Journal of Complex Networks (2020) **00**, 1–46
doi: 10.1093/comnet/cnaa026

Covariance and correlation measures on a graph in a generalized bag-of-paths formalism

GUILLAUME GUEX
SLI, Université de Lausanne, Lausanne 1002, Switzerland

SYLVAIN COURTAIN
LOURIM, Université catholique de Louvain, Louvain-la-neuve 1348, Belgium

AND

MARCO SAERENS[†]
ICTEAM, Université catholique de Louvain, Louvain-la-neuve 1348, Belgium
[†]Corresponding author. Email: marco.saerens@uclouvain.be

Edited by: Naoki Masuda

[Received on 24 December 2019; editorial decision on 17 July 2020; accepted on 22 July 2020]

This work derives closed-form expressions computing the expectation of co-presence and of number of co-occurrences of nodes on paths sampled from a network according to general path weights (a bag of paths). The underlying idea is that two nodes are considered as similar when they often appear together on (preferably short) paths of the network. The different expressions are obtained for both regular and hitting paths and serve as a basis for computing new covariance and correlation measures between nodes, which are valid positive semi-definite kernels on a graph. Experiments on semi-supervised classification problems show that the introduced similarity measures provide competitive results compared to other state-of-the-art distance and similarity measures between nodes.

Keywords: link analysis; network data analysis; complex networks; network science; graph mining; kernel on a graph; bag-of-paths model.

1. Introduction

1.1 *General introduction*

This work addresses the important problem of defining similarities and distances between nodes of a network based on its structure, faced in many applications such as link prediction, community detection, node classification and network visualization, among others [1–12]. It extends previous work on the *randomized shortest paths* and the *bag-of-paths* frameworks introduced in a series of papers [13–17], and most notably [18]. This effort was initially inspired by models developed in transportation science, especially [19, 20]. Of course, different meaningful notions of similarity between nodes can be defined, depending on the application (see [5, 12] for several examples). The most common one states that two nodes are considered as similar if (i) they are both close in the network (for instance in terms of shortest path distance) and highly inter-connected. In other words, two nodes are similar when they are highly

accessible from each other [21–23]. Another approach (ii) considers that two nodes are similar when they share some common properties, for instance based on features of their neighbourhoods [24–27], or when they often co-occur on trees or paths sampled from the network [18, 21]. This is the approach that will be investigated in this article by considering a general bag-of-paths approach. A third relevant idea states that (iii) two nodes could be considered as related when they play comparable roles in the network, for instance if they influence the network in a similar way [28, 29] or if they are structurally equivalent (see [12, 30] and references therein).

In this context, as discussed in [23] (see also [5, 14]), the specificity of our approach can be understood as follows. Most traditional distances or similarity measures between nodes are based on two common paradigms about the transfer of information, or more generally the movement, occurring in the network: optimal communication based on shortest paths and random communication based on a random walk on the graph. For instance, the shortest path distance is based on geodesics and the resistance distance ([31], proportional to the commute-time distance [32, 33]), on random walks. However, both the shortest path and the resistance distance suffer from some annoying drawbacks [5]: the shortest path distance does not integrate the amount of connectivity between the two nodes and produces many ties in unweighted networks, whereas random walks quickly loose the notion of proximity to the initial node when the graph becomes larger [34, 35].

Contrarily to these standard measures, the randomized shortest paths framework integrates both proximity and amount of connectivity for defining distance and betweenness measures between pairs of nodes based on approach (i) [5, 15, 16, 36, 37]. In its basic form, this model assumes that paths connecting the two nodes are chosen according to a Gibbs–Boltzmann probability distribution depending on a temperature parameter balancing the smoothness of the measure. When the temperature of the model is high, communication occurs through a random walk while, for low temperatures (close to zero), shorter paths are promoted. The model has been extended recently by adding a priori probabilities on the starting and ending nodes, thus allowing to weigh nodes [38, 39]. In the same spirit, the bag-of-paths based measures aim to quantify the similarity between the nodes by considering arbitrary paths between all pairs of nodes (and not only between one pair of predefined nodes) favouring low-cost, and thus short, paths [14, 18]. Similarity measures can then be derived by computing number of co-occurrences of nodes on paths, as in approach (ii). The present work expands on this idea by considering a more generic framework as well as deriving new measures based on node presence and on hitting paths in a systematic way. A nice property is that all the quantities of interest can be computed in closed form by standard matrix operations. Moreover, the introduced measures have nice intuitive interpretations, as explained in the next subsection.

Other such families of distances were recently introduced and studied, for instance in [40–42] by considering the co-occurrences of nodes in forests or walks on a graph, in [35, 43, 44] based on a generalization of the effective resistance in electric circuits, in [45, 46] by flow optimization with mixed L1–L2 norms, and in [13, 47] by considering network flows with entropy regularization. The originality of our work (see also [13]), in comparison with these previously developed methods, lies in the fact that we adopt a *bag-of-paths* formalism, that is, the quantities of interest are defined on the set of whole paths (or walks) appearing in the network.

For a comprehensive survey of related work on the design of similarity/distance measures on graphs and networks, see [14, 15, 18, 48] as well as [5]. However, three closely related and highly relevant works must be emphasized here. The first work [21, 22, 49] developed new similarity measures based on the co-occurrence of nodes on the same tree appearing on forests (sets of trees) sampled from the graph. The main quantity was called the relative forest accessibility between nodes and measures to which extend two nodes co-occur on the same tree and are therefore both close and highly connected.

The approach is similar to (and has inspired) our work but involves different motifs: trees instead of paths. More recently, distances derived from paths enumeration were investigated in [41]. Another interesting closely related work is [50] which introduced the concept of co-betweenness. The authors aimed to extend node betweenness centrality to sets of nodes in terms of shortest paths that pass through all nodes in a set. They then provided an expansion for group betweenness in terms of increasingly higher orders of co-betweenness. Moreover, they derived an efficient algorithm for computing the pairwise co-betweenness involving two nodes only, which generalizes the standard shortest-path node betweenness. Our work is similar in spirit but is based on paths sampled from the graph according to a general probability distribution favouring low-cost paths, instead of shortest paths. Finally, the recent DeepWalk [51] is based on simulating a large number of walks in the graph and uses a word2vec-like approach [52] to compute a p-dimensional representation of the nodes based on their co-occurrences in a sliding time frame. The algorithm provides a graph representation (or embedding) capturing the association between the nodes in terms of co-occurrence. In the present paper, the same idea is exploited, but the association measures between nodes (correlations and covariances) are computed in closed form instead of relying on computer-based path generation. The DeepWalk approach has become very popular in the deep learning community.

1.2 *The main intuition behind the models*

The idea behind the introduced node similarity measures is as follows. Let us assume we can enumerate all paths $\wp \in \mathcal{P}$ (the set of all paths) that can be sampled from a network G. Further consider that a measure of 'quality' of these paths can be easily computed, the weight $w(\wp)$, as in [41]. This weight reflects the reward of following the path and could be based on the length of the path, the total cost along the path, the time to follow the path, etc. Moreover, let us assume that paths are indexed and ordered by decreasing degree of quality, $(\wp_1, \wp_2, \dots)$. As an example let us consider a corpus of documents from which a network is extracted (see [8]). If the nodes of the network represent terms and the (weighted) links represent co-occurrences of terms within a given window in the text, then paths in this network correspond to sequences of words (sentences) where highly likely sequences have a higher weight.

The paths selection strategy, based on a sampling probability distribution derived from path weights, naturally favours high-weight paths and the overall quality level of the chosen paths could be monitored thanks to a temperature parameter (see Section 2.3.1). If the parameter is close to zero, only high-quality paths are considered whereas, for high values, all paths are more and more considered equally. The model can thus be considered as a bag of paths from which paths are drawn [14, 18]. Within this context, and as already mentioned, our similarity measure has the following interpretation: two nodes are considered as highly similar when they often co-occur on the same paths, when drawn thanks to a probability distribution favouring high-weight paths. In other words, two nodes are considered as related if they share the same paths.

Intuitively, this can be captured in the following way based on the enumeration of paths, although some tricks will be used in order to efficiently compute the quantities. Let $\mathbf{X}$ be a data matrix (the path-node matrix, inspired by the document-term matrix in information retrieval) with rows corresponding to paths, $\wp_i \in \mathcal{P}$, and columns to nodes, $j \in \{1 \cdots n\}$, of G. The entries i,j of this matrix are binary with a 1 if the node j appears on the path $\wp_i$ and zero otherwise.

Moreover, we further introduce a diagonal matrix weighting the paths, and containing as diagonal elements i,i the probability $\mathrm{P}(\wp_i)$ of choosing path $\wp_i$ according to the sampling distribution with $\sum_{i=1}^{\infty} \mathrm{P}(\wp_i) = 1$. Thus, as an example, these two matrices could be of the form:

$$\mathbf{X} = \begin{array}{c} \\ \wp_1 \\ \wp_2 \\ \wp_3 \\ \wp_4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & \cdots \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 0 & \cdots \\ 1 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}; \quad \mathbf{D} = \begin{array}{c} \\ \wp_1 \\ \wp_2 \\ \wp_3 \\ \wp_4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} \wp_1 & \wp_2 & \wp_3 & \wp_4 & \cdots \end{array} \\ \begin{bmatrix} \mathrm{P}(\wp_1) & 0 & 0 & 0 & \cdots \\ 0 & \mathrm{P}(\wp_2) & 0 & 0 & \cdots \\ 0 & 0 & \mathrm{P}(\wp_3) & 0 & \cdots \\ 0 & 0 & 0 & \mathrm{P}(\wp_4) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

In this example, we observe that node 1 and node 3 look similar as they appear on the same paths ($\wp_1$ and $\wp_3$). Conversely, nodes 1 and 2 are quite different.

In other words, nodes are characterized by their appearance on paths so that a binary feature vector indicating their presence on the different paths, $\mathbf{x}_j$, is associated to each node[1]. The $\mathbf{x}_j$ vectors can therefore be viewed as profile vectors characterizing each node j with respect to its presence on paths. These vectors form the column vectors of the data matrix $\mathbf{X}$. Alternatively, the data matrix could also contain the number of occurrences of each node on the different paths, instead of a binary presence value.

Now, a simple, but still meaningful, measure of the similarity between pairs of nodes i and j is simply the expected frequency of common presence on the paths. This quantity can be computed by taking the inner product $\mathbf{x}_i^{\mathrm{T}}\mathbf{D}\mathbf{x}_j$ for nodes i, j, or $\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}$ for all pairs of nodes. Such quantities will be studied and computed in this article in a general setting.

Note that this kind of similarity measure is closely related to contextual similarities used in information retrieval where two words are considered as related when they often appear in the same context (same sentence, window or document) [53], as illustrated by the recent, popular, word2vec method [52] and the emerging field of representation learning [54, 55]. Similarly, the context is represented here by paths of arbitrary length and the measures of association are computed directly in closed form from the structure of the graph and its edge weights. As an example, and as already mentioned, this idea has been exploited recently in the field of deep learning for computing a graph representation [51].

1.3 *Contributions and contents*

The paper derives closed-form expressions for computing similarity measures for two types of paths: *regular* (non-hitting) and *hitting* paths. More precisely, the derived similarities are the *covariance* and *correlation* kernels on a graph (which are positive semidefinite). Moreover, these kernels are defined for two different measures: based on (i) simple binary common *presence* of nodes on paths and (ii) *number of co-occurrences* on paths. In addition, various betweenness centrality measures are also derived within the framework.

The material presented in this article is therefore an extension of the previously published work [18] where a covariance and a correlation kernel were derived in the bag-of-paths framework for number of co-occurrences and regular paths only, and then slightly extended to hitting paths in [56]. In comparison with this previous work, the derivation of the different quantities in this work is more systematic, comprehensive and generic. Moreover, the results involving node presence (both for regular and hitting paths) are new and use a completely different technique than for those based on number of co-occurrences. Finally, the framework developed in this work is more general as it can be applied to a large class of weights defined on edges, whereas it was restricted to the bag-of-paths model based on Kullback–Leibler divergence

[1] Note that all vectors are considered as column vectors.

regularization in [14, 18]. The introduced framework contains the standard bag-of-paths model as a special case.

The article is organized as follows. First, the necessary background, notation and framework are detailed in Section 2. Then, Section 3 derives various important expressions computing the weights on sets of paths avoiding or containing some nodes. These results are derived for both regular and hitting paths and provide the basic support for the definition of the similarity measures. Section 4 develops betweenness centrality and node similarity measures based on the presence and the number of occurrences of nodes on paths, for both regular and hitting paths. Finally, those measures are assessed and compared in Section 5. Section 6 is the conclusion.

2. Framework and notation

2.1 *The generalized bag-of-paths formalism*

As already stated in the introduction, the standard *bag-of-paths* framework (BoP) [14, 18] sets up a *Gibbs–Boltzmann* distribution defining the probabilities of drawing a path from the *set of all paths in the graph*, also named the *bag of paths*, by assigning higher probabilities to short paths and lower probabilities to long paths. The standard bag-of-baths formalism will be described in more details in Section 2.3.1. For the moment, we will define a more generic framework, namely the *generalized bag-of-paths formalism*, inspired by walk distances [40–42]. A summary of main notation appears in Table 1.

2.1.1 *Paths, hitting-paths and sets of paths* Let $G = (\mathcal{V}, \mathcal{E})$ be a weighted, strongly connected, directed graph, with set of nodes $\mathcal{V} = \{1, 2, \ldots, n\}$ and set of edges $\mathcal{E} = \{(i, j)\}$ containing m elements. This graph is described by its *weighted adjacency matrix* $\mathbf{W} = (w_{ij})$ (or simply *weight matrix*), representing non-negative local affinities between nodes or rewards on edges, with $w_{ij} \geq 0$. Moreover, the weights must be equal to zero for missing links, $w_{ij} = 0$ when there is no link between i and j. This weight matrix is not simply the usual adjacency matrix $\mathbf{A}$ in general, but a function of it depending on the considered application like a substochastic transition matrix representing a killed random walk (see the end of Section 2.1 for some examples). The weight matrix will be used in order to define *path weights* and *path probabilities* on sets of paths. We will see later that it must satisfy some simple constraint in order to make sure that probabilities of drawing paths are well-defined: it should have a *spectral radius* strictly lower than 1, which implies that weights must be in the interval $w_{ij} \in [0, 1]$. This point will be discussed in Section 2.1.4.

A $\ell(\wp)$ or simply ℓ-length (regular) *path* or *walk* on the graph G, denoted by $\wp$, is a sequence of adjacent nodes $\wp = (i_0, i_1, \ldots, i_{\ell-1}, i_\ell)$, where $\ell \geq 0$ and $(i_\tau, i_{\tau+1}) \in \mathcal{E}$ for all $\tau = 0, \ldots, \ell - 1$. Note that, in this work, paths can contain repeated nodes; paths without any node repetition are called simple or self-avoiding paths. Moreover, 0-length paths are allowed by convention. We denote by the variable $\wp_{st}$ a path whose starting node is s and ending (target) node is t. Moreover, the node at position τ on path $\wp$ is referred to as $\wp(\tau)$. A *hitting path*, denoted by the superscript h in $\wp^{\mathrm{h}}$, is defined as a path such that the final node i_ℓ appears only once on the path, that is, $i_\ell \neq i_\tau$, $\forall \tau = 1, \ldots, \ell - 1$, and of 0 length if $i_0 = i_\ell$. In other words, for hitting paths, the final node can be considered as an absorbing node: it can only appear once, at the end of the path.

The *set of all paths* in G, also called the *bag-of-paths*, is denoted by $\mathcal{P}$. This set $\mathcal{P} = \cup_{s,t=1}^{n} \mathcal{P}_{st}$ is the union of the subsets $\mathcal{P}_{st}$ of all regular (non-hitting) paths starting in node s and ending in target node t. Several subsets of the bag of paths will be used in the sequel and, for the sake of clarity in further developments, we will use special symbols for these different subsets (see Table 1).

TABLE 1 *Summary of notation for the enumeration of paths in a graph G for both regular and hitting paths.*

$w_{ij} = [\mathbf{W}]_{ij}$	Element i, j of the weight matrix of graph G
$\wp$	A particular path visiting nodes $s = i_0, i_1, \ldots, i_\ell = t$
$w(\wp) = \prod_{\tau=0}^{\ell-1} w_{i_\tau, i_{\tau+1}}$	The total weight along path $\wp$ (product of edge weights)
$w(\mathcal{Q}) = \sum_{\wp \in \mathcal{Q}} w(\wp)$	The total weight for paths $\wp$ in set of paths $\mathcal{Q}$, $\wp \in \mathcal{Q}$
$\mathcal{P}_{st}(\ell)$	Set of regular paths connecting s to t in exactly ℓ steps
$\mathcal{P}_{st}$	Set of regular paths of arbitrary length connecting s to t
$\mathcal{P} = \cup_{s,t=1}^{n} \mathcal{P}_{st}$	Set of all regular paths of arbitrary length
$\mathcal{P}_{st}^{(+i)}$	Set of regular paths from s to t visiting intermediate node i
$\mathcal{P}_{st}^{(-i)}$	Set of regular paths from s to t avoiding node i
$\mathcal{P}_{st}^{(+\mathcal{I})}$	Set of regular paths from s to t visiting all nodes in set $\mathcal{I}$
$\mathcal{P}_{st}^{(-\mathcal{I})}$	Set of regular paths from s to t avoiding all nodes in set $\mathcal{I}$
$\mathcal{P}_{st}^{\mathrm{h}}$	Set of hitting paths of arbitrary length connecting s to t
$\mathcal{P}^{\mathrm{h}} = \cup_{s,t=1}^{n} \mathcal{P}_{st}^{\mathrm{h}}$	Set of all hitting paths of arbitrary length
$\mathcal{P}_{st}^{\mathrm{h}(+i)}$	Set of hitting paths from s to t visiting intermediate node i
$\mathcal{P}_{st}^{\mathrm{h}(-i)}$	Set of hitting paths from s to t avoiding node i
$\mathcal{P}_{st}^{\mathrm{h}(+\mathcal{I})}$	Set of hitting paths from s to t visiting all nodes in set $\mathcal{I}$
$\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{I})}$	Set of hitting paths from s to t avoiding all nodes in set $\mathcal{I}$

The superscript in $\mathcal{P}^{\mathrm{h}}$ will refer to the set of *hitting paths*, also named the *bag-of-hitting-paths*. Still another type of superscript has the form $\mathcal{P}^{(+\mathcal{I})}$ or $\mathcal{P}^{(-\mathcal{I})}$, where $\mathcal{I} \subset \mathcal{V}$ is a subset of nodes. This superscript indicates that the set is composed of paths *containing* (with a + symbol), or respectively *avoiding* (with a −), all of the nodes in $\mathcal{I}$. Note that we will usually simply write $\mathcal{P}^{(+i)}$ instead of $\mathcal{P}^{(+\{i\})}$ when there is only one node i in the set. Of course, all these notations can be combined. For example, $\mathcal{P}_{st}^{\mathrm{h}(-\{i,j\})}$ refers to the subset of hitting paths connecting s to t and avoiding nodes i and j.

Note that the set of all paths $\mathcal{P}$ is equipped with a composition rule for two paths where the ending node of one path corresponds to the starting node of the other: if $\wp_{sk} = (s, i_1, \ldots, k)$ and $\wp_{kt} = (k, j_1, \ldots, t)$, then $\wp_{sk} \circ \wp_{kt} = (s, i_1, \ldots, k, j_1, \ldots, t)$. By extension, let $\mathcal{P}_{sk} \circ \mathcal{P}_{kt}$ denote the set of all compositions between sets of paths $\mathcal{P}_{sk}$ and $\mathcal{P}_{kt}$.

2.1.2 *Paths weights* Paths weights are defined from the $n \times n$ non-negative *weighted adjacency matrix* or simply *weight matrix*, $\mathbf{W}$, of the graph (an example is provided later) and are used to define the bag-of-paths probabilities. The *weight* of a path $\wp = (i_0, \ldots, i_\ell)$, noted $w(\wp)$, is defined in the usual way [41] as the *product of the weights on its edges*,

$$w(\wp) \triangleq \prod_{\tau=0}^{\ell-1} w_{i_\tau, i_{\tau+1}}, \tag{2.1}$$

where we recall that ℓ is the length (number of edges) of the path. By convention, we assume that all 0-length paths have a weight of one. Note that this definition favours shorter paths over longer ones because the weight matrix will be restricted to irreducible sub-stochastic matrices, implying that $w_{ij} \in [0, 1]$.

Therefore, the weight of a path decreases with its length, at least in the long term. This type of weight matrix defines a transition probability matrix of a killed random walk (the random walker has a probability to stop his walk at each time step).

We also set the weight of any subset of the bag of paths $\mathcal{Q} \subseteq \mathcal{P}$ to the sum of the weights of its elements (paths),

$$w(\mathcal{Q}) \triangleq \sum_{\wp \in \mathcal{Q}} w(\wp). \tag{2.2}$$

Therefore, if two subsets, $\mathcal{Q}$ and $\mathcal{R}$, are disjoint, we have $w(\mathcal{Q} \cup \mathcal{R}) = w(\mathcal{Q}) + w(\mathcal{R})$.

2.1.3 *Bag-of-paths probabilities* We now consider the bag of paths $\mathcal{P}$ and we would like to draw paths from $\mathcal{P}$ with probabilities $\mathrm{P}(\wp)$ proportional to path weights. This is easily obtained by normalizing the weight of the path by the total weight of the bag-of-paths. In short, the *generalized bag-of-paths probabilities* of drawing a path $\wp$ are defined as

$$\mathrm{P}(\wp) \triangleq \frac{w(\wp)}{\sum_{\wp' \in \mathcal{P}} w(\wp')} = \frac{w(\wp)}{w(\mathcal{P})}. \tag{2.3}$$

Note that the generalized bag-of-paths probabilities $\mathrm{P}(\wp)$ are non-null if and only if the normalizing factor, the total weight of the bag-of-paths $w(\mathcal{P})$, is finite. This kind of measure is the object of interest of the present work. The necessary requirements on the weights in order to satisfy this property will be detailed in the next section.

