



A Simple Extension of the Bag-of-Paths Model Weighting Path Lengths by a Poisson Distribution

Sylvain Courtain^{1(✉)} and Marco Saerens²

¹ LOURIM, Université catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium
sylvain.courtain@uclouvain.be

² ICTEAM, Université catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium

Abstract. This work extends the bag-of-paths model [14] by introducing a weighting of the length of the paths in the network, provided by a Poisson probability distribution. The main advantage of this approach is that it allows to tune the mean path length parameter which is most relevant for the application at hand. Various quantities of interest, such as the probability of drawing a path from the bag of paths, or the join probability of sampling any path connecting two nodes of interest, can easily be computed in closed form from this model. In this context, a new distance measure between nodes of a network, considering a weighting factor on the length of the paths, is defined. Experiments on semi-supervised classification tasks show that the introduced distance measure provides competitive results compared to other state-of-the-art methods. Moreover, a new interpretation of the logarithmic communicability similarity measure [17] is proposed in terms of the new model.

Keywords: Network sciences · Link analysis · Distances between nodes · Bag-of-paths model

1 Introduction

1.1 General Introduction

The **bag-of-paths** (BoP) model is a framework finding its roots in transportation sciences [1], which was further expanded by exploiting basic concepts from statistical physics in the context of network data analysis [14, 19, 27, 33]. More precisely and in short, the BoP puts a Gibbs-Boltzmann probability distribution on the set of all possible paths between pairs of nodes, which depends on an inverse temperature parameter θ . The θ parameter allows to monitor the balance between *exploration* and *exploitation*. Indeed, when $\theta \rightarrow 0^+$, the likelihood of selecting a path is reduced to the probabilities generated by a random walk on the graph, or network, G (pure exploration). On the other hand, when $\theta \rightarrow \infty$, the probability distribution is biased, and only the least-cost paths are considered (pure exploitation). Classically, in this framework, the probability of

following a path on the network is obtained by minimizing the expected path cost regularized by the relative entropy, or Kullback-Leibler divergence, with respect to reference path probabilities, which leads to a Gibbs-Boltzmann probability distribution [33].

In this paper, we propose an extension of the BoP framework by weighting the paths by a Poisson probability distribution on their length and assuming the independence between path length and path likelihood. Interestingly, this simple model leads to a similarity measure between nodes which is closely related to the log-communicability measure (see [12, 17] for the communicability measure), and thus provides a new interpretation of this quantity. The present work is therefore part of a desire to broaden the scope of the BoP framework. Indeed, during this decade, several works have already proposed theoretical and practical extensions of this framework. Among them, we find the addition of capacity constraints on edges [8] and nodes [24] of the network, the addition of flow constraints on source and target nodes [7, 16], the use of the Tsallis divergence in place of Kullback-Leibler divergence for regularization [23], the application of the BoP model to fraud detection system [9], and collaborative recommendation [24].

The main advantage of this new extension is that the probability of a path can be expressed by taking explicitly into account the length ℓ of this path; it can be considered as a way of penalizing longer paths in the network and tuning the mean length depending on the application. In this context, path lengths are weighted by a Poisson distribution whose parameter is the mean path length. From this model, a new distance measure between nodes of the graph is introduced and compared to other distance measures, the *Poisson weighted surprisal distance*.

1.2 Contributions and Contents of the Paper

The main contributions of this paper are as follows:

- The definition of a new distance measure between nodes of a graph based on the introduction of a Poisson distribution weighting path lengths in the path likelihood.
- A new interpretation of the logarithmic communicability similarity measure introduced by Ivashkin and Chebotarev [17].
- An experimental comparison between the introduced distance and other baseline dissimilarities on kernel-based semi-supervised classification tasks.

The paper is organized as follows. First, the randomized shortest paths (RSP) and BoP frameworks are briefly presented in Sect. 2. Then, Sect. 3 introduces the extension of the BoP model taking the length of the paths into account. A new distance measure between nodes of a graph is derived from this model in the same section. This new distance measure is assessed and compared through a semi-supervised classification task in Sect. 4. Finally, Sect. 5 concludes the work.

2 The Standard Randomized Shortest Paths and Bag-of-Paths Models

In this section, we start by introducing the RSP model [20,27,33] and some quantities of interest. Then, we present the BoP model [14], which is an extension of the RSP framework where all paths in the network are considered, instead of single source-target paths only.

2.1 The Standard Randomized Shortest Paths Model

Let us consider a weighted, strongly connected, directed¹, graph G containing n nodes $\in \mathcal{V}$ (the set of nodes), and non-negative costs c_{ij} together with affinities a_{ij} (adjacency matrix) associated to (directed) edges. Recall that the RSP probability [20,27,33] of observing a path $\wp_{st}$ from *source* node s to *target* node t of length $\ell(\wp_{st})$ (or simply ℓ if there is no ambiguity), and thus node sequence $(i_0 = s, i_1, i_2, \dots, i_{\ell-1}, i_\ell = t)$ with i_τ being the index of the node at position τ in the sequence, is provided by a Gibbs-Boltzmann distribution

$$P_{\text{RSP}}(\wp_{st}) = \frac{P_{\text{RW}}(\wp_{st}) \exp(-\theta c(\wp_{st}))}{\sum_{\wp'_{st} \in \mathcal{P}_{st}} P_{\text{RW}}(\wp'_{st}) \exp(-\theta c(\wp'_{st}))} \quad (1)$$

where the total cost along a path $\wp$ is the sum of the individual edge costs $c(\wp_{st}) = c_{i_0 i_1} + c_{i_1 i_2} + \dots + c_{i_{\ell-1} i_\ell} = \sum_{\tau=1}^{\ell(\wp_{st})} c_{i_{(\tau-1)} i_\tau}$ and the natural, reference, random walk likelihood of the path is $P_{\text{RW}}(\wp_{st}) = \prod_{\tau=1}^{\ell(\wp_{st})} p_{i_{(\tau-1)} i_\tau}^{\text{RW}}$, usually with transition probabilities $p_{ij}^{\text{RW}} = a_{ij} / \sum_{j=1}^n a_{ij}$. In this expression, $\mathcal{P}_{st}$ is the set of all regular (non-absorbing, non-hitting, by default [20]) paths from node s to node t . Note that this probability distribution on the set of paths $\mathcal{P}_{st}$ can be derived from a maximal entropy argument. Moreover, for convenience, let us denote by $\wp_{st}(\tau)$ the index of the node at position $\tau \in [0, \ell]$ on path $\wp_{st}$. In our sequence defined above, it corresponds to $\wp_{st}(\tau) = i_\tau$. The transition probabilities matrix of the reference random walk on the graph $\mathbf{P}_{\text{RW}}$, containing the p_{ij}^{RW} , is a stochastic matrix.

