

## Similarities on graphs: Kernels versus proximity measures

Konstantin Avrachenkov, Pavel Chebotarev, Dmytro Rubanov

► **To cite this version:**

Konstantin Avrachenkov, Pavel Chebotarev, Dmytro Rubanov. Similarities on graphs: Kernels versus proximity measures. *European Journal of Combinatorics*, Elsevier, 2019, 80, pp.47-56. 10.1016/j.ejc.2018.02.002 . hal-02413478

**HAL Id: hal-02413478**

**<https://hal.inria.fr/hal-02413478>**

Submitted on 16 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Similarities on Graphs: Kernels versus Proximity Measures<sup>1</sup>

Konstantin Avrachenkov<sup>a</sup>, Pavel Chebotarev<sup>b,\*</sup>, Dmytro Rubanov<sup>a</sup>

<sup>a</sup>*Inria Sophia Antipolis, France*

<sup>b</sup>*Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Russia*

---

## Abstract

We analytically study proximity and distance properties of various kernels and similarity measures on graphs. This helps to understand the mathematical nature of such measures and can potentially be useful for recommending the adoption of specific similarity measures in data analysis.

---

## 1. Introduction

Until the 1960s, mathematicians studied only one distance for graph vertices, the shortest path distance [6]. In 1967, Gerald Sharpe proposed the electric distance [42]; then it was rediscovered several times. For some collection of graph distances, we refer to [19, Chapter 15].

Distances<sup>2</sup> are treated as dissimilarity measures. In contrast, similarity measures are maximized when every distance is equal to zero, i.e., when two arguments of the function coincide. At the same time, there is a close relation between distances and certain classes of similarity measures.

One of such classes consists of functions defined with the help of kernels on graphs, i.e., positive semidefinite matrices with indices corresponding to the nodes. Every kernel is the Gram matrix of some set of vectors in a Euclidean space, and Schoenberg's theorem [39, 40] shows how it can be transformed into the matrix of Euclidean distances between these vectors.

Another class consists of proximity measures characterized by the triangle inequality for proximities. Every proximity measure  $\kappa(x, y)$  generates a

---

\*Corresponding author

*Email addresses:* [k.avrachenkov@inria.fr](mailto:k.avrachenkov@inria.fr) (Konstantin Avrachenkov), [pavel14e@gmail.com](mailto:pavel14e@gmail.com) (Pavel Chebotarev), [dmytro.rubanov@inria.fr](mailto:dmytro.rubanov@inria.fr) (Dmytro Rubanov)

<sup>1</sup>This is the authors' version of a work that was submitted to Elsevier.

<sup>2</sup>In this paper, the term distance is used as a synonym of metric, i.e., every distance satisfies nonnegativity, symmetry, the identity of indiscernibles, and the triangle inequality.

distance [14] by means of the  $d(x, y) = \frac{1}{2}(\kappa(x, x) + \kappa(y, y)) - \kappa(x, y)$  transformation, and for some subclasses of proximity measures, the inverse transformation can also be specified.

Furthermore, it turns out that many similarity measures are transitional measures [8, 11], in particular, they satisfy the inequality  $s_{ij} s_{jk} \leq s_{ik} s_{jj}$  and so they can be transformed into proximities by means of the logarithmic transformation.

Distances and similarity measures on graphs are widely used in data analysis, especially, in graph-based supervised, semi-supervised, and unsupervised machine learning, see, e.g., [1, 2, 3, 5, 7, 21, 25, 26, 35, 36, 37, 47] and references therein. In a number of studies including [28, 33, 45], comparative ability of various measures to detect communities and predict links has been explored. However, in such studies, the authors do not focus on the mathematical properties of the measures under consideration.

The purpose of this paper is to start filling this gap. We consider a number of well-known and recently proposed similarity measures on graphs (including weighted graphs) defined in terms of one of the following basic matrices: the weighted adjacency matrix, the Laplacian matrix, and the (stochastic) Markov matrix. We explore their basic mathematical properties, in particular, we find out whether they belong to the classes of kernels or proximities and study the properties of distances related to them. This helps to reveal the nature of such measures and can be considered as a step towards a mathematical theory of similarity / dissimilarity measures on graphs.

## 2. Definitions and preliminaries

The *weighted adjacency matrix*  $W = (w_{ij})$  of a weighted undirected graph  $G$  with vertex set  $V(G) = \{1, \dots, n\}$  is the matrix with elements

$$w_{ij} = \begin{cases} \text{weight of edge } (i, j), & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

In what follows,  $G$  is connected.

The ordinary (or combinatorial) *Laplacian matrix*  $L$  of  $G$  is defined as follows:  $L = D - W$ , where  $D = \text{Diag}(W \cdot \mathbf{1})$  is the degree matrix of  $G$ ,  $\text{Diag}(\mathbf{x})$  is the diagonal matrix with vector  $\mathbf{x}$  on the main diagonal, and  $\mathbf{1} = (1, \dots, 1)^T$ . In most cases, the dimension of  $\mathbf{1}$  is clear from the context.

Informally, given a weighted graph  $G$ , a *similarity measure* on the set of its vertices  $V(G)$  is a function  $\kappa: V(G) \times V(G) \rightarrow \mathbb{R}$  that characterizes similarity (or affinity, or closeness) between the vertices of  $G$  in a meaningful manner and thus is intuitively and practically adequate for empirical applications [2, 20, 26, 36].