More generally, if we want the conditional probability of drawing a path from a subset $\mathcal{P}^a \subseteq \mathcal{P}$ knowing that we are in $\mathcal{P}^b \subseteq \mathcal{P}$ (with $\mathcal{P}^a \subseteq \mathcal{P}^b$), we use

$$\mathrm{P}(\mathcal{P}^a|\mathcal{P}^b) \triangleq \mathrm{P}(\wp \in \mathcal{P}^a|\wp \in \mathcal{P}^b) = \frac{w(\mathcal{P}^a)}{w(\mathcal{P}^b)}. \tag{2.4}$$

As probabilities restrained on the bag-of-hitting-paths form an important part of this work, we will often use the notation $\mathrm{P}^{\mathrm{h}}(\mathcal{Q}^{\mathrm{h}}) \triangleq \mathrm{P}(\mathcal{Q}^{\mathrm{h}}|\mathcal{P}^{\mathrm{h}})$ for any subset $\mathcal{Q}^{\mathrm{h}} \subseteq \mathcal{P}^{\mathrm{h}}$. In other words, we define the *generalized bag-of-hitting-paths probabilities* of drawing the hitting-path $\wp^{\mathrm{h}}$ by

$$\mathrm{P}^{\mathrm{h}}(\wp^{\mathrm{h}}) = \frac{w(\wp^{\mathrm{h}})}{\sum_{\tilde{\wp}^{\mathrm{h}} \in \mathcal{P}^{\mathrm{h}}} w(\tilde{\wp}^{\mathrm{h}})} = \frac{w(\wp^{\mathrm{h}})}{w(\mathcal{P}^{\mathrm{h}})}. \tag{2.5}$$

2.1.4 *Consistency condition on the weighted adjacency matrix* $\mathbf{W}$ As stated before, we require well-defined (non-null) bag-of-paths probabilities, which means a finite weight for the whole bag of paths $\mathcal{P}$. Note that the subsets of paths $\mathcal{P}_{st}$ with $s, t = 1, \ldots, n$ form a partition of the bag-of-paths, which implies [14, 18]

$$w(\mathcal{P}) = \sum_{s,t \in \mathcal{V}} w(\mathcal{P}_{st}) = \sum_{s,t \in \mathcal{V}} \sum_{\wp_{st} \in \mathcal{P}_{st}} w(\wp_{st}) = \sum_{s,t \in \mathcal{V}} \sum_{\tau=0}^{\infty} \sum_{i_1, i_2, \ldots, i_{\tau-1}=1}^{n} w_{si_1} w_{i_1 i_2} \cdots w_{i_{\tau-1} t} = \sum_{s,t \in \mathcal{V}} \left[\sum_{\tau=0}^{\infty} \mathbf{W}^{\tau} \right]_{st} \tag{2.6}$$

for paths of increasing length τ. This shows that $w(\mathcal{P})$ is finite if and only if $\sum_{\tau=0}^{\infty} \mathbf{W}^{\tau}$ converges. Let $\rho(\mathbf{W})$ be the *spectral radius* of this weighted adjacency matrix, that is, its largest eigenvalue in modulus [57],

$$\rho(\mathbf{W}) \triangleq \max_{\lambda \in \sigma(\mathbf{W})} |\lambda|, \tag{2.7}$$

where $\sigma(\mathbf{W})$ is the *spectrum* of $\mathbf{W}$. By the Perron–Frobenius theorem (see [57, 58]), the non-negativity of the components of $\mathbf{W}$ and its irreducibility imply that the largest eigenvalue is real and positive, and thus $\lambda_1 \triangleq \rho(\mathbf{W}) \geq 0$. λ_1 is also called the *Perron eigenvalue*.

The series $\sum_{\tau=0}^{\infty} \mathbf{W}^{\tau} = \mathbf{I} + \mathbf{W} + \mathbf{W}^2 + \cdots$ is called the *Neumann series* of $\mathbf{W}$ [57]. Concerning Neumann series, the following statements are equivalent [57]

(1) $\rho(\mathbf{W}) < 1$,

(2) $\lim_{\tau \to \infty} \mathbf{W}^{\tau} = 0$,

(3) $\sum_{\tau=0}^{\infty} \mathbf{W}^{\tau}$ converges.

In this case, $(\mathbf{I} - \mathbf{W})^{-1}$ exists with $\sum_{\tau=0}^{\infty} \mathbf{W}^{\tau} = (\mathbf{I} - \mathbf{W})^{-1}$, and we can define $\mathbf{Z} = (z_{ij})$, the *fundamental matrix* of the bag-of-paths system, as

$$\mathbf{Z} \triangleq (\mathbf{I} - \mathbf{W})^{-1}. \tag{2.8}$$

This matrix is already well-known when $\mathbf{W}$ represents the adjacency matrix of a graph—it corresponds (up to a constant term) to the celebrated Katz index [59] after rescaling the adjacency matrix in order to obtain a convergent series. It was, e.g., recently used in order to define distance measures between nodes based on walks [41].

Thus, by restricting ourselves to graphs with a weighted adjacency matrix verifying $\rho(\mathbf{W}) < 1$, we ensure that bag-of-paths probabilities are well-defined. Now, it is known that each irreducible (non-negative) substochastic matrix has this property (see [57], p. 685, exercise 8.3.7). Therefore, the property holds for *strongly connected graphs* with a *substochastic weight matrix* $\mathbf{W}$ which, instead otherwise stated, will be assumed for now. More generally and intuitively, $\rho(\mathbf{W}) < 1$ should still hold provided that the weight matrix is substochastic and each node of G is connected through directed links to at least one killing node—a node whose (weighted) outdegree is strictly less than one. Another interesting property of an irreducible substochastic weight matrix $\mathbf{W}$ is that all of its elements are smaller or equal than 1 because its row sums are lesser or equal to 1.

In the case of a strongly connected graph and a weight matrix whose spectral radius is smaller than 1, several important quantities defined on subsets of the bag-of-paths can be expressed through the components of the fundamental matrix, as will be shown in Section 3.

2.2 *Nodes (co-)presence and (co-)occurrences on paths*

The introduced similarity measures between nodes will be based on node presences and node occurrences on paths. We now introduce some notations related to these quantities that will be used all along the article.

Let the *presence* variable of node i on a given observed path $\wp$ be

$$\delta(i \in \wp) \triangleq \begin{cases} 1 & \text{if } i \in \wp, \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

with $\delta(i \in \wp)$ being an indicator function or variable [60], equal to 1 if the event $i \in \wp$ is true (node i belongs to path $\wp$) and 0 otherwise.

Moreover, let the *number of occurrences*, or simply *occurrences*, variable for node i on path $\wp = (i_0, \ldots, i_\ell)$ be

$$\eta(i \in \wp) \triangleq \sum_{\tau=0}^{\ell(\wp)} \delta(\wp(\tau) = i) = \sum_{\tau=0}^{\ell(\wp)} \delta_{ii_\tau}, \tag{2.10}$$

where $\wp(\tau)$ denotes the node at position τ and δ_{ii_τ} is the *Kronecker delta* [61] between i and i_τ, that is, $\delta_{ii_\tau} = 1$ if $i_\tau = i$ (node at position τ on path $\wp$ is equal to node i) and $\delta_{ii_\tau} = 0$ otherwise. These variables are also used to signify the presence of two nodes: let *co-presences* and number of *co-occurrences* of nodes i and j on path $\wp$ be, respectively, $\delta(i \in \wp)\delta(j \in \wp)$ and $\eta(i \in \wp)\eta(j \in \wp)$.

In this work, we will mainly be interested in computing *covariances* and *correlations* of presence and occurrence variables between nodes, defined with respect to either bag-of-paths or bag-of-hitting-paths probabilities. These covariances and correlations between nodes are semi-definite positive by definition as they are inner product, or Gram, matrices [62]. They will be used in Section 5 as *kernel matrices* in order to perform semi-supervised classification tasks. Note that expected values of presence and occurrences also define *centrality indices*, generalizing some other centrality measures such as the *betweenness centrality* [63, 64] or the *random walk centrality* [65, 66]. However, studying these centrality indices is out of the scope of this work, and some were already investigated in [67].

2.3 *Some examples of weight matrix*

2.3.1 *A particular case: the standard bag-of-paths framework* The standard bag-of-paths framework is a good example of such a weighting scheme [14, 18, 36]. In that context, graph G is represented by its weighted adjacency matrix, $\mathbf{A} = (a_{ij})$, with no special requirement except that $a_{ij} \geq 0$ and the fact that the graph is strongly connected. This also allows us to derive a *reference transition probabilities matrix* $\mathbf{P}_{\text{ref}} = \mathbf{D}^{-1}\mathbf{A}$ of a natural random walk on G with elements p_{ij}^{ref}, where $\mathbf{D}$ is the diagonal matrix containing node out-degrees. Moreover, we assume a *cost matrix* $\mathbf{C} = (c_{ij})$, which can be defined either independently from weights a_{ij}, or thanks to $c_{ij} = 1/a_{ij}$ or $c_{ij} = 1$ (among others). In this context, any observed path $\wp = (i_0, \ldots, i_\ell)$ induces a random-walk *likelihood* $\pi^{\text{ref}}(\wp)$, defined by $\pi^{\text{ref}}(\wp) \triangleq \Pi_{\tau=0}^{\ell-1} p_{i_\tau, i_{\tau+1}}^{\text{ref}}$ and a *cost* $c(\wp)$ defined by $c(\wp) \triangleq \sum_{\tau=0}^{\ell-1} c_{i_\tau, i_{\tau+1}}$. The bag-of-paths probabilities are constructed in order to favour paths of low cost subject to a Kullback–Leibler divergence (KL) regularization level by solving the following problem, minimizing free energy [14, 18]:

$$\left| \begin{array}{ll} \underset{\{\mathrm{P}(\wp)\}_{\wp \in \mathcal{P}}}{\text{minimize}} & \displaystyle\sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) c(\wp) + T \sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) \log(\mathrm{P}(\wp)/\mathrm{P}_{\text{ref}}(\wp)) \\ \text{subject to} & \displaystyle\sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) = 1, \end{array} \right. \tag{2.11}$$

where $T > 0$, the *temperature* is a free parameter monitoring the KL level, and $\mathrm{P}^{\text{ref}}(\wp)$ is the random walk reference probability of a path $\wp$ proportional to its likelihood, i.e., $\mathrm{P}^{\text{ref}}(\wp) = \pi^{\text{ref}}(\wp) / \sum_{\wp' \in \mathcal{P}} \pi^{\text{ref}}(\wp')$.

As for maximum entropy problems [68–70], solving this problem yields a *Gibbs–Boltzmann probability distribution* [14, 18]

$$\mathrm{P}^\star(\wp) = \frac{\pi^{\mathrm{ref}}(\wp)\exp(-\beta c(\wp))}{\sum_{\wp'\in\mathcal{P}}\pi^{\mathrm{ref}}(\wp')\exp(-\beta c(\wp'))}, \tag{2.12}$$

where $\beta \triangleq 1/T$ is the *inverse temperature* parameter. This solution allows us to choose paths according to $\mathrm{P}^{\mathrm{ref}}(\wp)$ when the temperature T is high, and increases the probability of choosing low-cost paths as the temperature decreases, up to eventually selecting shortest paths only when $T \to 0^+$.

Interestingly, for a path $\wp = (i_0, \ldots, i_\ell)$, the numerator in (2.12) can be written as $\pi^{\mathrm{ref}}(\wp)\exp(-\beta c(\wp)) = \prod_{\tau=0}^{\ell-1} p^{\mathrm{ref}}_{i_\tau, i_{\tau+1}} \exp(-\beta c_{i_\tau, i_{\tau+1}})$, which means that it corresponds to a generalized path weight $w(\wp)$ built from the matrix $\mathbf{W}$ defined by

$$\mathbf{W} \triangleq \mathbf{P}_{\mathrm{ref}} \circ \exp[-\beta\mathbf{C}], \tag{2.13}$$

where $\exp[-\beta\mathbf{C}]$ is the component-wise exponential and $\circ$ the Hadamard product. We observe that if the cost matrix has at least one non-zero value for a given edge, the matrix $\mathbf{W}$ is substochastic. Moreover, when the graph is strongly connected, we saw that the substochastic weight matrix verifies $\rho(\mathbf{W}) < 1$, which implies that the standard bag-of-paths model is indeed a particular case of the generalized one. In the case studies of Section 5, this standard bag-of-paths framework will be used in the considered applications.

2.3.2 *Another particular case: absorbing Markov chains* Let us now consider absorbing Markov chains. In this context, some nodes, called *absorbing nodes* $\mathcal{A}$, are trapping the random walker so that the corresponding rows of the transition matrix of the Markov chain $\mathbf{P}$ are set to 0, except on the diagonal where we find a 1 [71–75]. However, in that case, the matrix $(\mathbf{I} - \mathbf{P})$, which would be a good candidate for the fundamental matrix of the process, is not of full rank and its standard inverse is not defined. Then, one has to rely on pseudo-inverses which introduces an extra level of complexity to the theory of finite Markov chains [71–73].

One simple trick solving this issue is to consider *killing* absorbing nodes, which simply aims at setting the corresponding rows of the original transition matrix (considered as stochastic and irreducible) to 0 [5, 76]. Thus, we consider that the random walker stops its walk after reaching such a killing absorbing node. The matrix of this killed process, denoted by $\mathbf{W}$, is then substochastic and has a spectral radius lesser than one, even if the corresponding graph is not strongly connected any more. Then, the fundamental matrix is simply $\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}$ and the quantities of interest can be computed from its elements. For instance, the probability of being killed in node $t \in \mathcal{A}$ when starting a random walk from node $s \notin \mathcal{A}$ is

$$\frac{\sum_{\wp st\in\mathcal{P}_{st}} w(\wp_{st})}{\sum_{a\in\mathcal{A}}\sum_{\wp sa\in\mathcal{P}_{sa}} w(\wp_{sa})} = \frac{z_{st}}{\sum_{a\in\mathcal{A}} z_{sa}},$$

which relies on Result (R.1) stated later, and is quite intuitive. Let us now turn to the computation of the various quantities of interest.

3. Basic expressions for computing weights of various subsets of paths

We saw how the generalized bag-of-paths formalism revolves around computing weights associated to subsets of paths in order to compute quantities of interest. In this section, we assume a strongly connected

graph as well as $\rho(\mathbf{W}) < 1$. In this case, we show that weights of various paths subsets can be computed directly from the components of the fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}$ (see (2.8)). When expressed properly, these results are quite intuitive, although their derivations can be rather tedious. Therefore, all proofs are transferred to the Appendices A and B.

Note that these developments only concern subsets where the starting and ending nodes are known exactly. However, the weights of subsets with undetermined starting and ending location are easily found. Indeed, sets $\mathcal{P}_{st}$ form a partition of $\mathcal{P}$, implying $w(\mathcal{P}) = \sum_{s,t\in\mathcal{V}} w(\mathcal{P}_{st})$. All subsets defined by a particular superscript can be handled in exactly the same way, e.g., $w\big(\mathcal{P}^{\mathrm{h}(-i)}\big) = \sum_{s,t\in\mathcal{V}} w\big(\mathcal{P}_{st}^{\mathrm{h}(-i)}\big)$. The main results (numbered (R.1)–(R.18)) are presented in a sequential way, from the most obvious to the most complex. Moreover, each derived quantity usually depends on the previous ones.

3.1 *Key elementary relationships*

As a starting point, the following closed-form expressions, already known in the standard bag-of-paths and randomized shortest paths frameworks [14, 15], are derived in Appendix A for both regular and hitting paths,

$$z_{st} = w\left(\mathcal{P}_{st}\right) = [\mathbf{Z}]_{ij} = \sum_{\wp\in\mathcal{P}_{st}} w(\wp), \tag{R.1}$$

$$z_{st}^{\mathrm{h}} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}}\right) = \sum_{\wp\in\mathcal{P}_{st}^{\mathrm{h}}} w(\wp) = \frac{z_{st}}{z_{tt}}, \tag{R.2}$$

and we observe that $z_{tt}^{\mathrm{h}} = 1$ from (R.2). Moreover, we also have $z_{st} = z_{st}^{\mathrm{h}} z_{tt}$ which will be useful later. Thus, the sum of the weights of regular paths between two nodes is given by the corresponding element of the fundamental matrix (2.8).

3.2 *Computing weights for sets of paths containing or avoiding nodes (node presence)*

In this subsection, a number of *auxiliary z-quantities* computing total weights of subsets of paths are defined and calculated. They serve as building blocks for computing association measures between nodes, like covariance and correlation. We start with quantities involving visits to one intermediate node, then we consider visiting two nodes, and we finally extend the results to an arbitrary number of nodes. A summary of the relevant notation appears in Table 2.

3.2.1 *Path weights containing or avoiding one node*

Expressions computing the total weight for sets of paths containing or avoiding a particular node i can be further developed for both regular and hitting paths (see Appendix A.1 for all derivations and details). For regular paths,

$$z_{st}^{(+i)} \triangleq w\left(\mathcal{P}_{st}^{(+i)}\right) = \sum_{\wp\in\mathcal{P}_{st}} \delta(i \in \wp)\, w(\wp) = z_{si}^{\mathrm{h}} z_{it}, \tag{R.3}$$

and we recall that indicator function $\delta(i \in \wp) = 1$ when path $\wp$ contains node i and 0 otherwise. It can be observed that $z_{st}^{(+i)}$ reduces to z_{st} when $i = s$ and when $i = t$, which is natural.

TABLE 2 *Notation for the z auxiliary variables computing total weights of subsets of paths on G, for both regular and hitting paths.*

$w_{ij} = [\mathbf{W}]_{ij}$	Element i,j of the weight matrix of graph G
$z_{st} = [\mathbf{Z}]_{st} = w(\mathcal{P}_{st})$	Element s,t of the fundamental matrix $\mathbf{Z}$ computing the total weight on regular paths
$z_{st}^{\mathrm{h}} = w\left(\mathcal{P}_{st}^{\mathrm{h}}\right)$	Total weight of hitting paths connecting s to t
$z_{st}^{(+i)} = w\left(\mathcal{P}_{st}^{(+i)}\right)$	Total weight of regular paths connecting s to t and visiting intermediate node i
$z_{st}^{\mathrm{h}(+i)} = w\left(\mathcal{P}_{st}^{\mathrm{h}(+i)}\right)$	Total weight of hitting paths connecting s to t and visiting intermediate node i
$z_{st}^{(-i)} = w\left(\mathcal{P}_{st}^{(-i)}\right)$	Total weight of regular paths connecting s to t and avoiding node i
$z_{st}^{\mathrm{h}(-i)} = w\left(\mathcal{P}_{st}^{\mathrm{h}(-i)}\right)$	Total weight of hitting paths connecting s to t and avoiding node i
$z_{st}^{(+\{i,j\})} = w\left(\mathcal{P}_{st}^{(+\{i,j\})}\right)$	Total weight of regular paths connecting s to t and visiting intermediate nodes i,j
$z_{st}^{\mathrm{h}(+\{i,j\})} = w\left(\mathcal{P}_{st}^{\mathrm{h}(+\{i,j\})}\right)$	Total weight of hitting paths connecting s to t and visiting intermediate nodes i,j
$z_{st}^{(-\{i,j\})} = w\left(\mathcal{P}_{st}^{(-\{i,j\})}\right)$	Total weight of regular paths connecting s to t and avoiding nodes i,j
$z_{st}^{\mathrm{h}(-\{i,j\})} = w\left(\mathcal{P}_{st}^{\mathrm{h}(-\{i,j\})}\right)$	Total weight of hitting paths connecting s to t and avoiding nodes i,j
$z_{st}^{(+\mathcal{I})} = w\left(\mathcal{P}_{st}^{(+\mathcal{I})}\right)$	Total weight of regular paths connecting s to t and visiting intermediate nodes in set $\mathcal{I}$
$z_{st}^{\mathrm{h}(+\mathcal{I})} = w\left(\mathcal{P}_{st}^{\mathrm{h}(+\mathcal{I})}\right)$	Total weight of hitting paths connecting s to t and visiting intermediate nodes in set $\mathcal{I}$
$z_{st}^{(-\mathcal{I})} = w\left(\mathcal{P}_{st}^{(-\mathcal{I})}\right)$	Total weight of regular paths connecting s to t and avoiding nodes in set $\mathcal{I}$
$z_{st}^{\mathrm{h}(-\mathcal{I})} = w\left(\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{I})}\right)$	Total weight of hitting paths connecting s to t and avoiding nodes in set $\mathcal{I}$

From Equation (R.3), we further obtain, when avoiding i,

$$z_{st}^{(-i)} \triangleq w\left(\mathcal{P}_{st}^{(-i)}\right) = \sum_{\wp \in \mathcal{P}_{st}} (1 - \delta(i \in \wp))\, w(\wp)$$
$$= z_{st} - z_{st}^{(+i)} = z_{st} - z_{si}^{\mathrm{h}} z_{it}, \qquad \textbf{(R.4)}$$

and this time, by (R.2), $z_{st}^{(-i)} = 0$ when $i = s$ and when $i = t$, as should be. For hitting paths now,

$$z_{st}^{\mathrm{h}(-i)} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(-i)}\right) = \sum_{\wp \in \mathcal{P}_{st}^{\mathrm{h}}} (1 - \delta(i \in \wp))\, w(\wp)$$

$$= \begin{cases} \dfrac{z_{st}^{(-i)}}{z_{tt}^{(-i)}} \\ 0 \end{cases} = \begin{cases} \dfrac{z_{st}^{\mathrm{h}} - z_{si}^{\mathrm{h}} z_{it}^{\mathrm{h}}}{1 - z_{ti}^{\mathrm{h}} z_{it}^{\mathrm{h}}} & \text{if } i \neq t \\ 0 & \text{if } i = t, \end{cases} \tag{R.5}$$

where $z_{st}^{\mathrm{h}(-i)} = 0$ when $i = s$, as should be. Observe the similarity with (R.2). Note that when $s = t$, the result is equal to 1. Considering now the presence of node i,

$$z_{st}^{\mathrm{h}(+i)} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(+i)}\right) = \sum_{\wp \in \mathcal{P}_{st}^{\mathrm{h}}} \delta(i \in \wp)\, w(\wp) = \begin{cases} z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}} & \text{if } i \neq t \\ z_{st}^{\mathrm{h}} & \text{if } i = t, \end{cases} \tag{R.6}$$

and we naturally obtain $z_{st}^{\mathrm{h}(+i)} = z_{st}^{\mathrm{h}}$ when $i = s$.