This said, by defining $w_{ij} = p_{ij}^{\text{RW}} \exp(-\theta c_{ij})$ with inverse temperature parameter $\theta > 0$ in the RSP framework, the numerator of Eq. 1 can be rewritten as

$$\begin{aligned} P_{\text{RW}}(\wp_{st}) \exp(-\theta c(\wp_{st})) &= \prod_{\tau=1}^{\ell(\wp_{st})} (p_{i_{(\tau-1)} i_\tau}^{\text{RW}} \exp(-\theta c_{i_{(\tau-1)} i_\tau})) = \prod_{\tau=1}^{\ell(\wp_{st})} w_{i_{(\tau-1)} i_\tau} \\ &= w_{s i_1} w_{i_1 i_2} \dots w_{i_{(\ell-2)} i_{(\ell-1)}} w_{i_{(\ell-1)} t} \triangleq w(\wp_{st}) \end{aligned} \quad (2)$$

Because $\mathbf{P}_{\text{RW}}$ is a stochastic matrix and it is assumed that there is at least one edge cost which is strictly positive, $\mathbf{W} = (w_{ij}) = \mathbf{P}_{\text{RW}} \circ \exp(-\theta \mathbf{C})$ is sub-stochastic (all its row sums are lesser than or equal to 1 and there is at least one row sum strictly lesser than 1).

¹ If the graph G is undirected, each of its undirected edge $i \leftrightarrow j$ is considered as a set of two different directed edges going in opposite directions, $i \rightarrow j$ and $j \rightarrow i$.

Thus, this matrix represents a transition probability matrix between transient states, with a probability of giving up the walk in each node i provided by $1 - \sum_{j=1}^n w_{ij}$; we then say that the random walker is killed in i . In this context, Eq. 2 represents the likelihood of reaching (being killed in) the target node t through path φ_{st} when walking according to the sub-stochastic transition matrix $\mathbf{W}$ defining a killed random walk and starting from node s .

2.2 The Standard Bag-of-Paths Model

In a more general setting, let us now consider $\mathcal{P} = \cup_{i,j=1}^n \mathcal{P}_{ij}$ as the set of all paths in the network (a bag of paths, [14]). In other words, we now consider all possible paths between all pairs of nodes, instead of only paths connecting a predefined source-target pair of nodes s, t , in the RSP model. As already mentioned, this extension of the RSP is called the bag-of-paths model (BoP) considering all paths in the network [14].

Inspired from the RSP (see Eq. 2), in its basic form, the BoP model assumes that the probability of drawing a specific path $\varphi \in \mathcal{P}$ starting from some node i and reaching node j is *proportional to the likelihood of that path*,

$$\begin{aligned} P_{\text{BOP}}(\varphi \wedge \varphi(0) = i \wedge \varphi(\ell) = j) &\propto w_{i_0 i_1} w_{i_1 i_2} \cdots w_{i_{(\ell-2)} i_{(\ell-1)}} w_{i_{(\ell-1)} i_\ell} \delta_{i_0 i} \delta_{i_\ell j} \\ &= \delta(\varphi(0) = i) \delta(\varphi(\ell) = j) w(\varphi) \end{aligned} \quad (3)$$

where $\delta(\varphi(0) = i) = 1$ when the initial node $\varphi(0) = i$, and 0 otherwise.

In addition, we can also introduce *a priori weighting probabilities* of starting (in) in the node i and ending (out) in the node j [14]. In that situation, $\delta(\varphi(0) = i) \delta(\varphi(\ell) = j)$ is replaced by the product of the start and end probabilities, $\sigma_i^{\text{in}} \delta(\varphi(0) = i) \sigma_j^{\text{out}} \delta(\varphi(\ell) = j)$ with $\sum_{i=1}^n \sigma_i^{\text{in}} = \sum_{j=1}^n \sigma_j^{\text{out}} = 1$, providing

$$P_{\text{BOP}}(\varphi \wedge \varphi(0) = i \wedge \varphi(\ell) = j) \propto \sigma_i^{\text{in}} \delta(\varphi(0) = i) \sigma_j^{\text{out}} \delta(\varphi(\ell) = j) w(\varphi) \quad (4)$$

Note that this model does not imply $P_{\text{BOP}}(\varphi(0) = i) = \sigma_i^{\text{in}}$ nor $P_{\text{BOP}}(\varphi(\ell) = j) = \sigma_j^{\text{out}}$ – for such a model, see [16]. In order to be well-defined, the quantity (4) has to be properly normalized, $\sum_{\varphi \in \mathcal{P}} \sum_{i,j \in \mathcal{V}} P_{\text{BOP}}(\varphi \wedge \varphi(0) = i \wedge \varphi(\ell) = j) = 1$, which implies that the normalizing factor is $1 / \sum_{i,j=1}^n \sum_{\varphi \in \mathcal{P}_{ij}} \sigma_i^{\text{in}} \sigma_j^{\text{out}} w(\varphi)$, and thus

$$P_{\text{BOP}}(\varphi \wedge \varphi(0) = i \wedge \varphi(\ell) = j) = \begin{cases} \frac{\sigma_i^{\text{in}} \sigma_j^{\text{out}} w(\varphi)}{\sum_{i',j'=1}^n \sum_{\varphi' \in \mathcal{P}_{i'j'}} \sigma_{i'}^{\text{in}} \sigma_{j'}^{\text{out}} w(\varphi')} & \text{if } \varphi \in \mathcal{P}_{ij} \\ 0 & \text{else} \end{cases} \quad (5)$$

Let us now compute the a posteriori probability of drawing a path between i and j ,

$$P_{\text{BOP}}(\varphi \mid \varphi(0) = i \wedge \varphi(\ell) = j) = \frac{P_{\text{BOP}}(\varphi \wedge \varphi(0) = i \wedge \varphi(\ell) = j)}{P_{\text{BOP}}(\varphi(0) = i \wedge \varphi(\ell) = j)} \quad (6)$$