A *kernel on graph* is a graph similarity measure that has an inner product representation. All the inner product matrices (also called Gram matrices) with real entries are symmetric positive semidefinite matrices. On the other hand, any semidefinite matrix has a representation as a Gram matrix with respect to the Euclidean inner product [27].

We note that following [34, 44] we prefer to write *kernel on graph* rather than *graph kernel*, as the notion of “graph kernel” refers to a kernel between graphs [46].

A *proximity measure* (or simply *proximity*) [14] on a set  $A$  is a function  $\kappa: A \times A \rightarrow \mathbb{R}$  that satisfies the *triangle inequality for proximities*, viz.: For any  $x, y, z \in A$ ,  $\kappa(x, y) + \kappa(x, z) - \kappa(y, z) \leq \kappa(x, x)$ , and if  $z = y$  and  $y \neq x$ , then the inequality is strict.

A proximity  $\kappa$  is a  $\Sigma$ -proximity ( $\Sigma \in \mathbb{R}$ ) if it satisfies the *normalization condition*:  $\sum_{y \in A} \kappa(x, y) = \Sigma$  for any  $x \in A$ .

By setting  $z = x$  in the triangle inequality for proximities and using the arbitrariness of  $x$  and  $y$  one verifies that any proximity satisfies *symmetry*:  $\kappa(x, y) = \kappa(y, x)$  for any  $x, y \in A$ . Consequently, if  $\kappa(x, y)$  is a proximity, then  $-\kappa(x, y)$  is a protometric [18, 19].

Furthermore, any  $\Sigma$ -proximity has the *egocentrism* property:  $\kappa(x, x) > \kappa(x, y)$  for any distinct  $x, y \in A$  [14]. If  $\kappa(x, y)$  is represented by a matrix  $K = (K_{xy}) = (\kappa(x, y))$ , then egocentrism of  $\kappa(x, y)$  amounts to the *strict entrywise diagonal dominance* of  $K$ .

If  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are two points in the Euclidean space  $\mathbb{R}^n$ , then  $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$  is the squared distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Schoenberg’s theorem establishes a connection between positive semidefinite matrices (kernels) and matrices of Euclidean distances.

**Theorem 1 ([39, 40]).** *Let  $K$  be an  $n \times n$  symmetric matrix. Define the matrix*

$$\mathcal{D} = (d_{ij}) = \frac{1}{2}(\text{diag}(K) \cdot \mathbf{1}^T + \mathbf{1} \cdot \text{diag}(K)^T) - K, \quad (1)$$

where  $\text{diag}(K)$  is the vector consisting of the diagonal entries of  $K$ . Then there exists a set of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^n$  such that  $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$  ( $i, j = 1, \dots, n$ ) if and only if  $K$  is positive semidefinite.

In the case described in Theorem 1,  $K$  is the Gram matrix of  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Given  $K$ , these vectors can be obtained as the columns of the unique positive semidefinite real matrix  $B$  such that  $B^2 = B^T B = K$ .  $B$  has the expression  $B = U\Lambda^{1/2}U^*$ , where  $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ ,  $\Lambda^{1/2} = \text{Diag}(\lambda_1^{1/2}, \dots, \lambda_n^{1/2})$ , and  $A = U\Lambda U^*$  is the unitary decomposition of  $A$  [27, Corollary 7.2.11].

Connections between proximities and distances are established in [14].

**Theorem 2.** For any proximity  $\kappa$  on a finite set  $A$ , the function

$$d(x, y) = \frac{1}{2}(\kappa(x, x) + \kappa(y, y)) - \kappa(x, y), \quad x, y \in A \quad (2)$$

is a distance function  $A \times A \rightarrow \mathbb{R}$ .

This theorem follows from the proof of Proposition 3 in [14].

**Corollary 1.** Let  $\mathcal{D} = (d_{xy})$  be obtained by (1) from a square matrix  $K$ . If  $\mathcal{D}$  has negative entries or  $\sqrt{d_{xy}} + \sqrt{d_{yz}} < \sqrt{d_{xz}}$  for some  $x, y, z \in \{1, \dots, n\}$ , then the function  $\kappa(x, y) = K_{xy}$ ,  $x, y \in \{1, \dots, n\}$  is not a proximity.

**Proof.** If  $\sqrt{d_{xy}} + \sqrt{d_{yz}} < \sqrt{d_{xz}}$ , then  $d_{xy} + d_{yz} + 2\sqrt{d_{xy}d_{yz}} < d_{xz}$ , i.e., the function  $d(x, y) = d_{xy}$  violates the ordinary triangle inequality. Thus, it is not a distance, as well as in the case where  $\mathcal{D}$  has negative entries. Hence, by Theorem 2,  $\kappa$  is not a proximity.  $\square$

The following theorem describes a one-to-one correspondence between distances and  $\Sigma$ -proximities with a fixed  $\Sigma$  on the same finite set.