3.2.2 *Path weights containing or avoiding two nodes* In this subsection, we consider that $i \neq j$ and the expressions are only valid in this situation. When this is not the case, the problem is reduced to the task of finding only one node on paths, and its solution is provided in the previous section with, e.g., $w\left(\mathcal{P}_{st}^{(+\{i,i\})}\right) = w\left(\mathcal{P}_{st}^{(+i)}\right)$. In Appendix A.2, the expressions computing the total weight on sets of paths containing or avoiding two nodes of interest i and j, for both regular and hitting paths, are derived. The main results are summarized in (R.7)–(R.10),

$$z_{st}^{(+\{i,j\})} \triangleq w\left(\mathcal{P}_{st}^{(+\{i,j\})}\right) = \sum_{\wp \in \mathcal{P}_{st}} \delta(i \in \wp)\delta(j \in \wp)\, w(\wp)$$

$$= z_{sj}^{\mathrm{h}(+i)} z_{jt} + z_{si}^{\mathrm{h}(+j)} z_{it} \qquad \text{if } i \neq j. \tag{R.7}$$

Notice that this expression is coherent when $i = s, j = s, i = t$ and $j = t$ as it reduces to the expressions involving only one node (R.3). Three different expressions (R.8.1)–(R.8.3) can be derived for computing the next quantity,

$$z_{st}^{(-\{i,j\})} \triangleq w\left(\mathcal{P}_{st}^{(-\{i,j\})}\right) = \sum_{\wp \in \mathcal{P}_{st}} (1 - \delta(i \in \wp))(1 - \delta(j \in \wp))\, w(\wp)$$

$$= z_{st} - z_{si}^{\mathrm{h}(-j)} z_{it} - z_{sj}^{\mathrm{h}(-i)} z_{jt} \qquad \text{if } i \neq j \tag{R.8.1}$$

$$= z_{st}^{(-j)} - z_{si}^{\mathrm{h}(-j)} z_{it}^{(-j)} \qquad \text{if } i \neq j \tag{R.8.2}$$

$$= z_{st}^{(-i)} - z_{sj}^{\mathrm{h}(-i)} z_{jt}^{(-i)} \qquad \text{if } i \neq j. \tag{R.8.3}$$

The following results (R.9) and (R.10) for hitting paths assume that $i \neq j \neq t$. If either $i = t$ or $j = t$, the quantity (R.9) must be equal to 0 (the destination node t cannot be avoided). When $i = j$ (only one node is present or absent), the equations (R.5) and (R.6) must be used instead. Once again, three alternative expressions (R.9.1)–(R.9.3) can be derived from the more general result (R.9) for computing

the equivalent quantity for hitting paths,

$$z_{st}^{\mathrm{h}(-\{i,j\})} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(-\{i,j\})}\right) = \sum_{\wp \in \mathcal{P}_{st}^{\mathrm{h}}} (1-\delta(i\in\wp))(1-\delta(j\in\wp))\, w(\wp)$$

$$= \frac{z_{st}^{(-\{i,j\})}}{z_{tt}^{(-\{i,j\})}} \qquad \text{if } i,j \neq t \wedge i \neq j \qquad \textbf{(R.9)}$$

$$= 0 \qquad \text{if } i = t \vee j = t \qquad \textbf{(R.9)}$$

$$= \frac{z_{st}^{\mathrm{h}} - z_{si}^{\mathrm{h}(-j)} z_{it}^{\mathrm{h}} - z_{sj}^{\mathrm{h}(-i)} z_{jt}^{\mathrm{h}}}{1 - z_{ti}^{\mathrm{h}(-j)} z_{it}^{\mathrm{h}} - z_{tj}^{\mathrm{h}(-i)} z_{jt}^{\mathrm{h}}} \qquad \text{if } i,j \neq t \wedge i \neq j \qquad \textbf{(R.9.1)}$$

$$= \frac{z_{st}^{\mathrm{h}(-j)} - z_{si}^{\mathrm{h}(-j)} z_{it}^{\mathrm{h}(-j)}}{1 - z_{ti}^{\mathrm{h}(-j)} z_{it}^{\mathrm{h}(-j)}} \qquad \text{if } i,j \neq t \wedge i \neq j \qquad \textbf{(R.9.2)}$$

$$= \frac{z_{st}^{\mathrm{h}(-i)} - z_{sj}^{\mathrm{h}(-i)} z_{jt}^{\mathrm{h}(-i)}}{1 - z_{tj}^{\mathrm{h}(-i)} z_{jt}^{\mathrm{h}(-i)}} \qquad \text{if } i,j \neq t \wedge i \neq j. \qquad \textbf{(R.9.3)}$$

Note also that the result is equal to 1 when $s = t$. Finally, for node presence, we obtain

$$z_{st}^{\mathrm{h}(+\{i,j\})} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(+\{i,j\})}\right) = \sum_{\wp \in \mathcal{P}_{st}^{\mathrm{h}}} \delta(i\in\wp)\delta(j\in\wp)\, w(\wp)$$

$$= z_{si}^{\mathrm{h}(-\{j,t\})} z_{it}^{\mathrm{h}(+j)} + z_{sj}^{\mathrm{h}(-\{i,t\})} z_{jt}^{\mathrm{h}(+i)} \qquad \text{if } i \neq j, \qquad \textbf{(R.10)}$$

and, again, this expression is coherent when $i = s$, $j = s$, $i = t$ and $j = t$ as it reduces to (R.6). When $s = t$ and $i \neq j$, the result is simply 0. Now, if $i = j$, the quantity reduces to $z_{st}^{\mathrm{h}(+i)}$, which is not accounted for in (R.10).

3.2.3 *Path weights containing or avoiding sets of nodes* Up to now, we were mainly interested in computing weights on subsets of paths containing or avoiding one or two nodes. However, it is possible to deal with subsets with higher numbers of nodes through some recurrence formulae.

Let us assume we have a set of distinct nodes $\mathcal{I}$, where $s, t \notin \mathcal{I}$, and let $\mathcal{S}$ be a subset of $\mathcal{I}$, $\mathcal{S} \subset \mathcal{I}$. The weight of the regular paths avoiding all nodes in $\mathcal{S}$ is denoted as $z_{st}^{(-\mathcal{S})} = w\big(\mathcal{P}_{st}^{(-\mathcal{S})}\big)$ and the corresponding weight for hitting paths by $z_{st}^{\mathrm{h}(-\mathcal{S})} = w\big(\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{S})}\big)$. The same convention, but with a $+$ sign this time, is used for paths containing a set of nodes $\mathcal{S}$ (all nodes in $\mathcal{S}$ present). We would like to compute quantities like $z_{st}^{(+\mathcal{I})}$, $z_{st}^{\mathrm{h}(+\mathcal{I})}$, $z_{st}^{(-\mathcal{I})}$, and $z_{st}^{\mathrm{h}(-\mathcal{I})}$ for the set $\mathcal{I}$, in terms of some subsets $\mathcal{S}_i$ of $\mathcal{I}$.

First, let us compute the weights of avoiding paths. Let $\mathcal{S}_i \triangleq \mathcal{I} \setminus i$ with $i \in \mathcal{I}$, meaning that $\mathcal{I} = \mathcal{S}_i \cup \{i\}$. We obtain (see Appendix A.3)

$$z_{st}^{(-\mathcal{I})} \triangleq w\left(\mathcal{P}_{st}^{(-\mathcal{I})}\right) = z_{st}^{(-\mathcal{S}_i)} - z_{si}^{\mathrm{h}(-\mathcal{S}_i)} z_{it}^{(-\mathcal{S}_i)} \qquad \forall i \in \mathcal{I}, \qquad \textbf{(R.11)}$$

$$z_{st}^{\mathrm{h}(-\mathcal{I})} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{I})}\right) = \frac{z_{st}^{(-\mathcal{I})}}{z_{tt}^{(-\mathcal{I})}} = \frac{z_{st}^{\mathrm{h}(-\mathcal{S}_i)} - z_{si}^{\mathrm{h}(-\mathcal{S}_i)} z_{it}^{\mathrm{h}(-\mathcal{S}_i)}}{1 - z_{ti}^{\mathrm{h}(-\mathcal{S}_i)} z_{it}^{\mathrm{h}(-\mathcal{S}_i)}} \qquad \forall i \in \mathcal{I}. \qquad \textbf{(R.12)}$$

In addition, weights on sets of paths containing some predefined nodes in $\mathcal{I}$ can be obtained (see also Appendix A.3) from the previously computed quantities (R.11) and (R.12) thanks to

$$z_{st}^{(+\mathcal{I})} \triangleq w\left(\mathcal{P}_{st}^{(+\mathcal{I})}\right) = z_{st} + \sum_{\substack{\mathcal{S}\subseteq\mathcal{I}\\ \mathcal{S}\neq\varnothing}} (-1)^{|\mathcal{S}|} z_{st}^{(-\mathcal{S})} = \sum_{\mathcal{S}\subseteq\mathcal{I}} (-1)^{|\mathcal{S}|} z_{st}^{(-\mathcal{S})}, \tag{R.13}$$

$$z_{st}^{\mathrm{h}(+\mathcal{I})} \triangleq w\left(\mathcal{P}_{st}^{\mathrm{h}(+\mathcal{I})}\right) = z_{st}^{\mathrm{h}} + \sum_{\substack{\mathcal{S}\subseteq\mathcal{I}\\ \mathcal{S}\neq\varnothing}} (-1)^{|\mathcal{S}|} z_{st}^{\mathrm{h}(-\mathcal{S})} = \sum_{\mathcal{S}\subseteq\mathcal{I}} (-1)^{|\mathcal{S}|} z_{st}^{\mathrm{h}(-\mathcal{S})}, \tag{R.14}$$

where the summation on $\mathcal{S}\subseteq\mathcal{I}$ with $\mathcal{S}\neq\varnothing$ means a summation on all subsets $\mathcal{S}$ of $\mathcal{I}$, except the empty set. Moreover, by convention, $(-1)^0 = 1$.

Let us take an example with set $\mathcal{I} = \{i, j\}$ and regular paths. We easily obtain from (R.13)

$$z_{st}^{(+\{i,j\})} = z_{st} - z_{st}^{(-i)} - z_{st}^{(-j)} + z_{st}^{(-\{i,j\})},$$

and $z_{st}^{(-\{i,j\})}$ is computed with (R.11), $z_{st}^{(-\{i,j\})} = z_{st}^{(-j)} - z_{si}^{\mathrm{h}(-j)} z_{it}^{(-j)}$, which also corresponds to (R.8.2). It can be easily verified numerically that the obtained expression for $z_{st}^{(+\{i,j\})}$ provides the same results as (R.7).

We now turn to the computation of related quantities involving, this time, the number of visits (instead of presence) to some set of predefined nodes.

3.3 *Computing weights of node (co)-occurrences on sets of paths*

This section will derive equivalent results for the number of occurrences of nodes (and not simply the presence of the node as in the previous section) on paths of the graph. The first moments of occurrence variables of the type $\eta(i \in \wp)$ (the number of times node i is visited on path $\wp$) can be obtained, but in a very different way than in the previous subsections. Indeed, most of the results of this subsection are calculated by taking some partial derivatives with respect to path weights, as shown in Appendix B and already exploited in [18] for the standard bag-of-paths framework and regular paths. Note that we are not interested in paths avoiding nodes in this section as they correspond to paths with zero presence, which was covered in the previous section.

We now compute the weighted number of occurrences of some nodes on the sets of regular paths $\mathcal{P}$ and hitting paths $\mathcal{P}^{\mathrm{h}}$, expressed in terms of the fundamental matrix. This provides (see Appendix B for details), for regular paths between s and t,

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i \in \wp_{st})\, w(\wp_{st}) = z_{si} z_{it}, \tag{R.15}$$

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i \in \wp_{st})\eta(j \in \wp_{st})\, w(\wp_{st}) = z_{si} z_{ij} z_{jt} + z_{sj} z_{ji} z_{it} - \delta_{ij} z_{si} z_{jt}. \tag{R.16}$$

Note that this expression includes the $i = j$ case. For hitting paths, we have

$$\sum_{\wp_{st}^{\mathrm{h}}\in\mathcal{P}_{st}^{\mathrm{h}}} \eta(i \in \wp_{st}^{\mathrm{h}})\, w(\wp_{st}^{\mathrm{h}}) = z_{si}^{(-t)} z_{it}^{\mathrm{h}} + \delta_{it} z_{st}^{\mathrm{h}}, \tag{R.17}$$

and this quantity is equal to z^{h}_{st} when $i = t$. Moreover, for two nodes,

$$\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st}) \eta(j \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st}) = z^{(-t)}_{si} z^{(-t)}_{ij} z^{\mathrm{h}}_{jt} + z^{(-t)}_{sj} z^{(-t)}_{ji} z^{\mathrm{h}}_{it} - \delta_{ij} z^{(-t)}_{si} z^{\mathrm{h}}_{jt} + \delta_{jt} z^{(-t)}_{si} z^{\mathrm{h}}_{it} + \delta_{it} z^{(-t)}_{sj} z^{\mathrm{h}}_{jt} + \delta_{it}\delta_{jt} z^{\mathrm{h}}_{st}. \quad \textbf{(R.18)}$$

These various quantities will now be used in order to define centrality and association measures on nodes.

4. Betweenness and association measures

In this section, we will be interested in computing the expected value of (co-)presence as well as number of (co-)occurrences of nodes on paths (moments and co-moments), with respect to the generalized bag-of-paths and bag-of-hitting-paths probabilities. These quantities will allow us to compute the covariance and the correlation between node presence and occurrences. It is also shown that distance measures, extending the free energy distance [14, 15], can be derived from the quantities introduced so far. Note that still other similarity measures could be computed from the same expressions, like equivalents of cosine measure, Jaccard index, etc.

First, let us calculate the weights of the set of all paths $\mathcal{P}$ and the set of all hitting paths $\mathcal{P}^{\mathrm{h}}$, as these will be used as normalizing factors in order to compute our paths probabilities. We readily obtain from (R.1) and (R.2)

$$w(\mathcal{P}) = \sum_{s,t \in \mathcal{V}} w(\mathcal{P}_{st}) = z_{\bullet\bullet}, \qquad w(\mathcal{P}^{\mathrm{h}}) = \sum_{s,t \in \mathcal{V}} w(\mathcal{P}^{\mathrm{h}}_{st}) = z^{\mathrm{h}}_{\bullet\bullet}, \tag{4.1}$$

where $\bullet$ indicates summation over the corresponding index.

Then, recall that the probabilities of choosing a particular, regular, path $\wp$ (see Equation (2.3)) or a hitting path $\wp^{\mathrm{h}}$ (see Equation (2.5)) are simply

$$\mathrm{P}(\wp) = \frac{w(\wp)}{w(\mathcal{P})}, \qquad \mathrm{P}^{\mathrm{h}}(\wp) = \frac{w(\wp^{\mathrm{h}})}{w(\mathcal{P}^{\mathrm{h}})}. \tag{4.2}$$

Let us now compute the quantities of interest.

4.1 *Covariance and correlation for node presence*

From (R.1), (R.2), (R.3), (R.6), the expectation of node presence, regarding $\mathrm{P}(\wp)$ (regular paths framework) and $\mathrm{P}^{\mathrm{h}}(\wp^{\mathrm{h}})$ (hitting paths framework), is easily obtained thanks to

$$\mathbb{E}[\delta(i \in \wp)] = \sum_{\wp \in \mathcal{P}} \delta(i \in \wp)\, \mathrm{P}(\wp) = \frac{w\left(\mathcal{P}^{(+i)}\right)}{w(\mathcal{P})} = \frac{\sum_{s,t \in \mathcal{V}} w\left(\mathcal{P}^{(+i)}_{st}\right)}{\sum_{s,t \in \mathcal{V}} w\left(\mathcal{P}_{st}\right)} = \frac{z^{\mathrm{h}}_{\bullet i} z_{i\bullet}}{z_{\bullet\bullet}}, \tag{4.3}$$

$$\mathbb{E}^{\mathrm{h}}[\delta(i \in \wp^{\mathrm{h}})] = \sum_{\wp^{\mathrm{h}} \in \mathcal{P}^{\mathrm{h}}} \delta(i \in \wp^{\mathrm{h}})\, \mathrm{P}(\wp^{\mathrm{h}}) = \frac{w\left(\mathcal{P}^{\mathrm{h}(+i)}\right)}{w(\mathcal{P}^{\mathrm{h}})} = \frac{\sum_{\substack{t \in \mathcal{V} \\ t \neq i}} \left(z^{\mathrm{h}(-t)}_{\bullet i} z^{\mathrm{h}}_{it}\right) + z^{\mathrm{h}}_{\bullet i}}{z^{\mathrm{h}}_{\bullet\bullet}}, \tag{4.4}$$

where the last term is the contribution for $t = i$ as given in (R.6). These two quantities define *betweenness measures* based on node presence, quantifying to which extend each node is an important intermediary with respect to the communication (along paths) between pairs of nodes [67]. Communication along short paths is usually promoted by putting more weight on them, as in the standard bag of paths.

Similarly, from (R.7) and (R.10), the expected values of co-presence are

$$\mathbb{E}[\delta(i \in \wp)\delta(j \in \wp)] = \sum_{\wp \in \mathcal{P}} \delta(i \in \wp)\delta(j \in \wp)\,\mathrm{P}(\wp) = \frac{w\left(\mathcal{P}^{(+\{i,j\})}\right)}{w(\mathcal{P})} = \frac{z_{\bullet j}^{\mathrm{h}(+i)} z_{j\bullet} + z_{\bullet i}^{\mathrm{h}(+j)} z_{i\bullet} - \delta_{ij} z_{\bullet i}^{\mathrm{h}} z_{i\bullet}}{z_{\bullet\bullet}}, \tag{4.5}$$

$$\begin{aligned}\mathbb{E}^{\mathrm{h}}[\delta(i \in \wp^{\mathrm{h}})\delta(j \in \wp^{\mathrm{h}})] &= \sum_{\wp^{\mathrm{h}} \in \mathcal{P}^{\mathrm{h}}} \delta(i \in \wp^{\mathrm{h}})\delta(j \in \wp^{\mathrm{h}})\,\mathrm{P}(\wp^{\mathrm{h}}) = \frac{w\left(\mathcal{P}^{\mathrm{h}(+\{i,j\})}\right)}{w(\mathcal{P}^{\mathrm{h}})} \\ &= \frac{\sum_{\substack{t \in \mathcal{V} \\ t \neq i,j}} \left(z_{\bullet i}^{\mathrm{h}(-\{j,t\})} z_{it}^{\mathrm{h}(+j)} + z_{\bullet j}^{\mathrm{h}(-\{i,t\})} z_{jt}^{\mathrm{h}(+i)} - \delta_{ij} z_{\bullet i}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}} \right)}{z_{\bullet\bullet}^{\mathrm{h}}} + \frac{z_{\bullet i}^{\mathrm{h}(-j)} z_{ij}^{\mathrm{h}} + z_{\bullet j}^{\mathrm{h}(-i)} z_{ji}^{\mathrm{h}} + \delta_{ij} z_{\bullet i}^{\mathrm{h}}}{z_{\bullet\bullet}^{\mathrm{h}}},\end{aligned} \tag{4.6}$$

where terms like $-\delta_{ij} z_{\bullet i}^{\mathrm{h}} z_{i\bullet}$, $\delta_{ij} z_{\bullet i}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$ and $\delta_{ij} z_{\bullet i}^{\mathrm{h}}$ were added in order to take into account the special cases where $i = j$. Indeed, (R.7) and (R.10) concern only the situation where $i \neq j$ and must be extended in order to compute these diagonal terms. The details about these technicalities are transferred to Appendix C, for the interested readers.

These expected values are the building blocks needed to compute the *covariance* and the *correlation* measures between the common presence of two nodes on paths (the random variables $\delta(i \in \wp)$ and $\delta(j \in \wp)$). This provides, for regular paths,

$$\mathrm{cov}(\delta(i \in \wp), \delta(j \in \wp)) = \mathbb{E}[\delta(i \in \wp)\delta(j \in \wp)] - \mathbb{E}[\delta(i \in \wp)]\mathbb{E}[\delta(j \in \wp)], \tag{4.7}$$

$$\mathrm{cor}(\delta(i \in \wp), \delta(j \in \wp)) = \frac{\mathrm{cov}(\delta(i \in \wp), \delta(j \in \wp))}{\sqrt{\mathrm{cov}(\delta(i \in \wp), \delta(i \in \wp))\,\mathrm{cov}(\delta(j \in \wp), \delta(j \in \wp))}}, \tag{4.8}$$

and the expressions for hitting-paths probabilities are similar, with $\wp$ replaced by $\wp^{\mathrm{h}}$. The intuition behind these quantities is that two nodes are correlated when they often appear on the same, preferably short, paths.

4.2 *Covariance and correlation for the number of occurrences of nodes*

We now derive the same quantities for the number of occurrences of nodes on paths. From (R.15) and (R.17), we have for regular paths

$$\begin{aligned}\mathbb{E}[\eta(i \in \wp)] &= \sum_{\wp \in \mathcal{P}} \eta(i \in \wp)\mathrm{P}(\wp) = \sum_{\wp \in \mathcal{P}} \eta(i \in \wp)\frac{w(\wp)}{w(\mathcal{P})} \\ &= \sum_{s,t \in \mathcal{V}} \sum_{\wp_{st} \in \mathcal{P}_{st}} \eta(i \in \wp_{st})\frac{w(\wp_{st})}{w(\mathcal{P})} = \frac{z_{\bullet i} z_{i\bullet}}{z_{\bullet\bullet}},\end{aligned} \tag{4.9}$$

$$\mathbb{E}^{\mathrm{h}}[\eta(i \in \wp^{\mathrm{h}})] = \sum_{s,t \in \mathcal{V}} \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st}) \frac{w(\wp^{\mathrm{h}}_{st})}{w(\mathcal{P}^{\mathrm{h}})} = \frac{\sum_{t \neq i} z^{(-t)}_{\bullet i} z^{\mathrm{h}}_{it} + z^{\mathrm{h}}_{\bullet i}}{z^{\mathrm{h}}_{\bullet\bullet}}. \tag{4.10}$$

As before, these two quantities define *betweenness measures* based on node occurrences. For the expected values of co-occurrences, (R.16) and (R.18) provide

$$\begin{aligned}\mathbb{E}[\eta(i \in \wp)\eta(j \in \wp)] &= \sum_{s,t \in \mathcal{V}} \sum_{\wp_{st} \in \mathcal{P}_{st}} \eta(i \in \wp_{st})\eta(j \in \wp_{st}) \frac{w(\wp_{st})}{w(\mathcal{P})} \\ &= \frac{z_{\bullet i} z_{ij} z_{j\bullet} + z_{\bullet j} z_{ji} z_{i\bullet} - \delta_{ij} z_{\bullet i} z_{j\bullet}}{z_{\bullet\bullet}},\end{aligned} \tag{4.11}$$

$$\begin{aligned}\mathbb{E}^{\mathrm{h}}[\eta(i \in \wp^{\mathrm{h}})\eta(j \in \wp^{\mathrm{h}})] &= \sum_{s,t \in \mathcal{V}} \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st})\eta(j \in \wp^{\mathrm{h}}_{st}) \frac{w(\wp^{\mathrm{h}}_{st})}{w(\mathcal{P}^{\mathrm{h}})} \\ &= \frac{\sum_{\substack{t \in \mathcal{V} \\ t \neq i,j}} \left(z^{(-t)}_{\bullet i} z^{(-t)}_{ij} z^{\mathrm{h}}_{jt} + z^{(-t)}_{\bullet j} z^{(-t)}_{ji} z^{\mathrm{h}}_{it} - \delta_{ij} z^{(-t)}_{\bullet i} z^{\mathrm{h}}_{it} \right)}{z^{\mathrm{h}}_{\bullet\bullet}} \\ &\quad + \frac{z^{(-j)}_{\bullet i} z^{\mathrm{h}}_{ij} + z^{(-i)}_{\bullet j} z^{\mathrm{h}}_{ji} + \delta_{ij} z^{\mathrm{h}}_{\bullet i}}{z^{\mathrm{h}}_{\bullet\bullet}},\end{aligned} \tag{4.12}$$

which, again, allows us to compute the covariance and the correlation, but now for the number of co-occurrences of nodes on paths, with

$$\operatorname{cov}(\eta(i \in \wp), \eta(j \in \wp)) = \mathbb{E}[\eta(i \in \wp)\eta(j \in \wp)] - \mathbb{E}[\eta(i \in \wp))\mathbb{E}(\eta(j \in \wp)], \tag{4.13}$$

$$\operatorname{cor}(\eta(i \in \wp), \eta(j \in \wp)) = \frac{\operatorname{cov}(\eta(i \in \wp), \eta(j \in \wp))}{\sqrt{\operatorname{cov}(\eta(i \in \wp), \eta(i \in \wp)) \operatorname{cov}(\eta(j \in \wp), \eta(j \in \wp))}}. \tag{4.14}$$

The expressions considering hitting-paths probabilities are similar. It is well known that such covariance and correlation matrices are positive semi-definite (Gram matrices [62]) and are therefore valid kernels on a graph representing similarities between nodes (see [77–79] for kernels in general and [5, 48] for kernels on a graph).

