one, the undirected version of this single connected component is an MST [168].

Another method related to the maximum spanning tree is to use the k nearest neighbor graph of the correlation matrix, in which each ith node is connected at least to the k nodes with the highest correlation with the ith node [153]. Yet another choice, which is designed for general weighted networks, is the disparity filter, with which one preserves only statistically significant edges to generate the *network backbone* [169,170]. Note that, with these methods as well, some lower-correlation node pairs are retained as edges and some higher-correlation edges are discarded.

Application example: Wang et al. compared functional networks of the brain between children with attention-deficit hyperactivity disorder (ADHD) and healthy controls using fMRI data [171]. The brain scan lasted for eight minutes in the resting state, and the brain image was acquired every two seconds. The authors used a previously published popular brain atlas defined by three-dimensional coordinates of 90 anatomical regions of interest (45 per hemisphere), each of which defines a node of the network. After various steps of preprocessing the original data, they computed the Pearson correlation coefficient between the time series at each pair of nodes, separately for each child. Aware of the problem inherent in choosing a threshold value, which we discussed earlier in this section, the authors examined dichotomized networks for a range of threshold values $\theta \in [0.05, 0.5]$. As downstream analysis, they measured the small-world-ness of networks in terms of the so-called global efficiency and local efficiency [163]. A large global efficiency value and a small local efficiency value suggest that the network has a small-world property. They found that, while the networks of both ADHD and healthy control children were small-world, those of children with ADHD were somewhat less small-world than those of the controls across a wide range of θ; the difference was statistically significant for the local efficiency in some range of θ.

3.3. Persistent homology

In the previous section, we discussed the idea of integrating analysis of dichotomized networks over different threshold values to mitigate the effect of selecting a single threshold value. Topological data analysis, or more specifically, persistent homology, provides a systematic and mathematically founded suite of methods to do such analyses. In topological analysis in general, one focuses on properties of shapes that do not change under continuous deformations of them, such as

stretching and bending. Such topologically invariant properties include the numbers of connected components and of holes, which can be calculated through the so-called homology group. Persistent homology captures the changes in the network structure over multiple scales, or over a range of threshold values, clarifying topological features that are robust with respect to the threshold choice. In this broader perspective, networks are only a particular instance of the type of topological object under major consideration in topological data analysis, called simplicial complexes, because networks only consider pairwise interactions. One may want to consider the clique complex, in which each k-clique (i.e., complete subgraph with k nodes) in the network is defined to be a higher-order object called a $(k-1)$-simplex and belongs to the simplicial complex of the given data. Note that the clique complex contains all the edges of the original network as well because an edge is a 2-clique by definition. For reviews of topological data analysis including persistent homology, see [172–176].

To analyze correlation matrix data using persistent homology, we start with a point cloud, with each point corresponding to a node in a correlation network. Then, we introduce a distance between each pair of nodes, d_{ij}, where $1 \leq i, j \leq N$. By setting a threshold value θ', we obtain a dichotomized network, or, e.g., a clique complex, depending on the choice, denoted by $G_{\theta'}$. The simplicial complexes with varying θ' values forms a filtration, i.e., with the nestedness property

$$G_{\theta'_1} \subseteq G_{\theta'_2} \subseteq G_{\theta'_3} \subseteq \cdots, \text{ for } \theta'_1 \leq \theta'_2 \leq \theta'_3 \leq \cdots, \tag{10}$$

with the inclusion relationship in Eq. (10) referring to that of the edge set and that of any higher-order simplex object. The collection of clique complexes in Eq. (10) is called the Vietoris–Rips filtration. Simply put, with a larger threshold on the distance between nodes, the generated simplicial complex has more edges (and higher-order simplexes in the case of the clique complex). In practice, it suffices to consider the sequence of threshold values $\{\theta'_1, \theta'_2, \ldots\}$, such that $G_{\theta'_{m+1}}$, contains just one more edge (and some higher-order simplexes containing that edge) than $G_{\theta'_m}$; if there are multiple node pairs with exactly the same distance value, the corresponding multiple edges not in $G_{\theta'_m}$ simultaneously appear in $G_{\theta'_{m+1}}$. If θ' is taken to vary over all possible topologies, the simplicial complex $G_{\theta'_1}$ is composed of N isolated nodes, while the last one in the nested sequence is the complete graph (that is, precisely, the corresponding clique complex) of N nodes.

A large correlation should correspond to a small d_{ij}. One can realize this by setting $d_{ij} = f(\rho_{ij}^{\text{sam}})$, where f is a monotonically decreasing function. Then, the network interpretation of Eq. (10) simply states the nested relationship in which the edges existing in a dichotomized network are included in dichotomized networks with smaller thresholds. However, while this practice is common, d_{ij} is not mathematically guaranteed to be a distance metric, typically violating the triangle inequality if we use an arbitrary monotonically decreasing function f. Therefore, one often uses f that makes d_{ij} a distance metric, such as Eq. (6) or variants thereof. Then, the ensuing topological data analysis of correlation networks is underpinned by stronger mathematical foundations.

The next step is to calculate for each $G_{\theta'}$ the homology groups or associated quantities, such as the kth Betti number. The zeroth and first Betti numbers are the numbers of connected components and essentially different cycles, respectively. These and other topological features of $G_{\theta'}$ depend on θ'. For example, each connected component and cycle may appear (through closing loops) and disappear (through mergers with others) as one gradually increases the distance threshold θ' from $\theta' = 0$, at which all the nodes are isolated. One can precisely visualize the occurrence of the birth and death events of each component by the persistence barcodes or persistence diagrams. For example, the persistence diagram represents each connected component, cycles, two-dimensional voids, etc. as a point (x, y) in the two-dimensional space in which x represents the θ' value at which, e.g., a cycle appears and y ($\geq x$) represents the θ' value at which the same cycle disappears. If $y - x$ is large, that particular feature of the data is robust over changing scales because it is independent of the specific value of θ' within a relatively large range.

We are often motivated to compare different correlation matrices or networks, such as in the comparison of functional brain networks between the patient group and control group. Quantitative comparison between two persistent diagrams provides a threshold-free method to effectively compare dichotomized networks. A persistence diagram consists of a set of two-dimensional points (x, y). The bottleneck and Wasserstein distance, between persistence diagrams P_1 and P_2, which are commonly used, first consider the best matching between (x, y) in P_1 and that in P_2. For the obtained best matching, both distance metrics measure the distance between (x, y) in P_1 and that in P_2 in the Euclidean space and tally up the distance over all the points in different manners. For instance, the bottleneck distance is given by $d(P_1, P_2) = \inf_\gamma \sup_{(x,y)\in P_1} \|(x, y) - \gamma((x, y))\|_\infty$, where γ is a matching between P_1 and P_2.

Persistent homology has been applied to correlation networks in neuronal activity data [141,144,177,178], gene co-expression data [174,179,180], financial data [181,182], and to co-occurrence networks of characters in literature work and in academic collaboration [183]. Scalability of persistent homology algorithms remains a concern. However, it may be of less concern for correlation network analysis because the number of nodes allowed for correlation networks is typically limited by the length of the data, not by the speed of algorithms (see Section 3.1).

3.4. Weighted networks

A strategy for avoiding the arbitrariness in the choice of the threshold value and loss of information in dichotomizing is to use weighted networks, retaining the pairwise correlation value as the edge weight [73,140]. Although there are numerous settings in network science where negative edge weights are considered, they are generally more difficult to treat. (See Section 3.5.) As such, two common methods to create positively weighted networks are (1) using the absolute

value of the correlation coefficient as the edge weight and (2) ignoring negatively weighted edges and only using the positively weighted edges. Both methods dismiss some information contained in the original correlation matrix, i.e., the sign of the correlation or the magnitude of the negative pairwise correlation. Nonetheless, these transformations are widely used because many methods are available for analyzing general positively weighted networks, many of which are extensions of the corresponding methods for unweighted networks. One can also use methods that are specifically designed for weighted networks [184].

It should be noted that weighted networks share the problem of false positives due to indirect interaction between nodes with the unweighted networks created by dichotomization. We also note that, in contrast to thresholding (which may be followed by dichotomization), node pairs with any small correlation (i.e., correlation coefficient close to 0) are kept as edges in the case of the weighted network. This may increase the uncertainty of the generated network and hence of the subsequent network analysis results.

Thresholding operations in statistics literature to increase the sparsity of the estimated covariance matrix often produce weighted networks. This is in contrast to the dichotomization, which produces unweighted networks. Hard thresholding in statistics literature refers to coercing $C_{ij}^{\rm sam}$, with $i \neq j$, to 0 if $\left|C_{ij}^{\rm sam}\right| < \theta$ and keep the original $C_{ij}^{\rm sam}$ if $\left|C_{ij}^{\rm sam}\right| \geq \theta$ [127,185–187]. Soft thresholding [185,187,188] transforms $C_{ij}^{\rm sam}$ by a continuous non-decreasing function of $C_{ij}^{\rm sam}$, denoted by $f(C_{ij}^{\rm sam})$, such that

$$f(x) = \begin{cases} x-\theta & \text{if } x \geq \theta, \\ 0 & \text{if } -\theta < x < \theta, \\ x+\theta & \text{if } x \leq -\theta. \end{cases} \tag{11}$$

This assumption implies that, in contrast to hard thresholding, there is no discontinuous jump in the transformed edge weight at $C_{ij}^{\rm sam} = \pm\theta$. Both hard and soft thresholding, as well as a more generalized class of thresholding function $f(x)$ [187], do not imply dichotomization and therefore generate weighted networks. In numerical simulations, all these thresholding methods to generate weighted networks outperformed the sample covariance matrix in estimating true sparse covariance matrices [187]. The same study also found that there was no clear winner between hard or soft thresholding, while combination of them tended to perform somewhat better than other types of thresholding.

Adaptive thresholding refers to using threshold values that depend on (i,j). An example is to use Eq. (11), but with an (i,j)-dependent threshold value, denoted by θ_{ij}, in place of θ, with $\theta_{ij} = \overline{c}\sqrt{{\rm Var}_{ij} \cdot \log N/L}$, where $\overline{c}$ is a constant, and ${\rm Var}_{ij}$ is an estimate of the variance of $(X_i - \mu_i)(X_j - \mu_j)$ [189]. This adaptive thresholding theoretically converges faster to the real covariance matrix and performs better numerically than universal thresholding schemes (i.e., using a threshold value independent of (i,j)).

Tapering estimators also threshold and suppress $C_{ij}^{\rm sam}$ in an (i,j)-dependent manner. A tapering estimator for the (i,j) entry of the covariance matrix is given by $\overline{f}(i,j) \cdot \frac{L-1}{L} C_{ij}^{\rm sam}$, where $\overline{f}(i,j) = 1$ if $|i-j| \leq k/2$, $\overline{f}(i,j) = 2 - 2|i-j|/k$ if $k/2 < |i-j| < k$, and $\overline{f}(i,j) = 0$ if $|i-j| \geq k$. Here, tapering parameter k is an even integer, and $\frac{L-1}{L} C_{ij}^{\rm sam}$ is the maximum likelihood estimator of C_{ij}. The tapering estimator more strongly suppresses the (i,j) entries that are farther from the diagonal respecting the sparsity of the estimated covariance matrix. This estimator is optimal in terms of the rate of convergence to the true covariance matrix, and the value of k realizing the optimal rate differs between the two major matrix norms with which the estimation error is measured [190].

Application example: Chen et al. compared gene co-expression networks between people with schizophrenia and non-schizophrenic controls [191]. They used the Weighted Correlation Network Analysis (WGCNA) software (see Section 6), which is frequently used in gene co-expression network analysis, to construct weighted networks of genes. In short, it uses $\left|\rho_{ij}\right|^{\beta}$ as the weight of edge (i,j), where β is a parameter. Then, they compared gene modules, or communities of the network, between the schizophrenia and control groups. The module detection was carried out by an algorithm implemented in WGCNA. There was no significant difference between the schizophrenia and control groups in terms of the module structure (i.e., which genes are in each module). However, the eigengene — that is, the first principal component of the co-expression matrix within a module — of each of two modules was significantly associated with schizophrenia compared to the control. The hub genes of each of these two modules, *NOTCH2* and *MT1X*, were the largest contributor to the respective eigengenes. The authors further carried out biological analyses of these two genes to clarify where their expressions are upregulated and the functions in which these genes may be involved. The results were similar for a biopolar disorder data set.

3.5. Negative weights

Correlation matrices have negative entries in general. In the case of both unweighted and weighted correlation networks, we often prohibit negative edges either by coercing negative entries of the correlation matrix to zero or by taking the absolute value of the pairwise correlation before transforming the correlation matrix into a network. We prohibit negative edges for two main reasons. First, in some research areas, it is often difficult to interpret negative edges. In the case of multivariate financial time series, a negative edge implies that the price of the two assets tend to move in the opposite manner, which is not difficult to interpret. In contrast, when fMRI time series from two brain regions are negatively correlated, it does not necessarily imply that these regions are connected by inhibitory synapses, and it is not

straightforward to interpret negative edges in brain dynamics data [24,192]. Second, compared to weighted networks and directed networks, we do not have many established tools for analyzing networks in which positive and negative edges are mixed, i.e., signed networks. Signed network analysis is still emerging [193]

In fact, negative edges may provide useful information. For example, they benefit community detection because, while many positive edges should be within a community, negative edges might ideally connect different communities rather than lie within a community. Some community detection algorithms for signed networks exploit this idea [193,194]. Another strategy for analyzing signed network data is to separately analyze the network composed of positive edges and that composed of negative edges and then combine the information obtained from the two analyses. For example, the modularity, an objective function to be maximized for community detection, can be separately defined for the positive network and the negative network originating from a single signed network and then combined to define a composite modularity to be maximized [140,195]. While these methods are designed for general signed networks, they have been applied to brain correlation networks [140,194].

Another type of approach to signed weighted networks is nonparametric weighted stochastic block models [196,197], which are useful for modeling correlation matrix data. Crucially, this method separately estimates the unweighted network structure and the weight of each edge but in a unified Bayesian framework. By imposing a maximum-entropy principle with a fixed mean and variance on the edge weight, they assumed a normal distribution for the signed edge weight. Because the edge weight in the case of correlation matrices, i.e., the correlation coefficient, is confined between -1 and 1, an ad-hoc transformation to map $(-1, 1)$ to $(-\infty, \infty)$ such as $y = 2\,\mathrm{arctanh}\, x = \ln \frac{1+x}{1-x}$ is applied before fitting the model. One can assess the goodness of such an ad-hoc transformation by a posteriori comparison with different forms of functions to transform x to y using Bayesian model selection [197]. In this way, this stochastic block model can handle negative correlation values. With this method, one can determine community structure (i.e., blocks) including its number and hierarchical structure.

3.6. Partial correlation

A natural method with which to avoid false positives due to indirect interaction effects in the Pearson correlation matrix is to use the partial correlation coefficient (as in, e.g., [198–200]). This entails measuring the linear correlation between nodes i and j after partialing out the effect of the other $N-2$ nodes. Specifically, to calculate the partial correlation between nodes i and j, we first compute the linear regression of X_i on $\{X_1, \ldots, X_N\} \setminus \{X_i, X_j\}$, which we write as $X_i \approx \sum_{m=1;m\neq i,j}^{N} \beta_{i,m} X_m$, where $\beta_{i,m}$ is the coefficient of linear regression. Similarly, we regress X_j on $\{X_1, \ldots, X_N\} \setminus \{X_i, X_j\}$, which we write as $X_j \approx \sum_{m=1;m\neq i,j}^{N} \beta_{j,m} X_m$. The residuals for L samples are given by $\varepsilon_{i,\ell} = x_{i\ell} - \sum_{m=1;m\neq i,j}^{N} \beta_{i,m} x_{m\ell}$ and $\varepsilon_{j,\ell} = x_{j\ell} - \sum_{m=1;m\neq i,j}^{N} \beta_{j,m} x_{m\ell}$, where $\ell \in \{1, \ldots, L\}$. The partial correlation coefficient, denoted by $\overline{\rho}_{ij}^{\mathrm{par}}$, is the Pearson correlation coefficient between $\{\varepsilon_{i,1}, \ldots, \varepsilon_{i,L}\}$ and $\{\varepsilon_{j,1}, \ldots, \varepsilon_{j,L}\}$.

In fact, the partial correlation coefficient between i and j ($j \neq i$) is given by

$$\overline{\rho}_{ij}^{\mathrm{par}} = -\frac{\Omega_{ij}}{\sqrt{\Omega_{ii}\Omega_{jj}}}, \tag{12}$$

where $\Omega = C^{-1}$ is the precision matrix [133,134]. Eq. (12) implies that $\Omega_{ij} = 0$ is equivalent to the lack of partial correlation, i.e., $\overline{\rho}_{ij}^{\mathrm{par}} = 0$. This conditional independence property gives an interpretation of the precision matrix, Ω. Eq. (12) also implies that the partial correlation can be calculated only when C is of full rank, whose necessary but not sufficient condition is $L \geq N$. If C is rank-deficient, a natural estimator of the $N \times N$ partial correlation matrix $\overline{\rho}^{\mathrm{par}} = (\overline{\rho}_{ij}^{\mathrm{par}})$ is a pseudoinverse of C. However, the standard Moore–Penrose pseudoinverse is known to be a suboptimal estimator in terms of approximation error [201,202], while the pseudoinverse is useful for screening for hubs in partial correlation networks [202]. If C is of full rank, Ω as well as C is positive definite. Therefore, although Eq. (12) only holds true for $i \neq j$, if we denote the matrix defined by the right-hand side of Eq. (12) including the diagonal entries by $\tilde{\rho}$, then the diagonal entries of $\tilde{\rho}$ are equal to -1, and $\tilde{\rho}$ is negative definite. We can verify this by rewriting Eq. (12) as $\tilde{\rho} = -D^{-1/2}\Omega D^{-1/2}$, where $D = \mathrm{diag}(\Omega_{11}, \ldots, \Omega_{NN})$ is the diagonal matrix whose diagonal entries are Ω_{11}, …, Ω_{NN}. If we consider matrix $\overline{\rho}^{\mathrm{par}} = 2I + \tilde{\rho}$, where I is the identity matrix, as a partial correlation matrix to force its diagonal entries to 1 instead of -1, the eigenvalues of $\overline{\rho}^{\mathrm{par}}$ are upper-bounded by 2 [203].

By thresholding the partial correlation matrix, or using an alternative method, one obtains an unweighted or weighted partial correlation network, depending on whether we further dichotomize the thresholded matrix. Because the partial correlation avoids the indirect interaction affect, the network created from random partial correlation matrices yields, for example, smaller clustering coefficients [156] than if we had used the Pearson correlation matrix.

While it apparently sounds reasonable to partial out the effect of the other nodes to determine a pairwise correlation between two nodes, it is not straightforward to determine when the partial correlation matrix is better than the Pearson correlation one. First, Eq. (12) implies that extreme eigenvalues of $\overline{\rho}^{\mathrm{par}}$ are those of a normalized precision matrix. Because the precision matrix is the inverse of the covariance matrix C, extreme eigenvalues of $\overline{\rho}^{\mathrm{partial}}$ are derived from eigenvalues of C with small magnitudes. It is empirically known for, e.g., financial time series, that small-magnitude eigenvalues of the covariance matrices are buried in noise, i.e., not distinguishable from eigenvalues of random matrices [101,102], as we discuss later in Section 4. Therefore, the dominant eigenvalue of the precision matrix is strongly affected by noise [204].

Second, the entries of $\overline{\rho}^{\mathrm{par}}$ are more variable than those of Pearson correlation matrices. Specifically, if $(x_1, \ldots, x_N)$ obeys a multivariate normal distribution, the Fisher-transformed partial correlation of a sample partial correlation, i.e.,

$$z_{ij} = \frac{1}{2} \ln \left(\frac{1 + \overline{\rho}_{ij}^{\mathrm{par,sam}}}{1 - \overline{\rho}_{ij}^{\mathrm{par,sam}}} \right), \tag{13}$$

where $\overline{\rho}_{ij}^{\mathrm{par,sam}}$ is the sample partial correlation calculated through Eq. (12) with $\Omega = (C^{\mathrm{sam}})^{-1}$, approximately obeys the normal distribution with mean $\frac{1}{2} \ln \left(\frac{1+\overline{\rho}_{ij}^{\mathrm{par}}}{1-\overline{\rho}_{ij}^{\mathrm{par}}} \right)$ and standard deviation $[L-3-(N-2)]^{-1/2}$. This result dates back to Fisher (see e.g., [205]). In contrast, the corresponding result for the Fisher transformation of the Pearson correlation coefficient is that the transformed variable approximately obeys the normal distribution with mean $\frac{1}{2} \log \left(\frac{1+\rho_{ij}^{\mathrm{sam}}}{1-\rho_{ij}^{\mathrm{sam}}} \right)$ and standard deviation $(L-3)^{-1/2}$ [205]. Therefore, the partial correlation has more sampling variance than the Pearson correlation unless $L \gg N$.

