Variance and covariance of distributions on graphs

Karel Devriendt*
Mathematical Institute, University of Oxford, Oxford UK and Alan Turing Institute, London, UK

Samuel Martin-Gutierrez
Grupo de Sistemas Complejos, Universidad Politcnica de Madrid, Madrid, Spain

Renaud Lambiotte
Mathematical Institute, University of Oxford, Oxford UK

(Dated: August 24, 2020)

We develop a theory to measure the variance and covariance of probability distributions defined on the nodes of a graph, which takes into account the distance between nodes. Our approach generalizes the usual (co)variance to the setting of weighted graphs and retains many of its intuitive and desired properties. Interestingly, we find that a number of famous concepts in graph theory and network science can be reinterpreted in this setting as variances and covariances of particular distributions: we show this correspondence for Kemeny's constant, the Kirchhoff index, network modularity and Markov stability. As a particular application, we define the *maximum-variance problem* on graphs with respect to the effective resistance distance, and characterize the solutions to this problem both numerically and theoretically. We show how the maximum-variance distribution can be interpreted as a core-periphery measure, illustrated by the fact that these distributions are supported on the leaf nodes of tree graphs, low-degree nodes in a configuration-like graph and boundary nodes in random geometric graphs. Our theoretical results are supported by a number of experiments on a network of mathematical concepts, where we use the variance and covariance as analytical tools to study the (co-)occurrence of concepts in scientific papers with respect to the (network) relations between these concepts.

I. INTRODUCTION

The variance of a probability distribution is a fundamental concept in the toolkit of probability theory and statistics and is routinely applied throughout science, engineering and numerous practical settings. Intuitively speaking, the variance captures how spread-out the outcomes of a distribution are, and thus reflects the inherent variability in this distribution. In many practical cases however, probability distributions are defined on the nodes of a network: websites on the internet, individuals in a social network, neurons in the brain, etc. These nodes are the building blocks of a network, and when studying distributions or signals defined on nodes it is natural to take the underlying network structure into account. As the usual definition of variance can not take this structure into account, we thus lack a basic methodological tool when analysing distributions and signals on a graph.

In this article, we propose a measure of variance and covariance for distributions defined on a network, which take into account the underlying structure of the network by considering the distances between nodes. These distances provide a notion of what it means to be 'spread out' on the network, which in turn allows to define (co)variances of distributions on the network. Our proposed formulas for variance and covariance take a very simple mathematical form (as a quadratic product and matrix trace, respectively) yet still capture many of the intuitive and mathematical properties of the usual (co)variance. To illustrate our new measures in practice, we apply the proposed variance and covariance measures to the analysis of an empirical network of mathematical concepts with data from a collection of scientific papers. Our approach allows for a unified and intuitive treatment of the structural (relations between concepts) and functional data (usage of concepts in papers) in this system and we describe some qualitative and quantitative findings. As a second application, we show that the variance and covariance of some particular distributions correspond to previously known graph characteristics. In particular, we find an interpretation of Kemeny's constant and the Kirchhoff index (effective graph resistance) in terms of graph variances, and an interpretation of network modularity and Markov stability in terms of graph covariances. These correspondences offer a new interpretation and framework to understand these graph concepts, and provides theoretical support for the specific form of our introduced (co)variance measures.

We furthermore consider the *maximum variance problem* which asks to determine the largest possible variance on a given graph, and to characterize the distribution(s) that attains it. As an important theoretical contribution, we find a complete characterization of the maximum variance distribution when considering the *effective resistances* as a distance measure between nodes. Interestingly, we find an intuitive interpretation of this

* corresponding author: devriendt@maths.ox.ac.uk

maximum variance distribution (and its support) as a core-periphery measure. To illustrate this interpretation, we show that the maximum variance distribution can be used to identify the leaf nodes of a tree graph, the low-degree nodes in the case of a degree-determined (configuration) graph and boundary nodes in random geometric graphs. Finally, we provide a number of equivalent characterizations for the equations that determine the maximum variance distribution, which hints towards an algebraic graph characteristic with a broader applicability.

To the best of our knowledge, the introduced measures for variance and covariance are new in the context of distributions on networks. The effective resistance matrix however is very well-studied [1–3] and our variance quadratic form (3) and the related maximum variance problem (5) were defined and studied before in different contexts. Hjorth et al. [4] studied expression (5) for general metric spaces as a generalization of the diameter of these spaces and described the conditions for which this problem has a unique solution. Dankelmann defines the quadratic form in (3) for the geodesic graph distance and solves the corresponding maximum problem for tree and cycle graphs. These results are generalized from the geodesic graph distance to the effective resistance in [5, 6], where it is shown that the corresponding maximum variance problem can be solved efficiently.

The rest of this article is organized as follows: In Section II we introduce the relevant mathematical background about graphs and (joint) distributions on graphs and introduce our new variance and covariance measures. Section III treats an application of the (co)variance measures to a network of mathematical concepts. In Section IV, we show how our new measures relate to a number of existing concepts. In Section V we introduce and solve the maximum variance problem on a graph, with respect to the effective resistance distance. We characterize the maximum variance solutions, with a particular focus on the support of the the maximum variance distributions. In Section VI finally, we list a number of equivalent characterizations of 'locally optimal solutions' and conclude the article in Section VII with a summary of the results and outlook to further possible applications.

II. VARIANCE AND COVARIANCE ON GRAPHS

A. preliminaries

We start by introducing a number of preliminary notions about graphs and probability theory. A *graph* G consists of a set of n nodes $\mathcal{N}$ and a set of links $\mathcal{L}$ that connect pairs of nodes; a link between two nodes i and j thus corresponds to an element $(i,j) \in \mathcal{L}$. Every link is furthermore assigned a (positive, real) weight $c_{ij} > 0$, resulting in a *weighted graph*. Since graphs can be seen as weighted graphs where $c_{ij} = 1$ for all links, we will further work in the more general setting of weighted graphs. Furthermore, we will assume all graphs to be finite ($n < \infty$) and connected, which means that there is a path between every pair of nodes.

A *distribution on a graph* is a function $p : \mathcal{N} \to [0,1]$ which assigns a nonnegative number $p(i)$ to each node in the graph, such that all number add up to one $\sum_{i\in\mathcal{N}} p(i) = 1$. Distributions can be used to define a *random node* N, which is a random element of the node set, with probability to equal any of the nodes i given by the corresponding distribution $p(i)$. In other words, we can 'sample' the random node N and get any of the nodes of G with probability

$$\Pr[N = i] = p(i) \text{ for all } i \in \mathcal{N}.$$

For this reason, the value $p(i)$ of an element is called the *probability* of i. To specify the underlying distribution of a random node N, we often write $N \sim p$ and say that N is distributed according to p.

A *joint distribution on a graph* is a function $P : \mathcal{N}\times\mathcal{N} \to [0,1]$ which assigns a nonnegative number $P(i,j)$ to each pair of nodes in the graph, such that all numbers add up to one. A joint distribution can be used to define a random pair of nodes (N,M) which are two random elements of the node set, with probability of being sampled given by the joint distribution as

$$\Pr[(N,M) = (i,j)] = P(i,j) \text{ for all } (i,j) \in \mathcal{N}\times\mathcal{N}.$$

Any joint distribution P on node pairs also naturally leads to two (simple) distributions $\tilde{p}$ and $\tilde{q}$ on the nodes, defined by $\tilde{p}(i) = \sum_{j\in\mathcal{N}} P(i,j)$ and $\tilde{q}(j) = \sum_{i\in\mathcal{N}} P(i,j)$, which are called the *marginal distributions* of P.

In Section III we consider an application of these concepts in a practical setting, which serves as an illustration of the definitions above. We will consider a graph made up of nodes that represent mathematical concepts with links that reflect conceptual relations between pairs of concepts (inferred from Wikipedia hyperlinks) and study a collection of scientific papers that use these concepts. The occurrence of a subset of concepts in a given paper is translated to a distribution on the relevant nodes (concepts) in the network, and the frequency of pairs of concepts appearing together in a paper is represented as a joint distribution on the network.

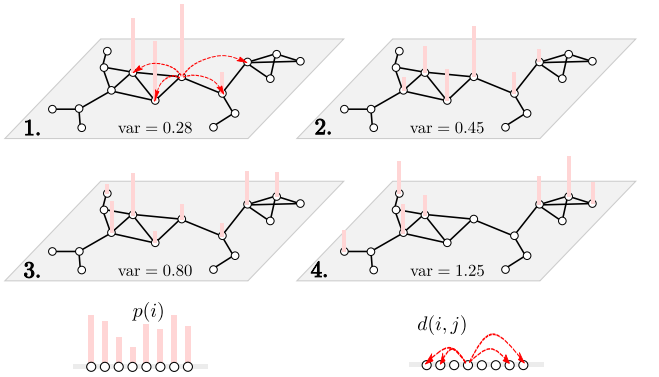
B. (co)variance with respect to node distances

As introduced, the interpretation of a graph distribution is twofold: we can consider it as a signal (function) on the nodes of a graph (see also [7]), or as representing a random node. In both cases, it is natural to ask whether a distribution is centered, or concentrated, on a small part of the graph and thus might be well understood and described by restricting our attention to this small part, or if instead the distribution is spread out over the graph. In the setting of distributions on the

real numbers (or other vector spaces), the variance is a natural quantifier of exactly these properties; it reflects how spread out a distribution is as the average squared difference between a random outcome of the distribution and the typical outcome (i.e. the mean). To generalize this standard notion of variance to graphs, we propose to take into account distances between nodes in the graph, i.e. a function $d : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$ that says something about 'how far' $d(i, j)$ two nodes i and j are in the graph. Given such a distance notion between nodes, we propose a generalization of variance to distributions on a graph:

$$\mathrm{var}(p) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} p(i)p(j)d^2(i, j) \tag{1}$$

As shown in Appendix A, definition (1) is a proper generalization of the usual variance, which is retrieved if $\mathcal{N} \subset \mathbb{R}$ with the usual Euclidean distance. We will later discuss a number of examples of distance functions that illustrate how expression (1) might be used in practice. Figure 1 illustrates how our notion of variance captures the variance of a graph distribution. In the case of joint



Distribution on the nodes
$p(i)$ = probability on i

Distance function between nodes
$d(i, j)$ = distance between i, j

FIG. 1. The variance of distributions on a graph can be defined with respect to a distance function between the nodes of the graph. This measure allows to compare how 'spread out' different distributions are on the network. In the example above, the distributions become more spread out over the network, going from distribution 1. → 4. with an increasing variance as a result (calculated via (3)).

distributions (i.e. pairs of random variables), the covariance is used much in the same way as the variance to quantify whether random pairs sampled according to this distribution are on average close together, or far apart. Making use of a distance notion d again, we propose a generalization of the covariance to joint distributions P on a graph:

$$\mathrm{cov}(P) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} \left[\tilde{p}(i)\tilde{q}(j) - P(i, j)\right] d^2(i, j) \tag{2}$$

In Appendix A we show that equation (2) indeed generalizes the usual covariance, which is retrieved if $\mathcal{N} \subset \mathbb{R}$ with the usual Euclidean distance.

C. Distance functions on graphs

Definitions (1) and (2) measure the variance and covariance of distributions on a graph with respect to a certain 'distance' between the nodes. The most famous distance on graphs is the shortest-path distance (or geodesic distance) where $d(i, j)$ is the length of the shortest path between two nodes i and j [8]. More than just capturing the intuitive notion of a distance between nodes, the geodesic distance also satisfies the mathematical properties of a metric. Classically, we say that a distance function d is *metric* (or simply, d is a metric) if it satisfies the following properties:

(i) $d(i, j) \geq 0$ for all i, j
(ii) $d(i, j) = 0$ if and only if $i = j$
(iii) $d(i, j) = d(j, i)$ for all i, j
(iv) $d(i, j) + d(j, k) \geq d(i, k)$ for all i, j, k

In other words, distances are non-negative and only zero between a node and itself, the distances are symmetric and they satisfy a triangle inequality. Metrics are an important example of distance functions on a graph, since they are extensively studied mathematically.
Another important metric between the nodes of a graph is the *effective resistance*. Similar to the geodesic distance, the effective resistance reflects the length of the paths between a pair of nodes. However, instead of only taking into account the length of the shortest path the effective resistance is influenced by all paths (and their lengths) between a pair of nodes, and becomes smaller as more paths are available. We can define the effective resistance using the Laplacian matrix of a graph, which is the $n \times n$ matrix with entries given by

$$(Q)_{ij} = \begin{cases} k_i \text{ if } i = j \\ -c_{ij} \text{ if } (i, j) \in \mathcal{L} \\ 0 \text{ otherwise} \end{cases}$$

where $k_i = \sum_{j \in \mathcal{N}} c_{ij}$ is called the *degree* of node i [9]. Importantly, for a connected graph, this Laplacian matrix is positive semidefinite with a single zero eigenvalue for the constant eigenvector $\mathbf{u} = (1, \ldots, 1)^T$ which means it has a pseudoinverse $Q^\dagger$ that satisfies $QQ^\dagger = Q^\dagger Q = I - \mathbf{u}\mathbf{u}^T/n$. Using this pseudoinverse, the *effective resistance* ω_{ij} is then defined as the quadratic product

$$\omega_{ij} \triangleq (e_i - e_j)^T Q^\dagger (e_i - e_j),$$

where the unit vectors have entries $(e_i)_k = 1$ if $k = i$ and zero otherwise. Originally defined in the context of electrical circuit, the effective resistance appears in many applications of graph theory; for an overview of properties and results of the effective resistance, we refer the readers to [1, 2, 10]. In a more applied context, the resistance distance is better known under the name commute distance, see for instance [11].
For our application, the most relevant property of effective resistances is that both $d(i, j) = \omega_{ij}$ and its square

root $d(i,j) = \sqrt{\omega_{ij}}$ are metric [2, 12], and can thus be used to measure the (co)variance of distributions. The algebraic properties of the effective resistance have the further advantage that they allow for a thorough theoretical investigation. When employing the square root effective resistance as a metric, we find that the variance and covariance can be written in matrix form as

$$\mathrm{var}_\omega(p) = \tfrac{1}{2}\mathbf{p}^T\Omega\mathbf{p}, \text{ and } \mathrm{cov}_\omega(P) = \tfrac{1}{2}\left[\tilde{\mathbf{p}}^T\Omega\tilde{\mathbf{q}} - \mathrm{tr}(P\Omega)\right] \tag{3}$$

with the $n \times n$ matrix Ω containing the effective resistances as entries $(\Omega)_{ij} = \omega_{ij}$, and with $\mathbf{p} = (p(i), \ldots, p(j))^T$ the vector containing the probability for all nodes, and similarly for $\tilde{\mathbf{p}}, \tilde{\mathbf{q}}$ and matrix $(P)_{ij} = P(i,j)$ containing the probabilities of all pairs of nodes. We will further also use p_i to denote an entry of the probability vector $\mathbf{p}$, which thus equals the probability $p(i)$.

