

STATISTICS OF DYNAMIC RANDOM NETWORKS: A DEPTH FUNCTION APPROACH.

DANIEL FRAIMAN^{1,2}, NICOLAS FRAIMAN³, AND RICARDO FRAIMAN⁴

ABSTRACT. The study of random graphs and networks had an explosive development in the last couple of decades. Meanwhile, there are just a few references about statistical analysis on graphs. In this paper we focus on graphs with a fixed number of labeled nodes (such as those used to model brain networks) and study some statistical problems in a nonparametric framework. We introduce natural notions of center and a depth function for graphs that evolve in time. This allows us to develop several statistical techniques including testing, supervised and unsupervised classification, and a notion of principal component sets in the space of graphs. Some examples and asymptotic results are given.

1. INTRODUCTION

The literature of random graphs and networks has grown exponentially during the last fifteen years. A huge number of different research lines have been developed in order to study the behavior of several stochastic models and real data networks. Some important results among those lines include the existence of stationary measures in dynamic models (or static but growing in size), characterizations of thresholds for giant components and connectivity, analysis of the spread of epidemics over fixed networks, and the development of new topological measures to characterize network structure (modules, motifs, etc.). In contrast, the study of the statistical properties of such models, is much less developed. Our contribution here is among this line, we discuss how some statistical methods can be adapted for analyzing a random sample of networks or the stochastic dynamics of a unique network.

The theory of random graphs is dominated by models where the label of each node is not relevant for the type of properties studied. Nevertheless, in the majority of real networks such like those modeling brain connections, financial markets, the internet, or protein interactions, the label of each node appears naturally and it is relevant. That is the reason we consider important to develop some statistical methods for the space of graphs, in which each node is distinguishable from other nodes. We study some statistical problems in the space of graphs of fixed size. Throughout the manuscript we refer to them as networks or graphs indistinctly.

We introduce a concept of depth, based on a natural distance in the space of graphs. This definition has the nice property that the probability measure of a random graph is determined by its depth function. We develop some statistical analysis tools based on the depth function introduced, and show that many standard problems in multivariate analysis can be easily adapted to the graph framework. Some of them are solved directly in the space of graphs while for others the depth

Key words and phrases. Depth — Graph estimation — Cluster analysis of graphs.

function is used, exploiting the fact that it determines the measure. Specifically, we address the following questions:

- (a) Given a random graph \mathbf{G} .
 1. How to define measures of centrality and variability?
 2. How to define a depth function in the space of graphs?
- (b) Given a sample of graphs G_1, G_2, \dots, G_ℓ .
 1. How to calculate their empirical measures of centrality and variability?
 2. Supervised and unsupervised classification.
 3. How to define a notion of principal components?
 4. How to test simple hypothesis?

We believe that the present approach, besides being simple, can be extended to other important statistical problems. We end by presenting some open problems that could be treated with similar ideas.

2. PROBABILITY AND GRAPH THEORY FRAMEWORK

A graph (or network), denoted by $G = (V, E)$, is an object described by a set V of nodes (vertices) and a set $E \subset V \times V$ of links (edges) between them. In what follows, we consider families of graphs defined over the same fixed finite set of n nodes. A graph is completely described by its adjacency matrix $A \in \{0, 1\}^{n \times n}$, where $A(i, j) = 1$ if and only if the link $(i, j) \in E$. If the matrix A is symmetric then the graph is undirected, otherwise we have a directed graph. In principle, we consider graphs with no loops, which have zero value at the diagonal. A path is a sequence of nodes such that each consecutive pair is a link. A graph is connected if there is a path between any pair of nodes.

2.1. Metric on the space of graphs. Given two graphs G_1, G_2 we consider the edit distance given by the minimum number of links we have to add and subtract in order to transform G_1 into G_2 . More precisely, if T_{ij} is the inversion operator of the link (i, j) , which interchanges 1 with 0 on the (i, j) entry of the adjacency matrix of the graph, and A_1, A_2 are the adjacency matrices of G_1 and G_2 respectively; the distance is defined as

$$(1) \quad d(G_1, G_2) = \min\{k : T_{i_1 j_1} T_{i_2 j_2} \dots T_{i_k j_k} A_1 = A_2\}.$$

Remark 1. Observe that the distance defined in (1) is nothing but the L^1 distance between the corresponding adjacency matrices A_1 and A_2 .

In what follows, the space of graphs with n nodes endowed with the metric given by eq. (1) is denoted by \mathcal{G} , while the total number of possible links for a graph in \mathcal{G} is denoted by $m = n(n-1)/2$.

2.2. Random graphs. We study dynamic random graphs evolving in (discrete) time. We reserve boldface typeface for random elements, for instance, \mathbf{G}_t stands for a random graph at time t . Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random graph with distribution μ is a function $\mathbf{G} : \Omega \rightarrow \mathcal{G}$ such that

$$\mathbb{P}(\mathbf{G} \in \mathcal{A}) = \int_{\mathcal{A}} d\mu(G) = \sum_{G \in \mathcal{G} \cap \mathcal{A}} p_G,$$

where $p_G = \mu(\mathbf{G} = G)$.

Using this definition it is easy to compute probabilities in some cases. For example, the probability that there exists a link between nodes i and j is

$$\mathbb{P}(\mathbf{A}(i, j) = 1) = \sum_{G \in \mathcal{G}_{ij}} p_G$$

where \mathcal{G}_{ij} is the set that contains all the graphs with a link between node i and j . In the same way we can calculate the probability of a connected graph,

$$\mathbb{P}(\mathbf{G} \text{ is connected}) = \sum_{G \in \mathcal{G}^\dagger} p_G.$$

where \mathcal{G}^\dagger is the set of connected graphs that belongs to \mathcal{G} .