From Eq. 5 and after a few algebraic manipulations, the denominator is

$$P_{\text{BOP}}(\wp(0) = i \wedge \wp(\ell) = j) = \sum_{\wp \in \mathcal{P}} P_{\text{BOP}}(\wp \wedge \wp(0) = i \wedge \wp(\ell) = j) = \frac{\sigma_i^{\text{in}} \sigma_j^{\text{out}} \sum_{\wp \in \mathcal{P}_{ij}} w(\wp)}{\sum_{i', j'=1}^n \sum_{\wp' \in \mathcal{P}_{i'j'}} \sigma_{i'}^{\text{in}} \sigma_{j'}^{\text{out}} w(\wp')}$$

Injecting this last result in Eq. 6 and further using Eq. 5 provides

$$P_{\text{BOP}}(\wp \mid \wp(0) = i \wedge \wp(\ell) = j) = \begin{cases} \frac{w(\wp)}{\sum_{\wp' \in \mathcal{P}_{ij}} w(\wp')} & \text{if } \wp \in \mathcal{P}_{ij} \\ 0 & \text{else} \end{cases} \quad (7)$$

which is nothing more than Eq. 1. Therefore, quite naturally, the RSP probability can be interpreted in the BoP framework as the a posteriori probability of drawing a path $\wp$ given that we know that the source node is i and the target node j .

3 Bag-of-Paths with Weights on Path Lengths

We now come to the main contribution of the paper which introduces a probability mass weighting the length of the paths. For the sake of simplicity, we start without considering priors on the start and end nodes. Let us first introduce the likelihood of a path including a Poisson distribution assigned to its path length. Then, the joint probability of drawing a path starting in node i and ending in j , $P_{\text{BOP}}(\wp(0) = i \wedge \wp(\ell) = j)$, is derived. This is an important measure quantifying how close and how highly inter-connected the two nodes are. In addition, a distance measure between nodes is derived from these probabilities. Finally, the results are easily extended to models weighted with probabilities on start and end nodes.

3.1 A Poisson Distribution on path Lengths

Let us now consider that the length $\ell(\wp)$ of each path $\wp$ is weighted according to a Poisson distribution in such a way that the likelihood of a path is modulated by the probability of its length. More specifically, from Eq. 3, the likelihood of drawing a path $i \rightsquigarrow j$ of a given length τ is assumed to be proportional to

$$P_{\text{BOP}}(\wp \wedge \wp(0) = i \wedge \wp(\ell) = j \wedge \ell(\wp) = \tau) \propto \delta(\wp(0) = i) \delta(\wp(\ell) = j) \delta(\ell(\wp) = \tau) f(\tau, \lambda) w(\wp) \quad (8)$$

with $f(\cdot)$ being the Poisson distribution, providing the probability mass of drawing a path of length τ with parameter (mean length) λ , $f(\tau, \lambda) = \lambda^\tau \exp(-\lambda) / \tau!$.

As before, this last Eq. 8 should be properly normalized in order to sum to 1 for defining a probability distribution on the set of all paths $\mathcal{P}$. However, we

will leave the normalization for later. According to this model, the probability of drawing a path $\wp$ is thus proportional to

$$P_{\text{BOP}}(\wp) \propto \sum_{i,j \in \mathcal{V}} \sum_{\tau=0}^{\infty} \delta(\wp(0) = i) \delta(\wp(\ell) = j) \delta(\ell(\wp) = \tau) f(\tau, \lambda) w(\wp) = f(\ell(\wp), \lambda) w(\wp) \quad (9)$$

which is the model that will be investigated in this paper. This choice is mainly motivated by the fact that (i) the model integrates a “scale parameter” λ modulating the region of influence in the network based on path lengths, and (ii) the resulting quantities of interest can be computed in closed form and are related to a previous similarity measure, as will be shown soon. Again, this model does not imply $P_{\text{BOP}}(\ell(\wp) = \tau)$ because the length of a path is also affected by the path cost and the reference random walk probabilities. This is due to the fact that this simple model assumes the independence² between path likelihood and path length, which is not the case in general. Future work will be devoted to the design of a more realistic, generative, model which does not make this assumption; the present work can therefore be considered as a first step in this direction.

3.2 Computing the Joint Probability of Drawing a Path Starting in i and Ending in j

Let us now compute the interesting quantity $\pi_{ij} \triangleq P_{\text{BOP}}(\wp(0) = i \wedge \wp(\ell) = j)$, that is, the joint probability of observing a path starting in i and ending in j when drawing paths from the bag of paths $\mathcal{P}$. A similar measure was derived in [14] and used in [11] in order to develop a paths-based modularity measure, but without any weighting by path length. From Eq. 8, we find

$$\begin{aligned} P_{\text{BOP}}(\wp(0) = i \wedge \wp(\ell) = j) &\propto \sum_{\wp \in \mathcal{P}} \sum_{\tau=0}^{\infty} \delta(\wp(0) = i) \delta(\wp(\ell) = j) \delta(\ell(\wp) = \tau) f(\tau, \lambda) w(\wp) \\ &= \sum_{\wp \in \mathcal{P}} \delta(\wp(0) = i) \delta(\wp(\ell) = j) f(\ell(\wp), \lambda) w(\wp) = \sum_{\wp \in \mathcal{P}_{ij}} f(\ell(\wp), \lambda) w(\wp) = \sum_{\tau=0}^{\infty} \sum_{\wp \in \mathcal{P}_{ij}^{(\tau)}} f(\tau, \lambda) w(\wp) \\ &= \sum_{\tau=0}^{\infty} f(\tau, \lambda) \sum_{\wp \in \mathcal{P}_{ij}^{(\tau)}} w(\wp) = \sum_{\tau=0}^{\infty} f(\tau, \lambda) [\mathbf{W}^{\tau}]_{ij} = \exp(-\lambda) \left[\sum_{\tau=0}^{\infty} \frac{\lambda^{\tau} \mathbf{W}^{\tau}}{\tau!} \right]_{ij} \propto [\expm(\lambda \mathbf{W})]_{ij} \end{aligned} \quad (10)$$

where $\expm$ is the matrix exponential and $\mathcal{P}_{ij}^{(\tau)}$ is the set of all paths connecting node i to node j in *exactly* τ steps. Remind also that $\sum_{\wp \in \mathcal{P}_{ij}^{(\tau)}} w(\wp) = [\mathbf{W}^{\tau}]_{ij}$ with $\mathbf{W} = \mathbf{P}_{\text{RW}} \circ \exp(-\theta \mathbf{C})$ in the standard BoP framework [14]. From this last equation, we easily deduce, after normalization,

$$\pi_{ij}(\lambda) = P_{\text{BOP}}(\wp(0) = i \wedge \wp(\ell) = j) = \frac{[\expm(\lambda \mathbf{W})]_{ij}}{\sum_{i',j'=1}^n [\expm(\lambda \mathbf{W})]_{i'j'}} \quad (11)$$

² As, for instance, a simple bag-of-words model in information retrieval or a naive Bayes classification model in data science.

