Diffusion and Superposition Distances for Signals Supported on Networks

Santiago Segarra, *Student Member, IEEE*, Weiyu Huang, *Student Member, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

***Abstract*—We introduce the diffusion and superposition distances as two metrics to compare signals supported in the nodes of a network. Both metrics consider the given vectors as initial temperature distributions and diffuse heat through the edges of the graph. The similarity between the given vectors is determined by the similarity of the respective diffusion profiles. The superposition distance computes the instantaneous difference between the diffused signals and integrates the difference over time. The diffusion distance determines a distance between the integrals of the diffused signals. We prove that both distances define valid metrics and that they are stable to perturbations in the underlying network. We utilize numerical experiments to illustrate their utility in classifying signals in a synthetic network as well as in classifying ovarian cancer histologies using gene mutation profiles of different patients. We also utilize diffusion as part of a label propagation method in semi-supervised learning to classify handwritten digits.**

***Index Terms*—Graph signals, networks, diffusion, superposition, signal classification.**

I. Introduction

NETWORKS, or graphs, are data structures that encode relationships between elements of a group and which, for this reason, play an important role in many disparate disciplines such as biology [1], [2] and sociology [3], [4] where relationships between, say, genes, species or individuals, are central. Often, networks have intrinsic value and are themselves the object of study. This is the case, e.g., when we are interested in distributed and decentralized algorithms in which agents iterate through actions that use information available either locally or at adjacent nodes to accomplish some sort of global outcome [5]–[7]. Equally often, the network defines an underlying notion of proximity, but the object of interest is a signal defined on top of the graph. This is the matter addressed in the field of graph signal processing, where the notions of frequency and linear filtering are extended to signals supported on graphs [8]–[12]. Examples of network-supported signals include gene expression patterns defined on top of gene networks [13] and brain activity signals supported on top of brain connectivity networks [14]. Indeed, one of the principal uses of networks of gene interactions is to determine how a change in the expression of a gene, or a group of genes, cascades through the network and alters the expression of other genes. Likewise, a brain connectivity network specifies relationships between areas of the brain, but it is the pattern of activation of these regions that determines the mental state of the subject.

In this paper we consider signals supported on graphs and address the challenge of defining a notion of distance between these signals that incorporates the structure of the underlying network. We want these distances to be such that two signals are deemed close if they are themselves close–in the examples in the previous paragraph we have gene expression or brain activation patterns that are similar–, or if they have similar values in adjacent or nearby nodes–the expressed genes or the active areas of the brain are not similar but they effect similar changes in the gene network or represent activation of closely connected areas of the brain. We define here the diffusion and superposition distances and argue that they inherit this functionality through their connection to diffusion processes.

Diffusion processes draw their inspiration from the diffusion of heat through continuous matter [15], [16]. The linear differential equation that models heat diffusion can be extended to encompass dynamics through discrete structures such as graphs or networks [17]–[21]. In the particular case of graphs, every node is interpreted as containing an amount of heat which flows from hot to cold nodes. The flow of heat is through the edges of the graph and such that the rate at which heat diffuses is proportional to both the heat difference between the nodes adjacent to the edge and the edge weight representing the proximity between these nodes. Diffusion processes in graphs are often used in engineering and science because they reach isothermal configurations in steady state. Driving the network to an isothermal equilibrium is tantamount to achieving a consensus action [22], [23], which, in turn, is useful in, e.g., problems in formation control [24] and flocking [25], as well as an important modeling tool in situations such as the propagation of opinions in social networks [26]–[28].

In this paper we do not exploit the asymptotic, but rather the transient behavior of diffusion processes. We regard the given vectors as initial heat configurations that generate different diffused heat profiles over time. The diffusion and superposition distances between the given vectors are defined as the difference between these heat profiles integrated over time. The superposition distance compares the instantaneous difference between the two evolving heat maps and integrates this difference over time. The diffusion metric integrates each of the heat profiles over time and evaluates the norm of the

Manuscript received November 26, 2014; revised March 15, 2015; accepted April 16, 2015. Date of publication May 06, 2015; date of current version June 29, 2015. This work was supported by NSF CCF-1217963. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Cedric Richard.

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: ssegarra@seas.upenn.edu; whuang@seas.upenn.edu; aribeiro@seas.upenn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSIPN.2015.2429471

difference between the two integrals. Both of these distances yield small values when the diffusion profiles are similar. This happens if the given vectors themselves are close or if they have similar values at nodes that are linked by edges with high similarity values.

A. Contributions and summary

The contributions of this paper are: (i) To design the superposition and diffusion distances and to prove their validity as metrics in the space of vectors supported on a given graph. (ii) To show that both distances are well behaved with respect to small perturbations in the underlying network. (iii) To illustrate their ability to identify vectors that are similar only after the network structure is accounted for. (iv) To demonstrate their value in two practical scenarios; the classification of ovarian cancer types from gene mutation profiles and the classification of handwritten arabic digits.

We begin the paper with a brief introduction of basic concepts in graph theory and metric geometry followed by a formal description of diffusion dynamics in networks (Section II). This preliminary discussion provides the necessary elements for a formal definition of the superposition and diffusion distances. In Section III we define the superposition distance between two signals with respect to a given graph and a given input norm. To determine this distance the signals are diffused in the graph, the input norm of their difference is computed for all times, and the result is discounted by an exponential factor and integrated over time. We show that the superposition distance is a valid metric between vectors supported in the node set of a graph.

The diffusion distance with respect to a given graph and a given input norm is introduced in Section IV as an alternative way of measuring the distance between two signals in a graph. In this case the diffused signals are also exponentially discounted and integrated over time but the input norm is taken after time integration. The diffusion distance is shown to also be a valid metric in the space of signals supported on a given graph and is further shown to provide a lower bound for the superposition distance. Different from the superposition distance, the diffusion distance can be reduced to a closed form expression with computational cost dominated by one matrix inversion. The superposition distance requires numerical integration of the time integral of the norm of a matrix exponential.

We further address stability with respect to uncertainty in the specification of the network (Section V). Specifically, we prove that when the input norm is either the 1-norm, the 2-norm, or the infinity-norm a small perturbation in the underlying network transports linearly to a small perturbation in the values of the superposition and diffusion distances. In Section VI we demonstrate that the diffusion and superposition distances can be applied to classify signals in graphs with better accuracy than comparisons that utilize traditional vector distances. We illustrate the differences using synthetic data (Section VI-A) and establish the practical advantages through the classification of ovarian cancer histologies from gene mutation profiles of different patients (Section VI-B). In Section VI-C, we utilize diffusion as part of a label propagation process and present its benefit through the classification of handwritten digits. Concluding remarks are presented in Section VII.

II. Preliminaries

A. Graphs and networks

We consider networks that are weighted, undirected, and symmetric. Formally, we define a network as a graph $G = (V, E, W)$, where $V = \{1, \ldots, n\}$ is a finite set of n nodes or vertices, $E \subseteq V \times V$ is a set of edges defined as ordered pairs (i, j), and $W : E \to \mathbb{R}_{++}$ is a map from the set of edges to the strictly positive reals, representing weights $w_{ij} > 0$ associated with each edge (i, j). Since the graph is undirected, we must have that the edge $(i, j) \in E$ if and only if $(j, i) \in E$. Since the graph is also symmetric, we must have $w_{ij} = w_{ji}$ for all $(i, j) \in E$. The edge (i, j) represents the existence of a relationship between i and j and we say that i and j are adjacent or neighboring. The weight $w_{ij} = w_{ji}$ represents the strength of the relationship, or, equivalently, the proximity or similarity between i and j. Larger edge weights are interpreted as higher similarity between the border nodes. The graphs considered here do not contain self loops, i.e., $(i, i) \notin E$ for any $i \in V$. We consider the usual definitions of the adjacency, Laplacian, and degree matrices for the weighted graph $G = (V, E, W)$; see e.g. [29, Chapter 1]. The adjacency matrix $A \in \mathbb{R}_{+}^{n \times n}$ is such that $A_{ij} = w_{ij}$ whenever i and j are adjacent, i.e., whenever $(i, j) \in E$ and such that for $(i, j) \notin E$ we have $A_{ij} = 0$. The degree matrix $D \in \mathbb{R}_{+}^{n \times n}$ is a diagonal matrix such that the i-th diagonal element $D_{ii} = \sum_j w_{ij}$ contains the sum of all the weights out of node i. The Laplacian matrix is defined as the difference $L := D - A \in \mathbb{R}^{n \times n}$. Since D is diagonal and the diagonal of A is null–because G does not have self loops–the components of the Laplacian matrix are explicitly given by

$$L_{ij} := \begin{cases} -A_{ij} & \text{if} \quad i \neq j, \\ \\ \sum_{k=1}^{n} A_{ik} & \text{if} \quad i = j. \end{cases} \tag{1}$$

Observe that the Laplacian is positive semidefinite [30] because it is diagonally dominant with positive diagonal elements.

B. Metrics and norms

Our goal in this paper is to define a metric to compare vectors defined on top of a graph. For reference, recall that for a given space X, a metric $d : X \times X \to \mathbb{R}_+$ is a function from pairs of elements in X to the nonnegative reals satisfying the following three properties for every $x, y, z \in X$:

Symmetry: $d(x, y) = d(y, x)$.

Identity: $d(x, y) = 0$ if and only if $x = y$.

Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

A closely related definition is that of a norm. In this case we need to have a given vector space Y and consider elements $v \in Y$. A norm $\| \cdot \|$ is a function $\| \cdot \| : Y \to \mathbb{R}_+$ from Y to the nonnegative reals such that, for all vectors $v, w \in Y$ and scalar constant β, it satisfies:

Positiveness: $\|v\| \geq 0$ with equality if and only if $v = \vec{0}$.

Positive homogeneity: $\|\beta\, w\| = |\beta|\ \|w\|$.

Subadditivity: $\|v + w\| \leq \|v\| + \|w\|$.

Norms are more stringent than metrics because they require the existence of a null element with null norm. However, whenever a norm is defined on a vector space Y it induces a distance in the same space as we formally state next [31, Chapter 1].

Lemma 1: Given any norm $\|\cdot\|$ on some vector space Y, the function $d : Y \times Y \to \mathbb{R}_+$ defined as $d(r, s) := \|r - s\|$ for all pairs $r, s \in Y$ is a metric.

