TRACKING NETWORK DYNAMICS: A SURVEY USING GRAPH DISTANCES[1]

BY CLAIRE DONNAT AND SUSAN HOLMES[2]

Stanford University

From longitudinal biomedical studies to social networks, graphs have emerged as essential objects for describing evolving interactions between agents in complex systems. In such studies, after pre-processing, the data are encoded by a set of graphs, each representing a system's state at a different point in time or space. The analysis of the system's dynamics depends on the selection of the appropriate analytical tools. In particular, after specifying properties characterizing similarities between states, a critical step lies in the choice of a distance between graphs capable of reflecting such similarities.

While the literature offers a number of distances to choose from, their properties have been little investigated and no guidelines regarding the choice of such a distance have yet been provided. In particular, most graph distances consider that the nodes are exchangeable—ignoring node "identities." Alignment of the graphs according to identified nodes enables us to enhance these distances' sensitivity to perturbations in the network and detect important changes in graph dynamics. Thus the selection of an adequate metric is a decisive—yet delicate—practical matter.

In the spirit of Goldenberg et al.'s seminal 2009 review [*Found. Trends Mach. Learn.* **2** (2010) 129–233], this article provides an overview of commonly-used graph distances and an explicit characterization of the structural changes that they are best able to capture. We show how these choices affect real-life situations, and we use these distances to analyze both a longitudinal microbiome dataset and a brain fMRI study. One contribution of the present study is a coordinated suite of data analytic techniques, displays and statistical tests using "metagraphs": a graph of graphs based on a chosen metric. Permutation tests can uncover the effects of covariates on the graphs' variability. Furthermore, synthetic examples provide intuition as to the qualities and drawbacks of the different distances. Above all, we provide some guidance on choosing one distance over another in different contexts. Finally, we extend the scope of our analyses from temporal to spatial dynamics and apply these different distances to a network created from worldwide recipes.

1. Introduction and motivation.

Motivation. From social sciences to biology, scientific communities across a wide number of disciplines have become increasingly interested in the study of networks, that is, graphs in which each entity or player is assigned to a node, and existing interactions or associations between entities are modeled by edges. If graphs provide a versatile framework for encapsulating structural information in datasets, they also come as an indispensable paradigm in a number of applications where the study of each individual node represented is either irrelevant or intractable. By way of illustration, brain connectome data represent brain activity by modeling neurons' activation patterns from a network perspective rather than by recording each individual neuron's activity. The focus of such studies typically concerns the system as a whole rather than at an atomic level. Similarly, in microbial ecology, communities of bacteria can be represented by a co-occurrence graph where each bacteria is a node and the edges are a (carefully-selected) function of bacteria co-abundances. The representation of biological samples as graphs provides a more informative framework than the bacterial counts themselves. Indeed, these graphs can be taken as essential summaries of the data, which can then be used to further investigate the associations highlighted by recent studies between "significant" bacterial communities and various medical conditions, such as obesity [Hullar and Lampe (2012), Turnbaugh et al. (2008, 2006)] or preterm birth [DiGiulio et al. (2015)].

1.1. *Applications*: *Microbiome and fMRI data*.   We illustrate our discussion on the analysis of distances between graphs with two main examples.

The 2011 Relman antibiotics dataset. This longitudinal microbiome study consists of a set of 162 bacterial samples taken from the gut of three distinct subjects (D, E, and F) at different points in time. The subjects were given two courses of antibiotics over ten months, yielding seven distinct treatment phases (pretreatment, first antibiotic course, week after stopping treatment 1, interim, second course of antibiotics, week after stopping treatment 2, and post-treatment phase). The goal of the study was to assess the antibiotics' effects on microbial communities. Already investigated in the literature [Dethlefsen and Relman (2011), Fukuyama et al. (2012)], we propose to tackle its analysis from a new network perspective. Our analysis provides complementary information to the previous by allowing the analysis of higher-order interactions between bacteria: Can we characterize prevalent communities of bacteria for each treatment phase? How do these communities react to the different drugs? The study of cooccurrence networks in microbiome samples is becoming increasingly popular [Barberán et al. (2012), Gerber (2014), Layeghifard, Hwang and Guttman (2017), Proulx, Promislow and Phillips (2005), Weiss et al. (2016)]. A critical step in the analysis is the transformation of the raw bacterial counts into a graph capturing such interactions. For each subject at a given treatment phase, we define a graph in which each node corresponds to a

specific bacteria species, and edges $\mathcal{E} = \{(i, j)\}$ capture pairwise "affinities" between bacteria i and j. Intuitively, these affinities capture symbiotic mechanisms between bacteria: Do they thrive simultaneously together or, on the contrary, does one bacteria tend to smother the development of another? While a plethora of methods have been suggested for inferring networks from the abundance matrix [see the Appendix Donnat and Holmes (2018) for more references], and because of the particular nature of the data (zero-inflated negative binomial counts), we use a thresholded-Kendall correlation as a measure of the tendency of two bacteria to thrive (or wither) together. Indeed, the Kendall correlation is a ranked-based correlation and less biased by the over-representation of zeros in the data. The analysis produces a set of twenty-one different graphs—one for each of the three subjects during each of the treatment phases—on the 2582 nodes representing the different species. More details regarding the explicit construction of these graphs are given in the Appendix [Donnat and Holmes (2018)].

Resting-State fMRI data. We will also use different distances to compare brain connectomes, which have radically different properties to the microbiome. The dataset that we study here consists in the resting-states fMRI data of 29 cocaine-dependent patients, and was published as part as a study of the effect of cocaine addiction on functional and structural connectivity [Kelly et al. (2011)].[3] The authors of the study show the existence of a statistically-significant reduction in the interhemispheric connectivity between cocaine-users and controls, highlighting the existence of an effect of substance-abuse on functional connectivity. In a similar spirit, we apply different network distances to this dataset in order to assess if the number of years under cocaine-dependence correlates with differences between the different connectomes. Each patient's raw fMRI data has been preprocessed through the FSL standard pipeline [more details can be found in the Appendix Donnat and Holmes (2018)]. This procedure yields a total of 116 nodes—each corresponding to a region of interest (ROI) in the brain—with values over 140 time points. To create a (weighted) graph from these filtered time series, we use thresholded Pearson correlation coefficients between nodes. Our approach is akin to Tétreault et al. (2016) who select a threshold by controlling the number of edges in the graphs: The threshold used here is the mean (across subject) of each correlation matrix's 97th quantile. On average, the graphs that we recover are thus sparse, with only 3% of the edges in a fully connected graph.

Given these sets of graphs, the crux of the analysis lies in the choice of a distance capable of identifying similar "graph-states," tailored to the data at hand. In the microbiome example, while the thousands of taxa that constitute the human microbiome allow for a rich variety of potentially different microbial communities, these communities usually involve a very small proportion of the taxa, and

---

[3]Data publicly available at the following link: http://fcon_1000.projects.nitrc.org/indi/ACPI/html/acpi_nyu_1.html.

the associated abundance matrices are typically sparse (here, 17% of the observations are nonzero). In the brain connectome setting, fMRI data is also usually characterized by a high amount of noise: Preprocessing steps, such as the template alignment and Gaussian smoothing performed to realign the brains and account for small head movements during the data retrieval, are known to blur even further the signal-to-noise ratio.

1.2. *Problem statement and notation*.

Problem statement. From now on, we assume that the raw data has been transformed into a set of graphs, which we treat as input. In this perspective, the definition of a distance between aligned graphs, that is, graphs defined on the same set of identified nodes takes on a significant importance. Indeed, distances constitute a stepping stone to any statistical analysis: Pairwise comparisons, cluster or variance analysis, or even the definition of a "mean" or "median," all follow naturally once a distance has been selected thus making the selection of an appropriate distance a crucial step in the analysis of graph data. While the problem of assessing the distance between two unlabeled graphs is somehow classical and has been well studied [Champin and Solnon (2003), Koutra et al. (2011), Shimada et al. (2016), Zager and Verghese (2008)], our focus is different. Indeed, since in our examples, the nodes have been endowed with a particular identity, permutation-invariant distances would discard potentially relevant information. One might even wish to leverage the information contained in the nodes' labeling to define a distance sensitive to the intensity of changes at key-node locations. While the literature provides us with a number of "off-the-shelf" graph distances—any of which being, in principle, suited for the task—these distances exhibit distinct properties and capture different types of structural changes. In particular, a decisive parameter lies in the identification of the correct scale for comparing graphs: At which scale is the information contained? Is the analysis interested in capturing local structural changes at an atomic level, or should it focus more broadly on identifying structural similarities at the level of node communities? This notion of scale is at the heart of this review, which we use to classify and investigate the properties of some of the many different distances and similarities between graphs. By organizing each section along the presentation of distances belonging to one of three scale categories (local, mesoscale, and global), we hope to provide templates to help the analyst in the selection of an appropriate distance given his/her desiderata.

Starting with structural and spectral distances—respectively local and global distances, two "scale extremes" which constitute the main bulk of metrics proposed in the literature—we highlight the dynamics and types of structural changes that these two main categories are best able to capture. An improved understanding of these metrics' properties suggest new dissimilarities tailored to specific scenarios. We then present a collection of distances best suited for analyses at the

mesoscale. In particular, we introduce a new set of graph distances based on spectral heat kernels, and argue that these dissimilarities are optimal in that they combine both local and global structural information. In each case, the performance of the different distances is assessed on both the microbiome and fMRI datasets, as well as on synthetic, controlled experiments. In Section 6, we compare the graph distances; the advantages and disadvantages of the different definitions are summarized in the decision tree displayed in Figure 11. Finally, we extend our illustration by applying the distances to the broader context of spatial analyses in an analysis of "recipe" networks.

Notation. Throughout this review, we write $G = (\mathcal{V}, \mathcal{E})$, the graph with vertices $\mathcal{V}$ and edges $\mathcal{E}$. We denote as $N = |\mathcal{V}|$ the number of nodes, $|\mathcal{E}|$ the number of edges, and we write $i \sim j$ if nodes i and j are neighbors. Our framework considers undirected binary graphs, with no self-loops (which we extend to the study of weighted graphs in our applications). A refers to the adjacency matrix of the graph, and D to its degree matrix:

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad D = \text{Diag}(d_i)_{i=1\cdots N} \quad \text{such that } d_i = \sum_{j=1}^{N} A_{ij}.$$

As we are restricting ourselves to undirected graphs, matrix A is symmetric: $A^T = A$. The Laplacian [Bollobás (1998)] of the graph is the matrix defined as $L = D - A$. The Laplacian is symmetric, and we consistently write its (real-valued) eigenvalue decomposition as $L = U\Lambda U^T$, where U is a unitary matrix, and $\Lambda = \text{Diag}(\lambda_i)$ is the diagonal matrix of the eigenvalues: $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$.

2. Quantifying local changes via structural distances.

Distances between graphs usually fall in one of two general categories, often considered as mutually exclusive: structural versus spectral distances. The first one captures local changes, that is, changes in the graph structure around each node and is thus especially well suited to data where local changes can induce radical changes. By way of illustration, different bonds in molecules can induce radically different properties (toxicity, solubility, etc.), while in brain networks, these distances can capture changes in the behavior of each individual region of interest (ROI). At the other extreme, the second collection of distances assesses the smoothness of the evolution of the overall graph structure by tracking changes in the eigenvalues of the graph Laplacian or its adjacency matrix. As such, these distances are better suited for analyses focusing on overall graph properties, that is, how nodes in the graph are globally organized and interact, rather than each of the nodes' individual function. In brain networks, for instance, spectral distances can be employed to assess global changes in connectivity: Is the treatment group characterized by a significantly different overall connectivity between areas of the brain compared to the control? We begin our review by analyzing properties of these two most popular types of distances in the two following sections.

2.1. *The Hamming distance.*

Definition. The Hamming distance—a special instance of the broader class of Graph-Edit distances—measures the number of edge deletions and insertions necessary to transform one graph into another. While it is widely used in many graph analyses, we will see in this subsection through tests and multidimensional scaling plots that the Hamming distance is nonetheless a blunt tool. We highlight some of its shortcomings.

More formally, let G and $\tilde{G}$ be two graphs on N nodes, with A and $\tilde{A}$ their corresponding adjacency matrices. The (normalized) Hamming distance is defined as

$$(2.1) \qquad d_H(G, \tilde{G}) = \sum_{i,j} \frac{|A_{ij} - \tilde{A}_{ij}|}{N(N-1)} = \frac{1}{N(N-1)} \|A - \tilde{A}\|_1.$$

This defines a metric between graphs, since it is a scaled version of the L_1 norm between the adjacency matrices A and $\tilde{A}$. Equation (2.1) defines a distance bounded between 0 and 1 over all graphs of size N.

General analysis framework We briefly outline here the framework that we use throughout the paper to analyze these graphs using each of the distances (or more generally, dissimilarity measures) that we present. For each dataset (microbiome and fMRI), we store the pairwise dissimilarities between graphs in a n-by-n dimensional matrix H, where $H_{ij} = d(G_i, G_j)$ and n is the number of graphs ($n = 21$ in the microbiome example and $n = 29$ in the fMRI dataset). We use this matrix H to analyze a set of different factors. In particular, in the microbiome example, each distance is used for the analysis of: (a) the graphs' variability from time frame to time frame, illustrated by plots of distances between consecutive graphs [Figure 1(B), (E)] and (b) similarities across subjects and across treatment phases, visualized by both heatmaps or clustermaps of the pairwise distances H between graphs [Figure 1(A), (E)] and their low-dimensional projections (multidimensional scaling MDS). Figure 1(B), (D) shows these two-dimensional projections. In the fMRI example, the projections show how dissimilarities capture the relationship between the connectivity graphs and the number of years that subjects spent under cocaine dependency.

Application. Figure 1 illustrates the results of the analysis of the microbiome study using the Hamming distance on both the bacterial (top row), as well as the fMRI data (bottom row-right).