These probabilities quantify to which extend the two nodes i and j are easily accessible in the network by taking into account both the proximity in the network (the two nodes are close) and their amount of inter-connectivity (many paths connecting the two nodes). This quantity is closely related to the *communicability measure* developed in [12]. The distance measure derived from this probability without weighting by path length in [14] was called the *surprisal distance* and provided competitive results on pattern recognition tasks [7, 14, 15, 23]. We will use this surprisal distance as a baseline in our experiments in Sect. 4. In matrix form, we have

$$\mathbf{\Pi}(\lambda) = \frac{\expm(\lambda \mathbf{W})}{\mathbf{e}^T \expm(\lambda \mathbf{W}) \mathbf{e}} \quad (12)$$

where $\mathbf{e}$ is a column vector full of 1's.

3.3 A Derived Distance Measure Between Nodes

A common procedure for deriving a dissimilarity from a probability mass is to take minus the logarithm of this probability. The resulting distance matrix between nodes, called the directed³ *Poisson weighted surprisal distance*, is consequently defined as

$$\mathbf{\Delta} = -\log \mathbf{\Pi}(\lambda) \quad (13)$$

with the log being the natural elementwise logarithm. It computes the *surprisal* of observing a path starting in i and ending in j , and is an extension of the corresponding measure introduced in [14] in the standard bag-of-paths framework. The resulting distance measure is the symmetrized quantity, $\frac{1}{2}(\mathbf{\Delta} + \mathbf{\Delta}^T)$. This new distance is closely related to the *log-communicability measure* introduced in [17]. The log-communicability measure is the logarithmic version of the well-known communicability measure [12] and is defined as $\mathbf{K} = \log(\expm(\alpha \mathbf{A}))$ where $\mathbf{A}$ is the adjacency matrix of G and α is a discounting parameter favoring shorter paths by giving them a larger weight. The main difference between these two measures is that the matrix exponential is taken on the adjacency matrix $\mathbf{A}$ in place of the $\mathbf{W}$ matrix in Eqs. 12 and 13. This observation therefore provides an interpretation to the log-communicability measure in the BoP framework, where the parameter α corresponds to the parameter λ .

3.4 Weighting Start and end Nodes

Considering now prior weighting factors σ_i^{in} , σ_j^{out} (which are assumed to be known) on the start and end nodes and applying the same reasoning as in previous section (see Eq. 10) provides

$$\begin{aligned} P_{\text{BoP}}(\varphi(0)=i \wedge \varphi(\ell)=j) &\propto \sum_{\varphi \in \mathcal{P}} \sum_{\tau=0}^{\infty} \sigma_i^{\text{in}} \delta(\varphi(0)=i) \sigma_j^{\text{out}} \delta(\varphi(\ell)=j) \delta(\ell(\varphi)=\tau) f(\tau, \lambda) w(\varphi) \\ &= \sigma_i^{\text{in}} \sigma_j^{\text{out}} \sum_{\varphi \in \mathcal{P}_{ij}} f(\ell(\varphi), \lambda) w(\varphi) \propto \sigma_i^{\text{in}} \sigma_j^{\text{out}} [\expm(\lambda \mathbf{W})]_{ij} \quad (14) \end{aligned}$$

³ In general, this matrix is non-symmetric.

In matrix form, this leads to

$$\Pi(\lambda) = \frac{\text{Diag}(\sigma_{\text{in}}) \expm(\lambda \mathbf{W}) \text{Diag}(\sigma_{\text{out}})}{\sigma_{\text{in}}^T \expm(\lambda \mathbf{W}) \sigma_{\text{out}}} \quad (15)$$

where σ_{in} and σ_{out} are column vectors containing the weights on the nodes and **Diag** is an operator which, when applied to a vector, defines a diagonal matrix with this vector as diagonal. In the experimental section, we will consider that the priors on start and end nodes are equal, $\sigma_{\text{in}} = \sigma_{\text{out}} = \sigma$. The distance is defined in the same way as before through Eq. 13. We will refer to this distance as the *Poisson weighted surprisal distance with priors*. Note that when the probabilities are uniform, $\sigma = \mathbf{e}/n$ where n is the number of nodes, the distance is equivalent (up to a scaling factor) to the *Poisson weighted surprisal distance* defined in the previous subsection (see Eq. 12).

4 Experiments

In this section, we will compare the performances, in terms of classification accuracy on a graph-based semi-supervised classification task, of the Poisson weighted surprisal distance and the Poisson weighted surprisal distance with priors on start/end nodes introduced in Subsects. 3.3 and 3.4 with other state-of-the-art methods. Notice that the compared methods, the experimental design, and the investigated datasets are very similar to those used in [7], so that our description closely follows this work.

4.1 Datasets

As already mentioned, we use the same collection of 14 network datasets as [7]; see Table 1 of this paper for details. This collection is composed of 9 subsets extracted from the 20 Newsgroup datasets [25, 32], 4 WebKB datasets [26] and the IMDB dataset [26]. All the networks are undirected and defined by an adjacency matrix $\mathbf{A}$ with a transition cost c_{ij} equal to the inverse of the affinity, $1/a_{ij}$, as for electrical networks [14]. For each dataset, a class label is associated to each node of G .

4.2 Investigated State-of-the-Art Methods

As part of our experiment, 6 dissimilarity measures and 3 kernel matrices are selected as baseline methods to assess the performances of our introduced methods. These selected methods already showed good performances in previous semi-supervised [7, 14, 18, 23] and unsupervised tasks [8, 17, 30, 32].