III. VARIANCE AND COVARIANCE IN A NETWORK OF KNOWLEDGE

An often-followed approach in the study of networked and complex systems is to divide the description of the system under study into an underlying structure - which can be a network or a more general structure such as a multilayer, temporal or higher-order network - and some processes that take place in the system [13–15]. Together, these *structural* and *functional* components make up a simplified model of the system that captures its key properties. In this framework, our proposed (co)variance measures can be used as a tool to describe features of functional aspects of the networks (e.g. distributions or signals) relative to the underlying structure of the network.

As an example application, we study a 'network of knowledge' made up of mathematical ideas and results with links between related concepts. The code and data of our analysis are available on GitHub [16]. We consider a list of mathematical concepts (theorems, lemmas, equations) compiled from four Wikipedia pages that list these concepts, and we infer links between the concepts from hyperlinks between their respective Wikipedia pages. More information about the data retrieval and filtering of the data set can be found in [17], where (a higher-order variant of) this network was investigated. The resulting network of concepts consists of $n = 1150$ nodes and $m = 4109$ links and is shown in Figure 2 below. We consider this network as the underlying structure of mathematical concepts and use it to investigate how these concepts are used in practice by their occurrences in scientific papers.

To investigate the functional aspect of the network of knowledge, we use a corpus of 140k+ papers from the arXiv and the mathematical concepts used therein. For each paper we count which of the mathematical concepts appear and represent this by a uniform distribution over the used concepts. Every paper i thus has a corresponding subset of concepts $\mathcal{V}_i$ and distribution $p^{(i)}$ uniform over this set of concepts. This representation of distributions on a graph allows to further analyse the data of these papers in our introduced variance framework.

FIG. 2. Hyperlink network of Wikipedia pages of the considered mathematical concepts (see [17]). The size of the nodes is proportional to their PageRank and the color coding corresponds to communities found using the Louvain algorithm.

A first question we consider is whether the mathematical papers contain 'coherent' sets of mathematical concepts, or not. In terms of variance, this question can be addressed by comparing the variance of the paper distributions $p^{(i)}$ to a null model, representing 'virtual papers'. Figure 3 below shows that the paper distributions generally have a smaller variance, compared to virtual papers made up of randomly sampled concepts according to their relative frequency over the full corpus. Intuitively, this reflects the idea that the practical use of mathematical concepts is related to the underlying connections between concepts, where a group of concepts is more likely to be considered in a paper if this group is coherent, as measured by a small variance of their corresponding uniform distribution. Furthermore, Figure 3 clearly indicates a range of 'typical' variance values which could be used to identify papers with exceptionally small (or high) variances. As a second application, the variance of paper distributions $p^{(i)}$ could be used as a ground for (qualitative) comparison between different fields of study as shown in Figure 4. More generally, this type of comparative analysis might be useful when there are different 'modes of operation' of a single network, giving rise to different functional signals.

On an aggregate level, we can study the corpus of papers by counting the co-occurrences of pairs of concepts over all papers. This gives a joint distribution P, where $P(i,j)$ is proportional to the frequency of co-occurrence of concepts i and j. To test how the function of the network of knowledge relates to its structure, we compare the covariance of this empirical distribution P on the given graph against two null models: in the first model the distribution P is left constant but the underlying

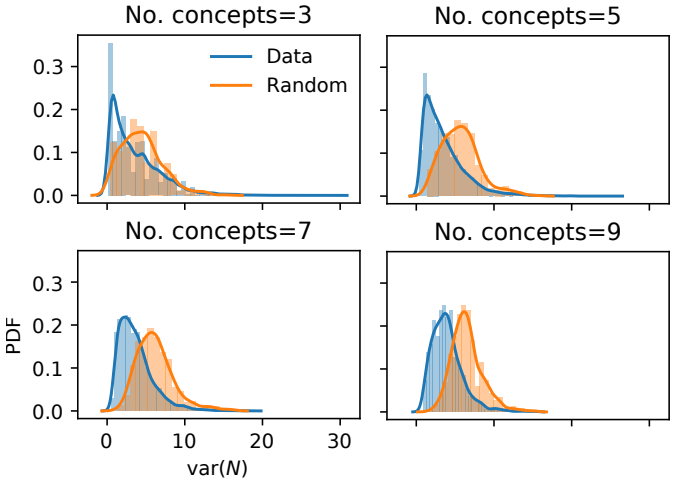


FIG. 3. Distribution of network variances of concept distributions $p^{(i)}$ in papers, calculated with respect to the geodesic distance in the Wikipedia network. In each panel, the probability density function (pdf) of variances is calculated for papers containing 3 (resp. 5,7,9) concepts present in the Wikipedia network, and is compared with the variance pdf for a collection of (null model) virtual papers with the same number of concepts. The empirical variance distributions seem to be concentrated on smaller variances compared to the null model variances.

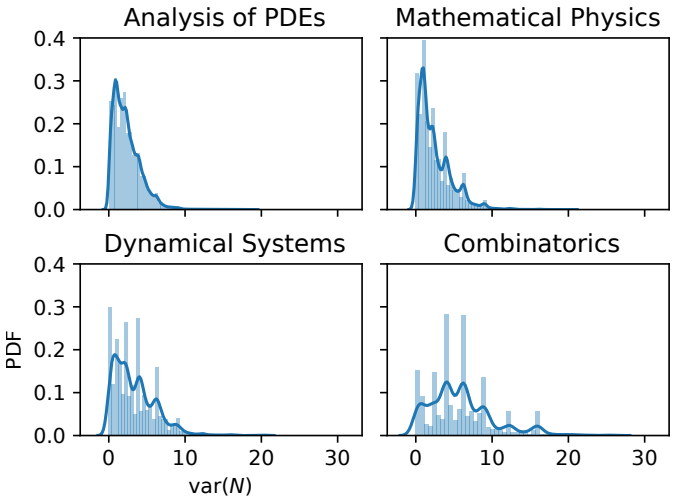


FIG. 4. Distribution of network variances of concept distributions $p^{(i)}$ for arXiv papers from different sub-fields.

graph is randomized, while in the second model the distribution P is randomized while the graph kept constant. Since we are comparing joint distributions with potentially different marginals, we consider a normalization of the covariance, which corresponds to the well-known concept of *correlation*:

$$\mathrm{corr}(P) = \frac{\mathrm{cov}(P)}{\sqrt{\mathrm{var}(\tilde{p})\,\mathrm{var}(\tilde{q})}}$$

Since the co-occurrence joint probability distribution is symmetric, we have $\tilde{p} = \tilde{q}$, which means that the correlation is simply calculated as $\mathrm{corr}(P)/\,\mathrm{var}(\tilde{p})$.

As a first null model, we perform a degree-preserving rewiring [18] of the Wikipedia network while keeping the joint distribution constant. As seen in the left panel of Figure 5, measuring the covariance of P with respect to these randomized graphs yields significantly lower correlation values; this would imply that the empirical correlation is not simply a consequence of the (local) degree distribution of the network. Our second null model consists of a marginal-preserving randomization of P while leaving the network intact: we pick two pairs of nodes (i,j) and (i',j') with non-zero joint probabilities and reshuffle their joint probabilities as

$$P \to P - \alpha(v_{ij}v_{ij}^T + v_{i'j'}v_{i'j'}^T - v_{ij'}v_{ij'}^T - v_{i'j}v_{i'j}^T)$$

where $v_{ij} = (e_i + e_j)(e_i + e_j)^T$ and with a uniform random value $\alpha < \min(P(i,j), P(i',j'))$. In other words, there is a shift of probability mass α from $(i,j) \to (i,j')$ and from $(i',j') \to (i',j)$. We repeat this procedure until all pairs of nodes have been involved, which produces a (symmetrically) randomized joint distribution P' with the same marginals as P. The right panel in Figure 5 shows that these randomized joint distributions are concentrated on significantly lower correlation values. Importantly, we remark that the empirical covariance is highly atypical for both null model classes and it is thus likely that these models are discarding too much structure to give a reliable baseline for the empirical covariance. A further development of appropriate null models for measuring the covariance on a network would be an interesting line of further research. We also calculated the co-occurrences for a number of sub-fields separately and report their covariances in Table I.

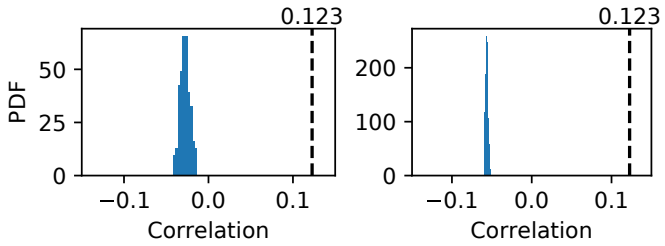


FIG. 5. Comparison between empirical covariance of the co-occurrence distribution P on the Wikipedia network of concepts and two null models. In the left panel, the covariance of P is calculated with respect to the shortest-path distance on the Wikipedia network (dashed line) and the shortest-path distance on 100 realizations of a degree-preserving randomization of the network (blue bars). In the right panel, the covariance of P (dashed line) and of 100 realizations of a marginal-preserving randomization of P (blue bars) are calculated with respect to the shortest-path distance on the Wikipedia network.

The above analysis illustrates a practical scenario where data is available in different modalities (i.e. structural and functional) and for which our variance and co-

TABLE I. Correlation of the joint probability distribution of concepts for papers of different sub-fields.

Field	Correlation
Analysis of PDEs	0.021
Mathematical Physics	0.065
Probability	0.042
Algebraic Geometry	0.051
Differential Geometry	0.054
Number Theory	0.058
Dynamical Systems	0.081
Combinatorics	0.073
Functional Analysis	0.067

variance measures enable a unified treatment of this system data. Our framework is of course not restricted to this specific example, but can be applied in the context of many other network problems in which a combination of structural and functional modalities are important. For example, in the field of neuroscience there is a lot of interest in trying to characterize the relationship between the structural network of the brain (how different areas are physically connected) and its functional network (how the signals generated by different areas are related) [19–21]. A possible approach in this setting could be to use the (normalized) functional adjacency matrix as a joint probability matrix. The covariance of this distribution with respect to the distances in the structural network could then be interpreted as a measure of correspondence between both modalities.

Another potential application can be found in social network analysis. In [22] the controversy (or social polarization) around a given issue is captured by partitioning a network of user interactions into two communities and measuring how well the highest degree nodes from both communities are connected (the lower the connectivity, the higher the polarization). Our framework could be used in this context by placing a uniform distribution over the high degree nodes and using the corresponding network variance as a proxy of polarization, since it directly measures how spread-out the distribution is. Other probability distributions may also prove useful.

Economic complexity has been thoroughly studied using networks of products, industries, countries, ... which in many cases are closely interrelated. Since there is a great interest in characterizing the diversity of such economic networks [23, 24], the network variance and covariance might be vital tools for this discipline. As an example, the variance of the distribution of exported products of a given country could be computed with respect to distances in the network of co-exported products, as defined in [25]; the resulting quantity can be interpreted as a measure of economic diversity of the country.

IV. CORRESPONDENCE BETWEEN (CO)VARIANCES AND EXISTING GRAPH MEASURES

Apart from using the variance to quantify the spread of distributions in real-world applications, we can also further study equations (1) and (2) theoretically. A first interesting result is that our proposed variance and covariance measures correspond to a number of existing concepts in graph theory and network science. This provides a new, probabilistic perspective on these concepts which adds to their understanding.

A. Kemeny's constant

A *random walk* on a graph is a process in which a 'random walker' makes its way across the nodes of a network by following the links in the network with certain probabilities. More precisely, a random walk describes a sequence of random nodes $\{N_t\}_{t\in\mathbb{N}}$ where the consecutive random nodes N_t - which represent the position of the random walker at timestep t - are related via a transition matrix T as

$$\Pr(N_{t+1} = j | N_t = i) = T_{ji}.$$

Taking into account that the walker can only cross links, we have that $T_{ji} > 0$ only if $(i, j) \in \mathcal{L}$, while conservation of probability requires that $\sum_{j\in\mathcal{N}} T_{ji} = 1$. Typically, we furthermore assume that at any node, the likelihood of the random walker to cross any of the possible links is proportional to its weight, which gives $T_{ji} = c_{ij}/k_i$. For this type of random walk on a connected graph, the distribution of N_t converges to a unique stationary distribution π, with probabilities $\pi(i) = k_i/(2m)$ where $m = \frac{1}{2}\sum_{i=1}^{n} k_i$, independently of the distribution of the initial position of the random walker N_0.
There are numerous interesting questions one can ask about random walks on a graph, with often very satisfying answers [26, 27]. One line of questions considers the time (number of steps) it takes for a random walker to travel through the graph in a certain way, e.g. from one node to another (hitting time), back and forth between two nodes (commute time) or visiting all nodes (covering time). The answers to these questions are all related to the effective resistance in some way, see for instance [26, 28–30]. Recently, another relation was discovered between the effective resistance and a random-walk characteristic. In [31], Kemeny discovered that the time κ_i it takes on average for a random walker to go from node i to a random node $J \sim \pi$ (with the stationary distribution) is independent of the node under consideration; in other words $\kappa_i = \kappa$ is a constant independent of i, which is now called Kemeny's constant. In [32], a new formula for Kemeny's constant was found in terms of the effective resistances of the underlying graph as

$$\kappa = m\pi^T \Omega \pi. \tag{4}$$

Clearly, this expression is closely related to our definition of the variance (3) with respect to the effective resistance; consequently, we find the following correspondence:

Proposition 1 *Kemeny's constant κ of a random walker is proportional to the steady-state variance of that random walker, measured with respect to the square root effective resistance, as*

$$\kappa = 2m \operatorname{var}_\omega(\pi)$$

Proof: Proposition 1 follows from expression (3) for the graph variance with respect to the square root effective resistance, and expression (4) derived in [32]. □

Proposition 1 follows directly from our definition of the graph variance, but has some interesting consequences. Firstly, it provides a new interpretation of the Kemeny constant as the variance of a distribution. This allows, for instance, to derive new bounds for the Kemeny constant based on the maximum variance results in Section V. Conversely, Kemeny's constant can serve as a reference point for the variance of other distributions on a graph; if a distribution has a higher (lower) variance than $\kappa/2m$ then we know that this distribution is more (resp. less) spread out than the stationary random walk distribution.