In some of our proofs we encounter norms induced in the vector space of matrices $\mathbb{R}^{n\times n}$ by norms defined in the vector space $\mathbb{R}^n$. For a given vector norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ the induced matrix norm $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}_+$ is defined as

$$\|A\| := \sup_{\|x\|=1} \|Ax\|. \tag{2}$$

I.e., the induced norm of matrix A is equal to the maximum achievable vector norm when multiplying A by a vector with unit norm. Apart from satisfying the three requirements in the definition of norms, induced matrix norms are compatible and submultiplicative [32, Section 2.3]. That they are submultiplicative means that for any given pair of matrices $A, B \in \mathbb{R}^{n\times n}$ the norm of the product does not exceed the product of the norms,

$$\|AB\| \leq \|A\|\,\|B\|. \tag{3}$$

That they are compatible means that for any vector $x \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n\times n}$ it holds,

$$\|Ax\| \leq \|A\|\,\|x\|. \tag{4}$$

I.e., the vector norm of the product Ax does not exceed the product of the norm of the vector x and the induced norm of the matrix A.

C. *Diffusion dynamics*

Consider an arbitrary graph $G = (V, E, W)$ with Laplacian matrix L and a vector $r = [r_1, \dots, r_n]^T \in \mathbb{R}^n$ where the component r_i of r corresponds to the node i of G. For a given constant $\alpha > 0$, define the time-varying vector $r(t) \in \mathbb{R}^n$ as the solution of the linear differential equation

$$\frac{\mathrm{d}\, r(t)}{\mathrm{d}\, t} = -\alpha\, L\, r(t), \qquad r(0) = r. \tag{5}$$

The differential equation in (5) represents heat diffusion on the graph G because $-L$ can be shown to be the discrete approximation of the continuous Laplacian operator used to describe the diffusion of heat in physical space [17]. The given vector $r = r(0)$ specifies the initial temperature distribution and $r(t)$ represents the temperature distribution at time t. The constant α is the thermal conductivity–which depends on the units used to measure the weights on the graph–and controls the heat diffusion rate. Larger α results in faster changing $r(t)$. The solution of (5) is given by

$$r(t) = e^{-\alpha\, L\, t}\, r, \tag{6}$$

where, for an arbitrary matrix $A \in \mathbb{R}^{n\times n}$, the matrix exponential e^A is defined as [30]

$$e^A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k. \tag{7}$$

Direct substitution is enough to confirm that indeed $r(t) = e^{-\alpha L t} r$ is a solution of (5). The expression in (6) allows us to compute the temperature distribution at any point in time given the initial heat configuration r and the structure of the underlying network through its Laplacian L. Notice that as time grows, $r(t)$ settles to an isothermal equilibrium–all nodes have the same temperature–if the graph is connected. It is instructive to rewrite (5) componentwise. If we focus on the variation of the i-th component of $r(t)$ and use the definition of L in (1) to replace $L_{ik} = -A_{ik}$ and $L_{ii} = \sum_{k=1}^{n} A_{ik}$, it follows that (5) implies

$$\frac{\mathrm{d}\, r_i(t)}{\mathrm{d}\, t} = \sum_{j=1}^{n} \alpha\, A_{ij}\, (r_j(t) - r_i(t)). \tag{8}$$

Further recalling that $A_{ij} = 0$ if i and j are not adjacent and that $A_{ij} = w_{ij}$ otherwise, we see that the sum in (8) entails multiplying each of the differences $r_j(t) - r_i(t)$ between adjacent nodes by the corresponding proximities w_{ij} on top of the constant thermal conductivity α. Thus, (8) is describing the flow of heat through edges of the graph. The flow of heat on an edge grows proportionally with the temperature differential $r_j(t) - r_i(t)$ as well as with the proximity w_{ij}. Nodes with larger proximity tend to equalize their temperatures faster, other things being equal. In particular, two initial vectors $r(0) = r$ and $s(0) = s$ result in similar temperature distributions across time if they are themselves similar–all r_i and s_i components are close–, or if they have similar initial levels at nodes with larger proximity–each component r_i need not be similar to s_i itself but might be similar to the component s_j of a neighboring node for which the edge weight w_{ij} is large. This latter fact suggests that the diffused vectors $r(t)$ and $s(t)$ define a notion of proximity between r and s associated with the underlying graph structure. We exploit this observation to define distances between signals supported on graphs in the following two sections.

III. Superposition distance

Given an arbitrary graph $G = (V, E, W)$ with Laplacian matrix L, an input vector norm $\|\cdot\|$, and two signals $r, s \in \mathbb{R}^n$ defined in the node space V, we define the superposition distance $d^L_{\text{sps}}(r, s)$ between r and s as

$$d^L_{\text{sps}}(r, s) := \int_0^{+\infty} e^{-t}\, \left\| e^{-\alpha\, L\, t}(r - s) \right\| \, \mathrm{d}t, \tag{9}$$

where $\alpha > 0$ corresponds to the diffusion constant in (5). As we mentioned in the discussion following (8), the distance $d^L_{\text{sps}}(r, s)$ defines a similarity between r and s that incorporates the underlying network structure. Indeed, notice that the term inside the input norm corresponds to the difference $r(t) - s(t)$ between the vectors that solve (5) for initial conditions r and s [cf. (6)]. This means that we are looking at the difference between the temperatures $r(t)$ and $s(t)$ at time t, which we then multiply by the dampening factor e^{-t} and integrate over all times. These temperatures are similar if r and s are similar, or, if r and s have similar values at similar nodes. The dampening factor gives more relative importance to the differences between

$r(t)$ and $s(t)$ for early times. This is necessary because after prolonged diffusion times the network settles into an isothermal equilibrium and the structural differences between r and s are lost.

Exploiting the same interpretation, we can define the superposition norm of a vector $v \in \mathbb{R}^n$ for a given graph with Laplacian matrix L and a given input norm $\|\cdot\|$ as

$$\|v\|_{\text{sps}}^L := \int_0^{+\infty} e^{-t} \left\|e^{-\alpha\, L\, t} v\right\| \, \mathrm{d}t. \tag{10}$$

Although we are referring to $d_{\text{sps}}^L(r,s)$ as the superposition distance between r and s and $\|v\|_{\text{sps}}^L$ as the superposition norm of v we have not proven that they indeed are valid definitions of distance and norm functions. As it turns out, they are. We begin by showing that $\|\cdot\|_{\text{sps}}^L$ is a valid norm as we claim in the following proposition.

Proposition 1: The function $\|\cdot\|_{\text{sps}}^L$ in (10) is a valid norm on $\mathbb{R}^n$ for every Laplacian L and every input norm $\|\cdot\|$.

Proof: As stated in Section II, we need to show positiveness, positive homogeneity and subadditivity of $\|\cdot\|_{\text{sps}}^L$. To show positive homogeneity, utilize the positive homogeneity of the input norm and the linearity of integrals to see that for every vector $v \in \mathbb{R}^n$ and scalar β, it holds

$$\begin{aligned}\|\beta v\|_{\text{sps}}^L &= \int_0^{+\infty} e^{-t} \, \|e^{-\alpha\, L\, t} \beta v\| \, \mathrm{d}t \\ &= |\beta| \int_0^{+\infty} e^{-t} \, \|e^{-\alpha\, L\, t} v\| \, \mathrm{d}t \\ &= |\beta| \, \|v\|_{\text{sps}}^L.\end{aligned} \tag{11}$$

In order to show subadditivity, pick arbitrary vectors $v, w \in \mathbb{R}^n$ and use the subadditivity of the input norm $\|\cdot\|$ and the linearity of integrals to see that

$$\begin{aligned}\|v+w\|_{\text{sps}}^L &= \int_0^{+\infty} e^{-t} \, \|e^{-\alpha\, L\, t} (v+w)\| \, \mathrm{d}t \\ &\leq \int_0^{+\infty} e^{-t} \left(\|e^{-\alpha\, L\, t} v\| + \|e^{-\alpha\, L\, t} w\|\right) \mathrm{d}t \\ &= \|v\|_{\text{sps}}^L + \|w\|_{\text{sps}}^L,\end{aligned} \tag{12}$$

To show positiveness, first observe that for every $v \in \mathbb{R}^n$ we have that $\|v\|_{\text{sps}}^L \geq 0$ since for every time t the argument of the integral in the definition (10) is the product of two nonnegative terms, an exponential and a norm which itself satisfies the positiveness property. The fact that $\|\vec{0}\|_{\text{sps}}^L = 0$ is an immediate consequence of the definition (10). Hence, we are only left to show that $\|v\|_{\text{sps}}^L \neq 0$ for $v \neq \vec{0}$. To show this, it suffices to prove that the argument of the integral in (10) is strictly positive for every time t which is implied by the fact that the matrix $e^{-\alpha\, L\, t}$ is strictly positive definite for every t. To see why this is true, notice that $-\alpha\, L\, t$ is a real symmetric matrix, thus, it is diagonalizable and has real eigenvalues. Consequently, the eigenvalues of $e^{-\alpha\, L\, t}$ are the exponentials of the eigenvalues of $-\alpha\, L\, t$ which are strictly positive. ■

If the superposition norm is a valid norm as shown by Proposition 1 it induces a valid metric as per the construction in Lemma 1. This induced metric is the superposition distance defined in (9) as we show in the following corollary.

Corollary 1: The function d_{sps}^L in (9) is a valid metric on $\mathbb{R}^n$ for every Laplacian L and every input norm $\|\cdot\|$.

Proof: Since $d_{\text{sps}}^L(r,s) = \|r-s\|_{\text{sps}}^L$ for all vectors $r, s \in \mathbb{R}^n$ and $\|\cdot\|_{\text{sps}}^L$ is a well-defined norm [cf. Proposition 1], Lemma 1 implies that d_{sps}^L is a metric on $\mathbb{R}^n$. ■

The distance d_{sps}^L incorporates the network structure to compare two signals r and s supported in a graph with Laplacian L. As a particular case the edge set E of the underlying graph G may be empty. In this case, the Laplacian $L = \mathbf{0}$ is identically null and we obtain from (9) that $d_{\text{sps}}^{\mathbf{0}}(r,s) = \|r-s\|$. This is consistent with the fact that when no edges are present, the network structure adds no information to aid in the comparison of r and s and the superposition distance reduces to the standard distance induced by the input norm. The same effect is obtained when the thermal conductivity α is set to zero.