Third, partial correlation matrices typically have more negative entries and smaller-magnitude entries than Pearson correlation matrices [205,206]. Combined with the larger variation of the sample partial correlation than the sample Pearson correlation discussed just above, the tendency that $\overline{\rho}_{ij}^{\mathrm{par}}$ has a smaller magnitude than ρ_{ij} poses a challenge of statistically validating the estimated partial correlation networks [205].

Studies in neuroscience have compared partial correlation networks with simple correlation networks and/or with the corresponding underlying structural networks [32,62,199,207–210] (see Section 2.2). Some studies found that the similarity between partial correlation networks and structural networks is higher than that between Pearson correlation networks and structural networks [62,211].

Application example: Wang, Xie, and Stanley analyzed correlation networks composed of stock market indices from 2005 to 2014 from 57 countries [212], widely covering the continents of the world, with each country corresponding to a node. They computed Pearson and partial correlation coefficients for the time series of the logarithmic returns, given by Eq. (1), between each pair of countries, converted the correlation coefficient value into a distance (see Eq. (6)), and constructed the minimal spanning trees, called the MST-Pearson and MST-Partial networks. These networks appeared to be scale-free (i.e., with a power-law-like degree distribution) trees. Among other things, they compared clusters and top centrality nodes (i.e., countries) between the MST-Pearson and MST-Partial. They observed that the results from the MST-Partial networks are more reasonable than those from the MST-Pearson construction in light of our general understanding of world economics.

3.7. Graphical lasso and variants

Estimating a true correlation matrix, which contains $N(N-1)/2$ unknowns, is an ill-founded problem unless the number of samples is sufficiently larger than $N(N-1)/2$. A strategy to overcome this problem is to impose sparsity of the estimated correlation network. A sparsity constraint enforces zeros on a majority of matrix entries to suppress the number of unknowns to be estimated relative to the number of samples. Imposing sparsity on estimated correlation networks is a major form of covariance selection. *Structural learning* refers to estimation of an unknown network from data and usually assumes that the given data obey a multivariate normal distribution and that the estimated network is sparse. For reviews with tutorials and examples on this topic, see [25,213,214].

The Gaussian graphical model assumes that the precision matrix from the data obeys a multivariate normal distribution and usually imposes sparsity of the precision matrix [134]. In addition to reducing the number of unknowns to be estimated, a motivation behind estimating a sparse precision matrix is that $\Omega_{ij} = 0$ is equivalent to the absence of conditional linear dependence of the signals at the ith and jth nodes given all the other $N-2$ variables, which is easy to interpret. The graphical lasso is an algorithm for learning the structure of a Gaussian graphical model [215–220]. The graphical lasso maximizes the likelihood of the multivariate normal distribution under a lasso penalty (i.e., ℓ_1 penalty), whose simplest version is of the form $\lambda \sum_{i,j=1}^{N} |\Omega_{ij}|$, where we recall that Ω_{ij} is the (i,j) entry of the precision matrix, and λ is a positive constant. This penalty term is added to the negative log likelihood to be minimized. If λ is large, it strongly penalizes positive $|\Omega_{ij}|$, and the minimization of the objective function yields many zeros of the estimated Ω_{ij}. The (i,j) pairs for which the estimated Ω_{ij} is nonzero form edges of the network; if there is no edge between i and j, they are conditionally independent, i.e., conditioned on the other $N-2$ variables. This conditional independence is also referred to as the pairwise Markov property because the distribution of X_i only depends on $\{X_j : (i,j)$ is an edge of the network$\}$. The pairwise Markov property is a special case of the global Markov property. The global Markov property dictates that $\{X_i : i \in A\}$ and $\{X_i : i \in B\}$ are conditionally independent given $\{X_i : i \in S\}$ if S is a cutset of A and B, that is, any path in the graph connecting a node in A and a node in B passes through S. The pairwise and global Markov properties are major cases of the Markov random field, which is defined by a set of random variables having Markov property specified by an undirected graph [134,221–224]. The network is sparse by design. One can extend the lasso penalty function in multiple ways, for example, by allowing λ to depend on (i,j) and automatically determining λ using an information criterion. Other ways to regularize the number of nonzero elements in the precision matrix than lasso penalty are also possible. (See e.g., [225–227].)

A neighborhood selection method is another algorithm to estimate a sparse Gaussian graphical model [228]. With this method, one first carries out lasso regression for each ith variable (i.e., node) to determine a tentative small set of i's neighbors. Second, if j is a tentative neighbor of i, and i is a tentative neighbor of j, then they are connected by an undirected edge (i, j). The method named "space" (Sparse PArtial Correlation Estimation) advances the aforementioned neighborhood selection method by proposing to minimize a single composite penalized loss function, not separately minimizing the penalized loss function for each variable [229]. The loss function of "space" is a weighted sum of the regression error (i.e., error in estimating X_i in terms of a linear combination of other X_j's) over all i's plus the usual ℓ_1 penalty term. These regression-based methods as well the graphical lasso algorithms permit the case in which the number of observations, L, is smaller than the number of variables, N.

Bayesian variants of graphical lasso provide the level of uncertainty in the estimated model. One such Bayesian approach assumes that we know node pairs that are not adjacent to each other, which is equivalent to imposing $\Omega_{ij} = 0$ for a given set of node pairs (i, j) [230]. Such a situation is possible when both the correlation matrix and structural network are available, as is common in MRI experiments in the brain. The Bayesian method uses the G-Wishart distribution, which is the Wishart distribution constrained by $\Omega_{ij} = 0$ for non-adjacent node pairs in the given graph (i.e., structural network) G, as the prior distribution of the precision matrix, Ω [231,232]. If G is the complete graph, then the G-Wishart distribution is the Wishart distribution. Then, it uses data $(x_{i\ell}) \in \mathbb{R}^{N\times L}$ to update the distribution using Bayes' rule, obtaining the posterior distribution of Ω, which is again a G-Wishart distribution but with updated parameter values. In other words, the G-Wishart distribution is a conjugate prior, giving us good intuitive pictures of how the prior distribution is transformed to the posterior distribution and mitigating the computational burden of the Bayesian method when N is not small.

Gaussian graphical models assume normality. One approach to relax this assumption is to use a so-called nonparanormal distribution [233]. The idea is to assume that the transformed random variables $(f_1(X_1), \ldots, f_N(X_N))$ obey a multivariate normal distribution, where $f_1, \ldots, f_N$ are monotone and differentiable functions. In this case, we say that the original variables $(X_1, \ldots, X_N)$ have a nonparanormal distribution or a Gaussian copula (given by $f_1, \ldots, f_N$, and the mean vector and the covariance matrix of the normal distribution after the transformation). The nonparanormal distribution is nonidentifiable. For example, if we replace $f_1(X_1)$ by a new function $\tilde{f}_1(X_1) \equiv 2f_1(X_1)$ and appropriately scale the first row and column of the covariance matrix and the first entry of mean vector used for the multivariate normal distribution that the transformed N variables obey, then one gets the same distribution of $(X_1, \ldots, X_N)$. Therefore, we further impose that the transformations $f_1, \ldots, f_N$ conserve the mean and variance of the original $X_1, \ldots, X_N$. Then, as in the case of Gaussian graphical models, $\Omega_{ij} = 0$ implies that X_i and X_j are conditionally independent of each other given all the other $N - 2$ variables, in addition to that $f(X_i)$ and $f(X_j)$ are conditionally independent of each other. There are methods to estimate from data the functions $f_1, \ldots, f_N$ as well as a sparse covariance matrix Ω, assuming a lasso penalty. Naturally, the nonparanormal distribution outperforms the graphical lasso when the true distribution of the data is not multivariate normal [233]. Later work proposed algorithms for estimating the nonparanormal distribution with optimal convergence rate [234,235]; the idea behind the improved algorithms is to avoid explicitly calculating $f_1, \ldots, f_N$ and to exploit statistics of rank correlation coefficient estimators. There are also other methods for estimating non-normal graphical models without relying on copula [236,237].

Extreme cases of non-Gaussian distribution of the random variables are discrete multivariate distributions, in particular when each X_i is binary (i.e., $\in \{-1, 1\}$.). In this case, the form of the distribution of $(X_1, \ldots, X_N)$ proportional to $e^{\boldsymbol{x}\Omega\boldsymbol{x}^\top}$, where $\boldsymbol{x} = (X_1 - \mu_1, \ldots, X_N - \mu_N)$, defines the Ising model. Note that the same form of distribution for continuous variables is the multivariate normal distribution, which we discussed above for graphical lasso. There are many methods for inferring the Ising model from observed multivariate binary data, which is the task referred to as the inverse Ising model and Boltzmann machine learning [238–242]. Although exact likelihood maximization is available, it is practical only for small N. Therefore, various methods aim to overcome this limitation by allowing some approximation. Similar to graphical models, inference algorithms for the Ising model respecting the sparsity of the underlying network have also been investigated. For example, ℓ_1-regularized methods based on pseudo-likelihood maximization were developed for estimating a sparse Ising Markov random field, or a graph in which the absence of the edge signifies the conditional independence between two nodes [243–245]. Decimation, or to recursively set the small edge weights to zero, combined with a stopping criterion and other heuristics, improves the performance of pseudo-likelihood maximization [246].

An alternative to the graphical lasso is to estimate sparse covariance matrices rather than sparse precision matrices under lasso penalty [247–252]. With this approach, the consequence of imposing sparsity, $C_{ij} = 0$, corresponds to marginal independence between X_i and X_j. Similar to the case of the graphical lasso, one regards (i, j) pairs for which the estimated C_{ij} is nonzero as edges of the network.

Most graphical lasso models and their variants do not model the estimation problem relative to a null model correlation matrix. However, by estimating a sparse correlation matrix that is different relative to a null model of correlation matrix (see Section 4 for null models), it was found that the estimated correlation matrix gives a better description of the given financial correlation matrix data than the graphical lasso and that the choice of the null model also affects the performance [252]. By construction, this method infers a set of edges that are not expected from the so-called correlation matrix configuration model (see Section 4 for details).

3.8. Statistical significance of correlation edges

A test of significance of an edge, run on each edge, may sound like a natural way to filter a network. However, this idea is not easily feasible because multiple comparisons with $N(N-1)/2$ estimates, each of whose significance would have to be tested, is not practical given that $N(N-1)/2$ is usually large [25]. If we require a significance level of 0.05, then Bonferroni correction for multiple comparisons implies that the edge statistic has to be tested to be significant with the p value less than $0.05/[N(N-1)/2]$ for the edge to be actually significant at the 0.05 significance level. Unless N is small, this condition is usually too harsh. Furthermore, the different edges are correlated with each other, particularly if they share a node. In contrast, the Bonferroni correction assumes that the different tests are independent. Extensions of the Bonferroni corrections, such as the Šidák and Holm corrections, do not resolve these issues.

The false discovery rate approach, more specifically, the Benjamini–Yekutieli procedure [253], provides a solution to these problems. This test is less stringent than family-wise error rate controls including the Bonferroni correction and allows dependency between different tests. To run this procedure to generate a correlation network, we first calculate the p value for each node pair (i,j). Second, we arrange all the p values in ascending order, which we denote by $p_{(1)}, p_{(2)}, \ldots, p_{(N(N-1)/2)}$. Then, we reject the null hypothesis (i.e., regard that the edge with the corresponding p value is significant) for $p_{(1)}, \ldots, p_{(M)}$, where

$$M = \max\left\{ i : p_{(i)} < \frac{0.05i}{\frac{N(N-1)}{2}\sum_{i'=1}^{N(N-1)/2}\frac{1}{i'}} \right\}. \tag{14}$$

This procedure is a thresholding method with an automatically set threshold value implied by the number of edges, M, determined by Eq. (14). We note that the type of correlation matrix (e.g., Pearson or partial correlation) does not matter once p values have been obtained.

In the above procedure, what does it mean to calculate a p value for a node pair? The answer varies. If we are given just one correlation matrix, which we assume in most of this article, the p value can be that of the Pearson correlation coefficient in a standard textbook, if the Pearson correlation matrix is given as input [254] (see [255–257] for different calculations of the p value when a single matrix is given as input). If we are given multiple correlation matrices forming one group, the one-sample t-test provides a p value for each (i,j) pair, quantifying whether the (i,j) correlation for the group is different from 0 [199]. If correlation matrices from two groups need to be compared, the two-sample t-test provides a p value for each (i,j). In this case, the generated network will be a difference network, in which the edges represent node pairs for which the correlation is significantly different between the two groups.

In neuroimaging research, the network-based statistic (NBS) is popularly used for controlling for the same multiple comparison problem [258]. In the NBS, one first calculates the p value for each (i,j), as in the Benjamini–Yekutieli procedure. Then, we keep (i,j) whose p values are smaller than an arbitrary threshold. If we just use those (i,j) pairs to form a network, we would suffer from the aforementioned problems (i.e., multiple comparisons and dependencies between edges). Instead, the NBS focuses on the connected components induced by the surviving edges, which are small in number in general, and tests the significance of the sizes (i.e., the number of nodes) of the connected components using a permutation test. The NBS has been extended to a threshold-free method [259]. NBS and many of its extensions are applicable to correlation matrix data beyond neuroscience. Note that the NBS is not a method to estimate a correlation network; it tactically avoids network estimation and the problem of multiple comparisons, while providing a statistically controlled downstream network analysis (i.e., test on the size of connected components). In this sense, the NBS casts a key question: why do we need to estimate a network first of all? We will discuss this topic in Section 7.1.

Covariance selection methods, such as the graphical lasso, do not explicitly test the significance of individual edges. The edges that survive these types of filtering methods should be regarded to be sufficiently strong to be included in the model [25].

3.9. Temporal correlation networks

Many empirical networks vary over time, including temporal correlation networks [260], and many methods have been developed for analyzing time-varying network data [110,260–262]. If the given data is a multivariate time series that is non-stationary, then correlation matrices computed from the first 10% of the time points may be drastically different from that computed from the last 10%. So, there is the possibility of greater adaptability and better generalizability when one uses a time series of correlation networks rather than just one. One can then apply various temporal network analysis tools to the obtained temporal correlation networks.

A simple method to create dynamic correlation networks from multivariate time series data is sliding-window correlation [263] (also called rolling-window correlation in the finance literature; see e.g. [264]). With this method, one considers time windows within the entire observation time horizon, $t = \{1, \ldots, t_{\max}\}$. These time windows may be overlapping or non-overlapping. Then, within each time window, one calculates the correlation matrix and network. If there are 100 time windows within $[1, t_{\max}]$, then this method creates a temporal network of length 100. Reliably computing a single correlation matrix and a static correlation network from multivariate time series requires a reasonable length (i.e., the number of time points) of a time window. Generation of a reliable dynamic correlation network requires

longer data because one needs a multitude of such reasonably long time windows. A limitation of sliding-window correlations is that they are susceptible to large variability if the size of the time window is small, whereas a large window size sacrifices sensitivity to temporal changes [265].

Early seminal reports analyzed temporal correlation networks of stock markets by tracking financial indices and central nodes of the static correlation network over more than a decade [166,266,267]. However, methods of dynamic correlation networks have been particularly developed in brain network analysis. In neuroimaging studies, in particular in fMRI studies, dynamic correlation networks are known as dynamic (or time-varying) functional connectivity networks [263, 268–271]. Temporal changes in functional (i.e., correlational) brain networks may represent neural signaling, behavioral performance, or changes in the cognitive state, for example. Patterns of time-varying functional networks may alter under a disease. One can also analyze stability of community structure of temporal correlation networks over time [254,272,273]. (See also [103–106,110] for temporal stability analyses of community detection in financial correlation networks.) Many variations of the method, such as on how to create sliding windows and how to cluster time windows to define discrete-state transition dynamics, and change-point detection are available in the field (e.g., see [271]). Many of these methods should be applicable to multivariate time series data to generate temporal correlation networks in other domains.

There are also methods for estimating dynamic precision matrices, proposed outside neuroscience [274–276]. For example, the time-varying graphical lasso (TVGL) formulates the inference of discrete-time time-varying precision matrix as a convex optimization problem [276]. The objective function is composed of maximization of the log likelihood under a lasso sparsity constraint and the constraint that the precision matrix at adjacent time points does not change too drastically, enforcing the temporal consistency.

4. Null models and random matrices

In the analysis of (correlation) networks, a good practice to verify any structural or dynamical measurement α is to compare the value of α in the given network with the value of α achieved in random networks. This allows one to determine whether the α value measured for the given network data is explainable purely by the gross structural properties (e.g. edge density, degree distributions) of the random graph family (when the α value is similar between the given and random networks) or it is the result of other distinctive features of the data (when the α value is statistically different between the given and random networks). Many discoveries in network science owe to the fact that key analyses have prudently implemented this practice by using or inventing appropriate null models of networks. Already in one of the earliest seminal papers on small-world networks, Watts and Strogatz compared the average path length and the clustering coefficient of empirical networks with those of the Erdős–Rényi random graph having the same number of nodes and edges [277].

In this section, we report on similar concepts and results for correlation matrices. The idea is to introduce various null hypotheses that would imply certain properties for the sample correlation matrix, and compare the empirical matrix with the null model in order to extract statistically significant properties. This discussion will lead us to consider on one hand models defined in analogy with null models for networks, and on the other hand genuine models for correlation matrices derived from RMT. Several papers have noted the need for proper null models specifically for correlation networks [33,103,156,278,279]. We should use correlation networks derived from a random correlation matrix as a null model. We stress that a random correlation matrix is different from a random network model (e.g., Erdős–Rényi model), because of the dependencies between entries. Similarly, many classes of random matrices are not appropriate null models for correlation networks, either. For example, a symmetric matrix whose all on-diagonal entries are 1 and off-diagonal entries are i.i.d. uniformly on $(-1, 1)$ is almost never a correlation matrix unless N is small [280]. In this section, we introduce several null models of correlation matrices. All the null models presented give distributions over correlation matrices. Then, using any method introduced in previous sections (e.g., by thresholding in various ways), one can define corresponding null models over correlation networks.

4.1. Models based on shuffling

A straightforward and traditional null model consists in shuffling the original data, $\{x_{i\ell}\}$, based on which the correlation matrix is calculated. This method is especially typical for multivariate time series data. In the simplest case, one randomizes all entries independently within each time series, thereby destroying all the cross-correlations while preserving the original values for each time series separately.

As a more constrained option, one preserves the power spectrum of the time series at each node while the time series is otherwise randomized [156,279,281,282]. More specifically, one Fourier transforms the time series at the ith node, randomize the phase, and carry out the inverse Fourier transform. Then, for the randomized multivariate time series, one calculates the correlation matrix, which is used as control.

Another method that preserves the full autocorrelation structure within each single time series, while randomizing cross-correlations among the N time series, has been proposed in [283,284]. The method is called the rotational random shuffling (RRS) model because it first imposes periodic boundary conditions (i.e., 'gluing' the last timestep to the first one) to turn each time series into a 'ring', and then randomly rotates the N rings with respect to each other while keeping each ring internally intact.

One can impose additional constraints on the randomization of time series depending on the properties other than the power spectrum that one wants to have the null model preserve [285].

