SIAM Review
Vol. 64, No. 2, pp. 343–359

Variance and Covariance of Distributions on Graphs*

Karel Devriendt†
Samuel Martin-Gutierrez‡
Renaud Lambiotte†

Abstract. 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. As a particular application, we define the *maximum variance problem* on graphs with respect to the effective resistance distance, and we characterize the solutions to this problem both numerically and theoretically. We show how the maximum variance distribution is concentrated on the boundary of the graph, and illustrate this in the case of 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.

Key words. network analysis, variance and covariance, diversity measure, effective resistance, geometric network, Wikipedia network, bibliographic network

AMS subject classifications. 60B99, 05C12, 05C69, 90C35, 05C82, 05C85

DOI. 10.1137/20M1361328

1. 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 that 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

*Received by the editors August 21, 2020; accepted for publication (in revised form) August 17, 2021; published electronically May 5, 2022.
https://doi.org/10.1137/20M1361328
Funding: The work of the first author was supported by The Alan Turing Institute under EPSRC grant EP/N510129/1. The work of the second author was supported by the Spanish Ministry of Science, Innovation and Universities (MICIU) under contract PGC2018-093854-B-100, by the Spanish Ministry of Education, Culture and Sport under grant FPU15/01461, and by the Austrian research agency (FFG) through project 873927. The work of the third author was supported by EPSRC grants EP/V013068/1 and EP/V03474X/1.

†Mathematical Institute, University of Oxford, Oxford, UK, and The Alan Turing Institute, London, UK (karel.devriendt@maths.ox.ac.uk, renaud.lambiotte@maths.ox.ac.uk).

‡Grupo de Sistemas Complejos, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Madrid, Spain, and Network Inequality Group, Complexity Science Hub Vienna, Wien, Austria (martin.gutierrez@csh.ac.at).

natural to take the underlying network structure into account. As the usual definition of variance cannot do so, we thus lack a basic methodological tool when analyzing distributions and signals on a graph.

In this article, we propose measures 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 us 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 (usage of concepts in papers) data 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, offering a new framework in which to interpret and understand them.

We furthermore consider the *maximum variance problem*, which seeks to determine the largest possible variance on a given graph and to characterize the distribution(s) that attains it. As a theoretical contribution, we find a complete characterization of the maximum variance distribution when considering the *effective resistance* as a distance measure between nodes. We show that this maximum variance distribution is concentrated on the *boundary* nodes of a graph and provide support for this intuition with experiments on random geometric graphs and further analytical results on a number of simple graphs.

To the best of our knowledge, the measures that we introduce for variance and covariance are new in the context of distributions on networks. Mathematically, our variance measure is a special case of Rao's well-known quadratic entropy [28, 33] which measures the diversity of distributions on general categorical variables with a notion of dissimilarity between the variables. This correspondence connects our work to the more general setting of *diversity measures* which are widely used, for instance, in ecology [30, 6]. The work of Leinster et al. on maximum diversity distributions [20, 19] is particularly relevant (see section 5). Additionally, the effective resistance matrix is well studied in the context of graph theory [2, 17, 37], and our variance quadratic form (2.3) and the related maximum variance problem (5.1) were previously defined and studied in different contexts. Hjorth et al. [15] studied (5.1) as a generalized diameter of metric spaces and described the conditions under which this problem has a unique solution. Dankelmann [7] defined the quadratic form in (2.3) for the geodesic graph distance and solved the corresponding maximization problem for tree and cycle graphs. These results are generalized from the geodesic graph distance to the effective resistance in [3, 11], 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 2 we introduce the relevant mathematical background on graphs and (joint) distributions on graphs and introduce our new variance and covariance measures. Section 3 gives an application of the (co)variance measures to a network of mathematical concepts. In section 4 we show how our new measures relate to a number of existing concepts. In section 5 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 maximum variance distributions. Section 6 concludes the article with a summary of the results and an outlook on further possible applications.

2. Variance and Covariance on Graphs.

2.1. 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*. The *degree*[1] of a node is defined as the combined weight of all links connected to a node, $k_i = \sum_j c_{ij}$. 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 numbers 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 the 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 3 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.

[1] The number of neighbors of a node i will be called the combinatorial degree to distinguish it from the "weighted" degree k_i.

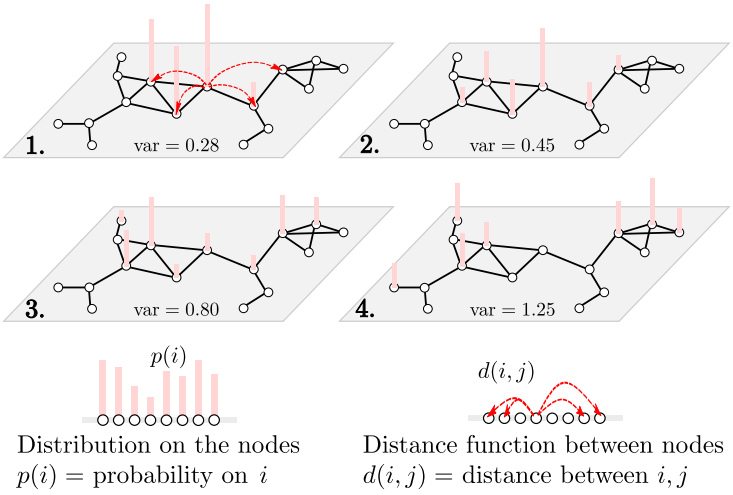


Fig. 1 *The variance of distributions on a graph can be defined with respect to a distance function between the nodes of the graph. This measure allows us to compare how "spread out" different distributions are on the network. In the example above, the distributions become more spread out over the network, going from distribution* 1. $\to$ 4. *with an increasing variance as a result; here, the variance is calculated with respect to the square root resistance distance as in* (2.3).