**Theorem 3 ([14]).** Let  $\mathbf{S}$  and  $\mathbf{D}$  be the set of  $\Sigma$ -proximities on  $A$  ( $|A| = n$ ;  $\Sigma \in \mathbb{R}$  is fixed) and the set of distances on  $A$ , respectively. Consider the mapping  $\psi(\kappa)$  defined by (2) and the mapping  $\varphi(d)$  defined by

$$\kappa(x, y) = d(x, \cdot) + d(y, \cdot) - d(x, y) - d(\cdot, \cdot) + \frac{\Sigma}{n}, \quad (3)$$

where  $d(x, \cdot) = \frac{1}{n} \sum_{y \in A} d(x, y)$  and  $d(\cdot, \cdot) = \frac{1}{n^2} \sum_{y, z \in A} d(y, z)$ . Then  $\psi(\mathbf{S}) = \mathbf{D}$ ,  $\varphi(\mathbf{D}) = \mathbf{S}$ , and  $\varphi(\psi(\kappa))$ ,  $\kappa \in \mathbf{S}$  and  $\psi(\varphi(d))$ ,  $d \in \mathbf{D}$  are identity transformations.

**Remark 1.** The  $K \rightarrow \mathcal{D}$  transformation (1) is the matrix form of (2). The matrix form of (3) is

$$K = -H\mathcal{D}H + \Sigma J, \quad (4)$$

where  $J = \frac{1}{n}\mathbf{1}\mathbf{1}^T$  and  $H = I - J$  is the centering matrix.

### 3. Kernel, proximity, and distance properties

#### 3.1. Adjacency matrix based kernels

Let us consider several kernels on graphs based on the weighted adjacency matrix  $W$  of a graph.

### 3.1.1. Katz kernel

The *Katz kernel* [31] (also referred to as walk proximity [15] and von Neumann<sup>3</sup> diffusion kernel [30, 43]) is defined<sup>4</sup> as follows:

$$K^{\text{Katz}}(\alpha) = \sum_{k=0}^{\infty} (\alpha W)^k = [I - \alpha W]^{-1},$$

with  $0 < \alpha < (\rho(W))^{-1}$ , where  $\rho(W)$  is the spectral radius of  $W$ .

It is easy to see that  $[I - \alpha W]$  is an M-matrix<sup>5</sup>, i.e., a matrix of the form  $A = qI - B$ , where  $B = (b_{ij})$  with  $b_{ij} \geq 0$  for all  $1 \leq i, j \leq n$ , while  $q$  exceeds the maximum of the moduli of the eigenvalues of  $B$  (in the present case,  $q = 1$ ). Thus,  $[I - \alpha W]$  is a symmetric M-matrix, i.e., a Stieltjes matrix. Consequently,  $[I - \alpha W]$  is positive definite and so is  $K^{\text{Katz}}(\alpha) = [I - \alpha W]^{-1}$ . Thus, by Schoenberg's theorem,  $K^{\text{Katz}}$  can be transformed by (1) into a matrix of squared Euclidean distances.

Moreover, the Katz kernel has the following properties:

If  $[I - \alpha W]$  is row diagonally dominant, i.e.,  $|1 - \alpha w_{ii}| \geq \alpha \sum_{j \neq i} |w_{ij}|$  for all  $i \in V(G)$  (by the finiteness of the underlying space, one can always choose  $\alpha$  small enough such that this inequality becomes valid) then

- $K^{\text{Katz}}(\alpha)$  satisfies the triangle inequality for proximities (see Corollary 6.2.5 in [32]), therefore, transformation (2) provides a distance on  $V(G)$ ;
- $K^{\text{Katz}}(\alpha)$  satisfies egocentrism (i.e., *strict entrywise* diagonal dominance; see also Metzler's property in [32]).

Thus, in the case of row diagonal dominance of  $[I - \alpha W]$ , the Katz kernel is a non-normalized proximity.

### 3.1.2. Communicability kernel

The *communicability kernel* [25, 22, 23] is defined as follows:

$$K^{\text{comm}}(t) = \exp(tW) = \sum_{k=0}^{\infty} \frac{t^k}{k!} W^k.$$

---

<sup>3</sup>M. Saerens [41] has remarked that a more suitable name could be *Neumann diffusion kernel*, referring to the *Neumann series*  $\sum_{k=0}^{\infty} T^k$  (where  $T$  is an operator) named after Carl Gottfried Neumann, while a connection of that to John von Neumann is not obvious (the concept of von Neumann kernel in group theory is essentially different).

<sup>4</sup>In fact, L. Katz considered  $\sum_{k=1}^{\infty} (\alpha W)^k$ .

<sup>5</sup>For the properties of M-matrices, we refer to [32].

(We shall use letter “ $t$ ” whenever some notion of time can be attached to the kernel parameter; otherwise, we shall keep using letter “ $\alpha$ ”.) It is an instance of symmetric exponential diffusion kernels [34]. Since  $K^{\text{comm}}$  is positive semidefinite, by Schoenberg’s theorem, it can be transformed by (1) into a matrix of squared Euclidean distances. However, this does not imply that  $K^{\text{comm}}$  is a proximity.

In fact, it is easy to verify that for the graph  $G$  with weighted adjacency matrix

$$W = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}, \quad (5)$$

$K^{\text{comm}}(1)$  violates the triangle inequality for proximities on the triple of vertices  $(1, \mathbf{2}, 3)$  (the “ $x$ ” element of the inequality is given in bold). On the other hand,  $K^{\text{comm}}(t) \rightarrow I$  as  $t \rightarrow 0$ , which implies that  $K^{\text{comm}}(t)$  with a sufficiently small  $t$  is a [non-normalized] proximity.