Among these 11 methods, seven are defined in the RSP or the BoP frameworks and therefore share a common parameter θ . This parameter allows to define dissimilarity measures which interpolate for most of them between the shortest paths distance when θ is high and the commute time distance (up to a scaling factor) when θ is close to zero. These seven methods are:

- The *Poisson weighted surprisal distance* (**PSurp**) is the distance introduced in Subsect. 3.3. This method has one extra parameter λ , which is the parameter of the Poisson distribution that needs to be tuned on the network.
- The *Poisson weighted surprisal distance with priors* is the distance introduced in Subsect. 3.4. This method has two additional parameters which are the prior vector σ and the parameter of the Poisson distribution λ . We tested two different types of prior vectors: the L1-normalized degree weights and the L1-normalized inverse degree weights defined as
 - *L1-normalized degree weights*: $\mathbf{d} \div \text{sum}(\mathbf{d})$ with $\mathbf{d} = \mathbf{A}^T \mathbf{e}$;
 - *L1-normalized inverse degree weights*: $(\mathbf{e} \div \mathbf{d}) \div \text{sum}(\mathbf{e} \div \mathbf{d})$.
 where $\div$ is the elementwise division. We refer to these methods as **PSurpDegree** and **PSurpInvDegree**.
- The *margin-constrained bag-of-paths surprisal distance* (**cBoP**) and the *margin-constrained bag-of-hitting-paths surprisal distance* (**cBoPH**) are two distances introduced in [16], and further studied in [7]. These methods have one additional parameter which is the positive weights vector σ . It has been shown in [7] that the L1-normalized degree weighting provides the best results in our current setup for both methods.
- The *free energy distance* (**FE**), the *randomized shortest paths dissimilarity* (**RSP**) and the *regular surprisal distance* (**Surp**) are the three dissimilarity measures developed in the origins of the RSP and the BoP frameworks [14, 20, 27]. These methods have no extra parameter.

The four other selected methods are:

- The *sigmoid commute time kernel* (**SCT**) and the *sigmoid corrected commute time kernel* (**SCCT**) [32], are obtained respectively by applying an elementwise sigmoid transformation [28] on the commute-time kernel [13] and on the correction of the commute-time kernel proposed in [31]. These methods have one parameter α which controls the sharpness of the sigmoid.
- The *logarithmic forest distance* (**LF**) is the distance proposed in [5] built on the matrix forest theorem [4]. This method has one parameter α which allows to interpolate up to a scaling factor between the shortest paths distance ($\alpha \rightarrow 0^+$) and the resistance distance ($\alpha \rightarrow +\infty$) [21].
- The *logarithmic communicability* (**logCom**) is the measure introduced in [17] corresponding to the logarithmic version of the communicability kernel⁴ [12] presented in Subsect. 3.3. Indeed, the authors of [17] observed that taking the elementwise logarithm of the communicability kernel improves significantly the results⁵. This method has one positive discounting parameter α favoring the paths between two nodes with a short length by giving them a larger weight.

A summary of all the investigated methods with their acronyms and the values tested for each parameter is available in Table 1.

⁴ The communicability kernel is also related to the exponential diffusion kernel [22].

⁵ The communicability kernel has also been investigated, but we have not included its results as these are worse compared to its logarithmic version.

4.3 Experimental Design

The supervised node classification strategy consists in transforming all the dissimilarity and similarity measures into kernel matrices $\mathbf{K}$ by removing all negative eigenvalues and then feeding them into a kernel SVM⁶ with several margin parameter values $c = \{10^{-2}, 10^{-1}, 1, 10, 100\}$. For each dissimilarity measure, three transformations are investigated: the classical multidimensional scaling (MDS)⁷ [2], the Gaussian transformation (Gauss) [28] and the centered Gaussian transformation (GaussCenter)⁸. For conciseness, we only present the results associated to the best kernel transformation according to Nemenyi tests [10] for each method (see Table 2 for details).

Table 1. Parameter ranges for the investigated methods.

Method acronym	Parameters	Parameters values
PSurp	θ, λ	$\theta = (0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10)$ $\lambda = (1, 2, 3, 5, 10)$
PSurpDegree	θ, λ	
PSurpInvDegree	θ, λ	
CBoP	θ	$\theta = (0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10)$
CBoPH	θ	
FE	θ	
RSP	θ	
Surp	θ	
SCT	α	
SCCT	α	$\alpha = (5, 10, 15, 20, 25, 30, 35, 40, 45, 50)$
LF	α	
logCom	α	

Moreover, to avoid large variability in the results, we assess all the methods by performing 10 runs of a 5×5 nested cross-validation methodology with different folds of labeled/unlabeled nodes on each run. In each external 5-folds cross-validation, 20% of the labels are used for training the model and the remaining 80% are hidden to test the performance. In each internal 5-folds cross-validation, the parameters are tuned on the external training fold by keeping 80% of the labeled nodes. Furthermore, both external and internal folds are kept identical for all methods inside a given run. Finally, the results reported in Table 2 are the average of the 50 accuracy scores of the external cross-validation folds.

4.4 Results and discussion

As already mentioned, the results of our semi-supervised classification experiments are reported in Table 2. For clarity, we highlighted the highest accuracy

⁶ We use the LIBSVM library [3] with the options ‘-s 0’ and ‘-t 4’.

⁷ Note that the kernels obtained through this transformation are already centered.

⁸ We apply the centering transformation $\mathbf{K} = \mathbf{H}\mathbf{K}\mathbf{H}$ to the Gaussian kernel where $\mathbf{H} = \mathbf{I} - \mathbf{e}\mathbf{e}^T/n$ is the centering matrix, $\mathbf{e}$ is a column vector full of 1’s, $\mathbf{I}$ is the identity matrix and n is the number of nodes.

Table 2. Classification accuracy in percent $\pm$ standard deviation for the various classification methods obtained on the different datasets. For each dataset and method, the final accuracy is obtained by following the experimental design detailed in Subsect. 4.3. The best performing method is highlighted in boldface for each dataset.