B. Kirchhoff index/effective graph resistance

Apart from serving as a graph metric, the effective resistance has been used in a variety of applications in different domains. An important example is the sum of all effective resistances

$$R_G = \frac{1}{2} \sum_{i,j \in \mathcal{N}} \omega_{ij}$$

which is used in mathematical chemistry as a 'fingerprint' of a network that represents a chemical compound - here, R_G is called the Kirchhoff index [2, 33] - and is used as a heuristic measure of robustness and connectedness in the characterization of (empirical) networks - here, it is called the effective graph resistance [10, 34]. Again, from expression (3) for the graph variance, we immediately find the following correspondence:

Proposition 2 *The Kirchhoff index or effective graph resistance R_G of a graph is proportional to the variance of the uniform distribution on that graph, measured with respect to the square root effective resistance, as*

$$R_G = n^2 \operatorname{var}_\omega(u/n)$$

Proof: Inserting the uniform distribution $(u/n)(i) = 1/n$ for all i, in equation (1) equals the formula for R_G up to the constant n^2, which proves Proposition 2. □

We remark that the uniform distribution also appears naturally as the stationary state of diffusion processes on a graph, see Appendix B.

C. Network modularity and Markov stability

Similar to the variance, we find that formula (2) for the graph covariance is related to some known expressions in graph theory and its applications.

An important theme across the multidisciplinary field of network science is the identification of groups of nodes in a graph that are tightly connected within, but poorly connected between the groups. These groups, known as communities, play an important role in many applications since the community structure says something about the higher-level or hierarchical organisation of a network. Consequently, there has been a large research effort to develop methods that uncover the underlying community structure of a given network, with many of these efforts centered around the concept of network modularity. First formulated by Newman and Girvan in [13], the *network modularity* $M(g)$ of a proposed partitioning $g : \mathcal{N} \to \{1, \ldots, k\}$ of the graph into k disjoint groups is a heuristic measure of how well these partitions exhibit the community structure, i.e. with many links within a group, but few between; the network modularity is defined as:

$$M(g) = \frac{1}{2m} \sum_{i,j \in \mathcal{N}} \left((A)_{ij} - \frac{k_i k_j}{2m} \right) \delta_{g(i)g(j)}$$

with adjacency matrix A which has entries $(A)_{ij} = c_{ij}$ for all links, and with the Kronecker delta defined as $\delta_{g(i)g(j)} = 1$ if and only if i and j are in the same group, i.e. such that $g(i) = g(j)$. Clearly this expression for modularity resembles the definition of covariance (2) and, indeed, if we let P be the joint distribution representing the ends of a randomly sampled link of the graph (proportional to its weight) as $P(i,j) = c_{ij}/(2m)$, and take the distance function $d_g(i,j) \triangleq (1 - \delta_{g(i)g(j)})$ then we find the correspondence:

Proposition 3 *The network modularity $M(g)$ of a node partitioning g is equal to the covariance of the ends of a random link of the graph with respect to the distance d_g, as*

$$M(g) = 2 \operatorname{cov}_{d_g}(P)$$

Proof: For the proposed joint distribution and distance function, we get the covariance expression

$$\operatorname{cov}_{d_g}(P) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} \left[\frac{k_i k_j}{(2m)^2} - \frac{c_{ij}}{2m} \right] (1 - \delta_{g(i)g(j)})^2$$

which by the fact that $c_{ij} = (A)_{ij}$ and $\delta^2_{g(i)g(j)} = \delta_{g(i)g(j)}$ for all i, j, and $\sum_{i,j} A_{ij} = \sum_{i,j} k_i k_j/(2m)$ gives equality with the network modularity $M(g)/2$. □

Proposition 3 shows that the usual interpretation of network modularity is well reproduced in the setting of covariances: for a proposed partitioning g, the covariance $\operatorname{cov}_{d_g}(P)$ measures how likely a random link

will fall within one of the groups instead of between two different groups. A high covariance thus reflects a good partitioning, in correspondence with the interpretation of a high modularity $M(g)$.

A recognized problem with the network modularity is that it suffers from a so-called 'resolution limit'. This limit impedes the detection of small but strongly connected groups of nodes [35] which is usually undesired when trying to describe the community structure of a network. One method to improve the detection of communities across different scales is the so-called Markov stability: the idea is to define the quality of a proposed partitioning g by measuring how likely a (continuous-time) random walker is to remain within the same group after a certain (continuous) time τ. The *Markov stability* $r_\tau(g)$ measures this likelihood and is defined as

$$r_\tau(g) = \operatorname{tr}\left(T_\tau \operatorname{diag}(\pi)\Delta_g\right) - \pi^T \Delta_g \pi$$

where $(\Delta_g)_{ij} = \delta_{g(i)g(j)}$ captures the proposed partitioning, T_τ determines the random walk process as $\Pr(N_{t+\tau} = j | N_t = i) = (T_\tau)_{ji}$ and with π the steady state distribution of this random walk process, i.e. satisfying $T_\tau \pi = \pi$. This transition matrix can be seen as a conditional probability matrix, such that $P_\tau \triangleq T_\tau \operatorname{diag}(\pi)$ is a joint distribution that reflects the probability that the random walker occupies node i in the stationary distribution and that it is at node j after a time τ. In other words distribution P_τ describes the distribution of a pair of nodes $(N_t, N_{t+\tau})$. Again, the formula for the Markov stability is reminiscent of the graph covariance (2) and we find the following correspondence:

Proposition 4 *The Markov stability $r_\tau(g)$ of a node partitioning g is equal to the covariance of the position of a stationary random walker and its position after time τ, with respect to the distance d_g, as*

$$r_\tau(g) = 2 \operatorname{cov}_{d_g}(P_\tau)$$

Proof: The proposed covariance can be written as

$$\operatorname{cov}_{d_g}(P_\mathcal{T}) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} \left(\pi(i)\pi(j) - (T_\tau)_{ij}\pi(i)\right)\left(1 - \delta_{g(i)g(j)}\right)^2$$

which by the fact that $\sum_{i,j} \pi(i)\pi(j) = \sum_{i,j} (T_\tau)_{ij}\pi(i)$ gives equality with the Markov stability $r_\tau(g)/2$. □

Interestingly, in [36] the Markov stability measure is derived based on a covariance matrix of the random vectors $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))^T$ and $\mathbf{X}(t+\tau)$, where $X_i(t)$ is a Bernoulli random variable, with probability $\Pr(X_i(t) = 1) = \Pr(N_t = i)$. The Markov stability is then found as the trace of this covariance matrix, after taking into account the proposed partitioning. By Proposition 4, this methodology is thus equivalent to simply measuring the covariance of the joint distribution (in contrast with the covariance of two random vectors) with respect to the metric d_g.

In Appendix B we briefly consider the variance for diffusion processes on a graph, which are closely related to the continuous-time random walk process used in the Markov stability.

V. MAXIMUM VARIANCE DISTRIBUTIONS

We now discuss a number of properties of the variance measured with respect to the square root effective resistance, and consider the problem of finding the maximum possible variance on a given graph. It will be useful to consider the variance as a function on the collection of all possible distributions (also called the probability simplex)

$$\Delta_n \triangleq \left\{ \mathbf{p} \in \mathbb{R}^n : p_i \geq 0 \text{ and } \sum_{i=1}^n p_i = 1 \right\}$$

as $\operatorname{var}_\omega : \Delta_n \to \mathbb{R}$ defined by $\operatorname{var}_\omega(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T \Omega \mathbf{p}$, where we fix an ordering of the nodes into the columns and rows of the effective resistance matrix Ω. We remark that distributions with $p_i = 1$ for a single element i (and thus zero otherwise) correspond to the extreme points of Δ_n, also called the *vertices*. We find the following characterization of the variance function for a certain graph on the probability simplex Δ_n:

Proposition 5 *The variance is strictly concave on Δ_n and bounded by $0 \leq \operatorname{var}_\omega(\mathbf{p}) \leq \sigma^2$. The minimum variance 0 is attained if and only if the distribution is a vertex of Δ_n, and the maximum variance σ^2 is attained by a unique distribution $\mathbf{p}^\star$.*

Proof: Strict concavity of the variance is proven in Appendix C based on the definition of the effective resistance in terms of the Laplacian matrix. This concavity means that for any two distinct distributions $\mathbf{p}$ and $\mathbf{q}$, the convex combinations of these distributions $\theta\mathbf{p} + (1-\theta)\mathbf{q}$ for $\theta \in (0,1)$ have a higher variance than the corresponding convex combination of their variances, in other words

$$\operatorname{var}_\omega(\theta\mathbf{p} + (1-\theta)\mathbf{q}) > \theta \operatorname{var}_\omega(\mathbf{p}) + (1-\theta)\operatorname{var}_\omega(\mathbf{q}).$$

For the vertices $\mathbf{p}_i$ of Δ_n we immediately find that $\operatorname{var}(\mathbf{p}_i) = \omega_{ii}/2 = 0$. Furthermore, since any other distribution (i.e. not equal to a vertex) can be written as a strict convex combination of vertices, the strict concavity of the variance says that no other distributions can have variance zero, showing that the minimum is only attained for vertices of Δ_n. Next, since all effective resistances and all possible distributions are finite, the variance must be bounded from above by some value σ^2 which is attained by at least one distribution $\mathbf{p}^\star$. If another distribution $\mathbf{q}^\star$ also were to attain this maximum we would find by strict concavity that $\operatorname{var}_\omega((\mathbf{p}^\star + \mathbf{q}^\star)/2) > \operatorname{var}_\omega(\mathbf{p}^\star)$; since this contradicts the

maximality of $\mathbf{p}^\star$, the maximum variance distribution is necessarily unique. □

Proposition 5 thus states that there is a unique solution to the *maximum variance problem*

$$\begin{aligned} &\text{maximize } \frac{1}{2}\mathbf{p}^T\Omega\mathbf{p} \\ &\text{subject to } \mathbf{p} \in \Delta_n. \end{aligned} \tag{5}$$

We write the *maximum variance* as σ^2, and the *maximum variance distribution* that attains it as $\mathbf{p}^\star$.
Since the objective function of (5) can be rewritten as $\mathrm{var}_\omega(\mathbf{p}) = -\mathbf{p}Q^\dagger\mathbf{p} + \zeta^T\mathbf{p}$ with $\zeta = \mathrm{diag}(Q^\dagger)$, where the pseudoinverse Laplacian is a positive semidefinite matrix, we know that the objective function is concave. Moreover, since the domain Δ_n is convex, the maximum variance problem (5) on a graph is a *convex quadratic program* and can be solved numerically in a number of steps which is polynomial in the size n of the graph [37]. In other words, the maximum variance problem can be solved efficiently. This result was shown before in [6] in context of calculating the average weighted resistance distance on a graph. Using a standard convex programming package CVX [38, 39], we calculate the maximum variance distribution on a number of empirical graphs, shown in Figure 6.

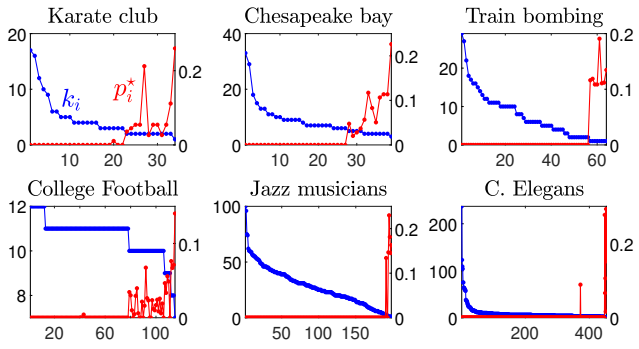


FIG. 6. Degree sequence (blue) and maximum variance distribution (red) on a number of real-world networks from the KONECT database [40]. The horizontal axis contains node indices i, sorted according to decreasing degree. For each node, the degree k_i and maximum variance probability $p_i^\star$ are given, with scales on the left and right vertical axis, respectively.

Apart from finding the maximum variance σ^2 and corresponding distribution $\mathbf{p}^\star$ numerically, we show in the next section that we can further describe the specific form of this distribution and its variance. This characterization allows to calculate exactly the maximum variance distribution (or its support) for a number of specific graph examples.

A. Characterizing maximum variance distributions

To further characterize the solution to the maximum variance problem we will first consider some necessary conditions that must be satisfied by any candidate optimal solution. In what follows, we assume that we know the maximum variance distribution $\mathbf{p}$ and derive three necessary conditions for this distribution. We then show that these conditions are also sufficient in the case of var_ω, and thus fully characterize the maximum variance distribution that solves problem (5).
A first necessary condition is a simple consequence of the fact that $\mathbf{p}$ is a distribution and thus needs to have non-negative entries and sum to one; in other words it must satisfy the (entrywise) inequality and equality

$$\mathbf{p} \geq 0 \text{ and } \mathbf{u}^T\mathbf{p} = 1. \tag{C_1}$$

The non-negativity condition may be satisfied with equality for some nodes (i.e. $p(i) = 0$ for some i), which can be captured in the *support* of the distribution as $\mathrm{supp}(\mathbf{p}) = \{i \in \mathcal{N} : p(i) > 0\}$. We will further abbreviate the support of our tentative optimal solution $\mathbf{p}$ by the set $\mathcal{V} \subseteq \mathcal{N}$ of size v.
Further optimality criteria can be derived from the fact that the maximal distribution $\mathbf{p}$ has the largest variance amongst all possible distributions, which means that any small perturbation of this distribution to another distribution must result in a decreasing variance. A first type of perturbation is a small transfer of probability mass ϵ between between two nodes of the support $i, j \in \mathcal{V}$, which yields the perturbed distribution $\mathbf{p}' = \mathbf{p} - \epsilon(e_i - e_j)$. Comparing the variance of this distribution to the original variance, we then must have

$$\mathrm{var}_\omega(\mathbf{p}') - \mathrm{var}_\omega(\mathbf{p}) \leq 0 \text{ for all } i, j \in \mathcal{V} \text{ and } \epsilon.$$

Introducing expression (3) for the variance and omitting the vanishing terms of order $O(\epsilon^2)$, this condition can be translated to the inequality $2\epsilon(e_i - e_j)^T\Omega\mathbf{p} \geq 0$. Since this must be true for all pairs of nodes in the support and all (small) values of ϵ, the inequality must hold with equality and we find the necessary condition

$$\Omega_{\mathcal{V}\mathcal{V}}\mathbf{p}_\mathcal{V} = \alpha\mathbf{u} \text{ for some } \alpha \in \mathbb{R}, \tag{C_2}$$

where $\Omega_{\mathcal{V}\mathcal{V}}$ is the $v \times v$ principal submatrix of the effective resistance matrix Ω, containing the resistances between pairs of nodes in the support, and similarly the $v \times 1$ vector $\mathbf{p}_\mathcal{V}$ containing the probabilities of nodes in the support. Equation (C_2) is the second necessary condition for the maximum variance distribution, and it is a *local optimality criterion* in the sense that it characterizes the optimal solution amongst all solutions with a given support $\mathcal{V}$. In the next section we show that equation (C_2) has a unique solution that is consistent with the normalization condition in $\mathbf{u}^T\mathbf{p} = 1$, which can be encoded by letting $\alpha = 2\,\mathrm{var}_\omega(\mathbf{p})$.