2.3. Centers, and scale measure. In this framework the expected distance from a graph H to a random graph \mathbf{G} can be computed as

$$\mathbb{E}(d(\mathbf{G}, H)) = \int_{\mathcal{G}} d(G, H) d\mu(G) = \sum_{G \in \mathcal{G}} d(G, H) p_G.$$

Definition 1. *The central (or median) subset of a random graph \mathbf{G} is the subset of networks fulfilling*

$$(2) \quad \mathcal{C}(\mathbf{G}) := \arg \min_{H \in \mathcal{G}} \mathbb{E}(d(\mathbf{G}, H)),$$

i.e., the Frechet center with respect to the metric d .

The notion of median subset corresponds to minimizing the expected L^1 -distance. A notion of mean subset can be defined by minimizing the expected L^2 -distance. In the Euclidean setup they corresponds to the L^1 -median and to the usual expected value respectively.

Given a sample G_1, \dots, G_ℓ of random networks in \mathcal{G} , applying definition (2) to the empirical distribution, the notion of empirical center is obtained. More precisely,

Definition 2. *The empirical central subset $\hat{\mathcal{C}}_\ell$ is defined as the subset of networks fulfilling*

$$(3) \quad \hat{\mathcal{C}}_\ell = \arg \min_{H \in \mathcal{G}} \frac{1}{\ell} \sum_{i=1}^{\ell} d(G_i, H).$$

In general, the subsets $\mathcal{C}(\mathbf{G})$ and $\hat{\mathcal{C}}_\ell$ contain only one graph, i.e., there exists a unique graph that minimizes equations (2) and (3) respectively. In this case, we call the unique central graph the *skeleton graph* and we denote it by $S(\mathbf{G})$ (or \hat{S}_ℓ in the empirical case). The following proposition gives necessary and sufficient conditions for uniqueness of the central graph, together with a complete characterization of the skeleton graph and the subsets $\mathcal{C}(\mathbf{G}), \hat{\mathcal{C}}_\ell$ when we have more than one solution. In the last case, there is a graph in $\mathcal{C}(\mathbf{G})$ (respectively in $\hat{\mathcal{C}}_\ell$) with the minimum number of links and another one with the maximum number of links. These are called the minimal and maximal centers respectively.

Proposition 1 (Characterization of the central set.).

- a) $\mathcal{C}(\mathbf{G})$ has a unique graph iff $\mathbb{P}(\mathbf{A}(i, j) = 1) \neq 1/2 \forall i, j$. The adjacency matrix of $S(\mathbf{G})$ satisfies $A_S(i, j) = 1$ iff

$$\mathbb{P}(\mathbf{A}(i, j) = 1) > 1/2.$$

- b) $\hat{\mathcal{C}}_\ell$ has a unique graph iff $(1/\ell) \sum_{k=1}^\ell A_k(i, j) \neq 1/2 \forall i, j$. The adjacency matrix of \hat{S}_ℓ satisfies $A_S(i, j) = 1$ iff

$$\frac{1}{\ell} \sum_{k=1}^\ell A_k(i, j) > 1/2.$$

- c) If for some pair (i, j) we have $\mathbb{P}(\mathbf{A}(i, j) = 1) = 1/2$, let the minimal center $S(\mathbf{G})$ be the graph whose adjacency matrix fulfills $A_S(i, j) = 1$ iff $\mathbb{P}(\mathbf{A}(i, j) = 1) > 1/2$ and the maximal center $L(\mathbf{G})$ be the graph whose adjacency matrix fulfills $A_L(i, j) = 1$ iff $\mathbb{P}(\mathbf{A}(i, j) = 1) \geq 1/2$. Then, the set $\mathcal{C}(\mathbf{G})$ contains exactly all subgraphs of $L(\mathbf{G})$ for which $S(\mathbf{G})$ is a subgraph. The same is true for the empirical version *mutatis mutandis*.

Since the space of graphs \mathcal{G} is finite, the following law of large numbers follows immediately.

Theorem 1. *Let (\mathcal{G}, d) be the space of graphs endowed with the metric d . Let \mathbf{G} be a random graph with law μ such that the central set $\mathcal{C}(\mathbf{G})$ has only one element $S(\mathbf{G})$. Let $\{\mathbf{G}_t, t \geq 1\}$ be an stationary and ergodic sequence of random networks with law μ . If \hat{S}_ℓ is any element of the empirical central set $\hat{\mathcal{C}}_\ell$, then almost surely*

$$\lim_{\ell \rightarrow \infty} d(S(\mathbf{G}), \hat{S}_\ell) = 0.$$

In other words, the set of empiric central graphs coincides with the singleton central element if ℓ is large enough.

Given a random graph, besides the center, it is desired to have a measure of the “homogeneity” (variability) of its distribution. The most natural notion of dispersion associated with our problem is given by

Definition 3. *The scale of the random graph G is defined as*

$$\sigma(\mathbf{G}) := \mathbb{E}(d(\mathbf{G}, S^*)).$$

where S^ is any of graph contained in $\mathcal{C}(\mathbf{G})$.*

The corresponding empirical scale measure $\hat{\sigma}_\ell$ based on the sample G_1, \dots, G_ℓ , is given by just replacing the expected distance by $\frac{1}{\ell} \sum_{i=1}^\ell d(G_i, S^*)$ in Definition 3. Using the inequality

$$\frac{1}{\ell} \sum_{i=1}^\ell |d(G_i, \hat{S}_\ell) - d(G_i, S(\mathbf{G}))| \leq \max_{H \in \mathcal{G}} |d(H, \hat{S}_\ell) - d(H, S(\mathbf{G}))|,$$

we can derive the strong consistency of $\hat{\sigma}$ to $\sigma(\mathbf{G})$ from Theorem 1.