Kernel $\rightarrow$ Dataset $\downarrow$	PSurp Gauss	PSurpDegree Gauss	PSurpInvDegree MDS	cBoP Gauss/Center	cBoPH Gauss	FE Gauss	RSP Gauss	Surp Gauss	SCT	SCCT	LF Gauss	logCom
WebKB-Cornell	59.56 $\pm$ 2.61	59.61 $\pm$ 2.59	58.53 $\pm$ 2.84	59.45 $\pm$ 2.99	58.79 $\pm$ 3.38	58.51 $\pm$ 2.81	58.30 $\pm$ 3.76	58.71 $\pm$ 2.75	58.92 $\pm$ 2.89	59.13 $\pm$ 2.94	59.96 $\pm$ 2.65	58.85 $\pm$ 2.88
WebKB-Texas	78.44 $\pm$ 3.14	78.55 $\pm$ 3.13	76.80 $\pm$ 3.27	78.87 $\pm$ 2.81	76.68 $\pm$ 3.26	77.11 $\pm$ 3.17	76.88 $\pm$ 2.93	76.91 $\pm$ 3.05	76.54 $\pm$ 3.49	76.80 $\pm$ 3.01	79.46 $\pm$ 2.72	79.04 $\pm$ 2.47
WebKB-Washington	70.42 $\pm$ 2.02	71.85 $\pm$ 2.04	70.16 $\pm$ 1.88	71.84 $\pm$ 2.04	70.12 $\pm$ 2.19	69.46 $\pm$ 2.02	69.78 $\pm$ 3.14	69.05 $\pm$ 2.20	69.49 $\pm$ 2.49	68.66 $\pm$ 2.18	71.87 $\pm$ 1.99	71.35 $\pm$ 2.41
WebKB-Wisconsin	77.08 $\pm$ 2.46	77.95 $\pm$ 2.14	75.68 $\pm$ 2.44	78.35 $\pm$ 1.97	76.42 $\pm$ 2.42	77.26 $\pm$ 2.24	75.69 $\pm$ 2.32	76.25 $\pm$ 2.32	77.70 $\pm$ 2.29	77.64 $\pm$ 2.03	79.48 $\pm$ 1.95	78.44 $\pm$ 2.26
IMDB	77.52 $\pm$ 1.47	77.95 $\pm$ 1.33	76.78 $\pm$ 1.58	77.86 $\pm$ 1.57	79.24 $\pm$ 1.48	78.95 $\pm$ 1.51	78.99 $\pm$ 1.21	79.58 $\pm$ 1.52	78.03 $\pm$ 1.65	77.41 $\pm$ 1.43	79.19 $\pm$ 1.36	77.29 $\pm$ 1.64
NewsGroup-2cl-1	96.36 $\pm$ 1.40	96.13 $\pm$ 1.12	96.83 $\pm$ 1.50	94.98 $\pm$ 1.28	95.63 $\pm$ 0.88	95.40 $\pm$ 1.21	95.66 $\pm$ 1.50	95.45 $\pm$ 0.88	95.84 $\pm$ 1.03	97.07 $\pm$ 0.67	94.95 $\pm$ 1.72	95.80 $\pm$ 1.31
NewsGroup-2cl-2	92.50 $\pm$ 1.82	92.41 $\pm$ 1.78	92.69 $\pm$ 1.44	91.56 $\pm$ 1.54	92.59 $\pm$ 1.60	92.27 $\pm$ 1.61	92.04 $\pm$ 1.60	92.32 $\pm$ 1.97	92.22 $\pm$ 1.41	92.50 $\pm$ 1.29	92.27 $\pm$ 1.54	92.32 $\pm$ 1.55
NewsGroup-2cl-3	96.48 $\pm$ 1.01	96.44 $\pm$ 0.99	96.61 $\pm$ 1.05	96.34 $\pm$ 1.06	96.62 $\pm$ 0.74	96.62 $\pm$ 0.78	96.47 $\pm$ 0.96	96.72 $\pm$ 0.84	96.00 $\pm$ 1.00	96.06 $\pm$ 0.75	95.53 $\pm$ 1.56	95.27 $\pm$ 1.59
NewsGroup-3cl-1	93.39 $\pm$ 1.12	93.56 $\pm$ 1.06	93.23 $\pm$ 1.29	92.85 $\pm$ 1.25	93.33 $\pm$ 1.00	93.24 $\pm$ 0.92	92.80 $\pm$ 1.32	93.29 $\pm$ 1.02	93.43 $\pm$ 1.02	93.94 $\pm$ 0.63	92.70 $\pm$ 1.13	93.28 $\pm$ 1.11
NewsGroup-3cl-2	93.32 $\pm$ 1.11	93.23 $\pm$ 1.05	93.56 $\pm$ 1.28	92.56 $\pm$ 1.14	93.23 $\pm$ 1.00	92.84 $\pm$ 1.05	92.65 $\pm$ 1.16	93.26 $\pm$ 0.97	92.25 $\pm$ 1.15	93.36 $\pm$ 0.85	92.35 $\pm$ 1.33	92.36 $\pm$ 1.00
NewsGroup-3cl-3	92.78 $\pm$ 1.35	93.28 $\pm$ 1.32	92.70 $\pm$ 1.22	92.30 $\pm$ 1.16	93.26 $\pm$ 1.02	92.78 $\pm$ 1.16	92.42 $\pm$ 1.27	92.98 $\pm$ 1.21	92.64 $\pm$ 1.12	93.11 $\pm$ 0.78	92.36 $\pm$ 1.43	91.61 $\pm$ 1.19
NewsGroup-5cl-1	89.37 $\pm$ 1.02	89.25 $\pm$ 1.07	89.39 $\pm$ 1.52	88.54 $\pm$ 1.24	88.94 $\pm$ 0.89	88.78 $\pm$ 1.02	88.57 $\pm$ 1.03	88.87 $\pm$ 0.96	86.96 $\pm$ 1.15	88.00 $\pm$ 1.03	87.95 $\pm$ 1.17	87.77 $\pm$ 1.46
NewsGroup-5cl-2	84.38 $\pm$ 1.42	84.46 $\pm$ 1.52	84.57 $\pm$ 1.51	82.73 $\pm$ 1.47	83.94 $\pm$ 1.32	83.26 $\pm$ 1.02	82.66 $\pm$ 1.64	83.70 $\pm$ 1.27	81.86 $\pm$ 1.59	82.96 $\pm$ 0.85	82.69 $\pm$ 1.84	83.16 $\pm$ 1.51
NewsGroup-5cl-3	83.68 $\pm$ 1.91	84.15 $\pm$ 1.51	83.93 $\pm$ 1.50	82.28 $\pm$ 1.61	83.54 $\pm$ 1.32	83.17 $\pm$ 1.17	82.43 $\pm$ 1.75	83.52 $\pm$ 1.20	82.17 $\pm$ 1.77	83.56 $\pm$ 1.18	82.50 $\pm$ 1.72	82.73 $\pm$ 1.79

in boldface for each dataset. From Table 2, we can observe that the LF outperforms all the methods on the WebKB datasets and that the Surp performs best on the IMDB dataset. Regarding the Newsgroups datasets, the results are more nuanced. More precisely, the PSurpDegree provides the best results on 3 of the Newsgroups datasets, the PSurpDegree and the SCCT on 2, and the Surp on 1. Overall, we can state that the best-performing method is dataset-dependent.