2.2. (Co)variance with Respect to Node Distances. As introduced above, the interpretation of a graph distribution is twofold: we can consider it as a signal (function) on the nodes of a graph (see also [32]) 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 whether 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 mean. To generalize this standard notion of variance to graphs, we propose to take into account distances between nodes in the graph, i.e., using 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 apart in the graph. Given such a distance notion between nodes, we propose a generalization of variance to distributions on a graph:

$$\operatorname{var}(p) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} p(i)p(j)d^2(i, j). \tag{2.1}$$

As shown in the supplementary material in section SM1, definition (2.1) 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 (2.1) 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 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:

$$\operatorname{cov}(P) = \frac{1}{2} \sum_{i,j \in \mathcal{N}} [\tilde{p}(i)\tilde{q}(j) - P(i,j)]\, d^2(i,j). \tag{2.2}$$

In section SM1 we show that (2.2) indeed generalizes the usual covariance, which is retrieved if $\mathcal{N} \subset \mathbb{R}$ with the usual Euclidean distance.

In the special case where (N, M) are independent random nodes with distributions p and q, respectively, the joint distribution equals $P(i,j) = p(i)q(j)$. As a consequence, we retrieve $\operatorname{cov}(P) = 0$ for independent random nodes. In the case where one random variable is a copy of the other, the joint distribution is a diagonal matrix with distribution given by the distribution of the random node on the diagonal, and we find $\operatorname{cov}(P) = \operatorname{var}(p)$. For more examples of joint distributions and their covariance, see section SM2 of the supplementary material.

2.3. Distance Functions on Graphs. Definitions (2.1) and (2.2) 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[2] of the shortest path between two nodes i and j. In addition to capturing the intuitive notion of a distance between nodes, the geodesic distance also satisfies the mathematical properties of a metric [5].

Another important metric between the nodes of a graph is the *effective resistance* [17, 2, 12, 27]. Similar to the geodesic distance, the effective resistance reflects the length of the paths between a pair of nodes. However, instead of only taking the shortest path into account, the effective resistance is influenced by all paths (and their lengths) between a pair of nodes, and becomes smaller as more paths are available. Due to this more integrative notion of distance and its nice mathematical properties, the effective resistance is often preferred over the shortest-path distance when studying networks. We write ω_{ij} for the effective resistance between two nodes i and j, and we define this resistance based on the Laplacian matrix of a graph. The Laplacian matrix Q of a graph with n nodes is an $n \times n$ matrix with entries $(Q)_{ii} = k_i$ on the diagonal, $(Q)_{ij} = -c_{ij}$ for all links (i,j), and zero otherwise, and can be used to define the effective resistance as

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

where the unit vectors have entries $(e_i)_k = 1$ if $k = i$, and zero otherwise, and where $Q^\dagger$ is the Moore–Penrose pseudoinverse of the Laplacian. For our application, the most relevant property is that both ω and its square root $\sqrt{\omega}$ are metrics between the nodes of a network [17, 9]. With the choice of the square root effective resistance as a distance between nodes, the variance and covariance can be written in matrix form as

$$\operatorname{var}_\omega(p) = \tfrac{1}{2}\mathbf{p}^T \Omega \mathbf{p}, \qquad \operatorname{cov}_\omega(P) = \tfrac{1}{2}\left[\tilde{\mathbf{p}}^T \Omega \tilde{\mathbf{q}} - \operatorname{tr}(P\Omega)\right], \tag{2.3}$$

with the $n \times n$ matrix Ω containing the effective resistances having entries $(\Omega)_{ij} = \omega_{ij}$; $\mathbf{p} = (p(i), \ldots, p(j))^T$ is the vector containing the probability for all nodes, and similarly for $\tilde{\mathbf{p}}, \tilde{\mathbf{q}}$, and matrix $(P)_{ij} = P(i,j)$ contains the probabilities of all pairs

[2] In the case of unweighted graphs, the length of a path 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 nodes acts as an affinity, where larger weights correspond to "better connected" nodes.