We finish this section presenting three examples to illustrate the proposed framework.

Example 2. In the Erdős–Rényi model each link is present with fixed probability p . If $p < 1/2$ the center is the empty graph G_\emptyset (the graph with no links), if $p > 1/2$ it is the complete graph K , and if $p = 1/2$ we get the entire space of graphs. The scale is given by $\sigma(\mathbf{G}) = (1/2 - |p - 1/2|)m$. This is intuitive given that the maximum scale is obtained at $p = 1/2$ and it is null at values of p equal 0 or 1.

Remark 2. A different empirical notion of center has been introduced in [1]. It restricts the search of the minimizer to the graphs in the sample, i.e., the center \hat{M}_ℓ is given by $\arg \min_j (1/\ell) \sum_{i=1}^\ell d(G_i, G_j)$. The population version corresponds to minimizing the expected distance over the support of the underlying distribution μ of \mathbf{G} , that is the center is defined as $\arg \min_{H \in \text{supp}(\mu)} \mathbb{E}(d(\mathbf{G}, H))$. If the support of μ is the whole space of graphs \mathcal{G} both notions coincide. Otherwise, in general this is not the case. However, like in the case of high dimensional data, maximizing just over the sample, is not a good strategy. Indeed, for example, it is easy to verify (using Hoeffding's inequality) that for the very simple Erdős–Rényi model with parameter $p < 1/2$,

$$\begin{aligned} \mathbb{P}(\hat{S}_\ell \neq G_\emptyset) &\leq 1 - \left(1 - e^{-2\ell(p-1/2)^2}\right)^m, \\ \mathbb{P}(\hat{M}_\ell \neq G_\emptyset) &\geq \prod_{i=1}^\ell \mathbb{P}(\mathbf{G}_i \neq G_\emptyset) = (1 - (1-p)^m)^\ell. \end{aligned}$$

Thus, \hat{S}_ℓ converges at a much better rate.

Example 3. An important distribution that arises in the space of graphs \mathcal{G} is the double exponential type distribution given by

$$(4) \quad \mu(\mathbf{G} = G) = ce^{-\lambda d(G, S_0)},$$

where the normalizing constant $c = e^{\lambda m}(1 + e^\lambda)^{-m}$, $\lambda > 0$ is a parameter, and S_0 is a particular graph. As in the double exponential distribution, this law is symmetric, it has an explicit symmetry center and mode (S_0), and has an exponential decay. It is a particular case of the so called Exponential Random Graph Model, and presents a unique central graph (Eq. (2)). It is easy to show that it verifies $S(\mathbf{G}) = S_0$, and $\sigma(\mathbf{G}) = m/(1 + e^\lambda)$.

Note also that the empirical center given in Eq. (3) can be seen as a *maximum likelihood* estimate of the center of the previous distribution. Indeed, if G_1, \dots, G_ℓ are i.i.d. random graphs with this μ distribution, the empirical center coincides with the maximum likelihood estimate of S_0 .

Example 4. Electroencefalographic (EEG) data correspond to brain electrical activity measured over time at n (number of electrodes) regions along the scalp. From this data, brain functional networks can be constructed (see for instance [11, 12]) just computing a measure of dependency between the time series and an appropriate threshold. The network finally is constructed including a link between a pair of brain nodes (electrodes) when the dependency measure is greater than the threshold. This functional network evolves in time. A toy model for modeling such brain functional networks is the following. Consider a 5×5 lattice where each lattice point (identified as a node) is enumerated from 1 to 25 (see Fig. 1). Let $\mathbf{x}_t \in \{1, 2, \dots, 25\}$ be an integer (node) random variable that evolves in a stochastic way. We define the probability that node i is linked with node j at time t given \mathbf{x}_t as,

$$(5) \quad \mathbb{P}(\mathbf{A}_t(i, j) = 1 \mid \mathbf{x}_t) = \frac{a}{(\text{dist}(i, \mathbf{x}_t) + \text{dist}(j, \mathbf{x}_t)) \text{dist}(i, j)}$$

where $\text{dist}(\cdot, \cdot)$ is the Euclidean distance, and $0 < a \leq 1$. We study this graph model when \mathbf{x}_t performs a nearest neighbor random walk in a given set of nodes. The nodes belonging to this set are represented in white color in all panels of Fig. 1.

For this model we calculate the skeleton graph under four different random walk scenarios for \mathbf{x}_t . The stationary central graph stabilizes once the graph has reached stationarity (the initial condition is lost). Given that \mathbf{x}_t evolves as an irreducible and aperiodic Markov chain, we can compute its stationary law, $\pi(k) = \lim_{t \rightarrow \infty} \mathbb{P}(\mathbf{x}_t = k)$ as $t \rightarrow \infty$. The stationary probability that nodes i and j are linked is

$$\lim_{t \rightarrow \infty} \mathbb{P}(\mathbf{A}_t(i, j) = 1) = \frac{a\pi(k)}{(\text{dist}(i, k) + \text{dist}(j, k)) \text{dist}(i, j)}.$$

From Prop. 1 and the previous equation the central graph is computed, and presented in Fig. 1. Panel (A) shows the central graph when \mathbf{x}_t moves only between nodes 12, 13 and 14, and Panel (B) when the nodes considered are 11, 12, 13, 14, and 15. In Panel (C), \mathbf{x}_t moves in the middle square, and in Panel (D) there is no restriction for \mathbf{x}_t . In the latter case we obtain a null central graph.