Note that for all these formulae, computing the results for hitting paths requires an iteration over nodes t, which is more computationally intensive than for the non-hitting case. Moreover, the computation time complexity of all derived kernels is in $O(n^3)$ where n is the number of nodes and when using basic matrix operations (without elaborated data structures nor parallel algorithms). Indeed, it is dominated by the computation of the fundamental matrix $\mathbf{Z}$ which requires a $n \times n$ matrix inversion. Once this matrix is computed, all the supplementary steps for computing the kernel matrices are standard matrix operations that do not exceed $O(n^3)$.

4.3 *A distance measure between nodes*

When the elements of the matrix $\mathbf{W}$ represent local affinities between linked nodes, a useful distance measure—that was called the free energy potential distance in the standard bag-of-paths framework

[14, 15]—can be computed from previous result (R.6). The directed distance is defined as

$$\phi(i,j) = -\log\left(z^{\mathrm{h}}_{st}\right). \tag{4.15}$$

It was shown in [14] that this quantity can be interpreted as minus log the probability of surviving during a killed random walk from i to j, for a random walker moving according to the sub-stochastic transition matrix $\mathbf{W}$. Notice that a closely related distance measure between nodes, called the walk distance, has been proposed in [41], by applying a transformation on (4.15).

Then, the *bag-of-paths*, free energy potential, distance measure can immediately be deduced from the previous quantity,

$$\Delta^{\mathrm{BOP}}_{ij} \triangleq \begin{cases} \frac{1}{2}\left(\phi(i,j) + \phi(j,i)\right) & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases} \tag{4.16}$$

The quantity is non-negative because it represents minus log of probabilities. Moreover, it can easily be shown that the triangle inequality is satisfied. Indeed, from result (R.3), $z^{\mathrm{h}}_{st} = z_{st}/z_{tt} \geq z^{(+i)}_{st}/z_{tt}$ (because $\mathcal{P}^{(+i)}_{st} \subseteq \mathcal{P}_{st}$). Then, from (R.3) and (R.2), the quantity on the right side of the inequality becomes $z^{(+i)}_{st}/z_{tt} = z^{\mathrm{h}}_{si} z_{it}/z_{tt} = z^{\mathrm{h}}_{si} z^{\mathrm{h}}_{it}$. Therefore, $z^{\mathrm{h}}_{st} \geq z^{\mathrm{h}}_{si} z^{\mathrm{h}}_{it}$ and taking $-\log$ of this inequality proves the triangle inequality for the directed distance, and thus also for the distance. In the standard bag-of-paths formalism, this distance provided competitive results in semi-supervised and clustering tasks [14, 80, 81].

5. Case study: application to semi-supervised classification

For illustration, the accuracy of the introduced methods will be compared on semi-supervised classification tasks. However, we have to emphasize that our goal here is not to propose new graph-based semi-supervised classification algorithms outperforming state-of-the-art techniques. Its aim is more to investigate if the introduced similarity measures are able to capture the community structure of networks in an accurate way. In our case, the main baseline method will be the free energy potential distance defined in the bag-of-paths framework, which performed best in a number of pattern recognition tasks [14, 80]. Thus, the experiments will tell us if the introduced similarity measures are competitive with respect to this free energy distance. But before going into the details of the experiments, let us first introduce the different similarity measures that are derived from the studied models.

Following the discussion in the introduction (Section 1.2), our introduced measures can also be interpreted as inner products in the (usually infinite-dimensional) vector space of paths in which each node $x \in \mathcal{V}$ has a coordinate[2] $\phi_i(x) = \phi(x, \wp_i)$, for paths $\wp_i \in \mathcal{P}$ which have been numbered as in Section 1.2, $(\wp_1, \wp_2, \dots)$ by decreasing weight, so that $\mathcal{P}$ becomes a totally ordered set. For two arbitrary nodes $x, y \in \mathcal{V}$, $\langle x|y\rangle = \sum_{i=1}^{\infty} \phi_i(x)\mathrm{P}(\wp_i)\phi_i(y)$ where $\phi_i(x)$ is a function of node x and path $\wp_i$; in our case, the presence of node x on path $\wp_i \in \mathcal{P}$ ($\phi_i(x) = \delta(x \in \wp_i)$) or the number of occurrences of node x on $\wp_i$ ($\phi_i(x) = \eta(x \in \wp_i)$).

Using this property, we can define eight different *kernel matrices* (see Table 3). All of these similarity measures will be used in a semi-supervised classification task and compared to eight other methods on various datasets, in order to investigate if these new kernel matrices are able to accurately capture

[2] Not to be confounded with the free energy $\phi(i,j)$.

TABLE 3 *The different similarity matrices compared in this experimental section. The first eight methods are the baselines and the remaining eight are the association (covariance and correlation) measures introduced in this paper.*

Acronym	Method and similarity matrix
BoPP	Bag-of-paths free energy potential distance
RSP	Randomized shortest path dissimilarity
SP	Shortest path distance
LF	Logarithmic forest distance
SoS	Sum-of-similarities method
Q	Modularity matrix
LogCom	Logarithmic communicability kernel
Katz	Neumann kernel
Cov	Covariance matrix (presence on regular paths)
Cor	Correlation matrix (presence on regular paths)
CovH	Covariance matrix (presence on hitting paths)
CorH	Correlation matrix (presence on hitting paths)
NCov	Covariance matrix (number of occurrences on regular paths)
NCor	Correlation matrix (number of occurrences on regular paths)
NCovH	Covariance matrix (number of occurrences on hitting paths)
NCorH	Correlation matrix (number of occurrences on hitting paths)

meaningful information from the graph structure. The kernels and the semi-supervised classification technique are described in detail in the following subsections.

5.1 *Investigated kernels and methods*

As most of the datasets used in this section are defined from their adjacency matrix $\mathbf{A} = (a_{ij})$, thus without costs assigned to edges, we define the cost matrix simply as $\mathbf{C} = (c_{ij}) = (1/a_{ij})$ [14], therefore interpreting the elements of the adjacency matrix as conductances and costs as resistances. As in the standard randomized shortest paths and bag-of-paths formalism [14–18], the weighted adjacency matrix is defined as (see Equation (2.13))

$$\mathbf{W} = \mathbf{P}_{\text{ref}} \circ \exp[-\beta \mathbf{C}], \text{ with } \mathbf{P}_{\text{ref}} = \mathbf{Diag}(\mathbf{Ae})^{-1}\mathbf{A},$$

where $\mathbf{P}_{\text{ref}}$ is the transition probabilities matrix of the natural random walk on the graph, $\mathbf{Diag}(\mathbf{v})$ is a diagonal matrix with vector $\mathbf{v}$ on its diagonal, $\mathbf{e}$ is a column vector full of 1's, $\circ$ is the element-wise product, and $\beta > 0$ is a hyperparameter (the inverse temperature) which will be tuned by internal cross-validation.

5.1.1 *Kernel matrices and their computation* The eight different kernel matrices (see Table 3 for the definition of the acronyms) that will be compared to baseline methods are defined as

$$\mathbf{K}_{\text{cov}} = (k_{ij}^{\text{cov}}) \qquad \triangleq (\text{cov}(\delta(i \in \wp), \delta(j \in \wp))), \tag{5.1}$$

$$\mathbf{K}_{\text{cor}} = (k_{ij}^{\text{cor}}) \triangleq (\text{cor}(\delta(i \in \wp), \delta(j \in \wp))), \tag{5.2}$$

$$\mathbf{K}_{\text{covH}} = (k_{ij}^{\text{covH}}) \triangleq \left(\text{cov}(\delta(i \in \wp^{\text{h}}), \delta(j \in \wp^{\text{h}}))\right), \tag{5.3}$$

$$\mathbf{K}_{\text{corH}} = (k_{ij}^{\text{corH}}) \triangleq \left(\text{cor}(\delta(i \in \wp^{\text{h}}), \delta(j \in \wp^{\text{h}})),\right) \tag{5.4}$$

$$\mathbf{K}_{\text{Ncov}} = (k_{ij}^{\text{Ncov}}) \triangleq (\text{cov}(\eta(i \in \wp), \eta(j \in \wp))), \tag{5.5}$$

$$\mathbf{K}_{\text{Ncor}} = (k_{ij}^{\text{Ncor}}) \triangleq (\text{cor}(\eta(i \in \wp), \eta(j \in \wp))), \tag{5.6}$$

$$\mathbf{K}_{\text{NcovH}} = (k_{ij}^{\text{NcovH}}) \triangleq \left(\text{cov}(\eta(i \in \wp^{\text{h}}), \eta(j \in \wp^{\text{h}}))\right), \tag{5.7}$$

$$\mathbf{K}_{\text{NcorH}} = (k_{ij}^{\text{NcorH}}) \triangleq \left(\text{cor}(\eta(i \in \wp^{\text{h}}), \eta(j \in \wp^{\text{h}}))\right). \tag{5.8}$$

The procedure for computing these kernels follows four different steps,

(1) First, calculate the fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}$ and choose the kernel that must be computed among (5.1)–(5.8).

(2) Then, depending on the kernel which is investigated, compute the needed auxiliary z quantities from Results (R.2)–(R.12) or (R.15)–(R.18).

(3) Depending on the kernel, compute the needed statistical moments and co-moments from Equations (4.3)–(4.6) or (4.9)–(4.12).

(4) Finally, compute the kernel from Equations (4.7) and (4.8) or (4.13) and (4.14).

5.1.2 *Other investigated methods* In this context, the introduced covariance and correlation kernels (5.1–5.8) will be compared to eight known methods, most of which were already investigated in previous semi-supervised and clustering tasks [14, 80–87]. Notice that the compared methods, the experimental setting, and the investigated datasets are very similar to [88], so that our description closely follows this work.

- The *bag-of-paths potential distance* [14], also known as the *free energy distance* [15], and the *randomized shortest path distance* [16, 67], both defined in the bag-of-paths framework. These methods are respectively denoted here by **BoPP** and **RSP**, and have one hyperparameter β. This hyperparameter corresponds to the inverse temperature parameter T presented in Section 2.3.1. It allows to define distances which interpolate between the shortest paths distance when $T \to 0^+$ and the commute time distance (up to a scaling factor) when T is high. In [14], it was found that the BoPP distance provided the best results overall (using almost the same methodology and datasets as those investigated in this paper). It is therefore a challenging competitor for the eight covariance and correlation kernels defined above.
- The *shortest path distance* between two nodes i and j corresponds to the total cost along the least cost paths, derived from the cost matrix $\mathbf{C}$. This method is denoted by **SP** and has no hyperparameter.
- The *logarithmic forest distance* introduced in [40], is based on the matrix forest theorem [21]. This method is denoted by **LF** and has one positive hyperparameter α. It allows to interpolate between the shortest paths distance when $\alpha \to 0^+$ and the commute time distance (up to a scaling factor) when α is high. Furthermore, for every $\alpha > 0$, it ensures the graph-geodetic property [5].

- The *Neumann kernel* [89], initially proposed in [59] by Katz as a method of computing degrees of influence, is defined as $\mathbf{K} = (\mathbf{I} - \alpha\mathbf{A})^{-1} - \mathbf{I}$ for an undirected graph. This method is denoted by **Katz** and has one hyperparameter α which has to be chosen positive and smaller than the inverse of the spectral radius of $\mathbf{A}$, $\rho(\mathbf{A})$, to guarantee convergence. This hyperparameter is a discounting factor which defines the attenuation in a link, where $\alpha = 0$ corresponds to a complete attenuation and $\alpha = 1$ to an absence of any attenuation [5].
- The *logarithmic communicability kernel* [82] is the logarithmic version of the exponential diffusion kernel [90], closely related to the communicability measure [91], and defined as $\mathbf{K} = \ln(\text{expm}(\alpha\mathbf{A}))$ where expm is the matrix exponential. This method is denoted by **LogCom** and has one positive hyperparameter α. This hyperparameter is also a discounting factor favouring the short-length paths between two nodes by giving them a larger weight [5].
- The *modularity matrix* $\mathbf{Q}$ [9, 92] defined as $\mathbf{Q} = \mathbf{A} - \frac{\mathbf{d}\mathbf{d}^{\mathrm{T}}}{\text{vol}}$ where $\mathbf{d}$ contains the node degrees (we assume an undirected graph) and vol is the volume of the graph. This method is denoted by $\mathbf{Q}$ and has no hyperparameter.
- The *sum-of-similarities* method [48, 93], which is based on the *regularized commute time kernel*. This method differs from the previous ones, as it uses a simple label propagation technique in order to classify nodes. It is both computationally efficient and competitive in terms of accuracy [14, 48, 93]. This method is denoted by **SoS** and has one hyperparameter α, which is also a discounting factor.

5.1.3 *Graph node features extraction and classification* The methodology for comparing the different measures on semi-supervised tasks closely follows[3] [14, 88] and was initially inspired by [83–87]. Therefore, the procedure is summarized; for a more complete description of the methods, see reference [14], Section 7.

The classification method consists in extracting five graph node feature vectors[4] from the investigated kernel and similarity/dissimilarity matrices by classical multidimensional scaling in order to use them as input to a linear support vector machine (SVM[5]) for classification. Thus, information is extracted from the structure of the graph in an unsupervised way. The main objective is to evaluate to which extent the different similarity measures are able to capture the information present in the graph structure in only a few dimensions (the 5 dominant eigenvectors of the similarity matrix). In other words, the question we are asking is: are the different measures able to perform efficient (for supervised classification purposes) feature extraction from the graph?

More precisely, first, following classical multidimensional scaling [95, 96], each dissimilarity matrix $\mathbf{D}$ is transformed into a kernel matrix by centring, $\mathbf{K} = -\frac{1}{2}\mathbf{H}\mathbf{D}^{(2)}\mathbf{H}$ where $\mathbf{H} = \mathbf{I} - \mathbf{e}\mathbf{e}^{\mathrm{T}}/n$ is the centring matrix, $\mathbf{e}$ is a column vector full of 1's and $\mathbf{D}^{(2)}$ is the matrix of (elementwise) squared distances. Then, the spectral decomposition of each kernel and similarity matrix is computed and the p dominant eigenvectors $\mathbf{u}_k$ (column vectors) are sorted by decreasing eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Eigenvectors corresponding to negative eigenvalues are then eliminated. At this stage, we tested two different options for extracting the node features.

[3] We thank the authors of this paper for providing the code and the datasets.
[4] We arbitrarily extract five dimensions but performed the same experiment with some more dimensions with similar results.
[5] We use the LIBLINEAR library [94] with the options '-s 1' and '-B 1'.

- Either the dominant eigenvectors $\mathbf{u}_k$ are weighted by the square root of their corresponding eigenvalues, $\sqrt{\lambda_k}$, and concatenated in order to build the data matrix $\mathbf{X}$ containing node features on its rows, $\mathbf{X} = [\sqrt{\lambda_1}\mathbf{u}_1, \sqrt{\lambda_2}\mathbf{u}_2, \dots]$, to be injected as input to a SVM. This is exactly multidimentional scaling limited to p dimensions.
- Or the eigenvectors are simply concatenated $\mathbf{X} = [\mathbf{u}_1, \mathbf{u}_2, \dots]$ and then each row of $\mathbf{X}$ is normalized, $\mathbf{x}_i \leftarrow \mathbf{x}_i / \|\mathbf{x}_i\|_2$, so that each node feature vector is of unit length. This corresponds to the projection of the feature vector on the sphere of radius 1 centred at the origin. This way of defining the node feature vectors removes the effect of the size of the vector, which works better when the angle between the node vectors is more relevant than their distance, like, for instance, for covariance measures.

For simplicity, we only report the results of the *best option* for each method, according to the Nemenyi test described later.

This classification setting is inspired by the work of Zhang et al. [86, 87] as well as Tang et al. [83–85] who compute the dominant eigenvectors (a 'latent space') of graph kernels or similarity matrices and then input them into a supervised classification method, such as a logistic regression or a SVM, to categorize the unlabelled nodes. Notice that these techniques based on similarities and eigenvectors extraction sometimes allow to scale to large graphs, depending on the kernel [97].

5.2 *Experimental settings*

As already mentioned, the experimental methodology is inspired by [14] and briefly summarized here; see this reference for further details.

5.2.1 *Datasets* The different classification methods will be compared on 14 well-known network datasets, already used in previous experimental comparisons. Note that all graphs are undirected.

- **WebKB (four datasets)**. These datasets [98] come from networks of co-citation between webpages of computer science departments of four different universities: **webKB-texas**, **webKB-washington**, **webKB-wisconsin** and **webKB-cornell**. Each page of these websites has been labelled manually to form six different classes: course, department, faculty, project, staff and student.
- **20 Newsgroups (nine subsets)**. The *Newsgroup* dataset consists of 20000 documents taken from 20 discussion groups of the Usenet diffusion list [99]. Nine subsets have been extracted from this data (for details, see [37, 100]): **news-2cl-1**, **news-2cl-2**, **news-2cl-3**, **news-3cl-1**, **news-3cl-2**, **news-3cl-3**, **news-5cl-1**, **news-5cl-2** and **news-5cl-3**. As their names suggest, there are different numbers of classes/topics in these datasets (e.g. **news-3cl-1** contains documents from three different topics considered as classes). For each of these datasets, a sparse graph structure was derived from the *term-document* matrix $\mathbf{T} \triangleq (t_{ij})$, with t_{ij} being the *tf-idf* score [101] of term i in document j. The corresponding adjacency matrix is then obtained by $\mathbf{A} = \mathbf{T}^\top\mathbf{T}$.
- **IMDB**. This dataset comes from the well-known *Internet Movie Database* [98]. Nodes represent movies and the adjacency matrix contains the number of production companies two movies have in common. There are only two labels in this dataset: box-office hit or not.

5.2.2 *Comparisons on semi-supervised classification* The different classification methods are compared in terms of classification accuracy on semi-supervised tasks [102, 103] where a subset of nodes of

the graph is kept unlabelled (their label is hidden to the classifier). The classification model then predicts the label of these unlabelled nodes and its prediction is compared to the true label which was hidden.

In order to reduce variance in accuracy, methods are tested by using a standard 5×5 nested cross-validation methodology. Each external cross-validation contains five folds, and methods are tested with a labelling rate of 20% (80% of the labels are hidden and then predicted by the classifier trained on the labelled 20%). Indeed, the most interesting setting in practical applications is the one where there are few labelled nodes compared to unlabelled. To tune hyperparameters, an internal 5-fold cross-validation on the training fold is performed, by taking $4/5$ of the training fold as labelled and $1/5$ as unlabelled. The average accuracy obtained from the external cross-validation folds is then computed for each method. The whole cross-validation procedure is repeated five times for different random permutations of the data, inducing different sets of labelled/unlabelled nodes. The *final accuracy* of the classifier on the investigated dataset is then obtained by averaging the results over the five repetitions, and is reported in Table 4.

Concerning the hyperparameters, the bag-of-paths-type methods and logarithmic forest distance (LF) are tuned among values of $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$, the Katz kernel and the sum-of-similarities (SoS) method $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and the logarithmic communicability kernel (LogCom) $\{0.01, 0.1, 0.5, 1, 5, 10\}$. For the SVM, the margin hyperparameter C is tuned on the set of values $\{10^{-2}, 10^{-1}, 1, 10, 100\}$. External and internal cross-validation folds are kept identical for all methods inside each repetition.

5.3 *Results and discussion*

As already mentioned, Table 4 reports average classification accuracy in percent for all the investigated methods on the different datasets, and labelling rates of 20%. The method performing best is highlighted in boldface for each dataset. Moreover, a simple Borda ranking of the techniques is performed and shown in Table 5. This ranking provides a score to each method equal to the sum of its ranks over all the datasets, where the methods are sorted in ascending order of classification accuracy. Thus, the higher the rank, the higher the accuracy on the dataset. The best method overall in this context is the one showing the highest Borda score.

As can be seen in Table 5, according to this ranking, the correlation matrix based on the presence of nodes on regular paths (Cor), the covariance matrix based on the presence of nodes on hitting paths (CovH) and the bag-of-paths free energy potential distance (BoPP) obtain the best results overall. Just below, the randomized shortest path dissimilarity (RSP), the logarithmic communicability kernel (LogCom) and the other covariance and correlation measures, except the covariance matrix based on the number of occurrences on regular paths (NCov), provide good results slightly superior to those obtained by the logarithmic forest distance (LF) baseline. We can also observe that the sum-of-similarities (SoS), the Neumann kernel (Katz), the modularity matrix (Q) and the shortest path distance (SP) obtain much worse results in comparison with the other methods.

However, when examining the raw results of Table 4, it can be seen that the best method is dataset-dependent and that no obvious pattern is present. Moreover, the differences between the different methods are often small. This is probably because the top methods have been selected because they already performed well in other comparisons. We observe on Table 4 that the BoPP is the best performing method on three datasets and CovH, NCovH on two datasets. Seven other methods are best only once. In practice, one might therefore think about using a model selection method by, for instance, keeping a validation set in order to select the best-performing model. This could help selecting the most appropriate technique on each dataset, provided that the number of training data is sufficient.