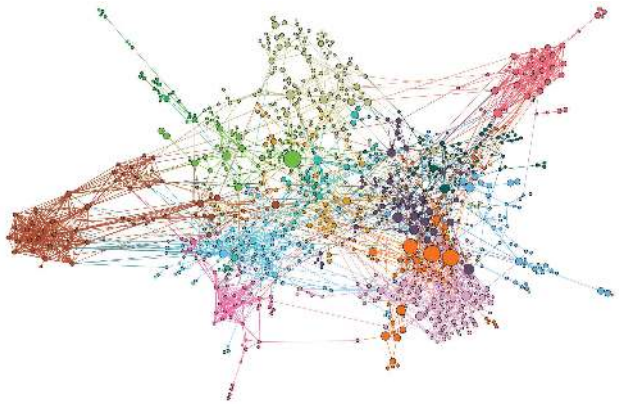


Fig. 2 *Hyperlink network of Wikipedia pages of the considered mathematical concepts (see* [29]*). The size of a node is proportional to its PageRank, and the color coding corresponds to communities found using the Louvain algorithm.*

of nodes. We will further also use p_i to denote an entry of the probability vector $\mathbf{p}$, which thus equals the probability $p(i)$. Replacing Ω by another distance matrix $(D)_{ij} = d^2(i,j)$ gives the general matrix form for variance and covariance.

3. Variance and Covariance in a Network of Knowledge. 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 for our analysis are available on GitHub [23]. 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 [29], 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. We consider this network to be the underlying structure of the mathematical concepts and use it to investigate how these concepts are used in practice by their occurrences in scientific papers.

To study 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 concepts used. Every paper i thus has a corresponding subset of concepts $\mathcal{V}_i$ and distribution $p^{(i)}$ uniform over this set of concepts.

A first question we consider is whether the mathematical papers contain "coherent" sets of mathematical concepts. In terms of variance, this question can be addressed by comparing the variance of the paper distributions $p^{(i)}$ with a null model, representing "virtual papers." Figure 3 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. This observation is confirmed by performing the one-sided Mann–Whitney U test, from which we find with high significance (p-values $<$ 1e-12) that the variance of a paper is typically smaller than what would be expected from the null model; more details on the test results are given in section SM3 of the supplementary material. Intuitively, this observation reflects the idea that the practical use of mathematical concepts is related

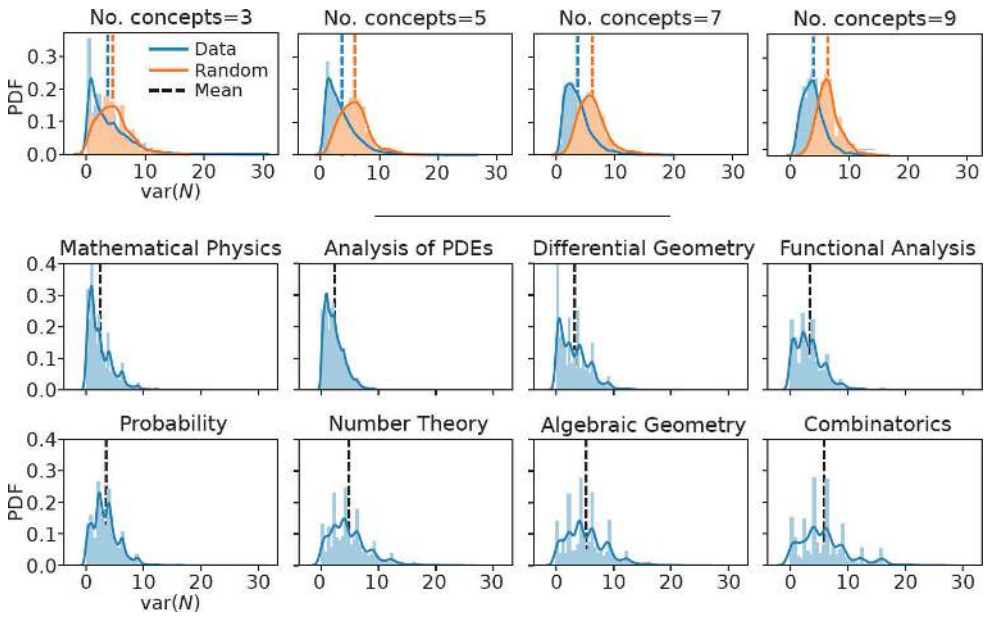
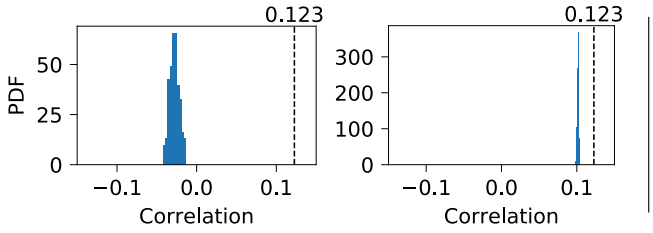