4.2. Models inspired by network analysis

Other null models are explicitly inspired by null models that are routinely used for networks. For instance, the H–Q–S algorithm, invented by Hirschberger, Qi, and Steuer [286] is an equivalent of the Erdős–Rényi random graph in general network analysis. Specifically, given the covariance matrix, C^{sam}, the H–Q–S algorithm generates random covariance matrices, C^{HQS}, under the following constraints. First, the expectation of each on-diagonal entry of C^{HQS} is equal to the average of the N on-diagonal entries of C^{sam}, denoted by $\mu_{\mathrm{on}} \equiv (1/N) \times \sum_{i=1}^{N} C_{ii}^{\mathrm{sam}}$. Second, the expectation and variance of each off-diagonal entry of C^{HQS} are equal to those of C^{sam} calculated on the basis of all the off-diagonal entries, denoted by μ_{off} and σ_{off}^2, respectively. Optionally, one can also constrain the variance of the on-diagonal entries of C^{HQS} [286] or use a fine-tuned heuristic variant of the algorithm [156]. To implement the most basic H-Q-S algorithm without constraining the variance of the on-diagonal entries of C^{HQS}, we set

$$L^{\mathrm{HQS}} \equiv \max\left(2, \lfloor \left(\mu_{\mathrm{on}}^2 - \mu_{\mathrm{off}}^2\right) / \sigma_{\mathrm{off}}^2 \rfloor\right), \tag{15}$$

where $\lfloor \cdot \rfloor$ is the largest integer that is not greater than the argument. Then, we draw $N \times L^{\mathrm{HQS}}$ variables, denoted by $x_{i\ell}$ (with $i \in \{1, \ldots, N\}$ and $\ell \in \{1, \ldots, L^{\mathrm{HQS}}\}$), independently from the identical normal distribution with mean $\sqrt{\mu_{\mathrm{off}}/L^{\mathrm{HQS}}}$ and variance $-\mu_{\mathrm{off}}/L^{\mathrm{HQS}} + \sqrt{\mu_{\mathrm{off}}^2/(L^{\mathrm{HQS}})^2 + \sigma_{\mathrm{off}}^2/L^{\mathrm{HQS}}}$. Then, the H–Q–S algorithm sets the covariance matrix by

$$C_{ij}^{\mathrm{HQS}} = \sum_{\ell=1}^{L^{\mathrm{HQS}}} x_{i\ell} x_{j\ell} \quad i, j \in \{1, \ldots, N\}. \tag{16}$$

It is known that $\langle C^{\mathrm{HQS}} \rangle_{ij} = \delta_{ij}\mu_{\mathrm{on}} + (1-\delta_{ij})\mu_{\mathrm{off}}$. Therefore, the expectation of the correlation matrix, ρ^{HQS}, is approximately given by $\langle \rho^{\mathrm{HQS}} \rangle_{ij} = \delta_{ij} + (1-\delta_{ij})\mu_{\mathrm{off}}/\mu_{\mathrm{on}}$. This highlights the completely homogeneous nature of the null model. For instance, while the degree of empirical correlation networks is usually heterogeneously distributed [108], this property is not captured by the H–Q–S algorithm [287].

One of the most popular heterogeneous null models for networks is the configuration model, i.e., a uniform ensemble of networks under the constraint that the degree of each node is conserved [279,288], either exactly or in expectation. By comparing given networks against a configuration model, one can reliably quantify and discuss various network properties such as network motifs [289], community structure [290], rich clubs [291], and core–periphery structure [292]. The rationale behind the use of the configuration model is that the node's degree is heterogeneously distributed for many empirical networks and that one generally wants to explore structural, dynamical, or other properties of networks that are not immediate outcomes of the heterogeneous degree distribution. One can extend the configuration model by imposing additional constraints that the network under discussion is supposed to satisfy, such as spatiality, i.e., the constraint that the nodes are embedded in a metric space and the probability of an edge depends on the distance between the two nodes [293]. See [288,294,295] for reviews of configuration models and best practices for generating random realizations from such models.

We should similarly test properties found for given correlation networks against appropriate null models. However, the usual configuration models are not appropriate as null models of correlation networks because they are significantly different from correlation networks derived from purely random data [87,103,110,156]. The expectation $\langle A_{ij} \rangle$ of the (i, j) entry of the adjacency matrix of the configuration model conditioned on the degrees of all nodes, at least in the idealized and unrealistic regime of weak heterogeneity of the degrees [294,295], is equal to

$$\langle A_{ij} \rangle = \frac{k_i k_j}{N\bar{k}}, \tag{17}$$

where $k_i = \sum_{j=1}^{N} A_{ij}$ is the degree of the ith node in the original network, A_{ij} is the entry of the *empirical* adjacency matrix of the original network (i.e., $A_{ij} = 1$ if there is an edge between the ith and jth nodes, and $A_{ij} = 0$ otherwise), and $\bar{k}$ is the average degree over all nodes. The above expected value also represents the probability of independently connecting the ith and jth nodes in realizations of the configuration model for networks. Note that, one can indeed realize graphs in the configuration model by sampling edges independently (at least if the constraint on the degree is 'soft', i.e. realized only as an ensemble average over realizations [295]). However, correlation matrices cannot be generated with independent entries, even in the null model of independent signals. This is because, even under the null hypothesis of independent realizations of the original time series, the correlation matrix constructed from such time series still obeys the 'metric' (or triangular) inequality in (9). We will elaborate more on this point below.

To see what Eq. (17) yields by merely replacing an empirical adjacency matrix with an empirical correlation matrix ρ_{ij}, we proceed as follows [103]. We assume that each empirical signal is standardized in advance such that $\mathrm{Var}(X_i) = 1$, $\forall i \in \{1, \ldots, N\}$. In this way, we do not need to distinguish between the correlation (i.e., ρ_{ij}) and covariance (i.e., C_{ij}) matrices. We express the 'degree' as

$$k_i = \sum_{j=1}^{N} \rho_{ij} = \sum_{j=1}^{N} C_{ij} = \sum_{j=1}^{N} \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_i, X_{\mathrm{tot}}), \tag{18}$$

where Cov represents the covariance and $X_{\rm tot} = \sum_{i=1}^{N} X_i$ is the 'total' signal. Then, we obtain

$$N\bar{k} = \sum_{i=1}^{N} k_i = \mathrm{Cov}(X_{\rm tot}, X_{\rm tot}) = \mathrm{Var}(X_{\rm tot}), \tag{19}$$

where $\mathrm{Var}(X_{\rm tot})$ is the variance of $X_{\rm tot}$. By inserting the above quantities into (17) with A_{ij} replaced by ρ_{ij}, we obtain for the expected correlation matrix

$$\langle \rho_{ij} \rangle = \frac{\mathrm{Cov}(X_i, X_{\rm tot})\mathrm{Cov}(X_j, X_{\rm tot})}{\mathrm{Var}(X_{\rm tot})} = \frac{\mathrm{Cov}(X_i, X_{\rm tot})}{\sqrt{\mathrm{Var}(X_{\rm tot})}\sqrt{\mathrm{Var}(X_i)}} \cdot \frac{\mathrm{Cov}(X_j, X_{\rm tot})}{\sqrt{\mathrm{Var}(X_{\rm tot})}\sqrt{\mathrm{Var}(X_j)}} = \rho(X_i, X_{\rm tot})\rho(X_j, X_{\rm tot}). \tag{20}$$

Technically, this expected matrix is a covariance matrix because it is symmetric, rank 1, and the only nonzero eigenvalue is positive [103,296]. To interpret the meaning of the above expression for $\langle \rho_{ij} \rangle$, we recall the definition of the conditional three-way partial Pearson correlation coefficient [3,133]:

$$\rho(X_i, X_j \mid X_{\rm tot}) = \frac{\rho(X_i, X_j) - \rho(X_i, X_{\rm tot})\rho(X_j, X_{\rm tot})}{\sqrt{1 - \rho(X_i, X_{\rm tot})^2}\sqrt{1 - \rho(X_j, X_{\rm tot})^2}}. \tag{21}$$

We therefore conclude that the expected correlation matrix in (20) is a correlation matrix of N signals that satisfy the conditional independence relationship

$$\rho(X_i, X_j \mid X_{\rm tot}) = 0 \tag{22}$$

$\forall i, j(\neq i) \in \{1, \ldots, N\}$ [109,110].

However, when one generates a correlation network from the configuration model, i.e. from a correlation matrix obeying (22), the generated network is far from a typical correlation network generated by random data, due to the triangular inequality mentioned above. To see an example of this, let us revisit the example briefly explained in Section 3.2. Let us consider purely random data in which we generate each sample $x_{i\ell}$, $i \in \{1, \ldots, N\}$, $\ell \in \{1, \ldots, L\}$ (where we recall that L is the number of samples for each node) as i.i.d. random numbers obeying a given distribution. Then, we calculate the sample correlation matrix for $\{x_{i\ell}\}$ and then a sample correlation network. This procedure immediately establishes a connection between null models for sample correlation matrices and RMT [7–9], some elements of which are discussed in Section 4.3. For a broad class of methods, including dichotomizing, the generated correlation network has a high clustering coefficient [156], precisely because of the inequality (9). Therefore, high clustering coefficients in correlation networks should not come as a surprise. In contrast, networks generated by the ordinary configuration model yield low clustering coefficients [297], disqualifying it as a null model for correlation networks [103]. If we use the usual configuration model as the null model, we would incorrectly conclude that a given correlation network has high clustering even if the network does not have particularly high clustering among correlation networks. The configuration model as null model also underperforms the simpler null model, called the uniform null model (which is analogous to the random regular graph and to the H–Q–S algorithm explained above) on benchmark problems of community detection when communities of different sizes are present in single networks and those communities are detected by modularity maximization [110].

A different configuration model, specifically designed for covariance matrices, can be defined as follows. This model preserves the expectation of each row sum excluding the diagonal entry, which is equivalent to each node's degree in the case of the adjacency matrix of a conventional network [278]. This algorithm, which we refer to as the correlation matrix configuration model, preserves the expectation of each diagonal entry of $C^{\rm sam}$, or the variance of each variable, and the expectation of each row sum excluding the diagonal entry, i.e., $\sum_{j=1;j\neq i}^{N} C_{ij}^{\rm sam}$ $\forall i$, corresponding to the degree of the ith node. Under these constraints, the correlation matrix configuration model uses the distribution of $x_{i\ell}$, determined from the maximum entropy principle. In fact, each $(x_{1\ell}, \ldots, x_{N\ell})^\top$, $\ell \in \{1, \ldots, L\}$, independently obeys an identical multivariate normal distribution whose mean is the zero matrix and covariance matrix is denoted by $C^{\rm cm}$. Therefore, the correlation matrix configuration model is the Wishart distribution $W_N(C, L)$ that we have introduced in Section 3.1, with $C = C^{\rm cm}$. The matrix $C^{\rm cm}$ is of the form $\left[(C^{\rm cm})^{-1}\right]_{ij} = -\left[\delta_{ij} \cdot 2\alpha_i + (1 - \delta_{ij})(\beta_i + \beta_j)\right]$, where δ_{ij} is the Kronecker delta symbol, and α_i and β_i are parameters to be determined. One determines the values of α_i and β_i using a gradient descent algorithm [278] or by reformulating the problem as a convex optimization problem and solving it [252].

4.3. Models based on random matrix theory

Another class of null models is based on powerful estimates provided by RMT for the expected spectral properties of a random sample correlation matrix, rather than for the expected matrix itself [7–9]. Note that the expected correlation matrix, under the null hypothesis of N independent signals, is the identity matrix whose entries we denote as $\rho_{ij}^{\rm MG1} = \delta_{ij}$ [103,110]. Equivalently, $\rho^{\rm MG1} = I$ where I is the $N \times N$ identity matrix. This null model corresponds to an expectation under white noise signals $\{x_{i\ell}\}$ that are independent for different nodes $i \in \{1, \ldots, N\}$ and samples $\ell \in \{1, \ldots, L\}$. However, the sample pairwise correlation $\rho_{ij}^{\rm sam}$ measured from the signals, even under the null hypothesis of independence, will be different from the identity matrix unless $L \to \infty$, assuming a finite N. To take into account the

effects of finite L, it is convenient to look at the distribution of the random sample correlation matrix. If we assume a standardized independent normal distribution for x_i $\forall i$, then ρ^{sam} obeys the Wishart distribution with $C = I$, i.e. $W_N(I, L)$. Note that the expectation of $W_N(I, L)$ is I. The ensemble of Wishart matrices is a well studied topic in RMT. It turns out that, in addition to the distribution $W_N(I, L)$ for the entire sample correlation matrix, one can accurately describe the limiting density of eigenvalues in the asymptotic limit $N \to \infty$ and $L \to \infty$, with $L/N \to Q > 1$. Note that $Q > 1$ is a necessary condition for the sample correlation matrix to be non-degenerate, as we already mentioned. The limiting eigenvalue density $p_Q(\lambda)$, known as the Marchenko–Pastur distribution, has the form

$$p_Q(\lambda) = \begin{cases} \frac{Q\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda} & \text{for } \lambda_- < \lambda < \lambda_+, \\ 0 & \text{otherwise,} \end{cases} \tag{23}$$

where

$$\lambda_\pm = \left(1 \pm \sqrt{Q^{-1}}\right)^2 \tag{24}$$

are the expected minimum (λ_-) and maximum (λ_+) eigenvalues. As $Q \to \infty$, which corresponds to an infinite number of samples per node, it holds true that $\lambda_\pm \to 1$. This result implies that all eigenvalues become unity, and it is because each sample correlation matrix becomes the identity matrix in that case, as we already mentioned; then all eigenvalues are equal to 1. The fact that Q is necessarily finite for empirical correlation matrices makes Eq. (23) particularly useful as a null model for the empirical eigenvalue distribution expected under the hypothesis of independence of the N observed time series. In particular, early studies of financial multivariate time series data found that only a small number of the largest eigenvalues of the empirical covariance matrix are found above λ_+ [101,102]. In other words, only a few leading eigenvalues and the associated eigenvectors outside the prediction of RMT are statistically significant relative to the null model. As we discuss below, the role of the largest empirical eigenvalue is however different from that of the next largest ones. Specifically, the former encodes a system-wide, overall positive correlation, while the latter represent 'mesoscopic' information arising from the presence of internally coherent *groups* of time series. Based on this interpretation, RMT has provided useful null models for covariance/correlation matrix data, in particular in financial time series data analysis; see [6] for a review. In neuroscience, RMT has been applied only more recently (e.g., for noise-cleaning in the estimation of correlation matrices [65] as we describe later in this section and for community detection [298] as we describe in Section 5.2) and less systematically. Nonetheless, it holds promise as a general and powerful tool for the analysis of large data in virtually any field, especially in the modern era of data science [9]. Also see [130] for a review of random correlation as opposed to covariance matrices.

Among many possible specific choices of null models for correlation matrices based on RMT, here we consider two models, which we denote by ρ^{MG2} and ρ^{MG3}, proposed in [103,298]. A given correlation matrix is symmetric and positive semidefinite and therefore can be decomposed as

$$\rho^{\mathrm{sam}} = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_{(i)} \boldsymbol{u}_{(i)}^\top, \tag{25}$$

where $\lambda_i (\geq 0)$ is the ith eigenvalue of ρ^{sam}, and $\boldsymbol{u}_{(i)}$ is the associated normalized right eigenvector. The null model correlation matrix ρ^{MG2} preserves the contribution of small eigenvalues to Eq. (25), which are regarded to be noisy and described by RMT, and is given by

$$\rho^{\mathrm{MG2}} = \sum_{i:\lambda_i \leq \lambda_+} \lambda_i \boldsymbol{u}_{(i)} \boldsymbol{u}_{(i)}^\top. \tag{26}$$

The boundary λ_+ originates from the Marchenko–Pastur distribution given by Eq. (23) and, as we mentioned, represents the expected largest eigenvalue *under the null hypothesis of independent signals*. Matrix ρ^{MG2} is not a correlation matrix because its diagonal elements are not equal to 1. However, this does not affect most network analyses because we usually ignore diagonal elements or self-loops in correlation networks. Matrix ρ^{MG2} represents a null model constructed only from the eigenvalues of the empirical correlation matrix that are deemed to be noisy. Therefore, comparing the empirical matrix against ρ^{MG2} singles out properties that cannot be traced back to noise [103]. Note that the difference

$$\rho^{\mathrm{sam}} - \rho^{\mathrm{MG2}} = \sum_{i:\lambda_i > \lambda_+} \lambda_i \boldsymbol{u}_{(i)} \boldsymbol{u}_{(i)}^\top \tag{27}$$

between the sample correlation matrix ρ^{sam} and the null model ρ^{MG2} is nothing but the sum of the dominant eigencomponents of ρ^{sam}. This matrix coincides with the output of the popular PCA technique, also called the eigenvalue clipping. The main difference between the generic use of that technique and the one based on ρ^{MG2} in Eq. (27) is the criterion (here based on RMT) for the selection of the number of principal components to retain.

By contrast, the matrix ρ^{MG3} also preserves the contribution of the largest eigenvalue in addition to that of the noisy eigenvalues [103–106]. The largest eigenvalue is useful to control for separately, if all the entries of the associated eigenvector are positive. It is because, in that case, it represents the effect of a common trend in the original time series,

giving an uninformative all-positive overall contribution (also called global mode, or market mode in the context of financial time series) to the sample correlation matrix. The matrix ρ^{MG3} is given by

$$\rho^{\mathrm{MG3}} = \lambda_{\max}\boldsymbol{u}_{(\max)}\boldsymbol{u}_{(\max)}^{\top} + \sum_{i:\lambda_i \le \tilde{\lambda}_+} \lambda_i \boldsymbol{u}_{(i)}\boldsymbol{u}_{(i)}^{\top}, \tag{28}$$

where max is the index for the dominant eigenvalue of ρ^{sam}, and λ_+ has been replaced by

$$\tilde{\lambda}_+ \equiv \left(1 - \frac{\lambda_{\max}}{N}\right) \lambda_+. \tag{29}$$

The above modification in the value of the threshold eigenvalue, with respect to the previous null model in Eq. (24), originates from the fact that sample correlation matrices always have trace equal to N, with all their diagonal entries being unity by construction. Therefore, the addition of a global mode represented by a large $\lambda_{\max}$ has the unavoidable effect of proportionally reducing all the other eigenvalues in such a way that the trace is preserved [101,298]. The value in Eq. (29) thus represents the expected largest eigenvalue *under the null hypothesis of independent signals plus a global mode*. Note that $\tilde{\lambda}_+ < \lambda_+$, which implies that any empirical eigenvalue λ_i smaller than λ_+ but larger than $\tilde{\lambda}_+$, i.e. $\tilde{\lambda}_+ < \lambda_i < \lambda_+$, is interpreted as noisy (hence discarded) under the null model ρ^{MG2} and as informative (hence retained) under the null model ρ^{MG3} [298]. Also note that all the entries of the dominant eigenvector, $\boldsymbol{u}_{(\max)}$, are positive, which is a necessary condition for the null model ρ^{MG3} to have a clear interpretation, if there is a sufficiently strong common trend affecting all the N signals. If this common trend is so strong that all the entries of the correlation matrix are positive, then the Perron–Frobenius theorem ensures the positivity of the dominant eigenvector. When this happens, the common trend obscures all the mutual correlations among the signals. Matrix ρ^{MG3} deliberately removes this global trend in addition to the noise, to reveal the underlying structure. Therefore, properties of correlation matrices or correlation networks that are not expected from ρ^{MG3} represent those not anticipated by the simultaneous presence of local noise and global trends. In other words, they reflect the presence of mesoscopic correlated structures such as correlation-induced communities [79,103–106,298]. As we will discuss in Section 5.2, one can indeed use matrix ρ^{MG3} to successfully detect communities of correlated signals.

We finally discuss another RMT-based model that, rather than representing a null model of the data, aims at providing the best fit, i.e., an optimal estimate, for the true (unobserved) correlation matrix ρ^{true}, starting from the sample correlation matrix ρ^{sam}. The model is called the optimal rotationally invariant estimator (RIE) for the correlation matrix [6,65]. Informally, the RIE estimator is the matrix ρ^{RIE} that, among the correlation matrices sharing the same eigenvectors with ρ^{sam}, achieves the minimum Hilbert–Schmidt distance $d_{\mathrm{HS}}(\rho, \rho^{\mathrm{true}}) = \mathrm{Tr}[(\rho - \rho^{\mathrm{true}})^2]$ from the true population matrix ρ^{true}. This task might seem impossible at first sight, because ρ^{true} is unobservable. However, it turns out that, for the minimization of $d_{\mathrm{HS}}(\rho, \rho^{\mathrm{true}})$ for large enough L, it is sufficient to know the spectral density $p^{\mathrm{true}}(\lambda)$ of ρ^{true}, which can be obtained from the spectral density $p^{\mathrm{sam}}(\lambda)$ of ρ^{sam}, thanks to a type of self-averaging property [6]. The final ingredient consists in modifying the eigenvalues $\{\lambda_i\}_{i=1}^N$ of ρ^{sam} to

$$\tilde{\lambda}_i \equiv \frac{\lambda_i}{|1 - Q^{-1} + Q^{-1} z_i s(z_i)|^2} \qquad i = 1, N \tag{30}$$

where $z_i \equiv \lambda_i - i\eta$ is the complexification of λ_i; η is a small parameter that should vanish in the $N \to \infty$ limit; $s(z) \equiv \mathrm{Tr}[(zI - \rho^{\mathrm{sam}})^{-1}]/N$ is the so-called Cauchy transform of the spectrum of ρ^{sam} [6]. A convenient choice for η in the case of finite N is $\eta = N^{-1/2}$ [6]. (However, see [65] and the application example below for an improved variant.) Using Eq. (30), we obtain the optimal RIE as

$$\rho^{\mathrm{RIE}} = \sum_{i=1}^{N} \tilde{\lambda}_i \boldsymbol{u}_{(i)}\boldsymbol{u}_{(i)}^{\top}. \tag{31}$$

For large L, the optimal RIE is expected to outperform, in terms of d_{HS}, any other estimator (including PCA, shrinkage, and other corrected sample estimators) that, like the RIE itself, modifies only the spectrum of ρ^{sam} and not its eigenvectors [6].