The solution to equation (C_2) is thus a candidate maximum variance distribution *if* this solution is a distribution. The question then remains which of these tentative distributions is the global optimum (i.e. for which $\mathcal{V}$)). For this reason, we consider a second type of perturbation which amounts to a small transfer of probability mass $\epsilon > 0$ from a node in the support $i \in \mathcal{V}$ to a node outside of the support $j \in \mathcal{V}^c$, yielding the perturbed distribution $\mathbf{p}'' = \mathbf{p} - \epsilon(e_i - e_j)$. Comparing the variance of this distribution to the original variance, we thus must have

$$\mathrm{var}_\omega(\mathbf{p}'') - \mathrm{var}_\omega(\mathbf{p}) < 0 \text{ for all } i \in \mathcal{V}, j \in \mathcal{V}^c \text{ and } \epsilon > 0,$$

which translates again to inequality $2\epsilon(e_i - e_j)^T \Omega \mathbf{p} > 0$ when introducing the variance and disregarding $O(\epsilon^2)$ terms. From the first two necessary conditions, we furthermore know that $e_i^T \Omega \mathbf{p} = 2\,\mathrm{var}_\omega(\mathbf{p})$ for any $i \in \mathcal{V}$, such that these inequalities can be written as the following (entrywise) inequality

$$\Omega_{\mathcal{V}^c\mathcal{V}} \mathbf{p}_\mathcal{V} < 2\,\mathrm{var}_\omega(\mathbf{p})\mathbf{u}, \qquad (C_3)$$

where $\Omega_{\mathcal{V}^c\mathcal{V}}$ is the $(n-v) \times v$ submatrix of the effective resistance matrix Ω, containing columns in $\mathcal{V}$ and rows in its complement $\mathcal{V}^c$. Equation (C_3) is the third necessary condition for the maximum variance distribution, and it is a *global optimality criterion* since it guarantees that the variance decreases when changing the support of $\mathbf{p}$ by transferring probability mass to nodes outside the support.

To summarize, we have identified three necessary conditions for the maximum variance distribution $\mathbf{p}^\star$ that solves problem (5) for *any* distance function; condition (C_1) guarantees that our solution is a distribution and condition (C_2) and (C_3) make sure that any local change in the distribution leads to a decrease in variance. In other words, these three conditions characterize a *local maximum* of the variance in the domain. Now, since the variance with respect to the resistance distance is concave this local solution also corresponds to the *global maximum*, i.e. the necessary conditions are also sufficient, and we find the following characterization:

Proposition 6 *The maximum variance distribution* $\mathbf{p}^\star$ *with respect to the square root effective resistance is the unique vector that satisfies conditions* (C_1)*,* (C_2) *and* (C_3)*.*

Proof: As the variance is a strictly concave function, we can bound the variance of any distribution $\mathbf{q}$ as

$$\mathrm{var}_\omega(\mathbf{q}) < \mathrm{var}_\omega(\mathbf{p}) + (\mathbf{q} - \mathbf{p})^T \nabla\, \mathrm{var}_\omega(\mathbf{p})$$

where the gradient of the variance equals $\nabla\, \mathrm{var}_\omega(\mathbf{p}) = \Omega\mathbf{p}$. Rewriting the difference between distributions we find that

$$\mathrm{var}_\omega(\mathbf{q}) < \mathrm{var}_\omega(\mathbf{p}) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}} p_i q_j (e_j - e_i)^T \Omega \mathbf{p}$$

When a distribution $\mathbf{p}$ satisfies conditions $(C_1)-(C_3)$, the terms in the sum are all non-positive and thus $\mathrm{var}_\omega(\mathbf{q}) < \mathrm{var}_\omega(\mathbf{p})$ holds for all $\mathbf{q} \in \Delta_n$. This shows that these conditions are not only necessary, but also sufficient for the vector $\mathbf{p}$ to be the global optimum $\mathbf{p}^\star$, which proves the proposition. We remark that an alternative proof via Lagrangian multipliers is quite instructive, in particular with regards to the role of slack variables in encoding the support $\mathcal{V}$ of a solution. □

Proposition 6 gives an exact description of what the maximum variance distribution looks like for a given graph. Somewhat surprisingly, this distribution can (and often will) have zero probability for a subset of the nodes. Intuitively, this means that when looking for a distribution that is most spread out, some nodes are too *central* in the network (i.e. too close on average to other nodes) and are assigned zero probability as a consequence. The remaining nodes which are in the *maximum variance support* $\mathcal{V}^\star$ on the other hand, are *peripheral* (or boundary) nodes whose distance to other nodes is sufficiently large to contribute to a high variance. By considering the maximum variance distribution, we thus have a procedure to 'single out' a subset of nodes $\mathcal{V}^\star$ from a given graph G, which can be interpreted as a set of peripheral nodes. In the following section, we further support this interpretation by characterizing $\mathcal{V}^\star$ on some simple graphs where this maximum variance support indeed corresponds to a peripheral set of nodes. For further discussions on core-periphery structure in networks, see for instance [41]

In Appendix G we describe a potential application of maximum variance supports in the context of a k-core decomposition of networks.

B. Maximum variance support on some particular graphs

A graph G is called a *tree graph* if there is exactly one path between every two nodes, or equivalently, if it contains no cycles. In many ways, tree graphs are the simplest possible graphs and their properties can be studied in great detail. The nodes of a tree graph fall in two categories: the *leaf nodes* [42] which have a single incident link and form the 'extremities' of the tree, and the *non-leaf nodes* that make up the 'core' of the tree; every path between two leaf nodes necessarily passes through one of the non-leaf nodes. This intuitively clear distinction between central and peripheral nodes in trees is well reproduced when considering the maximum variance support $\mathcal{V}^\star$ on tree graphs; we find the following condition:

Proposition 7 *On tree graphs, the maximum variance support* $\mathcal{V}^\star$ *is a subset of the leaf nodes.*

Proof: A proof of Proposition 7 was presented in [43] when considering the equivalent problem of maximizing $\mathbf{p}^T D \mathbf{p}$ for shortest-path distance matrix D (which equals the effective resistance matrix Ω in the case of tree graphs) over all distributions $\mathbf{p} \in \Delta_n$. For completeness,

we include a self-contained proof in Appendix D based on our expressions for the maximum variance distribution. Moreover, the proof and thus Proposition 7 is valid for any distance function d for which $d(i,j) = d(i,x)+d(x,j)$ whenever removing x from the graph disconnects i and j. □

As discovered already in [4, 43], the maximum variance support can be a strict subset of the leaf nodes in some cases. To illustrate this, we further restrict our attention to (weighted) star graphs as a special case of tree graphs. A star graph on $n+1$ nodes consists of a central (non-leaf) node $\{0\}$ connected to all the other (leaf) nodes, with link weights $c_{0i} = k_i$, and no further connections otherwise. The leaf node degrees $\mathbf{k} = \{k_1, \dots, k_n\}$ fully parametrize a weighted star and all relevant properties can be expressed in terms of these degrees. In particular, the effective resistance between any two leaf nodes i and j is given by $\omega_{ij} = k_i^{-1} + k_j^{-1}$ as shown in Appendix E. For star graphs, we then find the following exact characterization of the maximum variance support:

Proposition 8 *On a weighted star graph with leaf-node degrees* $\mathbf{k}$*, the maximum variance support* $\mathcal{V}^\star$ *is equal to the nodes with the* ℓ *smallest degrees, such that* $k_\ell \leq (\ell - 2)^{-1} \sum_{i=1}^{\ell} k_i < k_{\ell+1}$.

Proof: We prove Proposition 8 in Appendix E based on the necessary conditions (C_1)–(C_3) for the maximum variance distribution. □

Using Proposition 8 we find that the weighted star with degrees $\{k, k, k, k'\}$ is supported on the nodes with degree k, i.e. a strict subset of the leaf nodes, whenever $k' > 3k$, corresponding to an example given in [4]. While very simple and somewhat artificial, the problem of finding the maximum variance support on a star graph is in fact equivalent to the same problem on a more popular type of graph. For a (multi)set of degrees $\mathbf{k} = \{k_1, \dots, k_n\}$, the *configuration graph* $C_\mathbf{k}$ is a graph on n nodes where each node i is assigned one of the degrees k_i and has link weights $c_{ij} = k_i k_j/(2m-1)$ to all other nodes j. These graphs appear as the ensemble average of the popular *configuration random graphs* (see e.g. [13]) which are frequently used to study the influence of the degree sequence on various graph properties, or as null models in the analysis of empirical graphs; configuration graphs also appear in the mean-field analysis of dynamical processes on graphs such as epidemics [15]. In Appendix E we show that the effective resistance between any two nodes in $C_\mathbf{k}$ equals $\omega_{ij} = 2m/(2m-1)(k_i^{-1} + k_j^{-1})$ as in the star graph, which means that solving problem (5) on configuration graphs and star graphs with the same degree sequence $\mathbf{k}$ is equivalent, up to some constant factor. We thus have the following Corollary of Proposition 8

Corollary 1 *On configuration graphs with degree sequence* $\mathbf{k}$*, the maximum variance support* $\mathcal{V}^\star_\mathbf{k}$ *equals the nodes with the* ℓ *smallest nodes, where* $k_\ell \leq (\ell - 2)^{-1} \sum_{i=1}^{\ell} k_i < k_{\ell+1}$.

Proof: The proof of Corollary 1 is given in Appendix E by establishing the equivalence of the maximum variance problem (5) on weighted stars and configuration graphs with the same degree sequence $\mathbf{k}$. □

It was shown in [44, 45] that in certain parameter regimes of random geometric graphs and for realizations of the configuration model, the effective resistance between all node pairs will approximately be proportional to $k_i^{-1} + k_j^{-1}$, in which case $\mathcal{V}^\star_\mathbf{k}$ is thus particularly suitable as an approximate maximum variance support. Furthermore, this result could explain our observation in Section V C that nodes in the maximum variance support of a random geometric graphs seem more likely to be located close to the boundaries of the sampled domain.

Our results above focus on the case where the maximum variance support $\mathcal{V}^\star$ points to a subset of the total node set. In some cases however, a graph might be 'homogeneous' in the sense that all nodes have similar distances to the other nodes and no central or peripheral node sets stand out. In the example above for instance, when the degree sequence satisfies $k_{\max}/k_{\min} < v/(v-2)$, all nodes will be sufficiently similar with $\mathcal{V}^\star = \mathcal{N}$ as a result. An example of such homogeneous graphs are *node-transitive graphs* (more commonly called vertex-transitive graphs). A graph $G = (\mathcal{N}, \mathcal{L})$ is node transitive [46] if for every pair of nodes i and j, there exists a bijection $\pi : \mathcal{N} \to \mathcal{N}$ that maps i to j as $\pi(i) = \pi(j)$ and where π is an automorphism such that $(\pi(a), \pi(b)) \in \mathcal{L}$ if and only if $(a,b) \in \mathcal{L}$. In other words, no two nodes (and thus no subset of nodes) are 'distinguishable' from each other by simply considering their position in the network. Since all nodes in the network are indistinguishable, the corresponding Laplacian and resistance matrices are highly symmetrical from which the following result follows:

Proposition 9 *On node-transitive graphs, the maximum variance distribution is the uniform distribution* $\mathbf{p}^\star = \mathbf{u}/n$ *and the maximum variance support* $\mathcal{V}^\star$ *is the full node set* $\mathcal{N}$.

Proof: In [3, Thm. 14] it is shown that $\sum_{j\sim i} c_{ij}\omega_{ij} = 2 - 2/n$ for any node i in a node-transitive graph. Following Proposition 11 (see later) we thus find $p_i^\star = 1/n$ for every node and consequently $\mathcal{V}^\star = \mathcal{N}$. □

Some well-known examples of node-transitive graphs to which this result thus applies are complete graphs, cycles and hypercubes as well as all Cayley graphs (of which the former are all examples) [46]. We remark that a maximum variance distribution supported on the full node set has strong 'topological' implications for the underlying graph; Fiedler found that in this case, removing any set of r nodes can disconnect the graph into at most $r-1$ components and conjectured that, conversely, any graph with this connectivity property must have $\mathcal{V}^\star = \mathcal{N}$ [47, Thm. 3.4.18].

To conclude we observe that the maximum variance support only contains two nodes in the case of a path graph,

more precisely we find:

Proposition 10 *The maximum variance support contains exactly two nodes if and only if these nodes are the ends of a path graph.*

Proof: For a set of two nodes $\mathcal{V} = \{a, b\}$ we find that equation (C_2) yields $p_a = p_b = 1/2$ and that

$$\sum_{i\in\mathcal{V}} \omega_{ix} p_i - \sum_{i,j\in\mathcal{V}} \omega_{ik} p_i p_k = \frac{1}{2}(\omega_{ax} + \omega_{xb} - \omega_{ab}) \geq 0$$

for all $x \notin \mathcal{V}$. This shows that condition (C_3) is satisfied if and only if $\omega_{ax} + \omega_{xb} = \omega_{ab}$ for all $x \notin \mathcal{V}$, which is equivalent to a and b being the ends of a path graph. □
We remark that this contradicts an example proposed in [43] about a three-leaf tree graph with maximum variance support on two leaf nodes.

C. Maximum variance support in random geometric graphs

The results in the previous section provide theoretical support for the interpretation that the maximum variance support indicates some sort of periphery or boundary of a graph. Here, we provide another result in this line by considering random geometric graphs (RGGs), which are naturally embedded in some Euclidean space and have a well-defined boundary.
For a set of points $\mathcal{N} \subseteq \mathbb{R}^2$ in the plane, we define the ϵ-graph G_ϵ as a graph with nodes $\mathcal{N}$ and with pairs of nodes linked if and only if they are closer than a certain distance $\epsilon > 0$ from each other in the plane (the connection radius). A random geometric graph on a subset $\mathcal{X} \subseteq \mathbb{R}^2$ of the plane is then determined by sampling n points uniformly at random from $\mathcal{X}$ (or via a Poisson point process with a certain rate) and constructing the corresponding ϵ-graph. For more details and more general constructions, see for instance [48].
In Figure 7 we show experimentally that the maximum variance support nodes of a random geometric graph on a domain $\mathcal{X}$ tend to be located close to the boundary of this domain. Moreover, this observation seems to hold quite robustly for different parameter settings of the experiment (number of points, connection radius and shape of the domain). We cannot provide a conclusive theoretical explanation of the observations in Figure 7, but we believe it could be related to the results of von Luxburg et al. [44, 45]. Their work shows that in certain regimes of (ϵ, n), the effective resistances in a random geometric graph will be approximately $\omega_{ij} \approx k_i^{-1} + k_j^{-1}$, i.e. as in a configuration graph. Consequently, we might conjecture that the maximum variance support in this regime converges accordingly to the result described in Corollary 1 for configuration graphs, and that the maximum variance support will thus consist of low-degree nodes. As these low-degree nodes are more likely to be located at the boundary of the domain (due to the Poisson-distribution property of Poisson point processes) the maximum variance support nodes would thus indeed be more likely to be close to the domain boundaries. Developing a full theoretical understanding of our observations (potentially following our guesses above), perhaps extending them to RGGs on manifolds (the setting of [45]) and further investigating the role of curvature (points near a highly curved boundary will likely have small degrees) seems to be a particularly interesting venue for further research.