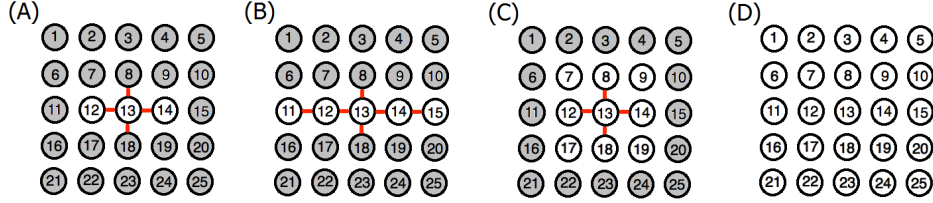


FIGURE 1. Inhomogeneous lattice graph model defined in Example 3. Scheme of the model for 4 different random walks scenarios for \mathbf{x}_t . From left to right panel each case corresponds to \mathbf{x}_t moving in the sets $\{12, 13, 14\}$, $\{11, 12, 13, 14, 15\}$, $\{7, 8, 9, 12, 13, 14, 17, 18, 19\}$, and $\{1, 2, \dots, 25\}$ respectively. Links (in red) corresponds to the skeleton graph $S(\mathbf{G})$.

The linking probability given by equation (5) favors links that are near to node \mathbf{x}_t . For the four cases studied here there exists a unique skeleton graph, $S(\mathbf{G})$, when the parameter $a = 1$. This central graph is represented with red links in each panel of Fig. 1. Note that in the case when \mathbf{x}_t evolves as a “complete” two dimensional random walk (right panel of Fig. 1), the skeleton graph is the null graph (which has no links). This graph model can be considered as an example of a Bickel model [13].

3. DEPTH FUNCTION

In this section we first introduce a notion of depth in the space of graphs. A depth function is a function that orders the space in terms of center-outward position. This idea has been introduced in the robust statistics literature. The most well known depth notions for the Euclidean space are the half-space depth [2], simplicial depth [3] the L^1 -depth [4, 5] and Mahalanobis depth. Several important applications to different statistical problems have been developed in the last years.

Given a fixed graph H and a sample of random graphs G_1, \dots, G_ℓ with the same distribution we consider the L^1 -depth notion with respect to the metric d , which in particular defines the central graph (also called spatial median) in our setup. *The central subset corresponds with the maximizing set of this depth.* More precisely,

Definition 4. We define the empirical depth at the graph $H \in \mathcal{G}$, as

$$\hat{D}(H) = m - \frac{1}{\ell} \sum_{t=1}^{\ell} d(G_t, H),$$

which corresponds to the population depth given by

$$D(H) = m - \mathbb{E}(d(\mathbf{G}, H)).$$

Observe that both the empirical and population depth are non-negative, and fulfill the main properties of a depth function given in [6]. Moreover, we have a simple explicit solution for the median center maximizing $D(H)$ given by Proposition 1. On the contrary in the Euclidean space, an algorithm is required to maximize the L^1 -depth. In fact, a fast monotonically convergent algorithm to calculate the L^1 -median of a data set in \mathbb{R}^d has been proposed in [7].

Figure 2 shows the empirical depth for a simulation study of the model presented in Example 3. One thousand random graphs were generated with the model and the empirical depth for eight particular graphs is shown in the panels. The empirical depth is shown below each graph, smaller values are observed for graphs that differ a lot from the maximum depth (central) graph shown in the right panel of first row.

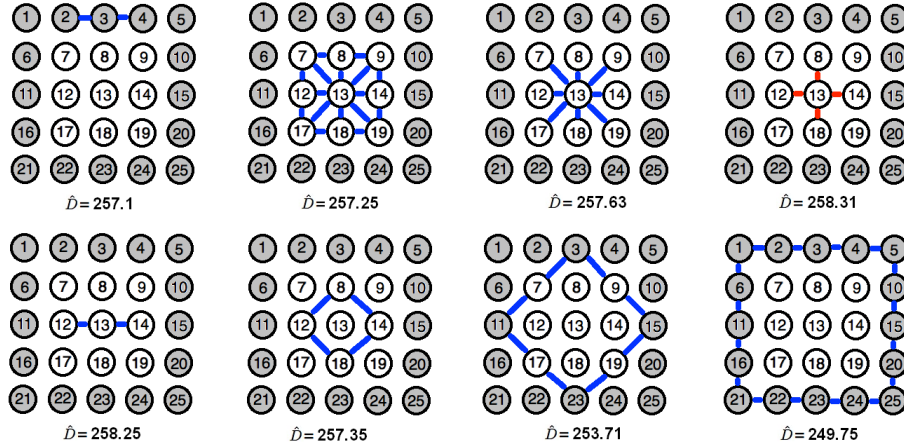


FIGURE 2. One thousand graphs were generated with the model presented in Example 3 with parameter $a = 1$ and in the case were x_t evolves as two dimensional random walk in the square set of nodes $\{7, 8, 9, 12, 13, 14, 17, 18, 19\}$. The empirical depth for eight selected graphs is shown below each graph. The empirical central graph \hat{S}_{1000} is shown in the right graph of first row.

An important property of the above definition of depth is that it determines the graph probability measure. This result, that follows from the invertibility of distance matrices given in [8], has an important impact in statistics and in particular in our setup since it allows to develop statistical methods based on the depth D and obtain results for the space of graphs \mathcal{G} . In general, depths do not determine measure, the only known result is for Tukey's half space depth when the measure is discrete and can be found in [10]. More precisely we have that,

Proposition 2. *Let \mathbf{G} be a random graph with distribution μ . Write $D_\mu(H) = m - \mathbb{E}_\mu(d(\mathbf{G}, H))$ to explicitly note the dependency on the distribution. Then,*

$$D_\mu(H) = D_\nu(H) \quad \forall H \in \mathcal{G} \quad \text{iff} \quad \mu = \nu.$$

4. UNSUPERVISED AND SUPERVISED CLASSIFICATION

Among others multivariate techniques, we now discuss the classification problem for graphs and its relationship with the depth function. We first study unsupervised classification, also called Cluster analysis.