Application example. Ibàñez-Berganza et al. [65] compared several noise-cleaning methods of estimation of covariance and precision matrices from both human brain activity (fMRI) time series and synthetic data of size comparable with that typically encountered in neuroscience. They assessed the reliability of each method via the test-set likelihood and, in the case of synthetic data, via the distance from the true precision matrix. The methods considered include the eigenvalue clipping (or PCA; see above), linear shrinkage (see Section 3.1), graphical lasso (see Section 3.7), FA [3,4], early-stopping gradient ascent algorithms, the RMT-based optimal RIE given by Eq. (31), and a variant of the last one where the parameter η is optimized by cross-validation on a grid of values rather than being set to $\eta = N^{-1/2}$ (see above). Their cross-validated RIE outperformed all the other estimators in the severe undersampling regime (i.e., small Q) typical of fMRI time series, highlighting the power of RMT for the analysis of neuroscience data. Notably, the cross-validated RIE was the only method that improved upon the raw sample correlation matrix ρ^{sam} in the estimation of the true correlation matrix ρ^{true}, in all the simulated regimes and especially for strongly correlated synthetic data. They finally proposed a simple algorithm based on an iterative likelihood gradient ascent, leading to accurate estimations in weakly correlated synthetic data sets. A Python code implementing all the methods used in the paper is available [299].

5. Network-analysis inspired analysis directly applied to correlation matrices

As we already mentioned, a straightforward way to use null models for correlation matrices as controls of correlation networks is to generate correlation networks from the correlation matrix generated by the null model or its distribution. In this section, we showcase another usage of null models for correlation matrices, which is to conduct analysis inspired by network analysis but directly on correlation matrix data with the help of null model correlation matrices. Crucially, this scenario does not involve transformation of a given correlation matrix into a correlation network. To explain the idea, consider financial time series analysis using correlation matrices. Portfolio optimization and RMT directly applied to correlation matrix data are among powerful techniques to analyze such data [6,8,9]. These methods do not suffer from difficulties in transforming correlation matrices into correlation networks because they do not carry out such a transformation. In contrast, a motivation behind carrying out such a transformation is that one can then use various network analysis methods. A strategy to take advantage of both approaches is to adapt network analysis methods for conventional networks to the case of correlation matrix data.

5.1. Degree

Many empirical networks show heterogeneous degree distributions such as a power-law-like distribution [297,300–302]; such networks are called scale-free networks. The same holds true for the weighted degree of many networks [303]. Correlation networks are no exception, not much depending on how one constructs a network from correlation matrix data [266,278,304–306].

If we do not transform the given correlation matrix into a network, the node's weighted degree represents how the node's signal, X_i, is close to the signal averaged over all the nodes, X_{total}, as shown in Eq. (18). Previous research showed that the weighted degree calculated in this manner is heterogeneously distributed for some empirical data, while the right tail of the distribution is not as fat as typical degree distributions for conventional empirical networks [278]. The results are qualitatively similar when one calculates the weighted degree of the ith node as $\sum_{j=1;j\neq i}^{N}|C_{ij}|$ or $\sum_{j=1;j\neq i;C_{ij}>0}^{N}C_{ij}$. Therefore, heterogeneous degree distributions of the correlation network are not an artifact of the thresholding or other operations for creating networks from correlation data, at least to some extent.

5.2. Community detection

Community structure in networks is a partition of the nodes into (generally non-overlapping) groups that are internally well connected and sparsely connected across. One can detect communities with many different algorithms, and one popular family of methods, despite some shortcomings [147], is modularity maximization [290,307], which aims at placing a higher-than-expected number of edges connecting nodes within the same community. Modularity with a resolution parameter γ is defined by

$$Q=\frac{1}{N\bar{k}}\sum_{i,j=1}^{N}\left(A_{ij}-\gamma\frac{k_ik_j}{N\bar{k}}\right)\delta_{g_i,g_j}, \tag{32}$$

where we remind that A_{ij} is the entry of the empirical adjacency matrix, g_i is the community in which the ith node is placed by the current partition, and δ is again the Kronecker delta. Note the presence of the term $k_ik_j/N\bar{k}$ coming from Eq. (17), which signifies the use of the configuration model as standard null model in the ordinary modularity for networks. Approximate maximization of Q by varying $g_1,\ldots,g_N$ for a given network identifies its community structure.

Community detection, in particular modularity maximization, is desirable for correlation network data, too. Given that the original correlation matrix has both positive and negative entries in general, a possible variant of modularity maximization for correlation networks is to maximize a modularity designed for signed networks. The modularity for signed networks may be defined as a weighted difference of the modularity calculated for the positive network (i.e., the weighted network only containing positively weighted edges) and the modularity calculated for the negative network (i.e., the weighted network only containing negatively weighted edges with the edge weight being the absolute value of the correlation) [140]. However, this procedure assumes that, in the null model, edges can be thought as independent (as in the model described by Eq. (17)) and that positive edges can be randomized independently of negative edges. We have seen that both assumptions are clearly not valid for correlation matrices.

One can bypass the analogy with networks by directly computing and maximizing a modularity function that is appropriately defined for correlation matrices [103]. A viable redefinition of modularity for correlation matrices is given by

$$Q^{\text{cor}}=\frac{1}{\mathcal{N}}\sum_{i,j=1}^{N}\left(\rho_{ij}^{\text{sam}}-\langle\rho_{ij}\rangle\right)\delta_{g_i,g_j}, \tag{33}$$

where $\mathcal{N}=\sum_{i,j=1}^{N}\rho_{ij}^{\text{sam}}$ is a normalization constant, which is inessential to the maximization problem as long as it is positive. Matrix $\langle\rho\rangle$ should be a proper null model for the correlation matrix. Approximate maximization of Q^{cor} provides

optimal communities in correlation matrices. The crucial ingredient is the choice of the null model, $\langle \rho_{ij} \rangle$. Depending on what features of the original data one desires to preserve, the use of any of the models described in Section 4 is in principle legitimate, e.g., the white-noise (identity matrix) model $\rho_{ij}^{\mathrm{MG1}} = \delta_{ij}$, the noise-only model ρ_{ij}^{MG2}, the noise+global model ρ_{ij}^{MG3}, or the correlation matrix configuration model C_{ij}^{cm}. However, note that in the summation in Eq. (33) some terms must be positive and some must be negative, but at the same time not dominated by noise, in order to identify a nontrivial community structure, i.e., one different from a single community enclosing all nodes. For example, if all the entries of the empirical correlation matrix, ρ^{sam}, are positive, then the null model ρ^{MG1} will keep all terms in the summation in Eq. (33) non-negative, and the resulting optimal partition will be a single community [103]. By contrast, the use of ρ^{MG2} removes the noisy component of the sample correlation matrix and allows to detect noise-filtered communities, unless a global trend is present. If a global trend is present, then all the entries of the filtered matrix in Eq. (27) are positive, preventing communities from being detected. When all the terms in the summation in Eq. (33) are non-negative, the resulting optimal partition is a single community [103], incidentally showing the limitation of PCA for the community detection task. In presence of such a global trend, one obtains the best results by using ρ^{MG3}, which uncovers group-specific correlations [103–106]. Having maximized the modularity guarantees that the identified community structure is 'optimally contrasted', with necessarily positive overall residual correlations (with respect to the null model) inside each community and necessarily negative overall residual correlations across different communities. Modularity maximization using ρ^{MG3} has successfully revealed nontrivial community structure in time series of financial stock prices [103,104,106], credit default swaps [105], single-cell gene expression profiles [79], and neuronal activity [298]. The last example is expanded below.

Application example: Almog et al. [298] applied RMT-based community detection, defined via the maximization of the modularity given by Eq. (33), to the empirical correlation matrix obtained from single-neuron time series of gene expression in the biological clock of mice. The recording was made from the suprachiasmatic nucleus (SCN), located in the hypothalamus. The biological clock is a highly synchronized brain region that is yet adaptable, for example, to external light stimuli and their seasonal variations. Therefore, the research focus was on the identification of both positive (excitatory, phase-coherent) and negative (inhibitory, phase-opposing) interactions among constituent neurons. They showed that methods based on dichotomization using a global threshold, as well as 'naive' community detection methods using the ordinary network-based modularity (i.e., Eq. (32)), fail to identify groups of neurons that are internally positively correlated, and negatively correlated across. On the other hand, the maximization of the RMT-based modularity (i.e., Eq. (33)), with the null model given by $\langle \rho \rangle = \rho^{\mathrm{MG3}}$ in Eq. (28), successfully found community structure by filtering out both the neuron-specific noise and the system-wide dependencies that obfuscate the presence of underlying modules in the SCN. Their study uncovered two otherwise undetectable, negatively correlated populations of neurons (specifically, a spatially inner population and an outer one, both with left–right symmetry), whose relative size and mutual interaction strength were found to depend on the photoperiod [298]. In particular, the average residual intra-community correlation was significantly higher in short photoperiods (e.g., winter) than in long photoperiods (e.g., summer). In contrast, the residual inter-community correlation was lower in short photoperiods than in long photoperiods. A MATLAB package for the calculation of the null models used in the paper is available [308].

5.3. Clustering coefficient

Clustering coefficients measure the abundance of triangles in a network [297,300–302]. A dominant definition of clustering coefficient for unweighted networks, denoted by $\tilde{C}$, is given by [277]

$$\tilde{C} = \frac{1}{N} \sum_{i=1}^{N} \tilde{C}_i, \tag{34}$$

where $\tilde{C}_i$ is the local clustering coefficient at the ith node given by

$$\tilde{C}_i = \frac{(\text{number of triangles involving the } i\text{th node})}{k_i(k_i - 1)/2}. \tag{35}$$

The denominator of the right-hand side of Eq. (35) is a normalization constant to ensure that $0 \le \tilde{C}_i \le 1$.

One often measures clustering coefficients for both unweighted and weighted correlation networks. There are various definitions of weighted clustering coefficients [309,310]. One definition [305] is given by $\tilde{C}^{\mathrm{wei,Z}} = N^{-1} \sum_{i=1}^{N} \tilde{C}_i^{\mathrm{wei,Z}}$, where

$$\tilde{C}_i^{\mathrm{wei,Z}} = \frac{1}{\max_{i'j'} w_{i'j'}} \frac{\sum_{\substack{1 \le j,\ell \le N \\ j,\ell \ne i}} w_{ij} w_{i\ell} w_{j\ell}}{\sum_{\substack{1 \le j,\ell \le N \\ j,\ell \ne i; j \ne \ell}} w_{ij} w_{i\ell}}, \tag{36}$$

and w_{ij} $(= w_{ji} \ge 0)$ is the weight of edge (i, j).

Many empirical networks show large unweighted or weighted clustering coefficient values, and correlation networks are no exception. However, as we pointed out in Section 3.2, a high clustering coefficient of the correlation network is at least partly due to pseudo correlation.

Given this background, clustering coefficients for correlation matrices were proposed using a similar idea to the case of modularity directly defined for correlation matrices [311]. Because correlation matrices are naturally clustered if we dichotomize on the Pearson correlation matrix, the authors used the three-way partial correlation coefficient or partial mutual information to partial out the effect of a common neighbor of nodes j and ℓ (say i) to quantify partial connection between j and ℓ. In other words, we measure the connectivity between neighbors of i by, for example, the partial correlation coefficient $\rho(X_j, X_\ell \mid X_i)$, which we abbreviate as $\rho_{j\ell|i}$; the partial correlation coefficient is defined by Eq. (21). Because there is no clear notion of neighborhood for correlation matrices, we need to consider all triplets of different nodes, (i, j, ℓ). Then, as for the definition of the original clustering coefficient for networks, they took the average of $\rho_{j\ell|i}$ over the ith node's neighbors j and ℓ to define a local clustering coefficient for i. For example, we define a local clustering coefficient for node i as a weighted average by

$$C_i^{\mathrm{cor,A}} = \frac{\sum_{\substack{1\le j<\ell\le N \\ j,\ell\neq i}} \left|\rho_{ij}\rho_{i\ell}\rho_{j\ell|i}\right|}{\sum_{\substack{1\le j<\ell\le N \\ j,\ell\neq i}} \left|\rho_{ij}\rho_{i\ell}\right|}. \tag{37}$$

Finally, as in the case of the clustering coefficient for networks, we define the global clustering coefficient by $C^{\mathrm{cor,A}} = \sum_{i=1}^{N} C_i^{\mathrm{cor,A}}/N$. This method borrows the idea of clustering coefficient from complex network studies and tailors it for correlation matrix data. Clustering coefficients $C_i^{\mathrm{cor,A}}$ and $C^{\mathrm{cor,A}}$ already partial out the effect of pseudo correlation between X_j and X_ℓ due to X_i. However, we can still compare the observed clustering coefficient values against those for null models to validate whether or not the clustering coefficient values for the given data are significantly different from those for the null model [278].

Application example: Masuda et al. searched for possible association of $C_i^{\mathrm{cor,A}}$, $C^{\mathrm{cor,A}}$, and similar clustering coefficients for correlation matrices with the age of human participants in fMRI experiments [311]. They used publicly available resting-state fMRI data from the brains of healthy adults with a wide range of ages. The nodes were defined by a commonly used brain atlas consisting of $N = 30$ regions of interest. They found that the global clustering coefficients, such as $C^{\mathrm{cor,A}}$, declined with age. The correlation between the age and $C^{\mathrm{cor,A}}$ (and a variant of $C^{\mathrm{cor,A}}$) was stronger than that between the age and conventional clustering coefficients for general unweighted and weighted networks combined with both Pearson and partial correlation networks. Furthermore, the proposed local clustering coefficients were more strongly negatively correlated with age than the conventional clustering coefficients for general networks.

6. Software

In this section, we introduce freely available code useful for analyzing correlation networks. Obviously, one can apply software for analyzing general unweighted or weighted networks after thresholding (and optionally further dichotomizing) the given correlation matrix data. There are numerous packages for unweighted and weighted network analysis, which we do not mention here without a few exceptions.

In gene correlation network analysis, Weighted Correlation Network Analysis (WGCNA) is a popular freely available R software package [73]. WGCNA provides various outputs, such as community structure, weighted clustering coefficients, and visualization. WGCNA transforms the original edge weight, denoted by w_{ij} for edge (i, j), by a so-called soft thresholding transformation, i.e., by $\left|w_{ij}\right|^{\beta}$, where $\beta \geq 1$ is a parameter[5], such that one obtains an unsigned weighted network. Phiclust [79] is another community-detection tool for correlations from single-cell gene expression data, derived from RMT (i.e., Wishart ensemble as described in Section 4). It can be used to identify cell clusters with non-random substructure, possibly leading to the discovery of previously overlooked phenotypes. Phiclust is written in R and is freely available at Github [312] and Zenodo under GNU General Public License V3.0 [313].

The Brain Connectivity Toolbox is a MATLAB toolbox for analyzing networks [314]. It is also implemented in Python. Despite the name, many of the functions provided by the Brain Connectivity Toolbox can be applied to general network data, and many people outside neuroscience also use it. In relation to correlation networks, this toolbox is particularly good at handling weighted and signed networks, such as their degree, clustering coefficients, and community structure.

The graph-tool module in Python provides powerful network analysis and visualization functionality [315]. In relation to correlation networks, graph-tool is particularly strong at network inference based on stochastic block models.

The bootnet package in R can be used for estimating psychological networks using graphical lasso, with a unique feature of being able to assess the accuracy of the estimated network [25]. This package can be used for other types of correlation matrix data as well. Also see [25,316] for various R packages for sparse and related estimation of the precision and covariance matrices. For example, the qgraph package can also generate networks using the graphical lasso and visualize Pearson and partial correlation matrices [317], with beginner-friendly tutorials (e.g., [318]).

Graphical lasso is also implemented in Python, through the GraphicalLassoCV function in the scikit-learn package [319, 320]. The sklearn.covariance package also contains other functions for covariance estimation such as covariance shrinkage. Table 2 of [21] lists other code packages for estimating graphical models as well as different models.

[5] The soft thresholding here, coined in [73], is different from the same term defined in Section 3.4. It also does not belong to thresholding in the sense used in this article (defined in Section 3.2).

The Covariance Estimators package in Python, developed by Lucibello and Ibàñez-Berganza [299], contains several utilities for denoising empirical covariance matrices. In particular, it implements various variants of PCA, linear shrinkage, graphical lasso, FA, early-stopping gradient ascent algorithms, and the RMT-based optimal RIE proposed in [65] (see Section 4.3).

Papers [103,279] contain references to Python, MATLAB, and R codes for generating null models of correlation matrices discussed in Section 4. For example, the "spatiotemporal modeling tools for Python" contains functions to generate null model correlation matrices such as the H–Q–S model (named Zalesky matching model in their package) and methods through generating surrogate time series [285] (see Section 4.1). Another package in the list is the Scola, in Python, which generates the H–Q–S model and the correlation matrix configuration model [252] (see Section 4.2). Finally, a MATLAB package by MacMahon [308] implements the null models based on RMT, ρ^{MG2}, and ρ^{MG3}, derived in [103] from the Wishart ensemble (see Section 4.3).

7. Outlook

7.1. Recommended practices

We have reviewed various techniques to obtain and analyze networks generated from correlation matrix data, which naturally arise across many domains. Sections 2 and 3 emphasize for readers that there is not a single dominant method. We also highlighted good practices and pitfalls of individual methods. Naïve applications of network generation and analysis methods to correlation matrix data can easily yield flawed results. We should be careful about both how to generate correlation networks and how to analyze them. We recommend the following practices.

First, in resonance with previous reports, we explained that a simplistic dichotomizing, which is widely used, is problematic for multiple reasons (see Section 3.2). Therefore, if you threshold your correlation matrices to create networks, which may or may not be followed by dichotomization, do either of the following: (i) report your downstream network analysis results across a wide range of the threshold value; (ii) use a method designed to investigate a range of thresholds altogether (e.g., persistent homology); (iii) devise and use a network index that is robust with respect to the choice of the threshold value; and/or (iv) still use the simplistic thresholding but combine it with a proper null model. We showed an example of (i) at the end of Section 3.2. All of these practices are heuristic, but they are substantially better than simply using a single threshold value, reporting network analysis results, and skipping the examination of robustness.

That said, we do not have much knowledge about item (iii) regarding which network indices one should use. For example, the values of many network indices probably depend on the threshold value, which directly affects the number of edges [156]. However, in most cases, we are interested in comparing network analysis results between different cases, such as between disease and control groups, between empirical and randomized data, or between individuals of different ages. Then, the group difference or ranking among individuals may be robust enough when one varies the threshold value. Investigating robustness of correlation network analysis outcomes with respect to the threshold warrants more work.

To illustrate item (iv), we recall that correlation networks have high clustering no matter what data we use. However, a proper null model (e.g., shuffling of $\{x_{i\ell}\}$) will also produce dichotomized correlation networks with high clustering. Therefore, by comparing the results with those for the null model, one can avoid wrong conclusions such as that almost all empirical correlation networks have high clustering. We recall that null models for networks, most famously the configuration model, are not a proper null model for correlation networks because they generally do not originate from correlation matrices and do not generally match key properties of typical correlation matrices. See Section 4 for proper choices.

Our second, alternative recommendation is to resort to other methods, such as weighted correlation networks, graphical lasso (which is a partial correlation network method) and its variants, and covariance shrinkage, which avoid thresholding. Nonetheless, these methods usually require some arbitrary decisions by the users, such as setting hyperparameter values. Therefore, assessing robustness with respect to such choices remains important. For example, with weighted correlation networks without thresholding, one usually chooses between whether to force negative correlation values to zero, to keep them by taking the absolute value of the correlation, or to treat them as signed networks. Few papers have investigated different cases to check robustness of the subsequent network analysis results. Furthermore, because these different operations have been main options for a long time, it may be beneficial to pursue quantification of weighted networks that would provide results that are robust with respect to this methodological choice.