Fig. 3 *(Top figure) Distribution of network variances of concept distributions $p^{(i)}$ in papers, calculated with respect to the geodesic distance in the Wikipedia network. In each panel, the probability density function (pdf) of variances is calculated for papers containing* 3 *(resp.,* $5, 7, 9$*) 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 are concentrated on smaller variances compared to the null model variances, as confirmed by the one-sided Mann–Whitney U test (see section* SM3 *of the supplementary material). (Bottom figure) Distribution of network variances of concept distributions $p^{(i)}$ for arXiv papers from different subfields. The fields are ordered from top left to bottom right according to concentration on increasing variances, as found from the Mann–Whitney U test.*

to the underlying connections between concepts, where a group of concepts is more likely to be considered in a paper if that 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 grounds for (qualitative) comparison between different fields of study. In the bottom plots of Figure 3, for instance, we rank the different subfields based on their variance distributions, where the distributions on the top left are concentrated on smaller variances than the distributions on the bottom right, as quantified by one-sided Mann–Whitney U tests (see GitHub [23] for the test statistics and p-values). 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 cooccurrence 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 graph



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

Fig. 4 *(Left figure) Comparison between empirical covariance of the cooccurrence 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 (solid bars). In the right panel, the covariance of P (dashed line) and of* 100 *realizations of a marginal-preserving randomization of P (solid bars) are calculated with respect to the shortest-path distance on the Wikipedia network. (Right table) Correlation of the joint probability distribution of concepts for papers of different subfields.*

is randomized, and in the second model the distribution P is randomized while the graph is 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*:

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

Since the cooccurrence joint probability distribution is symmetric, we have $\tilde{p} = \tilde{q}$, which means that the correlation is simply calculated as $\operatorname{cov}(P)/\operatorname{var}(\tilde{p})$.

As a first null model, we perform a degree-preserving rewiring [14] of the Wikipedia network while keeping the joint distribution constant. As seen in the left panel of Figure 4, 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 leaving the network intact: we pick two pairs of nodes (i, j) and (i', j') with nonzero 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 that is given as $\alpha < \min(P(i, j), P(i', j'))$. 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 addressed, which produces a (symmetrically) randomized joint distribution P' with the same marginals as P. The right panel in Figure 4 shows that these randomized joint distributions are concentrated on significantly lower correlation values. 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 cooccurrences for a number of subfields separately and report their correlations in the table in the right panel in Figure 4.

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

covariance 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 is important.

4. 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 (2.1) and (2.2) 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.

4.1. 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 [21, 22, 10]. 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

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

In other words, the probability that the walker leaves a node via a particular link is proportional to the weight of that link.[3] For this type of random walker 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.

In [16], 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. Following a recent result on Kemeny's constant in [36], we find the following relation to a graph variance.

PROPOSITION 4.1. *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 4.1 follows from expression (2.3) for the graph variance with respect to the square root effective resistance and the relation between Kemeny's constant and the resistance matrix $\kappa = m\pi^T\Omega\pi$, as derived in [36]. ∎

One immediate consequence of Proposition 4.1 and the interpretation of Kemeny's constant as a variance is that it implies an upper-bound of $\kappa \leq 2m\sigma^2$, using the maximum variance results from section 5.

4.2. 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

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

[3] Other types of random walks are possible and are defined based on different transition probabilities.

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 [17, 25]—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 [12, 13]. Again, from expression (2.3) for the graph variance, we immediately find the following correspondence.

PROPOSITION 4.2. *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:*

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

Proof. Inserting the uniform distribution $(u/n)(i) = 1/n$ for all i in (2.1) gives the formula for R_G up to the constant n^2, which proves Proposition 4.2. ∎

We remark that the uniform distribution also appears naturally as the stationary state of diffusion processes on a graph; see section SM10 of the supplementary material.

4.3. Network Modularity. Similar to the variance, we find that formula (2.2) for the graph covariance is related to some known expressions in graph theory and its applications.

In the network science community there has been a large and sustained research effort in developing methods to identify groups of nodes which are tightly connected within but poorly connected between groups, i.e., to uncover the so-called *community structure*. Many results on this problem have been centered around the concept of *network modularity* [24], which assigns a "score" $M(g)$ to each partition $g : \mathcal{N} \to \{1, \dots, k\}$ of the nodes into k communities,

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

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. In section SM4 of the supplementary material we show that the modularity function M can be interpreted as the covariance of an appropriate joint distribution and distance function.

PROPOSITION 4.3. *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 distribution $P(i,j) = c_{ij}/2m$ and with respect to the distance $d_g = (1 - \delta_{g(i)g(j)})$:*

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

Proof. See section SM4 of the supplementary material. ∎

Proposition 4.3 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 it is that 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 $M(g)$. In section SM4 we show that a generalization of the modularity function, the Markov stability [8], also has a natural interpretation as a covariance.