Note that the graph corresponding to (5) is a weighted path 1–2–3–4, and immediate intuition suggests the inequality  $d(1, 2) < d(1, 3) < d(1, 4)$  for a distance on its vertices. However,  $K^{\text{comm}}(3)$  induces a Euclidean distance for which  $d(1, 3) > d(1, 4)$ . For  $K^{\text{comm}}(4.5)$  we even have  $d(1, 2) > d(1, 4)$ . However,  $K^{\text{comm}}(t)$  with a small enough positive  $t$  satisfies the common intuition.

By the way, the Katz kernel behaves similarly: when  $\alpha > 0$  is sufficiently small, it holds that  $d(1, 2) < d(1, 3) < d(1, 4)$ , but for  $\alpha > 0.375$ , we have  $d(1, 3) > d(1, 4)$ . Moreover, if  $0.38795 < \alpha < (\rho(W))^{-1}$ , then  $d(1, 2) > d(1, 4)$  is true.

### 3.1.3. Double-factorial similarity

The *double-factorial similarity* [24] is defined as follows:

$$K^{\text{df}}(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!!} W^k.$$

As distinct from the communicability measure,  $K^{\text{df}}$  is not generally a kernel. Say, for the graph with weighted adjacency matrix (5),  $K^{\text{df}}(1)$  has two negative eigenvalues. Therefore  $K^{\text{df}}$  does not generally induce a set of points in  $\mathbb{R}^n$ , nor does it induce a natural Euclidean distance on  $V(G)$ .

Furthermore, in this example, matrix  $\mathcal{D}$  obtained from  $K^{\text{df}}(1)$  by (1) has negative entries. Therefore, by Corollary 1, the function  $\kappa(x, y) = K_{xy}^{\text{df}}(1)$ ,  $x, y \in V(G)$  is not a proximity.

However, as well as  $K^{\text{comm}}(t)$ ,  $K^{\text{df}}(t) \rightarrow I$  as  $t \rightarrow 0$ . Consequently, all eigenvalues of  $K^{\text{df}}(t)$  converge to 1, and hence,  $K^{\text{df}}(t)$  with a sufficiently

small positive  $t$  satisfies the triangle inequality for proximities. Thus,  $K^{\text{df}}(t)$  with a small enough positive  $t$  is a kernel and a [non-normalized] proximity.

### 3.2. Laplacian based kernels

#### 3.2.1. Heat kernel

The *heat kernel* is a symmetric exponential diffusion kernel [34] defined as follows:

$$K^{\text{heat}}(t) = \exp(-tL) = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} L^k,$$

where  $L$  is the ordinary Laplacian matrix of  $G$ .

$K^{\text{heat}}(t)$  is positive-definite for all values of  $t$ , and hence, it is a kernel. Then, by Schoenberg's theorem,  $K^{\text{heat}}$  induces a Euclidean distance on  $V(G)$ . For our example (5), this distance for all  $t > 0$  obeys the intuitive inequality  $d(1, 2) < d(1, 3) < d(1, 4)$ .

On the other hand,  $K^{\text{heat}}$  is not generally a proximity. E.g., for the example (5),  $K^{\text{heat}}(t)$  violates the triangle inequality for proximities on the triple of vertices  $(1, \mathbf{2}, 3)$  whenever  $t > 0.431$ . As well as for the communicability kernel,  $K^{\text{heat}}(t)$  with a small enough  $t$  is a proximity. Moreover, it is a 1-proximity, as in the above series expansion,  $L$  has row sums 0, while  $L^0 = I$  has row sums 1. Thus, the 1-normalization condition is satisfied for any  $t > 0$ .

#### 3.2.2. Normalized heat kernel

The *normalized heat kernel* is defined as follows:

$$K^{\text{n-heat}}(t) = \exp(-t\mathcal{L}) = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \mathcal{L}^k,$$

where  $\mathcal{L} = D^{-1/2}LD^{-1/2}$  is the normalized Laplacian,  $D$  being the degree matrix of  $G$  [16].

For this kernel, the main conclusions are the same as for the standard heat kernel. For the example (5),  $K^{\text{heat}}(t)$  violates the triangle inequality for proximities on the triple of vertices  $(1, \mathbf{2}, 3)$  when  $t > 1.497$ . It is curious to observe that the triangle inequality of the example (5) is violated starting with a larger value of  $t$  in comparison with the case of the standard heat kernel. An important distinction is that generally,  $\mathcal{L}$  has nonzero row sums. As a result,  $K^{\text{n-heat}}$  does not satisfy the normalization condition, and even for small  $t > 0$ ,  $K^{\text{n-heat}}$  is a non-normalized proximity.



### 3.2.3. Regularized Laplacian kernel

The *regularized Laplacian kernel*, or *forest kernel* is defined [12] as follows:

$$K^{\text{regL}}(t) = [I + tL]^{-1},$$

where  $t > 0$ .

As was shown in [13, 15], the regularized Laplacian kernel is a 1-proximity and a row stochastic matrix. Since  $[I + tL]$  is positive definite, so is  $[I + tL]^{-1}$ , and by Schoenberg's theorem,  $K^{\text{regL}}$  induces a Euclidean distance on  $V(G)$ .