Results: microbiome data. In the antibiotic study (top row of Figure 1), the Hamming distance is computed between the bacterial community graphs throughout the treatment course. Similar dynamics across subjects appear, the shapes of the curves of each subject match [Figure 1(C)] and show dips in the distances between consecutive graphs just after the antibiotic time courses. The MDS projection [Figure 1(B)] on the first components highlights the existence of a "treatment gradient": interim phases—located in the bottom right corner of the figure—are
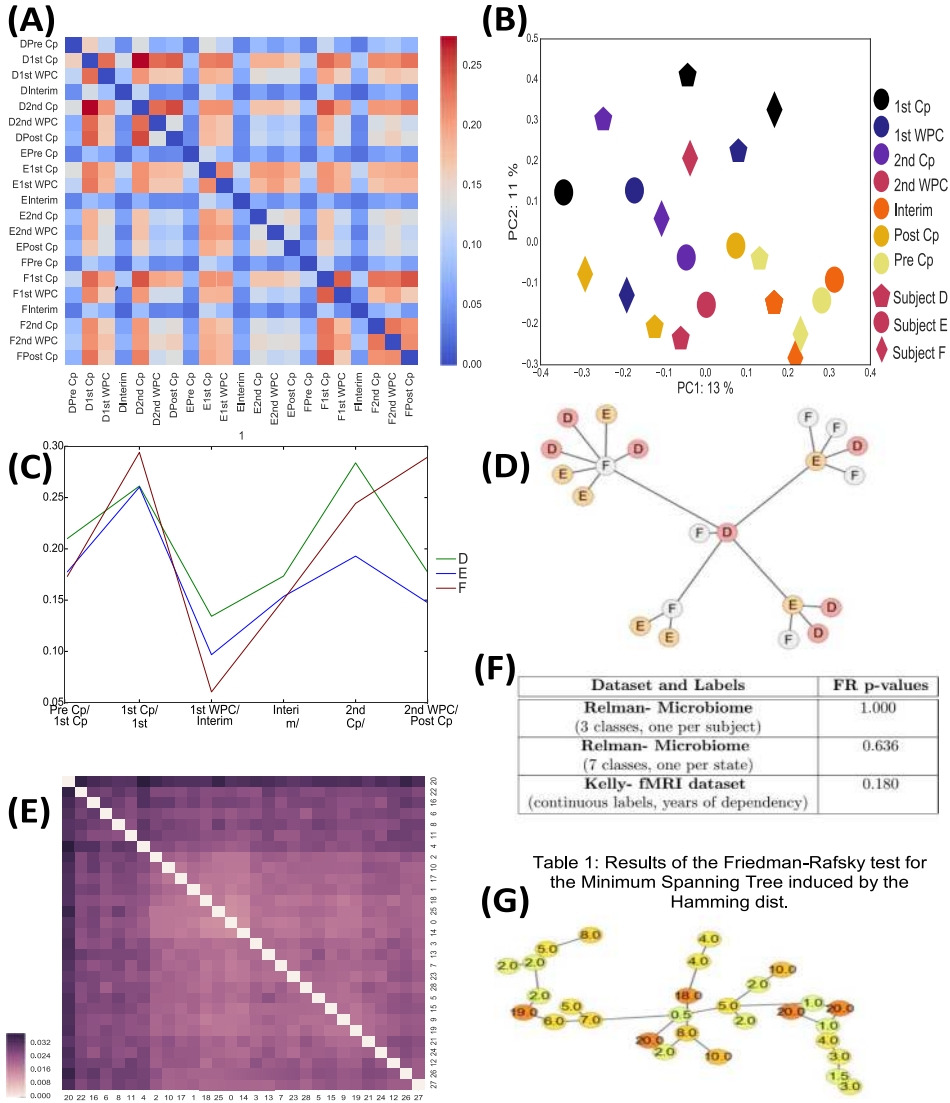
FIG. 1. *Hamming distance between bacterial graphs* (*top rows*), *and brain graphs* (*bottom row*). *Heatmap of the Hamming distances between Kendall-correlation-based bacterial graphs* (A) *and MDS projections on the first two principal components* (B). *Colors denote treatment phases, and shapes represent different subjects. Plots of the consecutive distances between bacterial graphs* (C). *Minimum Spanning tree between bacterial graphs induced by the Hamming distance* (G). *Friedman–Rafsky test for significance for the different datasets* (F). *Clustermap of the fMRI graphs* (E). *Minimum spanning tree between brain connectomes induced by the Hamming distance* (G).

closer to the pretreatment samples and far from the treatment phases (violet and black points at center-left of the figure), consistent with biological interpretation of the treatment effects. While the Hamming distance does not detect strong similarities between samples belonging to the same individual [no darker blue blocks along the diagonal of the heat map in Figure 1(A)], it is able to identify similar dynamic regimes across subjects as highlighted by the clustered MDS projections of points corresponding to the same treatment phase.

In order to quantify this effect, we ran a Friedman–Rafsky test. For a given value of k, we compute the k-nearest-neighbor "metagraph" (or graph of graphs) induced by the pairwise-dissimilarity matrix H. By representing H as a graph where each node (corresponding to the coabundance graph for a patient at a given treatment phase) is itself a graph, and edges reflect the k-strongest similarities between graphs, we simplify the relevant information. From now on, we refer to this induced k-nearest-neighbor graph of graphs as the k-nn metagraph. Having constructed the k-nn metagraph, as in a standard Friedman–Rafsky setting, we compute the number of its edges which connect graphs of the same class (i.e., in the microbiome dataset, treatment stage, or subject). We then permute the node labels, generating 50,000 graphs with the same topology, but where the edges randomly connect nodes independently of their class. We compare the original value to this synthetic null permutation distribution to get the associated p-value. This assesses the compatibility of the distance on a given set of labels: if the distance clusters together graphs belonging to the same category, then the p-value should be significantly small. The p-values are reported alongside the plots in Figure 1(F), where we have conducted this experiment with $k = 1$ (the nearest-neighbor graph is thus simply the minimum spanning tree). In this case, our test does not detect any statistically significant association between the edges in the metagraph and either the treatment stages or the subject labels. We note that increasing the number of neighbors considered ($k = 2, 3, \ldots$) in the metagraph does not uncover any meaningful statistical associations either.

Results: fMRI data. When applied to the resting state-fMRI data, the Hamming distance does not detect any clusters of closely related graphs [as shown by the uniform clustermap in Figure 1(E) and the uniform tSNE projections in Figure 1(D)]. We also adapt our previous Friedman–Rafksy test to handle continuous labels instead of discrete classes. This allows us to test the association of the k-nearest neighbor metagraph between patients and the amount of time that they have spent under cocaine dependency. The test statistic is now defined as the sum of the differences between labels (i.e., time under dependence) for all the edges in the graph. In this setting, a small score would indicate that brain networks are more similar to other networks with similar "time under dependence." As shown in both the figures and the plots, the Hamming distance does not detect any significant relationship between the relative distance of the graphs and their labels. We also run an analysis-of-variance type test. We split the graphs into two classes

(with roughly the same number of subjects): patients with less than 5 years un-
der cocaine dependency and patients with more than five years. We then compute
the ratio $\Delta = \frac{\bar{D}_{12}}{\frac{n_1}{n_1+n_2}\bar{D}_{11} + \frac{n_2}{n_1+n_2}\bar{D}_{22}}$ where $\bar{D}_{ij}$ denotes the average distance between
subjects of class i and j: under the null; this ratio should be centered around 1.
We assess the significance of this ratio via a permutation test, which here yields
a p-value of 0.62: In this case again, the Hamming distance does not detect any
significant difference between the graphs in the two classes.

Analysis. With a cost complexity of the order of $O(N^2)$, the Hamming distance
provides a straightforward way of comparing sequences of aligned graphs that only
takes into account the number of shared edges. It thus comes as no surprise that
this distance has been a long-time favorite in various graph comparison problems.
Graph embedding techniques—which provide a vector-valued representation for
each graph that captures its geometric properties—are a case in point: in Luqman,
Ramel and Lladós (2013), the authors define dissimilarities between subgraphs
through their graph-edit distance. Similarly, in Ferrer et al. (2013), the authors
introduce the notion of a "median graph" as the minimizer of the sum of pairwise
graph-edit distances.

While the Hamming distance is a perfectly valid first candidate graph distance
for any type of analysis, it is worth emphasizing that it only reveals some restricted
aspect of network dissimilarities.

The first trait to highlight is its uniform treatment of all changes in the graph
structure: All additions and deletions are assumed to have similar importance.
Changes in the network's core are treated equivalently to changes in the periph-
ery. We will analyze the consequences and limitations of this assumption in Sec-
tion 2.3. A second trait is Hamming's sensitivity to the density of the graphs. This
yields a limited capacity to recognize similar dynamical processes across graphs
with varying sparsity. As an example of the second point, let us consider a dynamic
regime in which, at every time point, each edge is randomly flipped independently
of the others: It either stays in the graph or disappears with probability p. The total
number of disappearances follows a binomial distribution with mean $p|\mathcal{E}|$. For an
identical perturbation mechanism, dense graphs are thus placed at higher distances
to each other—and are thus considered as more unstable—than sparse graphs. The
Hamming distance is unable to recognize that these graphs share in fact the same
level of relative variability, which can hinder some aspects of the analysis. Indeed,
the random deletion process at hand can be thought of as blurring noise applied
to a true underlying graph structure, and is a typical representation our inability
to observe all interactions between nodes in a complex system. In this case, it
seems more natural to specify the inherent variability of the data in terms of "noise
level" rather than "noise quantity," and our analysis should thus recognize similar
noise levels independently of the graphs' original sparsity. Similarly, the Hamming
distance tends to place nested graphs at a smaller distance to each other than other
metrics. Indeed, suppose that graph $\tilde{G}$ comprises 50% of the edges of graph G. The

Hamming distance between the two graphs is then simply $d(G, \tilde{G}) = \frac{0.5\|A\|}{N(N-1)}$, and does not correct for the size of the initial graph. We could nonetheless argue that this distance should be big (or at least close to 0.5), since the structure of the system is radically modified. Our microbiome study is a case in point [Figure 1(B)]: The variety of the microbiota involved in the interim phase jumps to almost twice its corresponding value in any of the antibiotics phases (the number of bacteria increases from around 210 bacteria to 420). The distances between the interim phase and the other phases are subsequently smaller than for any of the other phases, with a number of shared taxa.

The Hamming distance is thus a measure of the *amount of change* between two graphs. While this might be adequate for characterizing the evolution of a given system through time, it is nonetheless unfit for finding similarities in broader settings. Tasks such as comparing graph dynamics in the presence of different degree densities or recognizing instances of the same network family (Erdős–Rényi random graphs, preferential attachment graphs, etc.) indubitably require other metrics.

2.2. *The Jaccard distance*.   *Definition*. A potential solution to the aforementioned density-effect problem consists in using the Jaccard distance [Levandowsky and Winter (1971)], which includes a normalization with respect to the volume of the union graph

$$d_{\text{Jaccard}}(G, \tilde{G}) = \frac{|G \cup \tilde{G}| - |G \cap \tilde{G}|}{|G \cup \tilde{G}|}$$

$$(2.2) \qquad = \frac{\sum_{i,j} |A_{ij} - \tilde{A}_{ij}|}{\sum_{i,j} \max(A_{i,j}, \tilde{A}_{ij})}|$$

$$= \frac{\|A - \tilde{A}\|_1}{\|A + \tilde{A}\|_*},$$

where $\| \cdot \|_*$ denotes the nuclear norm of a matrix.

Equation (2.2) is known to define a proper distance between the graphs. A straightforward way to see this is to use the Steinhaus transform: For (X, d) a metric and c a fixed point, the transformation $\delta(x, y) = \frac{2d(x,y)}{d(x,c)+d(y,c)+d(x,y)}$ produces a metric. Apply this transformation, with d the Hamming distance and c the empty graph, to see

$$\delta(G, \tilde{G}) = \frac{2\|A - \tilde{A}\|_1}{\|A\|_1 + \|\tilde{A}\|_1 + \|A - \tilde{A}\|_1}$$

$$(*) \qquad = \frac{2(|G \cup \tilde{G}| - |G \cap \tilde{G}|)}{2|G \cup \tilde{G}|}$$

$$= d_{\text{Jaccard}}(G, \tilde{G}).$$

In particular, taking for instance $G = G_t$ and $\tilde{G} = G_{t+1}$ the graphs associated to the state of a system at two consecutive time points t and $t + 1$ (with $\mathcal{E}_{G_t}$ and $\mathcal{E}_{G_{t+1}}$ their respective set of undirected edges) and rewriting the left-hand side of (*), we have

$$d_{\text{Jaccard}}(G_t, G_{t+1}) = \frac{d_{\text{Hamming}}(G_t, G_{t+1})}{\frac{|\mathcal{E}_{G_t}| + |\mathcal{E}_{G_{t+1}}|}{2N(N-1)} + \frac{1}{2}d_{\text{Hamming}}(G_t, G_{t+1})}$$

(2.3)

$$\implies \quad d_{\text{Jaccard}}(G_t, G_{t+1}) = \frac{\frac{d_{\text{Hamming}}(G_t, G_{t+1})}{\bar{S}}}{1 + \frac{d_{\text{Hamming}}(G_t, G_{t+1})}{2\bar{S}}}$$

with $\bar{S} = \frac{|\mathcal{E}_{G_t}| + |\mathcal{E}_{G_{t+1}}|}{2N(N-1)}$ is the average sparsity of the two graphs.

Application. Figure 2 shows the result of the analysis carried out using the Jaccard distance. Since the edges in our graphs have been assigned different weights according to the intensity of the interaction between bacteria, we have used the version of the Jaccard distance extended to the weighted graph setting, defined as

$$d_{\text{Jaccard}}(G, \tilde{G}) = 1 - \frac{\sum_{i,j} \min(A_{ij}, \tilde{A}_{ij})}{\sum_{i,j} \max(A_{ij}, \tilde{A}_{ij})}.$$

This analysis yields different results to the Hamming distance [Figure 2(A), (D), (C), (D)]. We note that the treatment phases express more variability and are far from most of the other samples. The Friedman–Rafsky test for the microbiome
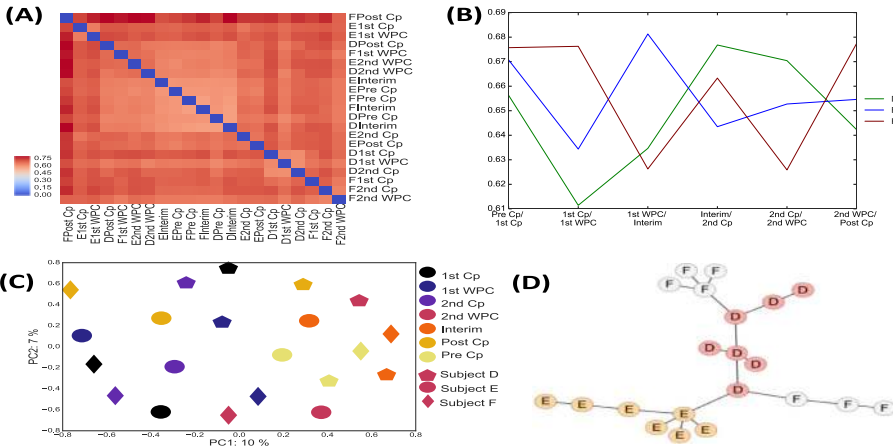

FIG. 2. *Application of the Jaccard distance to the microbiome study. Clustermap of the Jaccard distances between Kendall-correlation-based bacterial graphs* (A). *Plots of the consecutive distances between bacterial graphs* (B). *MDS projection of the bacterial* (C) *graphs on the first two principal axes. Colors denote treatment phases, and shapes represent different subjects. Minimum spanning tree between bacterial graphs* (D).

data highlighted a significant dependence of the 3-nn metagraph on the subject: With a p-value of 0.0002, this test shows that bacterial graphs corresponding to the same patient are significantly closer than under the random null model. This effect is further confirmed by running an analysis-of-variance type test and computing the statistics $\Delta = \frac{1}{3} \frac{\sum_{i \in \{D,E,F\}} \bar{D}_{i,i^c}}{\sum_{i \in \{D,E,F\}} \frac{n_i}{n_{\text{tot}}} \bar{D}_{i,i}}$ where $\bar{D}_{i,i^c}$ denotes the average distance between graphs in class i and graphs in any other class. Under the null, this statistic is centered at 1, and we evaluate its significance through a permutation test. This yields a p-value of 0.0018, highlighting the existence of a significant difference between graphs grouped according to their associated subject Id. This microbiome example is thus a case where the Jaccard distance is a better fit for our analysis: Whereas the Hamming fails to uncover any real similarity between bacterial graphs corresponding to the same subject, the Jaccard distance does capture the existence of greater similarities among graphs belonging to the same "block" (i.e., patient), a known effect in microbiome studies. However, when applied to the brain networks, the Jaccard distance displayed an almost uniform distance between all samples and did not recover any significant clustering or grouping of patients (with a p-value associated to the analysis-of-variance test of 0.61).