4.1. Unsupervised classification. Let us suppose that the probability measure μ is such that k groups or clusters of graphs can be identified. Two graphs in the same group are close together (similar), while graphs that belong to different groups are far apart. We want to identify each of the k groups. The most well known clustering methods are k -means or k -medioids, which are only based on the distances between the random elements. More precisely, the algorithm looks for the centers of the groups and then assign each data to its nearest center.

In our setting, what is important is to identify in a good way each of the k center graphs, that we denote by $S_1^*, S_2^*, \dots, S_k^*$. The strategy proposed here is the same to that of k -means (k -medioids in our case). We look the k graphs that maximize the depth of order k defined as:

$$D_k(H_1, \dots, H_k) = m - \mathbb{E} \left(\min_{i=1, \dots, k} d(H_i, \mathbf{G}) \right),$$

i.e., we look for subsets $\{S_1^*, \dots, S_k^*\}$ that satisfy

$$D_k(S_1^*, \dots, S_k^*) = \max_{H_1, \dots, H_k} D_k(H_1, \dots, H_k).$$

Then, each graph is assigned to its nearest center and we obtain a partition of the space. The asymptotic results for k -means and k -medioids given in [14] are valid for compact metric spaces, which covers our setup. In the empirical case we look for the empirical center graphs $\hat{S}_1^*, \hat{S}_2^*, \dots, \hat{S}_k^*$ that maximize the empirical depth of order k ,

$$\hat{D}_k(H_1, \dots, H_k) = m - \frac{1}{\ell} \sum_{j=1}^{\ell} \min_{i=1, \dots, k} d(H_i, G_j).$$

Example 5. A graph probability distribution where clusters of graphs can be identified is a mixture of k double exponential distributions introduced in Eq. (4), i.e.,

$$(6) \quad \mu(\mathbf{G} = G) = \sum_{i=1}^k p_i c_i e^{-\lambda_i d(G, S_i)}.$$

As previously, $c_i = e^{\lambda_i m} (1 + e^{\lambda_i})^{-m}$, $\lambda_i > 0$, S_i are graphs, and $p_i > 0$ with $p_1 + p_2 + \dots + p_k = 1$. In order to ensure the existence of k clusters, k peaks must be present in the law μ . One simple way to ensure this condition is

$$\min\{d(S_i, S_j) : i \neq j \in \{1 \dots k\}\} \gg 1 / \min\{\lambda_1, \dots, \lambda_k\}.$$

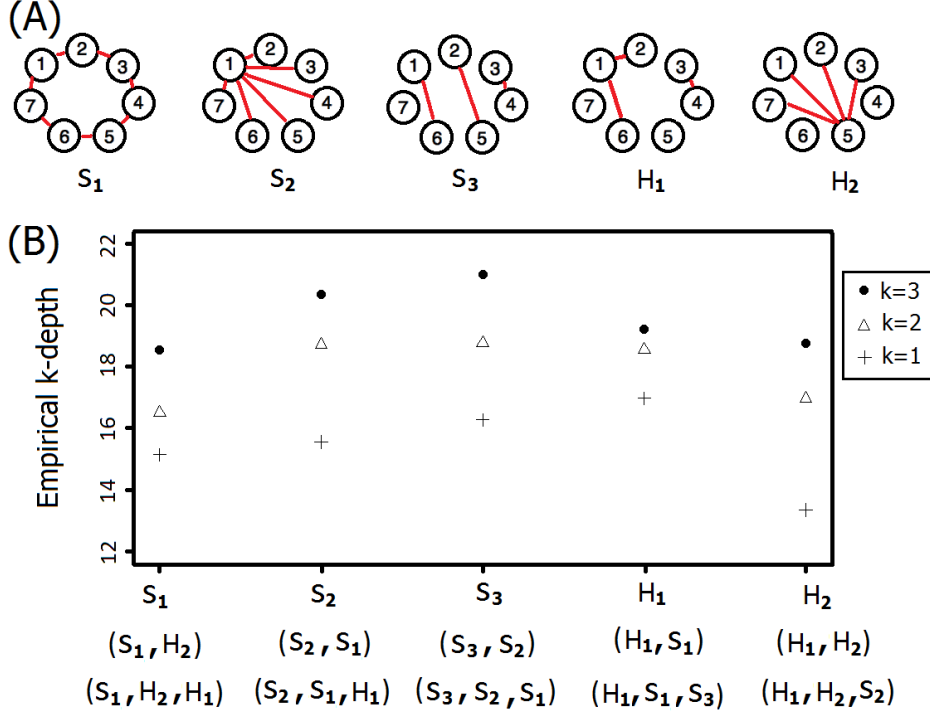


FIGURE 3. (A) S_1, S_2 and S_3 corresponds to the three mixture centers, while H_1 is the central graph (Eq. (2)) of the mixture. (B) Empirical depth of order k for $k = 1, 2, 3$ at the five graphs ($k=1$, crosses), pair of graphs ($k=2$, triangles) and triplets of graphs ($k=3$, circles). The corresponding graphs, pairs and triplets are display at the bottom.

Fig. 3 shows the results from a simulation of graphs of 7 nodes. One hundred graphs were generated with the law of Eq. (6) taking $k = 3$, $\lambda_1 = \lambda_2 = \lambda_3 = 10$, and $p_1 = p_2 = p_3 = 1/3$. The three centers, S_1, S_2, S_3 together with the center graph H_1 and an arbitrary fifth graph H_2 are shown in Fig. 3 (A). They verify that $d(S_1, S_2) = 9, d(S_1, S_3) = 8, d(S_2, S_3) = 7$, much larger than one. Fig. 3 (B) shows the empirical depth of order $k = \{1, 2, 3\}$ for 5 different graphs (crosses, $k = 1$), pair of graphs, (triangles $k = 2$) and triplets of graphs (circles $k = 3$). The maximum empirical depth occurs for $k = 3$ at $(S_1^*, S_2^*, S_3^*) = (S_1, S_2, S_3)$ as it is expected.