For the example (5), the induced distances corresponding to  $K^{\text{regL}}$  always satisfy  $d(1, 2) < d(1, 3) < d(1, 4)$ . Regarding the other properties of  $K^{\text{regL}}$ , we refer to [13, 3].

It is the first encountered example of similarity measure that satisfies the both distance and proximity properties for all values of the kernel parameter.

### 3.2.4. Absorption kernel

The *absorption kernel* [29] is defined as follows:

$$K^{\text{absorp}}(t) = [tA + L]^{-1}, \quad t > 0,$$

where  $A = \text{Diag}(\mathbf{a})$  and  $\mathbf{a} = (a_1, \dots, a_n)^T$  is called the *vector of absorption rates* and has positive components. As  $K^{\text{absorp}}(t^{-1}) = t(A + tL)^{-1}$ , this kernel is actually a generalization of the previous one.

Since  $[tA + L]$  is positive definite, Schoenberg's theorem attaches a matrix of squared Euclidean distances to  $K^{\text{absorp}}(t)$ .

$[tA + L]$  is a row diagonally dominant Stieltjes matrix, hence, by Corollary 6.2.5 in [32] we conclude that  $K^{\text{absorp}}$  satisfies the triangle inequality for proximities, i.e.,  $K^{\text{absorp}}$  is a proximity (but not generally a  $\Sigma$ -proximity).

## 3.3. Markov matrix based kernels and measures

### 3.3.1. Personalized PageRank

*Personalized PageRank* (PPR) similarity measure [38] is defined as follows:

$$K^{\text{PPR}}(\alpha) = [I - \alpha P]^{-1},$$

where  $P = D^{-1}W$  is a row stochastic (Markov) matrix,  $D$  is the degree matrix of  $G$ , and  $0 < \alpha < 1$ , which corresponds to the standard random walk on the graph.

In general,  $K^{\text{PPR}}(\alpha)$  is not symmetric, so it is not positive semidefinite, nor is it a proximity.

Moreover, the functions  $d(x, y)$  obtained from  $K^{\text{PPR}}$  by transformation<sup>6</sup>

$$d(x, y) = \frac{1}{2}(\kappa(x, x) + \kappa(y, y) - \kappa(x, y) - \kappa(y, x)) \quad (6)$$

need not generally be distances. Say, for

$$W = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix} \quad (7)$$

with  $K^{\text{PPR}}(\alpha)$ , one has  $d(1, 3) + d(3, 4) < d(1, 4)$  whenever  $\alpha > 0.9515$ .

$K^{\text{PPR}}$  has only positive eigenvalues. However, its symmetrized counterpart  $\frac{1}{2}(K^{\text{PPR}} + (K^{\text{PPR}})^T)$  may have a negative eigenvalue (say, with  $\alpha \geq 0.984$  for (5) or with  $\alpha \geq 0.98$  for (7)). Thus, it need not be positive semidefinite and, consequently, by Theorem 1,  $\mathcal{D}$  obtained from it by (1) (or from  $K^{\text{PPR}}$  by (6)) is not generally a matrix of squared Euclidean distances.

$K^{\text{PPR}}$  satisfies the normalization condition. For a small enough  $\alpha$ , it can be transformed (as well as  $K^{\text{comm}}$  and  $K^{\text{df}}$ ) into a distance matrix using (6).

On the other hand, one can slightly modify Personalized PageRank so it becomes a proximity. Rewrite  $K^{\text{PPR}}$  as follows:

$$[I - \alpha D^{-1}W]^{-1} = [D - \alpha W]^{-1}D.$$

Then consider

### 3.3.2. Modified Personalized PageRank

$$K^{\text{modifPPR}}(\alpha) = [I - \alpha D^{-1}W]^{-1}D^{-1} = [D - \alpha W]^{-1}, \quad 0 < \alpha < 1,$$

which becomes a non-normalized proximity by Corollary 6.2.5 in [32]. In particular, the triangle inequality becomes

$$\frac{K_{ii}^{\text{PPR}}(\alpha)}{d_i} - \frac{K_{ji}^{\text{PPR}}(\alpha)}{d_i} - \frac{K_{ik}^{\text{PPR}}(\alpha)}{d_k} + \frac{K_{jk}^{\text{PPR}}(\alpha)}{d_k} \geq 0,$$

---

<sup>6</sup>If  $K$  is symmetric, then (6) coincides with (2).

which looks like an interesting inequality for Personalized PageRank. Due to symmetry,  $K_{ij}^{\text{modifPPR}} = K_{ji}^{\text{modifPPR}}$ , and we obtain an independent proof of the following identity for Personalized PageRank [2]:

$$\frac{K_{ij}^{\text{PPR}}(\alpha)}{d_j} = \frac{K_{ji}^{\text{PPR}}(\alpha)}{d_i}.$$

Note that replacing the Laplacian matrix  $L = D - W$  with  $D - \alpha W$  is a kind of alternative regularization of  $L$ . Being diagonally dominant,

$$D - \alpha W = \bar{d}I - (\bar{d}I - D + \alpha W) \quad (8)$$

(where  $\bar{d}$  is the maximum degree of the vertices of  $G$ ) is a Stieltjes matrix. Consequently,  $D - \alpha W$  is positive definite and so is  $K^{\text{modifPPR}}(\alpha) = [D - \alpha W]^{-1}$ . Thus, by Schoenberg's theorem,  $K^{\text{modifPPR}}$  can be transformed by (1) into a matrix of squared Euclidean distances.