The computational cost of evaluating the superposition distance is significant in general. To evaluate $d_{\text{sps}}^L(r,s)$ we approximate the improper integral in (9) with a finite sum and evaluate the norm of the matrix exponential $\|e^{-\alpha\, L\, t}(r-s)\|$ at the points required by the appropriate discretization. Notice that the decaying exponential modulation in (9) renders the first time points more relevant for the approximation, thus, a finer discrete time grid should be used for smaller times. An alternative notion of distance for graph-supported signals that is computationally more tractable comes in the form of the diffusion distance that we introduce in the next section.

IV. Diffusion Distance

Given an arbitrary graph $G = (V, E, W)$ with Laplacian L, an input vector norm $\|\cdot\|$ and two signals $r, s \in \mathbb{R}^n$ defined in the node space V, the diffusion distance $d_{\text{diff}}^L(r,s)$ between r and s is given by

$$d_{\text{diff}}^L(r,s) := \left\| \int_0^{+\infty} e^{-t} \, e^{-\alpha\, L\, t} (r-s) \, \mathrm{d}t \right\|, \tag{13}$$

with $\alpha > 0$ corresponding to the diffusion constant in (5). As in the case of the superposition distance in (9), the diffusion distance incorporates the graph structure in determining the proximity between r and s through the solutions $r(t)$ and $s(t)$ of (5) for initial conditions r and s [cf. (6)]. The difference is that in the diffusion distance the input norm of the difference between $r(t)$ and $s(t)$ is taken *after* discounting and integration, whereas in the superposition distance the input norm is applied *before* discounting and integration. An interpretation in terms of heat diffusion is that the diffusion distance compares the total (discounted) energy that passes through each node. The superposition distance compares the energy difference at each point in time and integrates that difference over time. Both are reasonable choices. Computational aspects aside, whether the superposition or diffusion distance is preferable depends on the specific application.

A definite advantage of the diffusion distance is that the matrix integral in (13) can be resolved to obtain a closed solution that is more amenable to computation. To do so, notice that the primitive of the matrix exponential $e^{-t}e^{-\alpha Lt} = e^{-(I+\alpha L)t}$ is given by $-(I+\alpha L)^{-1}e^{-(I+\alpha L)t}$ to conclude that (13) is equivalent to

$$d_{\text{diff}}^{L}(r,s) = \left\|(I+\alpha L)^{-1}(r-s)\right\|. \qquad (14)$$

As in the case of the superposition distance of Section III, a vector norm can be defined based on the same heat diffusion interpretation used to define the distance in (13). Therefore, consider a given a graph with Laplacian L and a given input norm $\|\cdot\|$ and define the diffusion norm of the vector $v \in \mathbb{R}^n$ as

$$\|v\|_{\text{diff}}^{L} := \left\|\int_0^{+\infty} e^{-t}\, e^{-\alpha\, L\, t} v \,\mathrm{d}t\right\| = \left\|(I+\alpha L)^{-1}v\right\|, \quad (15)$$

where the second equality follows from the same primitive expression used in (14). We refer to $(I+\alpha L)^{-1}v$ as the diffused version of vector v.

The diffusion distance is a proper metric and the diffusion norm is a proper norm. We show first that $\|\cdot\|_{\text{diff}}^{L}$ is a valid norm as we formally state next.

Proposition 2: The function $\|\cdot\|_{\text{diff}}^{L}$ in (15) is a valid norm on $\mathbb{R}^n$ for every Laplacian L and every input norm $\|\cdot\|$.

Proof: To prove the validity of $\|\cdot\|_{\text{diff}}^{L}$ we need to show positiveness, positive homogeneity and subadditivity; see Section II. Positive homogeneity follows directly from the positive homogeneity of the input norm, i.e. for any vector $v \in \mathbb{R}^n$ and scalar β we have that

$$\begin{aligned}\|\beta v\|_{\text{diff}}^{L} &= \|(I+\alpha L)^{-1}\beta v\| \\ &= |\beta|\|(I+\alpha L)^{-1}v\| = |\beta|\|v\|_{\text{diff}}^{L}. \end{aligned} \qquad (16)$$

In order to show subadditivity, pick arbitrary vectors $v, w \in \mathbb{R}^n$ and use the subadditivity of the input norm $\|\cdot\|$ to see that

$$\begin{aligned}\|v+w\|_{\text{diff}}^{L} &= \|(I+\alpha L)^{-1}(v+w)\| \\ &\leq \|(I+\alpha L)^{-1}v\| + \|(I+\alpha L)^{-1}w\| \\ &= \|v\|_{\text{diff}}^{L} + \|w\|_{\text{diff}}^{L}. \end{aligned} \qquad (17)$$

Given the positiveness property of the input norm $\|\cdot\|$, to show positiveness of the diffusion norm $\|\cdot\|_{\text{diff}}^{L}$ it is enough to show that $(I+\alpha L)^{-1}v \neq \vec{0}$ for all vectors $v \in \mathbb{R}^n$ different from the null vector. This is implied by the fact that $(I+\alpha L)^{-1}$ is a positive definite matrix. To see why $(I+\alpha L)^{-1}$ is positive definite, first notice that L is positive semidefinite as stated in Section II. Consequently, αL is also positive semidefinite since $\alpha > 0$ and $I+\alpha L$ is positive definite since every eigenvalue of $I+\alpha L$ is a unit greater than the corresponding eigenvalues of αL, thus, strictly greater than 0. Finally, since inversion preserves positive definiteness, the proof is completed. ■

From Proposition 2 and Lemma 1 it follows directly that the diffusion distance defined in (13) is a valid metric as we prove next.

Corollary 2: The function d_{diff}^{L} in (13) is a valid metric on $\mathbb{R}^n$ for every Laplacian L and every input norm $\|\cdot\|$.

Proof: Since $d_{\text{diff}}^{L}(r,s) = \|r-s\|_{\text{diff}}^{L}$ for all vectors $r, s \in \mathbb{R}^n$ and $\|\cdot\|_{\text{diff}}^{L}$ is a well-defined norm [cf. Proposition 2], Lemma 1 implies that d_{diff}^{L} is a metric on $\mathbb{R}^n$. ■

As in the case of the superposition norm and distance, the diffusion norm and distance reduce to the input norm and its induced distance when $\alpha = 0$ or the edge set is empty. In that case we have $L = \mathbf{0}$ and it follows from the definitions in (15) and (13) that $\|v\|_{\text{diff}}^{L} = \|v\|_{\text{diff}}^{\mathbf{0}} = \|v\|$ and that $d_{\text{diff}}^{L}(r,s) = d_{\text{diff}}^{\mathbf{0}}(r,s) = \|r-s\|$. Notice also that for the particular case in which the input norm is $\|\cdot\|_2$, d_{diff}^{L} coincides with the Mahalanobis distance with covariance matrix $(I+\alpha L)^2$ [33].

The superposition and diffusion distances differ in the order in which the input norm and time integral are applied. It is therefore reasonable to expect some relationship to hold between their values. In the following proposition we show that the diffusion distance is a lower bound for the value of the superposition distance.

Proposition 3: Given any graph $G = (V, E, W)$ with Laplacian L, any two signals $r, s \in \mathbb{R}^n$ defined in V and any input vector norm $\|\cdot\|$, the diffusion distance $d_{\text{diff}}^{L}(r,s)$ defined in (13) is a lower bound on the superposition distance $d_{\text{sps}}^{L}(r,s)$ defined in (9)

$$d_{\text{sps}}^{L}(r,s) \geq d_{\text{diff}}^{L}(r,s). \qquad (18)$$

Proof: Since the exponential e^{-t} in (9) is nonnegative, we may replace it with its absolute value to obtain

$$\begin{aligned} d_{\text{sps}}^{L}(r,s) &= \int_0^{+\infty} \left|e^{-t}\right| \left\|e^{-\alpha\, L\, t}(r-s)\right\| \,\mathrm{d}t \\ &= \int_0^{+\infty} \left\|e^{-t}e^{-\alpha\, L\, t}(r-s)\right\| \,\mathrm{d}t, \end{aligned} \qquad (19)$$

where we used the positive homogeneity property of the input norm to write the second equality. Further using the subadditivity property of the input norm we may write

$$d_{\text{sps}}^{L}(r,s) \geq \left\|\int_0^{+\infty} e^{-t}e^{-\alpha\, L\, t}(r-s)\,\mathrm{d}t\right\|. \qquad (20)$$

The right hand side of (20) is the definition of the diffusion distance $d_{\text{diff}}^{L}(r,s)$ in (13). Making this substitution in (20) yields (18). ■

For applications in which the superposition distance is more appropriate, the diffusion distance is still valuable because, as it follows from Proposition 3, it can be used as a lower bound on the superposition distance. This lower bound is useful because computing the diffusion distance is less expensive than computing the superposition distance.

A. Discussion

In order to illustrate the superposition and diffusion distances and their difference with the standard vector distances, consider the undirected graph in Fig. 1 where the weight of each undirected edge is equal to 1. Define three different vectors supported in the node space and having exactly one component equal to 1 and the rest equal to 0. The vector r has its positive component for node x_1, colored in red, the vector g has its positive for node x_6, colored in green, and the vector y has its positive component for node x_7, colored in yellow.

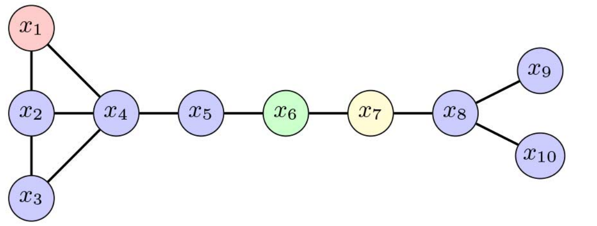


Fig. 1. Example of an underlying graph used to compute the superposition and diffusion distances. Three signals r, g and y are compared taking a value of 1 in the red, green, and yellow nodes, respectively, and zero everywhere else.