4.2. Supervised classification. In this case we have a training sample given by $(\mathbf{Y}_1, \mathbf{G}_1), \dots, (\mathbf{Y}_\ell, \mathbf{G}_\ell)$ where $\{\mathbf{G}_t : t \geq 1\}$ is a sequence of random graphs and $\mathbf{Y}_t : t \geq 1$ stands for the labels that indicates to which subpopulation (group) the individual belongs. For binary classification $\mathbf{Y}_t \in \{0, 1\}$ indicating sick or healthy for instance. The problem consist on predicting the label of a new observation only based on $\mathbf{G}_{\ell+1}$ and the training sample.

The most simple and well known nonparametric classification is k -nearest neighbors. The method just looks for the k nearest neighbors of $\mathbf{G}_{\ell+1}$ among the training sample $\mathbf{G}_t : 1 \leq t \leq \ell$ and assigns the label by majority vote within the labels of the k -nearest neighbors. Again it is a method just based on distances and can be applied in our setup. The method is asymptotically optimal as long as $k = k(\ell) \rightarrow \infty$ and $k/\ell \rightarrow 0$ as $\ell \rightarrow \infty$ (see for instance [15]).

5. PRINCIPAL COMPONENTS

Principal components is an important statistical tool when analyzing data, particularly for high dimensional and functional data. The objective of this technique is to reduce the dimension p of the data using linear combinations of the variables. This is done by projecting the data onto the $k \ll p$ dimensional subspace which minimize the distance to the original random vector. Equivalently, the principal components can be defined iteratively. The first is the direction on which the projection of the random element has maximal variance. The next one, maximizes the variance of the projection on the orthogonal subspace to the first one and so on. The absence of projections in metric spaces makes the extension non trivial. In what follows we introduce a method in such direction for random elements in the space \mathcal{G} .

Let G_\emptyset be the empty graph and write $|H| = d(H, G_\emptyset)$ for the number of links in H . Given $G, H \in \mathcal{G}$ define the intersection graph $G \wedge H$ as the graph with only the common links to both. Note that $|G \wedge H|$ is nothing but the inner product between the adjacency matrices of G and H .

Given a random graph \mathbf{G} , we define the first *principal component* as the set of graphs \mathcal{Q}_1 that maximize the variance of the following “projection”

$$(7) \quad \mathcal{Q}_1 = \arg \max_{Q \in \mathcal{G}} \text{Var} \left(\frac{|\mathbf{G} \wedge Q|}{|Q|} \right).$$

If $\mathcal{Q}_1 = \{Q_1, \dots, Q_p\}$ is the set of solutions then the *principal component space* \mathcal{S}_1 generated by \mathcal{Q}_1 is defined as the set of all the geodesic curves in the space \mathcal{G} joining each Q_j with the complete graph, denoted by K . Recall that, given a, b, x arbitrary points in a metric space, we say that x belongs to a geodesic from a to b if $d(a, b) = d(a, x) + d(x, b)$. So, if $d(G, H) = k$ then there are $k!$ geodesics in \mathcal{G} . To define the *second principal component* \mathcal{Q}_2 we consider the same problem, but now we maximize the variance within the “orthogonal” subset $\mathcal{G} \setminus \mathcal{S}_1$.

Observe that $H \in \mathcal{G} \setminus \mathcal{S}_1$ iff H has no links in common with any element in \mathcal{Q}_1 , i.e., it has no link in common with the graph $\tilde{Q}_1 := Q_1 \vee \dots \vee Q_p$, which contains all links present in at least one Q_j , $j = 1, \dots, p$. In this sense we refer to $\mathcal{G} \setminus \mathcal{S}_1$ as the orthogonal subset. This particular graph, \tilde{Q}_1 , can be considered as the most informative graph to visualize \mathcal{S}_1 . Note that the $\mathcal{G} \setminus \mathcal{S}_1$ has cardinality $2^{m-|\tilde{Q}_1|}$. The next principal components are defined analogously.

To define the corresponding empirical version of Eq. (7) let

$$\Delta_\ell(Q) = \frac{1}{\ell} \sum_{k=1}^{\ell} \frac{(|G_k \wedge Q| - \Lambda_\ell(Q))^2}{|Q|^2},$$

where $\Lambda_\ell(Q) = (1/\ell) \sum_{k=1}^{\ell} |G_k \wedge Q|$. The empirical principal component is given by

$$\hat{\mathcal{Q}}_1 = \{Q \in \mathcal{G} : \Delta_\ell^{\max} - \Delta_\ell(Q) < \varepsilon_\ell\},$$

where $\Delta_\ell^{\max} = \max_{Q \in \mathcal{G}} \Delta_\ell(Q)$ and $\varepsilon_\ell \rightarrow 0$.