Our third recommendation is to avoid transforming correlation matrix data into networks but yet carry out downstream analysis analogous to network analysis. However, we recommend doing so only for properties whose definition does not make (implicit) assumptions that are violated by correlation matrices, such as the assumption of independent matrix entries that the ordinary modularity function in Eq. (32) implicitly makes. In presence of such unverified assumptions, we recommend either appropriately revising the definition of the property, e.g. as in the modified modularity in Eq. (33), or dismissing the property altogether. In Section 5, we showcased some such methods including two example analyses. This type of analysis is available for at least the degree, modularity maximization, and clustering coefficients. Then, one can evade thresholding or thoroughly examining various threshold values. Future work should generalize this approach to other structural properties and analysis methods formulated for networks. Examples include various node centrality measures, motifs, community detection methods apart from modularity maximization, rich clubs, fractals, and

simplicial complexes. In many cases, the configuration model is a standard choice of null model when constructing a network algorithm, such as a community detection algorithm. However, while we now have a reasonably long list of null models for correlation matrices (see Section 4), it is not known whether the configuration model for covariance matrices or a different null model described in Section 4 should be a standard choice when constructing an algorithm for correlation matrices inspired by network analysis. Answering this question needs systematic comparison studies of downstream analyses across various null models for correlation matrices.

7.2. Directions for future research

In this section, we identify some promising directions for future research.

Reproducibility of correlation networks arising from common approaches is a major practical issue, as has been pointed out in the literature in psychology and psychotherapy (e.g. [51,57]) and microbiome analysis [92]. In these research fields and others, it is often the case that only relatively few samples are available given the size of matrix and network, N, we wish to analyze. Especially if the number of samples, L, is smaller than or close to N, the partial correlation matrix and spectra of random matrices carry large variation, in part because they are inherently rank deficient. From the statistical inference point of view, it is not sound to try to infer many parameters, such as entries of the covariance matrix, from relatively few observations. The recognition of such phenomena led to the idea of covariance selection and various other methods. The amount of data needed for reliably estimating correlation networks, i.e., power analysis in statistics, should be further pursued for various correlation matrix data [25]. The development of methods to help practitioners use correlation networks better (e.g., by providing uncertainty quantification or clarifying the various noise trade-offs) can be transformational. Despite these challenges, there is a pressing need to understand complex systems of a very high dimension (i.e., $N \gg L$) with correlational data. One approach to this problem is to formulate estimation of large correlation networks as a computational rather than statistical challenge, as a problem to be solved under runtime and memory constraints, and to search feasible solutions in combination with machine learning [1]. How to reconcile the statistical and computational types of approach and deepen usage of machine learning and artificial intelligence techniques to correlation network analysis may be a beneficial research direction.

A key outstanding question is the treatment of low-correlation edges. On one hand, we have surveyed attempts to remove "noise" edges (for example by thresholding or graphical lasso), which is supposed to improve the overall signal-to-noise ratio of the graph representation. Sparser models are more parsimonious, easier to process quickly and with a lower memory footprint, and amenable to a range of network science analysis tools. On the other hand, one can argue that getting rid of low-correlation edges risks losing valuable information (see Section 3.2). In fact, it has been shown in the neuroscience context that the low-correlation edges alone can have substantial predictive power [161,321,322].

A related question is how to use *a priori* domain knowledge to choose appropriate preprocessing steps, such as what threshold to apply and whether or not to dichotomize. For example, dichotomizing may be more appropriate when the *a priori* belief is that nodes are either coordinating or not, with no appreciable difference in the degree of coordination when two nodes coordinate. As another example, one could use RMT on domain-relevant distributions to compare the dominant eigenvectors or eigenvalues before and after thresholding. This exercise may provide guidance on when thresholding is unlikely to have adverse effects.

Another viable alternative to the current focus on trying to recover and analyze the most accurate possible correlation matrix may be to treat the constructed correlation network as inherently uncertain and to regard it as a sample from a distribution of possible matrices as part of the analysis. For example, when assessing community structure, it may make sense to focus on structures that appear consistently across samples of correlation networks drawn from the estimated distribution, even when exact correlation values (perhaps especially the weaker correlations) considerably vary from sample to sample. Although there are established ways to do this for general networks [323], modeling of correlation networks and their substructures by their probability distributions is still a new idea [324] and needs further development. Such approaches may leverage existing work on null models for correlation networks, for example, as priors when forming a posterior distribution to sample from. On the other hand, some studies have documented the stability of the detected correlation-induced communities across time and their robustness under change of temporal resolution [103–106,298].

There are many multilayer network data sets, including multilayer correlation matrix data sets, and various data-analysis methods for them [110,325–329]. Examples include brain activity, where different layers of correlation matrices correspond to, for example, different frequency bands [330–333], or brain activity during different tasks [334]. In gene co-expression networks, different layers correspond to, for example, different levels of co-expression between gene pairs [335] or different tissue types such as different organs [336]. Overlaying different methods to construct correlation networks from one data set in each layer is another method to construct multilayer correlation matrices [337]. There are methods for analyzing multilayer correlation matrices and networks such as multilayer community detection algorithms [110,336]. However, methods that exploit the mathematical nature of correlation matrix data are still scarce and left for future work. Furthermore, multilayer networks are a major representation of temporal network data, where each layer corresponds to one time window [110,260,338]. Therefore, methods for analyzing multilayer correlation network data will also contribute to analysis of time-varying correlation network data.

Similar to other studies, we emphasize the potentially negative effects of thresholding, motivating our explanation of other methods for constructing correlation networks. However, thresholding also has positive effects such as reducing false

positives by discarding edges with small weights including the case of correlation networks. Such positive effects of thresholding may be manifested in multilayer data. For example, aggregating layers in a multilayer network and dichotomizing the aggregated edge weight can enhance detectability of multilayer communities compared to no dichotomizing under certain conditions [339,340]. Furthermore, some layers in a multilayer network may be more informative than other layers. While these arguments are for general multilayer networks, many of them may directly apply to multilayer correlation matrices.

For a given N, the set of covariance matrices constitutes the positive semidefinite cone, which is convex. Similarly, the set of full-rank correlation matrices, which is a strict subset of full-rank covariance matrices, is called the elliptope [131,132]. Positive semidefinite cones and elliptopes are manifolds and have their own geometric structure, which have been suggested to be useful for measuring the similarity between pairs of covariance or correlation matrices. Quantitative comparison of two covariance and correlation matrices is useful for various tasks such as fingerprinting of individuals, anomaly detection, and change-point detection in multivariate time series data. A straightforward way to measure the distance between two covariance/correlation matrices is to use a common matrix norm such as the Frobenius norm (i.e., $\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\rho_{ij}^{(1)}-\rho_{ij}^{(2)}\right|^2}$ in the case of correlation matrices, where $\rho^{(1)}$ and $\rho^{(2)}$ are two correlation matrices). However, research (in particular in neuroimaging studies) has suggested that geodesic distance measures respecting the structure of these manifolds is better at, for example, fingerprinting of individuals from fMRI data [341–343]. In these geodesic distance measures, one considers the tangent space at a given point x on the manifold, which corresponds to a correlation/covariance matrix. The so-called exponential map provides a one-to-one mapping from the straight line segment on the tangent space, which is essentially the Euclidean space, to the geodesic from x to y on the manifold. The logarithm map is the inverse of the exponential map. The geodesic distance between x and y is the length of the geodesic and has a practical matrix algebraic formula. Multiple reasonable definitions of such geodesic distances exist [342]. See [342,344,345] for mathematical expositions. Although these techniques are not for correlation networks but for matrices, they may potentially benefit understanding and algorithms for correlation networks. For example, can we understand null models of correlation matrices as a projection onto a submanifold of the entire elliptope? What are geometric meanings of thresholding, dichotomizing, and other operations to create correlation networks? Do we benefit by measuring distances between correlation networks rather than between correlation matrices?

We mentioned examples of microbiome and bibliometric co-occurrence networks as variants of correlation networks in sections 2.5 and 2.8, respectively. For example, let $x_{i\ell}=1$ if the ith researcher is an author of the ℓth paper in the database and $x_{i\ell}=0$ otherwise. Then, the number of papers that the ith and jth researchers have coauthored, which are co-occurrence events, is equal to $\sum_{\ell=1}^{L}x_{i\ell}x_{j\ell}$, where L is the number of papers in the entire data. This quantity is a non-standardized covariance, which one can analyze as a correlation network. In fact, the original data, i.e., $N\times L$ matrix $(x_{i\ell})$, is a bipartite network, in which one part consists of N nodes representing the researchers, and the other part consists of L nodes representing the papers. An edge exists between an author node and a paper node if and only if $x_{i\ell}=1$. Then, we can interpret the $N\times N$ co-occurrence, or correlation, network whose (i,j) entry is given by $\sum_{\ell=1}^{L}x_{i\ell}x_{j\ell}$ as a one-mode projection of the bipartite network. This viewpoint provides research opportunities. It is known that one-mode projection introduces various biases [346]. For example, one-mode projection inflates the clustering coefficient [347–349], which is in fact consistent with the finding that correlation networks even from random data would have a high clustering coefficient (Section 3.2). One strategy to avoid such biases is to analyze the data as a bipartite network [346]. Therefore, bipartite network analysis may be equally useful for understanding correlation structure of continuous-valued data, i.e., an $N\times L$ matrix $(x_{i\ell})$, $x_{i\ell}\in\mathbb{R}$, which we have mostly dealt with in the present article. Establishing mapping of the continuous-valued data to a bipartite network is a first natural step toward this goal. A framework that both handles bipartite signed networks and provides a statistically validated one-mode projection [350] may be fruitfully extended to this direction.

We have confined the present review to pairwise correlation. It is possible to quantify co-fluctuations among more than two nodes and analyze their system-wide organization by hypergraph [351] or simplicial complex [352,353] techniques. In fact, pursuit of higher-order correlation among three or more variables is itself not new; various information-theoretic and statistical methods have been available to quantify higher-order correlation in data for more than two decades, particularly in computational neuroscience [354–359]. As a network represents a set of pairwise correlations, a hypergraph or simplicial complex [360–362] (also see Section 3.3 for references for the latter) represents a set of higher-order correlations. Further developing methods for analyzing data with higher-order correlation using hypergraph and simplicial complex techniques would be fruitful.

One complication that has not received enough attention is that many in-practice comparisons involve ensembles of observed networks rather than single networks. This is the case in most fMRI studies where networks are used. When working with an ensemble of networks, one must make various decisions, such as whether or not to ensure that edge density is constant across networks (see Section 3.2 for the absolute versus proportional threshold). The development of mathematical theories for how to construct correlation-based networks for ensembles may be helpful because most null models and other tools are only oriented toward single networks. Multilayer approach and geometric approaches to correlation networks and matrices, both of which cater to between-network/matrix comparisons, are promising paths toward this goal.

A graphon is a symmetric measurable function $W:[0,1]^2\to[0,1]$. Given W, we generate dense graphs in which there is an edge between the ith and jth nodes with probability $W(u_i,u_j)$, where each u_i $\forall i$ independently obeys the uniform

density on [0, 1] [363]. In network science, assigning a node weight, either from a probability distribution or empirical data, and connecting two nodes probabilistically as a function of the two node weights has been a major method to generate networks [364–370]. Basic correlation networks are equivalent to an extension of this class of network models where u_i is an L-dimensional vector of the ith feature from L samples, and W is a criterion with which to determine the edge. In fact, similar to the construction of correlation networks by dichotomizing, dichotomizing functions have been used as W to generate networks with power-law degree distributions from scalar node weights that do not have to obey long-tailed distributions [367,369,371–373]. Importing mathematical frameworks and methods from graphon-related models, such as the limit of a sequence of dense networks, to correlation network analysis may be an interesting idea. Those frameworks may be able to provide null models for correlation networks or give more mathematical foundations of correlation networks.

7.3. Final words

We have seen that there are various research fields in which they collect and analyze correlation networks. We have also seen that some particular analysis techniques are heavily studied in one field, but others are preferred in different fields. For example, random matrices for sample correlation matrices [8,9] have particularly been used in financial data studies, including econophysics [6,101–106], while they have been applied much less, and only more recently, in neuroscience [65,298]. As another example, a majority of studies of temporal correlation networks has been done in network neuroscience under the name of dynamic functional connectivity/networks. However, very often, methods for analyzing correlation networks developed in one research field do not rely on particularities of the field and are therefore transferable to other research fields. While such cross-fertilization has been ongoing and advocated [21], we emphasize that much more of it will be useful for furthering correlation network analysis and applications. By the same token, studies directly comparing objectives and performances of methods used in different research domains will be valuable.

Cross-fertilization is also desirable within theoretical fields. Statisticians and non-statisticians tend not to know each other's work and publish in different types of journals. Statisticians tend to start with research questions that are ideally asked and answered by statistical hypothesis testing or Bayesian methods. Therefore, they would develop methods for correlation networks with which the data analysis results can be statistically tested. In contrast, non-statistically-focused researchers including many applied mathematicians, physicists, and computer scientists tend to focus on network analysis techniques which may be heuristic or reflect analogies to other processes, many of which have been proven to be powerful in different settings. Because there are already many useful network analysis methods, if one can exploit them in correlation data analysis, these types of researchers, including the present authors, would expect that it is a beneficial connection. We tried to cover both types of approaches to correlation networks as much as possible in this article. We believe that more discussion between these different perspectives on correlation networks will drive further developments of both types of approach. See for example [374] for related discussion.

CRediT authorship contribution statement

Naoki Masuda: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Zachary M. Boyd:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization. **Diego Garlaschelli:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition. **Peter J. Mucha:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Sarah Muldoon for discussion. N.M. acknowledges support from National Institute of General Medical Sciences, United States (under grant no. R01GM148973), the Japan Science and Technology Agency (JST) Moonshot R&D (under grant no. JPMJMS2021), the National Science Foundation, United States (under grant nos. 2052720 and 2204936), and JSPS KAKENHI (under grant nos. JP 21H04595 and 23H03414). P.J.M. and Z.M.B. acknowledge support from the Army Research Office, United States (under MURI award W911NF-18-1-0244). P.J.M. also acknowledges support from the National Institute of Diabetes and Digestive and Kidney Diseases, United States (under grant no. R01DK125860) and from the National Science Foundation, United States (under grant no. 2140024). Z.M.B. also acknowledges support from the National Science Foundation, United States (under grant no. 2137511). D.G. acknowledges support by the European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR), project 'SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics' - grant IR0000013 (n. 3264, 28/12/2021) (https://pnrr.sobigdata.it/), project 'RECON-NET - Reconstruction, Resilience and Recovery of Socio-Economic

Networks' - grant EP_FAIR_005, PE0000013 "FAIR" (PNRR M4C2 Investment 1.3), and project 'MNESYS - A Multiscale integrated approach to the study of the nervous system in health and disease' - grant PE0000006 (DN. 1553 11.10.2022). The content is solely the responsibility of the authors and does not necessarily represent the official views of any agency supporting this work.

References

[1] M. Becker, H. Nassar, C. Espinosa, I.A. Stelzer, D. Feyaerts, E. Berson, N.H. Bidoki, A.L. Chang, G. Saarunya, A. Culos, D. De Francesco, R. Fallahzadeh, Q. Liu, Y. Kim, I. Marić, S.J. Mataraso, S.N. Payrovnaziri, T. Phongpreecha, N.G. Ravindra, N. Stanley, S. Shome, Y. Tan, M. Thuraiappah, M. Xenochristou, L. Xue, G. Shaw, D. Stevenson, M.S. Angst, B. Gaudilliere, N. Aghaeepour, Large-scale correlation network construction for unraveling the coordination of complex biological systems, Nat. Comput. Sci. 3 (2023) 346–359.
[2] I.T. Jolliffe, Principal component analysis, second ed., Springer, New York, NY, 2002.
[3] T.W. Anderson, An introduction to multivariate statistical analysis, third ed., John Wiley & Sons, Hoboken, NJ, 2003.
[4] S.A. Mulaik, Foundations of factor analysis, second ed., Taylor & Francis, Boca Raton, FL, 2010.
[5] H. Markowitz, Portfolio selection, J. Financ. 7 (1952) 77–91.
[6] J. Bun, J.P. Bouchaud, M. Potters, Cleaning large correlation matrices: Tools from Random Matrix Theory, Phys. Rep. 666 (2017) 1–109.
[7] M.L. Mehta, Random Matrices, third ed., Elsevier, San Diego, CA, 2004.
[8] G. Livan, M. Novaes, P. Vivo, Introduction to Random Matrices, Springer, Cham, Switzerland, 2018.
[9] M. Potters, J.-P. Bouchaud, A First Course in Random Matrix Theory, Cambridge University Press, Cambridge, UK, 2021.
[10] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (2008) 084102.
[11] X.-Y. Qian, Y.-M. Liu, Z.-Q. Jiang, B. Podobnik, W.-X. Zhou, H.E. Stanley, Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces, Phys. Rev. E 91 (2015) 062816.
[12] N. Yuan, Z. Fu, H. Zhang, L. Piao, E. Xoplaki, J. Luterbacher, Detrended partial-cross-correlation analysis: A new method for analyzing correlations in complex system, Sci. Rep. 5 (2015) 8143.
[13] D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R.N. Mantegna, E. Ben-Jacob, Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market, PLoS ONE 5 (2010) e15032.
[14] D.Y. Kenett, T. Preis, G. Gur-Gershgoren, E. Ben-Jacob, Dependency network and node influence: Application to the study of financial markets, Int. J. Bifu. Chaos 22 (2012) 1250181.
[15] D. Zhou, A. Gozolchiani, Y. Ashkenazy, S. Havlin, Teleconnection paths via climate network direct link detection, Phys. Rev. Lett. 115 (2015) 268501.
[16] L. Chen, R. Liu, Z.-P. Liu, M. Li, K. Aihara, Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers, Sci. Rep. 2 (2012) 342.
[17] S. Chen, E.B. O'Dea, J.M. Drake, B.I. Epureanu, Eigenvalues of the covariance matrix as early warning signals for critical transitions in ecological systems, Sci. Rep. 9 (2019) 2572.
[18] N.G. MacLaren, P. Kundu, N. Masuda, Early warnings for multi-stage transitions in dynamics on networks, J. R. Soc. Interface 20 (2023) 20220743.
[19] T. Watanabe, N. Masuda, F. Megumi, R. Kanai, G. Rees, Energy landscape and dynamics of brain activity during human bistable perception, Nat. Comm. 5 (2014) 4765.
[20] T. Ezaki, T. Watanabe, M. Ohzeki, N. Masuda, Energy landscape analysis of neuroimaging data, Phil. Trans. R. Soc. A 375 (2017) 20160287.
[21] D. Borsboom, M.K. Deserno, M. Rhemtulla, S. Epskamp, E.I. Fried, R.J. McNally, D.J. Robinaugh, M. Perugini, J. Dalege, G. Costantini, A.-M. Isvoranu, A.C. Wysocki, C.D. van Borkulo, R. van Bork, L.J. Waldorp, Network analysis of multivariate data in psychological science, Nat. Rev. Methods Prim. 1 (2021) 58.
[22] V. Kukreti, H.K. Pharasi, P. Gupta, S. Kumar, A perspective on correlation-based financial networks and entropy measures, Front. Phys. 8 (2020) 323.
[23] M. Tumminello, F. Lillo, R.N. Mantegna, Correlation, hierarchies, and networks in financial markets, J. Econ. Behav. Org. 75 (2010) 40–58.
[24] F. De Vico Fallani, J. Richiardi, M. Chavez, S. Achard, Graph analysis of functional brain networks: Practical issues in translational neuroscience, Phil. Trans. R. Soc. B 369 (2014) 20130521.
[25] S. Epskamp, D. Borsboom, E.I. Fried, Estimating psychological networks and their accuracy: A tutorial paper, Behav. Res. 50 (2018) 195–212.
[26] M.-E. Lynall, D.S. Bassett, R. Kerwin, P.J. McKenna, M. Kitzbichler, U. Muller, E. Bullmore, Functional connectivity and brain networks in schizophrenia, J. Neurosci. 30 (2010) 9477–9487.
[27] X. Liu, Y. Liu, J. Chen, X. Liu, PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization, IEEE Trans. Circ. Syst. Video Technol. 32 (2022) 7505–7517.
[28] M. Lee, C. Park, S. Cho, S. Lee, Superpixel group-correlation network for co-saliency detection, in: Proceedings of the International Conference on Image Processing, ICIP, 2022, pp. 806–810.
[29] Z. Tu, Z. Li, C. Li, J. Tang, Weakly alignment-free RGBT salient object detection with deep correlation network, IEEE Trans. Image Process. 31 (2022) 3752–3764.
[30] U. von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (2007) 395–416.
[31] M. Tumminello, C. Coronnello, F. Lillo, S. Micciche, R.N. Mantegna, Spanning trees and bootstrap reliability estimation in correlation-based networks, Int. J. Bifu. Chaos 17 (2007) 2319–2329.
[32] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, M.W. Woolrich, Network modelling methods for FMRI, NeuroImage 54 (2011) 875–891.
[33] K. Faust, J. Raes, Microbial interactions: From networks to models, Nat. Rev. Microbiol. 10 (2012) 538–550.
[34] Y.X.R. Wang, H. Huang, Review on statistical methods for gene network reconstruction using expression data, J. Theoret. Biol. 362 (2014) 53–61.
[35] A. Fukushima, M. Kusano, H. Redestig, M. Arita, K. Saito, Metabolomic correlation-network modules in *Arabidopsis* based on a graph-clustering approach, BMC Syst. Biol. 5 (2011) 1.
[36] T. Millington, M. Niranjan, Construction of minimum spanning trees from financial returns using rank correlation, Phys. A 566 (2021) 125605.
[37] A.J. Butte, I.S. Kohane, Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements, in: Pacific Symposium on Biocomputing, World Scientific, Singapore, 2000, pp. 418–429.
[38] P.E. Meyer, F. Lafitte, G. Bontempi, minet: A r/bioconductor package for inferring large transcriptional networks using mutual information, BMC Bioinformatics 9 (2008) 461.