We note that Personalized PageRank can be generalized by using non-homogeneous restart [4], which will lead to the discrete-time analog of the absorption kernel. However, curiously enough, the discrete-time version has a smaller number of proximity-distance properties than the continuous-time version.

### 3.3.3. PageRank heat similarity measure

*PageRank heat similarity measure* [17] is defined as follows:

$$K^{\text{heatPPR}}(t) = \exp(-t(I - P)).$$

Basically, the properties of this measure are similar to those of the standard Personalized PageRank. Say, for the example (7) with  $K^{\text{heatPPR}}$ , one has  $d(1, 2) + d(2, 3) < d(1, 3)$  whenever  $t > 1.45$ .

### 3.4. Logarithmic similarity measures and transitional properties

Given a strictly positive similarity measure  $s(x, y)$ , the function  $\kappa(x, y) = \ln s(x, y)$  is the corresponding *logarithmic similarity*.

Using Theorem 2 it can be verified [8] that whenever  $S = (s_{ij}) = (s(i, j))$  produces a strictly positive *transitional measure* on  $G$  (i.e.,  $s_{ij} s_{jk} \leq s_{ik} s_{jj}$  for all vertices  $i, j$ , and  $k$ , while  $s_{ij} s_{jk} = s_{ik} s_{jj}$  if and only if every path from  $i$  to  $k$  visits  $j$ ), we have that the logarithmic similarity  $\kappa(x, y) = \ln s(x, y)$  produces a *cutpoint additive distance*, viz., a distance that satisfies  $d(i, j) + d(j, k) = d(i, k)$  iff every path from  $i$  to  $k$  visits  $j$ :

$$d(i, j) = \frac{1}{2}(\kappa(i, i) + \kappa(j, j) - \kappa(i, j) - \kappa(j, i)) = \ln \sqrt{\frac{s(i, i) s(j, j)}{s(i, j) s(j, i)}}. \quad (9)$$

In the case of digraphs, five transitional measures were indicated in [8], namely, *connection reliability*, *path accessibility* with a sufficiently small parameter, *walk accessibility*, and two versions of *forest accessibility*; the undirected counterparts of the two latter measures were studied in [10] and [9], respectively.

**Proposition 1.**  $K^{\text{absorp}}$ ,  $K^{\text{PPR}}$ , and  $K^{\text{modifPPR}}$  produce transitional measures.

**Proof.** For  $K^{\text{absorp}}(t) = [tA + L]^{-1}$ , let  $h = \max_i \{a_i t + d_i - w_{ii}\}$ , where  $d_i$  is the degree of vertex  $i$ . Then  $K^{\text{absorp}}(t) = [hI - (hI - tA - D + W)]^{-1} = [I - W']^{-1} h^{-1}$ , where  $W' = h^{-1}(hI - tA - D + W)$  is nonnegative with row sums less than 1. Hence,  $K^{\text{absorp}}(t)$  is positively proportional to the matrix  $[I - W']^{-1}$  of walk weights of the graph with weighted adjacency matrix  $W'$ .

Similarly, by (8),  $K^{\text{modifPPR}}(\alpha) = [D - \alpha W]^{-1} = [I - W'']^{-1} \bar{d}^{-1}$ , where  $W'' = \bar{d}^{-1}(\bar{d}I - D + \alpha W)$  is nonnegative with row sums less than 1. Consequently,  $K^{\text{modifPPR}}(\alpha)$  is proportional to the matrix of walk weights of the graph whose weighted adjacency matrix is  $W''$ .

Finally,  $K^{\text{PPR}}(\alpha)$  is the matrix of walk weights of the digraph with weighted adjacency matrix  $\alpha P$ .

Since by [8, Theorem 6], any finite matrix of walk weights of a weighted digraph produces a transitional measure, so do  $K^{\text{absorp}}$ ,  $K^{\text{PPR}}$ , and  $K^{\text{modifPPR}}$ .  $\square$

Thus, as by Proposition 1 and the results of [8],  $K^{\text{Katz}}$ ,  $K^{\text{regL}}$ ,  $K^{\text{absorp}}$ ,  $K^{\text{PPR}}$ , and  $K^{\text{modifPPR}}$  produce transitional measures, we have that the corresponding *logarithmic* dissimilarities (9) are cutpoint additive distances.

Furthermore, if  $S = (s_{ij}) = (s(i, j))$  produces a strictly positive transitional measure on  $G$ , then, obviously,  $\kappa(x, y) = \ln s(x, y)$  satisfies  $\kappa(y, x) + \kappa(x, z) - \kappa(y, z) \leq \kappa(x, x)$ , which coincides<sup>7</sup> with the triangle inequality for proximities whenever  $s(x, y)$  is symmetric. Therefore, as  $K^{\text{Katz}}$ ,  $K^{\text{regL}}$ ,  $K^{\text{absorp}}$ , and  $K^{\text{modifPPR}}$  are symmetric, we obtain that the corresponding logarithmic similarities  $\kappa(x, y) = \ln s(x, y)$  are proximities.