TABLE 4 *Classification accuracy in percent for the various classification methods obtained on the different datasets. For each dataset and method, the final accuracy is obtained by averaging over five repetitions of a standard cross-validation procedure. Each repetition consists of a nested cross-validation with five external folds (test sets, for validation) on which the accuracy of the classifier is averaged, and five internal folds (for parameter tuning). The best performing method is highlighted in boldface for each dataset.*

Classif. method: **Dataset**:	BoPP	RSP	SP	LF	SoS	Q	LogCom	Katz
webKB-texas	73.35	75.82	64.75	75.28	73.26	73.01	75.45	64.98
webKB-washington	68.15	70.35	56.47	**71.19**	61.72	62.52	69.92	67.25
webKB-wisconsin	74.33	74.02	63.99	74.45	73.88	73.42	74.99	73.46
webKB-cornell	57.20	57.31	48.76	58.24	57.43	50.71	**59.16**	52.88
imdb	75.31	76.68	75.22	76.41	**78.12**	74.37	76.28	73.75
news-2cl-1	96.79	96.14	93.46	96.83	91.09	95.85	96.18	95.84
news-2cl-2	91.21	90.16	90.15	89.62	87.70	91.22	91.18	91.49
news-2cl-3	95.98	95.66	95.90	95.14	94.20	95.78	95.43	95.50
news-3cl-1	92.53	93.08	93.18	93.00	88.93	93.02	94.00	93.56
news-3cl-2	93.45	92.56	89.53	91.33	88.01	92.63	92.17	93.30
news-3cl-3	**93.66**	93.06	91.71	91.19	88.29	91.20	91.49	90.37
news-5cl-1	**88.70**	87.81	86.34	86.63	84.84	77.04	86.30	79.49
news-5cl-2	**82.32**	81.81	78.42	80.49	78.80	75.97	79.04	69.25
news-5cl-3	80.27	**80.30**	73.11	79.45	79.22	76.51	78.65	68.65
Classif. method: **Dataset**:	Cov	Cor	CovH	CorH	NCov	NCor	NCovH	NCorH
webKB-texas	75.52	76.57	75.76	76.48	75.63	**76.90**	75.54	76.65
webKB-washington	69.47	67.21	66.59	67.30	68.09	66.44	68.57	66.62
webKB-wisconsin	74.25	74.37	**75.56**	74.07	73.21	74.14	73.13	73.85
webKB-cornell	55.75	58.86	58.83	58.13	52.90	58.02	53.32	58.97
imdb	76.04	77.00	77.57	76.57	76.42	76.67	76.39	76.67
news-2cl-1	96.61	96.83	96.73	96.53	96.05	96.80	96.05	**97.26**
news-2cl-2	91.42	91.07	**91.86**	90.84	91.66	90.67	91.83	91.17
news-2cl-3	96.49	96.45	96.45	96.30	96.53	96.42	**96.75**	96.55
news-3cl-1	93.72	93.76	93.74	93.68	93.43	93.80	**94.06**	93.60
news-3cl-2	93.72	93.11	93.39	93.04	**93.79**	92.98	93.78	92.87
news-3cl-3	91.69	91.51	91.45	91.81	91.90	91.76	91.85	91.92
news-5cl-1	83.28	86.38	83.43	86.45	82.91	86.33	82.48	86.18
news-5cl-2	75.03	77.07	75.06	77.56	73.18	77.25	75.02	77.50
news-5cl-3	78.39	76.56	78.35	76.56	76.86	76.26	76.85	76.22

In order to rate globally the results of each method, a non-parametric Friedman–Nemenyi statistical test [104] allowing to make comparisons across all the 14 datasets is investigated. At first, we run a Friedman test [105, 106] on our results. This tells us whether at least one classification method is

TABLE 5 *Overall position of the different classification techniques (see Table 3) according to Borda's method performed across all datasets (the higher the score, the better).*

Method	Position	Score
Cor	1	152
CovH	2	149
BoPP	3	147
NCorH	4	144
RSP	5	143
LogCom	6	137
CorH	7	136
NCor	7	136
Cov	8	131
NCovH	8	131
LF	9	128
NCov	10	121
SoS	11	76
Katz	12	64
Q	13	61
SP	14	60

significantly different from the others. We obtain a p-value of 5×10^{-6}. This p-value is lower than the threshold α of 0.05 and we can reject the null hypothesis, at least at this level.

Then, we perform a multiple comparison with the Nemenyi test [104, 107]. The results of this test are illustrated in Figure 1 and are quite similar to those provided by the Borda ranking. The figure confirms that the BoPP, the Cor and the CovH all provide good results, which are significantly superior to the results obtained by the shortest path distance SP. We can also observe that the Cor and CovH perform better than the modularity matrix Q. Furthermore, the Cor also outperforms the Neumann kernel (Katz). With regards to the other covariance and correlation measures, we cannot say that they are significantly different from the other baselines. This is partly because the Friedman–Nemenyi test is rather conservative, especially when comparing many different models.

Therefore, to obtain more precise information concerning the relative performance of the methods, we decided to also perform a Wilcoxon signed-ranks test [104, 108] for matched data with a threshold α of 0.05. The results of these tests, comparing the results of the methods on the 14 different datasets, are presented in Table 6. They show that all the introduced covariance and correlation methods, the BoPP, the RSP, the LF and the LogCom are significantly better than the modularity matrix Q, the shortest path distance SP and the Neumann kernel (Katz). The tests also show that all these methods except the Cov, the NCov and the NCovH obtain results significantly better than the SoS.

In summary, the experiments performed on the investigated datasets showed that most of the introduced covariance and correlation measures achieve reasonably good results on a graph features extraction task for semi-supervised classification, in comparison with the eight already known techniques. However, from Table 6, we cannot conclude that the introduced methods perform better or worse than the other baselines (BoPP, RSP, LF and LogCom). Indeed, none of these methods are significantly different after

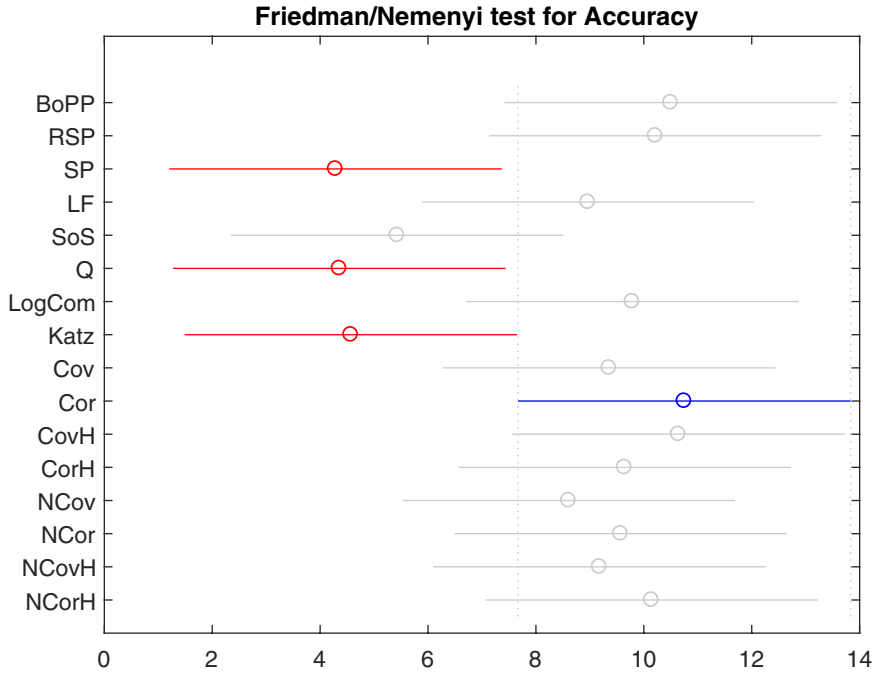

FIG. 1. Mean ranks and 95% Nemenyi confidence intervals for the 16 methods (see Table 3) across the 14 datasets. Two methods are considered as significantly different if their confidence intervals do not overlap. The higher the rank, the best the method. The worse method (SP) and the best method (Cor) overall are highlighted.

performing the pairwise statistical tests. Moreover, only some of these methods (Cor and CovH) are globally better ranked than them. Nevertheless, we can state that they perform reasonably well on the investigated datasets, which is already a good result as some of the baselines performed well in previous experimental comparisons, on both semi-supervised classification and clustering tasks [14, 80–82] .

6. Conclusion

This article derived a series of useful mathematical expressions for computing the expectation of (co-) presence and of the number of (co-) occurrences of nodes on paths sampled from a network. These quantities can then be used for defining similarity measures between nodes as well as betweenness centrality measures. Most of the derived similarity measures are positive semi-definite and can therefore be considered as kernels on a graph and used with kernel-based methods for solving pattern recognition tasks such as node clustering and supervised classification. The main intuition behind these measures is that two nodes are considered as closely related if they often co-occur on the same (preferably short) paths in the network. Experiments on semi-supervised classification tasks have shown that the introduced quantities provide competitive results on the investigated datasets, within a clear theoretical framework.

Note that, as in [5, 36], the different methods could easily be extended in order to provide similarity and centrality measures between groups of nodes, which is left for further work.

Another application that is left for future work is the task of embedding graphs in low-dimensional spaces [55, 109]. Indeed, such a representation can easily be deduced from our measures by applying

TABLE 6 *The p-values provided by a pairwise Wilcoxon signed-rank test for the results presented in Table 4. The p-values inferior to* 0.05 *are highlighted.*

Classif. method: **Classif. method**:	BoPP	RSP	SP	LF	SoS	Q	LogCom	Katz
BoPP	1.0000	0.8077	**0.0006**	0.5830	**0.0031**	**0.0012**	0.7148	**0.0040**
RSP	0.8077	1.0000	**0.0012**	0.1353	**0.0012**	**0.0067**	0.3910	**0.0134**
SP	**0.0006**	**0.0012**	1.0000	**0.0107**	0.3575	0.1937	**0.0017**	0.7148
LF	0.5830	0.1353	**0.0107**	1.0000	**0.0017**	**0.0353**	0.7148	**0.0203**
SoS	**0.0031**	**0.0012**	0.3575	**0.0017**	1.0000	0.9032	**0.0040**	0.6257
Q	**0.0012**	**0.0067**	0.1937	**0.0353**	0.9032	1.0000	**0.0052**	0.8077
LogCom	0.7148	0.3910	**0.0017**	0.7148	**0.0040**	**0.0052**	1.0000	**0.0040**
Katz	**0.0040**	**0.0134**	0.7148	**0.0203**	0.6257	0.8077	**0.0040**	1.0000
Cov	0.6698	0.1937	**0.0166**	0.5830	0.0785	**0.0023**	0.2676	**0.0002**
Cor	0.6257	0.7148	**0.0031**	0.6848	**0.0134**	**0.0004**	0.8077	**0.0023**
CovH	0.8552	0.6698	**0.0166**	0.7148	**0.0203**	**0.0023**	0.9032	**0.0009**
CorH	0.4263	0.5830	**0.0017**	1.0000	**0.0134**	**0.0004**	0.3910	**0.0031**
NCov	0.2166	0.1189	**0.0295**	0.2676	0.2412	**0.0134**	0.1531	**0.0023**
NCor	0.3575	0.6257	**0.0023**	0.9515	**0.0166**	**0.0009**	0.5016	**0.0067**
NCovH	0.2958	0.1353	**0.0203**	0.2676	0.1531	**0.0031**	0.2412	**0.0004**
NCorH	0.5416	0.6257	**0.0017**	0.7609	**0.0134**	**0.0009**	0.6698	**0.0067**
Classif. method: **Classif. method**:	Cov	Cor	CovH	CorH	NCov	NCor	NCovH	NCorH
BoPP	0.6698	0.6257	0.8552	0.4263	0.2166	0.3575	0.2958	0.5416
RSP	0.1937	0.7148	0.6698	0.5830	0.1189	0.6257	0.1353	0.6257
SP	**0.0166**	**0.0031**	**0.0166**	**0.0017**	**0.0295**	**0.0023**	**0.0203**	**0.0017**
LF	0.5830	0.6848	0.7148	1.0000	0.2676	0.9515	0.2676	0.7609
SoS	0.0785	**0.0134**	**0.0203**	**0.0134**	0.2412	**0.0166**	0.1531	**0.0134**
Q	**0.0023**	**0.0004**	**0.0023**	**0.0004**	**0.0134**	**0.0009**	**0.0031**	**0.0009**
LogCom	0.2676	0.8077	0.9032	0.3910	0.1531	0.5016	0.2412	0.6698
Katz	**0.0002**	**0.0023**	**0.0009**	**0.0031**	**0.0023**	**0.0067**	**0.0004**	**0.0067**
Cov	1.0000	0.4631	0.3258	0.8077	0.0676	0.6257	0.2958	0.5016
Cor	0.4631	1.0000	0.7148	0.1726	0.1531	0.1189	0.2676	0.5830
CovH	0.3258	0.7148	1.0000	0.7148	0.0676	0.7148	0.2676	0.9032
CorH	0.8077	0.1726	0.7148	1.0000	0.2412	0.5830	0.5016	0.8077
NCov	0.0676	0.1531	0.0676	0.2412	1.0000	0.2958	0.3396	0.1726
NCor	0.6257	0.1189	0.7148	0.5830	0.2958	1.0000	0.4263	0.4631
NCovH	0.2958	0.2676	0.2676	0.5016	0.3396	0.4263	1.0000	0.3258
NCorH	0.5016	0.5830	0.9032	0.8077	0.1726	0.4631	0.3258	1.0000

classical multidimensional scaling to the defined kernels. Interestingly, the recently introduced DeepWalk technique [51] (an example of representation learning applied to graph data) and variants seem to provide accurate low-dimensional representations of the nodes of the network and its structure. Quoting [51],

'DeepWalk uses local information obtained from truncated random walks to learn latent representations by treating walks as the equivalent of sentences' (in natural language processing). This is similar in spirit to the techniques studied in this article with an important difference: our measures are computed in closed form through matrix operations whereas DeepWalk uses sampled paths and an artificial neural network to build the representation. Therefore, we plan to work on a thorough experimental comparison between representation learning based techniques and our proposed methods, on semi-supervised classification and graph embedding tasks. In this context it would also be interesting to try to directly estimate the elements of the fundamental matrix $\mathbf{Z}$ by sampling paths according to a killed random walk with transition matrix[6] $\mathbf{W}$. This would allow to scale on larger graphs, the main advantage of DeepWalk.

Still another interesting avenue would be to examine the properties of the investigated networks and to try to understand why a method works better on some data sets, based on these properties. Indeed, each considered measure captures a different structural property of the network in a low-dimensional feature vector. In this context, it would also be interesting to combine different extracted feature vectors (coming from different measures) in order to verify if their combination improves the performance.

It would also be interesting to extend the bag-of-paths framework to other probability distributions on paths, beyond the Gibbs–Boltzmann distribution. Changing the probability distribution to, for instance, distributions belonging to the more general exponential family [110], could be attempted.

However, the main drawback of the introduced techniques is the fact that they require the inversion of a $n \times n$ matrix so that they do not scale well on large graphs. Thus, another interesting further work would be the study of co-occurrence measures, defined this time on truncated paths, such as in [93, 111]. This should allow to scale to medium to large sparse graphs.

Acknowledgements

This work was partially supported by the Immediate and the Brufence projects funded by InnovIris (Brussels Region), as well as former projects funded by the Walloon region, Belgium. We thank these institutions for giving us the opportunity to conduct both fundamental and applied research. We also appreciated the remarks of the anonymous reviewers which helped us to improve significantly the manuscript.

References

1. Barabási, A.-L. (2016) *Network Science*. Cambridge, MA, USA:Cambridge University Press.
2. Chiang, M. (2012) *Networked Life*. Cambridge, MA, USA:Cambridge University Press.
3. Chung, F. R. & Lu, L. (2006) *Complex Graphs and Networks*. Providence, RI, USA:American Mathematical Society.
4. Estrada, E. (2012) *The Structure of Complex Networks*. Oxford, UK:Oxford University Press.
5. Fouss, F., Saerens, M. & Shimbo, M. (2016) *Algorithms and Models for Network Data and Link Analysis*. Cambridge, MA, USA:Cambridge University Press.
6. Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data: Methods and Models*, Springer Series in Statistics. New York City, NY, USA:Springer.
7. Lewis, T. G. (2009) *Network Science*. Hoboken, NJ, USA:Wiley.
8. Mihalcea, R. & Radev, D. (2011) *Graph-based Natural Language Processing and Information Retrieval*. Cambridge, MA, USA:Cambridge University Press.
9. Newman, M. (2018) *Networks*, 2nd edn. Oxford, UK:Oxford University Press.

[6] Indeed, it was shown that the fundamental matrix corresponds to the probability of surviving during a killed random walk [14].

10. Silva, T. & Zhao, L. (2016) *Machine Learning in Complex Networks*. New York City, NY, USA:Springer.

11. Thelwall, M. (2004) *Link Analysis: An Information Science Approach*. Amsterdam, NL:Elsevier.

12. Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge, MA, USA:Cambridge University Press.

13. Bavaud, F. & Guex, G. (2012) Interpolating between random walks and shortest paths: a path functional approach. *International Conference on Social Informatics* (Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J. & Guéret, G. eds). New York City, NY, USA:Springer, pp. 68–81.

14. Françoisse, K., Kivimäki, I., Mantrach, A., Rossi, F. & Saerens, M. (2017) A bag-of-paths framework for network data analysis. *Neural Netw.*, **90**, 90–111.

15. Kivimäki, I., Shimbo, M. & Saerens, M. (2014) Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A*, **393**, 600–616.

16. Saerens, M., Achbany, Y., Fouss, F. & Yen, L. (2009) Randomized shortest-path problems: two related models. *Neural Comput.*, **21**, 2363–2404.

17. Yen, L., Mantrach, A., Shimbo, M. & Saerens, M. (2008) A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. *Proceedings of the 14th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*. New York City, NY, USA: Association for Computing Machinery, pp. 785–793.

18. Mantrach, A., Yen, L., Callut, J., Francoise, K., Shimbo, M. & Saerens, M. (2010) The sum-over-paths covariance kernel: a novel covariance between nodes of a directed graph. *IEEE Trans. Patt. Anal. Mach. Intell.*, **32**, 1112–1126.

19. Akamatsu, T. (1996) Cyclic flows, Markov process and stochastic traffic assignment. *Transport. Res. B*, **30**, 369–386.

20. Dial, R. (1971) A probabilistic multipath assignment model that obviates path enumeration. *Transport. Res.*, **5**, 83–111.

21. Chebotarev, P. & Shamis, E. (1997) The matrix-forest theorem and measuring relations in small social groups. *Autom. Remote Control*, **58**, 1505–1514.

22. Chebotarev, P. & Shamis, E. (1998) On proximity measures for graph vertices. *Autom. Remote Control*, **59**, 1443–1459.

23. Lebichot, B. & Saerens, M. (2018) A bag-of-paths node criticality measure. *Neurocomputing*, **275**, 224–236.

24. de Oliveira Werneck, R., Raveaux, R., Tabbone, S. & da Silva Torres, R. (2019) Learning cost function for graph classification with open-set methods. *Patt. Recogn. Lett.*, **128**, 8–15.

25. Jouili, S. & Tabbone, S. (2009) Graph matching based on node signatures. *Proceedings of the 7th International Workshop on Graph-Based Representations in Pattern Recognition (IAPR-TC-15)* (Torsello, A., Escolano Ruiz, F. & Brun, L. eds). New York City, NY, USA:Springer, pp. 154–163.

26. Leicht, E. A., Holme, P. & Newman, M. E. J. (2006) Vertex similarity in networks. *Phys. Rev. E*, **73**, 026120.

27. Yang, Y., Pei, J. & Al-Barakati, A. (2017) Measuring in-network node similarity based on neighborhoods: a unified parametric approach. *Knowl. Inf. Syst.*, **53**, 43–70.

28. Nadler, B., Lafon, S., Coifman, R. & Kevrekidis, I. (2006) Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. *Adv. Neural Inform. Process. Syst.*, **18**, 955–962.

29. Yen, L., Saerens, M. & Fouss, F. (2011) A link analysis extension of correspondence analysis for mining relational databases. *IEEE Trans. Knowl. Data Eng.*, **23**, 481–495.

30. Jin, R., Lee, V. E. & Hong, H. (2011) Axiomatic ranking of network role similarity. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. New York City, NY, USA: Association for Computing Machinery, pp. 922–930.

31. Klein, D. J. & Randić, M. (1993) Resistance distance. *J. Math. Chem.*, **12**, 81–95.

32. Chandra, A. K., Raghavan, P., Ruzzo, W. L., Smolensky, R. & Tiwari, P. (1989) The electrical resistance of a graph captures its commute and cover times. *Annual ACM Symposium on Theory of Computing*. New York City, NY, USA: Association for Computing Machinery, pp. 574–586.

33. Fouss, F., Pirotte, A., Renders, J.-M. & Saerens, M. (2007) Random-walk computation of similarities between nodes of a graph, with application to collaborative recommendation. *IEEE Trans. Knowl. Data Eng.*, **19**, 355–369.

34. von Luxburg, U., Radl, A. & Hein, M. (2010) Getting lost in space: large sample analysis of the commute distance. *Advances in Neural Information Processing Systems 23: Proceedings of the Neural Information Processing Systems conference (NIPS 2010)*. pp. 2622–2630.

35. von Luxburg, U., Radl, A. & Hein, M. (2014) Hitting and commute times in large random neighborhood graphs. *J. Mach. Learn. Res.*, **15**, 1751–1798.

36. Lebichot, B., Kivimaki, I., Francoisse, K. & Saerens, M. (2014) Semi-supervised classification through the bag-of-paths group betweenness. *IEEE Trans. Neural Netw. Learn. Syst.*, **25**, 1173–1186.

37. Yen, L., Fouss, F., Decaestecker, C., Francq, P. & Saerens, M. (2007) Graph nodes clustering based on the commute-time kernel. *Proceedings of the 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007). Lecture notes in Computer Science, LNCS* (Zhou, Z.-H., Li, H. & Yang, Q. eds). New York City, NY, USA: Springer, pp. 1037–1045.

38. Guex, G. (2016) Interpolating between random walks and optimal transportation routes: flow with multiple sources and targets. *Physica A*, **450**, 264–277.

39. Guex, G., Kivimaki, I. & Saerens, M. (2019) Randomized optimal transport on a graph: framework and new distance measures. *Netw. Sci.*, **7**, 88–122.

40. Chebotarev, P. (2011) A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. *Discrete Appl. Math.*, **159**, 295–302.

41. Chebotarev, P. (2012) The walk distances in graphs. *Discrete Appl. Math.*, **160**, 1484–1500.

42. Chebotarev, P. (2013) Studying new classes of graph metrics. *Proceedings of the 1st International Conference on Geometric Science of Information (GSI '13)* (F. Nielsen & F. Barbaresco eds), vol. 8085 of Lecture Notes in Computer Science. New York City, NY, USA:Springer, pp. 207–214.

43. Alamgir, M. & von Luxburg, U. (2011) Phase transition in the family of p-resistances. *Advances in Neural Information Processing Systems 24: Proceedings of the NIPS 2011 conference*. (Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. & Weinberger, K. Q. eds). Cambridge, MA, USA:MIT Press, pp. 379–387.

44. Herbster, M. & Lever, G. (2009) Predicting the labelling of a graph via minimum p-seminorm interpolation. *Proceedings of the 22nd Annual Conference on Learning Theory (COLT2009)*.