[39] X. Guo, H. Zhang, T. Tian, Development of stock correlation networks using mutual information and financial big data, PLoS ONE 13 (2018) e0195941.
[40] R. Salvador, J. Suckling, C. Schwarzbauer, E. Bullmore, Undirected graphs of frequency-dependent functional connectivity in whole brain networks, Phil. Trans. R. Soc. B 360 (2005) 937–946.
[41] P. Fiedor, Partial mutual information analysis of financial networks, Acta Phys. Pol. A 127 (2015) 863–867.
[42] J. Pearl, Causal inference, J. Mach. Learn. Res. Work. Conf. Proc. 6 (2010) 39–58.
[43] G. Briganti, M. Scutari, R.J. McNally, A tutorial on bayesian networks for psychopathology researchers, Psychol. Methods 28 (2023) 947–961.
[44] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M.D. Mahecha, J. Muñoz-Marí, E.H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, J. Zscheischler, Inferring causation from time series in Earth system sciences, Nat. Commun. 10 (2019) 2553.
[45] P. Maurage, A. Heeren, M. Pesenti, Does chocolate consumption really boost nobel award chances? The peril of over-interpreting correlations in health studies, J. Nutr. 143 (2013) 931–933.
[46] J.M. Rohrer, Thinking clearly about correlations and causation: Graphical causal models for observational data, Adv. Methods Pr. Psychol. Sci. 1 (2018) 27–42.
[47] O. Ryan, L.F. Bringmann, N.K. Schuurman, The challenge of generating causal hypotheses using network models, Struct. Equ. Model. 29 (2022) 953–970.
[48] G.M. Harari, N.D. Lane, R. Wang, B.S. Crosier, A.T. Campbell, S.D. Gosling, Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges, Perspect. Psychol. Sci. 11 (2016) 838–854.
[49] A.L. McGowan, F. Sayed, Z.M. Boyd, M. Jovanova, Y. Kang, M.E. Speer, D. Cosme, P.J. Mucha, K.N. Ochsner, D.S. Bassett, E.B. Falk, D.M. Lydon-Staley, Dense sampling approaches for psychiatry research: Combining scanners and smartphones, Biol. Psychiatry 93 (2023) 681–689.
[50] D. Borsboom, A.O.J. Cramer, Network analysis: An integrative approach to the structure of psychopathology, Annu. Rev. Clin. Psychol. 9 (2013) 91–121.
[51] E.I. Fried, A.O.J. Cramer, Moving forward: Challenges and directions for psychopathological network theory and methodology, Pers. Psychol. Sci. 12 (2017) 999–1020.
[52] E.I. Fried, C.D. van Borkulo, A.O.J. Cramer, L. Boschloo, R.A. Schoevers, D. Borsboom, Mental disorders as networks of problems: A review of recent insights, Soc. Psychiatry Psychiatr. Epidemiol. 52 (2017) 1–10.
[53] A.L. McGowan, Z.M. Boyd, Y. Kang, L. Bennett, P.J. Mucha, K.N. Ochsner, D.S. Bassett, E.B. Falk, D.M. Lydon-Staley, Within-person temporal associations among self-reported physical activity, sleep, and well-being in college students, Psychosom. Med. 85 (2023) 141–153.
[54] S. Epskamp, L.J. Waldorp, R. Mõttus, D. Borsboom, The Gaussian graphical model in cross-sectional and time-series data, Multivar. Behav. Res. 53 (2018) 453–480.
[55] W. Lutz, B. Schwartz, S.G. Hofmann, A.J. Fisher, K. Husen, J.A. Rubel, Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study, Sci. Rep. 8 (2018) 7819.
[56] D.C.R. van Zelst, E.M. Veltman, D. Rhebergen, P. Naarding, A.A.L. Kok, N.R. Ottenheim, E.J. Giltay, Network structure of time-varying depressive symptoms through dynamic time warp analysis in late-life depression, Int. J. Geriatr. Psychiatry 37 (2022) 1–12.
[57] M.K. Forbes, A.G.C. Wright, K.E. Markon, R.F. Krueger, Evidence that psychopathology symptom networks have limited replicability., J. Abnorm. Psychol. 126 (2017) 969–988.
[58] E. Bullmore, O. Sporns, Complex brain networks: Graph theoretical analysis of structural and functional systems, Nat. Rev. Neurosci. 10 (2009) 186–198.
[59] D.S. Bassett, O. Sporns, Network neuroscience, Nat. Neurosci. 20 (2017) 353–364.
[60] J. Chung, E. Bridgeford, J. Arroyo, B.D. Pedigo, A. Saad-Eldin, V. Gopalakrishnan, L. Xiang, C.E. Priebe, J.T. Vogelstein, Statistical connectomics, Annu. Rev. Stat. Appl. 8 (2021) 463–492.
[61] H.-J. Park, K. Friston, Structural and functional brain networks: From connections to cognition, Science 342 (2013) 1238411.
[62] R. Liégeois, A. Santos, V. Matta, D. Van De Ville, A.H. Sayed, Revisiting correlation-based functional connectivity and its relationship with structural connectivity, Netw. Neurosci. 4 (2020) 1235–1251.
[63] R.C. Craddock, S. Jbabdi, C.-G. Yan, J.T. Vogelstein, F.X. Castellanos, A. Di Martino, C. Kelly, K. Heberlein, S. Colcombe, M.P. Milham, Imaging human connectomes at the macroscale, Nat. Methods 10 (2013) 524–539.
[64] G. Varoquaux, R.C. Craddock, Learning and comparing functional connectomes across subjects, NeuroImage 80 (2013) 405–415.
[65] M. Ibáñez-Berganza, C. Lucibello, F. Santucci, T. Gili, A. Gabrielli, Noise cleaning the precision matrix of short time series, Phys. Rev. E 108 (2023) 024313.
[66] B. Biswal, F.Z. Yetkin, V.M. Haughton, J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI, Magn. Reson. Med. 34 (1995) 537–541.
[67] G.L. Colclough, M.W. Woolrich, P.K. Tewarie, M.J. Brookes, A.J. Quinn, S.M. Smith, How reliable are MEG resting-state connectivity metrics? NeuroImage 138 (2016) 284–293.
[68] A. Alexander-Bloch, J.N. Giedd, E. Bullmore, Imaging structural co-variance between human brain regions, Nat. Rev. Neurosci. 14 (2013) 322–336.
[69] A.C. Evans, Networks of anatomical covariance, NeuroImage 80 (2013) 489–504.
[70] J. Seidlitz, F. Váša, M. Shinn, R. Romero-Garcia, K.J. Whitaker, P.E. Vértes, K. Wagstyl, P. Kirkpatrick Reardon, L. Clasen, S. Liu, A. Messinger, D.A. Leopold, P. Fonagy, R.J. Dolan, P.B. Jones, I.M. Goodyer, the NSPN Consortium, A. Raznahan, E.T. Bullmore, Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation, Neuron 97 (2018) 231–247.
[71] J.Y. Hansen, G. Shafiei, R.D. Markello, K. Smart, S.M.L. Cox, M. Nørgaard, V. Beliveau, Y. Wu, J.-D. Gallezot, É. Aumont, S. Servaes, S.G. Scala, J.M. DuBois, G. Wainstein, G. Bezgin, T. Funck, T.W. Schmitz, R.N. Spreng, M. Galovic, M.J. Koepp, J.S. Duncan, J.P. Coles, T.D. Fryer, F.I. Aigbirhio, C.J. McGinnity, A. Hammers, J.-P. Soucy, S. Baillet, S. Guimond, J. Hietala, M.-A. Bedard, M. Leyton, E. Kobayashi, P. Rosa-Neto, M. Ganz, G.M. Knudsen, N. Palomero-Gallagher, J.M. Shine, R.E. Carson, L. Tuominen, A. Dagher, B. Misic, Mapping neurotransmitter systems to the structural and functional organization of the human neocortex, Nat. Neurosci. 25 (2022) 1569–1581.
[72] S. Horvath, J. Dong, Geometric interpretation of gene coexpression network analysis, PLoS Comput. Biol. 4 (2008) e1000117.
[73] P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis, BMC Bioinformatics 9 (2008) 559.
[74] B.H. Junker, F. Schreiber (Eds.), Analysis of Biological Networks, John Wiley & Sons, Inc., Hoboken, NJ, 2008.
[75] M. Vidal, M.E. Cusick, A.-L. Barabási, Interactome networks and human disease, Cell 144 (2011) 986–998.
[76] C. Gaiteri, Y. Ding, B. French, G.C. Tseng, E. Sibille, Beyond modules and hubs: The potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders, Genes Brain Behav. 13 (2014) 13–24.
[77] G. Fiscon, F. Conte, L. Farina, P. Paci, Network-based approaches to explore complex biological systems towards network medicine, Genes 9 (2018) 437.
[78] S. van Dam, U. Võsa, A. van der Graaf, L. Franke, J.P. de Magalhães, Gene co-expression analysis for functional classification and gene–disease predictions, Brief. Bioinfo. 19 (2018) 575–592.

[79] M. Mircea, M. Hochane, X. Fan, S.M. Chuva de Sousa Lopes, D. Garlaschelli, S. Semrau, Phiclust: A clusterability measure for single-cell transcriptomics reveals phenotypic subpopulations, Genome Biol. 23 (2022) 18.
[80] K.S. Parker, J.D. Wilson, J. Marschall, P.J. Mucha, J.P. Henderson, Network analysis reveals sex- and antibiotic resistance-associated antivirulence targets in clinical uropathogens, ACS Infect. Dis. 1 (2015) 523–532.
[81] Z. Zou, R.F. Potter, W.H. McCoy 4th, J.A. Wildenthal, G.L. Katumba, P.J. Mucha, G. Dantas, J.P. Henderson, *E. coli* catheter-associated urinary tract infections are associated with distinctive virulence and biofilm gene determinants, JCI Insight 8 (2023) e161461.
[82] W.-C. Chou, A.-L. Cheng, M. Brotto, C.-Y. Chuang, Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer, BMC Genomics 15 (2014) 300.
[83] R. Dobrin, J. Zhu, C. Molony, C. Argman, M.L. Parrish, S. Carlson, M.F. Allan, D. Pomp, E.E. Schadt, Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease, Genome Biol. 10 (2009) R55.
[84] Y. Xiang, J. Zhang, K. Huang, Mining the tissue-tissue gene co-expression network for tumor microenvironment study and biomarker prediction, BMC Genomics 14 (2013) S4.
[85] L.J.A. Kogelman, J. Fu, L. Franke, J.W. Greve, M. Hofker, S.S. Rensen, H.N. Kadarmideen, Inter-tissue gene co-expression networks between metabolically healthy and unhealthy obese individuals, PLoS ONE 11 (2016) e0167519.
[86] N.J. Hudson, A. Reverter, B.P. Dalrymple, A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation, PLoS Comput. Biol. 5 (2009) e1000382.
[87] R. Steuer, On the analysis and interpretation of correlations in metabolomic data, Brief. Bioinfo. 7 (2006) 151–158.
[88] E.K. Silverman, H.H.H.W. Schmidt, E. Anastasiadou, L. Altucci, M. Angelini, L. Badimon, J.-L. Balligand, G. Benincasa, G. Capasso, F. Conte, A. Di Costanzo, L. Farina, G. Fiscon, L. Gatto, M. Gentili, J. Loscalzo, C. Marchese, C. Napoli, P. Paci, M. Petti, J. Quackenbush, P. Tieri, D. Viggiano, G. Vilahur, K. Glass, J. Baumbach, Molecular networks in Network Medicine: Development and applications, Wiley Interdiscip. Rev.: Syst. Biol. Med. 12 (2020) e1489.
[89] D. Camacho, A. de la Fuente, P. Mendes, The origin of correlations in metabolomics data, Metabolomics 1 (2005) 53–63.
[90] J.I. Robinson, W.H. Weir, J.R. Crowley, T. Hink, K.A. Reske, J.H. Kwon, C.-A.D. Burnham, E.R. Dubberke, P.J. Mucha, J.P. Henderson, Metabolomic networks connect host-microbiome processes to human *Clostridioides difficile* infections, J. Clin. Invest. 129 (2019) 3792–3806.
[91] H. Hirano, K. Takemoto, Difficulty in inferring microbial community structure based on co-occurrence network approaches, BMC Bioinformatics 20 (2019) 329.
[92] M. Goberna, M. Verdú, Cautionary notes on the use of co-occurrence networks in soil ecology, Soil Biol. Biochem. 166 (2022) 108534.
[93] J.M. Diamond, Assembly of species communities, in: M.L. Cody, J.M. Diamond (Eds.), Ecology and Evolution of Communities, Harvard University Press, Cambridge, MA, 1975, pp. 342–444.
[94] J.X. Hu, C.E. Thomas, S. Brunak, Network biology concepts in complex disease comorbidities, Nat. Rev. Genet. 17 (2016) 615–629.
[95] C.A. Hidalgo, N. Blumm, A.-L. Barabási, N.A. Christakis, A dynamic network approach for the study of human phenotypes, PLoS Comput. Biol. 5 (2009) e1000353.
[96] K.-I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.-L. Barabási, The human disease network, Proc. Natl. Acad. Sci. USA 104 (2007) 8685–8690.
[97] A. Rzhetsky, D. Wajngurt, N. Park, T. Zheng, Probing genetic overlap among complex human phenotypes, Proc. Natl. Acad. Sci. USA 104 (2007) 11694–11699.
[98] D.-S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, A.-L. Barabási, The implications of human metabolic network topology for disease comorbidity, Proc. Natl. Acad. Sci. USA 105 (2008) 9880–9885.
[99] X. Zhou, J. Menche, A.-L. Barabási, A. Sharma, Human symptoms–disease network, Nat. Commun. 5 (2014) 4212.
[100] R.N. Mantegna, H.E. Stanley, An introduction to econophysics, Cambridge University Press, Cambridge, UK, 1999.
[101] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (1999) 1467–1470.
[102] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. 83 (1999) 1471–1474.
[103] M. MacMahon, D. Garlaschelli, Community detection for correlation matrices, Phys. Rev. X 5 (2015) 021006.
[104] A. Almog, F. Besamusca, M. MacMahon, D. Garlaschelli, Mesoscopic community structure of financial markets revealed by price and sign fluctuations, PLoS ONE 10 (2015) e0133679.
[105] I. Anagnostou, T. Squartini, D. Kandhai, D. Garlaschelli, Uncovering the mesoscale structure of the credit default swap market to improve portfolio risk modelling, Quant. Finance 21 (2021) 1501–1518.
[106] S.M. Zema, G. Fagiolo, T. Squartini, D. Garlaschelli, Mesoscopic structure of the stock market and portfolio optimization, J. Econ. Interact. Coord. (2024).
[107] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1999) 193–197.
[108] G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Topology of correlation-based minimal spanning trees in real and model markets, Phys. Rev. E 68 (2003) 046130.
[109] D.J. Fenn, M.A. Porter, P.J. Mucha, M. McDonald, S. Williams, N.F. Johnson, N.S. Jones, Dynamical clustering of exchange rates, Quant. Finance 12 (2012) 1493–1520.
[110] M. Bazzi, M.A. Porter, S. Williams, M. McDonald, D.J. Fenn, S.D. Howison, Community detection in temporal multilayer networks, with an application to correlation networks, Multiscale Model. Simul. 1 (2016) 1–41.
[111] M. Tumminello, F. Lillo, J. Piilo, R.N. Mantegna, Identification of clusters of investors from their real trading activity in a financial market, New J. Phys. 14 (2012) 013041.
[112] S. Ranganathan, M. Kivelä, J. Kanniainen, Dynamics of investor spanning trees around dot-com bubble, PLoS ONE 13 (2018) e0198807.
[113] D. Kercher, Inconsistent Correlation and Momenta: A New Approach to Portfolio Allocation (M.S. thesis), Brigham Young University, 2023.
[114] R.K.Y. Low, E. Tan, The role of analyst forecasts in the momentum effect, Int. Rev. Financ. Anal. 48 (2016) 67–84.
[115] E. Yan, Y. Ding, Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other, J. Amer. Soc. Info. Sci. Tech. 63 (2012) 1313–1326.
[116] J.-P. Qiu, K. Dong, H.-Q. Yu, Comparative study on structure and correlation among author co-occurrence networks in bibliometrics, Scientometrics 101 (2014) 1345–1360.
[117] M.E.J. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (2001) 016132.
[118] C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, V. Loreto, Collective dynamics of social annotation, Proc. Natl. Acad. Sci. USA 106 (2009) 10511–10515.
[119] F. Luo, J.Z. Wang, E. Promislow, Exploring local community structures in large networks, in: Proceedings of 2006 IEEE/WIC/ACM International Conference on Web Intelligence, WI'06, 2006, pp. 233–239.
[120] A.A. Tsonis, P.J. Roebber, The architecture of the climate network, Phys. A 333 (2004) 497–504.
[121] A.A. Tsonis, K.L. Swanson, P.J. Roebber, What do networks have to do with climate? Bull. Amer. Meteorol. Soc. 87 (2006) 585–595.
[122] J.F. Donges, Y. Zou, N. Marwan, J. Kurths, The backbone of the climate network, EPL 87 (2009) 48007.
[123] J. Fan, J. Meng, J. Ludescher, X. Chen, Y. Ashkenazy, J. Kurths, S. Havlin, H.J. Schellnhuber, Statistical physics approaches to the complex Earth system, Phys. Rep. 896 (2021) 1–84.