For the traditional vector metrics, the distances between each of the vectors r, g and y are the same. In the case when, e.g., the ℓ_2 distance is used as input metric, we have that $\|r - g\|_2 = \|g - y\|_2 = \|y - r\|_2 = \sqrt{2}$. In the case of the ℓ_1 and ℓ_∞ distances we have that $\|r - g\|_1 = \|g - y\|_1 = \|y - r\|_1 = 2$ and $\|r - g\|_\infty = \|g - y\|_\infty = \|y - r\|_\infty = 1$. However, by observing the network in Fig. 1, it is intuitive that signals g and y should be more alike than they are to r since they affect nodes that are closely related. E.g., if we think of the vectors r, g and y as signaling faulty nodes in a communication network, it is evident that the impact of nodes x_6 and x_7 failing would disrupt the communication between the right and left components of the graph, whereas the failure of x_1 would entail a different effect. This intuition is captured by the diffusion and superposition distances. Indeed, if we fix $\alpha = 1$ and we use the ℓ_2 norm as input norm to the diffusion distance, we have that the distance between the vectors that signal faults at x_6 and x_7 are [cf. (14)]

$$d^L_{\text{diff}}(g, y) = \|(I + L)^{-1}(g - y)\|_2 = 0.418, \tag{21}$$

where L is the Laplacian of the graph in Fig. 1. However, the diffusion distances from these green and yellow vectors to the red vector that signals a fault at node x_1 are

$$\begin{aligned} d^L_{\text{diff}}(r, g) &= \|(I + L)^{-1}(r - g)\|_2 = 0.664, \\ d^L_{\text{diff}}(r, y) &= \|(I + L)^{-1}(r - y)\|_2 = 0.698. \end{aligned} \tag{22}$$

The distances in (22) are larger than the distance in (21) signaling the relative similarity of the g and y vectors with respect to the r vector. The differences are substantial–almost 60% increase–, thus allowing identification of g and y as somehow separate from r. Further observe that the distance between r and g is slightly smaller than the distance between r and y. This is as it should be, because node x_1 is closer to node x_6 than to node x_7 in the underlying graph.

Repeating the exercise, but using the superposition distance instead [cf. (9)], we obtain that $d^L_{\text{sps}}(r, g) = 0.701$, $d^L_{\text{sps}}(r, y) = 0.742$, and $d^L_{\text{sps}}(g, y) = 0.456$. Although the numbers are slightly different, the qualitative conclusions are the same as those obtained for the diffusion distance. We can tell that g and y are more like each other than they are to r, and we can tell that g is slightly closer to r than y is. Also note that the diffusion distances are smaller than the superposition distances between the corresponding pairs, i.e., $d^L_{\text{sps}}(r, g) \geq d^L_{\text{diff}}(r, g)$, $d^L_{\text{sps}}(r, y) \geq d^L_{\text{diff}}(r, y)$, and $d^L_{\text{sps}}(g, y) \geq d^L_{\text{diff}}(g, y)$. This is consistent with the result in Proposition 3.

To further illustrate the intuitive idea behind the diffusion and superposition distances, Fig. 2 plots the evolution of the diffused signals $r(t)$, $g(t)$ and $y(t)$ for each of the respective initial conditions r, g, and y. At time $t = 0$ each of the signals is concentrated at one specific node. The signals are, as a consequence, equally different to each other. At very long times, the signals are completely diffused and therefore indistinguishable. For intermediate times, the signal distributions across nodes for the green and yellow signals are more similar than between the green and red or yellow and red signals. This difference between the evolution of the diffused signals results in different values for the superposition and diffusion distances.

Remark 1: Computation of the diffusion distance using the closed form expression in (14) requires the inversion of the $n \times n$ identity plus Laplacian matrix followed by multiplication with the difference vector $r - s$. The cost of this computation is of order n^3, but is much smaller when the matrix L is sparse, as is typically the case. Further observe that most computations can be reused when computing multiple distances, because the vectors change, but the matrix inverse $(I + \alpha L)^{-1}$ stays unchanged.

V. Stability

The superposition and diffusion distances depend on the underlying graphs through their Laplacian L. It is therefore important to analyze how a perturbation of the underlying network impacts both distances. We prove in this section that these distances are well behaved with respect to perturbations of the underlying graph. I.e., we show that if the network perturbation is small, the change in the diffusion and superposition distances is also small. We think of a perturbation of a given network as noise added to its edge weights, thus, we quantify the network perturbation as the matrix p-norm of the difference between the Laplacians of the original and perturbed networks. We focus our analysis on the most frequently used norms where $p \in \{1, 2, \infty\}$. We begin with a formal statement for the case of the superposition distance defined by (9).

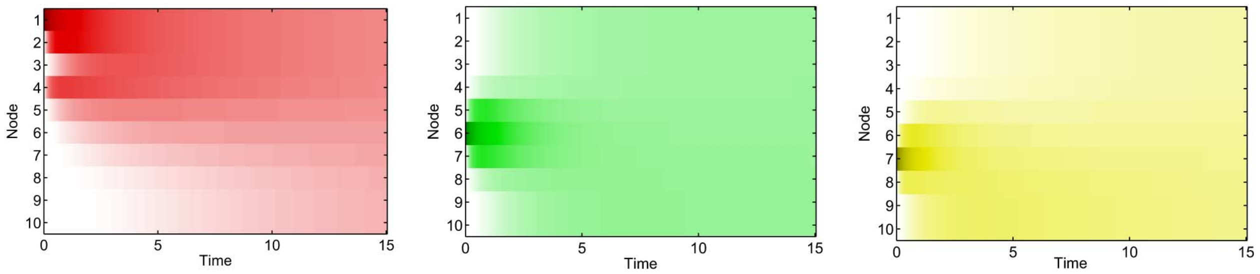
Proposition 4: Given any graph with Laplacian L, an input ℓ_p norm $\|\cdot\|_p$ with $p \in \{1, 2, \infty\}$, and bounded signals s and r on the network with $\|s\|_p \leq \gamma$ and $\|r\|_p \leq \gamma$, if we perturb the network such that the resulting Laplacian $L' = L + E$ where the perturbation E is such that $\|E\|_p \leq \epsilon \|L\|_p < 1$, then

$$\left\| d^{L'}_{\text{sps}}(s, r) - d^L_{\text{sps}}(s, r) \right\| \leq 2\gamma \|L\|_p \epsilon. \tag{23}$$

Proof: See Appendix A. ■

Proposition 4 guarantees that for any two vectors, the difference between their superposition distances computed based on different underlying graphs is bounded by a term which is bilinear in a bound on the magnitude of the input vectors γ and a bound on the difference between the Laplacians of both underlying graphs $\|E\|_p \leq \epsilon \|L\|_p$. This implies that vanishing perturbations on the underlying network have vanishing effects on the distance between two signals defined on the network.

Similarly to the case of the superposition distance, perturbations have limited effect on the diffusion metric defined in (13) as shown next.



(a) Diffusion of r

(b) Diffusion of g

(c) Diffusion of y

Fig. 2. Heat maps of the diffused signals for r, g, and y as diffusion evolves for every node in the network in Fig. 1. Darker colors represent stronger signals. The heat maps of g and y are more similar, entailing smaller diffusion and superposition distances.

Proposition 5: For the same setting described in Proposition 4, we have that

$$\left\| d^{L'}_{\text{diff}}(s,r) - d^{L}_{\text{diff}}(s,r) \right\| \leq 2\gamma \|L\|_p \epsilon + o(\epsilon). \tag{24}$$

Proof: See Appendix B. ■

In contrast to Proposition 4, the bound in (24) contains higher order terms that depend on the magnitude of the perturbation. Hence, since the other terms of the bound in (24) tend to zero super linearly, we may divide (24) by $\epsilon\|L\|_p$ and compute the limit as the perturbation vanishes

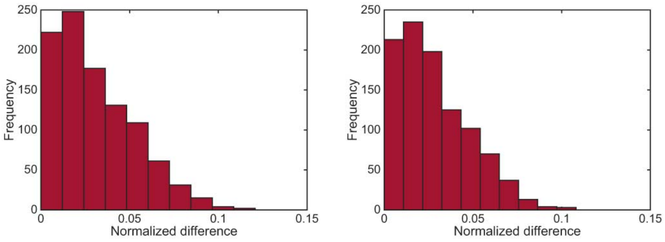
$$\lim_{\epsilon \to 0} \frac{\left\| d^{L'}_{\text{diff}}(s,r) - d^{L}_{\text{diff}}(s,r) \right\|}{\epsilon \|L\|_p} \leq 2\gamma, \tag{25}$$

which implies that for small perturbations the difference in diffusion distances grows linearly.

When constructing the underlying graph to compare signals in a real-world application, noisy information can be introduced. This means that the similarity weight between two nodes in the underlying graph contains inherent error. Propositions 4 and 5 show that the superposition and diffusion distances are impervious to these minor perturbations.

In order to illustrate the stability results presented, consider again the underlying network in Fig. 1. We perturb this network by multiplying every edge weight–originally equal to 1–by a random number uniformly picked from $[0.95, 1.05]$ and then compute the diffusion and superposition distances between vectors r and g with the perturbed graph as underlying network. For these illustrations we pick the input norm to be ℓ_2 and observe that $\gamma = 1$ given the definitions of r and g. In Fig. 3 we plot histograms of the absolute value of the difference in the distances when using the original and the perturbed graphs as underlying networks normalized by the norm of the perturbation for 1000 repetitions of the experiment. From (23) we know that this value should be less than 2 for the superposition distance and from (25) we know this should also be the case for the diffusion distance for vanishing perturbations. Indeed, as can be seen from Fig. 3, all perturbations are below the threshold of 2 by a considerable margin. This stability property is essential for the practical utility of the diffusion and superposition distances as seen in the next section.

Remark 2: In Propositions 4 and 5 we focus our analysis on the input norms $\|\cdot\|_p$ for $p \in \{1, 2, \infty\}$ because these norms lead to the simple bounds in (23) and (24). The simplicity of these bounds is derived from the fact that $\|e^{-Lt}\|_p \leq 1$ and $\|(I+L)^{-1}\|_p \leq 1$ for the values of p previously mentioned. For other matrix norms satisfying (3) and (4), including all induced matrix norms, the equivalence of norms guarantees that bounds analogous to those in (23) and (24) must exist, but with more involved constant terms.