45. Li, Y., Zhang, Z.-L. & Boley, D. (2011) The routing continuum from shortest-path to all-path: A unifying theory. *Proceedings of the 31st International Conference on Distributed Computing Systems (ICDCS '11)*. IEEE Computer Society, pp. 847–856.

46. Li, Y., Zhang, Z.-L. & Boley, D. (2013) From shortest-path to all-path: The routing continuum theory and its applications. *IEEE Trans. Parallel Distrib. Syst.*, **25**, 1745–1755.

47. Guex, G. & Bavaud, F. (2015) Flow-based dissimilarities: shortest path, commute time, max-flow and free energy. *Data Science, Learning by Latent Structures, and Knowledge Discovery* (B. Lausen, S. Krolak-Schwerdt, & M. Bohmer eds), vol. 1564 of Studies in Classification, Data Analysis, and Knowledge Organization. New York City, NY, USA:Springer, pp. 101–111.

48. Fouss, F., Françoisse, K., Yen, L., Pirotte, A. & Saerens, M. (2012) An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Netw.*, **31**, 53–72.

49. Chebotarev, P. & Agaev, R. (2002) Forest matrices around the Laplacian matrix. *Linear Algebra Appl.*, **356**, 253–274.

50. Kolaczyk, E., Chua, D. & Barthelemy, M. (2009) Group betweenness and co-betweenness: inter-related notions of coalition centrality. *Soc. Netw.*, **31**, 190–203.

51. Perozzi, B., Al-Rfou, R. & Skiena, S. (2014) DeepWalk: online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14. New York City, NY, USA: Association for Computing Machinery, pp. 701–710.

52. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013) Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26: Proceedings*

of the NIPS 2013 Conference. (Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q. eds). Cambridge, MA, USA:MIT Press, pp. 3111–3119.
53. Harispe, S., Ranwez, S., Janaqi, S. & Montmain, J. (2015) *Semantic Similarity from Natural Language and Ontology Analysis*. San Rafael, CA, USA:Morgan & Claypool Publishers.
54. Bengio, Y., Courville, A. & Vincent, P. (2013) Representation learning: a review and new perspectives. *IEEE Trans. Patt. Anal. Mach. Intell.*, **35**, 1798–1828.
55. Zhang, D., Yin, J., Zhu, X. & Zhang, C. (2020) Network representation learning: a survey. *IEEE Trans. Big Data*, **6**, 3–28.
56. Devooght, R. (2013) Bag of paths framework for graph mining. *Master's Thesis*, Universite Libre de Bruxelles, Ecole Polytechnique, Belgium, Supervisors: Profs H. Bersini and M. Saerens.
57. Meyer, C. D. (2000) *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA:SIAM.
58. Langville, A. & Meyer, C. (2006) *Google's PageRank and Beyond*. Princeton, NJ, USA:Princeton University Press.
59. Katz, L. (1953) A new status index derived from sociometric analysis. *Psychmetrika*, **18**, 39–43.
60. Bain, L. & Engelhardt, M. (1992) *Introduction to Probability and Mathematical Statistics*, 2nd edn. Belmont, CA, USA:Duxbury.
61. Kaplan, W. (2003) *Advanced Calculus*, 3rd edn. London, UK:Pearson.
62. Olver, P. & Shakiban, C. (2006) *Applied Linear Algebra*. London, UK:Pearson.
63. Freeman, L. (1977) A set of measures of centrality based on betweenness. *Sociometry*, **40**, 35–41.
64. Freeman, L. (1978-1979) Centrality in social networks conceptual clarification. *Soc. Netw.*, **1**, 215 – 239.
65. Brandes, U. & Fleischer, D. (2005) Centrality measures based on current flow. *Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science (STACS)*. (Diekert, V. & Durand, B eds). New York City, NY, USA: Springer, pp. 533–544.
66. Newman, M. (2005) A measure of betweenness centrality based on random walks. *Soc. Netw.*, **27**, 39–54.
67. Kivimäki, I., Lebichot, B., Saramäki, J. & Saerens, M. (2016) Two betweenness centrality measures based on randomized shortest paths. *Sci. Rep.*, **6**, srep19668.
68. Cover, T. M. & Thomas, J. A. (2006) *Elements of Information Theory*, 2nd edn. Hoboken, NJ, USA:Wiley.
69. Kapur, J. N. (1989) *Maximum-Entropy Models in Science and Engineering*. Hoboken, NJ, USA:Wiley.
70. Kapur, J. N. & Kesavan, H. K. (1992) *Entropy Optimization Principles with Applications*. San Diego, CA, USA:Academic Press.
71. Doyle, P. G. & Snell, J. L. (1984) *Random Walks and Electric Networks*. The Mathematical Association of America. Washington D.C., USA.
72. Grinstead, C. & Snell, J. L. (1997) *Introduction to Probability*. The Mathematical Association of America, 2nd edn. Washington D.C., USA.
73. Kemeny, J. G. & Snell, J. L. (1976) *Finite Markov Chains*. New York City, NY, USA:Springer.
74. Norris, J. (1997) *Markov Chains*. Cambridge, MA, USA:Cambridge University Press.
75. Taylor, H. M. & Karlin, S. (1998) *An Introduction to Stochastic Modeling*, 3rd edn. San Diego, CA, USA:Academic Press.
76. Klenke, A. (2014) *Probability Theory, A Comprehensive Course*. New York City, NY, USA:Springer.
77. Gartner, T. (2008) *Kernels for Structured Data*. Singapore:World Scientific.
78. Schölkopf, B. & Smola, A. J. (2001) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA:MIT press.
79. Shawe-Taylor, J. & Cristianini, N. (2004) *Kernel Methods for Pattern analysis*. Cambridge, MA, USA:Cambridge University Press.
80. Sommer, F., Fouss, F. & Saerens, M. (2016) Comparison of graph node distances on clustering tasks. *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2016)*. Lecture Notes in Computer Science, vol. 9886. (Villa, A. E. P., Masulli, P. & Pons Rivero, A. J. eds). New York City, NY, USA:Springer, pp. 192–201.
81. Sommer, F., Fouss, F. & Saerens, M. (2017) Modularity-driven kernel k-means for community detection. *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2017).* Lecture Notes in

Computer Science, vol. 10614. (Lintas, A., Rovetta, S., Verschure, P. F. M. J. & Villa, A. E. P. eds). New York City, NY, USA:Springer, pp. 423–433.

82. Ivashkin, V. & Chebotarev, P. (2016) Models, Algorithms and Technologies for Network Analysis:Do logarithmic proximity measures outperform plain ones in graph clustering? *International Conference on Network Analysis*. (Kalyagin, V. A., Koldanov, P. A. & Pardalos, P. eds). New York City, NY, USA:Springer, pp. 87–105.

83. Tang, L. & Liu, H. (2009) Relational learning via latent social dimensions. *Proceedings of the ACM conference on Knowledge Discovery and Data Mining (KDD 2009)*. New York City, NY, USA: Association for Computing Machinery, pp. 817–826.

84. Tang, L. & Liu, H. (2009) Scalable learning of collective behavior based on sparse social dimensions. *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM 2009)*. New York City, NY, USA: Association for Computing Machinery, pp. 1107–1116.

85. Tang, L. & Liu, H. (2010) Toward predicting collective behavior via social dimension extraction. *IEEE Intell. Syst.*, **25**, 19–25.

86. Zhang, D. & Mao, R. (2008) Classifying networked entities with modularity kernels. *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008)* New York City, NY, USA:ACM, pp. 113–122.

87. Zhang, D. & Mao, R. (2008) A new kernel for classification of networked entities. *Proceedings of 6th International Workshop on Mining and Learning with Graphs*, New York City, NY, USA: Association for Computing Machinery.

88. Courtain, S., Leleux, P., Kivimaki, I., Guex, G. & Saerens, M. (2020) Randomized shortest paths with net flows and capacity constraints. Accepted for publication in Information Sciences.

89. Schölkopf, B. & Smola, A. J. (2002) *Learning with Kernels*. The MIT Press.

90. Kondor, R. I. & Lafferty, J. (2002) Diffusion kernels on graphs and other discrete structures. *Proceedings of the 19th International Conference on Machine Learning*. (Sammut, C. & Hoffmann, A.G. eds). New York City, NY, USA:Association for Computing Machinery pp. 315–322.

91. Estrada, E. & Hatano, N. (2008) Communicability in complex networks. *Phys. Rev. E*, **77**, 036111.

92. Newman, M. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, **103**, 8577–8582.

93. Mantrach, A., Zeebroeck, N. V., Francq, P., Shimbo, M., Bersini, H. & Saerens, M. (2011) Semi-supervised classification and betweenness computation on large, sparse, directed graphs. *Patt. Recogn.*, **44**, 1212–1224.

94. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. & Lin, C.-J. (2008) LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.*, **9**, 1871–1874.

95. Borg, I. & Groenen, P. (1997) *Modern Multidimensional Scaling: Theory and Applications*. New York City, NY, USA:Springer.

96. Cox, T. & Cox, M. (2001) *Multidimensional Scaling,* 2nd edn. Chapman and Hall.

97. Devooght, R., Mantrach, A., Kivimäki, I., Bersini, H., Jaimes, A. & Saerens, M. (2014) Random walks based modularity: application to semi-supervised learning. *Proceedings of the 23rd International World Wide Web Conference (WWW '14)*. New York City, NY, USA: Association for Computing Machinery, pp. 213–224.

98. Macskassy, S. A. & Provost, F. (2007) Classification in networked data: a toolkit and a univariate case study. *J. Mach. Learn. Res.*, **8**, 935–983.

99. Dua, D. & Graff, C. (2019) UCI Machine Learning Repository, http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science.

100. Yen, L., Fouss, F., Decaestecker, C., Francq, P. & Saerens, M. (2009) Graph nodes clustering with the sigmoid commute-time kernel: a comparative study. *Data Knowl. Eng.*, **68**, 338–361.

101. Manning, C. D., Raghavan, P. & Schütze, H. (2008) *Introduction to Information Retrieval*. Cambridge, MA, USA:Cambridge University Press.

102. Chapelle, O., Scholkopf, B. & Zien, A. (2006) *Semi-supervised Learning*. Cambridge, MA, USA:MIT Press.

103. Subramanya, A. & Pratim Talukdar, P. (2014) *Graph-based Semi-supervised Learning*. San Rafael, CA, USA:Morgan & Claypool Publishers.

104. DEMŠAR, J. (2006) Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, **7**, 1–30.
105. FRIEDMAN, M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.*, **32**, 675–701.
106. FRIEDMAN, M. (1990) A comparison of alternative tests of significance for the problem of m rankings. *Ann. Math. Stat.*, **11**, 86–92.
107. NEMENYI, P. (1963) Distribution-free multiple comparisons. *Ph.D. Thesis*, Princeton University, United States.
108. WILCOXON, F. (1945) Individual comparisons by ranking methods. *Biometrics Bull.*, **1**, 80–83.
109. CAI, H., ZHENG, V. & CHANG, K. (2018) A comprehensive survey of graph embedding: problems, techniques, and applications. *IEEE Trans. Knowl. Data Eng.*, **30**, 1616–1637.
110. MCCULLAGH, P. & NELDER, J. (1989) *Generalized Linear Models*. London, UK:Chapman & Hall.
111. SARKAR, P. & MOORE, A. (2007) A tractable approach to finding closest truncated-commute-time neighbors in large graphs. *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI '07)*, pp. 335–343.
112. BRUALDI, R. (2009) *Introductory Combinatorics*, 5th edn. London, UK:Pearson.

Appendix: proofs of the main results

This appendix contains the proofs of the results stated in Section 3 as well as a discussion of Equations (4.5) and (4.6) appearing in Section 4. See Tables 1 and 2 for the notation.

A. Proofs of results of Sections 3.1 and 3.2 (node presence)

A.1 *Results involving one intermediate node*

(R.1).

Let $\mathcal{P}_{st}(\ell)$ be the set of all regular paths $\wp_{st}^{\ell}$ from s to t of length exactly equal to ℓ. We easily obtain that

$$\sum_{\wp_{st}^{\ell}\in\mathcal{P}_{st}(\ell)} w(\wp_{st}^{\ell}) = \sum_{i_1,i_2,\ldots,i_{\ell-1}=1}^{n} w_{si_1}w_{i_1i_2}\cdots w_{i_{\ell-1}t} = [\mathbf{W}^{\ell}]_{st}. \tag{A.1}$$

Therefore, from (2.8),

$$\begin{aligned} w(\mathcal{P}_{st}) &= \sum_{\wp_{st}\in\mathcal{P}_{st}} w(\wp_{st}) = \sum_{\ell=0}^{\infty}\sum_{\wp_{st}^{\ell}\in\mathcal{P}_{st}(\ell)} w(\wp_{st}^{\ell}) \\ &= \sum_{\ell=0}^{\infty}[\mathbf{W}^{\ell}]_{st} = \left[\sum_{\ell=0}^{\infty}\mathbf{W}^{\ell}\right]_{st} = [\mathbf{Z}]_{st} = z_{st}. \end{aligned} \tag{R.1}$$

(R.2).

Observe that there is a bijection between $\mathcal{P}_{st}$ and $\mathcal{P}_{st}^{\mathrm{h}}\circ\mathcal{P}_{tt}$. In other words, any regular path from s to t can be decomposed uniquely into a hitting sub-path from s to t, where node t is reached for the first time, followed by a regular sub-path (a cycle) connecting t to itself. We then have $w(\wp_{st}) = w(\wp_{st}^{\mathrm{h}})w(\wp_{tt})$ for

the weight of corresponding paths. This implies that

$$w(\mathcal{P}^{\mathrm{h}}_{st})w(\mathcal{P}_{tt}) = \Bigg(\sum_{\wp^{\mathrm{h}}_{st}\in\mathcal{P}^{\mathrm{h}}_{st}} w(\wp^{\mathrm{h}}_{st})\Bigg)\Bigg(\sum_{\wp_{tt}\in\mathcal{P}_{tt}} w(\wp_{tt})\Bigg)$$
$$= \sum_{\wp^{\mathrm{h}}_{st}\in\mathcal{P}^{\mathrm{h}}_{st}}\sum_{\wp_{tt}\in\mathcal{P}_{tt}} w(\wp^{\mathrm{h}}_{st})w(\wp_{tt}) = \sum_{\wp_{st}\in\mathcal{P}_{st}} w(\wp_{st}) = w(\mathcal{P}_{st}),$$

and, from (R.1), we get the result (R.2), $z^{\mathrm{h}}_{st} = z_{st}/z_{tt}$. Note that this implies, for $s = t$, $z^{\mathrm{h}}_{tt} = 1$.

(R.3).

Similarly to (R.2), there exists a bijection between $\mathcal{P}^{\mathrm{h}}_{si} \circ \mathcal{P}_{it}$ and $\mathcal{P}^{(+i)}_{st}$. Indeed, any path from s to t visiting i can be decomposed uniquely into a hitting sub-path from s to i, where node i is visited for the first time, followed by a regular sub-path from i to t. Consequently, we have $w(\wp_{st}) = w(\wp^{\mathrm{h}}_{si})w(\wp_{it})$ for corresponding paths. Thus,

$$w\left(\mathcal{P}^{(+i)}_{st}\right) = \sum_{\wp_{st}\in\mathcal{P}_{st}} \delta(i\in\wp_{st})w(\wp_{st}) = \sum_{\wp^{\mathrm{h}}_{si}\in\mathcal{P}^{\mathrm{h}}_{si}}\sum_{\wp_{it}\in\mathcal{P}_{it}} w(\wp^{\mathrm{h}}_{si})w(\wp_{it}) = z^{\mathrm{h}}_{si}z_{it}. \tag{R.3}$$

(R.4).

From (R.3), we observe that

$$w\left(\mathcal{P}^{(-i)}_{st}\right) = \sum_{\wp_{st}\in\mathcal{P}_{st}} \delta(i\notin\wp_{st})w(\wp_{st}) = \sum_{\wp_{st}\in\mathcal{P}_{st}} (1-\delta(i\in\wp_{st}))w(\wp_{st}) = z_{st} - z^{\mathrm{h}}_{si}z_{it}, \tag{R.4}$$

and the result is also valid for $s = t$.

(R.5).

Let us now consider results involving hitting paths. The derivation is similar to the proof of (R.4) and (R.2). Considering $i \neq t$,

$$\begin{aligned} w\left(\mathcal{P}^{(-i)}_{st}\right) &= \sum_{\wp_{st}\in\mathcal{P}_{st}} \delta(i\notin\wp_{st})w(\wp_{st}) \\ &= \sum_{\wp^{\mathrm{h}}_{st}\in\mathcal{P}^{\mathrm{h}}_{st}}\sum_{\wp_{tt}\in\mathcal{P}_{tt}} \delta(i\notin\wp^{\mathrm{h}}_{st})w(\wp^{\mathrm{h}}_{st})\,\delta(i\notin\wp_{tt})w(\wp_{tt}) \\ &= w\left(\mathcal{P}^{\mathrm{h}(-i)}_{st}\right) w\left(\mathcal{P}^{(-i)}_{tt}\right). \end{aligned}$$

Therefore, by isolating $w\left(\mathcal{P}^{\mathrm{h}(-i)}_{st}\right)$ and using (R.4) as well as (R.2),

$$w\left(\mathcal{P}^{\mathrm{h}(-i)}_{st}\right) = \frac{z^{(-i)}_{st}}{z^{(-i)}_{tt}} = \frac{z_{st} - z^{\mathrm{h}}_{si}z_{it}}{z_{tt} - z^{\mathrm{h}}_{ti}z_{it}} = \frac{z^{\mathrm{h}}_{st} - z^{\mathrm{h}}_{si}z^{\mathrm{h}}_{it}}{1 - z^{\mathrm{h}}_{ti}z^{\mathrm{h}}_{it}}, \tag{R.5}$$

where we divided the numerator and the denominator by z_{tt}. Now, if $i = t$, we trivially have that $\delta(i \notin \wp_{st}^{\mathrm{h}}) = 0$ and therefore $z_{st}^{\mathrm{h}(-t)} = 0$.

(R.6).

Still for hitting paths, if $i \neq t$, we obtain from $w\left(\mathcal{P}_{st}^{\mathrm{h}}\right) = w\left(\mathcal{P}_{st}^{\mathrm{h}(+i)}\right) + w\left(\mathcal{P}_{st}^{\mathrm{h}(-i)}\right)$ and from the previous result (R.5),

$$\begin{aligned} w\left(\mathcal{P}_{st}^{\mathrm{h}(+i)}\right) &= w\left(\mathcal{P}_{st}^{\mathrm{h}}\right) - w\left(\mathcal{P}_{st}^{\mathrm{h}(-i)}\right) = z_{st}^{\mathrm{h}} - z_{st}^{\mathrm{h}(-i)} \\ &= \frac{(1 - z_{ti}^{\mathrm{h}} z_{it}^{\mathrm{h}}) z_{st}^{\mathrm{h}}}{1 - z_{ti}^{\mathrm{h}} z_{it}^{\mathrm{h}}} - \frac{z_{st}^{\mathrm{h}} - z_{si}^{\mathrm{h}} z_{it}^{\mathrm{h}}}{1 - z_{ti}^{\mathrm{h}} z_{it}^{\mathrm{h}}} \\ &= \frac{(z_{si}^{\mathrm{h}} - z_{st}^{\mathrm{h}} z_{ti}^{\mathrm{h}}) z_{it}^{\mathrm{h}}}{1 - z_{ti}^{\mathrm{h}} z_{it}^{\mathrm{h}}} = z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}. \end{aligned} \tag{R.6}$$

Moreover, if $i = t$, obviously $z_{st}^{\mathrm{h}(+i)} = z_{st}^{\mathrm{h}}$.

A.2 *Results involving two intermediate nodes*

(R.7).

Assuming $i \neq j$, let $\mathcal{P}_{st}^{(i<j)}$ be the set of all regular paths containing i and j with node i appearing first on the path, that is, before the first visit to node j. Then

$$w\left(\mathcal{P}_{st}^{(+\{i,j\})}\right) = w\left(\mathcal{P}_{st}^{(i<j)}\right) + w\left(\mathcal{P}_{st}^{(j<i)}\right).$$

We can observe that there exists a bijection between $\mathcal{P}_{st}^{(i<j)}$ and $(\mathcal{P}_{si}^{\mathrm{h}(-j)} \circ \mathcal{P}_{ij}^{\mathrm{h}} \circ \mathcal{P}_{jt})$ with corresponding weights. Indeed, each path $\wp$ from s to t, visiting node i first and then node j, can be decomposed into a first hitting sub-path from s to the first occurrence of node i on $\wp$, followed by a second hitting sub-path from i to the first occurrence of node j, and finally a third regular sub-path from j to t. Thus,

$$w\left(\mathcal{P}_{st}^{(i<j)}\right) = w\left(\mathcal{P}_{si}^{\mathrm{h}(-j)}\right) w\left(\mathcal{P}_{ij}^{\mathrm{h}}\right) w\left(\mathcal{P}_{jt}\right) = z_{si}^{\mathrm{h}(-j)} z_{ij}^{\mathrm{h}} z_{jt}.$$

We therefore obtain from (R.6), for $i \neq j$,

$$w\left(\mathcal{P}_{st}^{(+\{i,j\})}\right) = z_{si}^{\mathrm{h}(-j)} z_{ij}^{\mathrm{h}} z_{jt} + z_{sj}^{\mathrm{h}(-i)} z_{ji}^{\mathrm{h}} z_{it} = z_{sj}^{\mathrm{h}(+i)} z_{jt} + z_{si}^{\mathrm{h}(+j)} z_{it}, \tag{R.7}$$

which is the desired result.

(R.8).

We will derive three different expressions for computing the quantity. For the first expression, by assuming $i \neq j$ and using (R.1), (R.3) and (R.7), we obtain

$$w\left(\mathcal{P}_{st}^{(-\{i,j\})}\right) = \sum_{\wp_{st} \in \mathcal{P}_{st}} (1 - \delta(i \in \wp_{st}))(1 - \delta(j \in \wp_{st}))\, w(\wp_{st})$$

$$
\begin{aligned}
&= z_{st} - z^{\mathrm{h}}_{si} z_{it} - z^{\mathrm{h}}_{sj} z_{jt} + w\left(\mathcal{P}^{(+\{i,j\})}_{st}\right) \\
&= z_{st} - z^{\mathrm{h}}_{si} z_{it} - z^{\mathrm{h}}_{sj} z_{jt} + \left(z^{\mathrm{h}(+i)}_{sj} z_{jt} + z^{\mathrm{h}(+j)}_{si} z_{it}\right) \\
&= z_{st} - \left(z^{\mathrm{h}}_{si} - z^{\mathrm{h}(+j)}_{si}\right) z_{it} - \left(z^{\mathrm{h}}_{sj} - z^{\mathrm{h}(+i)}_{sj}\right) z_{jt} \qquad \text{(A.2)} \\
&= z_{st} - z^{\mathrm{h}(-j)}_{si} z_{it} - z^{\mathrm{h}(-i)}_{sj} z_{jt}. \qquad \text{(R.8.1)}
\end{aligned}
$$

Note that expression (A.2) is obtained by rearranging the terms. Moreover, the last expression (R.8.1) is either obtained by direct calculation, or by observing that $z^{\mathrm{h}}_{si} = z^{\mathrm{h}(+j)}_{si} + z^{\mathrm{h}(-j)}_{si}$.