Discussion. The Jaccard distance adjusts for graph density by including in its normalization the average sparsity of the two graphs. As such, it reflects the amount of change with respect to the initial graph structure. To highlight the benefits of this property, let us consider a dynamic regime in which the total number of edges stays fixed, but at each time point, each edge is replugged with probability p in a previously vacant connection: The overall number of edges remains identical, but each flipped edge induces an increase in the Hamming distance of size $\frac{4}{N(N-1)}$. Hence, the average Hamming distance between G_t and G_{t+1} admits a closed-form expression of the type

$$d_{\text{Hamming}}(G_t, G_{t+1}) = \frac{4p|\mathcal{E}|}{N(N-1)} = 4ps,$$

where $s = |\mathcal{E}|/(N(N-1))$ is the sparsity of the original graph. By equation (2.3), the Jaccard distance can be written as $d_{\text{Jaccard}}(G_t, G_{t+1}) = \frac{4p}{1+2p} = 2[1 - \frac{1}{1+2p}]$, where the later expression is a strictly increasing function of p. The Jaccard distance is thus independent of the sparsity and defines a one-to-one mapping between the rate of change p and the observed distance. In contrast, the effect of p is confounded in the Hamming distance by the influence of the sparsity.

This simple example shows that the Jaccard distance is better suited to comparing different dynamics, where the rate of edge rewiring is the main quantity of interest. Another one of its advantages with respect to Hamming is that it provides a more interpretable notion of graph distances. Indeed, the Jaccard distance can be understood as the proportion of edges that have been deleted or added with respect to the total number of edges appearing in either network: A Jaccard distance close to 1 indicates an entire remodeling of the graph structure between time t and $t + 1$.

In the microbiome study at hand, the Jaccard distance reveals more within-subject variability than Hamming distance, where the blue and red blocks in Figure 1(A) highlighted contrasted dissimilarities between graphs: Here, while there exists a strong subject effect, on the whole, the almost-uniform clustermap in Figure 2(A) shows that samples within subject are still highly variable.

2.3. *Shortcomings of local approaches.* While the Hamming and Jaccard distances provide straightforward ways of analyzing a graph's dynamics or evolution over time, such measures appear short-sighted. Indeed, these metrics focus on the direct neighborhood of each node, and fail to capture the "bigger picture" and information on the evolution of the graph as a whole. Figure 3 shows an example where a network G_0 undergoes two different dynamic processes, yielding distinct graph structures with similar Hamming distances to the original. In this setup, it is possible to argue that G_1 and G_2 are more similar to each other, since the maximal path length between any two nodes is 2, whereas information percolates less rapidly across the network in the third. Conversely, from another perspective, we could also argue that we should have $d(G_1, G_3) \leq d(G_1, G_2)$, since the two first share a higher number of nodes with identical degree or since they have the same number of spanning trees. This example is meant to show that distances can be adapted to capture specific aspects of a network's properties. In Koutra et al. (2016), Koutra and coauthors advocate that dissimilarity scores between graphs should satisfy the following four characteristics:

1. *Edge-importance*: modifications of the graph structure yielding disconnected components should be penalized more.

2. *Edge-submodularity*: a specific change is more important in a graph with a few edges than in a denser graph on the same nodes.

3. *Weight awareness*: the impact on the similarity measure increases with the weight of the modified edge.
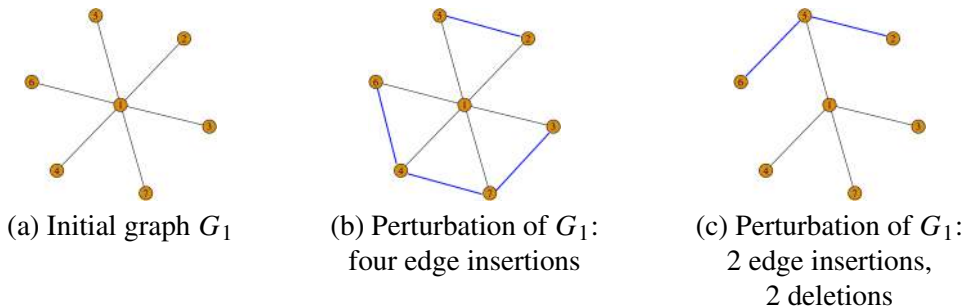

(a) Initial graph G_1 | (b) Perturbation of G_1: four edge insertions | (c) Perturbation of G_1: 2 edge insertions, 2 deletions

FIG. 3. *Two modifications of the same initial graphs (displayed in Figure 3(a), such that the Hamming distance with the original is $d_H(G_1, G_2) = d_H(G_1, G_3) = \frac{4}{21}$, and the Jaccard distances are $d_J(G_1, G_2) = \frac{2}{5}$ and $d_J(G_1, G_3) = \frac{1}{4}$. The average shortest path lengths are 1.71 for the initial graph G_1, 1.51 for G_2, and 2 for G_3.*

   4. *Focus awareness*: random changes in graphs are less important than targeted changes of the same extent.

These can serve as guidelines and be modified and enriched by the data analyst depending on the application at hand. The Jaccard and Hamming distance treat all edges uniformly, irrespective of their status (thus violating criteria 1 and 2 for instance).

   3. Comparing graph structures: A spectral approach. We now turn to the class of spectral distances. Spectral distances are global measures defined using the eigenvalues of either the adjacency matrix A or of some version of the Laplacian L. Consistent with the notation introduced in Section 1, the (combinatorial) Laplacian of the graph is defined as $L = D - A$, where D is the diagonal matrix such that D_{ii} is the degree of node i. Another popular choice consists in using the normalized Laplacian, defined as $\tilde{L} = I - D^{-1/2} A D^{-1/2}$.

   As mentioned in the first section, spectral distances are more suited to analyses where the critical information is contained at a global scale, rather than at a local one. For instance, in brain connectome data, such distances are suited to analyses focusing on the overall connectivity patterns between regions of the brain: Does cocaine usage decrease the connectivity between different regions? In microbiome studies, they can quantify changes in the overall microbial ecosystem: For instance, does the introduction of a given treatment radically change a microbiome ecosystem by destroying multiple coexisting communities and favoring the rise of a dominant bacteria?

   But why should eigenvalues characterize the state of a graph better? Let us first provide some intuition for this spectral approach. The eigenvalues of a graph characterize its topological structure, and in particular the way that energy or information localized at a particular node can be propagated over the graph. As such, they are related to the stability of the complex system that the graph represents. In quantum chemistry for instance, hydrocarbons are typically represented by graphs, whose adjacency matrix' eigenvalues correspond to energy levels of its electrons. In physics, the eigenvalues of the Laplacian represent the vibrational frequencies of the heat equation. The analysis of the spectral properties of a graph thus provide considerable insight into the dynamics of the system as a whole.

   In this section, we brush an overview of various spectral distances that can be used for analyses pertaining global graph properties. Such distances have been well studied and developed in the literature [Banerjee and Jost (2008), Jost and Joy (2002), Jurman et al. (2012)]. In Jurman, Visintainer and Furlanello (2011), for instance, the authors provide an interesting review of several such spectral distances. This deviates slightly from our original setup: Spectral distances are permutation invariant and do not take into account the fact that nodes have been endowed with a particular identity. In fact, such distances can be used to compare any set of graphs, provided that they all share the same number of nodes. Spectral distances are unable to distinguish between isospectral graphs and are in fact pseudo-distances

rather than actual distances. However, as the probability of having distinct graphs with identical eigenspectra quickly dwindles as the number of nodes increases, spectral distances are also viable candidates for studying the dynamics of a given complex system through time. While most current algorithms can compute eigenvalue decompositions in $O(N^3)$ steps, some computational tricks bring down this cost to $O(N^2)$ [Thüne (2012)] thus making this spectral approach an appealing, computationally tractable alternative for defining graph dissimilarities [Banerjee (2008), Gu et al. (2016)].

3.1. ℓ_p distances on the eigenvalues.

Definition. We begin by introducing a general class of versatile spectral distances. A first natural candidate for comparing two graphs based on their eigenvalue decomposition is to choose a representation of the graph (typically its adjacency matrix, combinatorial or normalized Laplacian, etc.) and to simply consider the ℓ_p distance between functions of their eigenspectra. For the simplest case, where the function is just the identity, this would be of the form

$$d(G, \tilde{G})^p = \sum_{i=0}^{N-1} |\lambda_i - \tilde{\lambda}_i|^p.$$

For any (almost everywhere) differentiable function of the graph's eigenvalues $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{N-1}$, we can write

$$(3.1) \qquad d_f(G, \tilde{G})^p = \sum_{i=0}^{N-1} |f(\lambda_i + \varepsilon_i) - f(\lambda_i)|^p \approx \sum_{i=0}^{N-1} |f'(\lambda_i)|^p |\varepsilon_i|^p.$$

An important step thus consists in picking an adequate representation of the graph: how to decide between using the graph's adjacency matrix, its Laplacian or normalized Laplacian. These representations and the relationship between their eigenvalues and properties of the graph (average degree, Cheeger constant, etc.) have been well investigated in the literature [Cvetković (2012)], yet no consensus as to which representation yields more accurate results for comparing graphs has been established. To try and resolve this issue, we suggest the following guidelines:

- *leveraging the representation's physical interpretation.* As underlined in the introductory paragraph of this section, both the eigenvalues of the Laplacian and those of the adjacency matrix can be related to physical properties of a system and can thus be considered as characteristics of its states. Whenever such a physical interpretation exists, a good choice thus lies in the selection of the corresponding representation.
- *opting for the most robust alternative.* The adjacency matrix does not downweight any changes and treats all nodes equivalently. On the other hand, the eigenspectrum of the Laplacian accounts for the degree of the nodes and is known to be robust to most perturbations: A "small" perturbation of the graph,

that is, a perturbation that has very little impact on the graph's overall connectivity will only induce a small change in the eigenvalues [Spielman (2007)], thus making them a more attractive alternative for comparing graph structures.

- *choosing a stable representation.* The literature remains divided on which version of the Laplacian to pick. We believe that the eigenvalues of the normalized Laplacian being bounded between 0 and 2 make it a more stable and preferable representation.

Application. Let us look at the spectral distances on our microbiome and brain data. Figure 4 shows the results for the ℓ_2 distance using two different functions of the combinatorial Laplacian eigenspectrum in equation (3.1) on our datasets: the low-pass filters with randomly chosen parameters $f(\lambda) = e^{-0.1\lambda}$ (Microbiome, top row of Figure 4) and $f(\lambda) = e^{-1.2\lambda}$ (rs-fMRI, bottom row of Figure 4). The first interesting observation that we make is that these distances produce different results than the previous distances. In particular, we note that for $f(\lambda) = e^{-0.1\lambda}$, the nearest-neighbor metagraph is significantly associated with the treatment-stage labels: The p-value associated to the analysis-of-variance type test yields a value of 0.025, which is confirmed by the MDS projections [Figure 4(B)] showing a clear
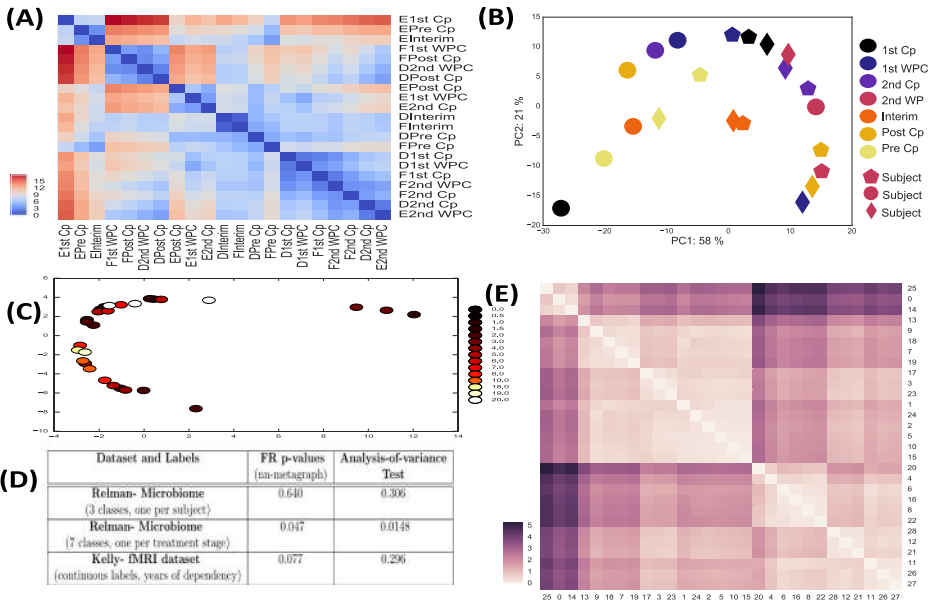

FIG. 4. *Application of ℓ_2 spectral distances using two functions of the eigenspectra as in equation* (3.1): *low-pass filters $f(\lambda) = e^{-0.1\lambda}$ on the Microbiome (top row) and $f(\lambda) = e^{-1.2\lambda}$ on the fMRI dataset (bottom row). Clustermap of the corresponding distances between bacterial graphs (A)/brain connectomes (E). MDS projection of the bacterial (B)/fMRI (C) graphs on the first two principal axes. Colors denote treatment phases/years of dependency. Pvalue of the FR-test for the 1-nn metagraph across the different datasets, for the low-pass filter $f(\lambda) = e^{-1.2\lambda}$ (D).*

grouping of the graphs per subject. This effect subsides as the scaling value increases while its association with the treatment stage becomes predominant. For $f(\lambda) = e^{-1.2\lambda}$, the analysis-of-variance test (with stages as labels) yields a p-value of 0.015. We also note that this distance is the only one which recovers some meaningful associations between the nn-metagraph of the fMRI data and the age under dependency (Figure 4D). We also note that the choice of the representation matters: The fMRI dataset analyzed through the scope of spectral distances based on the adjacency matrix failed to reveal any significant effect. Note that in the microbiome example and the synthetic experiments detailed in Section 5, the choice of one representation over another was mitigated, with both the Laplacian and the adjacency matrix yielding comparable results. Moreover, as underlined above, the choice of the function itself can lead to the discovery of different effects.

To understand this phenomenon, we build upon the signal processing analogy developed in Shuman et al. (2013). In this paper, Shuman and coauthors show that the eigenvalues of the Laplacian can be interpreted as the analog of a signal's frequencies in the temporal domain. In this case, each node's label (or feature representation) can be understood as the value of a signal propagating over the graph at that precise node location. Low eigenvalues and their corresponding eigenvectors are analogous to slowly-varying low-frequency signals over the graph: If two vertices are connected by an edge with a large weight, the values of the signal at those locations are likely to be similar. By contrast, the eigenvectors associated to high eigenvalues vary more rapidly across edges [Shuman, Ricaud and Vandergheynst (2016), Tremblay and Borgnat (2014)]. Hence, "low" eigenvectors encapsulate local information about the structure of the graph (yielding results akin to the Jaccard distance in the microbiome example) while higher values of αs cover a larger portion of the spectrum and allow the incorporation of more global information.