5. 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 largest 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 $\mathrm{var}_\omega : \Delta_n \to \mathbb{R}$ defined by $\mathrm{var}_\omega(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\Omega\mathbf{p}$, 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 5.1. *The variance is strictly concave on Δ_n and bounded by $0 \leq \mathrm{var}_\omega(\mathbf{p}) \leq \sigma^2$. 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 section SM5 of the supplementary material 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 combination of these distributions $\theta\mathbf{p} + (1-\theta)\mathbf{q}$ for $\theta \in (0,1)$ has a higher variance than the corresponding convex combinations of their variances; in other words,

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

For the vertices $\mathbf{p}_i$ of Δ_n we immediately find that $\mathrm{var}(\mathbf{p}_i) = \omega_{ii}/2 = 0$. 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 distribution 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 $\mathbf{q}^\star$ were also to attain this maximum, we would find by strict concavity that $\mathrm{var}_\omega((\mathbf{p}^\star + \mathbf{q}^\star)/2) > \mathrm{var}_\omega(\mathbf{p}^\star)$; since this contradicts the maximality of $\mathbf{p}^\star$, the maximum variance distribution is necessarily unique. ∎

Proposition 5.1 thus states that there is a unique solution to the *maximum variance problem*

$$\begin{aligned} &\text{maximize } \frac{1}{2}\mathbf{p}^T\Omega\mathbf{p} \\ &\text{subject to } \mathbf{p} \in \Delta_n. \end{aligned} \tag{5.1}$$

We write the *maximum variance* as σ^2 and the *maximum variance distribution* that attains it as $\mathbf{p}^\star$. The support of the maximum variance distribution will be denoted by $\mathcal{V}^\star = \{i : \mathbf{p}_i^\star > 0\}$.

Numerical Solution. From Proposition 5.1 we know that the objective function of (5.1) is concave. Moreover, the optimization domain Δ_n is convex, which means that the maximum variance problem (5.1) 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 [4]. In other words, the maximum variance problem can be solved efficiently. This result was shown before in [11] in the context of calculating the average weighted resistance distance on a graph. In section SM8 of the supplementary material, we discuss how the maximum variance problem can be solved in practice using standard convex programming packages.

Analytical Solution. Apart from finding the maximum variance σ^2 and corresponding distribution $\mathbf{p}^\star$ numerically, we can further characterize this solution in terms of a number of necessary and sufficient conditions for $\mathbf{p}^\star$. As we will show, these conditions allow us to calculate exactly the maximum variance distribution (or its support) for a number of specific graph examples.

PROPOSITION 5.2. *The maximum variance distribution* $\mathbf{p}^\star$ *with respect to the square root effective resistance is the unique vector that satisfies the conditions*

$$(C_1) \qquad \mathbf{p} \in \Delta_n \text{with support } \mathcal{V},$$
$$(C_2) \qquad \Omega_{\mathcal{V}\mathcal{V}}\mathbf{p}_{\mathcal{V}} = 2\,\mathrm{var}_\omega(\mathbf{p})\mathbf{u},$$
$$(C_3) \qquad \Omega_{\mathcal{V}^c\mathcal{V}}\mathbf{p}_{\mathcal{V}} < 2\,\mathrm{var}_\omega(\mathbf{p})\mathbf{u},$$

where $(\Omega)_{\mathcal{A}\mathcal{B}}$ *denotes a submatrix of* Ω *with row (resp., column) indices in the set* $\mathcal{A}$ *(resp.,* $\mathcal{B}$*),* $(\mathbf{p})_{\mathcal{V}}$ *is a subvector of* $\mathbf{p}$ *with indices in* $\mathcal{V}$*, and with all-one vector* $\mathbf{u} = (1, \ldots, 1)^T$.

Proof. Proposition 5.2 is proven in section SM6 of the supplementary material in two ways: first, we derive (C_1), (C_2), and (C_3) as necessary conditions for the maximum variance distribution, and then we show that they are also sufficient by concavity of the variance. As a second derivation, we show that the conditions are equivalent to the Karush–Kuhn–Tucker conditions of the optimization problem (5.1). ∎

While conditions (C_1)–(C_3) might not look very intuitive at first sight, we can interpret them as follows: condition (C_1) is a *basic feasibility* criterion which states that $\mathbf{p}^\star$ must be a valid distribution supported on some set of nodes $\mathcal{V}$. Condition (C_2) is a *local optimality* criterion (local with respect to $\mathcal{V}$) which guarantees that the variance of $\mathbf{p}^\star$ can only decrease if we slightly change its distribution without changing the support $\mathcal{V}$. Condition (C_3) is finally a *global optimality* criterion (global with respect to $\mathcal{V}$) and guarantees that slightly changing the distribution of $\mathbf{p}^\star$ and changing its support can only result in a decreasing variance. Clearly these are necessary conditions, but as shown in section SM6, they are also sufficient and thus characterize the maximum variance distribution.

Maximum Variance Support $\mathcal{V}^\star$. Proposition 5.2 gives an exact description of what the maximum variance distribution looks like for a given graph; more details on the solution to (C_2) specifically can be found in section SM9 of the supplementary material. Somewhat surprisingly, the maximum variance 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 that 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 nodes near the boundary or periphery of the graph.[4] In section SM7 of the supplementary material we provide further theoretical support for this interpretation by characterizing $\mathcal{V}^\star$ on some simple graphs: trees, weighted stars, configuration graphs, node-transitive graphs, and paths, for which the maximum variance support indeed corresponds to the intuitive set of peripheral nodes.