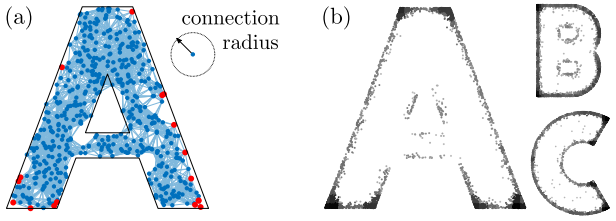


FIG. 7. The maximum variance support of random geometric graphs (RGGs) in a bounded domain is more likely to contain nodes which are located near the boundaries. Panel (a) shows one realization of an RGG on $n = 500$ nodes in a domain shaped as the letter 'A', with connection radius as indicated. The nodes in the maximum variance support are colored red, and are all located near the boundary of the domain. Panel (b) summarizes the maximum variance support node locations calculated in 250 independent RGG realizations on the given domains. A darker (resp. lighter) shading indicates a higher (lower) density of maximum variance nodes. This figure seems to confirm that the maximum variance support nodes are more likely to be located near the domain boundaries and, in particular, near corners with 'high curvature'.

VI. ALTERNATIVE FORMULATIONS OF LOCALLY OPTIMAL SOLUTIONS

To conclude the article, we present a number of equivalent formulations for the local optimality condition (C_2). These reformulations are relevant in the context of our article since they determine the form of the maximum variance solution $\mathbf{p}^\star$, but more importantly, the vector $\mathbf{p}$ that solves (C_2) (with $\alpha = 2\,\mathrm{var}_\omega(\mathbf{p})$) uniquely for a given graph seems to be a natural algebraic object on this graph, worthy of a more thorough characterization. The many different formulations in Proposition 11 below are an attest to the different settings in which the vector $\mathbf{p}$ seems to play a role.
We first introduce two results of Miroslav Fiedler from which a distinct *geometric* and *algebraic* characteristic of $\mathbf{p}$ will follow. In his book [47], Fiedler collects a broad range of results that follow from a geometric perspective on graphs. At the core of this perspective is the fact that each graph has an embedding $\mathbf{m} : \mathcal{N} \to \mathbb{R}^{n-1}$ into Euclidean space such that

$$\|\mathbf{m}(i) - \mathbf{m}(j)\|^2 = \omega_{ij} \text{ for all } i, j \in \mathcal{N},$$

where the embedded nodes are the vertices of a hyperacute simplex (see [47, 49] for more details). In other

words $\mathbf{m}$ is an isometric embedding of the node set with respect to the square root resistance distance. This embedding can be extended to functions on the nodes as $\mathbf{m}(\mathbf{q}) = \sum_{i\in\mathcal{N}} q_i \mathbf{m}(i)$ which means that we can interpret node functions as 'coordinates' of points in $\mathbb{R}^{n-1}$. From this perspective, we will see that $\mathbf{p}$ is the coordinate of a distinguished point of the simplex with vertices $\mathbf{m}(\mathcal{N})$.
A second place where equation (C_2) appears in the work of Fiedler is in a matrix identity between the effective resistance and Laplacian matrices. Writing the local optimality condition (C_2) in block-matrix form, we find $\begin{pmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega \end{pmatrix} \begin{pmatrix} -2\sigma^2 \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}$ with zero vector $\mathbf{0} = (0, \dots, 0)^T$. The following result of Fiedler shows that this equation can be completed to a full (inverse) matrix identity:

Theorem 1 (Fiedlers identity) *The Laplacian matrix Q and resistance matrix Ω of a graph satisfy the matrix identity*

$$\begin{pmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega \end{pmatrix}^{-1} = -\frac{1}{2} \begin{pmatrix} 4\sigma^2 & -2\mathbf{p}^T \\ -2\mathbf{p} & Q \end{pmatrix} \tag{6}$$

where $\Omega\mathbf{p} = 2\sigma^2\mathbf{u}$.

Proof: See proof in [47, Thm. 1.4.1]. The validity of equation (6) can also be checked by multiplying both matrices, which results in the identity matrix. □
Using the embedding $\mathbf{m}$ of a graph and Fiedlers identity (6) we now find the following reformulations of equation (C_2):

Proposition 11 *The following are equivalent characterizations for a vector $\mathbf{p} \in \mathbb{R}^n$ and its variance $\mathrm{var}_\omega(\mathbf{p})$:*

(i) $\Omega\mathbf{p} = 2\,\mathrm{var}_\omega(\mathbf{p})\mathbf{u}$. In other words, $\mathbf{p}$ solves equation (C_2) with $\alpha = 2\,\mathrm{var}_\omega(\mathbf{p})$.

(ii) $\mathbf{p} = \mathrm{argmax}_{\mathbf{q}} \left\{ \frac{1}{2}\mathbf{q}^T\Omega\mathbf{q} \text{ s.t. } \mathbf{u}^T\mathbf{q} = 1 \right\}$, with corresponding optimal value $\mathrm{var}_\omega(\mathbf{p})$

(iii) $\|\mathbf{m}(\mathbf{p}) - \mathbf{m}(i)\|^2 = \mathrm{var}_\omega(\mathbf{p})$ for all $i \in \mathcal{N}$ and isometric embedding $\mathbf{m}$. In other words, $\mathbf{m}(\mathbf{p})$ is the *circumcenter* of the sphere going through vertices $\mathbf{m}(\mathcal{N})$, with *circumradius* $\sqrt{\mathrm{var}_\omega(\mathbf{p})}$.

(iv) $\mathbf{p} = \frac{1}{2}Q\zeta + \mathbf{u}/n$, with corresponding variance $\mathrm{var}_\omega(\mathbf{p}) = \frac{1}{4}\zeta^T Q\zeta + \mathbf{u}^T\zeta/n$

(v) $p_i = 1 - \frac{1}{2}\sum\limits_{j\sim i} c_{ij}\omega_{ij}$ for all $i \in \mathcal{N}$

Proof: The proof is given in Appendix F. We remark that the characterizations (i) and (iii) are discussed in [47], and the form (v) appears in [50]. In formulation (v), the normalization of p_i is guaranteed by $\sum_{j\sim i} c_{ij}\omega_{ij} = n - 1$, which is known as Foster's Theorem [51]. □
While formulated in terms of solutions of the local optimality condition on the full node set $\mathcal{N}$, Proposition 11 above in fact describes the solutions for general $\mathcal{V}$ due to an important recursive structure of the effective resistance: for any resistance matrix Ω of a graph G, the submatrix $\Omega_{\mathcal{V}\mathcal{V}}$ is again the resistance matrix of a graph G'. More precisely, if we write the Laplacian matrix in block-form $Q = \begin{pmatrix} Q_{\mathcal{V}\mathcal{V}} & Q_{\mathcal{V}\mathcal{V}^c} \\ Q_{\mathcal{V}^c\mathcal{V}} & Q_{\mathcal{V}^c\mathcal{V}^c} \end{pmatrix}$ for some node-set $\mathcal{V}$ and its complement $\mathcal{V}^c$, then the *Schur complement of Q with respect to $\mathcal{V}$* is defined as

$$Q/\mathcal{V}^c \triangleq Q_{\mathcal{V}\mathcal{V}} - Q_{\mathcal{V}\mathcal{V}^c}(Q_{\mathcal{V}^c\mathcal{V}^c})^{-1}Q_{\mathcal{V}^c\mathcal{V}} \tag{7}$$

and has the important properties that it is again a Laplacian matrix of some graph G' and that this graph has resistance matrix $\Omega' = \Omega_{\mathcal{V}\mathcal{V}}$, see [47, 52]. Consequently, the solutions described in Proposition 11 are valid for the local optimum on any set $\mathcal{V}$ by considering the Laplacian $Q/\mathcal{V}^c$ and effective resistance matrix $\Omega_{\mathcal{V}\mathcal{V}}$ for which $\mathcal{V}$ is the full node set. Due to this recursive structure of Laplacians and the effective resistance, we can intuitively understand the method of solving the maximum variance problem as calculating the variance of the locally optimal solution $\mathbf{p}_\mathcal{V}$ for each set $\mathcal{V}$ by (C_2) and then choosing the feasible ones (i.e. distributions) by (C_1) and characterizing the globally optimal one by condition (C_3). In this perspective, we have a last result that relates solutions of (C_2) and their variances between different sets of nodes:

Proposition 12 *The solution $\mathbf{p}$ to equation (C_2) for the set $\mathcal{V} = \mathcal{N}$ and $\mathbf{p}'$ for the sets $\mathcal{V} = \mathcal{N}\backslash\{x\}$ are related by*

$$\begin{cases} \mathbf{p}' = \mathbf{p} - \frac{p_x}{k_x}\sum_{j\sim x} c_{xj}(e_x - e_j) \\ \mathrm{var}_\omega(\mathbf{p}') = \mathrm{var}_\omega(\mathbf{p}) - \frac{p_x^2}{k_x} \end{cases}$$

Proof: Proposition 12 is proven in Appendix F making use of the Schur complement of the graph Laplacian and Fiedlers identity. □
In particular, Proposition 12 thus shows that if a locally optimal distribution is non-negative for a set $\mathcal{V}$ as $\mathbf{p}_\mathcal{V} \geq 0$, then the locally optimal solutions for all subsets $\mathcal{W} \subset \mathcal{V}$ will be non-negative as well, by positivity of the factor $p_x c_{xj}/k_x$. Furthermore, we find that the variance var_ω of locally optimal distributions is non-increasing when considering subsets.

VII. CONCLUSION

In this paper, we have introduced two new measures which allow to calculate the variance and covariance of distributions defined on the nodes of a network, as a generalization of the standard (co)variance. These measures take into account the underlying structure of the network in the form of distance between nodes, thus providing a tool in the study of functional properties of the network (distributions, signals, ...) relative to the underlying structure of the network. Furthermore, the variance and covariance take the simple form of a quadratic product and matrix trace, respectively, which are easily calculated using standard linear algebra solvers.

To support the specific form of our introduced variance and covariance measures, we show that their definitions

specialize to a number of known graph characteristics and heuristics developed in network science and that there is a conceptual correspondence between our (co)variance measures and these graph characteristics. Specifically, we show that Kemeny's constant, the Kirchhoff index, network modularity and Markov stability are obtained as the (co)variance of specific distributions and distance functions on a network.

We furthermore use the variance and covariance in a practical scenario where both structural and functional data of a networked system are known. We analyse a 'network of knowledge' consisting of mathematical concepts as nodes and links inferred from hyperlinks on Wikipedia. Based on a corpus of 140k+ papers on the arXiv, we then analyse how this network of knowledge is used in practice. We translate the occurrences and co-occurrences of concepts in these papers into distributions and joint distributions on the network, and show that the corresponding variance is smaller, on average, compared to a 'virtual paper' null model and similarly, that the corresponding covariance is larger, on average, compared to a randomized null model. Beyond this particular setting, our framework has many potential applications in fields like neuroscience, to characterize the relationship between structural and functional brain networks in terms of covariance; in economics, where variance can be used as a measure of economic diversity, or in social networks, where the variance of certain distributions could be interpreted as a measure of polarization.

In the second part of this paper, we consider the variance on a graph measured with respect to the square root effective resistance distance. In this case, the variance is a strictly concave function over the set of possible distributions Δ_n and there is a unique maximizing distribution. We give a detailed description of the *maximum variance problem* and the numerical and theoretical approaches to find the maximum variance distribution $\mathbf{p}^\star$, and an interpretation of this distribution as a core-periphery indicator. We highlight the interesting observation that, in general, the maximum variance distribution is supported on a subset of the nodes $\mathcal{V}^\star$. We furthermore characterize this maximum variance support $\mathcal{V}^\star$ on tree graphs, weighted stars, configuration graphs and node-transitive graphs to further support our interpretation of this set as a collection of peripheral nodes. Additional evidence is found by considering random geometric graphs where we find in experiments that the maximum variance support nodes are most likely situated near the boundaries of the domain of the geometric graphs. In Appendix B and G we describe two potential further applications of our variance and covariance framework.
Finally, we discuss a number of alternative characterizations of the vector determined by equation (C_2) which appears to be a natural algebraic object associated to graphs, appearing in many interesting contexts such as graph variances and Fiedlers graph-simplex correspondence. While the theoretical analyses in this article focus on measuring variance with respect to the effective resistance, we stress that definitions (1) and (2) work for any distance function on a graph, and in many cases other distances than the resistance distance are likely to be more natural.