Example 6. Distributions with spherical symmetry satisfy that there does not exist a principal direction, since all directions are equally informative. The distribution given in Eq. (4) is one such example. However, a mixture of two distributions of that form breaks the symmetry. Moreover, the graph distribution

$$\mu_1(\mathbf{G} = H) = p_1 c_1 e^{-\lambda_1 d(H, S_1)} + (1 - p_1) c_2 e^{-\lambda_2 d(H, S_2)},$$

with $S_1 \neq S_2$, and $0 < p_1 < 1$ has a “direction” of maximum variance and it is the one that “connects” S_1 with S_2 . In this example, we take $p_1 = p_2 = 0.5$, and $\lambda_1 = \lambda_2 = 10$. The centers are given in Fig. 4, Panel (A). We calculate the first two principal components. The first one consist on three graphs with only one link, namely $a = 1 \leftrightarrow 2$, $b = 1 \leftrightarrow 3$ and $c = 1 \leftrightarrow 4$. Panel (B) shows the variance of the random sample of graphs once it has been projected by a graph in \mathcal{G} , i.e., $\Delta_\ell(Q)$ as a function of Q is shown. Graphs belonging to the first principal space \mathcal{S}_1 are shown as red points, while all other graphs (green points) belong to the second principal space \mathcal{S}_2 .

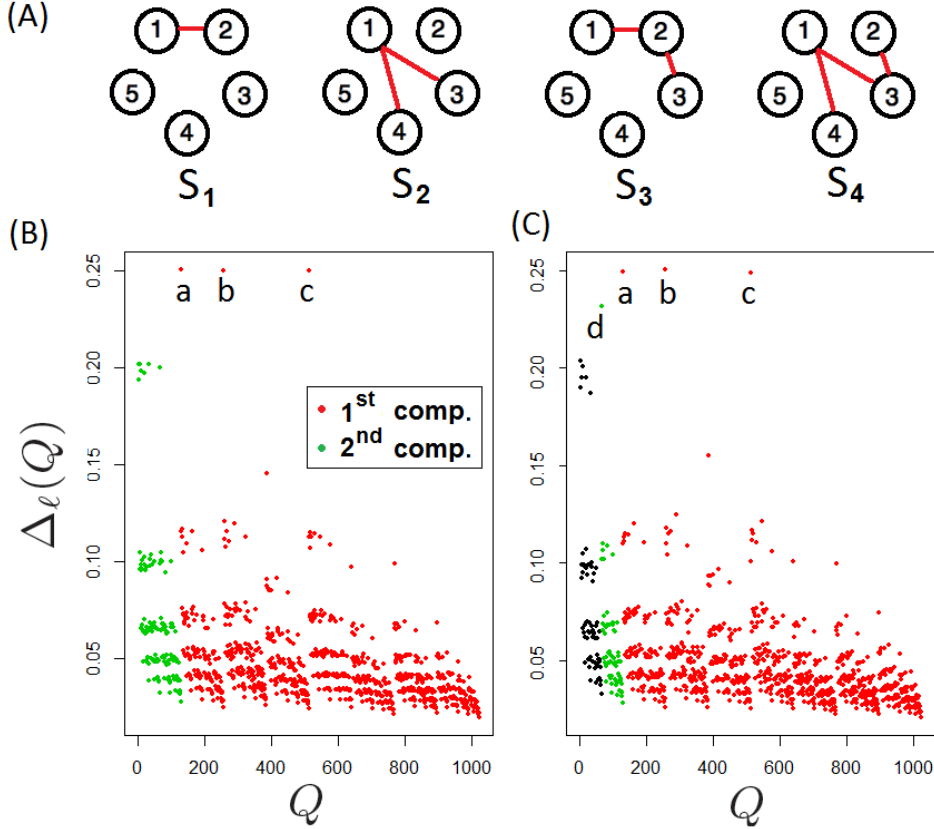


FIGURE 4. (A) Graphs used in the definition of μ_1 and μ_2 . $\Delta_\ell(Q)$ as a function of Q for a random sample of 1000 graphs with law (B) μ_1 , and (C) μ_2 . The letters a, b , and c indicates the graphs in $\hat{\mathcal{Q}}_1$, d correspond to the graph in $\hat{\mathcal{Q}}_2$. Graphs are ordered for good visualization.

Example 7. Here we consider a mixture of 4 exponentials

$$\mu_2(\mathbf{G} = H) = \sum_{i=1}^4 p_i c_i e^{-\lambda_1 d(H, S_i)},$$

where $(p_1, p_2, p_3, p_4) = (0.4, 0.4, 0.1, 0.1)$, $\lambda_1 = 10$, and the four centers are shown in Fig. 4, Panel (A). The first principal component consist on the same graphs a, b, c of Example 5, and the second principal component is the one link graph $d = 2 \leftrightarrow 4$. The projected variance on each of this graph and the rest of graphs can be observed on Panel (C). Red points corresponds to the graphs in \mathcal{S}_1 , green points to those in \mathcal{S}_2 , while the rest to the next principal components.

6. LIMIT THEOREMS FOR THE DEPTH FUNCTION

The empirical depth function $\hat{D}(H)$ converges almost surely to the population version $D(H)$ uniformly. Indeed, as a consequence of the ergodic theorem we have the following theorem.

Theorem 8 (Uniform convergence of the depth function.). *Given a stationary ergodic sequence of random graphs $\{\mathbf{G}_t : t \geq 1\}$ with common law μ . Then, almost surely*

$$\max_{H \in \mathcal{G}} |\hat{D}(H) - D_\mu(H)| \rightarrow 0.$$

Moreover, the asymptotic distribution of the depth process is Gaussian. We have the following result.