(a) Diffusion distance

(b) Superposition distance

Fig. 3. Histogram of the absolute value of the normalized difference, i.e. $|d^{L'}(g,r) - d^{L}(g,r)|/\|E\|_2$, for the diffusion and superposition distances. For this particular network and perturbations, the difference is considerably lower than the theoretical upper bound of 2.

VI. Applications

We illustrate the advantages of the superposition and diffusion distances developed in Sections III and IV respectively through numerical experiments in both synthetic (Section VI-A) and real-world data (Sections VI-B and VI-C).

A. *Classification of synthetic signals on networks*

The diffusion and superposition distances lead to better classification of signals on networks compared to traditional vector distances such as the Euclidean ℓ_2 metric. Consider the network presented in Fig. 4(a) containing three clusters–blue, red, and green–where nodes within each cluster are highly connected and there exist few connections between nodes in different clusters. This network was generated randomly, where an undirected edge between a pair of nodes in the same cluster is formed with probability 0.4 and its weight is picked uniformly between 1 and 3. In addition, three edges were added with weight 1 between random pairs of nodes in different clusters. We consider three types of signals on this network. The strength of all signals is equal to 1 on three nodes in the network and 0 on the remaining ones. Among the three nodes with value 1

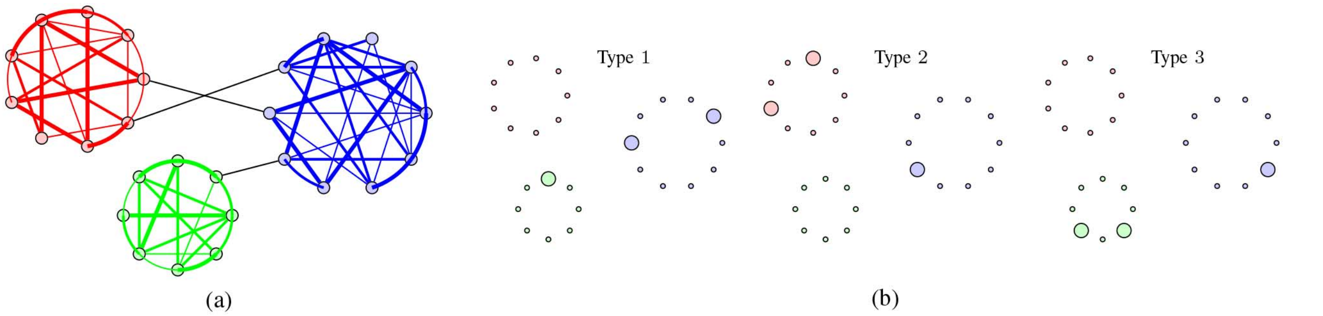


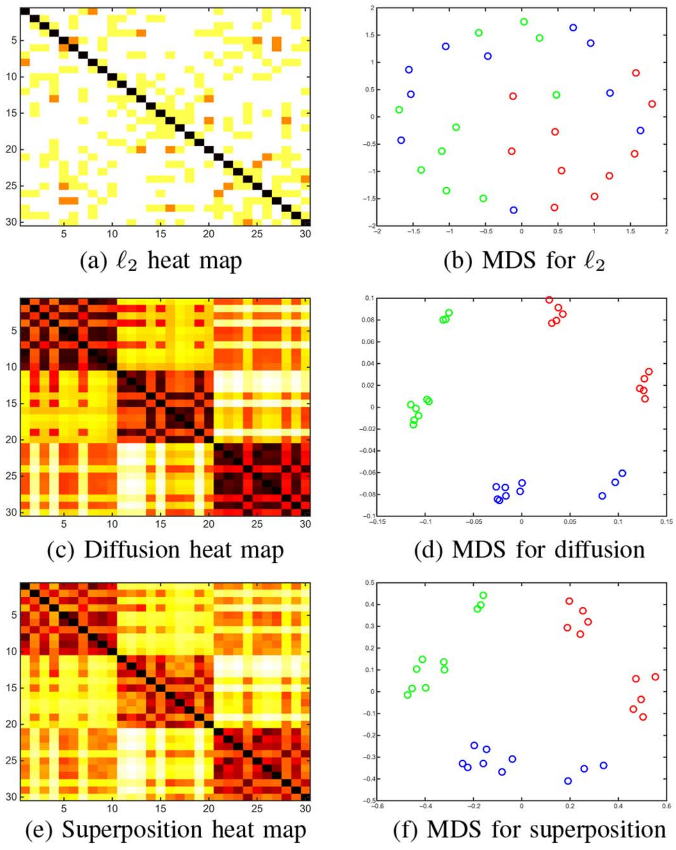
Fig. 4. (a) Three-cluster network on which signals to be classified are defined. The width of the links is proportional to the weights of the corresponding edges. (b) Sample signals for the three types considered. Type 1 signals have stronger presence in the blue cluster, type 2 in the red, and type 3 in the green cluster.

for the first type of signals, two of them are randomly selected from the blue cluster and the remaining one is randomly chosen from the other clusters. Similarly, for the second type of signals, exactly two out of the three nodes with positive value belong to the red cluster and the remaining one is chosen randomly between the blue and green clusters. Finally, the third type of signal has two positive values on the green cluster and the third value randomly chosen from the rest of the network. Sample signals for each type are illustrated in Fig. 4(b) where positive signal values are denoted by larger nodes.

We generate ten signals of each type and measure the distance between them with the superposition, diffusion, and ℓ_2 metrics. For the superposition and diffusion metrics we use ℓ_2 as input norm and $\alpha = 1$. The use of each metric generates a different metric space with the thirty signals as the common underlying set of points. In order to illustrate these higher dimensional spaces, in Fig. 5 (left) we present heat maps of the distance functions, where darker colors represent closer signals. It is clear that for the diffusion and superposition distances, three blocks containing ten points each appear along the diagonal in exact correspondence with the three types of signals. In contrast, the heat map corresponding to the ℓ_2 metric does not present any clear structure. To further illustrate these implications, in Fig. 5 (right) we present 2D multi dimensional scaling (MDS) [34] representations of the three metric spaces. The points corresponding to type 1 signals are represented as blue circles, type 2 as red circles, and type 3 as green circles. The MDS representations for diffusion and superposition are fundamentally different from the one obtained for ℓ_2. For the latter, the circles of different colors are spread almost randomly on the plane, with no clear clustering structure. For diffusion and superposition, in contrast, signals of different colors are clearly separated so that any clustering method is able to recover the original signal type.

B. Ovarian cancer histology classification

We demonstrate that the diffusion distance can provide a better classification of histology subtypes for ovarian cancer patients than the traditional ℓ_2 metric. To do this, we consider 240 patients diagnosed with ovarian cancer corresponding to two different histology subtypes [35]: serous and endometrioid. Our objective is to recover the histology subtypes from patients' genetic profiles.



(a) ℓ_2 heat map (b) MDS for ℓ_2

(c) Diffusion heat map (d) MDS for diffusion

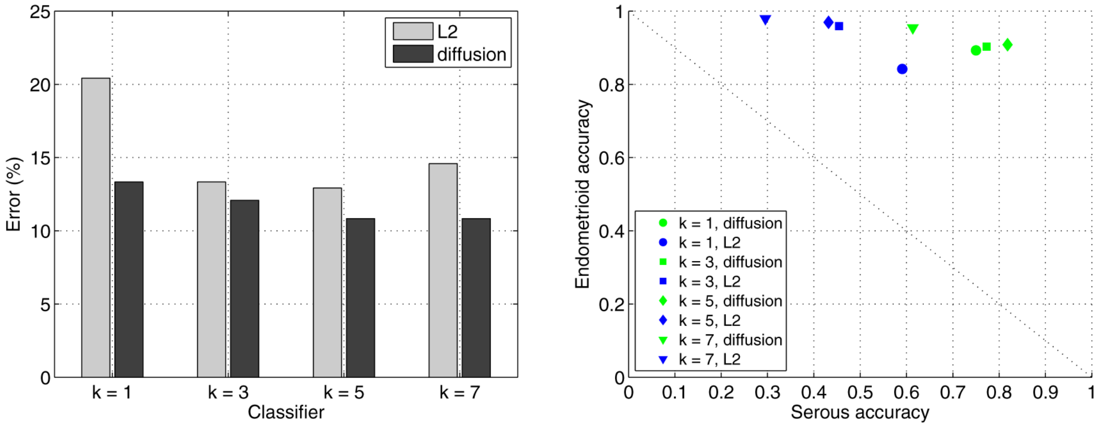
(e) Superposition heat map (f) MDS for superposition

Fig. 5. Heat maps (left) and 2D multi dimensional scaling (MDS) [34] representations (right) for the metric spaces generated by the ℓ_2 (top), diffusion (middle) and superposition (bottom) distances. The diffusion and superposition metrics perfectly classify the signals into the three types while ℓ_2 does not reveal any clear classification.

For each patient i, her genetic profile consists of a binary vector $v_i \in \{0,1\}^{2458}$ where, for each of the 2458 genes studied, v_i contains a 1 in position k if patient i presents a mutation in gene k and a 0 otherwise. One way of building a metric in the space of 240 patients is by quantifying the distance between patients i and j as the ℓ_2 distance between their genetic profiles,

$$d_{\ell_2}(i,j) = \|v_i - v_j\|_2. \tag{26}$$

In this approach, every gene is considered orthogonal to each other and compared separately across patients. An alternative approach is to take into account the relational information



(a)

(b)

Fig. 6. Histology classification of ovarian cancer patients based on k nearest neighbors with respect to the ℓ_2 and diffusion distances of their genetic profile. (a) Light bars denote the error when patients are classified using the ℓ_2 distance while the dark bars denote the error when diffusion distance is used for different k-NN classifiers. The diffusion distance reduces the classification error consistently across classifiers. (b) Accuracy of serous subtype vs. endometrioid subtype. Classifiers using diffusion (green) are closer to the top right corner, i.e. perfect classification, than those using the ℓ_2 distance (blue).

across genes when comparing patients. In order to do so, we apply the diffusion distance on an underlying gene-to-gene network built based on publicly available data [36]. In order to build this network, we first extract the pairwise gene-gene interactions from [36] using the *NCI_Nature* database. After normalization, every edge weight is contained between 0 and 1, which we interpret as a probability of interaction between genes. We assign to each path the probability obtained by multiplying the probabilities in the edges that form the path. For every pair of genes in the network, we compute a similarity value between them corresponding to the maximum probability achievable by a path that links both genes. Finally, we apply normalization and thresholding operations to obtain the gene-to-gene network that we use in our experiments. Observe that the gene-to-gene network contains accepted relations between genes in humans in general and is not patient dependent, hence, it defines a common underlying network for all subjects being compared. Thus, denoting as L the Laplacian of the gene-to-gene network and using the ℓ_2 as input norm we compute the diffusion distances between patients i and j as [cf. (14)]

$$d^L_{\text{diff}}(i,j) = \|(I+\alpha L)^{-1}(v_i - v_j)\|_2, \qquad (27)$$

where α was set to 15, however, results are robust to this particular choice. Given that in Section VI-A we obtained similar performance between the diffusion and superposition distances, combined with the fact that the latter is computationally expensive, we do not implement the superposition distance in this data set.