Therefore, in order to analyze the results more thoroughly, we perform a nonparametric Friedman-Nemenyi statistical test followed by multiple Wilcoxon signed-rank tests at a level of confidence of 95% ($\alpha = 0.05$) [10]. The p -value of the Friedman test is 8.8×10^{-4} which is inferior to our α meaning that the results of at least one method are significantly different from the others. As the Friedman test is positive, we performed the Nemenyi test and reported the results in Fig. 1. The Nemenyi test shows that the PSurpDegree outperforms significantly the RSP and the SCT only.

To refine our analysis and potentially provide more information about the relative performances of the methods, we performed multiple Wilcoxon signed-rank tests and reported the results in Table 3. As regards the PSurpDegree, the Wilcoxon tests show that it obtains results significantly superior to those of all the methods except the PSurp, the PSurpInvDegree, and the LF. The tests also highlight that the RSP is outperformed by the PSurp, the PSurpInvDegree, the cBoPH, the FE, and the Surp. The SCT, for its part, obtains results significantly inferior to the PSurp, the cBoPH, and the SCCT. Finally, the tests indicate that the PSurp performs significantly better than the FE.

Globally, the experiments showed that the two variants of the introduced distance achieve good results on our experimental setup and the investigated datasets compared to the other baselines, especially the PSurpDegree. Moreover, as already mentioned, the PSurpDegree obtains results significantly superior to the regular Surp and FE which is a good achievement as these distances performed quite well in previous semi-supervised [7, 14, 23] and unsupervised tasks [8, 29, 30, 32].

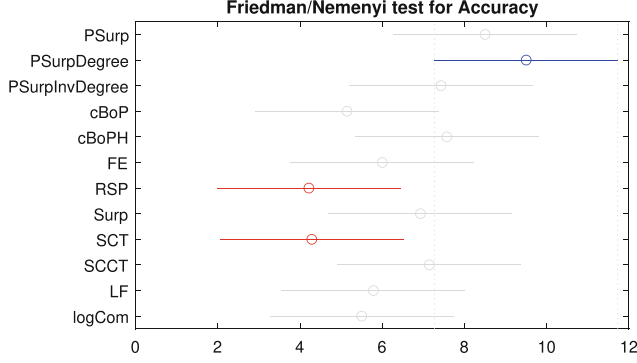


Fig. 1. Mean ranks and 95% Nemenyi confidence intervals for the 12 methods across the 14 datasets. Two methods are considered as significantly different if their confidence intervals do not overlap. The x-axis unit is the average rank of the methods; the higher the rank, the best the method. The best method (PSurpDegree) and the worse methods (RSP and SCT) overall are highlighted.

Table 3. The p -values provided by a pairwise Wilcoxon signed-rank test for the results presented in Table 2. The p -values inferior to 0.05 are highlighted.

Kernel → Kernel ↓	PSurp	PSurpDegree	PSurpInvDegree	cBoP	cBoPH	FE	RSP	Surp	SCT	SCCT	LF	logCom
PSurp	1.0000	0.1040	0.4631	0.1726	0.1531	0.0494	0.0085	0.0906	0.0134	0.2958	0.7148	0.0906
PSurpDegree	0.1040	1.0000	0.1040	0.0067	0.0295	0.0085	0.0040	0.0203	0.0002	0.0353	0.2166	0.0040
PSurpInvDegree	0.4631	0.1040	1.0000	0.8552	0.7609	0.4263	0.0245	0.4263	0.1189	1.0000	0.9515	0.5016
cBoP	0.1726	0.0067	0.8552	1.0000	0.6257	0.8552	0.9032	0.7148	0.2412	0.7148	0.3575	0.5416
cBoPH	0.1531	0.0295	0.7609	0.6257	1.0000	0.0906	0.0009	0.1726	0.0295	0.6698	0.8552	0.5016
FE	0.0494	0.0085	0.4263	0.8552	0.0906	1.0000	0.0295	0.2166	0.0676	0.8552	1.0000	0.8552
RSP	0.0085	0.0040	0.0245	0.9032	0.0009	0.0295	1.0000	0.0134	0.4263	0.1937	0.4631	0.5416
Surp	0.0906	0.0203	0.4263	0.7148	0.1726	0.2166	0.0134	1.0000	0.1531	1.0000	0.9032	0.6698
SCT	0.0134	0.0002	0.1189	0.2412	0.0295	0.0676	0.4263	0.1531	1.0000	0.0353	0.0580	0.2958
SCCT	0.2958	0.0353	1.0000	0.7148	0.6698	0.8552	0.1937	1.0000	0.0353	1.0000	0.9515	0.3910
LF	0.7148	0.2166	0.9515	0.3575	0.8552	1.0000	0.4631	0.9032	0.0580	0.9515	1.0000	0.2676
LogCom	0.0906	0.0040	0.5016	0.5416	0.5016	0.8552	0.5416	0.6698	0.2958	0.3910	0.2676	1.0000

5 Conclusion and further Work

This paper presented a new bag-of-paths model where the lengths of the paths are weighted according to a Poisson probability distribution. From this model, a new distance measure between nodes of the graph was introduced, the Poisson weighted surprisal distance. Moreover, a new interpretation of the logarithmic communicability kernel inspired by this new framework was proposed.

Experiments on semi-supervised classification tasks showed that the introduced distance is competitive with the other state-of-the-art methods. More specifically, the experiments also highlight that the introduction of node degree weighting in the Poisson weighted surprisal distance with priors provides results significantly superior to the ones obtained by its counterpart defined in the standard bag-of-paths framework, the regular surprisal distance.

One considered further work is to derive the randomized shortest path dissimilarity [20] and the free energy distance [14] within this new formulation. Another one aims at developing a more realistic generative model of paths alleviating the assumption of independence between path likelihood and path length. Still another one aims at investigating a generalized Poisson distribution (see for instance [6]) on the path lengths which avoids the drawback of having a mean equal to the variance, as the Poisson distribution does. Finally, additional experiments on pattern recognition tasks, such as clustering, collaborative recommendation, and link prediction, will be performed in order to try to confirm our findings.