Discussion. We continue upon the signal processing analogy to find an appropriate choice of the function f:

- If the goal of the analysis is to capture *the importance of the changes in the connectivity of the overall graph structure*, the distance should put more emphasis on the first eigenvectors. An adequate choice for f would be thus to select f to act as a low-pass filter: putting more weight on changes occurring in small eigenvalues, and discounting the effect of changes at the higher end of the spectrum. The strength of the modulation of the eigenvalues by the filter depends on the analysis. For instance, for λ_3 the third eigenvalue, taking $f : x \rightarrow e^{-\alpha x/\lambda_3}$ ensures associating a weight of at most $\frac{\alpha}{\lambda_3}\varepsilon^{-\alpha}$ in equation (3.1) on changes in eigenvalues greater or equal to λ_3. In the case where $\lambda_2 \ll \lambda_3$, this gives more importance to changes occurring in λ_2. In our microbiome study, as previously highlighted, we recover more structure in the dataset by focusing on the lower part of the spectrum (Figure 4) and discarding the signal carried by the higher frequency, "noisier" eigenvalues.

- Supposing that one is interested in the *overall change in the "graph's frequencies" at every level* induced by the perturbation, one might actually prefer to take a function that would not discriminate against any value of the eigenfunction, but simply look at the amplitude of the change in eigenvalue. In that case, f could simply be taken to be the identity.

This section has shown the possibility of crafting a distance based on the Laplacian eigenspectrum, tailored to the requirements and objectives of the analysis. However, choosing "an optimal" kernel function for the problem at hand requires domain knowledge or additional insight into the problem thus requiring more thought than the straightforward Hamming distance.

3.2. *Spanning tree dissimilarities.*

Definition. Inspired by Kelmans (1976, 1997), who characterized transformations by their "ability to destroy," we now introduce a similarity which reflects the number of spanning trees that are destroyed or created by the transformation of one graph to another.

The Matrix-Tree theorem provides us with a convenient way of computing the number of spanning trees for a connected graph: denoting $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}$ the eigenvalues of the graph Laplacian $L = D - A$, we have

$$\mathcal{T}_G = N_{\text{Spanning tree of } G} = \frac{1}{N} \prod_{i=1}^{N-1} \lambda_i.$$

A dissimilarity between two graphs G and $\tilde{G}$ can be defined by comparing the quantities

$$(3.2) \qquad d_{\text{ST}}(G, \tilde{G}) = \left| \log(\mathcal{T}_G) - \log(\mathcal{T}_{\tilde{G}}) \right|.$$

On an intuitive level, spanning trees are a reflection of the graph's interconnectedness and robustness to change: to draw an analogy with electric current, this amounts to quantifying the effect of one edge deletion on the impedance of the system: How easily does the current still manage to flow?

Application. In this case, the results obtained using the spanning tree dissimilarity (denoted as ST dissimilarity in the rest of the text) are comparable to the results provided by the low-pass filter spectral distance described in the previous subsection. We also observe an interesting phenomenon: The nearest-neighbor metagraph induced by the ST dissimilarity on the microbiome data is fairly consistent with the treatment stages (with an associated FR-p-value of 0.16), but as we increase the number of neighbors, this effect becomes rapidly insignificant. The k-nearest neighbor graphs are significantly associated to the subject labels (below the 5% threshold). This effect is confirmed by the test described in Section 2.2, which yields a significant p-value of 0.035. This indicates that the ST dissimilarity does capture both similarities between treatment phases as well as across subjects. This
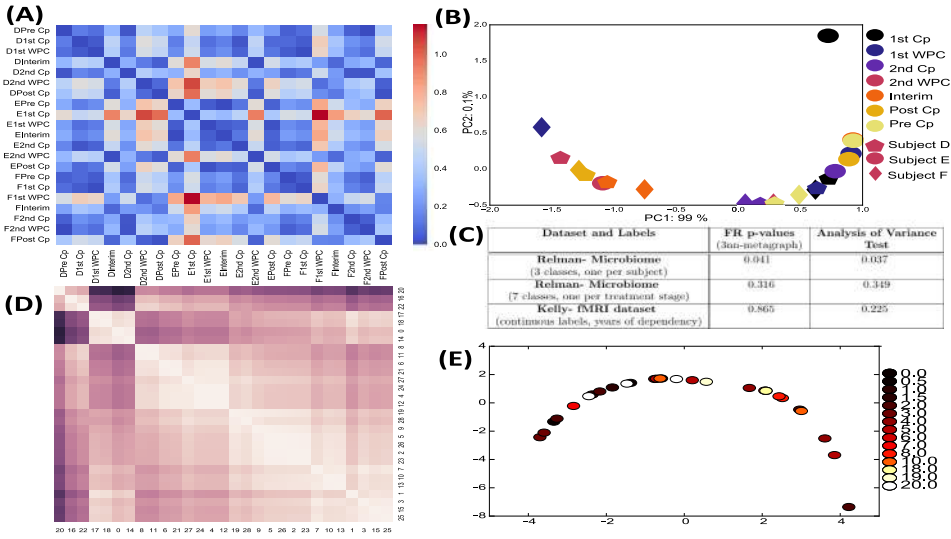
FIG. 5. *Application of the Spanning tree dissimilarity.* (*Top*) *Microbiome Data/*(*Bottom*) *fMRI Data. Heatmap of the corresponding dissimilarity between Kendall-correlation-based bacterial graphs* (A). *MDS projection of the bacterial graphs on the first two principal axes* (B). *Colors denote treatment phases, and shapes represent different subjects. P-values associated to Friedman–Rafsky test of the consistency of the 3-nn metagraph with the labeling of the nodes and analysis-of-variance test* (C). *Clustermap for the fMRI data* (D). *MDS projections of the brain connectomes on the first two principal components* (E).

effect can be visualized in Figure 5(B): The MDS projections of the microbiome graphs along the first 2 principal coordinates follow a curve indicative of a gradient in higher dimensions. Points belonging to the same subject seem relatively close. As before, the Spanning tree distance recovers some structure in the fMRI datasets [lighter blocks along the diagonal in Figure 5(D)], although there is no evidence that these clusters are associated with the time under dependency.

Discussion. Suppose that graph G undergoes a "small" perturbation, yielding a new graph $\tilde{G} = \mathcal{T}(G)$. We know that the eigenvalues of $\tilde{G}$ can be written as a perturbed version of the eigenvalues of G, that is,

$$\forall i, \qquad \tilde{\lambda}_i = \lambda_i + \varepsilon_i.$$

Hence, we can write

$$\tilde{\mathcal{T}}_G = N_{\text{Spanning tree of } \tilde{G}}$$

$$= \frac{1}{N} \prod_{i=1}^{N-1} \tilde{\lambda}_i$$

(3.3)

$$= \mathcal{T}_G \times \left[1 + \sum_{i=1}^{N-1} \frac{\varepsilon_i}{\lambda_i} + \sum_{i,j=1}^{N-1} \frac{\varepsilon_i \varepsilon_j}{\lambda_i \lambda_j} + \cdots \right].$$

Combining (3.3) and (3.2) yields

$$(3.4) \qquad d_{\mathrm{ST}}(G, \tilde{G}) = \left| \log\left(1 + \sum_{i=1}^{N-1} \frac{\varepsilon_i}{\lambda_i} + \sum_{i,j=1}^{N-1} \frac{\varepsilon_i \varepsilon_j}{\lambda_i \lambda_j} + \cdots \right) \right|.$$

The impact of the change is thus inversely proportional to the value of the eigenvalues. This is an attractive property for weakly connected graphs (i.e., that have small λ_1), where the addition or deletion of a critical edge can have a huge impact on the graph's overall connectivity. Conversely, changes in larger eigenvalues have less impact: To continue with the temporal frequency analogy drawn in the previous section, this similarity automatically discounts changes that are related to noise, and accentuates the impact of changes on low eigenvalues which are considered to be more reflective of the graph's structure. We emphasize that, once again, this defines a pseudo-distance between eigenspectra (or dissimilarity score between graphs), rather than a distance on the graphs themselves. This effect is shown in Figure 5, which exhibit results close to the low-pass filter approach developed in the previous subsection. The advantage of the ST dissimilarity is that it does not require the specification of a particular ad-hoc low-pass kernel on the eigenspectrum. This comes at an increased price in terms of the variability of the results: the effect of perturbations is measured with respect to the inverse of the eigenvalues [equation (3.4)], this distance is less stable than the low-pass filter spectral distance. We will study this in more depth in our synthetic experiments in Section 5.

3.3. *Distances based on the eigenspectrum distributions*.

3.3.1. *General framework.* Rather than focusing on the graph's eigenspectra, another alternative proposed by Jurman, Visintainer and Furlanello (2011), Gu et al. (2016) considers *continuous spectral distributions*. The continuous spectral distribution is obtained from each graph by computing the graph's eigenvalues and considering a kernelized version of its eigenvalue distribution. For a Gaussian kernel, the spectral distribution is defined as

$$\rho_G(x) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda_i)^2}{2\sigma^2}}.$$

A pseudo-distance between graphs is based on the distance between spectrum distributions, which, in the case of the ℓ_1-distance used in Gu et al. (2016), yields the following expression:

$$d(G, \tilde{G}) = \int \left| \rho_G(x) - \rho_{G'}(x) \right| dx.$$

In their 2016 article [Gu et al. (2016)], Gu and coauthors show that, in the limit of an infinite number of nodes, these distances have the added benefit of distinguishing between different types of graphs (Erdős–Rényi versus Preferential Attachment, etc.). As such, these distances are able to recognize important geometrical

information in the overall graph structure. We now investigate a variant of such a class of distances. Proposed in Ipsen and Mikhailov (2002), it has been shown to exhibit interesting properties [Jurman et al. (2015)]: the Ipsen–Mikhailov distance.

3.3.2. *Definition of the IM distance.*

Definition. First introduced by Ipsen [Ipsen and Mikhailov (2002)] for graph reconstruction and later extended to the broader "graph-comparison" problem by Jurman and coauthors [Jurman, Visintainer and Furlanello (2011), Jurman et al. (2015, 2016)], the Ipsen–Mikhailov distance is a spectral measure which relates a network on N nodes to a system with N molecules connected by elastic strings. The connections are dictated by the graph's adjacency matrix A and the system can thus be described by a set of N equilibrium equations:

$$\frac{\partial^2 x_i}{\partial t^2} + \sum_{j \neq i} A_{ij}(x_i - x_j) = 0.$$

In this setting, the eigenvalues of the Laplacian matrix of the network are interpreted as the squares of the vibrational frequencies ω_i of the system: $\lambda_i = \omega_i^2$ with $\lambda_0 = \omega_0 = 0$.

The Ipsen–Mikhailov distance characterizes the difference between two graphs by comparing *their spectral densities*, rather than the raw eigenvalues themselves. The spectral density of a graph is defined as the sum of Lorenz distributions,

$$\rho(\omega, \gamma) = K \sum_{i=1}^{N-1} \frac{\gamma}{\gamma^2 + (\omega - \omega_i)^2},$$

where γ is a parameter common to all vibrational frequencies that we will have to determine, and K is the normalization constant defined such that $\int_0^\infty \rho(\omega, \gamma) \, d\omega = 1$. This spectral distance between two graphs A and B is defined as

(3.5) $$\varepsilon_\gamma(A, B) = \sqrt{\int_0^\infty \left[\rho_A(\omega, \gamma) - \rho_B(\omega, \gamma)\right]^2 d\omega}.$$

The latter expression depends on the choice of the scale parameter γ. Jurman and coauthors [Jurman, Visintainer and Furlanello (2011)] set $\gamma = \bar{\gamma}$ as the unique solution of

$$\varepsilon_{\bar{\gamma}}(\mathcal{E}_N, \mathcal{F}_N) = 1.$$

So the IM distance is bounded between 0 and 1 and its upper bound is attained only for $\{A, B\} = \{\mathcal{E}_N, \mathcal{F}_N\}$ where $\mathcal{E}_N$ denotes the empty graph and $\mathcal{F}_N$ the complete graph on N nodes. In the Appendix [Donnat and Holmes (2018)], we investigate a closed-form formula for these parameters.

3.3.3. *The Hamming–Ipsen–Mikhailov distance*.

Definition. So far, none of these spectral distances have used the fact that particular nodes can be matched. There is no way of discriminating changes (i.e., emphasizing changes in areas of the graph deemed important to the analyst), or of accounting for rare—but existing—isospectral graphs.

To bridge the two approaches, Jurman et al. (2015) propose a distance that is a weighted linear combination of the Ipsen–Mikhailov and the normalized Hamming:

$$d_{\text{HIM}}^{\xi} = \frac{1}{\sqrt{1+\xi}}\sqrt{\text{IM}^2 + \xi H^2}.$$

Discussion. This distance benefits from the advantages of both the Hamming and the Ipsen–Mikhailov distances by combining local and global information. Note that, since it is a linear combination of a distance with a nonnegative quantity, this defines a proper distance between graphs. The parameter ξ provides additional flexibility to the metric by allowing to favor one type of information over another. However, empirically, we have observed this distance to be computationally expensive, and thus difficult to apply to the study of large graphs and/or large datasets.

Application. Figure 6 shows the results of the analysis using the HIM distance on our microbiome study. The MDS projection [Figure 6(A)] seem to highlight a similarity between graphs corresponding to the same treatment. The Friedman–Rafsky test on the minimum spanning tree with the treatment phases as labels is significant, with a p-value of 0.00048. This is further confirmed by the analysis of variance test described in Section 2.2 with the stages as labels, yielding a p-value of $p = 10^{-5}$. As also shown by Figure 6(C), the HIM is able to make the best of both the Hamming and spectral distances, and is thus able to spot more structure in the datasets. Overall, because these spectral distances are "unlocalized" and make no use of the nodes' identities, they are suited to the comparison of graphs' overall structure without any prior on where "critical" changes occur in the spectrum.
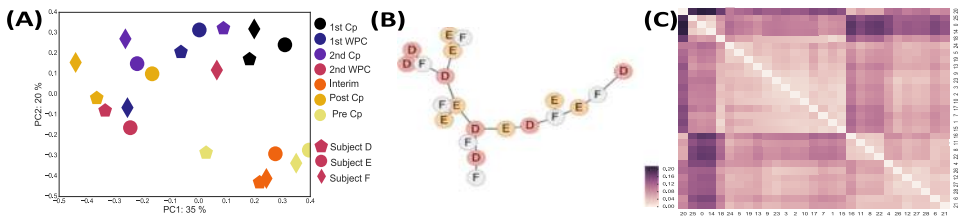

FIG. 6.   *Application of the Hamming–Ipsen–Mikhailov distance. MDS projection of the bacterial graphs on the first two principal axes* (A). *Colors denote treatment phases, and shapes represent different subjects. Minimum Spanning Tree induced on the bacterial graphs by the HIM distance* (B). *Application of the HIM distance to the fMRI data set: clustermap of the different distances between connectomes* (C).