In order to evaluate the classification power of both approaches–ℓ_2 and diffusion distance–we perform 240-fold cross validation for a k nearest neighbors (k-NN) classifier. More precisely, for a particular patient, we look at the k nearest patients as given by the metric being evaluated and assign to this patient the most common cancer histology among the k nearest patients. We then compare the assigned histology with her real cancer histology and evaluate the accuracy of the classifier. Finally, we repeat this process for the 240 women considered and obtain a global classification accuracy for both approaches.

In Fig. 6(a) we show the reduction in histology classification error when using the diffusion distance (27) compared to using the ℓ_2 distance (26) when comparing genetic profiles. The four groups of bars correspond to classifiers built using different numbers of neighbors $k \in \{1,3,5,7\}$. Notice that the reduction in error is consistent across all classifiers analyzed with an average error reduction of over 21%, unveiling the value of incorporating the network information in the classification process.

To further analyze the obtained results, in Fig. 6(b) we present the accuracy obtained for the serous subtype versus the accuracy obtained for the endometrioid subtype for different classifiers based on the diffusion (green) and ℓ_2 (blue) distances. Points on the top right corner of the plot are ideal, obtaining perfect classification for both subtypes. When using diffusion, accuracies shift towards the ideal position since the accuracies for the serous subtypes increase by 20% to 40% whereas the accuracies for endometrioid subtypes decrease by less than 5%. Furthermore, among the 240 patients analyzed, there are 196 of them with endometrioid subtype and only 44 with serous subtype. Hence, a nearest neighbor classifier based on an uninformative distance would tend to have a high classification accuracy for the former but a low one for the latter. This is the case for the ℓ_2 metric. The diffusion distance, in contrast, by exploiting the gene-to-gene interaction can overcome this limitation.

C. Handwritten digit recognition

Diffusion distance can be instrumental in the classification of digits via semi-supervised learning. To illustrate this, consider the well-known MNIST handwritten digit database [37]. Each observation consists of a square gray-scaled image of a handwritten digit with 28 × 28 pixels. Consequently, we can think of each observation as a vector $x \in \mathbb{R}^{784}$ where the value of each component corresponds to the intensity of the associated pixel. A subset of these images–the training set–are labeled, i.e. we know the digit that the image represents. The rest of the images–the testing set–are unlabeled and our objective is to correctly identify the digits they represent. Given n the total number of images–labeled or unlabeled–, we define $X \in \mathbb{R}^{784\times n}$ as $X = [x_1, x_2, \ldots, x_n]$ so that each column in X corresponds to the pixels of one digit.

K nearest neighbors is a simple conventional approach used to classify the digits. In order to implement it, we first compute the ℓ_2 pairwise distance between all the vectors x_i. Equivalently, if we denote by e_i the i-th canonical vector–all entries of e_i are zero except the i-th entry which is 1–the ℓ_2 distance between digits i and j can be written as

$$d_{\ell_2}(i,j) = \|X(e_i - e_j)\|_2. \tag{28}$$

To obtain the estimated label of an image in the testing set, we look at the labels of the k closest images among those in the training set as given by (28) and pick the mode of these labels, i.e., the most popular one.

An alternative k-NN approach can be designed using diffusion by defining a graph G_τ whose nodes are the handwritten digits. To do this, we draw an edge–with weight 1–between two digits i and j in G_τ if the ℓ_2 pairwise distance (28) is less than a threshold τ. We can interpret digit i as being represented by the signal e_i on G_τ, with value 1 at node i and 0 elsewhere. The diffused version of e_i is given by $(I + \alpha L_\tau)^{-1} e_i$ [cf. (15)] where L_τ is the Laplacian of G_τ. We can then quantify the distance between two diffused digits i and j as

$$d_{\text{diff}}^{L_\tau}(i,j) = \|X(I + \alpha L_\tau)^{-1}(e_i - e_j)\|_2. \tag{29}$$

We can then train a k-NN classifier based on the distance between the diffused digits and compare the results with the conventional k-NN based on the ℓ_2 distance without diffusion. Notice that $d_{\text{diff}}^{L_\tau}(i,j)$ reduces to $d_{\ell_2}(i,j)$ when $L_\tau = \mathbf{0}$.

In Fig. 8 we present the attribution error comparison between both approaches when performing a binary attribution task between hard-to-distinguish digits: 3 vs. 5, 3 vs. 8, and 5 vs. 8. For each of these cases, we use the entire MNIST training set and testing set with $k \in \{3, 5, 7\}$. It is immediate to see that the diffusion approach outperforms the traditional k-NN in the three tasks. To see why this is the case, in Fig. 7 (top) we present two handwritten images that correspond to threes but are misclassified as fives by the traditional k-NN method. In Fig. 7 (bottom) we present their representations after diffusion in G_τ. It is clear that diffusion averages out irregularities found in particular handwritten digits and drives them towards a canonical representation of the number 3.

If we replicate the comparison for a ten class classification problem, i.e. for all digits between 0 and 9, diffusion still improves the accuracy by reducing the error rates from 4.43% to 4.21% (training set of 8600 digits, testing set of 1400 digits and $k = 3$). Moreover, further accuracy improvements can be obtained by combining the traditional and the diffused k-NN methods by choosing the most popular label among the k nearest neighbors in the traditional approach and the $k+1$ nearest neighbors in the diffused approach. The error rate is further reduced to 3.93%. We pick k neighbors from one approach and $k+1$ from the other to obtain an odd total number of neighbors, reducing the possibility of a multimodal distribution of labels.

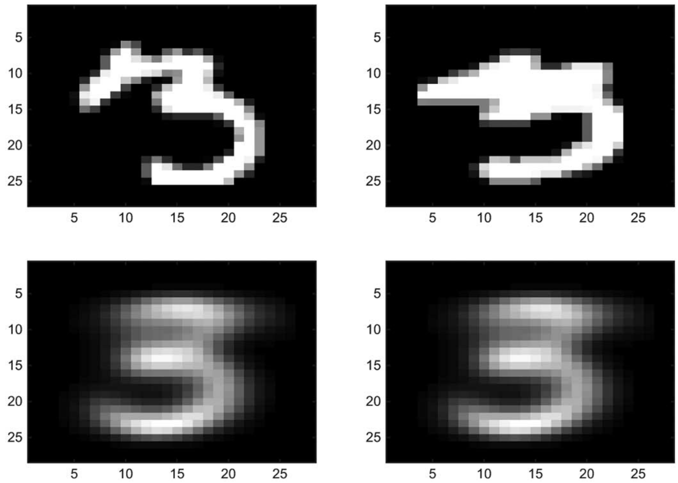


Fig. 7. Two instances of handwritten threes (top) which are interpreted as fives by the classical k-NN approach and their corresponding diffused image (bottom). Diffusion averages out irregularities, achieving higher classification accuracy.

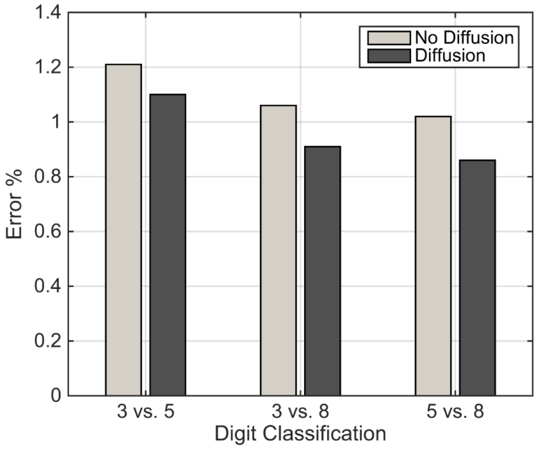


Fig. 8. Error rates for three binary classification problems of written digits given by the traditional and diffused k-NN approaches. Error is reduced by diffusion in the three cases.

For the cases where $k \in \{5, 7\}$, similar results are obtained where we see still see the benefit of using diffusion which is further boosted by combining the traditional and the diffused k-NN methods.

Notice that this application of the diffusion distance is fundamentally different from the one presented in Section VI-B. In the ovarian cancer case, the nodes in the network represent genes and each signal on the network represents a patient. In contrast, in the current case, both the nodes in the network and the signals represent handwritten digits. This approach can be used in general for label propagation problems in graphs.

VII. Conclusion

The superposition and diffusion distances, as metrics to compare signals in networks, were introduced. Both metrics rely on the temporal heat map induced by the diffusion of signals across the network. The superposition distance quantifies the instantaneous difference between the diffused signals while the diffusion distance evaluates the accumulated effect across time. Both distances were shown to be stable with respect to perturbations in the underlying network, however, due to its closed form, the diffusion distance was found to be more suitable for implementation. We showed how both distances can be used to obtain a better classification of signals in networks both in synthetic settings as well as in a real-world classification of cancer histologies. Finally, we illustrated the use of diffusion as part of a label propagation process to classify handwritten digits.

Appendix A

Proof of Proposition 4

The next lemma is central to the proof of Proposition 4.

Lemma 2: Given the Laplacian L for some undirected network, the matrix exponential of nonpositive multiples of the Laplacian $e^{-\tau L}$ with $\tau \geq 0$ is a doubly stochastic matrix.