ACKNOWLEDGEMENTS

KD was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. SMG acknowledges support by the Spanish Ministry of Science, Innovation and Universities (MICIU) under Contract No. PGC2018-093854-B-I00 and Spanish Ministry of Education, Culture and Sport (Grant No. FPU15/01461)

[1] R. B. Bapat, *Graphs and Matrices* (Springer London, 2010).
[2] D. J. Klein and M. Randić, Resistance distance, Journal of Mathematical Chemistry **12**, 81 (1993).
[3] J. Zhou, Z. Wang, and C. Bu, On the resistance matrix of a graph, The Electronic Journal of Combinatorics **23**, 10.37236/5295 (2016).
[4] P. Hjorth, P. Lisoněk, S. Markvorsen, and C. Thomassen, Finite metric spaces of strictly negative type, Linear Algebra and its Applications **270**, 255 (1998).
[5] R. B. Bapat and S. K. Neogy, On a quadratic programming problem involving distances in trees, Annals of Operations Research **243**, 365 (2014).
[6] D. Dubey and S. K. Neogy, On solving a non-convex quadratic programming problem involving resistance distances in graphs, Annals of Operations Research **287**, 643 (2018).
[7] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Processing Magazine **30**, 83 (2013).
[8] In the case of unweighted graphs, the length of a path simply equals the number of links contained in the path. In the case of weighted graphs, the length of a path is defined as the sum $\sum_{(i,j)\in\mathcal{P}} c_{ij}^{-1}$ over all links $\mathcal{P}$ in the path. From this perspective, the weight between two links acts as an affinity, where larger weights correspond to 'better connected' nodes.
[9] The number of neighbours of a node i will be called the combinatorial degree to distinguish it from the 'weighted' degree k_i.
[10] W. Ellens, F. Spieksma, P. V. Mieghem, A. Jamakovic, and R. Kooij, Effective graph resistance, Linear Algebra and its Applications **435**, 2491 (2011).
[11] H. Qiu and E. R. Hancock, Clustering and embedding using commute times, IEEE Transactions on Pattern Anal-

ysis and Machine Intelligence **29**, 1873 (2007).
[12] A. D. Gvishiani and V. A. Gurvich, Metric and ultrametric spaces of resistances, Russian Mathematical Surveys **42**, 235 (1987).
[13] M. Newman, *Networks* (Oxford University Press, 2018).
[14] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
[15] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, Epidemic processes in complex networks, Reviews of Modern Physics **87**, 925 (2015).
[16] S. Martin-Gutierrez, R. Lambiotte, and K. Devriendt, Variance and covariance of distributions on graphs, `https://github.com/rlambiot/variance` (2020).
[17] V. Salnikov, D. Cassese, R. Lambiotte, and N. S. Jones, Co-occurrence simplicial complexes in mathematics: identifying the holes of knowledge, Applied Network Science **3**, 10.1007/s41109-018-0074-3 (2018).
[18] C. G. M. Mihail, The Markov chain simulation method for generating connected power law random graphs, in *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments*, Vol. 111 (SIAM, Philadelphia, 2003) p. 16.
[19] H.-J. Park and K. Friston, Structural and Functional Brain Networks: From Connections to Cognition, Science **342**, 10.1126/science.1238411 (2013).
[20] F. Abdelnour, M. Dayan, O. Devinsky, T. Thesen, and A. Raj, Functional brain connectivity is predictable from anatomic network's Laplacian eigen-structure, NeuroImage **172**, 728 (2018).
[21] C. O. Becker, S. Pequito, G. J. Pappas, M. B. Miller, S. T. Grafton, D. S. Bassett, and V. M. Preciado, Spectral mapping of brain functional connectivity from diffusion imaging, Scientific Reports **8**, 1411 (2018).
[22] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, Quantifying controversy on social media, Trans. Soc. Comput. **1** (2018).
[23] A. Alshamsi, F. L. Pinheiro, and C. A. Hidalgo, Optimal diversification strategies in the networks of related products and of related research areas, Nature Communications **9**, 1328 (2018).
[24] D. Hartmann, M. R. Guevara, C. Jara-Figueroa, M. Aristarn, and C. A. Hidalgo, Linking economic complexity, institutions, and income inequality, World Development **93**, 75 (2017).
[25] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, The product space conditions the development of nations, Science **317**, 482 (2007).
[26] D. Levin and Y. Peres, *Markov Chains and Mixing Times* (American Mathematical Society, 2017).
[27] L. Lovsz, Random walks on graphs: A survey, in *Combinatorics, Paul Erdős is eighty* (János Bolyai Math. Soc., Budapest, Hungary, 1993) pp. 353–398.
[28] P. G. Doyle and J. L. Snell, Random walks and electric networks, Mathematical Association of America (1984).
[29] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, The electrical resistance of a graph captures its commute and cover times, Computational Complexity **6**, 312 (1996).
[30] C. St. J. A. Nash-Williams, Random walk and electric currents in networks, Mathematical Proceedings of the Cambridge Philosophical Society **55**, 181194 (1959).
[31] J. G. Kemeny, Generalization of a fundamental matrix, Linear Algebra and its Applications **38**, 193 (1981).
[32] X. Wang, J. L. Dubbeldam, and P. Van Mieghem, Kemeny's constant and the effective graph resistance, Linear Algebra and its Applications **535**, 231 (2017).
[33] J. L. Palacios, Closed-form formulas for Kirchhoff index, International Journal of Quantum Chemistry **81**, 135 (2001).
[34] A. Ghosh, S. Boyd, and A. Saberi, Minimizing effective resistance of a graph, SIAM Review **50**, 37 (2008).
[35] S. Fortunato and M. Barthélemy, Resolution limit in community detection, Proceedings of the National Academy of Sciences **104**, 36 (2007).
[36] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Stability of graph communities across time scales, Proceedings of the National Academy of Sciences **107**, 12755 (2010).
[37] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).
[38] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, `http://cvxr.com/cvx` (2014).
[39] M. Grant and S. Boyd, Graph implementations for nonsmooth convex programs, in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag Limited, 2008) pp. 95–110.
[40] J. Kunegis, Konect: The Koblenz network collection, in *Proceedings of the 22nd International Conference on World Wide Web* (Association for Computing Machinery, New York, NY, USA, 2013) p. 13431350.
[41] P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, Core-periphery structure in networks (revisited), SIAM Review **59**, 619 (2017).
[42] A standard result of graph theory states that every tree has at least 2 leaf nodes; see for instance [53].
[43] P. Dankelmann, Average distance in weighted graphs, Discrete Mathematics **312**, 12 (2012).
[44] U. von Luxburg, A. Radl, and M. Hein, Getting lost in space: Large sample analysis of the resistance distance, in *Advances in Neural Information Processing Systems* (2010) pp. 2622–2630.
[45] U. von Luxburg, A. Radl, and M. Hein, Hitting and commute times in large random neighborhood graphs, Journal of Machine Learning Research **15**, 1751 (2014).
[46] C. Godsil and G. Royle, *Algebraic Graph Theory* (Springer New York, 2001).
[47] M. Fiedler, *Matrices and Graphs in Geometry* (Cambridge University Press, 2009).
[48] M. Penrose, *Random Geometric Graphs* (Oxford University Press, 2003).
[49] K. Devriendt and P. Van Mieghem, The simplex geometry of graphs, Journal of Complex Networks **7**, 469 (2019).
[50] R. Bapat, Resistance matrix of a weighted graph, MATCH - Communications in Mathematical and in Computer Chemistry **50** (2004).
[51] R. M. Foster, The average impedance of an electrical network, Contributions to Applied Mechanics (Reissner Anniversary Volume) , 333 (1949).
[52] F. Dörfler and F. Bullo, Kron reduction of graphs with applications to electrical networks, IEEE Transactions on Circuits and Systems I: Regular Papers **60**, 150 (2013).
[53] B. Bollobás, *Graph Theory* (Springer New York, 1979).
[54] L. Blumenthal, *Theory and Applications of Distance Geometry*, AMS Chelsea Publishing Series (Chelsea Publishing Company, 1970).

[55] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, Large scale networks fingerprinting and visualization using the k-core decomposition, in *Advances in neural information processing systems* (2006) pp. 41–50.

Appendix A: Standard variance and covariance as a special case of the graph (co)variance (1) and (2).

We consider the definition of variance and covariance for distributions on $\mathcal{N} \subset \mathbb{R}$, a finite subset of the real line $\mathbb{R}$, and joint distributions on $\mathcal{N} \times \mathcal{N}$. Both the variance and covariance make use of the *expectation* operator $\mathbb{E} : \mathcal{N} \to \mathbb{R}$, which is defined as

$$\mathbb{E}(N) \triangleq \sum_{i \in \mathcal{N}} p(i) i \text{ for some } N \sim p. \tag{A1}$$

This operator can be extended to functions on $\mathcal{N}$ as $\mathbb{E}(f(N)) = \sum p(i) f(i)$ for some $f : \mathcal{N} \to \mathbb{R}$, which can be interpreted as the *expected* outcome of the function f when applied to the random variable N.
Using the expectation operator, the *variance* of a distribution p is then defined as

$$\text{var}(p) = \mathbb{E}\left([N - \mathbb{E}(N)]^2\right) \text{ with } N \sim p. \tag{A2}$$

In other words, the variance is the expected squared difference between a random outcome of distribution p and its expected outcome. Importantly, for this definition to make sense we need a 'difference' $(i - \mathbb{E}(N))$ between an element of the set and its average, and a way to 'square' this difference $(i - \mathbb{E}(N))^2$; both are possible when $\mathcal{N}$ is a subset of the real line.
For a joint distribution P, the expectation operator on a function $f : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$ is defined similarly as $\mathbb{E}(f(N, M)) = \sum_{i,j \in \mathcal{N}} P(i, j) f(i, j)$ for a random pair of nodes $(N, M) \sim P$. The *covariance* of a joint distribution P is then defined as

$$\text{cov}(P) = \mathbb{E}\left([N - \mathbb{E}(N)][M - \mathbb{E}(M)]\right) \text{ with } (N, M) \sim P. \tag{A3}$$

We will now show that the usual definitions (A2) and (A3) for the variance and covariance of distributions on a subset of the real line correspond to the 'metric' definitions (1) and (2), respectively when taking the Euclidean distance $d^2(i, j) = \|i - j\|_2^2$.
Introducing expression (A1) for the (linear) expectation operator into the variance definition (A2), we find

$$\begin{aligned}
\text{var}(p) &= \sum_{i \in \mathcal{N}} p(i) \left[i - \sum_{j \in \mathcal{N}} p(j) j \right]^2 \\
&= \sum_{i \in \mathcal{N}} p(i) \left[i^2 - 2i \sum_{j \in \mathcal{N}} p(j) j + \sum_{j,k \in \mathcal{N}} p(j) p(k) jk \right] \\
&= \sum_{i,j \in \mathcal{N}} p(i) p(j) \left[i^2 - ij \right] \\
&= \frac{1}{2} \sum_{i,j \in \mathcal{N}} p(i) p(j) \left[i^2 - 2ij + j^2 \right] \\
&= \frac{1}{2} \sum_{i,j \in \mathcal{N}} p(i) p(j) d^2(i, j)
\end{aligned}$$

as required. For the covariance, we start from the following identities:

$$\begin{cases} \text{cov}(P) = \mathbb{E}(NM) - \mathbb{E}(N)\mathbb{E}(M) \\ \mathbb{E}([N - M]^2) = \mathbb{E}(N^2) - 2\mathbb{E}(NM) + \mathbb{E}(M^2) \end{cases}$$

which follow from the definition of covariance (A3) and the expectation operator for functions of pairs of random variables. Combining both identities, we find

$$\begin{aligned}
\text{cov}(P) = \frac{1}{2} \big(&\mathbb{E}(N^2) - 2\mathbb{E}(N)\mathbb{E}(M) \\
&+ \mathbb{E}(M^2) - \mathbb{E}([N - M]^2) \big).
\end{aligned}$$

Expanding the first three terms in this expression, we find

$$\begin{aligned}
&\mathbb{E}(N^2) - 2\mathbb{E}(N)\mathbb{E}(M) + \mathbb{E}(M^2) \\
&= \sum_{i \in \mathcal{N}} \tilde{p}(i) i^2 - 2 \sum_{i \in \mathcal{N}} \tilde{p}(i) i \sum_{j \in \mathcal{N}} \tilde{q}(j) j + \sum_{j \in \mathcal{N}} \tilde{q}(j) j^2 \\
&= \sum_{i \in \mathcal{N}} \tilde{p}(i) i \left(i - \sum_{j \in \mathcal{N}} \tilde{q}(j) j \right) + \sum_{j \in \mathcal{N}} \tilde{q}(j) j \left(j - \sum_{i \in \mathcal{N}} \tilde{p}(i) i \right) \\
&= \sum_{i \in \mathcal{N}} \tilde{p}(i) i \left(\sum_{j \in \mathcal{N}} \tilde{q}(j) (i - j) \right) \\
&\quad + \sum_{j \in \mathcal{N}} \tilde{q}(j) j \left(\sum_{i \in \mathcal{N}} \tilde{p}(i) (j - i) \right) \\
&= \sum_{i,j \in \mathcal{N}} \tilde{p}(i) \tilde{q}(j) (i - j)^2.
\end{aligned}$$

Consequently, the covariance (A3) can be written as

$$\text{cov}(P) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} (\tilde{p}(i) \tilde{q}(j) - P(i, j))(i - j)^2$$

which corresponds to expression (2) for the Euclidean distance.

Appendix B: Variance of diffusion processes and an alternative distance measure

A well-studied dynamical process on networks is the diffusion process, where a time-dependent state $\mathbf{p}_t$ is defined on the nodes of a graph with Laplacian Q and evolves according to the diffusion equation (also called heat equation)

$$\frac{d}{dt}\mathbf{p}_t = -Q\mathbf{p}_t \text{ with solution } \mathbf{p}_t = e^{-Qt}\mathbf{p}_0 \text{ for } t > 0$$

for some initial state $\mathbf{p}_0$. From properties of the Laplacian matrix it follows that if the initial state of the process is a distribution then all further states will be distributions as well, i.e. $\mathbf{p}_0 \in \Delta_n \Rightarrow \mathbf{p}_t \in \Delta_n$ for all $t > 0$, with the uniform distribution $p_\infty = \mathbf{u}/n$ as eventual stationary state.

We now consider the variance $\mathrm{var}_\omega(\mathbf{p}_t)$ of the time-evolving distribution with respect to the square root effective resistance (3). Using the fact that $\Omega Q = 2I - 2\mathbf{u}\mathbf{p}^T$ (e.g. following (6)) with $\mathbf{p}$ as in Proposition 11, we find that this variance evolves as

$$\frac{d}{dt}\mathrm{var}_\omega(\mathbf{p}_t) = 2\|\mathbf{p}_t\|_2^2 - 2\mathbf{p}^T\mathbf{p}_t.$$

In this expression, the first term causes the variance to increase, while the second term causes a decreasing variance depending on how similar (in the inner-product sense) the probability at time t is to the vector $\mathbf{p}$. This effectively separates the probability simplex into two regions

$$\Delta_n^+ = \{\mathbf{q} \in \Delta_n : \|\mathbf{q}\|^2 > \mathbf{p}^T\mathbf{q}\}, \text{ and}$$
$$\Delta_n^- = \{\mathbf{q} \in \Delta_n : \|\mathbf{q}\|^2 < \mathbf{p}^T\mathbf{q}\}$$

where the variance either increases or decreases for the diffusion process.

In the particular case of node-transitive graphs, we find a simplified evolution of the variance $d\,\mathrm{var}_\omega(\mathbf{p}_t)/dt = 2\|\mathbf{p}_t\|^2 - 2/n$ due to the fact that $\mathbf{p} = \mathbf{u}/n$ (see proof of Proposition 9). Since $\|\mathbf{q}\|^2 \geq 1/n$ for all $\mathbf{q} \in \Delta_n$ with equality only for the uniform distribution, this shows that the diffusion process results in a strictly increasing variance up to the stationary state. In the case of arbitrary graphs, a similar evolution is reproduced if we consider the variance with respect to an alternative dissimilarity d in definition (1) as $\mathrm{var}_{Q^\dagger}(\mathbf{p}) = -\mathbf{p}^T Q^\dagger \mathbf{p}$. The time-evolution in this case is given by

$$\frac{d}{dt}\mathrm{var}_{Q^\dagger}(\mathbf{p}_t) = 2\|\mathbf{p}_t\|_2^2 - 2/n \geq 0$$

with equality if and only if $\mathbf{p}_t = \mathbf{p}_\infty = \mathbf{u}/n$. In other words, the diffusion process results in a time-increasing variance $\mathrm{var}_{Q^\dagger}$ up to the uniform stationary distribution, when the variance is maximised. Note the important differences between the two choices var_ω and $\mathrm{var}_{Q^\dagger}$, whose maximum distribution accumulates probability at the extremities of the graph in the former case, and spreads it uniformly in the latter case.