On an aside note, both the IM and HIM distances were the lengthiest to compute perhaps restricting their scope of use to the comparison of small sets of reasonably-sized graphs.

3.4. *The polynomial approach.*

Definition and motivation. The previous spectral distances all shared a common problem: They require an explicit computation of the graph's eigenvalues—which, computational tricks aside, still generally has complexity $O(N^3)$—and is sensitive to global properties of the graph (as captured by the eigenvalues). Structural distances (Hamming and Jaccard), however, were too short-sighted and concentrated on changes in each node's direct neighborhood. Another interesting type of distances would thus be at an intermediate scale, and compare changes in local neighborhoods. For instance, changes in sparse regions of the graphs might be more informative than perturbations in very dense ones. Following Koutra and coauthors' [Koutra et al. (2016)] guidelines for distance selection, a "good" similarity score should be able to capture such nuances and attribute more weight to changes in areas of the graphs deemed more critical by the data analyst.

In this new setup, a possible solution is to work directly with the powers of the graphs' adjacency matrix A^k. Indeed, the powers of the adjacency matrix relate directly to a graph's local topology through the coefficients A^k_{ij}, which corresponds to the number of paths (possibly with cycles) that start at i and arrive at j in k hops. Hence, by design, these powers are inherently local. The coefficients A^k_{ij} can be thought of as a characterization of the connectivity between two nodes with respect to the k-hop neighborhoods: nodes i and j at distance greater than k hops have connectivity index $A^k_{ij} = 0$, whereas nodes within each other's k-hop neighborhood will typically have high connectivity index A^k_{ij} if the neighborhood is dense, and lower A^k_{ij} if the region is sparse. As such, the powers of the adjacency matrix seem to offer an attractive starting point to quantify changes on the mesoscale.

Typically, for each neighborhood (centered around a node a), perturbations should be assigned weights that are monotonically decreasing functions of the distance: A perturbation has higher impact in the local neighborhood if it is closer to the center than the periphery. In this spirit, denoting as $A = Q\Lambda_A Q^T$ the eigenvalue decomposition of the adjacency matrix A of a given graph, a proposed similarity score is defined with a polynomial $P(x) = x + \frac{1}{(N-1)^\alpha}x^2 + \cdots + \frac{1}{(N-1)^{\alpha(K-1)}}x^K$ of the adjacency

$$P(A) = QWQ^T,$$

where $W = \Lambda_A + \frac{1}{(N-1)^\alpha}\Lambda_A^2 + \cdots + \frac{1}{(N-1)^{\alpha(K-1)}}\Lambda_A^K$.

The distance between two graphs G_1 and G_2 can simply be computed by comparing the polynomials of their associated adjacency matrices A_1 and A_2:

$$d_{\text{pol1}}(G_1, G_2) = \frac{1}{N^2} \| P(A_1) - P(A_2) \|_{2,2},$$

(3.6)

$$\text{where } \| M_1 - M_2 \|_{2,2} = \left(\sum_{i,j} |M_{i,j}^1 - M_{i,j}^2|^2 \right)^{\frac{1}{2}}.$$

In a way, this distance is a straightforward extension of the Hamming distance to the mesoscale: rather than looking at perturbations at the atomic level counting the number of removed and inserted edges without assessing the effect of the perturbation on the overall structure, this polynomial distance compares neighborhoods of larger sizes, and thus attempt to capture the effect of perturbation at an intermediate scale. The weighting factor α is a way of discounting "peripheral" changes in neighborhoods of larger sizes with respect to neighborhoods of smaller size. We note that equation (3.6) is just a proposed class of polynomial distances, but this set of distance can be more broadly customized to a specific problem at hand, including domain knowledge to choose the size of the neighborhood, etc.

Application. Figure 7 shows the application of the polynomial distance to the microbiome data. Similar to the Hamming distance, the polynomial distance does detect significant similar dynamics across subjects [closely matching curves in Figure 7(C)]. In this case, the polynomial approach seems a weak compromise between structural and spectral distances, and does not benefit from any of their advantages: the Polynomial distance is neither significantly associated to states or subjects (as per the associated Friedman–Rafsky and analysis-of-variance type tests).
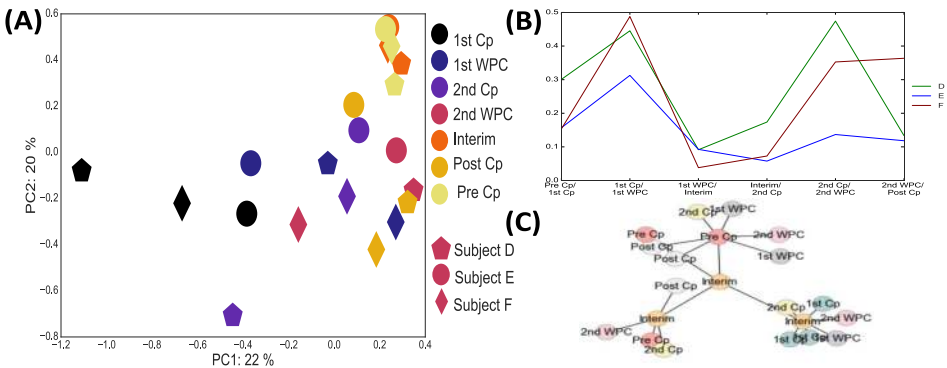

FIG. 7.   *Application of the polynomial dissimilarity to the microbiome bacterial graphs, for $K = 3$, $\alpha = 0.9$. Heatmap of the corresponding dissimilarity (A). MDS projection of the bacterial graphs (B) on the first two principal axes. Colors denote treatment phases, and shapes represent different subjects. Plots of the consecutive distances between bacterial graphs (C).*

Comparison of polynomial, spectral, and structural distances. The main advantage of the polynomial distances over the Hamming and Jaccard distances is that the former takes into account the properties of each node past each node's immediate neighborhood. Indeed, it is possible to show that the effect of a perturbation (i.e., the addition of one edge) on the graph can be directly related to properties of the graph at a higher order than simply the one-hop neighborhood. By construction, polynomials of order k reflect the effect of the perturbation on k-hop neighborhoods. They can be expressed in terms of polynomials of the degree of the nodes of the added edge, as well as the size of the intersection of their neighborhoods (up to size k). Hence, while its form (powers of the Laplacian) make it an intrinsically local distance, the polynomial distance is a first step toward bridging purely structural and spectral distances, by extending the Hamming distance to neighborhoods of greater depth. This does come at an increased price: The real-application studies show this distance to be blurred by both the perturbation of the organizational structure of the microbiome from one phase to the next and the variability of the bacteria across subjects.

4. Quantifying change at the mesoscale. Most of the distances described in the previous sections can be considered as extremes: Structural distances have proven to be too "local" and agnostic to perturbations' effects on a given complex system's organization as a whole. Spectral distances are global and fail to use the information captured in the nodes' identities. However, in several cases, the information of interest to the analyst lies somewhere in between. In brain connectome data for instance, while the individual neuron's activities might be too noisy to yield any significant results, the neighborhoods (in this case, larger regions of the brain) shed some light on the effect of such or such drug on the nervous system. Similarly, in microbiome networks, analyses focusing on small modules of bacteria might alleviate some of the noise due to "reading errors." Both of these examples are case in point where an analysis of the graph at the neighborhood level (or "mesoscale") adds strength to the analysis.

On the other hand, polynomial distances—which we had originally proposed as a scalable version of eigenvalue-based distances—quantify changes with respect to the k-hop neighborhoods, and have shown promising properties in both real and synthetic experiments: in the case where nodes' identities hold some insightful information, this extension of standard structural metrics seems to have brought a solution, trading off between the locality of the changes and their impact on the organization of the system as a whole. This indicates that considerable insight can be gained by comparing graphs at this intermediate "neighborhood" scale. This approach thus calls for the need for characterizing topological properties of these neighborhoods. In this section, we investigate the graph comparison through a "glocal" lens [borrowing an expression from Jurman et al. (2015)], extending the class of mesoscale polynomial distances introduced in Section 3.4 by suggesting two alternative characterizations of neighborhoods' topological properties.

4.1. *Quantifying interactions*: *Connectivity-based distances.*   We begin with a simple intuitive distance based on some measure of nodes' pairwise interactions. Indeed, as previously emphasized, we want a distance that: (a) preserves information about each node's identity and (b) incorporates information characterizing nodes by their relationship to the whole graph, rather than uniquely with respect to their direct neighbors. A general framework is to consider the dissimilarities between graphs at consecutive time points (G_t and G_{t+1}) as

$$(4.1) \qquad d_{\text{centrality}}(G_t, G_{t+1}) = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (s_{ij}^{(t+1)} - s_{ij}^{(t)})^p \right)^{1/p},$$

where s_{ij}^t is some measure of the interaction or affinity between nodes i and j in graph G_t. This dissimilarity metric thus quantifies how much the different interactions have changed from one graph to the other. This approach satisfies our constraints: It is both local and respects nodes' identities while accounting for the whole graph structure by summing over all pairwise "interaction" scores.

In the simplest, most intuitive case, we can simplify this expression by using centrality measures as our $s_{i,j}$'s. Indeed, centrality measures (betweenness, harmonic, etc.) can typically be used to characterize nodes as either belonging the core or the periphery of the graph, and thus encode global topological information on the status of the node within the graph. These metrics are thus natural candidates to characterize "mesoscopic" changes. More formally, in this setting, denoting $c_i^{(t)}$ as the betweenness-centrality of each node i in the graph at time t, one defines a distance between two graphs G_t and G_{t+1} as

$$(4.2) \qquad d_{\text{centrality}}(G_t, G_{t+1}) = \sqrt{\sum_{i=1}^{n} (c_i^{(t+1)} - c_i^{(t)})^2}.$$

One of the positive aspects of this metric is that centrality measures are "integrated" quantities; measuring the number of paths that typically pass through a given node. As such, similarly to eigenvalues, these metrics are more robust to small perturbations in the graph structure than the Hamming distance. Moreover, each "drift" measure in equation (4.2) is interpretable: A change in centrality can be understood as a drift of the node away from (or toward) the core of the network. However, the problems associated with this approach are two-fold. First of all, one has to choose a single "good" centrality measure (harmonic, betweenness, etc.), which may require domain-knowledge, since this captures a specific aspect of the network's evolution. Moreover, the computation of betweenness centrality on unweighted graphs typically requires algorithms with complexity $O(|\mathcal{V}||\mathcal{E}|)$. This approach is thus unfortunately, like the IM distance, difficult to extend to larger graphs.

In order to make this approach more tractable, recent work has proposed using approximation algorithms to compute alternative interaction metrics in equation

(4.1). For instance, in Papadimitriou, Dasdan and Garcia-Molina (2010) Papadimitriou and coauthors suggest five different scalable dissimilarities. In Koutra et al. (2016), Koutra and coauthors propose a low-dimensional approximation of these scores based on loopy belief propagation algorithm—yielding a method (DeltaCon) able to approximate equation (4.1) with a computational complexity linear on the number of edges in the graphs.

4.2. *Heat spectral wavelets.* Another alternative is to derive characterizations of each node's topological properties through a signal processing approach: The values of the nodes constitute a signal over the graph, which can be filtered by modulating the graph's spectrum. This yields a different similarity than in the eigenvalue-based setting: whereas in the previous section, eigenvalue distances simply computed a distance between the modulation of two graphs' eigenvalues; here, the eigenvalues are modulated and combined with their respective eigenvectors to yield a "filtered" representation of the graph's signal. Such an approach could follow work initially done by Monning and coauthors, who, in their recent 2016 paper [Monnig and Meyer (2018)], build upon the DeltaCon similarity to create a (proper) distance between graphs: they introduce the *Resistance Perturbation* index, a metric based on the eigenvalues and eigenvectors of a modified version of the graph Laplacian. In this section, we focus on a closer analogy to signal processing and use recent work in the graph signal processing literature to derive such characterizations. In this subsection, we focus on an approach inspired by Donnat et al. (2017) for the purpose of structural role identification. In that paper, inspired by the emerging field of graph signal processing [Shuman et al. (2013)], the authors suggest using heat spectral wavelets to characterize each node's local topology for the purpose of structural role identification. To use a concrete analogy, this method operates in a way similar to sonar detection: Each node probes the network by diffusing a heat wavelet, and the way that the network responds to each of these probes, that is, the different heat prints that are obtained for each node is taken as a signature for each of the nodes' topological neighborhoods (Figure 8).

More formally, denoting $L = U\Lambda U^T$ as the Laplacian's eigenvalue decomposition, where $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$, the heat scaling-wavelet [Shuman, Ricaud and Vandergheynst (2016)] $\Psi_{\cdot,a}^{(\tau)}$ centered at node a with scale τ is defined as the column vector of the matrix $Ue^{-\tau\Lambda}U^T$:

$$\Psi_{\cdot,a}^{(\tau)} = Ue^{-\tau\Lambda}U^T\delta_a \quad \Longrightarrow \quad \forall m,$$

$$\Psi_{m,a}^{(\tau)} = \delta_m^T Ue^{-\tau\Lambda}U^T\delta_a = \sum_{j=0}^{N-1} e^{-s\lambda_j}U_{aj}U_{mj},$$

where δ_m is the indicator vector associated to node m, and $e^{-\tau\Lambda}$ is the diagonal matrix $\text{Diag}(e^{-\tau\lambda_0}, e^{-\tau\lambda_1}, \ldots, e^{-\tau\lambda_{N-1}})$. In Donnat et al. (2017), the authors propose to define a structural signature for each node as the unordered set of coefficients:

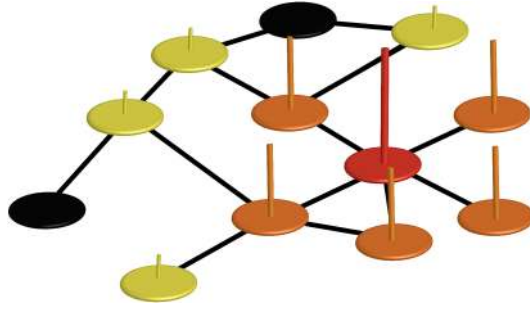$$(4.3) \qquad\qquad\qquad \chi_a = \{\Psi_{m,a}^{(\tau)}\}.$$

FIG. 8. *Representation of the heat kernel: the source node (red) has the highest temperature (interpreted here as a "signal value")—illustrated by the high red bar. The heat wave diffuses over the neighborhood (illustrated by the varying intensities of the nodes' colors and height of the bars representing the dwindling temperature/signal strengths).*

By comparing these distributions, one captures information on the connectedness and centrality of each node within the network, thereby providing a way to encompass in a Euclidean vector all the necessary information to characterize nodes' topological status within the graph. Moreover, in order to compute these wavelets in a tractable fashion that extends to large graphs, Hammond and coauthors [Hammond, Vandergheynst and Gribonval (2011)] suggest the use of Chebychev polynomial approximations. The cost of computing the wavelet transforms becomes simply $O(K|\mathcal{E}|)$—making spectral wavelets an attractive approach for characterizing structural roles. While these signatures were initially devised to detect structural similarities across a network, they can also be employed to characterize similarities across a set of aligned graphs. In this setting, network similarity between graphs G_t and G_s is defined by comparing each node's topological signature in G_t with its counterpart in G_s. Here, a large dissimilarity between graphs indicates either an important "volume" of change (as in the Hamming distance) or that some nodes have undergone important topological changes. It thus captures changes at both the fine and intermediary scales.