As mentioned in the introduction, our formula for variance also appears in the context of *diversity measures*, where it is a special case of Rao's quadratic entropy [28] and relates to the Rényi entropy of order 2 when all nodes are assumed indistinguishable (i.e., in a complete graph). In the work of Leinster et al. [20, 19], these measures were further generalized to a family of diversity measures div_q on Δ_n that encompasses several other well-known measures such as the Gini–Simpson index, and our variance fits in this framework[5] as $\mathrm{div}_2(\mathbf{p}) = (1 - 2\,\mathrm{var}(\mathbf{p}))^{-1}$. This relation is particularly relevant in the context of problem (5.1), since it was shown [19, Thm. 6.3.2] that every similarity space has a unique distribution that *maximizes the diversity measure* div_q *for all* $q \in [0, \infty]$ *simultaneously*, which thus implies that the maximum variance distribution $\mathbf{p}^\star$ maximizes a whole family of diversity measures.

5.1. Maximum Variance Support in Random Geometric Graphs. The results in the previous section and the supplementary material 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 vein 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, [26].

In Figure 5 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 5, but we believe it could be related to the results of von Luxburg, Radl, and Hein [34, 35]. 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 (see section SM7). Consequently, we might conjecture that the maximum variance support in this regime converges according to the result described in Corollary SM7.3 for configuration graphs in the supplementary material, and that the maximum variance support will thus consist of low-degree nodes. As

[4] Our use of the term "peripheral" to describe the (geometric) notion of a boundary is different from its use in the context of core-periphery structures, where it is reserved to describe nodes a certain mesoscale connectivity structure in a graph. While in some cases these might overlap, this is certainly not guaranteed; tree graphs, for instance, can lack a core-periphery structure, but they clearly possess a boundary made up of the leaf nodes.

[5] Instead of describing diversity with respect to distances between points, the framework of Leinster et al. starts from similarities between points. As mentioned in [19], one possible mapping is "similarity = 1-distance," giving rise to a similarity matrix of $\mathbf{u}\mathbf{u}^T - \Omega$ in our case.

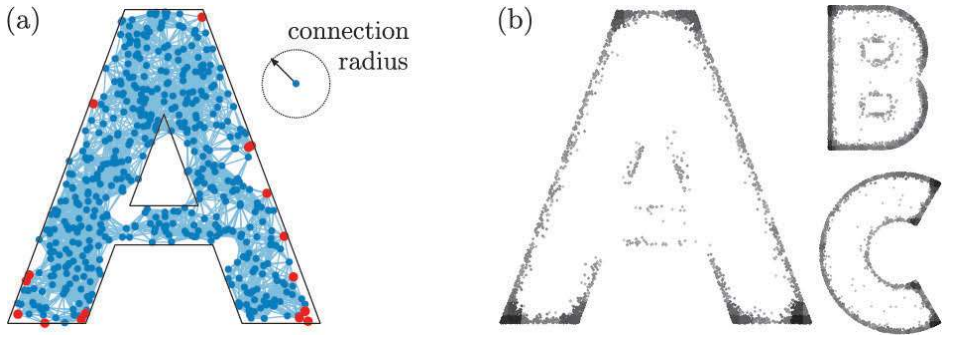


Fig. 5 *The maximum variance support of random geometric graphs (RGGs) in a bounded domain is more likely to contain nodes which are located near the boundaries. Panel* (a) *shows one realization of an RGG on* $n = 500$ *nodes in a domain shaped as the letter "A," with connection radius as indicated. 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."*

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 [35]), 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.

5.2. Application: k-Core Decomposition from the Maximum Variance Support. The *k-core decomposition* is a network analysis tool which is used to visualize large graphs, identify clusters of important nodes, or map the hierarchical structures in a network [1, 31]. 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 repeated until no nodes remain.

We now consider an adaptation of the k-core decomposition by consecutively removing the boundary nodes, defined as the maximum variance support. This procedure gives a recursive definition of the kth core $\mathcal{C}_k \subseteq \mathcal{N}$ as

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

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 $c_i = \text{argmax}_k\{i \in \mathcal{C}_k\}$. Figure 6 illustrates the k-core decomposition on some net-

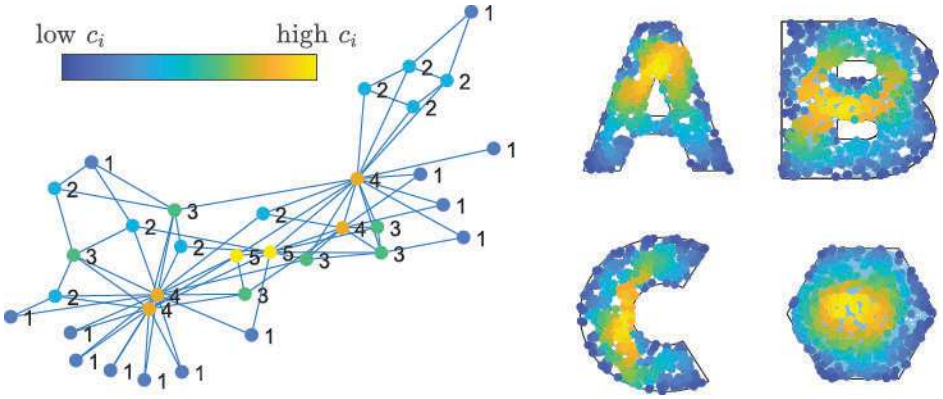


Fig. 6 *The iterated core-periphery decomposition divides the nodes of a graph into layers of increasing centrality or "coreness." The figure above shows the coreness c_i of nodes in the karate club network (obtained from the KONECT database* [18]*) and four random geometric graphs (see section* 5.1*).*

works, which shows that our adapted decomposition can fulfill a similar visualization or summarization role as the standard one.