Let us derive still another expression for this quantity. As a regular path from s to t visiting node j can be decomposed uniquely into a hitting path from s to j (visiting j for the first time) followed by a regular path from j to t, we have

$$
\begin{aligned}
w\left(\mathcal{P}^{(-\{i,j\})}_{st}\right) &= \sum_{\wp_{st} \in \mathcal{P}_{st}} (1 - \delta(i \in \wp_{st}))(1 - \delta(j \in \wp_{st}))\, w(\wp_{st}) \\
&= \sum_{\wp_{st} \in \mathcal{P}_{st}} (1 - \delta(i \in \wp_{st}))\, w(\wp_{st}) \\
&\quad - \sum_{\wp_{st} \in \mathcal{P}_{st}} (1 - \delta(i \in \wp_{st}))\delta(j \in \wp_{st})\, w(\wp_{st}) \\
&= w\left(\mathcal{P}^{(-i)}_{st}\right) - \sum_{\wp_{st} \in \mathcal{P}_{st}} \delta(i \notin \wp_{st})\delta(j \in \wp_{st})\, w(\wp_{st}) \\
&= w\left(\mathcal{P}^{(-i)}_{st}\right) - \sum_{\wp^{\mathrm{h}}_{sj} \in \mathcal{P}^{\mathrm{h}}_{sj}} \sum_{\wp_{jt} \in \mathcal{P}_{jt}} \delta(i \notin \wp^{\mathrm{h}}_{sj}) w(\wp^{\mathrm{h}}_{sj})\, \delta(i \notin \wp_{jt}) w(\wp_{jt}) \\
&= w\left(\mathcal{P}^{(-i)}_{st}\right) - w\left(\mathcal{P}^{\mathrm{h}(-i)}_{sj}\right) w\left(\mathcal{P}^{(-i)}_{jt}\right) \\
&= z^{(-i)}_{st} - z^{\mathrm{h}(-i)}_{sj} z^{(-i)}_{jt}. \qquad \text{(R.8.3)}
\end{aligned}
$$

It can be shown in the same way that, symmetrically,

$$
w\left(\mathcal{P}^{(-\{i,j\})}_{st}\right) = z^{(-j)}_{st} - z^{\mathrm{h}(-j)}_{si} z^{(-j)}_{it}. \qquad \text{(R.8.2)}
$$

(R.9).

Let us consider hitting paths now. It is obvious that $w\left(\mathcal{P}^{\mathrm{h}(-\{i,j\})}_{st}\right) = 0$ if $i = t$ or $j = t$. Therefore, we assume $t \neq i \neq j \neq t$. Similarly to (R.2), we decompose paths $\wp_{st}$ in $\wp^{\mathrm{h}}_{st} \circ \wp_{tt}$,

$$
\begin{aligned}
w\left(\mathcal{P}^{(-\{i,j\})}_{st}\right) &= \sum_{\wp_{st} \in \mathcal{P}_{st}} \delta(i \notin \wp_{st})\delta(j \notin \wp_{st})\, w(\wp_{st}) \\
&= \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \delta(i \notin \wp^{\mathrm{h}}_{st})\delta(j \notin \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st}) \sum_{\wp_{tt} \in \mathcal{P}_{tt}} \delta(i \notin \wp_{tt})\delta(j \notin \wp_{tt})\, w(\wp_{tt}) \\
&= z^{\mathrm{h}(-\{i,j\})}_{st} z^{(-\{i,j\})}_{tt}.
\end{aligned}
$$

Consequently, $z_{st}^{\mathrm{h}(-\{i,j\})} = z_{st}^{(-\{i,j\})}/z_{tt}^{(-\{i,j\})}$ which proves (R.9); thus, with the help of (R.8.1)–(R.8.3),

$$\begin{aligned} z_{st}^{\mathrm{h}(-\{i,j\})} &= \frac{z_{st} - z_{si}^{\mathrm{h}(-j)} z_{it} - z_{sj}^{\mathrm{h}(-i)} z_{jt}}{z_{tt} - z_{ti}^{\mathrm{h}(-j)} z_{it} - z_{tj}^{\mathrm{h}(-i)} z_{jt}} && \text{(obtained from (R.8.1))} \\ &= \frac{z_{st}^{(-i)} - z_{sj}^{\mathrm{h}(-i)} z_{jt}^{(-i)}}{z_{tt}^{(-i)} - z_{tj}^{\mathrm{h}(-i)} z_{jt}^{(-i)}} && \text{(obtained from (R.8.3))} \\ &= \frac{z_{st}^{(-j)} - z_{si}^{\mathrm{h}(-j)} z_{it}^{(-j)}}{z_{tt}^{(-j)} - z_{ti}^{\mathrm{h}(-j)} z_{it}^{(-j)}}, && \text{(obtained from (R.8.2))} \end{aligned}$$

and by dividing numerators and denominators by, respectively, z_{tt}, $z_{tt}^{(-i)}$ and $z_{tt}^{(-j)}$, we get the results (R.9.1)–(R.9.3).

(R.10).

This result can be obtained by developing expressions in terms of z_{kl} and z_{kl}^{h}, then grouping them differently. However, we will use a more intuitive argument here, similar to the one used for deriving (R.7). Assuming $t \neq i \neq j \neq t$, let $\mathcal{P}_{st}^{\mathrm{h}(i<j)}$ be the set of all hitting paths containing i and j with node i appearing first (before node j) on the paths. We have

$$w\left(\mathcal{P}_{st}^{\mathrm{h}(+\{i,j\})}\right) = w\left(\mathcal{P}_{st}^{\mathrm{h}(i<j)}\right) + w\left(\mathcal{P}_{st}^{\mathrm{h}(j<i)}\right).$$

As for (R.7), observe that there is a bijection between $\mathcal{P}_{st}^{\mathrm{h}(i<j)}$ and $(\mathcal{P}_{si}^{\mathrm{h}(-\{j,t\})} \circ \mathcal{P}_{ij}^{\mathrm{h}(-t)} \circ \mathcal{P}_{jt}^{\mathrm{h}})$, with corresponding weights. Thus,

$$w\left(\mathcal{P}_{st}^{\mathrm{h}(i<j)}\right) = w\left(\mathcal{P}_{si}^{\mathrm{h}(-\{j,t\})}\right) w\left(\mathcal{P}_{ij}^{\mathrm{h}(-t)}\right) w\left(\mathcal{P}_{jt}^{\mathrm{h}}\right) = z_{si}^{\mathrm{h}(-\{j,t\})} z_{ij}^{\mathrm{h}(-t)} z_{jt}^{\mathrm{h}}.$$

Moreover, from (R.6),

$$\begin{aligned} w\left(\mathcal{P}_{st}^{\mathrm{h}(+\{i,j\})}\right) &= z_{si}^{\mathrm{h}(-\{j,t\})} z_{ij}^{\mathrm{h}(-t)} z_{jt}^{\mathrm{h}} + z_{sj}^{\mathrm{h}(-\{i,t\})} z_{ji}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}} \\ &= z_{si}^{\mathrm{h}(-\{j,t\})} z_{it}^{\mathrm{h}(+j)} + z_{sj}^{\mathrm{h}(-\{i,t\})} z_{jt}^{\mathrm{h}(+i)}. \end{aligned} \tag{R.10}$$

When $j = t$, $w\big(\mathcal{P}_{st}^{\mathrm{h}(+\{i,j\})}\big) = w\big(\mathcal{P}_{st}^{\mathrm{h}(+i)}\big) = z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$, from (R.6). It corresponds to the expression (R.10) above knowing that the first term gives $z_{si}^{\mathrm{h}(-\{t,t\})} z_{it}^{\mathrm{h}(+t)} = z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$ and that the second term is zero. The same applies for $i = t$.

A.3 *Results involving any number of intermediate nodes*

(R.11).

Recall that $\mathcal{S}_i = \mathcal{I} \setminus i$ for any $i \in \mathcal{I}$ or, in other words, $\mathcal{I} = \mathcal{S}_i \cup \{i\}$. Similarly to the proof of (R.3), we notice that there is a bijection between the set of paths $\mathcal{P}_{si}^{\mathrm{h}(-\mathcal{S}_i)} \circ \mathcal{P}_{it}^{(-\mathcal{S}_i)}$ and the set of paths in $\mathcal{P}_{st}^{(-\mathcal{S}_i)}$

containing i. Thus,

$$w\left(\mathcal{P}_{st}^{(-\mathcal{I})}\right) = w\left(\mathcal{P}_{st}^{(-\mathcal{S}_i\cup\{i\})}\right) = \sum_{\wp_{st}^{(-\mathcal{S}_i)}\in\mathcal{P}_{st}^{(-\mathcal{S}_i)}} \left(1-\delta(i\in\wp_{st}^{(-\mathcal{S}_i)})\right) w\left(\wp_{st}^{(-\mathcal{S}_i)}\right)$$
$$= z_{st}^{(-\mathcal{S}_i)} - \sum_{\wp_{si}^{\mathrm{h}(-\mathcal{S}_i)}\in\mathcal{P}_{si}^{\mathrm{h}(-\mathcal{S}_i)}} \sum_{\wp_{it}^{(-\mathcal{S}_i)}\in\mathcal{P}_{it}^{(-\mathcal{S}_i)}} w\left(\wp_{si}^{\mathrm{h}(-\mathcal{S}_i)}\right) w\left(\wp_{it}^{(-\mathcal{S}_i)}\right)$$
$$= z_{st}^{(-\mathcal{S}_i)} - z_{si}^{\mathrm{h}(-\mathcal{S}_i)} z_{it}^{(-\mathcal{S}_i)}. \tag{R.11}$$

(R.12).

Similarly to (R.2), with the bijection between $(\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{I})} \circ \mathcal{P}_{tt}^{(-\mathcal{I})})$ and $\mathcal{P}_{st}^{(-\mathcal{I})}$, we have

$$w\left(\mathcal{P}_{st}^{(-\mathcal{I})}\right) = w\left(\mathcal{P}_{st}^{\mathrm{h}(-\mathcal{I})}\right) w\left(\mathcal{P}_{tt}^{(-\mathcal{I})}\right) = z_{st}^{\mathrm{h}(-\mathcal{I})} z_{tt}^{(-\mathcal{I})}.$$

So, $z_{st}^{\mathrm{h}(-\mathcal{I})} = z_{st}^{(-\mathcal{I})} / z_{tt}^{(-\mathcal{I})}$, and we get the result (R.12) by using (R.11) and dividing the numerator and the denominator by $z_{tt}^{(-\mathcal{S}_i)}$.

(R.13)–(R.14).

We already know how to calculate weights of sets of paths avoiding a set of nodes. Conversely, let us now compute weights on path containing a set of nodes $\mathcal{I} = \mathcal{S}_i \cup \{i\}$ from these quantities. Indeed, $z_{st}^{(+\mathcal{I})} = w\left(\mathcal{P}_{st}^{(+\mathcal{I})}\right) = w\left(\mathcal{P}_{st}^{(+\mathcal{S}_i\cup\{i\})}\right)$ can be obtained through the *inclusion–exclusion principle* as follows (see e.g. [112]). In this context, the properties that are studied are the (mutually exclusive) presence or absence of nodes on the paths.

Observe that the total weight of the union of all sets of paths from s to t *avoiding at least one node in* $\mathcal{I}$, $w\left(\bigcup_{i\in\mathcal{I}} \mathcal{P}_{st}^{(-i)}\right)$, can be computed by ([112], Eq. (6.3)[7])

$$w\left(\bigcup_{i\in\mathcal{I}}\mathcal{P}_{st}^{(-i)}\right) = \sum_{\substack{\mathcal{S}\subseteq\mathcal{I}\\ \mathcal{S}\neq\varnothing}} (-1)^{(|\mathcal{S}|-1)} w\left(\mathcal{P}_{st}^{(-S)}\right) = -\sum_{\substack{\mathcal{S}\subseteq\mathcal{I}\\ \mathcal{S}\neq\varnothing}} (-1)^{|\mathcal{S}|} w\left(\mathcal{P}_{st}^{(-S)}\right),$$

where the summation on $\mathcal{S}\subseteq\mathcal{I}$ with $\mathcal{S}\neq\varnothing$ means a summation on all subsets $\mathcal{S}$ of $\mathcal{I}$, except the empty set. In the last equation, if we denote some subset $\mathcal{S} = \{i_1, i_2, \ldots, i_m\}$ (where nodes are distinct), then the set of paths avoiding all nodes in $\mathcal{S}$ is $\mathcal{P}_{st}^{(-S)} = \mathcal{P}_{st}^{(-i_1)} \cap \mathcal{P}_{st}^{(-i_2)} \cap \ldots \cap \mathcal{P}_{st}^{(-i_m)} = \bigcap_{i\in\mathcal{S}} \mathcal{P}_{st}^{(-i)}$. Now, it is clear that the set of paths visiting all nodes in $\mathcal{I}$ (the set we are interested in) is equal to the set of all paths between s and t minus the set of paths avoiding at least one node in $\mathcal{I}$ (the quantity in the last equation).

Therefore, from the additivity of weights on disjoint subsets, $w\left(\mathcal{P}_{st}\right) = w\left(\mathcal{P}_{st}^{(+\mathcal{I})}\right) + w\left(\bigcup_{i\in\mathcal{I}} \mathcal{P}_{st}^{(-i)}\right)$. We deduce that $w\left(\mathcal{P}_{st}^{(+\mathcal{I})}\right) = w\left(\mathcal{P}_{st}\right) - w\left(\bigcup_{i\in\mathcal{I}} \mathcal{P}_{st}^{(-i)}\right)$; thus, knowing that $w\left(\mathcal{P}_{st}\right) = z_{st}$ and $w\left(\mathcal{P}_{st}^{(-S)}\right) = z_{st}^{(-S)}$, we get the result (R.13). The same reasoning applies to hitting paths in order to derive (R.14).

[7] This equation computes the weight of objects in a set which have at least one of the m properties associated to these objects. In our case, the objects are s-t paths and the property is to avoid a node in $\mathcal{I}$.

B. Proofs of results of Section 3.3 (number of co-occurrences)

First, let us recall that the number of *occurrences of an edge* (i,j) on path $\wp$ is $\eta((i,j) \in \wp_{st})$. Observe also that *node occurrences* can be computed from edge occurrences by using

$$\eta(i \in \wp_{st}) = \sum_{j \in \mathcal{V}} \eta((i,j) \in \wp_{st}) + \delta_{it} = \sum_{j \in \mathcal{V}} \eta((j,i) \in \wp_{st}) + \delta_{is}, \tag{B.1}$$

where δ_{it} is the Kronecker delta. As a matter of fact, the total number of visits will be missing one unit if the node i is at the end (respectively the beginning) of the path as only outgoing (respectively ingoing) edges are counted in (B.1). Notice that these expressions are also valid for hitting paths.

Thus, in order to compute $\eta(i \in \wp_{st})$, we need to compute $\eta((i,j) \in \wp_{st})$ in terms of the elements of the fundamental matrix $\mathbf{Z}$. To this end, let us first prove a preliminary result.

B.1 *Number of occurrences in terms of partial derivatives of path weights*

From the definition of $w(\wp_{st})$ (Equation (2.1)),

$$w(\wp_{st}) = \prod_{\tau=0}^{\ell-1} w_{i_\tau i_{\tau+1}} = \prod_{\alpha,\beta=1}^{n} (w_{\alpha\beta})^{\eta((\alpha,\beta)\in\wp_{st})}.$$

Taking the partial derivative of this expression provides

$$\begin{aligned}
\frac{\partial(w(\wp_{st}))}{\partial w_{ij}} &= \eta((i,j) \in \wp_{st})\,(w_{ij})^{\eta((i,j)\in\wp_{st})-1} \prod_{\substack{\alpha,\beta=1 \\ (\alpha,\beta)\neq(i,j)}}^{n} (w_{\alpha\beta})^{\eta((\alpha,\beta)\in\wp_{st})} \\
&= \eta((i,j) \in \wp_{st})\,\frac{(w_{ij})^{\eta((i,j)\in\wp_{st})}}{w_{ij}} \prod_{\substack{\alpha,\beta=1 \\ (\alpha,\beta)\neq(i,j)}}^{n} (w_{\alpha\beta})^{\eta((\alpha,\beta)\in\wp_{st})} \\
&= \eta((i,j) \in \wp_{st})\frac{w(\wp_{st})}{w_{ij}}.
\end{aligned} \tag{B.2}$$

From this last result, the following relationship holds

$$\eta((i,j) \in \wp_{st})\, w(\wp_{st}) = w_{ij}\frac{\partial(w(\wp_{st}))}{\partial w_{ij}}. \tag{B.3}$$

Taking once more the partial derivative of $w(\wp_{st})$ with respect to w_{kl}, $(k,l) \neq (i,j)$, yields

$$\begin{aligned}
\frac{\partial^2(w(\wp_{st}))}{\partial w_{kl}\partial w_{ij}} &= \eta((i,j) \in \wp_{st})\,(w_{ij})^{\eta((i,j)\in\wp_{st})-1}\eta((k,l) \in \wp_{st})\,(w_{kl})^{\eta((k,l)\in\wp_{st})-1} \\
&\quad \times \prod_{\substack{\alpha,\beta=1 \\ (\alpha,\beta)\neq\{(i,j),(k,l)\}}}^{n} (w_{\alpha\beta})^{\eta((\alpha,\beta)\in\wp_{st})}
\end{aligned}$$

$$= \eta((i,j) \in \wp_{st})\eta((k,l) \in \wp_{st}) \frac{w(\wp_{st})}{w_{ij}w_{kl}}. \tag{B.4}$$

Therefore, for $(k,l) \neq (i,j)$,

$$\eta((i,j) \in \wp_{st})\eta((k,l) \in \wp_{st})\, w(\wp_{st}) = w_{ij}w_{kl} \frac{\partial^2(w(\wp_{st}))}{\partial w_{kl}\partial w_{ij}}.$$

Now, proceeding in the same way for the case $(i,j) = (k,l)$, we obtain from Equation (B.2)

$$\frac{\partial^2(w(\wp_{st}))}{(\partial w_{ij})^2} = \eta((i,j) \in \wp_{st})\eta((i,j) \in \wp_{st}) \frac{w(\wp_{st})}{(w_{ij})^2} - \eta((i,j) \in \wp_{st}) \frac{w(\wp_{st})}{(w_{ij})^2}. \tag{B.5}$$

After multiplying this last equation by $(w_{ij})^2$ and then using (B.3), this provides an additional term, leading to the following, more general, expression,

$$\eta((i,j) \in \wp_{st})\eta((k,l) \in \wp_{st})\, w(\wp_{st}) = w_{ij}w_{kl} \frac{\partial^2(w(\wp_{st}))}{\partial w_{kl}\partial w_{ij}} + \delta_{ik}\delta_{jl}w_{ij} \frac{\partial(w(\wp_{st}))}{\partial w_{ij}}, \tag{B.6}$$

which is also valid when $(i,j) = (k,l)$. Equations (B.3) and (B.6) are the first needed preliminary results. Note that these two expressions also hold for hitting paths.

B.2 *Number of occurrences in terms of the fundamental matrix*

From here, results for the non-hitting and hitting paths will differ. Before proceeding, we still need some other useful results showing that

$$\frac{\partial z_{st}}{\partial w_{ij}} = z_{si}z_{jt}, \tag{B.7}$$

$$\frac{\partial^2 z_{st}}{\partial w_{kl}\partial w_{ij}} = z_{si}z_{jk}z_{lt} + z_{sk}z_{li}z_{jt}. \tag{B.8}$$

These expressions are obtained by using the standard formula computing the partial derivative of a matrix inverse, $\frac{\partial \mathbf{M}^{-1}}{\partial m_{ij}} = -\mathbf{M}^{-1}\frac{\partial \mathbf{M}}{\partial m_{ij}}\mathbf{M}^{-1}$ (details and similar approaches can be found in [14, 15, 18, 67]). When applying this formula to the fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}$,

$$\begin{aligned}\frac{\partial z_{st}}{\partial w_{ij}} &= \mathbf{e}_s^\top \frac{\partial \mathbf{Z}}{\partial w_{ij}}\mathbf{e}_t = -\mathbf{e}_s^\top \mathbf{Z}\frac{\partial(\mathbf{I}-\mathbf{W})}{\partial w_{ij}}\mathbf{Z}\mathbf{e}_t \\ &= \mathbf{e}_s^\top \mathbf{Z}\mathbf{e}_i\mathbf{e}_j^\top \mathbf{Z}\mathbf{e}_t = z_{si}z_{jt},\end{aligned}$$

which is the first result (B.7). Taking once more the partial derivative provides (B.8),

$$\frac{\partial^2 z_{st}}{\partial w_{kl}\partial w_{ij}} = \frac{\partial(z_{si}z_{jt})}{\partial w_{kl}} = \frac{\partial(z_{si})}{\partial w_{kl}}z_{jt} + z_{si}\frac{\partial(z_{jt})}{\partial w_{kl}} = z_{sk}z_{li}z_{jt} + z_{si}z_{jk}z_{lt}.$$

We are now ready to prove the next results (R.15)–(R.18).

B.2.1 *Results for regular paths*

(R.15)–(R.16).