[124] J. Heitzig, J.F. Donges, Y. Zou, N. Marwan, J. Kurths, Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes, Eur. Phys. J. B 85 (2012) 38.
[125] S. Scarsoglio, F. Laio, L. Ridolfi, Climate dynamics: A network-based approach for the analysis of global precipitation, PLoS ONE 8 (2013) e71129.
[126] M. van der Mheen, H.A. Dijkstra, A. Gozolchiani, M. den Toom, Q. Feng, J. Kurths, E. Hernandez-Garcia, Interaction network based early warning indicators for the Atlantic MOC collapse, Geophys. Res. Lett. 40 (2013) 2714–2719.
[127] P.J. Bickel, E. Levina, Covariance regularization by thresholding, Ann. Stat. 36 (2008) 2577–2604.
[128] M. Pourahmadi, Covariance estimation: The GLM and regularization perspectives, Statist. Sci. 26 (2011) 369–387.
[129] J. Fan, Y. Liao, H. Liu, An overview of the estimation of large covariance and precision matrices, Econ. J. 19 (2016) C1–C32.
[130] R.B. Holmes, On random correlation matrices, SIAM J. Matrix Anal. Appl. 12 (1991) 239–272.
[131] J.A. Tropp, Simplicial faces of the set of correlation matrices, Discrete Comput. Geom. 60 (2018) 512–529.
[132] Y. Thanwerdas, X. Pennec, Geodesics and curvature of the quotient-affine metrics on full-rank correlation matrices, Lecture Notes in Comput. Sci. 12829 (2021) 93–102.
[133] J. Whittaker, Graphical Models in Applied Multivariate Statistics, John Wiley & Sons, Chichester, UK, 1990.
[134] S.L. Lauritzen, Graphical models, Clarendon Press, Oxford, UK, 1996.
[135] A.P. Dempster, Covariance selection, Biometrics 28 (1972) 157–175.
[136] O. Ledoit, M. Wolf, The power of (non-)linear shrinking: A review and guide to covariance matrix estimation, J. Financ. Econ. 20 (2022) 187–218.
[137] O. Ledoit, M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, J. Empir. Financ. 10 (2003) 603–621.
[138] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2004) 365–411.
[139] Y. Chen, A. Wiesel, Y.C. Eldar, A.O. Hero, Shrinkage algorithms for MMSE covariance estimation, IEEE Trans. Signal Proc. 58 (2010) 5016–5029.
[140] M. Rubinov, O. Sporns, Weight-conserving characterization of complex functional brain networks, NeuroImage 56 (2011) 2068–2079.
[141] H. Lee, H. Kang, M.K. Chung, B.-N. Kim, D.S. Lee, Persistent brain network homology from the perspective of dendrogram, IEEE Trans. Med. Imaging 31 (2012) 2267–2277.
[142] K.A. Garrison, D. Scheinost, E.S. Finn, X. Shen, R.T. Constable, The (in)stability of functional brain network measures across thresholds, NeuroImage 118 (2015) 651–661.
[143] F. De Vico Fallani, V. Latora, M. Chavez, A topological criterion for filtering information in complex brain networks, PLoS Comput. Biol. 13 (2017) e1005305.
[144] M.K. Chung, H. Lee, A. DiChristofano, H. Ombao, V. Solo, Exact topological inference of the resting-state brain networks in twins, Netw. Neurosci. 3 (2019) 674–694.
[145] M.W. Cole, S. Pathak, W. Schneider, Identifying the brain's most globally connected regions, NeuroImage 49 (2010) 3132–3148.
[146] D. Scheinost, J. Benjamin, C.M. Lacadie, B. Vohr, K.C. Schneider, L.R. Ment, X. Papademetris, R.T. Constable, The intrinsic connectivity distribution: A novel contrast measure reflecting voxel level functional connectivity, NeuroImage 62 (2012) 1510–1519.
[147] L. Peel, T.P. Peixoto, M. De Domenico, Statistical inference links data and theory in network science, Nat. Commun. 13 (2022) 6794.
[148] M.P. van den Heuvel, S.C. de Lange, A. Zalesky, C. Seguin, B.T.T. Yeo, R. Schmidt, Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations, NeuroImage 152 (2017) 437–449.
[149] S. Achard, E. Bullmore, Efficiency and cost of economical brain functional networks, PLoS Comput. Biol. 3 (2007) e17.
[150] M.P. van den Heuvel, C.J. Stam, M. Boersma, H.E. Hulshoff Pol, Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain, NeuroImage 43 (2008) 528–539.
[151] J. Wang, L. Wang, Y. Zang, H. Yang, H. Tang, Q. Gong, Z. Chen, C. Zhu, Y. He, Parcellation-dependent small-world brain functional networks: A resting-state fMRI study, Hum. Brain Mapp. 30 (2009) 1511–1523.
[152] B.C.M. van Wijk, C.J. Stam, A. Daffertshofer, Comparing brain networks of different size and connectivity density using graph theory, PLoS ONE 5 (2010) e13701.
[153] A.F. Alexander-Bloch, N. Gogtay, D. Meunier, R. Birn, L. Clasen, F. Lalonde, R. Lenroot, J. Giedd, E.T. Bullmore, Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia, Front. Syst. Neurosci. 4 (2010) 147.
[154] E. Langford, N. Schwertman, M. Owens, Is the property of being positively correlated transitive? Am. Stat. 55 (2001) 322–325.
[155] J. Gillis, P. Pavlidis, The role of indirect connections in gene networks in predicting function, Bioinformatics 27 (2011) 1860–1866.
[156] A. Zalesky, A. Fornito, E. Bullmore, On the use of correlation as a measure of network connectivity, NeuroImage 60 (2012) 2096–2106.
[157] M. Barthélemy, Spatial networks, Phys. Rep. 499 (2011) 1–101.
[158] M. Barthelemy, Morphogenesis of Spatial Networks, Springer, Cham, Switzerland, 2018.
[159] N. Langer, A. Pedroni, L. Jäncke, The problem of thresholding in small-world network analysis, PLoS ONE 8 (2013) e53199.
[160] C.E. Ginestet, T.E. Nichols, E.T. Bullmore, A. Simmons, Brain network analysis: Separating cost from topology using cost-integration, PLoS ONE 6 (2011) e21570.
[161] D.S. Bassett, B.G. Nelson, B.A. Mueller, J. Camchong, K.O. Lim, Altered resting state complexity in schizophrenia, NeuroImage 59 (2012) 2196–2207.
[162] S.M.H. Hosseini, F. Hoeft, S.R. Kesler, GAT: A graph-theoretical analysis toolbox for analyzing between-group differences in large-scale structural and functional brain networks, PLoS ONE 7 (2012) e40709.
[163] V. Latora, M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. 87 (2001) 198701.
[164] D.S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, E. Bullmore, Adaptive reconfiguration of fractal small-world human brain functional networks, Proc. Natl. Acad. Sci. USA 103 (2006) 19518–19523.
[165] N. Vandewalle, F. Brisbois, X. Tordoir, Non-random topology of stock markets, Quant. Finance 1 (2001) 372–374.
[166] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, Dynamic asset trees and portfolio analysis, Eur. Phys. J. B 30 (2002) 285–288.
[167] M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, A tool for filtering information in complex systems, Proc. Natl. Acad. Sci. USA 102 (2005) 10421–10426.
[168] R. Mastrandrea, A. Gabrielli, F. Piras, G. Spalletta, G. Caldarelli, T. Gili, Organization and hierarchy of the human functional brain network lead to a chain-like core, Sci. Rep. 7 (2017) 4888.
[169] M.Á. Serrano, M. Boguñá, A. Vespignani, Extracting the multiscale backbone of complex weighted networks, Proc. Natl. Acad. Sci. USA 106 (2009) 6483–6488.
[170] V. Gemmetto, A. Cardillo, D. Garlaschelli, Irreducible network backbones: unbiased graph filtering via maximum entropy, 2017, arXiv preprint arXiv:1706.00230.
[171] L. Wang, C. Zhu, Y. He, Y. Zang, Q. Cao, H. Zhang, Q. Zhong, Y. Wang, Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder, Hum. Brain Mapp. 30 (2009) 638–649.
[172] R. Ghrist, Barcodes: The persistent topology of data, Bull. Amer. Math. Soc. 45 (2008) 61–75.

[173] D. Horak, S. Maletić, M. Rajković, Persistent homology of complex networks, J. Stat. Mech. 2009 (2009) P03034.
[174] N. Otter, M.A. Porter, U. Tillmann, P. Grindrod, H.A. Harrington, A roadmap for the computation of persistent homology, EPJ Data Sci. 6 (2017) 17.
[175] M.E. Aktas, E. Akbas, A.E. Fatmaoui, Persistence homology of networks: Methods and applications, Appl. Netw. Sci. 4 (2019) 61.
[176] A.E. Sizemore, J.E. Phillips-Cremins, R. Ghrist, D.S. Bassett, The importance of the whole: Topological data analysis for the network neuroscientist, Netw. Neurosci. 3 (2019) 656–673.
[177] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P.J. Hellyer, F. Vaccarino, Homological scaffolds of brain functional networks, J. R. Soc. Interface 11 (2014) 20140873.
[178] C. Giusti, E. Pastalkova, C. Curto, V. Itskov, Clique topology reveals intrinsic geometric structure in neural correlations, Proc. Natl. Acad. Sci. USA 112 (2015) 13455–13460.
[179] A.N. Duman, H. Pirim, Gene coexpression network comparison via persistent homology, Int. J. Genom. 2018 (2018) 7329576.
[180] D. Shnier, M.A. Voineagu, I. Voineagu, Persistent homology analysis of brain transcriptome data in autism, J. R. Soc. Interface 16 (2019) 20190531.
[181] G. Leibon, S. Pauls, D. Rockmore, R. Savell, Topological structures in the equities market network, Proc. Natl. Acad. Sci. USA 105 (2008) 20589–20594.
[182] M. Gidea, Topological data analysis of critical transitions in financial networks, in: 3rd International Winter School and Conference on Network Science: NetSci-X 2017 3, Springer, Cham, Switzerland, 2017, pp. 47–59.
[183] B. Rieck, U. Fugacci, J. Lukasczyk, H. Leitte, Clique community persistence: A topological visual analysis approach for complex networks, IEEE Trans. Vis. Comput. Graphics 24 (2017) 822–831.
[184] S. Horvath, Weighted Network Analysis, Springer, New York, NY, 2011.
[185] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[186] N. El Karoui, Operator norm consistent estimation of large-dimensional sparse covariance matrices, Ann. Stat. 36 (2008) 2717–2756.
[187] A.J. Rothman, E. Levina, J. Zhu, Generalized thresholding of large covariance matrices, J. Amer. Statist. Assoc. 104 (2009) 177–186.
[188] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.
[189] T. Cai, W. Liu, Adaptive thresholding for sparse covariance matrix estimation, J. Amer. Statist. Assoc. 106 (2011) 672–684.
[190] T.T. Cai, C.-H. Zhang, H.H. Zhou, Optimal rates of convergence for covariance matrix estimation, Ann. Stat. 38 (2010) 2118–2144.
[191] C. Chen, L. Cheng, K. Grennan, F. Pibiri, C. Zhang, J.A. Badner, E.S. Gershon, C. Liu, Two gene co-expression modules differentiate psychotics and controls, Mol. Psychiatry 18 (2013) 1308–1314.
[192] K.R.A. Van Dijk, T. Hedden, A. Venkataraman, K.C. Evans, S.W. Lazar, R.L. Buckner, Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization, J. Neurophysiol. 103 (2010) 297–321.
[193] J. Tang, Y. Chang, C. Aggarwal, H. Liu, A survey of signed network mining in social media, ACM Comput. Surv. 49 (2016) 42.
[194] L. Zhan, L.M. Jenkins, O.E. Wolfson, J.J. GadElkarim, K. Nocito, P.M. Thompson, O.A. Ajilore, M.K. Chung, A.D. Leow, The significance of negative correlations in brain connectivity, J. Comp. Neurol. 525 (2017) 3251–3265.
[195] S. Gómez, P. Jensen, A. Arenas, Analysis of community structure in networks of correlated data, Phys. Rev. E 80 (2009) 016114.
[196] C. Aicher, A.Z. Jacobs, A. Clauset, Learning latent block structure in weighted networks, J. Comp. Netw. 3 (2015) 221–248.
[197] T.P. Peixoto, Nonparametric weighted stochastic block models, Phys. Rev. E 97 (2018) 012306.
[198] A. de la Fuente, N. Bing, I. Hoeschele, P. Mendes, Discovery of meaningful associations in genomic data using partial correlation coefficients, Bioinformatics 20 (2004) 3565–3574.
[199] R. Salvador, J. Suckling, M.R. Coleman, J.D. Pickard, D. Menon, E. Bullmore, Neurophysiological architecture of functional magnetic resonance images of human brain, Cereb. Cortex 15 (2005) 1332–1342.
[200] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, H. Benali, Partial correlation for functional brain interactivity investigation in functional MRI, NeuroImage 32 (2006) 228–237.
[201] M. Goldstein, A.F.M. Smith, Ridge-type estimators for regression analysis, J. R. Stat. Soc. Ser. B 36 (1974) 284–291.
[202] A. Hero, B. Rajaratnam, Hub discovery in partial correlation graphs, IEEE Trans. Info. Th. 58 (2012) 6064–6078.
[203] R. Artner, P.P. Wellingerhof, G. Lafit, T. Loossens, W. Vanpaemel, F. Tuerlinckx, The shape of partial correlation matrices, Commun. Stat. 51 (2022) 4133–4150.
[204] T. Aste, T. Di Matteo, Dynamical networks from correlations, Phys. A 370 (2006) 156–161.
[205] D.R. Williams, Learning to live with sampling variability: Expected replicability in partial correlation networks, Psychol. Methods 27 (2022) 606–621.
[206] T. Millington, M. Niranjan, Partial correlation financial networks, Appl. Netw. Sci. 5 (2020) 11.
[207] G. Varoquaux, A. Gramfort, J.-B. Poline, B. Thirion, Brain covariance selection: better individual functional connectivity models using population prior, Adv. Neural Inf. Process. Syst. 23 (2010) 2334–2342.
[208] S. Ryali, T. Chen, K. Supekar, V. Menon, Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty, NeuroImage 59 (2012) 3852–3861.
[209] M.R. Brier, A. Mitra, J.E. McCarthy, B.M. Ances, A.Z. Snyder, Partial covariance based functional connectivity computation using Ledoit-Wolf covariance regularization, NeuroImage 121 (2015) 29–38.
[210] U. Pervaiz, D. Vidaurre, M.W. Woolrich, S.M. Smith, Optimising network modelling methods for fMRI, NeuroImage 211 (2020) 116604.
[211] F. Santucci, A. Jimenez-Marin, A. Gabrielli, P. Bonifazi, M. Ibáñez-Berganza, T. Gili, J.M. Cortes, Partial correlation as a tool for mapping functional-structural correspondence in human brain connectivity, 2025, bioRxiv: https://doi.org/10.1101/2024.10.16.618230.
[212] G.-J. Wang, C. Xie, H.E. Stanley, Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks, Comput. Econ. 51 (2018) 607–635.
[213] M. Drton, M.H. Maathuis, Structure learning in graphical modeling, Annu. Rev. Stat. Appl. 4 (2017) 365–393.
[214] S. Epskamp, E.I. Fried, A tutorial on regularized partial correlation networks, Psychol. Methods 23 (2018) 617–634.
[215] M. Yuan, Y. Lin, Model selection and estimation in the Gaussian graphical model, Biometrika 94 (2007) 19–35.
[216] O. Banerjee, L. El Ghaoui, A. d'Aspremont, Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data, J. Mach. Learn. Res. 9 (2008) 485–516.
[217] J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, Biostatistics 9 (2008) 432–441.
[218] J. Fan, Y. Feng, Y. Wu, Network exploration via the adaptive LASSO and SCAD penalties, Ann. Appl. Stat. 3 (2009) 521–541.
[219] R. Foygel, M. Drton, Extended Bayesian information criteria for Gaussian graphical models, Adv. Neural Inf. Process. Syst. 23 23 (2010) 604–612.
[220] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, Boca Raton, FL, 2015.
[221] R. Kindermann, J.L. Snell, Markov Random Fields and Their Applications, American Mathematial Society, Providence, RI, 1980.
[222] H. Rue, L. Held, Gaussian Markov Random Fields, Chapman & Hall/CRC, New York, NY, 2005.
[223] S.Z. Li, Markov random field modeling in image analysis, third ed., Springer-Verlag, London, UK, 2009.
[224] C. Wang, N. Komodakis, N. Paragios, Markov random field modeling, inference & learning in computer vision & image understanding: A survey, Comput. Vis. Image Underst. 117 (2013) 1610–1627.

[225] J. Schäfer, K. Strimmer, An empirical Bayes approach to inferring large-scale gene association networks, Bioinformatics 21 (2005) 754–764.
[226] H. Li, J. Gui, Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks, Biostatistics 7 (2006) 302–317.
[227] A. d'Aspremont, O. Banerjee, L. El Ghaoui, First-order methods for sparse covariance selection, SIAM J. Matrix Anal. Appl. 30 (2008) 56–66.
[228] N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the Lasso, Ann. Stat. 34 (2006) 1436–1462.
[229] J. Peng, P. Wang, N. Zhou, J. Zhu, Partial correlation estimation by joint sparse regression models, J. Amer. Statist. Assoc. 104 (2009) 735–746.
[230] M. Hinne, L. Ambrogioni, R.J. Janssen, T. Heskes, M.A.J. van Gerven, Structurally-informed Bayesian functional connectivity analysis, NeuroImage 86 (2014) 294–305.
[231] A. Atay-Kayis, H. Massam, A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models, Biometrika 92 (2005) 317–335.
[232] A. Dobra, A. Lenkoski, A. Rodriguez, Bayesian inference for general Gaussian graphical models with application to multivariate lattice data, J. Amer. Statist. Assoc. 106 (2011) 1418–1433.
[233] H. Liu, J. Lafferty, L. Wasserman, The nonparanormal: Semiparametric estimation of high dimensional undirected graphs, J. Mach. Learn. Res. 10 (2009) 2295–2328.
[234] H. Liu, F. Han, M. Yuan, J. Lafferty, L. Wasserman, High-dimensional semiparametric Gaussian copula graphical models, Ann. Stat. 40 (2012) 2293–2326.
[235] L. Xue, H. Zou, Regularized rank-based estimation of high-dimensional nonparanormal graphical models, Ann. Stat. 40 (2012) 2541–2571.
[236] R.E. Morrison, R. Baptista, Y. Marzouk, Beyond normality: Learning sparse probabilistic graphical models in the non-Gaussian setting, Adv. Neural Inf. Process. Syst. 31 (2017) 2356–2366.
[237] R. Baptista, R. Morrison, O. Zahm, Y. Marzouk, Learning non-Gaussian graphical models via hessian scores and triangular transport, J. Mach. Learn. Res. 25 (2024) 85.
[238] T. Mora, W. Bialek, Are biological systems poised at criticality? J. Stat. Phys. 144 (2011) 268–302.
[239] R.R. Stein, D.S. Marks, C. Sander, Inferring pairwise interactions from biological data using maximum-entropy probability models, PLoS Comput. Biol. 11 (2015) e1004182.
[240] H.C. Nguyen, R. Zecchina, J. Berg, Inverse statistical problems: from the inverse ising problem to data science, Adv. Phys. 66 (2017) 197–261.
[241] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Machine learning and the physical sciences, Rev. Modern Phys. 91 (2019) 045002.
[242] P. Mehta, M. Bukov, C.-H. Wang, A.G.R. Day, C. Richardson, C.K. Fisher, D.J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. 810 (2019) 1–124.
[243] J. Bento, A. Montanari, Which graphical models are difficult to learn? in: Proceedings of the 22nd International Conference on Neural Information Processing Systems, 2009, pp. 1303–1311.
[244] P. Ravikumar, M.J. Wainwright, J.D. Lafferty, High-dimensional ising model selection using ℓ_1-regularized logistic regression, Ann. Stat. 38 (2010) 1287–1319.
[245] E. Aurell, M. Ekeberg, Inverse Ising inference using all the data, Phys. Rev. Lett. 108 (2012) 090201.
[246] A. Decelle, F. Ricci-Tersenghi, Pseudolikelihood decimation algorithm improving the inference of the interaction network in a general class of Ising models, Phys. Rev. Lett. 112 (2014) 070603.
[247] J.Z. Huang, N. Liu, M. Pourahmadi, L. Liu, Covariance matrix selection and estimation via penalised normal likelihood, Biometrika 93 (2006) 85–98.
[248] J. Bien, R.J. Tibshirani, Sparse estimation of a covariance matrix, Biometrika 98 (2011) 807–820.
[249] H. Liu, L. Wang, T. Zhao, Sparse covariance matrix estimation with eigenvalue constraints, J. Comput. Graph. Stat. 23 (2014) 439–459.
[250] H. Wang, Coordinate descent algorithm for covariance graphical lasso, Stat. Comput. 24 (2014) 521–529.
[251] H. Wang, Scaling it up: Stochastic search structure learning in graphical models, Bayesian Anal. 10 (2015) 351–377.
[252] S. Kojaku, N. Masuda, Constructing networks by filtering correlation matrices: A null model approach, Proc. R. Soc. A 475 (2019) 20190578.
[253] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. (2001) 1165–1188.
[254] D.S. Bassett, N.F. Wymbs, M.A. Porter, P.J. Mucha, J.M. Carlson, S.T. Grafton, Dynamic reconfiguration of human brain networks during learning, Proc. Natl. Acad. Sci. USA 108 (2011) 7641–7646.
[255] S. Achard, R. Salvador, B. Whitcher, J. Suckling, E. Bullmore, A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs, J. Neurosci. 26 (2006) 63–72.
[256] J. Kang, F.D. Bowman, H. Mayberg, H. Liu, A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs, NeuroImage 141 (2016) 431–441.
[257] B. Ma, Y. Wang, S. Ye, S. Liu, E. Stirling, J.A. Gilbert, K. Faust, R. Knight, J.K. Jansson, C. Cardona, L. Röttjers, J. Xu, Earth microbial co-occurrence network reveals interconnection pattern across microbiomes, Microbiome 8 (2020) 82.
[258] A. Zalesky, A. Fornito, E.T. Bullmore, Network-based statistic: Identifying differences in brain networks, NeuroImage 53 (2010) 1197–1207.
[259] H.C. Baggio, A. Abos, B. Segura, A. Campabadal, A. Garcia-Diaz, C. Uribe, Y. Compta, M.J. Marti, F. Valldeoriola, C. Junque, Statistical inference in brain graphs using threshold-free network-based statistics, Hum. Brain Mapp. 39 (2018) 2289–2302.
[260] N. Masuda, R. Lambiotte, A Guide to Temporal Networks, second ed., World Scientific, Singapore, 2020.
[261] P. Holme, J. Saramäki, Temporal networks, Phys. Rep. 519 (2012) 97–125.
[262] P. Holme, Modern temporal network theory: A colloquium, Eur. Phys. J. B 88 (2015) 234.
[263] R. Hindriks, M.H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N.K. Logothetis, G. Deco, Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? NeuroImage 127 (2016) 242–256.
[264] Z. Adams, R. Füss, T. Glück, Are correlations constant? Empirical and theoretical results on popular correlation models in finance, J. Bank. Financ. 84 (2017) 9–24.
[265] M.A. Lindquist, Y. Xu, M.B. Nebel, B.S. Caffo, Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach, NeuroImage 101 (2014) 531–546.
[266] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Phys. Rev. E 68 (2003) 056110.
[267] J.-P. Onnela, K. Kaski, J. Kertész, Clustering and information in correlation based financial networks, Eur. Phys. J. B 38 (2004) 353–362.
[268] R.M. Hutchison, T. Womelsdorf, E.A. Allen, P.A. Bandettini, V.D. Calhoun, M. Corbetta, S. Della Penna, J.H. Duyn, G.H. Glover, J. Gonzalez-Castillo, D.A. Handwerker, S. Keilholz, V. Kiviniemi, D.A. Leopold, F. de Pasquale, O. Sporns, M. Walter, C. Chang, Dynamic functional connectivity: Promise, issues, and interpretations, NeuroImage 80 (2013) 360–378.
[269] V.D. Calhoun, R. Miller, G. Pearlson, T. Adalı, The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery, Neuron 84 (2014) 262–274.
[270] M. Filippi, E.G. Spinelli, C. Cividini, F. Agosta, Resting state dynamic functional connectivity in neurodegenerative conditions: A review of magnetic resonance imaging findings, Front. Neurosci. 13 (2019) 657.