More formally, this dissimilarity between graphs amounts to a ℓ_2 distance between each node's structural embedding:

$$d(G_t, G_{t+1}) = \frac{1}{N} \sum_{a \in \mathcal{V}} \|r_a^{(t)} - r_a^{(t+1)}\|_2 = \frac{1}{N} \sum_{a \in \mathcal{V}} \delta_a \Delta^T \Delta \delta_a = \frac{1}{N} \text{Tr}[\Delta^T \Delta]$$

(4.4)

$$\text{where } \Delta = U_t e^{-\tau \Lambda_t} U_t^T - \tilde{U_{t+1}} e^{-\tau \Lambda_{t+1}} U_{t+1}^T.$$

To formalize the link with the previous subsection, we argue that these wavelet coefficients are in fact robust integrated centrality scores. Indeed heat kernels can be understood as a robust *page rank score* at each node [Chung (2007)]. By design, the heat kernel integrates over the neighborhoods (the size of which depends on the scale of the kernel) and is thus less sensitive to small perturbations. Hence, these

wavelet coefficients provide a tractable alternative to the centrality measures proposed in Section 4.1. As such, they benefit from these measures' interpretability, while being generalizable to larger networks.

Discussion. The advantages of this method can be summarized as follows.[4]

tractability: As already noted, the cost of computing the wavelets via a polynomial approximation is linear in the number of edges, making it a suitable approach for large sparse graphs.

granularity: Since this metric compares each node's status in the two graphs, this approach benefits from granular information which allows the possibility of identifying the nodes that have undergone the most drastic changes. This is particularly useful in a number of applications where the identification of the area of the graph which changed the most is also of interest (what bacteria radically changed, which neurons adopted a completely different role in the graph, etc.).

inclusion of "mesoscopic" information: The wavelets allow us to compare neighborhoods of the node at different scales automatically. This enables a less short-sighted representation of the overall graph structure than standard structural distances.

inclusion of "multiscale" information: The topological signatures of each node can be further enriched to contain multiscale information: In Donnat et al. (2017), a multiscale topological signature associated to scales $\{s_1, \ldots, s_j\}$ is defined as the concatenation of the representations: $\chi_a = [\Psi_a^{s_1}, \Psi_a^{s_2}, \ldots, \Psi_a^{s_j}]$. In this case, the heat-distance between two graphs is simply computed by replacing r_a in equation (4.4) by this new value χ_a. This allows a more robust representation of the topological role assumed by each node. An application of distances based on this multiscale signature on the recipes network is presented in Section 7.

4.3. *Application to the microbiome and fMRI study.* Figure 9 shows the results of the analysis of the microbiome study using a heat based distance with $\tau = 1.2$. Interestingly, this distance is one of the few showing a significant link between the years under dependency and the graphs in the fMRI datasets [Friedman–Rafsky for the 5-nn metagraph has a p-value below the 0.05 threshold, Figure 9(B)]. However, in the microbiome dataset, this distance is dominated by a clear subject effect [Figure 9(B), (D), and (E)]. This is further confirmed by the analysis-of-variance test with the subjects as labels described in Section 2.2 yields a p-value below 10^{-4}. We also note that this distance clearly indicates similar dynamics across subjects [Figure 9(C)].

---

[4]The connectivity-based distances from Section 4.1 also have these properties.
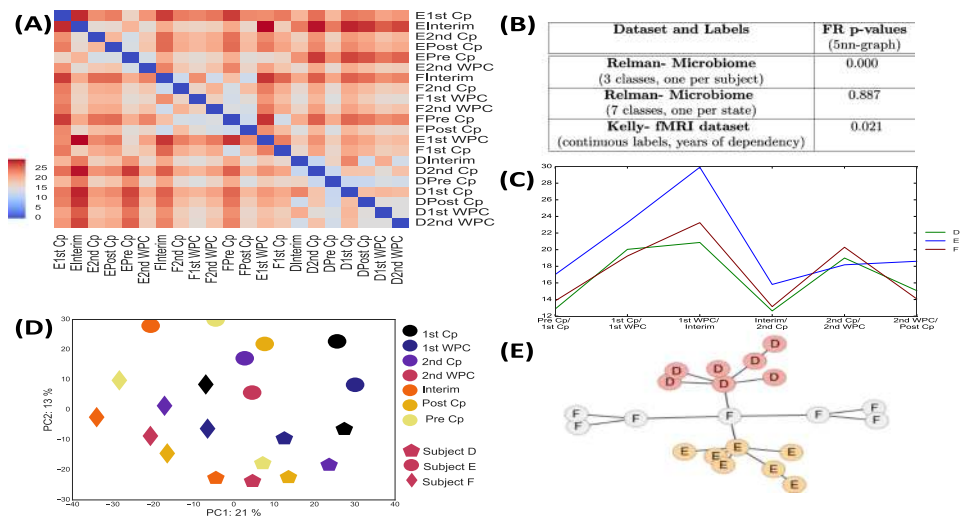
FIG. 9. *Application of the heat wavelet characteristic distance to the microbiome bacterial graphs,
for* $\tau = 1.2$. *Heatmap of the corresponding dissimilarity* (A). *P-values of the Friedman–Rafsky test
on the 5-nearest-neighbor graphs induced by the heat distance, on each dataset. Plots of the consecutive distances between bacterial graphs* (C). *MDS projection of the bacterial graphs* (D) *on the
first two principal axes. Colors denote treatment phases, and shapes represent different subjects.
Minimum Spanning tree induced on the bacterial graphs* (E).

5. Synthetic experiments. We use several synthetic experiments on model
graphs to enhance our understanding of the different distances' behavior and relative advantages. Synthetic experiments[5] have the benefit of offering a controlled
environment for testing the different distances' sensitivity relative to:

- *the graph topology*: we test the different distances on three types of graphs:
  - the Erdős–Rényi model (with $N = 81$ nodes, and a probability of connection
    $p = 0.1$);
  - a Preferential Attachment (PA) graph on $N = 81$ nodes (with $\alpha = 1$);
  - a Stochastic Block Model (SBM) graph, with 3 equally sampled communities
    and connection matrix:

$$C = \begin{pmatrix} 0.4 & 0.1 & 0.001 \\ 0.1 & 0.2 & 0.01 \\ 0.001 & 0.01 & 0.5 \end{pmatrix}.$$

These sets of network families present different global and local degree densities, and will allow us to assess the impact of the topology on the analysis of
network dynamics. The Erdős–Rényi graphs are denser than the preferential attachment graphs, which have an almost star-like structure with a few hubs. The

---

[5]The code for all synthetic and real experiments developed in this review is public and available
at: https://github.com/donnate/TrackingNetworks.

SBM model is somewhere between the two: There are three relatively dense cliques with only a few edges connecting them.

- *the perturbation mechanism*: In our first set of experiments, three initial graphs are generated according to a given topology: ER, preferential attachment, and SBM. We simulate network dynamics as follows: at each time step, for each graph, $\eta\%$ of the edges are removed and reconnected elsewhere (at random, following a preferential attachment model). In an attempt to replicate real life situations, we add to this procedure a "background" depletion/thickening process: edges are deleted with probability 0.015 and "formerly absent" edges are added with probability 0.015. How do the distances behave in this simple setting? We expect the curves for the distances depending on topological properties to be stable for denser graphs. In this case, modifications rarely impact the structure of the graph, but structural distances are large, because many edges are being moved. On the other hand, we might observe more instability in the plots for sparser graph structures, where the deletion of a critical edge can have a much stronger impact on the overall connectivity of the graph.

- *changes in the intensity of the perturbation mechanism*: In our second set of experiments, we want to assess metrics' sensitivity to changes in the dynamical process. At time $T = 0$, we generate on initial graph according to one of our three proposed topologies and simulate a dynamical mechanism as before in which, at each time step, 8.5% of edges are randomly rewired, and edges are randomly deleted or added with probability 0.015. At $T = 6$, the perturbation mechanism increases its rewiring probability to 34%, and its random deletion/addition probabilities to 6%. At $T = 13$, the process reverts back to its original characteristics. This induces three distinct time blocks in the time series. The aim here is to see which distances show the existence of a change-point in the graphs' dynamics.

In order to analyze the results, we:

1. quantify different distances' ability to cluster graphs belonging to the same time series: in the first set of experiments, for a low level of noise η, the distance should recognize graphs belonging to the same time series. To quantify this effect, we use agglomerative clustering on the distance matrix to recover 3 different clusters. We then compute the homogeneity and completeness of these clusters.

2. assess the consistency of the ordering of the graph induced by each distance with the time series: each graph at time t should be closer to its "parent" graph at time $t - 1$ and "child" graph at time $t + 1$.

3. estimate the ability of each distance to spot changes in the dynamic regime. To this effect, we visualize the heatmaps of the different distances, and compute the ratios $r_1 = \frac{\bar{D}_{12}}{\sqrt{D_{11}D_{22}}}$ and $r_2 = \frac{\bar{D}_{32}}{\sqrt{D_{33}D_{22}}}$, where $\bar{D}_{ij}$ denotes the average "between" (or "within", if $i = j$) distance between graphs in time chunk i and graphs in time chunk j.

The legend in each of the subsequent figures indicate the correspondence between curves and distances.

We note that in the SDM case, the eigenvalue-based distances, both the Laplacian and adjacency-based representation yield comparable results. We summarize our findings on the synthetic experiments as follows:

- *Smooth dynamical process*: We observe that the performance of the different distances is not consistent across topologies. In particular, the dynamic regime associated to the Preferential Attachment graph seems to yield more instability, and none of the distances are able to cluster graphs belonging to the same time series correctly. Nor do they yield an ordering of the graphs consistent with their ordering in the time series. On the other hand, in the case of the ER and SBM graphs, all structural distances (Hamming, Jaccard, polynomial), as well as some of the eigenvalue-based ones (IM) exhibit perfect recovery of the different clusters and the temporal ordering. This is probably due to the higher graph degree densities. Thus, the addition of an edge perturbs the overall structure less. In this case, it seems that the heat distance as well as the low pass filter emphasize similarities across graphs at a similar "state in time": The MDS projection shows a curve with closely clustered points representing graphs at identical times.
- *Change-point detection problem*: The structural distances exhibit better behavior (as shown by the clear blocks along the diagonal). We note that the ST dissimilarity and the polynomial distance have high r_1 and r_2 ratios, making them perfect candidates for detecting a change in regime.
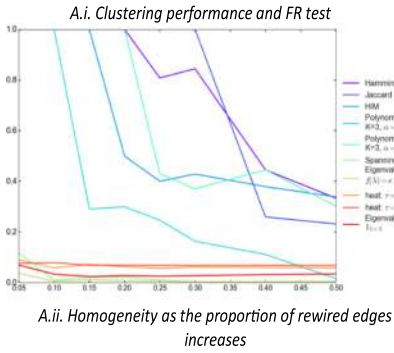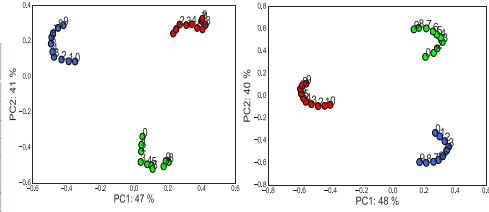
6. Summary: Discussion of the different distances.   We now conclude with a discussion of the strengths and relative advantages of the different distances using the results of both the synthetic and data-based analyses.

Overall, structural distances seem particularly well fitted for tracking temporal evolutions through time when the nodes' Ids are well-defined and hold a particular importance in the network. Indeed, these distances focus on some measure of the volume of edges that change from one graph to the other, and as such, are especially able to recognize graphs belonging to the same trajectory. This is useful in the microbiome case, where the Jaccard distance was able to recognize a strong "subject effect." The Hamming distance enriched by taking into consideration higher order information enables us to account for larger neighborhoods using the polynomial distances. In real datasets, unfortunately this distance still appears to suffer from the same drawback: All changes are treated equivalently across nodes, and the distance is blurred by the numerous changes, hindering its ability to correctly capture subtle similarities between graphs. On the other hand, as shown by the synthetic experiments (Figure 10, and Figures D.1 and D.2 of the Supplementary Material [Donnat and Holmes (2018)]), global distances, or mesoscale distance with large neighborhood scopes (e.g., large scaling factor τ in

**(I) Clustering Performance experiment (3 graphs, noise level σ=0.1)**

A. Clustering by graph (3 labels)

Dissimilarity	FR p-values (nn-graph)	Homogeneity	Completeness
Hamming	0.000	1.000	1.000
Jaccard	0.000	1.000	1.000
HIM	0.000	1.000	1.000
Spanning Tree	0.854	0.128	0.173
Polynomial $K=3, \alpha=0.5$	0.000	1.000	1.000
Heat, $\tau=0.1$	0.541	0.090	0.191
Eigenvalue-based $f(\lambda)=e^{-1.2\lambda}$	0.674	6.73e-02	7.62e-02

A.i. Clustering performance and FR test

B.i. Hamming Distance

B.ii. Polynomial (K=3, α=0.9)

A.ii. Homogeneity as the proportion of rewired edges increases

Dissimilarity	FR p-values
Hamming	0.000
Jaccard	0.000
HIM	0.000
Spanning Tree	0.000
Polynomial $K=3, \alpha=0.5$	0.000
Heat, $\tau=0.1$	0.011
Eigenvalue-based $f(\lambda)=e^{-3\lambda}$	0.267
Eigenvalue-based $f(\lambda)=1_{\lambda<2}\lambda$	1.10^{-4}

B.iv. FR-test of the consistency between the MST and time

B.iii. Eigen. Distance $f(\lambda)=\lambda 1_{\lambda<2}$
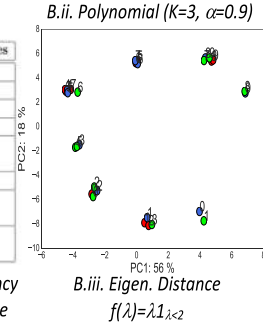
B. MDS projection on the first two principal components

(II). Change-Point Detection experiment (T=7 and T=13)

C.i. Ratio $r_1=D_{12}/(D_{11}D_{22})^{1/2}$. Green line indicates the boundary $r_1=1$

C.ii. Heatmap: Hamming distance

C.iii. Heat distance (τ=0.1)

C.iv. Polynomial distance (K=3, α=0.9)

C.v. Ratio $r_2=D_{23}/(D_{33}D_{22})^{1/2}$. Green line indicates the boundary $r_2=1$
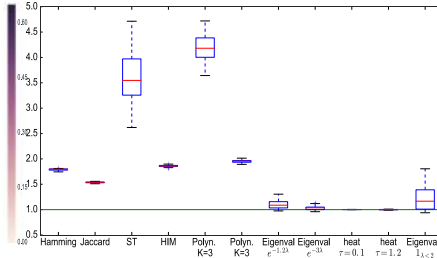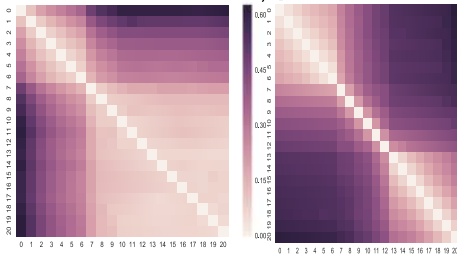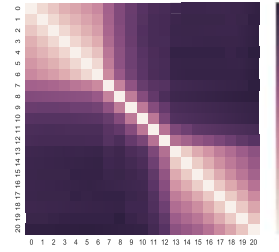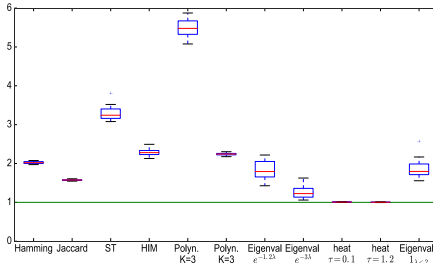
FIG. 10. *Results for the Stochastic Block Model topology. Top Row: Comparison of the smooth dynamics (no change point), with 0.05% edges rewired at each time step. Bottom Row: Change point detection experiment.*

the heat distance) seem less fit for the task of recognizing graph trajectories: These distances are more focused on comparing graph properties at a higher level.