Theorem 9 (Asymptotic normality of the depth process.). *Given a strictly stationary α -mixing sequence of graphs $\{\mathbf{G}_t : t \geq 1\}$ in \mathcal{G} with common distribution μ fulfilling $\sum_{n=1}^{\infty} \alpha(n) < \infty$, and \mathbf{G} a random graph with the same distribution. Fix an ordering $(G_j)_{j=1, \dots, 2^m}$ of the elements of the space \mathcal{G} . Pick $\beta \in \mathbb{R}^{2^m}$, $\|\beta\| = 1$, and define*

$$\begin{aligned} \mathbf{Z}_\ell &= \left(\hat{D}(G_j) - D_\mu(G_j) \right)_{j=1, \dots, 2^m} \\ \mathbf{Y}_k &= \left(d(\mathbf{G}_k, G_j) - \mathbb{E}(d(\mathbf{G}_k, G_j)) \right)_{j=1, \dots, 2^m}. \end{aligned}$$

a) *If in addition*

$$\sum_{k=1}^{\infty} \beta^T \mathbb{E}(\mathbf{Y}_1^T \mathbf{Y}_k) \beta > 0, \quad \forall \beta \text{ with } \|\beta\| = 1.$$

Then, $\sqrt{\ell} \beta^T \mathbf{Z}_\ell$ converges weakly as $\ell \rightarrow \infty$ to a normal distribution with mean zero and with the same variance as $\beta^T \mathbf{Y}_1$.

b) *As a consequence, the asymptotic law of*

$$KS := \max_{H \in \mathcal{G}} \sqrt{\ell} |\hat{D}(H) - D_\mu(H)|,$$

is derived from a) and the Continuous Mapping Theorem.

c) *If the common distribution of the sequence of graphs is $\nu \neq \mu$, then $KS \rightarrow \infty$ as $\ell \rightarrow \infty$.*

Remark 3 (Hypothesis Testing). Using items b) and c) of Theorem 9 we can derive universally consistent tests for a given distribution on the space of graphs. A similar result can be obtained for the two sample problem. This last problem has also been addressed in [16] recently.

7. DISCUSSION

Herein we have studied dynamic random graphs with a fix number of nodes in a general framework. Some classical statistical problems such as clustering and principal component analysis were addressed. All the statistics defined here have been constructed using a natural distance between graphs and its corresponding L^1 -depth notion. From a theoretical point of view, we believe that the framework presented here can be the building block to construct more sophisticated statistical parametric and non parametric models. There are lots of possible applications for the results presented here. Of particular interest are the so called correlation networks. This networks are constructed from n time series computing a measure of dependency between them in a time window, and an appropriate varying or fixed threshold. Each network for each sliding time window is finally constructed including a link when the dependence between two of this time series is greater than the threshold. In this way a random sample G_1, G_2, \dots, G_ℓ of graphs is obtained. This method of constructing networks is becoming a standard procedure in the areas of finance and neuroscience. Another important application is the classification of patients (e.g., high or low-risk to have a particular cognitive disorder) from their (fMRI, MEG, or EEG) resting state functional networks.

7.1. Open problems. We end by proposing a short list of statistical problems that can be solved with similar techniques.

- Canonical correlation can be performed using the same ideas introduced for principal components.
- Other classification methods can be considered, as well as aggregation methods. How can the random forest procedure be adapted to this setup?
- Principal curves: Finding geodesic curves in the space \mathcal{G} with a fixed length that minimize an adequate fit notion.
- Consider minimum distance estimators for some parametric models.

REFERENCES

- [1] Jiang X, Munger A, Bunke H. (2001). On Median Graphs: Properties, Algorithms, and Applications. *IEEE TPAMI*, **23**, 1144-1151.
- [2] Tukey, J (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians, Vancouver, 1974*, **2**, 523-531.
- [3] Liu R (1988). On a notion of simplicial depth. *Proc. Natl. Acad. Sci. USA*, **97**, 1732-1734.
- [4] Brown, B. (1983) Statistical uses of the spatial median. *J. R. Stat. Soc. B*, **45**, 25-30.
- [5] Small C (1996). A survey of multidimensional medians. *Int. Stat. Rev.*, **58**, 263-277.
- [6] Zuo Y, Serfling R (2000). General notions of statistical depth function. *The Annals of Statistics*, **28**, 461-482.
- [7] Vardi Y, Zhang C (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA*, **97**, 1423-1426.
- [8] Auer J (1995). An elementary proof of the invertibility of Distance matrices. *Linear and Multilinear Algebra*, **40**, 119-124.
- [9] Rosenblatt M (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA*, **42**, 43-47.
- [10] Cuesta-Albertos J, Nieto-Reyes A (2008). The Tukey and the random Tukey depths characterize discrete distributions. *J. Multivariate Anal.*, **10**, 2304-2311.
- [11] Eguiluz V, Chialvo D, Cecchi G, Baliki M, Apkarian A (2005). Scale-free brain functional networks. *Phys. Rev. Lett.*, **94**, 018102 (2005).
- [12] Fraiman D, Saunier G, Martins E, Vargas C (2014). Biological Motion Coding in the Brain: Analysis of Visually Driven EEG Functional Networks. *PloS One*, 0084612.

- [13] Bickel P, Chenb A (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, **106**, 21068-21073.
- [14] Pollard Strong consistency of k -means clustering (1981). *Ann. Statist.*, **9**, 135-140.
- [15] Devroye L, Györfi L, Lugosi G (1996). A Probabilistic Theory of Pattern Recognition (Springer).
- [16] Cerqueira A, et al. (2014). Hypothesis testing for random graphs (preprint).

¹DEPARTAMENTO DE MATEMÁTICA Y CIENCIAS, UNIVERSIDAD DE SAN ANDRÉS, BUENOS AIRES, ARGENTINA.

²CONSEJO NACIONAL DE INVESTIGACIONES CIENTÍFICAS Y TECNOLÓGICAS (CONICET), ARGENTINA.

E-mail address: `dfraiman@udesa.edu.ar`

³DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA, USA.

E-mail address: `nfraiman@math.upenn.edu`

⁴CENTRO DE MATEMÁTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD DE LA REPÚBLICA, URUGUAY.

E-mail address: `rfraiman@cmat.edu.uy`