For regular paths, following (B.3), (R.1) and (B.7), we find for edge occurrences

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((i,j)\in\wp_{st})\, w(\wp_{st}) = \sum_{\wp_{st}\in\mathcal{P}_{st}} \frac{\partial w(\wp_{st})}{\partial w_{ij}} w_{ij} = \frac{\partial \left(\sum_{\wp_{st}\in\mathcal{P}_{st}} w(\wp_{st})\right)}{\partial w_{ij}} w_{ij}$$
$$= \frac{\partial z_{st}}{\partial w_{ij}} w_{ij} = z_{si} w_{ij} z_{jt}. \tag{B.9}$$

Moreover, from (B.6), (R.1) and (B.7) and (B.8),

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((i,j)\in\wp_{st})\eta((k,l)\in\wp_{st})\, w(\wp_{st})$$
$$= \sum_{\wp_{st}\in\mathcal{P}_{st}} \frac{\partial^2 w(\wp_{st})}{\partial w_{kl}\partial w_{ij}} w_{ij} w_{kl} + \delta_{ik}\delta_{jl} \sum_{\wp_{st}\in\mathcal{P}_{st}} \frac{\partial w(\wp_{st})}{\partial w_{ij}} w_{ij}$$
$$= \frac{\partial^2 z_{st}}{\partial w_{kl}\partial w_{ij}} w_{ij} w_{kl} + \delta_{ik}\delta_{jl} \frac{\partial z_{st}}{\partial w_{ij}} w_{ij}$$
$$= z_{si} w_{ij} z_{jk} w_{kl} z_{lt} + z_{sk} w_{kl} z_{li} w_{ij} z_{jt} + \delta_{ik}\delta_{jl} z_{si} w_{ij} z_{jt}. \tag{B.10}$$

This is almost the expected result, as we are mainly interested in the closely related quantities involving node occurrences instead of edge occurrences, $\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i\in\wp_{st})\, w(\wp_{st})$ and $\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i\in\wp_{st})\eta(k\in\wp_{st})\, w(\wp_{st})$. For the first expression, using (R.1), (B.1), (B.9) and $(\mathbf{I}-\mathbf{W})\mathbf{Z}=\mathbf{I}$, that is, $\sum_{i\in\mathcal{V}} w_{si} z_{it} = z_{st} - \delta_{st}$, finally provides (R.15)

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i\in\wp_{st})\, w(\wp_{st}) = \sum_{\wp_{st}\in\mathcal{P}_{st}} \left(\sum_{j\in\mathcal{V}} \eta((i,j)\in\wp_{st}) + \delta_{it}\right) w(\wp_{st})$$
$$= \sum_{j\in\mathcal{V}} \underbrace{\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((i,j)\in\wp_{st})\, w(\wp_{st})}_{z_{si} w_{ij} z_{jt}} + \delta_{it} \underbrace{\sum_{\wp_{st}\in\mathcal{P}_{st}} w(\wp_{st})}_{z_{st}}$$
$$= \sum_{j\in\mathcal{V}} z_{si} w_{ij} z_{jt} + \delta_{it} z_{st}$$
$$= z_{si}(z_{it} - \delta_{it}) + \delta_{it} z_{st} = z_{si} z_{it}. \tag{R.15}$$

Let us now compute the second expression,

$$\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i\in\wp_{st})\eta(k\in\wp_{st})\, w(\wp_{st})$$
$$= \sum_{\wp_{st}\in\mathcal{P}_{st}} \left(\sum_{j\in\mathcal{V}} \eta((i,j)\in\wp_{st}) + \delta_{it}\right)\left(\sum_{l\in\mathcal{V}} \eta((k,l)\in\wp_{st}) + \delta_{kt}\right) w(\wp_{st})$$

$$
\begin{aligned}
&= \sum_{\wp_{st}\in\mathcal{P}_{st}} \Big(\sum_{j\in\mathcal{V}} \eta((i,j)\in\wp_{st})\Big)\Big(\sum_{l\in\mathcal{V}} \eta((k,l)\in\wp_{st})\Big)\, w(\wp_{st}) \\
&\quad + \sum_{\wp_{st}\in\mathcal{P}_{st}} \Big(\sum_{j\in\mathcal{V}} \eta((i,j)\in\wp_{st})\Big)\delta_{kt}\, w(\wp_{st}) + \sum_{\wp_{st}\in\mathcal{P}_{st}} \Big(\sum_{l\in\mathcal{V}} \eta((k,l)\in\wp_{st})\Big)\delta_{it}\, w(\wp_{st}) \\
&\quad + \sum_{\wp_{st}\in\mathcal{P}_{st}} \delta_{it}\delta_{kt}\, w(\wp_{st}) \\
&= \sum_{j\in\mathcal{V}}\sum_{l\in\mathcal{V}} \Big(\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((i,j)\in\wp_{st})\eta((k,l)\in\wp_{st})\, w(\wp_{st})\Big) \\
&\quad + \delta_{kt}\sum_{j\in\mathcal{V}} \Big(\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((i,j)\in\wp_{st})\, w(\wp_{st})\Big) + \delta_{it}\sum_{l\in\mathcal{V}} \Big(\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta((k,l)\in\wp_{st})\, w(\wp_{st})\Big) \\
&\quad + \delta_{it}\delta_{kt}\Big(\sum_{\wp_{st}\in\mathcal{P}_{st}} w(\wp_{st})\Big). \qquad \text{(B.11)}
\end{aligned}
$$

From this last result, further using Equations (B.9), (B.10), (R.1) and, again, $\sum_{i\in\mathcal{V}} w_{si}z_{it} = z_{st} - \delta_{st}$ provides (R.16),

$$
\begin{aligned}
&\sum_{\wp_{st}\in\mathcal{P}_{st}} \eta(i\in\wp_{st})\eta(k\in\wp_{st})\, w(\wp_{st}) \\
&\qquad = \sum_{j\in\mathcal{V}}\sum_{l\in\mathcal{V}} \big(z_{si}w_{ij}z_{jk}w_{kl}z_{lt} + z_{sk}w_{kl}z_{li}w_{ij}z_{jt} + \delta_{ik}\delta_{jl}z_{si}w_{ij}z_{jt}\big) \\
&\qquad\quad + \delta_{kt}\sum_{j\in\mathcal{V}} \big(z_{si}w_{ij}z_{jt}\big) + \delta_{it}\sum_{l\in\mathcal{V}} (z_{sk}w_{kl}z_{lt}) + \delta_{it}\delta_{kt}\, z_{st} \\
&\qquad = \big(z_{si}(z_{ik}-\delta_{ik})(z_{kt}-\delta_{kt}) + z_{sk}(z_{ki}-\delta_{ki})(z_{it}-\delta_{it}) + \delta_{ik}z_{si}(z_{it}-\delta_{it})\big) \\
&\qquad\quad + \delta_{kt}z_{si}(z_{it}-\delta_{it}) + \delta_{it}z_{sk}(z_{kt}-\delta_{kt}) + \delta_{it}\delta_{kt}\, z_{st} \\
&\qquad = z_{si}z_{ik}z_{kt} + z_{sk}z_{ki}z_{it} - \delta_{ik}z_{si}z_{kt}. \qquad \text{(R.16)}
\end{aligned}
$$

B.2.2 *Results for hitting paths*

(R.17)–(R.18).

We proceed in the same way for hitting paths. By using the expressions (B.7)–(B.8) for computing the partial derivative of $z^{\mathrm{h}}_{st} = z_{st}/z_{tt}$ (see Equation (R.2)), and using (R.4), we get

$$
\frac{\partial z^{\mathrm{h}}_{st}}{\partial w_{ij}} = \frac{\partial z_{st}}{\partial w_{ij}} z_{tt}^{-1} + z_{st}\frac{\partial z_{tt}^{-1}}{\partial w_{ij}} = \big(z_{si} - z^{\mathrm{h}}_{st}z_{ti}\big)\, z^{\mathrm{h}}_{jt} = z_{si}^{(-t)}z^{\mathrm{h}}_{jt}, \qquad \text{(B.12)}
$$

$$
\frac{\partial^2 z^{\mathrm{h}}_{st}}{\partial w_{kl}\partial w_{ij}} = (z_{si} - z^{\mathrm{h}}_{st}z_{ti})(z_{jk} - z^{\mathrm{h}}_{jt}z_{tk})z^{\mathrm{h}}_{lt} + (z_{sk} - z^{\mathrm{h}}_{st}z_{tk})(z_{li} - z^{\mathrm{h}}_{lt}z_{ti})z^{\mathrm{h}}_{jt}
$$

$$= z^{(-t)}_{si} z^{(-t)}_{jk} z^{\mathrm{h}}_{lt} + z^{(-t)}_{sk} z^{(-t)}_{li} z^{\mathrm{h}}_{jt}. \tag{B.13}$$

Observe the similarity with the results for regular paths, (B.7) and (B.8).

As for the previous subsection (see Equations (B.9), (B.10)), by using (B.3)–(B.6), (R.2), and the previous expressions for the partial derivatives of z^{h}_{st}, we obtain the equivalent of (B.9) and (B.10) for hitting paths,

$$\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta((i,j) \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st}) = \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \frac{\partial w(\wp^{\mathrm{h}}_{st})}{\partial w_{ij}} w_{ij} = \frac{\partial z^{\mathrm{h}}_{st}}{\partial w_{ij}} w_{ij} = z^{(-t)}_{si} w_{ij} z^{\mathrm{h}}_{jt}, \tag{B.14}$$

and

$$\begin{aligned}
\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta((i,j) \in \wp^{\mathrm{h}}_{st}) \eta((k,l) \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st}) &= \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \frac{\partial w(\wp^{\mathrm{h}}_{st})}{\partial w_{kl} \partial w_{ij}} w_{ij} w_{kl} + \delta_{ik}\delta_{jl} \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \frac{\partial w(\wp^{\mathrm{h}}_{st})}{\partial w_{ij}} w_{ij} \\
&= \frac{\partial z^{\mathrm{h}}_{st}}{\partial w_{kl} \partial w_{ij}} w_{ij} w_{kl} + \delta_{ik}\delta_{jl} \frac{\partial z^{\mathrm{h}}_{st}}{\partial w_{ij}} w_{ij} \\
&= z^{(-t)}_{si} w_{ij} z^{(-t)}_{jk} w_{kl} z^{\mathrm{h}}_{lt} + z^{(-t)}_{sk} w_{kl} z^{(-t)}_{li} w_{ij} z^{\mathrm{h}}_{jt} + \delta_{ik}\delta_{jl} z^{(-t)}_{si} w_{ij} z^{\mathrm{h}}_{jt}.
\end{aligned} \tag{B.15}$$

We observe that these results are surprisingly similar to the equivalent results for regular paths, displayed in Equations (B.9) and (B.10).

But, as before, we are in fact interested in the closely related quantities involving node occurrences, $\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st})$ and $\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st}) \eta(k \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st})$. For the first expression, using (R.1), (R.2), (B.1), (B.14) and $\sum_{i \in \mathcal{V}} w_{si} z_{it} = z_{st} - \delta_{st}$, provides result (R.17):

$$\begin{aligned}
\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st}) &= \sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \left(\sum_{j \in \mathcal{V}} \eta((i,j) \in \wp^{\mathrm{h}}_{st}) + \delta_{it} \right) w(\wp^{\mathrm{h}}_{st}) \\
&= \sum_{j \in \mathcal{V}} \underbrace{\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta((i,j) \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st})}_{z^{(-t)}_{si} w_{ij} z^{\mathrm{h}}_{jt}} + \delta_{it} \underbrace{\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} w(\wp^{\mathrm{h}}_{st})}_{z^{\mathrm{h}}_{st}} \\
&= \sum_{j \in \mathcal{V}} z^{(-t)}_{si} w_{ij} z^{\mathrm{h}}_{jt} + \delta_{it} z^{\mathrm{h}}_{st} = z^{(-t)}_{si} \sum_{j \in \mathcal{V}} w_{ij} \frac{z_{jt}}{z_{tt}} + \delta_{it} z^{\mathrm{h}}_{st} \\
&= z^{(-t)}_{si} \frac{(z_{it} - \delta_{it})}{z_{tt}} + \delta_{it} z^{\mathrm{h}}_{st} = z^{(-t)}_{si} z^{\mathrm{h}}_{it} + \delta_{it} z^{\mathrm{h}}_{st},
\end{aligned} \tag{R.17}$$

where we used $z^{(-t)}_{si} \delta_{it} = 0$ (see (R.5)) in the last equality.

Let us now compute the second expression by proceeding in the same way as in the previous section for non-hitting paths (see Equation (B.11)),

$$\sum_{\wp^{\mathrm{h}}_{st} \in \mathcal{P}^{\mathrm{h}}_{st}} \eta(i \in \wp^{\mathrm{h}}_{st}) \eta(k \in \wp^{\mathrm{h}}_{st})\, w(\wp^{\mathrm{h}}_{st})$$

$$
\begin{aligned}
&= \sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \Big(\sum_{j \in \mathcal{V}} \eta((i,j) \in \wp_{st}^{\mathrm{h}}) + \delta_{it} \Big) \Big(\sum_{l \in \mathcal{V}} \eta((k,l) \in \wp_{st}^{\mathrm{h}}) + \delta_{kt} \Big) w(\wp_{st}^{\mathrm{h}}) \\
&= \sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \Big(\sum_{j \in \mathcal{V}} \eta((i,j) \in \wp_{st}^{\mathrm{h}}) \Big) \Big(\sum_{l \in \mathcal{V}} \eta((k,l) \in \wp_{st}^{\mathrm{h}}) \Big) w(\wp_{st}^{\mathrm{h}}) \\
&\quad + \sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \Big(\sum_{j \in \mathcal{V}} \eta((i,j) \in \wp_{st}^{\mathrm{h}}) \Big) \delta_{kt}\, w(\wp_{st}^{\mathrm{h}}) + \sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \Big(\sum_{l \in \mathcal{V}} \eta((k,l) \in \wp_{st}^{\mathrm{h}}) \Big) \delta_{it}\, w(\wp_{st}^{\mathrm{h}}) \\
&\quad + \sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \delta_{it} \delta_{kt}\, w(\wp_{st}^{\mathrm{h}}).
\end{aligned} \tag{B.16}
$$

Further using Equations (B.14), (B.15), (R.1), (R.4) provides

$$
\begin{aligned}
&\sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \eta(i \in \wp_{st}^{\mathrm{h}}) \eta(k \in \wp_{st}^{\mathrm{h}})\, w(\wp_{st}^{\mathrm{h}}) \\
&= \sum_{j,l \in \mathcal{V}} \Big(z_{si}^{(-t)} w_{ij} z_{jk}^{(-t)} w_{kl} z_{lt}^{\mathrm{h}} + z_{sk}^{(-t)} w_{kl} z_{li}^{(-t)} w_{ij} z_{jt}^{\mathrm{h}} + \delta_{ik} \delta_{jl} z_{si}^{(-t)} w_{ij} z_{jt}^{\mathrm{h}} \Big) \\
&\quad + \delta_{kt} \sum_{j \in \mathcal{V}} \Big(z_{si}^{(-t)} w_{ij} z_{jt}^{\mathrm{h}} \Big) + \delta_{it} \sum_{l \in \mathcal{V}} \Big(z_{sk}^{(-t)} w_{kl} z_{lt}^{\mathrm{h}} \Big) + \delta_{it} \delta_{kt}\, z_{st}^{\mathrm{h}} \\
&= z_{si}^{(-t)} \Big(\Big(\sum_{j \in \mathcal{V}} w_{ij} z_{jk}^{(-t)} \Big) \Big(\sum_{l \in \mathcal{V}} w_{kl} z_{lt}^{\mathrm{h}} \Big) \Big) + z_{sk}^{(-t)} \Big(\Big(\sum_{l \in \mathcal{V}} w_{kl} z_{li}^{(-t)} \Big) \Big(\sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}} \Big) \Big) \\
&\quad + \delta_{ik} z_{si}^{(-t)} \Big(\sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}} \Big) + \delta_{kt} z_{si}^{(-t)} \Big(\sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}} \Big) + \delta_{it} z_{sk}^{(-t)} \Big(\sum_{l \in \mathcal{V}} w_{kl} z_{lt}^{\mathrm{h}} \Big) \\
&\quad + \delta_{it} \delta_{kt}\, z_{st}^{\mathrm{h}}.
\end{aligned} \tag{B.17}
$$

Before going further, note that we encounter expressions like $\sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}}$ and $\sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{(-t)}$, each multiplied by $z_{si}^{(-t)}$. Let us compute these expressions. We already know that $(\mathbf{I} - \mathbf{W})\mathbf{Z} = \mathbf{I}$, or elementwise, $\sum_{j \in \mathcal{V}} w_{sj} z_{jt} = z_{st} - \delta_{st}$. Therefore, from (R.2),

$$
z_{si}^{(-t)} \sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}} = \frac{z_{si}^{(-t)}}{z_{tt}} \sum_{j \in \mathcal{V}} w_{ij} z_{jt} = \frac{z_{si}^{(-t)}}{z_{tt}} (z_{it} - \delta_{it}) = \frac{z_{si}^{(-t)}}{z_{tt}} z_{it} = z_{si}^{(-t)} z_{it}^{\mathrm{h}}, \tag{B.18}
$$

and the last equality holds because $z_{si}^{(-t)}$ is equal to 0 when $i = t$ (see Equation (R.4) and the following discussion) so that $\delta_{it} z_{si}^{(-t)} = 0$. It means that this simplification is only valid when $z_{si}^{(-t)}$ is present. Moreover, in the same way, following Equation (R.4) and using the last expression (B.18),

$$
\begin{aligned}
z_{si}^{(-t)} \sum_{j \in \mathcal{V}} w_{ij} z_{jk}^{(-t)} &= z_{si}^{(-t)} \sum_{j \in \mathcal{V}} w_{ij} (z_{jk} - z_{jt}^{\mathrm{h}} z_{tk}) = z_{si}^{(-t)} \Big(\sum_{j \in \mathcal{V}} w_{ij} z_{jk} - z_{tk} \sum_{j \in \mathcal{V}} w_{ij} z_{jt}^{\mathrm{h}} \Big) \\
&= z_{si}^{(-t)} \big((z_{ik} - \delta_{ik}) - z_{tk} z_{it}^{\mathrm{h}} \big) = z_{si}^{(-t)} \big(z_{ik}^{(-t)} - \delta_{ik} \big).
\end{aligned} \tag{B.19}
$$

Injecting successively the expressions (B.18) and (B.19) in Equation (B.17) provides (R.18),

$$\begin{aligned}
&\sum_{\wp_{st}^{\mathrm{h}} \in \mathcal{P}_{st}^{\mathrm{h}}} \eta(i \in \wp_{st}^{\mathrm{h}}) \eta(k \in \wp_{st}^{\mathrm{h}})\, w(\wp_{st}^{\mathrm{h}}) \\
&\quad = z_{si}^{(-t)} z_{kt}^{\mathrm{h}} \sum_{j \in \mathcal{V}} \left(w_{ij} z_{jk}^{(-t)} \right) + z_{sk}^{(-t)} z_{it}^{\mathrm{h}} \sum_{l \in \mathcal{V}} \left(w_{kl} z_{li}^{(-t)} \right) \\
&\qquad + \delta_{ik} z_{si}^{(-t)} z_{it}^{\mathrm{h}} + \delta_{kt} z_{si}^{(-t)} z_{it}^{\mathrm{h}} + \delta_{it} z_{sk}^{(-t)} z_{kt}^{\mathrm{h}} + \delta_{it} \delta_{kt}\, z_{st}^{\mathrm{h}} \\
&\quad = z_{si}^{(-t)} \left(z_{ik}^{(-t)} - \delta_{ik} \right) z_{kt}^{\mathrm{h}} + z_{sk}^{(-t)} \left(z_{ki}^{(-t)} - \delta_{ki} \right) z_{it}^{\mathrm{h}} \\
&\qquad + \delta_{ik} z_{si}^{(-t)} z_{kt}^{\mathrm{h}} + \delta_{kt} z_{si}^{(-t)} z_{it}^{\mathrm{h}} + \delta_{it} z_{sk}^{(-t)} z_{kt}^{\mathrm{h}} + \delta_{it} \delta_{kt} z_{st}^{\mathrm{h}} \\
&\quad = z_{si}^{(-t)} z_{ik}^{(-t)} z_{kt}^{\mathrm{h}} + z_{sk}^{(-t)} z_{ki}^{(-t)} z_{it}^{\mathrm{h}} \\
&\qquad - \delta_{ik} z_{si}^{(-t)} z_{kt}^{\mathrm{h}} + \delta_{kt} z_{si}^{(-t)} z_{it}^{\mathrm{h}} + \delta_{it} z_{sk}^{(-t)} z_{kt}^{\mathrm{h}} + \delta_{it} \delta_{kt} z_{st}^{\mathrm{h}}.
\end{aligned} \tag{R.18}$$

Note that, contrary to regular paths (see Equation (R.16)), in this case, some terms containing Kronecker deltas do not cancel out, leading to a more complex expression.

C. Discussion of Equations (4.5) and (4.6)

This appendix provides some additional details about the derivation of Equations (4.5) and (4.6), in particular the additional diagonal terms containing Kronecker delta.

Recall that terms like $-\delta_{ij} z_{\bullet i}^{\mathrm{h}} z_{i\bullet}$, $\delta_{ij} z_{\bullet i}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$ and $\delta_{ij} z_{\bullet i}^{\mathrm{h}}$ were added to Equations (4.5) and (4.6) in order to take into account the special cases where $i = j$. Indeed, the basic Equations (R.7) and (R.10) concern only the situation where $i \neq j$ and must be extended in order to compute these diagonal terms.

In the first case (4.5) (regular paths), when $i = j$, the expression in (R.7) provides $2 z_{si}^{\mathrm{h}(+i)} z_{it}$. However, the result should instead provide Equation (R.3), that is, the expression for the total weight computed on paths containing only one node i. But this expression (R.3) is equal to $z_{si}^{\mathrm{h}(+i)} z_{it}$ so that we have to remove one time this quantity to (R.7) in order to obtain (R.3) when $i = j$. Hence, the subtraction of $z_{si}^{\mathrm{h}(+i)} z_{it}$ appearing in Equation (4.5).

For the second case (4.6) (hitting paths), the expression in (R.10) provides 0 when $i = j$ (and $i \neq t$). However, the result should instead be Equation (R.6), the corresponding expression for the total weight computed on hitting paths containing only one node i. This expression (R.6) is equal to $z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$ so that we have to add this quantity to (R.10) in order to obtain (R.6) when $i = j$ (and $i \neq t$). Hence, the addition of $z_{si}^{\mathrm{h}(-t)} z_{it}^{\mathrm{h}}$ appearing in the first line of Equation (4.6).

The last line of Equation (4.6) aims at taking care of the special cases $i = t$ and $j = t$, which are also prohibited in (R.10). In the situation $i = t$, i is automatically part of the paths and the quantity $z_{st}^{\mathrm{h}(+\{i,j\})}$ should reduce to $z_{st}^{\mathrm{h}(+\{t,j\})} = z_{st}^{\mathrm{h}(+j)} = z_{si}^{\mathrm{h}(+j)}$, that is, to (R.6 with $t \neq i$). Hence the addition of the term $z_{sj}^{\mathrm{h}(-i)} z_{ji}^{\mathrm{h}}$ in the last line of (4.6). Symmetrically, the same holds for the case $j = t$, with the introduction of $z_{\bullet i}^{\mathrm{h}(-j)} z_{ij}^{\mathrm{h}}$ in the last line of (4.6).

Finally, as both $z_{sj}^{\mathrm{h}(-i)} z_{ji}^{\mathrm{h}}$ and $z_{\bullet i}^{\mathrm{h}(-j)} z_{ij}^{\mathrm{h}}$ are equal to zero when $i = j$, we still need to handle the case $i = j = t$ and add the corresponding contribution. In that situation, the contribution is $z_{si}^{\mathrm{h}(+j)} = z_{si}^{\mathrm{h}(+i)}$, that is, (R.6 with $t = i$) which provides z_{si}^{h}. This corresponds to the last term of Equation (4.6).