Appendix C: Strict concavity of the variance

A function f on a set $\mathcal{X}$ is called *strictly convex* if it satisfies

$$f(\theta x + (1-\theta)y) > \theta f(x) + (1-\theta)f(y)$$

for all distinct $x, y \in \mathcal{X}$ and $\theta \in (0,1)$. We now prove that $\mathrm{var}_\omega : \Delta_n \to \mathbb{R}$ is strictly concave.

Proof: For any $\theta \in [0,1]$ and $\mathbf{p}, \mathbf{q} \in \Delta_n$ we can write

$$\begin{aligned}
&\theta\mathbf{p}^T\Omega\mathbf{p} + (1-\theta)\mathbf{q}^T\Omega\mathbf{q} - [\theta\mathbf{p} + (1-\theta)\mathbf{q}]^T\Omega[\theta\mathbf{p} + (1-\theta)\mathbf{q}] \\
&= \theta(1-\theta)\left[\mathbf{p}^T\Omega\mathbf{p} + \mathbf{q}^T\Omega\mathbf{q} - 2\mathbf{p}^T\Omega\mathbf{q}\right] \\
&= \theta(1-\theta)(\mathbf{p}-\mathbf{q})^T\Omega(\mathbf{p}-\mathbf{q}) \\
&= -2\theta(1-\theta)(\mathbf{p}-\mathbf{q})^T Q^\dagger(\mathbf{p}-\mathbf{q}) \leq 0.
\end{aligned}$$

In the last line, we first use the fact that by definition of the effective resistance we have $\Omega = \mathbf{u}\zeta^T + \zeta\mathbf{u}^T - 2Q^\dagger$ where $\zeta = \mathrm{diag}(Q^\dagger)$, such that $\mathbf{x}^T\Omega\mathbf{x} = -2\mathbf{x}^T Q^\dagger\mathbf{x}$ for all vectors $\mathbf{x} \perp \mathbf{u}$, and thus in particular for $(\mathbf{p}-\mathbf{q})$. Since the Laplacian matrix is positive semidefinite, its pseudoinverse $Q^\dagger$ is positive semidefinite as well, resulting in the inequality and thus establishing concavity of the variance. Moreover, since the (pseudoinverse) Laplacian has a single zero eigenvalue corresponding to the constant eigenvector and since $(\mathbf{p}-\mathbf{q}) \perp \mathbf{u}$ for distinct distributions, the inequality is strict. □

We remark that the concavity of the variance is not a generic property for all metrics d. For some metric on the nodes d with matrix $(D)_{ij} = d^2(i,j)$ we can write

$$\begin{aligned}
&\theta\,\mathrm{var}_d(\mathbf{p}) + (1-\theta)\,\mathrm{var}_d(\mathbf{p}) - \mathrm{var}_d(\theta\mathbf{p} + (1-\theta)\mathbf{q}) \\
&\quad = \theta(1-\theta)(\mathbf{p}-\mathbf{q})^T D(\mathbf{p}-\mathbf{q}).
\end{aligned}$$

This shows that the variance var_d is concave if and only if

$$(\mathbf{p}-\mathbf{q})^T D(\mathbf{p}-\mathbf{q}) \leq 0 \text{ for all } \mathbf{p} \neq \mathbf{q} \in \Delta_n. \qquad \text{(C1)}$$

A result from *distance geometry* states that criterion (C1) together with the conditions $d(i,i) = 0$ and $d(i,j) = d(j,i)$ for all i, j is satisfied if and only the metric space $(\mathcal{N}, \sqrt{d})$ can be embedded isometrically in Euclidean space [54], i.e. with a point $\mathbf{m}(i) \in \mathbb{R}^r$ for each i such that $\sqrt{d(i,j)} = \|\mathbf{m}(i) - \mathbf{m}(j)\|^2$. In Section VI we discuss how this 'embedding criterion' is satisfied for the effective resistance as discovered by Miroslav Fiedler [47, Thm. 1.2.4]. Similar criteria for distance matrices were also studied in [4].

Appendix D: Maximum variance support for tree graphs

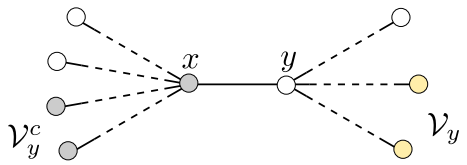
We show Proposition 7 which says that the maximum variance distribution on tree graphs is supported on the leaf nodes. This result was mentioned in [4] without proof, and was proven in [43] based on a perturbation

argument of the maximum variance distribution for an equivalent problem formulation.

Here, we present a proof by showing that having a non-leaf node in the maximum variance support contradicts the necessary conditions (C_1)–(C_3) for the maximum variance distribution. For our proof, we will make use of the fact that the effective resistance on trees corresponds to the shortest path distance (see e.g. [2]) and thus that

$$\omega_{ij} = \omega_{ix} + \omega_{xj} \tag{D1}$$

whenever the (unique) path from i to j passes through node x or, equivalently, when removing node x from the graph disconnects i and j. We will consider an arbitrary non-leaf node x in the tree graph and one of its neighbours $y \sim x$. We write $\mathcal{N}_y$ for the connected set of nodes after removal of x that contains node y, and write $\mathcal{V}_y = \mathcal{N}_y \cap \mathcal{V}^\star$ (which could be empty) and $\mathcal{V}_y^c = \mathcal{V}^\star \backslash \mathcal{V}$ (which always contains x) as in the figure below. Follow-



ing (D1), we thus have $\omega_{iy} = \omega_{ix} + \omega_{xy}$ for all $i \in \mathcal{V}_y^c$ and since removing y disconnects $\mathcal{V}_y$ from x we also have $\omega_{iy} = \omega_{ix} - \omega_{xy}$ for all $i \in \mathcal{V}_y$. We now give the proof by contradiction.

Proof of Proposition 7: Assume that a non-leaf node x has non-zero probability in the maximum variance distribution $\mathbf{p}^\star$, i.e. $x \in \mathcal{V}^\star$, and let y be a neighbour of x. We first show that $\mathcal{V}_y$ is non-empty. Assume otherwise (in particular $y \notin \mathcal{V}^\star$), then we can write

$$\begin{aligned}
&\sum_{i \in \mathcal{V}^\star (= \mathcal{V}_y^c)} p_i^\star \omega_{iy} - \sum_{k \in \mathcal{V}^\star} p_k^\star \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ik} \\
&= \sum_{i \in \mathcal{V}_y^c} p_i^\star (\omega_{ix} + \omega_{xy}) - \sum_{k \in \mathcal{V}^\star} p_k^\star \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ik} \\
&= \omega_{xy} - \sum_{k \in \mathcal{V}^\star} p_k^\star \sum_{i \in \mathcal{V}^\star} p_i^\star (\omega_{ik} - \omega_{ix}) \\
&= \omega_{xy}
\end{aligned}$$

where the last step uses (C_2) as $\sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ik} = \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ix}$. Invoking the global optimality condition we then find $(C_3) \Leftrightarrow \omega_{xy} < 0$ which contradicts the fact that $\omega_{xy} > 0$. Consequently, for every neighbour y of x the set $\mathcal{V}_y$ must be non-empty.

Next, we consider two cases for the neighbour y.

(Case 1) If $y \notin \mathcal{V}^\star$ then we can write

$$\begin{aligned}
&\sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{iy} - \sum_{k \in \mathcal{V}^\star} p_k^\star \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ik} \\
&= \sum_{i \in \mathcal{V}_y} p_i^\star (\omega_{ix} - \omega_{xy}) + \sum_{i \in \mathcal{V}_y^c} p_i^\star (\omega_{ix} + \omega_{xy}) \\
&\quad - \sum_{k \in \mathcal{V}^\star} p_k^\star \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ik} \\
&= \omega_{xy} \left(\sum_{i \in \mathcal{V}_y^c} p_i^\star - \sum_{i \in \mathcal{V}_y} p_i^\star \right) \\
&= \omega_{xy} \left(1 - 2 \sum_{i \in \mathcal{V}_y} p_i^\star \right)
\end{aligned}$$

The global optimality condition then gives that

$$(C_3) \Leftrightarrow \sum_{i \in \mathcal{V}_y} p_i^\star > \frac{1}{2} \text{ for all } y \sim x \text{ with } y \notin \mathcal{V}^\star$$

(Case 2) If $y \in \mathcal{V}^\star$ then we can write

$$\begin{aligned}
&\sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{ix} - \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{iy} \\
&= \sum_{i \in \mathcal{V}_y} p_i^\star (\omega_{iy} + \omega_{ix}) + \sum_{i \in \mathcal{V}_y^c} p_i^\star (\omega_{iy} - \omega_{ix}) - \sum_{i \in \mathcal{V}^\star} p_i^\star \omega_{iy} \\
&= \sum_{i \in \mathcal{V}_y^c} p_i^\star - \sum_{i \in \mathcal{V}_y} p_i^\star \\
&= 1 - 2 \sum_{i \in \mathcal{V}_y^c} p_i^\star
\end{aligned}$$

The local optimality condition then gives

$$(C_2) \Leftrightarrow \sum_{i \in \mathcal{V}_y} p_i^\star = \frac{1}{2} \text{ for all } y \sim x \text{ with } y \in \mathcal{V}^\star$$

Since each neighbour is either in the maximum variance support (Case 2) or not (Case 1), we have that

$$1 = \sum_{i \in \mathcal{V}^\star} p_i^\star = p_x^\star + \sum_{y \sim x} \sum_{i \in \mathcal{V}_y} p_i^\star > \sum_{y \sim x} \frac{1}{2}$$

which shows that x can have at most one neighbour. This contradicts our initial assumption that x is a non-leaf node and thus shows that no maximum variance distribution (satisfying (C_1)–(C_3)) can have a non-leaf node in the support, hence $\mathcal{V}^\star$ must be a subset of the leaf nodes. □

Appendix E: Maximum variance distribution on weighted stars and configuration graphs

We show Proposition 8 and Corollary 1 that characterize the maximum variance distributions on weighted

star graphs and configuration graphs, respectively. We first derive the solution for the weighted star using properties of tree graphs, and then show an equivalence to the maximum variance problem on configuration graphs which leads to the corollary.

We start with the following result about the solution to equation (C_2) for (the leaf nodes of) a weighted star:

Lemma 1 *For any subset $\mathcal{V}$ of the leaf nodes of a weighted star, the solution $\mathbf{p}$ to equation (C_2) is given by*

$$p_i = \frac{1}{2} - \frac{(v-2)k_i}{4m_{\mathcal{V}}} \text{ for all } i \in \mathcal{V} \tag{E1}$$

where $2m_{\mathcal{V}} = \sum_{i\in\mathcal{V}} k_i$.

Proof: We can assume without loss of generality that $\mathcal{V}$ is the full set of leaf nodes. If $\mathcal{V}$ is a subset of the leaf nodes, we can remove those leaf nodes not in $\mathcal{V}$ and obtain a weighted star (parametrized by degrees $\{k_i\}_{i\in\mathcal{V}}$) with leaf nodes $\mathcal{V}$, to which the Lemma is then applicable. As pointed out in Appendix D, the effective resistance on tree graphs (and thus weighted stars) equals the shortest path distance. For the leaf nodes in a weighted star, which are connected by a single, weighted link to node 0, we thus find that $\omega_{io} = c_{io}^{-1} = k_i^{-1}$. Furthermore, for pairs of leaf nodes we find

$$\omega_{ij} = \omega_{i0} + \omega_{0j} = k_i^{-1} + k_j^{-1} \text{ for leaf nodes } i, j$$

since removing $\{0\}$ disconnects all leaf nodes. If we let $\tilde{\mathbf{k}} = (k_1^{-1}, \dots, k_n^{-1})$ we can thus write the resistance matrix between leaf nodes $\mathcal{V}$ as

$$\Omega_{\mathcal{V}\mathcal{V}} = \tilde{\mathbf{k}}\mathbf{u}^T + \mathbf{u}\tilde{\mathbf{k}}^T - 2\,\mathrm{diag}(\tilde{\mathbf{k}}).$$

We now show that the local optimality condition $\Omega_{\mathcal{V}\mathcal{V}}\mathbf{p} = (2\sigma^2)\mathbf{u}$ for the maximum variance solution $\mathbf{p}$ is satisfied by the proposed solution. Introducing solution (E1) in vector form into condition (C_2), we find

$$\begin{aligned}
\Omega_{\mathcal{V}\mathcal{V}}\mathbf{p} &= (\tilde{\mathbf{k}}\mathbf{u}^T + \mathbf{u}\tilde{\mathbf{k}}^T - 2\,\mathrm{diag}(\tilde{\mathbf{k}}))\left[\frac{1}{2}\mathbf{u} - \frac{(v-2)}{4m_{\mathcal{V}}}\mathbf{k}\right] \\
&= \left[\frac{v}{2} - \frac{(v-2)}{2} - 1\right]\tilde{\mathbf{k}} \\
&\quad + \left[\frac{\tilde{\mathbf{k}}^T\mathbf{u}}{2} - \frac{v(v-2)}{4m_{\mathcal{V}}} + \frac{(v-2)}{2m_{\mathcal{V}}}\right]\mathbf{u} \\
&= \left[\frac{\tilde{\mathbf{k}}^T\mathbf{u}}{2} - \frac{(v-2)^2}{2m_{\mathcal{V}}}\right]\mathbf{u}
\end{aligned}$$

as required, which completes the proof. Alternatively, (E1) can be derived from equation $\Omega_{\mathcal{V}\mathcal{V}}\mathbf{p} = 2\sigma^2\mathbf{u}$ by introducing the found effective resistance matrix. □

We can now continue to prove Proposition 8, by showing that the solution $\mathbf{p}$ to equation (C_2) described in Lemma 1 together with conditions (C_1) and (C_3) are equivalent to the conditions described in the proposition. We assume all degrees to be ordered in *increasing* order, i.e. with $k_i \le k_{i+1}$ for all $i < n$.

Proof of Proposition 8: From Lemma 1 we know that the local optimality condition (C_2) is equivalent to (E1). We now consider when the two other necessary and sufficient conditions are met.