Proof: Since $L = D - A$, all off-diagonal components of $-\tau L$ are nonnegative, making $-\tau L$ a Metzler matrix [38]. Since the exponentials of Metzler matrices are nonnegative [38, Theorem 8.2], we are guaranteed that all elements of $e^{-\tau L}$ are nonnegative. From the power series of matrix exponentials, we have

$$e^{-\tau L} = \sum_{k=0}^{\infty} \frac{1}{k!}(-\tau L)^k = I - \tau L + \frac{\tau^2 L^2}{2} - \frac{\tau^3 L^3}{3!} + \cdots . \tag{30}$$

If we are able to show that all rows and columns of L^k add up to 0 for any integer $k \geq 1$, then we know that all rows and columns of $\sum_{k=1}^{\infty}(-\tau L)^k/k!$ also add up to 0. Therefore, when we add the identity matrix to this summation to obtain the exponential $e^{-\tau L}$ as in (30) we are guaranteed that the rows and columns sum up to 1. Combining this with the non negativity of $e^{-\tau L}$ implies doubly stochasticity, as wanted. To see that the rows and columns of L^k indeed add up to 0 for any integer $k \geq 1$, denote by $\vec{1}$ and $\vec{0}$ the vectors of all-ones and all-zeros, respectively. Then, by the definition of the graph Laplacian (1), it follows that $\vec{1}^T L = L\vec{1} = \vec{0}$ which immediately implies that $\vec{1}^T L^k = L^k \vec{1} = \vec{0}$ for all $k \geq 1$. ■

We now use Lemma 2 to show Proposition 4.

Proof of Proposition 4: Given the definition of L', from (9) we have that

$$d_{\text{sps}}^{L'}(s,r) = \int_0^{\infty} e^{-t} \left\| e^{-(L+E)t}(s-r) \right\|_p dt, \tag{31}$$

where without loss of generality we assume $\alpha = 1$. If $\alpha \neq 1$, then $\alpha L'$ defines a Laplacian and we can think of the distance $d_{\text{sps}}^{\alpha L'}(s,r)$ where the new α parameter is equal to 1. If we focus on the input norm $\|\cdot\|_p$ inside the integral in (31), we may add and subtract $e^{-Lt}(s-r)$ to obtain

$$\begin{aligned} &\left\| e^{-(L+E)t}(s-r) \right\|_p \\ &\quad = \left\| \left(e^{-(L+E)t} - e^{-Lt}\right)(s-r) + e^{-Lt}(s-r) \right\|_p \\ &\quad \leq \left\| \left(e^{-(L+E)t} - e^{-Lt}\right)(s-r) \right\|_p + \left\| e^{-Lt}(s-r) \right\|_p , \end{aligned} \tag{32}$$

where we used the subadditivity property of the input norm. To further bound the first term on the right hand side of (32) we apply the compatibility property of p-norms (4) followed by the subadditivity property to obtain that

$$\begin{aligned} &\left\| \left(e^{-(L+E)t} - e^{-Lt}\right)(s-r) \right\|_p \\ &\quad \leq \left\| e^{-(L+E)t} - e^{-Lt} \right\|_p \|(s-r)\|_p \\ &\quad \leq \left\| e^{-(L+E)t} - e^{-Lt} \right\|_p (\|s\|_p + \|r\|_p) . \end{aligned} \tag{33}$$

In order to bound the first term on the right hand side of (33), we use a well-known result in matrix exponential analysis [30], [39] that allows us to write the difference of matrix exponentials in terms of an integral,

$$\begin{aligned} \left\| e^{-(L+E)t} - e^{-Lt} \right\|_p &= \left\| \int_0^t e^{-L(t-\tau)} E e^{-(L+E)\tau} d\tau \right\|_p \\ &\leq \int_0^t \left\| e^{-L(t-\tau)} E e^{-(L+E)\tau} \right\|_p d\tau \\ &\leq \|E\|_p \int_0^t \left\| e^{-L(t-\tau)} \right\|_p \\ &\qquad \times \left\| e^{-(L+E)\tau} \right\|_p d\tau, \end{aligned} \tag{34}$$

where the first inequality follows from subadditivity of the input p-norm and the second one from submultiplicativity (3).

We now bound each of the three terms on the right hand side of (34). For the first term, $\|E\|_p \leq \epsilon \|L\|_p$ by assumption. From Lemma 2, the doubly stochasticity of $e^{-L(t-\tau)}$ implies that $\|e^{-L(t-\tau)}\|_1 = \|e^{-L(t-\tau)}\|_\infty = 1$. For $p = 2$, notice that $-L(t-\tau)$ is a negative semi-definite matrix with an eigenvalue at 0. Since the eigenvalues of $e^{-L(t-\tau)}$ are equal to the exponentials of the eigenvalues of $-L(t-\tau)$, it follows that the largest eigenvalue of $e^{-L(t-\tau)}$ is 1 and hence $\|e^{-L(t-\tau)}\|_2 = 1$. For the term $\left\|e^{-(L+E)\tau}\right\|_p$, notice that $L + E = L'$ is in itself a Laplacian, meaning that we can follow the aforementioned argument and upper bound this term by 1. Substituting these bounds in (34) and solving the integral yields

$$\left\| e^{-(L+E)t} - e^{-Lt} \right\|_p \leq \epsilon \|L\|_p \, t. \tag{35}$$

Further substitution in (33) combined with the fact that $\|s\|_p \leq \gamma$ and $\|r\|_p \leq \gamma$, results in

$$\left\| \left(e^{-(L+E)t} - e^{-Lt}\right)(s-r) \right\|_p \leq 2\gamma\epsilon \|L\|_p \, t. \tag{36}$$

By substituting this result in (32) and inputing the resultant inequality in the integral in (31) we conclude that

$$\begin{aligned} d_{\text{sps}}^{L'}(s,r) \leq & \int_0^{\infty} t e^{-t} 2\gamma\epsilon \|L\|_p dt \\ &+ \int_0^{\infty} e^{-t} \left\| e^{-Lt}(s-r) \right\|_p dt. \end{aligned} \tag{37}$$

Notice that the rightmost summand in (37) is exactly equal to $d_{\text{sps}}^{L}(r,s)$ [cf. (9)]. Thus, solving the integral in the first summand we get that

$$d_{\text{sps}}^{L'}(s,r) - d_{\text{sps}}^{L}(s,r) \leq 2\gamma\epsilon \|L\|_p. \tag{38}$$

Following the same methodology but starting from the definition of $d^L_{\text{sps}}(s,r)$, it can be shown that

$$d^L_{\text{sps}}(s,r) - d^{L'}_{\text{sps}}(s,r) \leq 2\gamma\epsilon\|L\|_p. \tag{39}$$

Finally, by combining (38) and (39), we obtain (23), concluding the proof. ■

Appendix B

Proof of Proposition 5

In the proof of Proposition 5 we use two lemmas. The first one is similar to Lemma 2 and shows that $(I+L)^{-1}$ is doubly stochastic.

Lemma 3: Given the Laplacian L for some undirected network, the inverse of the Laplacian plus identity matrix $(I+L)^{-1}$ is a doubly stochastic matrix.

Proof: Since all the off-diagonal entries of $I+L$ are less than or equal to zero, $I+L$ is a Z-matrix [40]. Moreover, due to the fact that all eigenvalues of $I+L$ have positive real parts, $I+L$ is an M-matrix. Since the inverse of an M-matrix is elementwise nonnegative [41], $(I+L)^{-1}$ is a nonnegative matrix. Thus, to show doubly stochasticity, we only need to prove that all rows and columns of $(I+L)^{-1}$ add up to 1. Recall that $\vec{1}$ and $\vec{0}$ stand for the vectors of all-ones and all-zeros, respectively, and that $L\vec{1}=\vec{0}$ [cf. (1)] Thus, we may write $(I+L)\vec{1}=\vec{1}$ from which we have that

$$\vec{1} = (I+L)^{-1}\,(I+L)\vec{1} = (I+L)^{-1}\,\vec{1}, \tag{40}$$

showing that all the rows of $(I+L)^{-1}$ sum up to 1. Similarly, it can be shown that all the columns of $(I+L)^{-1}$ sum up to 1, concluding the proof. ■

The second lemma is a statement about the stability of inverse matrices.

Lemma 4: If A is nonsingular and $\|A^{-1}E\|_p < 1$, then $A+E$ is nonsingular and it is guaranteed that

$$\left\|(A+E)^{-1} - A^{-1}\right\|_p \leq \frac{\|E\|_p\|A^{-1}\|_p^2}{1-\|A^{-1}E\|_p}. \tag{41}$$

Proof: See [32, Theorem 2.3.4]. ■

We now use Lemmas 3 and 4 to show Proposition 5.

Proof of Proposition 5: Given the definition of L', from (14) we have that

$$d^{L'}_{\text{diff}}(s,r) = \left\|(I+L+E)^{-1}(s-r)\right\|_p. \tag{42}$$

As in the proof of Proposition 4, we can assume that $\alpha=1$ without loss of generality. Subtracting and adding $(I+L)^{-1}(s-r)$ from (42) and applying the subadditivity property of the p-norm implies

$$\begin{aligned} d^{L'}_{\text{diff}}(s,r) \leq{}& \left\|\left((I+L+E)^{-1}-(I+L)^{-1}\right)(s-r)\right\|_p \\ &+ \left\|(I+L)^{-1}(s-r)\right\|_p, \end{aligned} \tag{43}$$

where the second term in the sum is exactly $d^L_{\text{diff}}(s,r)$ [cf. (14)]. Therefore we may write

$$\begin{aligned} &d^{L'}_{\text{diff}}(s,r) - d^L_{\text{diff}}(s,r) \\ &\quad\leq \left\|\left((I+L+E)^{-1}-(I+L)^{-1}\right)(s-r)\right\|_p. \end{aligned} \tag{44}$$

By applying compatibility of p-norms (4) followed by the subadditivity property we obtain that

$$\begin{aligned} &d^{L'}_{\text{diff}}(s,r) - d^L_{\text{diff}}(s,r) \\ &\quad\leq \left\|\left((I+L+E)^{-1}-(I+L)^{-1}\right)\right\|_p \|(s-r)\|_p \\ &\quad\leq \left\|\left((I+L+E)^{-1}-(I+L)^{-1}\right)\right\|_p (\|s\|_p + \|r\|_p) \end{aligned} \tag{45}$$

Given that $I+L$ is nonsingular we have to show that $\|(I+L)^{-1}E\|_p < 1$ in order to be able to apply Lemma 4 with $A=(I+L)$ and further bound (45).