References

1. Akamatsu, T.: Cyclic flows, Markov process and stochastic traffic assignment. *Transp. Res. B* **30**(5), 369–386 (1996)
2. Borg, I., Groenen, P.J.F.: *Modern Multidimensional Scaling*. SSS, Springer, New York (2005). <https://doi.org/10.1007/0-387-28981-X>
3. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011)
4. Chebotarev, P., Shamis, E.: The matrix-forest theorem and measuring relations in small social groups. *Autom. Remote Control* **58**(9), 1505–1514 (1997)
5. Chebotarev, P.: A class of graph-geodesic distances generalizing the shortest-path and the resistance distances. *Discret. Appl. Math.* **159**(5), 295–302 (2011)
6. Consul, P.C.: *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker (1989)
7. Courtain, S., Guex, G., Kivimäki, I., Saelens, M.: Relative entropy-regularized optimal transport on a graph: a new algorithm and an experimental comparison. ArXiv preprint [arXiv:0912.0238v9](https://arxiv.org/abs/0912.0238v9) (2021)
8. Courtain, S., Leleux, P., Kivimäki, I., Guex, G., Saelens, M.: Randomized shortest paths with net flows and capacity constraints. *Inf. Sci.* **556**, 341–360 (2020)
9. Courtain, S., Lebichot, B., Kivimäki, I., Saelens, M.: Graph-based fraud detection with the free energy distance. In: Cherifi, H., Gaito, S., Mendes, J.F., Moro, E., Rocha, L.M. (eds.) *COMPLEX NETWORKS 2019*. SCI, vol. 882, pp. 40–52. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-36683-4_4
10. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006)
11. Devooght, R., Mantrach, A., Kivimäki, I., Bersini, H., Jaimes, A., Saelens, M.: Random walks based modularity: application to semi-supervised learning. In: *Proceedings of the 23rd International World Wide Web Conference (WWW 2014)*, pp. 213–224 (2014)
12. Estrada, E., Hatano, N.: Communicability in complex networks. *Phys. Rev. E* **77**(3), 036111 (2008)
13. Fouss, F., Pirotte, A., Renders, J.M., Saelens, M.: Random-walk computation of similarities between nodes of a graph, with application to collaborative recommendation. *IEEE Trans. Knowl. Data Eng.* **19**(3), 355–369 (2007)
14. Francoisse, K., Kivimäki, I., Mantrach, A., Rossi, F., Saelens, M.: A bag-of-paths framework for network data analysis. *Neural Netw.* **90**, 90–111 (2017)
15. Guex, G., Courtain, S., Saelens, M.: Covariance and correlation Kernels on a graph in the generalized bag-of-paths formalism. *J. Complex Netw.* **8**(6), 1–46 (2021)

16. Guex, G., Kivimäki, I., Saerens, M.: Randomized optimal transport on a graph: framework and new distance measures. *Netw. Sci.* **7**(1), 88–122 (2019)
17. Ivashkin, V., Chebotarev, P.: Do logarithmic proximity measures outperform plain ones in graph clustering? In: Kalyagin, V.A., Nikolaev, A.I., Pardalos, P.M., Prokopyev, O.A. (eds.) *NET 2016. SPMS*, vol. 197, pp. 87–105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56829-4_8
18. Ivashkin, V., Chebotarev, P.: Dissecting graph measures performance for node clustering in LFR parameter space (2021). Preprint available at <https://openreview.net/forum?id=HkUfnZFt1Rw>
19. Kivimäki, I., Lebichot, B., Saramäki, J., Saerens, M.: Two betweenness centrality measures based on randomized shortest paths. *Sci. Rep.* **6**(1), 1–15 (2016)
20. Kivimäki, I., Shimbo, M., Saerens, M.: Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A* **393**, 600–616 (2014)
21. Klein, D.J., Randic, M.: Resistance distance. *J. Math. Chem.* **12**(1), 81–95 (1993)
22. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete structures. In: *Proceedings of the 19th International Conference on Machine Learning (ICML 2002)*, pp. 315–322 (2002)
23. Leleux, P., Courtain, S., Guex, G., Saerens, M.: Sparse randomized shortest paths routing with Tsallis divergence regularization. *Data Min. Knowl. Disc.* **35**(3), 986–1031 (2021). <https://doi.org/10.1007/s10618-021-00742-y>
24. Leleux, P., Courtain, S., François, K., Saerens, M.: Design of biased random walks on a graph with application to collaborative recommendation. Submitted for publication (2020)
25. Lichman, M.: UCI machine learning repository (2013)
26. Macskassy, S.A., Provost, F.: Classification in networked data: a toolkit and a univariate case study. *J. Mach. Learn. Res.* **8**, 935–983 (2007)
27. Saerens, M., Achbany, Y., Fouss, F., Yen, L.: Randomized shortest-path problems: two related models. *Neural Comput.* **21**(8), 2363–2404 (2009)
28. Schölkopf, B., Smola, A.: *Learning with Kernels*. MIT Press, Cambridge (2002)
29. Sommer, F., Fouss, F., Saerens, M.: Comparison of graph node distances on clustering tasks. In: Villa, A.E.P., Masulli, P., Pons Rivero, A.J. (eds.) *ICANN 2016. LNCS*, vol. 9886, pp. 192–201. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44778-0_23
30. Sommer, F., Fouss, F., Saerens, M.: Modularity-driven Kernel k -means for community detection. In: Lintas, A., Rovetta, S., Verschure, P.F.M.J., Villa, A.E.P. (eds.) *ICANN 2017. LNCS*, vol. 10614, pp. 423–433. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68612-7_48
31. von Luxburg, U., Radl, A., Hein, M.: Getting lost in space: Large sample analysis of the commute distance. In: *Advances in Neural Information Processing Systems: Proceedings of the NIPS 2010 Conference*, vol. 23, pp. 2622–2630 (2010)
32. Yen, L., Fouss, F., Decaestecker, C., Francq, P., Saerens, M.: Graph nodes clustering with the sigmoid commute-time kernel: a comparative study. *Data Knowl. Eng.* **68**(3), 338–361 (2009)
33. Yen, L., Mantrach, A., Shimbo, M., Saerens, M.: A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, pp. 785–793 (2008)