6. Conclusion. In this paper, we have introduced two new measures which allow us 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 forms of a quadratic product and matrix trace, respectively, which are easily calculated using standard linear algebra solvers.

To support the specific forms of our introduced variance and covariance measures, we have shown 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.

We have furthermore used the variance and covariance in a practical scenario where both structural and functional data of a networked system are known. We analyzed 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 analyzed how this network of knowledge is used in practice. We translated the occurrences and cooccurrences of concepts in these papers into distributions and joint distributions on the network, and showed that the corresponding variance is smaller, on average, compared with a "virtual paper" null model and, similarly, that the corresponding covariance is larger, on average, compared with 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; or 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 considered 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 gave a detailed description of the *maximum variance problem* and the numerical and theoretical approaches to finding the maximum variance distribution $\mathbf{p}^\star$. We highlighted the interesting observation that, in general, the maximum variance distribution is supported on a subset of the nodes $\mathcal{V}^\star$, which can be interpreted as a set of peripheral/boundary nodes of the underlying graph. This interpretation is supported by analyzing a number of simple graphs such as trees and configuration-like graphs. Additional experimental evidence is found by considering random geometric graphs where we found that the maximum variance support nodes are most likely situated near the boundaries of the domain of the geometric graphs. Following the interpretation of the maximum variance support as a set of peripheral nodes, we proposed an application of $\mathcal{V}^\star$ in calculating a k-core decomposition of networks.

While the theoretical analyses in this article focus on measuring variance with respect to the effective resistance, we stress that definitions (2.1) and (2.2) work for any distance function on a graph, and in many cases distances other than the resistance distance are likely to be more natural.

REFERENCES

[1] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, *K-core decomposition of internet graphs: Hierarchies, self-similarity and measurement biases*, Netw. Heterog. Media, 3 (2008), pp. 371–393, https://doi.org/10.3934/nhm.2008.3.371. (Cited on p. 356)

[2] R. B. Bapat, *Graphs and Matrices*, Springer, London, 2010, https://doi.org/10.1007/978-1-84882-981-7. (Cited on pp. 344, 347)

[3] R. B. Bapat and S. K. Neogy, *On a quadratic programming problem involving distances in trees*, Ann. Oper. Res., 243 (2014), pp. 365–373, https://doi.org/10.1007/s10479-014-1743-y. (Cited on p. 344)

[4] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004, https://doi.org/10.1017/cbo9780511804441. (Cited on p. 354)

[5] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, Redwood City, CA, 1990. (Cited on p. 347)

[6] M. W. Cadotte, K. Carscadden, and N. Mirotchnick, *Beyond species: Functional diversity and the maintenance of ecological processes and services*, J. Appl. Ecol., 48 (2011), pp. 1079–1087, https://doi.org/10.1111/j.1365-2664.2011.02048.x. (Cited on p. 344)

[7] P. Dankelmann, *Average distance in weighted graphs*, Discrete Math., 312 (2012), pp. 12–20, https://doi.org/10.1016/j.disc.2011.02.010. (Cited on p. 344)

[8] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, *Stability of graph communities across time scales*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 12755–12760, https://doi.org/10.1073/pnas.0903215107. (Cited on p. 352)

[9] K. Devriendt, *Effective resistance is more than distance: Laplacians, simplices and the Schur complement*, Linear Algebra Appl., 639 (2022), pp. 24–49. https://arxiv.org/abs/2010.04521, 2020. (Cited on p. 347)

[10] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, Mathematical Association of America, 1984. (Cited on p. 351)

[11] D. Dubey and S. K. Neogy, *On solving a non-convex quadratic programming problem involving resistance distances in graphs*, Ann. Oper. Res., 287 (2018), pp. 643–651, https://doi.org/10.1007/s10479-018-3018-5. (Cited on pp. 344, 354)

[12] W. Ellens, F. Spieksma, P. V. Mieghem, A. Jamakovic, and R. Kooij, *Effective graph resistance*, Linear Algebra Appl., 435 (2011), pp. 2491–2506, https://doi.org/10.1016/j.laa.2011.02.024. (Cited on pp. 347, 352)

[13] A. Ghosh, S. Boyd, and A. Saberi, *Minimizing effective resistance of a graph*, SIAM Rev., 50 (2008), pp. 37–66, https://doi.org/10.1137/050645452. (Cited on p. 352)

[14] C. Gkantsidis, M. Mihail, and E. Zegura, *The Markov chain simulation method for generating connected power law random graphs*, in Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments, SIAM, Philadelphia, 2003, pp. 16–25. (Cited on p. 350)