Due to doubly stochasticity [cf. Lemma 3], we have that $\|(I+L)^{-1}\|_1 = \|(I+L)^{-1}\|_\infty = 1$. Moreover, $\|(I+L)^{-1}\|_2 = 1$ comes from the fact that the smallest eigenvalue of $(I+L)$ and hence the largest eigenvalue of $(I+L)^{-1}$ is equal to 1. Consequently, we may write

$$\|(I+L)^{-1}E\|_p \leq \|(I+L)^{-1}\|_p\|E\|_p < 1, \tag{46}$$

for $p\in\{1,2,\infty\}$, as wanted, where the first inequality follows from submultiplicativity (3). Hence, applying Lemma 4 with $A=(I+L)$ yields

$$\left\|(I+L+E)^{-1}-(I+L)^{-1}\right\|_p \leq \frac{\|E\|_p\|(I+L)^{-1}\|_p^2}{1-\|(I+L)^{-1}E\|_p}. \tag{47}$$

Recalling that $\|(I+L)^{-1}\|_p = 1$ for any $p\in\{1,2,\infty\}$ allows us to further bound (47) to obtain

$$\left\|(I+L+E)^{-1}-(I+L)^{-1}\right\|_p \leq \frac{\|E\|_p}{1-\|E\|_p} \leq \frac{\epsilon\|L\|_p}{1-\epsilon\|L\|_p}, \tag{48}$$

where we used that $\|E\|_p \leq \epsilon\|L\|_p < 1$ for the last inequality.

Utilizing the Taylor series of $1/(1-\epsilon\|L\|_p)$ and substituting (48) into (45) combined with the fact that $\|s\|_p \leq \gamma$ and $\|r\|_p \leq \gamma$ we have that

$$d^{L'}_{\text{diff}}(s,r) - d^L_{\text{diff}}(s,r) \leq \sum_{n=1}^{\infty} 2\gamma(\epsilon\|L\|_p)^n = 2\gamma\|L\|_p\epsilon + o(\epsilon). \tag{49}$$

In a similar manner but starting from the definition of $d^L_{\text{diff}}(s,r)$, it can be shown that

$$d^L_{\text{diff}}(s,r) - d^{L'}_{\text{diff}}(s,r) \leq 2\gamma\|L\|_p\epsilon + o(\epsilon). \tag{50}$$

Finally, by combining (49) and (50), we obtain (24) and the proof concludes. ■

References

[1] D. Bu *et al.*, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic Acids Res.*, vol. 31, no. 9, pp. 2443–2450, 2003.

[2] E. Lieberman, C. Hauert, and M. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–316, 2005.

[3] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, p. 036104, 2006.

[4] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.

[5] J. Kleinberg, "Complex networks and decentralized search algorithms," in *Proc. Int. Congr. Math. (ICM)*, 2006, vol. 3, pp. 1019–1044.

[6] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," in *Proc. 36th Annu. ACM Symp. Theory Comput.*, New York, NY, USA, 2004, pp. 561–568.

[7] N. Lynch, *Distributed Algorithms*. San Mateo, CA, USA: Morgan Kaufmann, 1996.

[8] J. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.

[9] B. Miller, N. Bliss, and P. Wolfe, "Toward signal processing theory for graphs and non-euclidean data," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 5414–5417.

[10] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[11] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," arXiv preprint arXiv:1210.4752, 2012.

[12] S. Narang and A. Ortega, "Downsampling graphs using spectral theory," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2011, pp. 4208–4211.

[13] R. Mittler, S. Vanderauwera, M. Gollery, and F. V. Breusegem, "Reactive oxygen gene network of plants," *Trends Plant Sci.*, vol. 9, no. 10, pp. 490–498, 2004.

[14] O. Sporns, *Networks of the Brain*. Cambridge, MA, USA: MIT Press, 2011.

[15] A. Luikov, *Analytical Heat Diffusion Theory*. New York, NY, USA: Academic, 1968.

[16] E. Eckert and R. Drake, *Analysis of Heat and Mass Transfer*. Bristol, PA, USA: Hemisphere, 1987.

[17] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete input spaces," in *Proc. 9th Int. Conf. Mach. Learn. (ICML)*, 2002, vol. 2, pp. 315–322.

[18] P. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*, vol. 28. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[19] M. Freidlin and A. D. Wentzell, "Diffusion processes on graphs and the averaging principle," *Ann. Probab.*, vol. 21, no. 4, pp. 2215–2245, 1993.

[20] A. Szlam, M. Maggioni, and R. Coifman, "Regularization on graphs with function-adapted diffusion processes," *J. Mach. Learn. Res.*, vol. 9, pp. 1711–1739, 2008.

[21] A. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*, vol. 2777, B. Schalkopf and M. Warmuth, Eds. New York, NY, USA: Springer, 2003, pp. 144–158.

[22] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proc. Amer. Control Conf.*, 2005, pp. 1859–1864.

[23] J. A. F. R. Olfati-Saber and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Mar. 2007.

[24] W. Ren, "Consensus based formation control strategies for multi-vehicle systems," in *Proc. Amer. Control Conf.*, 2006, pp. 4237–4242.

[25] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Stable flocking of mobile agents, part I: Fixed topology," *Proc. IEEE Conf. Decis. Control*, 2003, pp. 2010–2015.

[26] M. H. DeGroot, "Reaching a consensus," *J. Amer. Stat. Assoc.*, vol. 69, pp. 118–121, 1974.

[27] J. C. Dittmer, "Consensus formation under bounded confidence," *Nonlinear Anal.*, vol. 47, 2001.

[28] S. Segarra and A. Ribeiro, "Hierarchical clustering and consensus in trust networks," in *Proc. IEEE 5th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2013, pp. 85–88.

[29] F. Chung, *Spectral Graph Theory*, vol. 92. Providence, RI, USA: American Mathematical Society, 1997.

[30] R. Bellman, *Introduction to Matrix Analysis*, vol. 960. Philadelphia, PA, USA: SIAM, 1970.

[31] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*, vol. 33. Providence, RI, USA: American Mathematical Society, 2001.

[32] G. Golub and C. V. Loan, *Matrix Computations*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1989.

[33] R. D. Maesschalck, D. Jouan-Rimbaud, and D. Massart, "The mahalanobis distance," *Chemom. Intell. Lab. Syst.*, vol. 50, no. 1, pp. 1–18, 2000.

[34] M. A. A. Cox and T. F. Cox, "Multidimensional scaling," in *Handbook of Data Visualization*. New York, NY, USA: Springer, 2008, pp. 315–347.

[35] M. Hofree, J. Shen, H. Carter, A. Gross, and T. Ideker, "Network-based stratification of tumor mutations," *Nature Methods*, vol. 10, pp. 1108–1115, 2013.

[36] E. G. Cerami *et al.*, "Pathway commons, a web resource for biological pathway data," *Nucleic Acids Res.*, vol. 39, no. suppl 1, pp. D685–D690, 2011 [Online]. Available: http://nar.oxfordjournals.org/content/39/suppl_1/D685.abstract

[37] Y. Lecun and C. Cortes. (1998). *The MNIST Database of Handwritten Digits* [Online]. Available: http://yann.lecun.com/exdb/mnist/

[38] R. Varga, *Matrix Iterative Analysis*. New York, NY, USA: Springer, 2000.

[39] C. V. Loan, "The sensitivity of the matrix exponential," *SIAM J. Numer. Anal.*, vol. 14, no. 6, pp. 971–981, 1977.

[40] D. M. Young, *Iterative Solution of Large Linear Systems*. New York, NY, USA: Academic, 1971.

[41] T. Fujimoto and R. Ranade, "Two characterizations of inverse-positive matrices: The Hawkins-Simon condition and the Le Chatelier-Braun principle," *Electron. J. Linear Algebra*, vol. 11, pp. 59–65, 2004.

Santiago Segarra (S'12) received the B.Sc. (Hons.) degree in industrial engineering from the Instituto Tecnológico de Buenos Aires (ITBA), Argentina, in 2011, and the M.Sc. degree in electrical engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2014. He is currently pursuing the Ph.D. degree in electrical and systems engineering at the University of Pennsylvania. His research interests include network theory, data analysis, machine learning, and graph signal processing. Mr. Segarra was the recipient of the ITBA's 2011 Award to the Best Undergraduate Thesis in industrial engineering and the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina.

Weiyu Huang (S'15) received the B.Eng. (Hons.) degree in electronics and telecommunication from the Australian National University (ANU), Canberra, Australia, in 2012. He is currently pursuing the Ph.D. degree in electrical and systems engineering at the University of Pennsylvania, Philadelphia, PA, USA. From 2011 to 2013, he was a Telecommunication Engineer and Policy Officer with the Australian Communication and Media Authority. His research interests include network theory, pattern recognition, graph signal processing, and the study of networked data arising in human, social, and technological networks. Mr. Huang was the recipient of the University Medal offered by the ANU.

Alejandro Ribeiro (S'02–M'07) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, Uruguay, in 1998, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2005 and 2007, respectively. From 1998 to 2003, he was a Member of the Technical Staff with Bellsouth Montevideo. After his M.Sc. and Ph.D. studies in 2008, he joined the University of Pennsylvania (Penn), Philadelphia, where he is currently the Rosenbluth Associate Professor with the Department of Electrical and Systems Engineering. His research interests include applications of statistical signal processing to the study of networks and networked phenomena, wireless networks, network optimization, learning in networks, networked control, robot teams, and structured representations of networked data structures. Dr. Ribeiro is a Fulbright Scholar and a Penn Fellow. He was the recipient of the 2014 O. Hugo Schuck Best Paper Award, the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and Student Paper Awards at the 2013 American Control Conference (as Adviser), as well as the 2005 and 2006 International Conferences on Acoustics, Speech, and Signal Processing.