The data-based experiments also show the advantages of using global spectral information. This is especially useful in real-life "noisy" settings where the correspondence between nodes from one graph to another is only approximative. In the fMRI dataset typically, the role played by one node in a graph can in fact be assumed by its neighbor in another. In that case, while these graphs will be classified as dissimilar by structural distances, spectral and mesoscale distances are able to recognize the similarity. This seems to explain the results obtained by the spectral and heat distances on the fMRI dataset, where some interesting associations between the one-nearest neighbor metagraph and the number of years under dependency are detected. The heat-based distance is promising avenue for achieving "glocality," that is, finding an appropriate middle ground between the very local, structural distances and the very global spectral ones. Indeed, the scaling factor controls for the propensity of the distance to take into account local information. In order to provide more intuition to this phenomenon, let us come back to the signal processing analogy of the Laplacian's eigenvalues: In this setting, the small eigenvalues of Laplacian correspond to low frequency signals, enforcing neighboring nodes to share similar values, while large eigenvalues correspond to fast varying signal across the graph's edges and can be compared to high-frequency noise. In this setting, the heat-kernel distances act by filtering out these noisy "high frequencies" of the signal over the graph and keeping only its low-frequency components, that is, information about neighborhoods. When applied to the microbiome example, for instance, heat-distances with lower scaling factors were able to recover the strong subject effect in the data. However, as the scaling factor τ increases, more global information is taken into account and, similar to the eigenvalue-based distances, the heat-kernel distance is then able to recover meaningful treatment stage effects.

In order to help the reader navigate all the choices we have illustrated in the current review, we have provided a decision tree (Figure 11) that summarizes the different distances, highlighting their relative advantages and drawbacks.

7. Case study for spatial dynamics: Worldwide recipe networks. In this final section, we illustrate an interesting extension from temporal to spatial dynamics through the study of a more entertaining final example: a worldwide recipe network. Each cuisine is modeled by a graph in which nodes represent ingredients, and edges measure their cooccurrence frequency in various recipes. The motivating intuition behind this graph-based cooccurrence is that cuisines can be better characterized by typical associations of ingredients. For instance, the Japanese cuisine might be characterized by a higher associativity of ingredients such as "rice" and "nori" than Greek cuisine. In our analysis, the graphs were obtained by processing the 57,691 recipes scraped from three different American culinary websites (*allrecipes*, *epicurious*, and *menupan.com*) in Ahn et al. (2011) as part as a study on
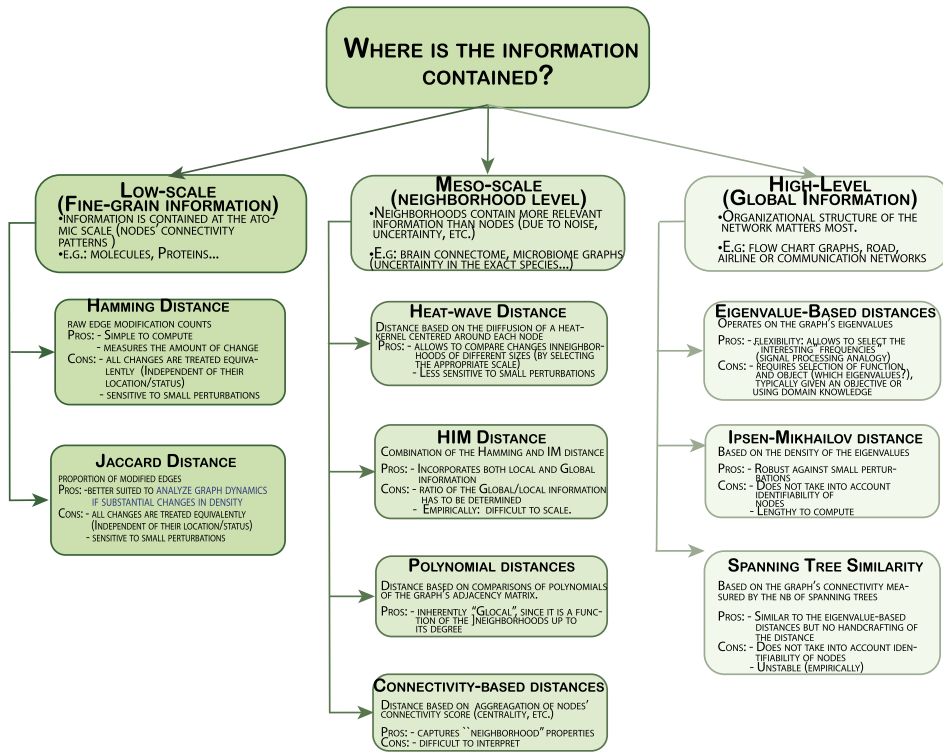
FIG. 11. *Summary of the distances detailed in the paper.*

food-pairing associations. We use the cooccurrence counts of 1530 different ingredients for 49 different cuisines (Chinese, American, French, etc.).[6] Each cuisine is then characterized by its own cooccurrence graph. The weight on the edge is the frequency of cooccurrence of the two ingredients in a given cuisine. The final graph for a given cuisine thus consists in a collection of disconnected nodes (ingredients that never appear in a single recipe for that cuisine) and a weighted connected component. The construction of these graphs is further discussed in the Appendix [Donnat and Holmes (2018)].

The goal of this analysis is to show which meaningful similarities can be captured by our different distances, and to highlight which distances are better suited to the comparison of the different cuisine graphs in this very sparse and unbalanced setting. In this case, natural groupings of cuisines are intuitive, and the results are thus easy to benchmark. Here, we use our inferred pairwise distance matrix between ingredient cooccurrence networks and evaluate our results by plotting both the heatmaps of the pairwise distances and constructing "3-nearest-cuisine metagraphs" (*k*-nn metagraphs, with $k = 3$) for each type of distance. In this graph,

---

[6]The data can be downloaded on the following link: http://yongyeol.com/pub/.

each node corresponds to a given cuisine c. The neighbors of cuisine c correspond to its three-nearest neighbors with respect to a given pairwise similarity matrix. This yields a directed graph of order 3, which we treat here as undirected, hence the degree of each node can be greater than 3 if the node is among the three nearest-neighbors of several cuisines. This provides a way of filtering the information contained in the distance matrix, and quickly visualizing whether the similarities recovered by the distance make intuitive sense.

Structural distances. We begin by analyzing the similarities captured by the Hamming and Jaccard distances. We note that these two structural distances yield very different results. The Jaccard 3-nearest-cuisine summary graph exhibits an interesting tri-cephalic structure [Figure 12(b)]: Almost every node in the graph
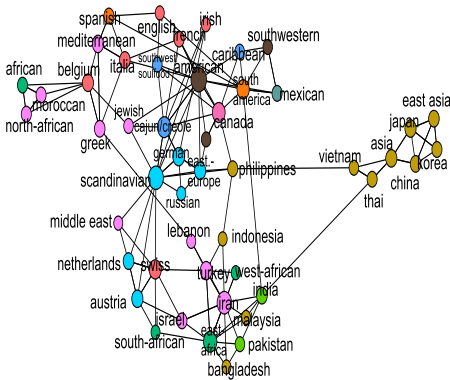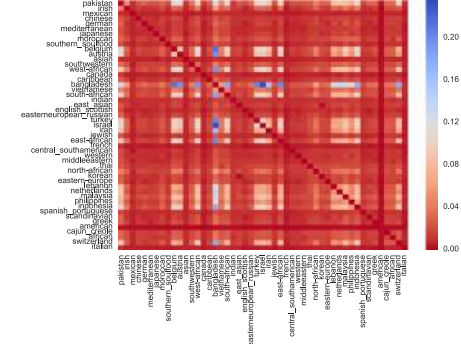

(a) 3-nearest-neighbor proximity graph between cuisines (Hamming distance)
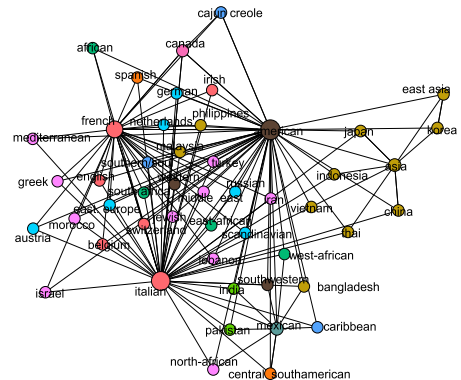
(b) 3-nearest-neighbor proximity graph between cuisines (Jaccard distance)
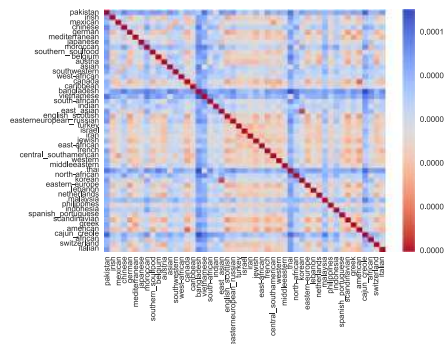
(c) Pairwise distances between cuisines (Hamming distance)

(d) Pairwise distances between cuisines (Jaccard distance)

FIG. 12.   *Comparison of the pairwise distances and three-nearest cuisine summary graphs (Left column: Hamming distance. Right column: Jaccard). The three-nearest cuisine summary graph is constructed by representing each cooccurrence network by a node and linking it to its three nearest neighbors according to a given pairwise similarity matrix.*

is connected to three main hubs (American, French, and Italian). This shows that the Jaccard similarity mostly captures the proportion of shared cooccurrences (as opposed to other network properties). Indeed, here, the American, Italian, and French cuisines have the largest connected components, hence the overlap with the other cuisines' connected components is greater. As such, the Jaccard distance fails to recover more subtle structure in the food network. At the other extreme, the Hamming distance recovers more structure than the Jaccard distances: It manages to recover clusters corresponding to East Asian and East European cuisines. We note here that the similarities are linked with the number of shared ingredients between two cuisine. In particular, the Bangladesh cuisine—whose connected component comprises only 22 ingredients) is uniformly far from the other graphs [Figure 12(c)]. The Hamming distance only reflects the overlap in connected components without accounting for the components relative sizes.

Spectral distances. Spectral distances also struggle to recover similarities between cuisines. In this example, the IM and HIM yielded the same 3-nearest-neighbor graphs (Figure 13). We note that this graph consists of two connected components, and does not follow the expected clustering of cuisines: the Mediterranean cuisines for instance (pink nodes in Figure 13) are scattered in each of the two clusters. This might be due to the fact that the IM distance fails in the presence of disconnected graphs, where the eigenvalue 0 has a high order of multiplicity for every graph: In particular, the Bangladesh graph, where 0 has multiplicity 1508, is very distant from the others. We note that Bangladesh sits unusually far from America, where 0 has order of multiplicity 1189. For the polynomial distances (Section 3.4), we have taken parameters $\alpha = 0.9$ and $K = 5$ (a study of $\alpha = 0.5$ and $K = 3$ has achieved the same results). We have also computed a eigenspectrum-based distance (Section 3.1) with $f(x) = e^{-0.9x}$ (using the adjacency matrix of the graph). The polynomial distance seems to recover clusters that are almost consistent with geographical proximities of the different cuisines. However, the lack
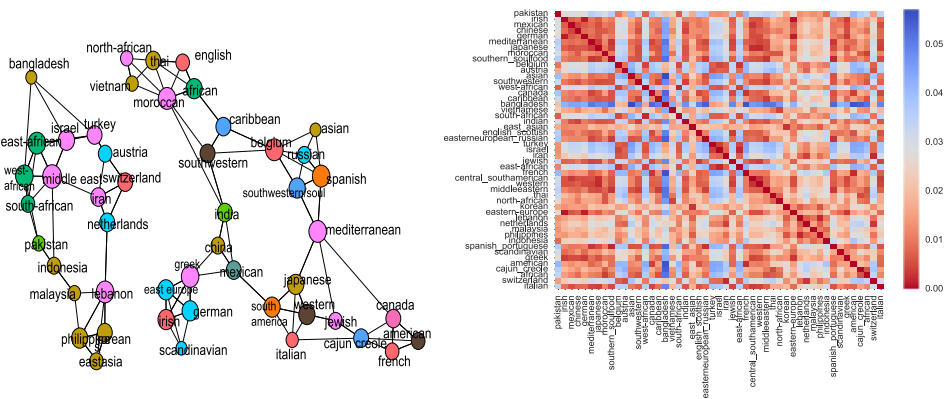

FIG. 13. *Ipsen–Mikhailov distance.* (*right*) 3-*nearest-neighbor proximity graph between cuisines* (*left*) *Pairwise distances between cuisines.*

(a) Pairwise distances between cuisines (polynomial distance)

(b) 3-nearest-neighbor proximity graph between cuisines (polynomial distance)

(c) Pairwise distances between cuisines (Eigen distance)