$K^{\text{PPR}}$  is not generally symmetric, however, it can be observed that  $\tilde{K}^{\text{PPR}}$  such that  $\tilde{K}_{ij}^{\text{PPR}} = \sqrt{K_{ij}^{\text{PPR}} K_{ji}^{\text{PPR}}}$  is symmetric and produces the same *logarithmic* distance (9) as  $K^{\text{PPR}}$ . Hence, the logarithmic similarity  $\kappa(x, y) = \ln \tilde{K}_{xy}^{\text{PPR}}$  is a proximity.

At the same time, the above logarithmic similarities are not kernels, as the corresponding matrices have negative eigenvalues.

---

<sup>7</sup>On various alternative versions of the triangle inequality, we refer to [18].

This implies that being a proximity is not a stronger property than being a kernel. By Corollary 1, the square root of the distance induced by a proximity is also a distance. However, this square rooted distance need not generally be Euclidean, thus, Theorem 1 is not sufficient to conclude that the initial proximity is a kernel.

It can be verified that all logarithmic measures corresponding to the similarity measures under study preserve the natural order of distances  $d(1, 2) < d(1, 3) < d(1, 4)$  for the example (5).

#### 4. Conclusion

We have considered similarity measures on graphs based upon three fundamental graph matrices: the adjacency matrix, the Laplacian matrix, and the (stochastic) Markov matrix. For each measure, we examine if it is a kernel, if it is a proximity measure, and if it is a transitional measure: these classes are not nested. Regularized Laplacian turns out to be a similarity measure satisfying most of the properties, whereas the logarithmic similarity transformation appears to be useful as a tool for obtaining cutpoint additive distances. We are currently working on an understanding of what consequences the established properties have for machine learning algorithms.

#### Acknowledgements

The work of KA and DR was supported by the joint Bell Labs Inria ADR “Network Science” and the work of PC was supported by the Russian Science Foundation (project no.16-11-00063 granted to IRE RAS).

#### References

- [1] Avrachenkov, K., Mishenin, A., Gonçalves, P. and Sokol, M., 2012. Generalized optimization framework for graph-based semi-supervised learning. In: *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 966–974).
- [2] Avrachenkov, K., Gonçalves, P. and Sokol, M., 2013. On the choice of kernel and labelled data in semi-supervised learning methods. In: *Proceedings of WAW 2013* (pp. 56–67).
- [3] Avrachenkov, K., Chebotarev, P. and Mishenin, A., 2017. Semi-supervised learning with regularized Laplacian. *Optimization Methods and Software*, 32(2), pp. 222–236.

- [4] Avrachenkov, K., van der Hofstad, R. and Sokol, M., 2014. Personalized PageRank with node-dependent restart. In: *Proceedings of International Workshop on Algorithms and Models for the Web-Graph* (pp. 23–33).
- [5] Backstrom, L. and Leskovec, J., 2011. Supervised random walks: predicting and recommending links in social networks. In: *Proceedings of ACM WSDM 2011* (pp. 635–644).
- [6] Buckley, F., Harary, F., 1990. *Distance in Graphs*, Addison-Wesley, Redwood City, CA.
- [7] Chapelle, O., Schölkopf, B. and Zien, A. 2006. *Semi-Supervised Learning*, MIT Press.
- [8] Chebotarev, P., 2011. The graph bottleneck identity. *Advances in Applied Mathematics*, 47(3), pp. 403–413.
- [9] Chebotarev, P., 2011. A Class of graph-geodetic distances generalizing the shortest-path and the resistance distances. *Discrete Applied Mathematics*, 159(5), pp. 295–302.
- [10] Chebotarev, P., 2012. The walk distances in graphs, *Discrete Applied Mathematics*, 160(10-11), pp. 1484–1500.
- [11] Chebotarev, P., 2013. Studying new classes of graph metrics. In: *Proceedings of the SEE Conference “Geometric Science of Information” (GSI-2013)*, F. Nielsen, F. Barbaresco (eds.), Lecture Notes in Computer Science, LNCS 8085, Springer, Berlin (pp. 207–214).
- [12] Chebotarev, P.Yu. and Shamis, E.V., 1995. On the proximity measure for graph vertices provided by the inverse Laplacian characteristic matrix. In: *Abstracts of the conference “Linear Algebra and its Application”*, 10-12 June 1995, The Institute of Mathematics and its Applications, in conjunction with the Manchester Center for Computational Mathematics, Manchester, UK (pp. 6–7). URL <http://www.ma.man.ac.uk/~higham/1aa95/abstracts.ps>
- [13] Chebotarev, P.Yu. and Shamis, E.V., 1997. The matrix-forest theorem and measuring relations in small social groups. *Autom. Remote Control*, 58(9), pp. 1505–1514.
- [14] Chebotarev, P.Yu. and Shamis, E.V., 1998. On a duality between metrics and  $\Sigma$ -proximities. *Autom. Remote Control*, 59(4), pp. 608–612. Erratum: 59(10), p. 1501.