[15] P. Hjorth, P. Lisoněk, S. Markvorsen, and C. Thomassen, *Finite metric spaces of strictly negative type*, Linear Algebra Appl., 270 (1998), pp. 255–273, https://doi.org/10.1016/s0024-3795(97)00242-5. (Cited on p. 344)
[16] J. G. Kemeny, *Generalization of a fundamental matrix*, Linear Algebra Appl., 38 (1981), pp. 193–206, https://doi.org/10.1016/0024-3795(81)90020-3. (Cited on p. 351)
[17] D. J. Klein and M. Randić, *Resistance distance*, J. Math. Chem., 12 (1993), pp. 81–95, https://doi.org/10.1007/bf01164627. (Cited on pp. 344, 347, 352)
[18] J. Kunegis, *KONECT: The Koblenz network collection*, in Proceedings of the 22nd International Conference on World Wide Web, ACM, New York, 2013, pp. 1343–1350, https://doi.org/10.1145/2487788.2488173. (Cited on p. 357)
[19] T. Leinster, *Entropy and Diversity: The Axiomatic Approach*, Cambridge University Press, Cambridge, UK, 2021. (Cited on pp. 344, 355)
[20] T. Leinster and C. A. Cobbold, *Measuring diversity: The importance of species similarity*, Ecology, 93 (2012), pp. 477–489, https://doi.org/10.1890/10-2402.1. (Cited on pp. 344, 355)
[21] D. A. Levin, Y. Peres, E. L. Wilmer, J. Propp, and D. B. Wilson, *Markov Chains and Mixing Times*, 2nd ed., AMS, Providence, RI, 2017. (Cited on p. 351)
[22] L. Lovász, *Random walks on graphs: A survey*, in Combinatorics, Paul Erdős Is Eighty, János Bolyai Math. Soc., Budapest, Hungary, 1993, pp. 353–398. (Cited on p. 351)
[23] S. Martin-Gutierrez, R. Lambiotte, and K. Devriendt, *Variance and Covariance of Distributions on Graphs*, https://github.com/rlambiot/variance, 2020. (Cited on pp. 348, 349)
[24] M. Newman, *Networks*, Oxford University Press, Oxford, UK, 2018, https://doi.org/10.1093/oso/9780198805090.001.0001. (Cited on p. 352)
[25] J. L. Palacios, *Closed-form formulas for Kirchhoff index*, Internat. J. Quantum Chem., 81 (2001), pp. 135–140, https://doi.org/10.1002/1097-461X(2001)81:2<135::AID-QUA4>3.0.CO;2-G. (Cited on p. 352)
[26] M. Penrose, *Random Geometric Graphs*, Oxford University Press, Oxford, UK, 2003, https://doi.org/10.1093/acprof:oso/9780198506263.001.0001. (Cited on p. 355)
[27] H. Qiu and E. R. Hancock, *Clustering and embedding using commute times*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 1873–1890, https://doi.org/10.1109/TPAMI.2007.1103. (Cited on p. 347)
[28] C. R. Rao, *Diversity: Its measurement, decomposition, apportionment and analysis*, Sankhyā Ser. A, 44 (1982), pp. 1–22. (Cited on pp. 344, 355)
[29] V. Salnikov, D. Cassese, R. Lambiotte, and N. S. Jones, *Co-occurrence simplicial complexes in mathematics: Identifying the holes of knowledge*, Appl. Network Sci., 3 (2018), art. 37, https://doi.org/10.1007/s41109-018-0074-3. (Cited on p. 348)
[30] D. Schleuter, M. Daufresne, F. Massol, and C. Argillier, *A user's guide to functional diversity indices*, Ecol. Monogr., 80 (2010), pp. 469–484, https://doi.org/10.1890/08-2225.1. (Cited on p. 344)
[31] S. B. Seidman, *Network structure and minimum degree*, Soc. Networks, 5 (1983), pp. 269–287, https://doi.org/10.1016/0378-8733(83)90028-X. (Cited on p. 356)
[32] 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 Process. Mag., 30 (2013), pp. 83–98, https://doi.org/10.1109/MSP.2012.2235192. (Cited on p. 346)
[33] A. Stirling, *A general framework for analysing diversity in science, technology and society*, J. Roy. Soc. Interface, 4 (2007), pp. 707–719, https://doi.org/10.1098/rsif.2007.0213. (Cited on p. 344)
[34] 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. (Cited on p. 355)
[35] U. von Luxburg, A. Radl, and M. Hein, *Hitting and commute times in large random neighborhood graphs*, J. Mach. Learn. Res., 15 (2014), pp. 1751–1798, http://jmlr.org/papers/v15/vonluxburg14a.html. (Cited on pp. 355, 356)
[36] X. Wang, J. L. Dubbeldam, and P. Van Mieghem, *Kemeny's constant and the effective graph resistance*, Linear Algebra Appl., 535 (2017), pp. 231–244, https://doi.org/10.1016/j.laa.2017.09.003. (Cited on p. 351)
[37] J. Zhou, Z. Wang, and C. Bu, *On the resistance matrix of a graph*, Electron. J. Combinatorics, 23 (2016), art. P1.41, https://doi.org/10.37236/5295. (Cited on p. 344)