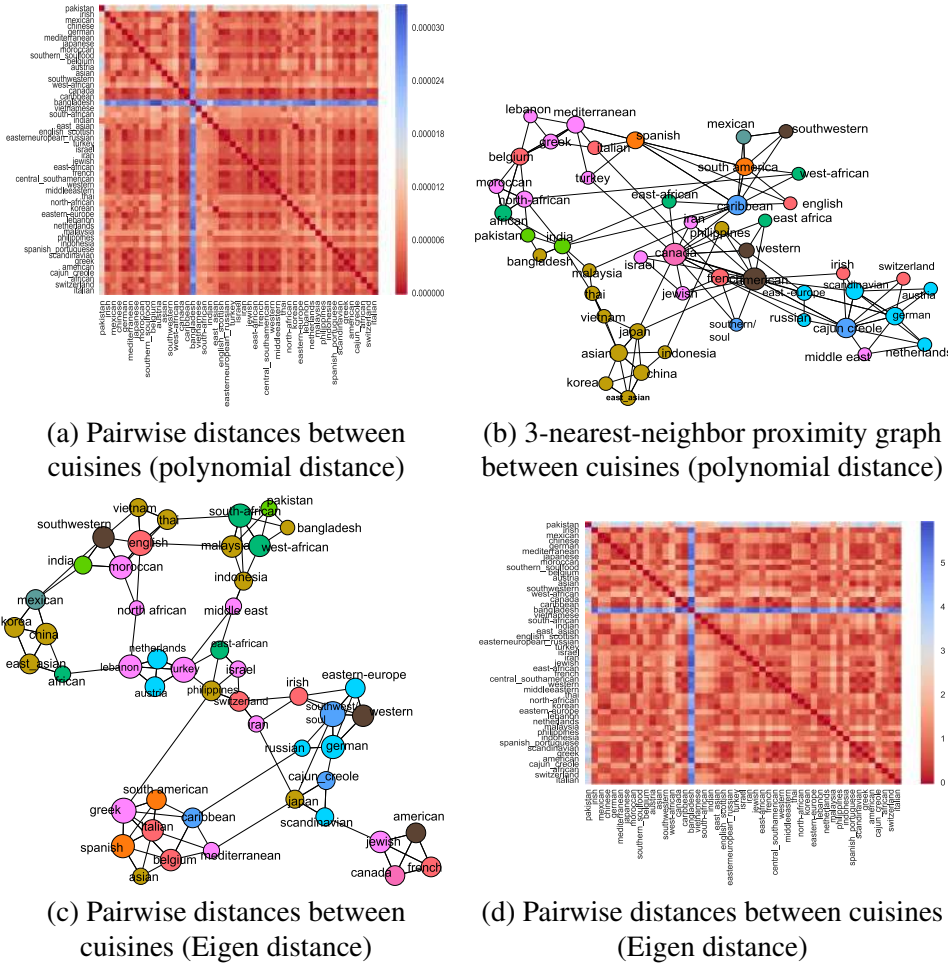(d) Pairwise distances between cuisines (Eigen distance)

FIG. 14. *Pairwise distances between cuisines for various spectral distances. Top row*: *Ipsen–Mikhailov distance.* 2*nd row*: *Polynomial distance* (*Section* 3.4), *with* $\alpha = 0.9$ *and* $K = 5$. *Bottom row*: *Eigenspectrum-based distance* (*Section* 3.1) *with* $f(x) = e^{-0.9x}$.

of structure (no block elements or pronounced groupings) apparent from the heat maps [Figure 14(a) and (d)] highlights the fact that spectral distances struggle to find definite patterns in this dataset. We thus conclude that a distance based on eigenvalues seems to achieve very unconvincing results for the study of graphs with many disconnected components: In this case, comparing the structure of the graph is insufficient, and we need to include information contained in the nodes' labels.

Wavelet distances. In this case, we have computed the heat wavelet signatures for each node according to their multiscale version described in Section 4. The scale s was chosen to take values in $\{1, 2, \ldots, 29\}$.
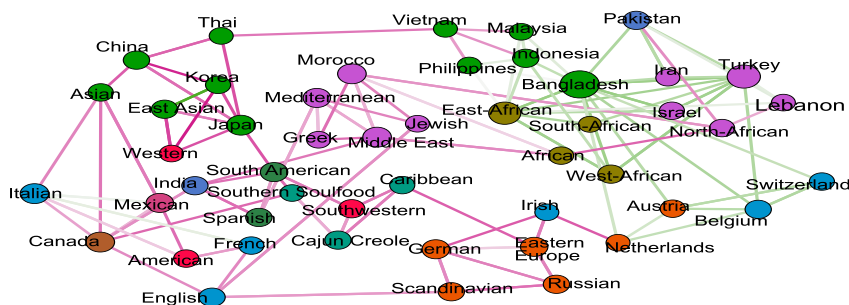
FIG. 15. *Proximity graph between cuisines (heat-wavelet based distance).*

Figure 15 shows the 3-nearest cuisine metagraph that this distance yields. We see that the graph that we are able to recover is consistent with geographical proximities would expect. We note for instance the clusters of Scandinavian cuisines and southwestern European cuisines, as well as a high proximity of Mediterranean cuisines and Asian cuisines. It is interesting to note that this approach puts Bangladesh cuisine with high centrality. This is due to the limited number of recipes that we have for Bangladesh cuisine yielding more homogeneous and higher edges weights, and thus seemingly closer distance to the other graphs.

8. Conclusion. We have given an overview of different metrics and similarity measures for comparing graphs for which we have node labels. Graphs created in real applications rarely have exchangeable nodes and our main focus here has been to supplement the ample literature on permutation invariant graph distances with more refined ones that account for these node identities. Our main focus has been to reflect on the types of changes in dynamic and spatial scenarios that these distances are best suited for. We have provided highlights of these performances on both synthetic and real-data graph analyses.

We have illustrated the use of these metrics for doing statistics on graph-objects in much the same way we did for binary rooted trees in Chakerian and Holmes (2012). Finding the right distance for the problem at hand can enable us to further our analyses by constructing the Fréchet mean graph or decomposing the sums of squares of distances between graphs enabling the type of analysis-of-variance type tests that we illustrated in this article.

Distances are useful in assessing many sources of variability in a dataset and as we have shown, can even detect the existence of change-points in dynamics of complex systems such as microbial communities. Pairwise dissimilarity matrices can be used to draw heatmaps (and visualizing the existence—or lack-there-of—of structure in a dataset). Multidimensional scaling embeddings of graphs in Euclidean space allow us to detect latent clusters or gradients.

However, much remains to be done to construct a complete framework for quantifying differences between graphs. For instance, we have not used subgraphs and

motif counts that could also be useful in quantifying such similarities as was suggested in Bonato et al. (2014). Another possible perspective which we have not covered here focuses on the use of graph kernels [Vishwanathan et al. (2010)] to define similarities between graphs.

Finally, we have only taken into account simple node identifiers, whereas incorporating more information about node covariates and edge lengths would enable higher resolution studies and enhanced change-point detection.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Tracking network dynamics: A survey using graph distances" (DOI: 10.1214/18-AOAS1176SUPP; .pdf). This supplement contains Appendices A–E as referred to in the text.

REFERENCES

AHN, Y.-Y., AHNERT, S. E., BAGROW, J. P. and BARABÁSI, A.-L. (2011). Flavor network and the principles of food pairing. *Sci. Rep.* **1** 196.

BANERJEE, A. (2008). The spectrum of the graph Laplacian as a tool for analyzing structure and evolution of networks. Ph.D. thesis, Univ. Leipzig.

BANERJEE, A. and JOST, J. (2008). Spectral plot properties: Towards a qualitative classification of networks. *Netw. Heterog. Media* **3** 395–411. MR2395239

BARBERÁN, A., BATES, S. T., CASAMAYOR, E. O. and FIERER, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6** 343–351.

BOLLOBÁS, B. (1998). *Modern Graph Theory*. *Graduate Texts in Mathematics* **184**. Springer, New York. MR1633290

BONATO, A., GLEICH, D. F., KIM, M., MITSCHE, D., PRALAT, P., TIAN, Y. and YOUNG, S. J. (2014). Dimensionality of social networks using motifs and eigenvalues. *PLoS ONE* **9** e106052.

CHAKERIAN, J. and HOLMES, S. (2012). Computational tools for evaluating phylogenetic and hierarchical clustering trees. *J. Comput. Graph. Statist.* **21** 581–599. MR2970909

CHAMPIN, P.-A. and SOLNON, C. (2003). Measuring the similarity of labeled graphs. In *International Conference on Case-Based Reasoning* 80–95. Springer, Berlin.

CHUNG, F. (2007). The heat kernel as the PageRank of a graph. *Proc. Natl. Acad. Sci. USA* **104** 19735–19740.

CVETKOVIĆ, D. (2012). Spectral recognition of graphs. *Yugosl. J. Oper. Res.* **22** 145–161. MR3007483

DETHLEFSEN, L. and RELMAN, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci. USA* **108** 4554–4561.

DIGIULIO, D. B., CALLAHAN, B. J., MCMURDIE, P. J., COSTELLO, E. K., LYELL, D. J., ROBACZEWSKA, A., SUN, C. L., GOLTSMAN, D. S., WONG, R. J., SHAW, G. et al. (2015). Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl. Acad. Sci. USA* **112** 11060–11065.

DONNAT, C. and HOLMES, S. (2018). Supplement to "Tracking network dynamics: A survey using graph distances." DOI:10.1214/18-AOAS1176SUPP.

DONNAT, C., ZITNIK, M., HALLAC, D. and LESKOVEC, J. (2017). Spectral graph wavelets for structural role similarity in networks. Preprint. Available at ArXiv:1710.10321.

FERRER, M., BARDAJÍ, I., VALVENY, E., KARATZAS, D. and BUNKE, H. (2013). Median graph computation by means of graph embedding into vector spaces. In *Graph Embedding for Pattern Analysis* 45–71. Springer, Berlin.

FUKUYAMA, J., MCMURDIE, P. J., DETHLEFSEN, L., RELMAN, D. A. and HOLMES, S. (2012). Comparisons of distance methods for combining covariates and abundances in microbiome studies. In *Biocomputing* 2012 213–224. World Scientific, Singapore.

GERBER, G. K. (2014). The dynamic microbiome. *FEBS Lett.* **588** 4131–4139.

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

GU, J., JOST, J., LIU, S. and STADLER, P. F. (2016). Spectral classes of regular, random, and empirical graphs. *Linear Algebra Appl.* **489** 30–49. MR3421836

HAMMOND, D. K., VANDERGHEYNST, P. and GRIBONVAL, R. (2011). Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmon. Anal.* **30** 129–150. MR2754772

HULLAR, M. A. and LAMPE, J. W. (2012). The gut microbiome and obesity. In *Obesity Treatment and Prevention*: *New Directions* **73** 67–79. Karger, Basel.

IPSEN, M. and MIKHAILOV, A. S. (2002). Evolutionary reconstruction of networks. *Phys. Rev. E* (3) **66** 6–9. DOI:10.1103/PhysRevE.66.046109.

JOST, J. and JOY, M. P. (2002). Evolving networks with distance preferences. *Phys. Rev. E* (3) **66** 036126.

JURMAN, G., VISINTAINER, R. and FURLANELLO, C. (2011). An introduction to spectral distances in networks. *Frontiers Artificial Intelligence Appl.* **226** 227–234. DOI:10.3233/978-1-60750-692-8-227.

JURMAN, G., VISINTAINER, R., RICCADONNA, S., FILOSI, M. and FURLANELLO, C. (2012). A glocal distance for network comparison. Preprint. Available at ArXiv:1201.2931.

JURMAN, G., VISINTAINER, R., FILOSI, M., RICCADONNA, S. and FURLANELLO, C. (2015). The HIM glocal metric and kernel for network comparison and classification. In *Proceedings of the* 2015 *IEEE International Conference on Data Science and Advanced Analytics*, *DSAA* 2015 **7** 46109. DOI:10.1109/DSAA.2015.7344816.

JURMAN, G., FILOSI, M., RICCADONNA, S., VISINTAINER, R. and FURLANELLO, C. (2016). Differential network analysis and graph classification: A glocal approach. 1–13.

KELLY, C., ZUO, X.-N., GOTIMER, K., COX, C. L., LYNCH, L., BROCK, D., IMPERATI, D., GARAVAN, H., ROTROSEN, J., CASTELLANOS, F. X. et al. (2011). Reduced interhemispheric resting state functional connectivity in cocaine addiction. *Biological Psychiatry* **69** 684–692.

KELMANS, A. K. (1976). Comparison of graphs by their number of spanning trees. *Discrete Math.* **16** 241–261. MR0463000

KELMANS, A. K. (1997). Transformations of a graph increasing its Laplacian polynomial and number of spanning trees. *European J. Combin.* **18** 35–48. MR1427603

KOUTRA, D., PARIKH, A., RAMDAS, A. and XIANG, J. (2011). Algorithms for graph similarity and subgraph matching. In *Proc. Ecol. Inference Conf.*

KOUTRA, D., SHAH, N., VOGELSTEIN, J. T., GALLAGHER, B. and FALOUTSOS, C. (2016). Delta-Con: Principled massive-graph similarity function with attribution. *ACM Transactions on Knowledge Discovery from Data* (*TKDD*) **10** 28.

LAYEGHIFARD, M., HWANG, D. M. and GUTTMAN, D. S. (2017). Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol.* **25** 217–228.

LEVANDOWSKY, M. and WINTER, D. (1971). Distance between sets. *Nature* **234** 34–35.

LUQMAN, M. M., RAMEL, J.-Y. and LLADÓS, J. (2013). Multilevel analysis of attributed graphs for explicit graph embedding in vector spaces. In *Graph Embedding for Pattern Analysis* 1–26. Springer, Berlin.

MONNIG, N. D. and MEYER, F. G. (2018). The resistance perturbation distance: A metric for the analysis of dynamic networks. *Discrete Appl. Math.* **236** 347–386. MR3739796

PAPADIMITRIOU, P., DASDAN, A. and GARCIA-MOLINA, H. (2010). Web graph similarity for anomaly detection. *Journal of Internet Services and Applications* **1** 19–30.

PROULX, S. R., PROMISLOW, D. E. and PHILLIPS, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution* **20** 345–353.

SHIMADA, Y., HIRATA, Y., IKEGUCHI, T. and AIHARA, K. (2016). Graph distance for complex networks. *Sci Rep* **6** 34944.

SHUMAN, D. I., RICAUD, B. and VANDERGHEYNST, P. (2016). Vertex-frequency analysis on graphs. *Appl. Comput. Harmon. Anal.* **40** 260–291. MR3440174

SHUMAN, D., NARANG, S., FROSSARD, P., ORTEGA, A. and VANDERGHEYNST, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **30** 83–98.

SPIELMAN, D. A. (2007). Spectral graph theory and its applications. In *Foundations of Computer Science*, 2007. *FOCS*'07. 48*th Annual IEEE Symposium on* 29–38. IEEE Press, New York.

TÉTREAULT, P., MANSOUR, A., VACHON-PRESSEAU, E., SCHNITZER, T. J., APKARIAN, A. V. and BALIKI, M. N. (2016). Brain connectivity predicts placebo response across chronic pain clinical trials. *PLoS Biology* **14** e1002570.

THÜNE, M. (2012). Eigenvalues of matrices and graphs. Ph.D. thesis, Univ. Leipzig.

TREMBLAY, N. and BORGNAT, P. (2014). Graph wavelets for multiscale community mining. *IEEE Trans. Signal Process.* **62** 5227–5239. MR3268107

TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. and GORDON, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444** 1027–131.

TURNBAUGH, P. J., BÄCKHED, F., FULTON, L. and GORDON, J. I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host & Microbe* **3** 213–223.

VISHWANATHAN, S. V. N., SCHRAUDOLPH, N. N., KONDOR, R. and BORGWARDT, K. M. (2010). Graph kernels. *J. Mach. Learn. Res.* **11** 1201–1242. MR2645450

WEISS, S., VAN TREUREN, W., LOZUPONE, C., FAUST, K., FRIEDMAN, J., DENG, Y., XIA, L. C., XU, Z. Z., URSELL, L., ALM, E. J. et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal* **10** 1669–1681.

ZAGER, L. A. and VERGHESE, G. C. (2008). Graph similarity scoring and matching. *Appl. Math. Lett.* **21** 86–94. MR2435232

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
390 SERRA MALL
STANFORD, CALIFORNIA 94305
USA
E-MAIL: cdonnat@stanford.edu
          susan@stat.stanford.edu