- [15] Chebotarev, P.Yu. and Shamis, E.V., 1998. On proximity measures for graph vertices. *Autom. Remote Control*, 59(10), pp. 1443–1459.
- [16] Chung, F., 1997. *Spectral graph theory*. American Math. Soc.
- [17] Chung, F., 2007. The heat kernel as the pagerank of a graph. *Proceedings of the National Academy of Sciences*, 104(50), pp. 19735–19740.
- [18] Deza, M. and Chebotarev, P., 2011. Protometrics. arXiv preprint arXiv:1112.4829.
- [19] Deza, M.M., Deza, E., 2016. *Encyclopedia of Distances.*, 4th ed., Springer, Berlin–Heidelberg.
- [20] Dhillon, I. S., Fan, J. and Guan, Y., 2001. Efficient clustering of very large document collections. *Data mining for scientific and engineering applications*, 2, pp. 357–381.
- [21] Dhillon, I. S., Guan, Y. and Kulis, B., 2004. Kernel k-means: spectral clustering and normalized cuts. In *Proceedings of ACM KDD 2004* (pp. 551–556).
- [22] Estrada, E. and Hatano, N., 2007. Statistical-mechanical approach to subgraph centrality in complex networks. *Chem. Phys. Lett.*, 439, pp. 247–251.
- [23] Estrada, E. and Hatano, N., 2008. Communicability in complex networks. *Physical Review E*, 77(3), 036111.
- [24] Estrada, E. and Silver, G., 2017. Accounting for the role of long walks on networks via a new matrix function. *Journal of Mathematical Analysis and Applications*, 449, pp. 1581–1600.
- [25] Fouss, F., Yen L., Pirotte, A., and Saerens, M., 2006. An experimental investigation of graph kernels on a collaborative recommendation task. In *Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*. IEEE (pp. 863–868).
- [26] Fouss, F., Saerens, M., and Shimbo, M., 2016. *Algorithms and Models for Network Data and Link Analysis*. Cambridge University Press.
- [27] Horn, R.A. and Johnson, C.R., 2013. *Matrix Analysis* (2nd Edition). Cambridge University Press.

- [28] Ivashkin, V. and Chebotarev P., 2017. Do logarithmic proximity measures outperform plain ones in graph clustering? In: *Models, Algorithms, and Technologies for Network Analysis*, V.A. Kalyagin et al., eds. Proceedings in Mathematics & Statistics, Vol. 197, Springer, , pp. 87–105.
- [29] Jacobsen, K.A. and Tien, J.H., 2018. A generalized inverse for graphs with absorption. *Linear Algebra and its Applications*, 537, pp. 118–147.
- [30] Kandola, J., Shawe-Taylor, J. and Cristianini, N., 2003. Learning semantic similarity. In *Advances in Neural Information Processing Systems 15 (NIPS 15)*. MIT Press (pp. 673-680).
- [31] Katz, L., 1953. A new status index derived from sociometric analysis. *Psychometrika*, 18(1), pp. 39–43.
- [32] Kirkland, S.J. and Neumann, M., 2012. *Group Inverses of M-matrices and Their Applications*. CRC Press.
- [33] Kivimäki, I., Shimbo, M. and Saerens, M., 2014. Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A: Statistical Mechanics and its Applications*, 393, pp. 600–616.
- [34] Kondor, R.I. and Lafferty, J., 2002. Diffusion kernels on graphs and other discrete input spaces. In: *Proceedings of ICML* (pp. 315–322).
- [35] Lenart, C., 1998. A generalized distance in graphs and centered partitions. *SIAM Journal on Discrete Mathematics*, 11(2), pp. 293–304.
- [36] Liben-Nowell, D. and Kleinberg, J., 2007. The link-prediction problem for social networks. *Journal of the Association for Information Science and Technology*, 58(7), pp. 1019–1031.
- [37] Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B., 2001. An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*, 12(2), pp. 181–202.
- [38] Page, L., Brin S., Motwani R. and Winograd, T., 1999. The PageRank citation ranking: Bringing order to the web, Technical Report, Stanford InfoLab.
- [39] Schoenberg, I.J., 1935. Remarks to Maurice Fréchet’s article “Sur la définition axiomatique d’une classe d’espace distanciés vectoriellement applicable sur l’espace de Hilbert”. *Ann. Math.* 36(3), pp. 724–732.



- [40] Schoenberg, I.J., 1938. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3), pp. 522–536.
- [41] Saerens, M., 2016. Personal communication.
- [42] Sharpe, G.E., 1967. Solution of the  $(m + 1)$ -terminal resistive network problem by means of metric geometry. In: *Proceedings of the First Asilomar Conference on Circuits and Systems*, Pacific Grove, CA (pp. 319–328).
- [43] Shawe-Taylor, J. and Cristianini, N., 2004. *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press.
- [44] Smola, A.J. and Kondor, R., 2003. Kernels and regularization on graphs. In: *Learning Theory and Kernel Machines* (pp. 144–158).
- [45] Sommer, F., Fouss, F. and Saerens, M., 2016. Comparison of graph node distances on clustering tasks, In: *Lecture Notes in Computer Science*, LNCS 9886, Springer, pp. 192-201.
- [46] Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R. and Borgwardt, K.M., 2010. Graph kernels. *Journal of Machine Learning Research*, 11(Apr), pp. 1201–1242.
- [47] Zhou, D., Schölkopf, B. and Hofmann, T., 2004. Semi-supervised learning on directed graphs. In: *Proceedings of NIPS* (pp. 1633–1640).