[271] D.J. Lurie, D. Kessler, D.S. Bassett, R.F. Betzel, M. Breakspear, S. Kheilholz, A. Kucyi, R. Liégeois, M.A. Lindquist, A.R. McIntosh, R.A. Poldrack, J.M. Shine, W.H. Thompson, N.Z. Bielczyk, L. Douw, D. Kraft, R.L. Miller, M. Muthuraman, L. Pasquini, A. Razi, D. Vidaurre, H. Xie, V.D. Calhoun, Questions and controversies in the study of time-varying functional connectivity in resting fMRI, Netw. Neurosci. 4 (2020) 30–69.
[272] D.S. Bassett, N.F. Wymbs, M.P. Rombach, M.A. Porter, P.J. Mucha, S.T. Grafton, Task-based core-periphery organization of human brain dynamics, PLoS Comput. Biol. 9 (2013) e1003171.
[273] U. Braun, A. Schäfer, H. Walter, S. Erk, N. Romanczuk-Seiferth, L. Haddad, J.I. Schweiger, O. Grimm, A. Heinz, H. Tost, A. Meyer-Lindenberg, D.S. Bassett, Dynamic reconfiguration of frontal brain networks during executive cognition in humans, Proc. Natl. Acad. Sci. USA 112 (2015) 11678–11683.
[274] J. Nakajima, M. West, Bayesian analysis of latent threshold dynamic models, J. Busi. Econ. Stat. 31 (2013) 151–164.
[275] J. Nakajima, M. West, Dynamic network signal processing using latent threshold models, Digit. Signal Proc. 47 (2015) 5–16.
[276] D. Hallac, Y. Park, S. Boyd, J. Leskovec, Network inference via the time-varying graphical lasso, in: Proceedings of the 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 205–213.
[277] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[278] N. Masuda, S. Kojaku, Y. Sano, Configuration model for correlation matrices preserving the node strength, Phys. Rev. E 98 (2018) 012312.
[279] F. Váša, B. Mišić, Null models in network neuroscience, Nat. Rev. Neurosci. 23 (2022) 493–504.
[280] W. Böhm, K. Hornik, Generating random correlation matrices by the simple rejection method: Why it does not work, Statist. Probab. Lett. 87 (2014) 27–30.
[281] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, Phys. D 58 (1992) 77–94.
[282] T. Schreiber, A. Schmitz, Surrogate time series, Phys. D 142 (2000) 346–382.
[283] H. Iyetomi, Y. Nakayama, H. Yoshikawa, H. Aoyama, Y. Fujiwara, Y. Ikeda, W. Souma, What causes business cycles? Analysis of the Japanese industrial production data, J. Jpn. Int. Econ. 25 (2011) 246–272.
[284] H. Iyetomi, Y. Nakayama, H. Aoyama, Y. Fujiwara, Y. Ikeda, W. Souma, Fluctuation-dissipation theory of input-output interindustrial relations, Phys. Rev. E 83 (2011) 016103.
[285] M. Shinn, A. Hu, L. Turner, S. Noble, K.H. Preller, J.L. Ji, F. Moujaes, S. Achard, D. Scheinost, R.T. Constable, J.H. Krystal, F.X. Vollenweider, D. Lee, A. Anticevic, E.T. Bullmore, J.D. Murray, Functional brain networks reflect spatial and temporal autocorrelation, Nat. Neurosci. 26 (2023) 867–878.
[286] M. Hirschberger, Y. Qi, R.E. Steuer, Randomly generating portfolio-selection covariance matrices with specified distributional characteristics, European J. Oper. Res. 177 (2007) 1610–1625.
[287] A. Fornito, A. Zalesky, M. Breakspear, Graph analysis of the human connectome: Promise, progress, and pitfalls, NeuroImage 80 (2013) 426–444.
[288] B.K. Fosdick, D.B. Larremore, J. Nishimura, J. Ugander, Configuring random graph models with fixed degree sequences, SIAM Rev. 60 (2018) 315–355.
[289] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827.
[290] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2010) 75–174.
[291] V. Colizza, M.A.S. A. Flammini, A. Vespignani, Detecting rich-club ordering in complex networks, Nat. Phys. 2 (2006) 110–115.
[292] S. Kojaku, N. Masuda, Core-periphery structure requires something else in the network, New J. Phys. 20 (2018) 043012.
[293] P. Expert, T.S. Evans, V.D. Blondel, R. Lambiotte, Uncovering space-independent communities in spatial networks, Proc. Natl. Acad. Sci. USA 108 (2011) 7663–7668.
[294] T. Squartini, R. Mastrandrea, D. Garlaschelli, Unbiased sampling of network ensembles, New J. Phys. 17 (2015) 023052.
[295] T. Squartini, D. Garlaschelli, Maximum-Entropy Networks, Springer, Cham, Switzerland, 2017.
[296] E. Valdano, A. Arenas, Exact rank reduction of network models, Phys. Rev. X 9 (2019) 031050.
[297] M.E.J. Newman, Networks, second ed., Oxford University Press, Oxford, UK, 2018.
[298] A. Almog, M.R. Buijink, O. Roethler, S. Michel, J.H. Meijer, J.H.T. Rohling, D. Garlaschelli, Uncovering functional signature in neural systems via random matrix theory, PLoS Comput. Biol. 15 (2019) e1006934.
[299] C. Lucibello, M. Ibàñez-Berganza, Covariance estimators, 2024, GitHub. (Accessed 24 September 2024), https://github.com/CarloLucibello/covariance-estimators.
[300] G. Caldarelli, Scale-Free Networks, Oxford University Press, Oxford, UK, 2007.
[301] A.L. Barabási, Network Science, Cambridge University Press, Cambridge, UK, 2016.
[302] F. Menczer, S. Fortunato, C.A. Davis, A First Course in Network Science, Cambridge University Press, Cambridge, UK, 2020.
[303] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (2004) 3747–3752.
[304] V.M. Eguíluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, A.V. Apkarian, Scale-free brain functional networks, Phys. Rev. Lett. 94 (2005) 018102.
[305] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. 4 (2005) 17.
[306] D.S. Bassett, E. Bullmore, B.A. Verchinski, V.S. Mattay, D.R. Weinberger, A. Meyer-Lindenberg, Hierarchical organization of human cortical networks in health and schizophrenia, J. Neurosci. 28 (2008) 9239–9248.
[307] M.A. Porter, J.-P. Onnela, P.J. Mucha, Communities in networks, Not. the AMS 56 (2009) 1082–1097, 1164–1166.
[308] M. MacMahon, Random matrix theory (RMT) filtering of financial time series for community detection, 2024, MATLAB Central File Exchange. (Accessed 15 July 2024). http://www.mathworks.com/matlabcentral/fileexchange/49011-random-matrix-theory-rmt-filtering-of-financial-time-series-for-community-detection.
[309] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, J. Kertész, Generalizations of the clustering coefficient to weighted complex networks, Phys. Rev. E 75 (2007) 027105.
[310] Y. Wang, E. Ghumare, R. Vandenberghe, P. Dupont, Comparison of different generalizations of clustering coefficient and local efficiency for weighted undirected graphs, Neural Comput. 29 (2017) 313–331.
[311] N. Masuda, M. Sakaki, T. Ezaki, T. Watanabe, Clustering coefficients for correlation networks, Front. Neuroinfo. 12 (2018) 7.
[312] Phiclust: A clusterability measure for scRNA-seq data, 2023, (Accessed 6 November 2023).https://github.com/semraulab/phiclust.
[313] M. Mircea, M. Hochane, X. Fan, S.M. Chuva de Sousa Lopes, D. Garlaschelli, S. Semrau, Phiclust: A clusterability measure for single-cell transcriptomics reveals phenotypic subpopulations, 2023, (Accessed 13 November 2023), https://zenodo.org/record/5785793#.Ybs5wn3MK3I.
[314] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: Uses and interpretations, NeuroImage 52 (2010) 1059–1069.
[315] T.P. Peixoto, The graph-tool python library, 2023, http://dx.doi.org/10.6084/m9.figshare.1164194, (Accessed 14 November 2023), http://figshare.com/articles/graph_tool/1164194.
[316] M.O. Kuismin, M.J. Sillanpää, Estimation of covariance and precision matrix, network structure, and a view toward systems biology, WIREs Comput. Stat. 9 (2017) e1415.
[317] S. Epskamp, A.O.J. Cramer, L.J. Waldorp, V.D. Schmittmann, D. Borsboom, qgraph: Network visualizations of relationships in psychometric data, J. Stat. Softw. 48 (2012) 1–18.

[318] R.R. Gabriel, Estimating a psychometric network with qgraph, 2023, (Accessed 24 August 2023) https://reisrgabriel.com/blog/2021-08-10-psych-network/.
[319] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[320] Scikit-learn package, 2022, (Accessed 18 May 2022), https://scikit-learn.org/stable/install.html.
[321] M.W. Cole, T. Yarkoni, G. Repovs, A. Anticevic, T.S. Braver, Global connectivity of prefrontal cortex predicts cognitive control and intelligence, J. Neurosci. 32 (2012) 8988–8999.
[322] E. Santarnecchi, G. Galli, N.R. Polizzotto, A. Rossi, S. Rossi, Efficiency of weak brain connections support general cognitive functioning, Hum. Brain Mapp. 35 (2014) 4566–4582.
[323] J.G. Young, G.T. Cantwell, M.E.J. Newman, Bayesian inference of network structure from unreliable data, J. Compl. Netw. 8 (2020) cnaa046.
[324] S. Raimondo, M. De Domenico, Measuring topological descriptors of complex networks under uncertainty, Phys. Rev. E 103 (2021) 022311.
[325] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. 544 (2014) 1–122.
[326] M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks, J. Comp. Netw. 2 (2014) 203–271.
[327] G. Bianconi, Multilayer Networks, Oxford University Press, Oxford, UK, 2018.
[328] O. Artime, B. Benigni, G. Bertagnolli, V. d'Andrea, R. Gallotti, A. Ghavasieh, S. Raimondo, M. De Domenico, Multilayer network science, Cambridge University Press, Cambridge, UK, 2022.
[329] M. De Domenico, More is different in real-world multilayer networks, Nat. Phys. 19 (2023) 1247–1262.
[330] M.J. Brookes, P.K. Tewarie, B.A.E. Hunt, S.E. Robson, L.E. Gascoyne, E.B. Liddle, P.F. Liddle, P.G. Morris, A multi-layer network approach to MEG connectivity analysis, NeuroImage 132 (2016) 425–438.
[331] P. Tewarie, A. Hillebrand, B.W. van Dijk, C.J. Stam, G.C. O'Neill, P. Van Mieghem, J.M. Meier, M.W. Woolrich, P.G. Morris, M.J. Brookes, Integrating cross-frequency and within band functional networks in resting-state MEG: A multi-layer network approach, NeuroImage 142 (2016) 324–336.
[332] J.M. Buldú, M.A. Porter, Frequency-based brain networks: From a multiplex framework to a full multilayer description, Netw. Neurosci. 2 (2017) 418–441.
[333] M. De Domenico, S. Sasai, A. Arenas, Mapping multiplex hubs in human functional brain networks, Front. Neurosci. 10 (2016) 326.
[334] M. De Domenico, Multilayer modeling and analysis of human brain networks, GigaScience 6 (2017) gix004.
[335] R. Dorantes-Gilardi, D. García-Cortés, E. Hernández-Lemus, J. Espinal-Enríquez, Multilayer approach reveals organizational principles disrupted in breast cancer co-expression networks, Appl. Netw. Sci. 5 (2020) 47.
[336] M. Russell, A. Aqil, M. Saitou, O. Gokcumen, N. Masuda, Gene communities in co-expression networks across different tissues, PLoS Comput. Biol. 19 (2023) e1011616.
[337] N. Musmeci, V. Nicosia, T. Aste, T. Di Matteo, V. Latora, The multiplex dependency structure of financial markets, Complexity 2017 (2017) 9586064.
[338] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (2010) 876–878.
[339] D. Taylor, S. Shai, N. Stanley, P.J. Mucha, Enhanced detectability of community structure in multilayer networks through layer aggregation, Phys. Rev. Lett. 116 (2016) 228301.
[340] D. Taylor, R.S. Caceres, P.J. Mucha, Super-resolution community detection for layer-aggregated multilayer networks, Phys. Rev. X 7 (2017) 031056.
[341] M. Venkatesh, J. Jaja, L. Pessoa, Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification, NeuroImage 207 (2020) 116398.
[342] K. You, H.-J. Park, Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity, NeuroImage 225 (2021) 117464.
[343] K. Abbas, M. Liu, M. Venkatesh, E. Amico, A.D. Kaplan, M. Ventresca, L. Pessoa, J. Harezlak, J. Goñi, Geodesic distance on optimally regularized functional connectomes uncovers individual fingerprints, Brain Conn. 11 (2021) 333–348.
[344] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis. 66 (2006) 41–66.
[345] M. Rahim, B. Thirion, G. Varoquaux, Population shrinkage of covariance (PoSCE) for better individual brain functional-connectivity estimation, Med. Image Anal. 54 (2019) 138–148.
[346] M. Latapy, C. Magnien, N. Del Vecchio, Basic notions for the analysis of large two-mode networks, Soc. Netw. 30 (2008) 31–48.
[347] M.E.J. Newman, Scientific collaboration networks. I. Network construction and fundamental results, Phys. Rev. E 64 (2001) 016131.
[348] J.-L. Guillaume, M. Latapy, Bipartite structure of all complex networks, Info. Proc. Lett. 90 (2004) 215–221.
[349] J.J. Ramasco, S.N. Dorogovtsev, R. Pastor-Satorras, Self-organization of collaboration networks, Phys. Rev. E. 70 (2004) 036106.
[350] A. Gallo, F. Saracco, T. Squartini, Statistically validated projection of bipartite signed networks, 2025, Preprint https://arxiv.org/abs/2502.08567v2.
[351] L. Gallo, L. Lacasa, V. Latora, F. Battiston, Higher-order correlations reveal complex memory in temporal hypergraphs, Nat. Commun. 15 (2024) 4754.
[352] A. Santoro, F. Battiston, G. Petri, E. Amico, Higher-order organization of multivariate time series, Nat. Phys. 19 (2023) 221–229.
[353] A. Santoro, F. Battiston, M. Lucas, G. Petri, E. Amico, Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior, Nat. Commun. 15 (2024) 10244.
[354] S. Amari, H. Nakahara, S. Wu, Y. Sakai, Synchronous firing and higher-order interactions in neuron pool, Neural Comput. 15 (2003) 127–142.
[355] E. Schneidman, S. Still, M.J. Berry, W. Bialek, Network information and connected correlations, Phys. Rev. Lett. 91 (2003) 238701.
[356] S. Yu, H. Yang, H. Nakahara, G.S. Santos, D. Nikolić, D. Plenz, Higher-order interactions characterized in cortical activity, J. Neurosci. 31 (2011) 17514–17526.
[357] H. Shimazaki, S.-i. Amari, E.N. Brown, S. Grün, State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data, PLoS Comput. Biol. 8 (2012) e1002385.
[358] T.F. Varley, M. Pope, J. Faskowitz, O. Sporns, Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex, Commun. Biology 6 (2023) 451.
[359] T.F. Varley, M. Pope, M. Grazia Puxeddu, J. Faskowitz, O. Sporns, Partial entropy decomposition reveals higher-order information structures in human brain activity, Proc. Natl. Acad. Sci. USA 120 (2023) e2300888120.
[360] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.G. Young, G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Phys. Rep. 874 (2020) 1–92.
[361] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M.M. Murray, T.P. Peixoto, F. Vaccarino, G. Petri, The physics of higher-order interactions in complex systems, Nat. Phys. 17 (2021) 1093–1098.
[362] G. Bianconi, Higher-Order Networks, Cambridge University Press, Cambridge, UK, 2021.
[363] L. Lovász, Large Networks and Graph Limits, American Mathematical Society, Providence, RI, 2012.
[364] G. Bianconi, A.-L. Barabási, Bose-Einstein condensation in complex networks, Phys. Rev. Lett. 86 (2001) 5632–5635.

[365] G. Bianconi, A.-L. Barabási, Competition and multiscaling in evolving networks, Europhys. Lett. 54 (2001) 436–442.
[366] K.-I. Goh, B. Kahng, D. Kim, Universal behavior of load distribution in scale-free networks, Phys. Rev. Lett. 87 (2001) 278701.
[367] G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Muñoz, Scale-free networks from varying vertex intrinsic fitness, Phys. Rev. Lett. 89 (2002) 258702.
[368] M. Boguna, R. Pastor-Satorras, Class of correlated random networks with hidden variables, Phys. Rev. E 68 (2003) 036112.
[369] N. Masuda, H. Miwa, N. Konno, Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights, Phys. Rev. E 70 (2004) 036124.
[370] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks., Sci. Rep. 2 (2012) 469.
[371] N. Masuda, H. Miwa, N. Konno, Geographical threshold graphs with small-world and scale-free properties, Phys. Rev. E 71 (2005) 036108.
[372] N. Masuda, N. Konno, VIP-club phenomenon: Emergence of elites and masterminds in social networks, Soc. Netw. 28 (2006) 297–309.
[373] G.T. Cantwell, Y. Liu, B.F. Maier, A.C. Schwarze, C.A. Serván, J. Snyder, G. St-Onge, Thresholding normally distributed data creates complex networks, Phys. Rev. E 101 (2020) 062302.
[374] M.A. Porter, S.D. Howison, The role of network analysis in industrial and applied mathematics, 2018, Preprint https://arxiv.org/abs/1703.06843v3.