Introducing expression (E1) for the solution to the local optimality condition into the positivity condition (C_1) for $\mathbf{p}$, we find that

$$(C_1) \Leftrightarrow k_i \le \frac{\sum_{j\in\mathcal{V}} k_j}{v-2} \text{ for all } i \in \mathcal{V}.$$

Next, if $\mathcal{V}$ is a subset of the leaf nodes we let $j \notin \mathcal{V}$ be one of the leaf nodes not in the support, and write

$$\begin{aligned}
&\sum_{i\in\mathcal{V}} p_i\omega_{ij} - \sum_{i,r\in\mathcal{V}} p_i p_r \omega_{ir} \\
&= \sum_{i\in\mathcal{V}} p_i(k_i^{-1} + k_j^{-1}) - \sum_{i,r\in\mathcal{V}} p_i p_r (k_i^{-1} + k_r^{-1}) + 2\sum_{i\in\mathcal{V}} p_i^2 k_i^{-1} \\
&= k_j^{-1} - \sum_{i\in\mathcal{V}} p_i k_i^{-1} + 2\sum_{i\in\mathcal{V}} p_i^2 k_i^{-1} \\
&= k_j^{-1} - 2\sum_{i\in\mathcal{V}} p_i k_i^{-1}\left(\frac{1}{2} - p_i\right) \\
&= k_j^{-1} - \frac{(v-2)}{2m_{\mathcal{V}}}\sum_{i\in\mathcal{V}} p_i = k_j^{-1} - \frac{(v-2)}{2m_{\mathcal{V}}}.
\end{aligned}$$

Consequently, the global optimality condition (C_3) is equivalent to

$$(C_3) \Leftrightarrow \frac{\sum_{j\in\mathcal{V}} k_j}{v-2} < k_i \text{ for all } i \notin \mathcal{V}$$

Taking into account the most extreme cases, the optimality conditions (C_1)–(C_3) for the maximum variance set $\mathcal{V}^\star$ are thus

$$\max_{i\in\mathcal{V}^\star} k_i \le \frac{\sum_{j\in\mathcal{V}^\star} k_j}{v-2} < \min_{i\notin\mathcal{V}^\star} k_i$$

which in particular shows that $k_i < k_j$ for all $i \in \mathcal{V}^\star$ and $j \notin \mathcal{V}^\star$ and thus that $\mathcal{V}^\star$ can only be equal to a set of nodes with the ℓ smallest degrees. More precisely, making use of the ordering we find that $\mathcal{V}^\star = \{i \in \mathcal{N} : k_i \le k_\ell\}$ where k_ℓ satisfies

$$k_\ell \le \frac{\sum_{i=1}^{\ell} k_i}{\ell - 2} < k_{\ell+1}$$

and defining $k_{n+1} = \infty$ for consistency. Since the maximum variance support is a subset of the leaf nodes (by Proposition 7), the above expression fully characterizes the maximum variance support $\mathcal{V}^\star$. □

We now show that the result for weighted star graphs immediately implies the result for configuration graphs, based on the following correspondence between effective resistances in both graphs

Lemma 2 *The effective resistance ω_{ij} between pairs of nodes in a configuration graph and ω'_{ij} between pairs of leaf nodes in a weighted star parametrized by the same degree sequence* $\mathbf{k}$ *is proportional:*

$$\omega_{ij} = \frac{2m}{2m-1}\omega'_{ij} \tag{E2}$$

where i and j correspond to nodes with the same degrees in both graphs.

Proof: Our proof will make use of the Schur complement. An important property of the Schur complement of a Laplacian matrix is that it leaves the effective resistances invariant [52], see also Section VI. More precisely, if Ω is the effective resistance matrix of a Laplacian matrix Q, then the Schur complement $Q/\mathcal{V}^c$ with respect to some set $\mathcal{V}$ has effective resistance matrix $\Omega_{\mathcal{V}\mathcal{V}}$ equal to the submatrix of Ω with rows and columns in $\mathcal{V}$.
The Laplacian Q' of the weighted star can be written in block-form as

$$Q' = \begin{pmatrix} 2m & -\mathbf{k}^T \\ -\mathbf{k} & \operatorname{diag}(\mathbf{k}) \end{pmatrix}$$

where the first row and column correspond to node $\{0\}$ in the star. Taking the Schur complement with respect to the leaf nodes $\{0\}^c$ we find

$$Q'/\{0\} = \operatorname{diag}(\mathbf{k}) - \mathbf{k}\mathbf{k}^T/2m = \frac{2m}{2m-1}Q$$

with Q the Laplacian of the configuration graph. Since the Schur complement leaves the effective resistances invariant, we thus find that

$$\Omega'_{\mathcal{V}\mathcal{V}} = \frac{2m-1}{2m}\Omega$$

as required. □
Corollary 1 now follows as a simply corollary of Lemma 1 and Proposition 8 combined:
Proof of Corollary 1: Since the maximum variance distribution on a weighted star is always supported on the leaf nodes, the maximum variance problem (5) for the distance matrix Ω' of the weighted star is the same as the maximum variance problem on the distance matrix $\Omega'_{\mathcal{V}\mathcal{V}}$, which is equivalent to the optimization problem on the rescaled distance matrix Ω of the configuration graph, up to a scaling of $\frac{m}{m-1}$ of the maximum variance solution $2\sigma^2$. The criterion for the maximum variance support $\mathcal{V}^\star$ for weighted star graphs given in Proposition 8 thus also describes the maximum variance support for configuration graphs as in Corollary 1. □

Appendix F: Reformulations of the local optimality condition

We prove Proposition 11 which describes equivalent characterizations for a vector $\mathbf{p} \in \mathbf{R}^n$.

Proof of Proposition 11: (i)⇔(ii) Using a similar derivation as in Section V A which considers a transfer of probability ϵ between nodes i and j, we find that a necessary condition for any solution $\mathbf{p}$ to the optimization problem in (ii) is that $\Omega\mathbf{p} = 2\operatorname{var}_\omega(\mathbf{p})\mathbf{u}$, in other words (ii)⇒(i). Since there is a unique solution to (i) this means it is also necessary and (i)⇒(ii).
(i)⇔(iii) To prove this result, we make use of a particular embedding $\mathbf{m}$ based on the $(n-1)$ eigenvectors $\mathbf{z}_k \in \mathbb{R}^n$ with $\mathbf{z}_k \perp \mathbf{u}$ and non-zero eigenvalues μ_k of the Laplacian Q as (see [49])

$$(\mathbf{m}(i))_k = (\mathbf{z}_k)_i \mu_k^{-1/2} \text{ for all } i \in \mathcal{N} \text{ and } k < n.$$

This embedding satisfies $\mathbf{m}(i)^T\mathbf{m}(j) = (Q^\dagger)_{ij}$ for all $i, j \in \mathcal{N}$, which means we can write the distance from a point $\mathbf{m}(\mathbf{p})$ to a vertex as

$$\|\mathbf{m}(\mathbf{p}) - \mathbf{m}(i)\|^2 = (e_i - \mathbf{p})^T Q^\dagger (e_i - \mathbf{p})$$

We then find that

$$\begin{aligned} &\|\mathbf{m}(\mathbf{p}) - \mathbf{m}(i)\|^2 - \|\mathbf{m}(\mathbf{p}) - \mathbf{m}(j)\|^2 \\ &= e_i^T Q^\dagger e_i - e_j^T Q^\dagger e_j - 2\mathbf{p}^T Q^\dagger (e_i - e_j) \\ &= (e_i - e_j)^T \Omega \mathbf{p} \end{aligned}$$

which shows that the distance from $\mathbf{m}(\mathbf{p})$ is the same to all vertices $\mathbf{m}(i)$ if and only if $\Omega\mathbf{p}$ is a constant vector. In this case, the radius of the circumsphere can furthermore be calculated as

$$\begin{aligned} \|\mathbf{m}(\mathbf{p}) - \mathbf{m}(i)\|^2 &= (\mathbf{p} - e_i)^T \mathbf{Q}^\dagger (\mathbf{p} - e_i) \\ &= -\frac{1}{2}(\mathbf{p} - e_i)^T \Omega (\mathbf{p} - e_i) \\ &= -\frac{1}{2}\left(\mathbf{p}^T\Omega\mathbf{p} - 2e_i^T\Omega\mathbf{p}\right) \\ &= \frac{1}{2}\mathbf{p}^T\Omega\mathbf{p} = \operatorname{var}_\omega(\mathbf{p}) \end{aligned}$$

This confirms that (i)⇔(iii)
(i)⇔(iv) Using the definition of the effective resistance matrix, we can write

$$\begin{aligned} &\Omega\mathbf{p} - 2\operatorname{var}_\omega(\mathbf{p})\mathbf{u} \\ &= \left[(I - \mathbf{u}\mathbf{u}^T/n)\zeta - 2Q^\dagger\mathbf{p}\right] \\ &\quad - \left[2\operatorname{var}_\omega(\mathbf{p}) - \zeta^T\mathbf{p} - \mathbf{u}^T\zeta/n\right]\mathbf{u}. \end{aligned}$$

Since the first term in square brackets is in $\mathbf{u}^\perp$ while the second term is parallel to $\mathbf{u}$ we know that (i) holds if and only if both terms are zero, and thus

$$2Q^\dagger\mathbf{p} = (I - \mathbf{u}\mathbf{u}^T/n)\zeta \text{ and } \operatorname{var}_\omega(\mathbf{p}) = \zeta^T\mathbf{p} - \mathbf{u}^T\zeta/n$$

which is satisfied if and only (iv) holds, i.e. when $\mathbf{p} = \frac{1}{2}Q\zeta + \mathbf{u}/n$ and with the corresponding variance $\operatorname{var}_\omega(\mathbf{p}) = \frac{1}{4}\zeta^T Q\zeta + \mathbf{u}^T\zeta/n$.
(i)⇔(v) Using Fiedlers identity (6) to encode (i), we can write

$$\begin{pmatrix} -2\operatorname{var}_\omega(\mathbf{p}) & \mathbf{p}^T \\ \mathbf{p} & -\frac{1}{2}Q \end{pmatrix}\begin{pmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega \end{pmatrix} = I \Rightarrow \mathbf{p}\mathbf{u}^T = \frac{1}{2}Q\Omega - I.$$

Hence, taking the i^{th} diagonal of this expression and using the fact that $\omega_{ii} = 0$, we find

$$p_i = 1 - \frac{1}{2}\sum_{j\sim i} c_{ij}\omega_{ij}$$

which proves (i)⇔(iv) and thus completes the proof. □
We now derive Proposition 12 which relates the solution $\mathbf{p}$ of (C_2) for the full node set, to the solution for subsets.
Proof of Proposition 12: We let $\mathcal{V} = \mathcal{N}\backslash\{x\}$ and from Fiedlers identity for the resistance matrix $\Omega_{\mathcal{V}\mathcal{V}}$ we find

$$p'_i = e_1^T \begin{pmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega_{\mathcal{V}\mathcal{V}} \end{pmatrix}^{-1} e_i.$$

where $\mathbf{p}'$ solves condition (C_2) for the set $\mathcal{V}$. This inverse resistance matrix is the inverse of a submatrix of the matrix $\left(\begin{smallmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega \end{smallmatrix}\right)$ which appears in Fiedlers identity (6) for the full node set. Using the Schur complement formula for block-matrix inversion (see also (7)), we find that

$$\begin{pmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \Omega_{\mathcal{V}\mathcal{V}} \end{pmatrix}^{-1} = \begin{pmatrix} -2\sigma^2 & \mathbf{p}_{\mathcal{V}}^T \\ \mathbf{p}_{\mathcal{V}} & \frac{-1}{2}Q_{\mathcal{V}\mathcal{V}} \end{pmatrix} + \frac{2}{k_x}\begin{pmatrix} p_x \\ \frac{1}{2}\mathbf{c}_x \end{pmatrix}\begin{pmatrix} p_x \\ \frac{1}{2}\mathbf{c}_x \end{pmatrix},$$

with vector $(\mathbf{c}_x)_i = c_{ix}$. Consequently, for p'_i and σ'^2 we find

$$\begin{cases} p'_i = p_i + c_{ix}/k_x \\ \mathrm{var}_\omega(\mathbf{p}') = \mathrm{var}_\omega(\mathbf{p})^2 - \dfrac{p_x^2}{k_x} \end{cases}$$

as required. □
Remark: To illustrate the convenience of Proposition 12 we note that it allows for a particularly simple proof of Lemma 1. In any weighted star on n leaf nodes, the solution to (C_2) gives $p'_i = 1/2$ for each leaf node and $p'_0 = (2-v)/2$ (e.g. by Proposition 11 (v)). Thus if we take the Schur complement with respect to $\{0\}^c$ we find by Proposition 12 that $p_i = 1/2 - (v-2)k_i/(4m_{\mathcal{V}})$ as required.

Appendix G: Application: iterated core-periphery decomposition

For large graphs, the k-core decomposition is used as a tool to visualise graphs, identify clusters of important nodes or map the hierarchical structures in a network [55]. This decomposition divides the nodes of a network into (overlapping) subgraphs by recursively deleting nodes with the smallest degree from a graph until all remaining nodes have at least a certain degree k: starting from $k = 1$, all nodes are in the 1-core; then all nodes of degree one are removed consecutively, until all remaining nodes have degree two, which makes up the 2-core. This process is then repeated until no nodes remain.
We now consider an adaptation of the k-core decomposition by consecutively removing the peripheral nodes, defined as the maximum variance support. This procedure gives a recursive definition of the k^{th} core $\mathcal{C}_k \subseteq \mathcal{N}$ as

$$\begin{cases} 1\text{-core: } \mathcal{C}_1 = \mathcal{N}\backslash\mathcal{V}^\star(G) \\ k\text{-core: } \mathcal{C}_k = \mathcal{C}_{k-1}\backslash\mathcal{V}^\star(G_{k-1}) \text{ for } k > 1 \end{cases}$$

where G_k is the induced subgraph on the nodes $\mathcal{C}_k$. When a subgraph G_k is disconnected, the maximum variance support is calculated for each connected component separately. The coreness c_i of a node is determined by the 'highest' core it is part of, as $c_i = \max_k\{i \in \mathcal{C}_k\}$. Figure 8 below illustrates the k-core decomposition on two networks, which shows that our adapted decomposition can fulfill a similar visualisation or summarization role as the standard one.

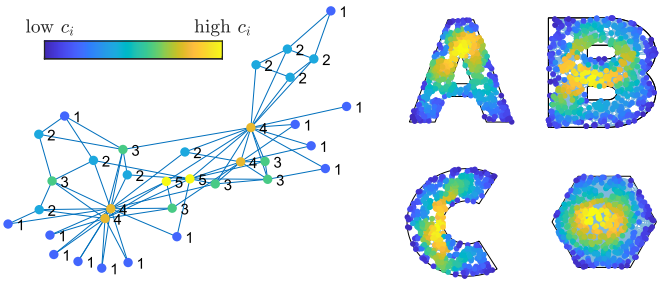


FIG. 8. The iterated core-periphery decomposition divides the nodes of a graph into layers of increasing centrality or 'coreness'. The figure above shows the coreness c_i of nodes in the Karate club network (obtained from the KONECT database [40]) and four random geometric graphs (see Section V C).