POLITECNICO MILANO 1863

The Euclidiscrete Distance: a new metric for the analysis of network data

Marco Teodori
matr. 920335

Supervisor: Simone Vantini
Co-supervisor: Anna Calissano

Academic Year 2020/21

This is a Thesis work for the Master's Degree in Mathematical Engineering.

The Euclidiscrete Distance: a new metric for the analysis of network data

Marco Teodori
Politecnico di Milano

Supervisor: Simone Vantini
MOX- Dept. of Mathematics, Politecnico di Milano

Co-supervisor: Anna Calissano
MOX- Dept. of Mathematics, Politecnico di Milano

October 7, 2021

Abstract

Weighted networks are rather complex data objects, so that it is not trivial to extract statistics or summary indices from a population of networks, especially if the nodes of the networks are unlabelled. Proper assumptions on the distribution of the values of the edges, and on the geometry of a so called *Graph Space*, are needed in order to construct the best explanatory statistics for such framework. In particular, we introduce a new metric with both a discrete and a continuous part, suited for networks having zero-valued edges with positive probability. Such metric is the basis of a loss function we introduce, called *Euclidiscrete Distance*, which for example is minimized by a peculiar network being the *Fréchet Mean* of a dataset. We see also that such metric can be reinterpreted to be used for unlabelled networks, providing an opportune graph matching algorithm.

Keywords: Network-valued/Graph-valued data, Graph Space, Network/Graph metric, Mean Network/Graph, Weighted networks, Graph matching.

1 Introduction

Network Data analysis is a branch of Object oriented data analytics involving networks, which are more complex than single numbers or vectors. Networks are particular objects represented by graphs with a certain number of nodes, connected by some edges. There exist many phenomena from several fields which may be explained by a population of networks. Indeed, scientific studies and works about networks are now quite common, but handling such complex objects becomes challenging soon. In particular, we focus on the naturally non-Euclidean structure of the notion of "distance" between networks, since both nodes and edges may both have attributes coming from a continuous probability distribution, but be absent in the datum with positive probability. The presence of a discrete component inside eventual distance measures has already been studied [Wang and Marron, 2007]. Moreover, we might have to deal with unlabelled networks, for which each node has no label and the corresponding datum can be represented in different ways, or permutations. Accordingly, our aim is to provide a proper distance measure between (labelled or unlabelled) networks. As an application of such distance, we are going to calculate the so called *Fréchet Mean* network of a population, namely a graph which minimizes an overall distance function from our data. This concept has been well explained by [Calissano et al., 2020], together with the definition of the space in which we intend our data to live in (see also [Fréchet, 1948] for completeness).

1.1 State of the art

We focused on the already present literature about multivariate regression to find a possible extension to our object of interest. *Multivariate* because our response variable should be at least as complex as a vector, and *regression* since the notion of mean is bound to the one of intercept. Indeed, even if in this work no regression is applied, the concept of regression always relies on a minimization of some error or distance measure, which is what we are actually looking for. An exhaustive definition of the topic can be found in [Johnson et al., 2014] (chapter 7.7). Moreover, weighted networks can be interpreted as if the attribute of a non-existent edge were just missing. Thus, several approaches to the issue can be examined, from the theory of randomness of missing data [Bojinov et al., 2020] to tobit models Anastasopoulos et al. [2012]. Consequently, the reference model we considered to be the most suitable with the geometry of networks is the so called *zero-augmented* model, treated by [Yau et al., 2002], [Wright et al., 2017]. This choice relies on the fact that the presence or absence of an edge in a weighted network is captured by a Bernoulli distribution, whereas the probability distribution of the value of the link (when present) is generally continuous.

In order to face the computational cost which might arise from such complex models, having a network object in response, a simplified version of them should be considered, using a frequentist approach to maximize the overall likelihood of the result, and discarding random effects.

Furthermore, we will try to unravel the problem of dependence between the probability distributions of the attributes of each network. A good idea, for example, might be to assume independence between occurrences and values of *distant* edges. Such approach has been discussed by [van der Pol, 2019], who introduced an Exponential Random Graph Model (ERGM) in which only dependences within specific structures (as 2-stars, 3-stars or triangles) are considered. Other methods linger on the correlation among components of a multivariate Bernoulli distribution, which has to be treated carefully, since the Pearson's correlations between two Bernoulli distributions generally cannot assume all values in the interval $[-1; 1]$ [Geenens, 2019]. A model which might be taken into account for multivariate Bernoulli distribution is the so called Ising Model [Dai et al., 2013], in which only correlations between nodes/edges of order 2 do exist.

We conclude this introduction by saying that a detailed analysis of such methods is important in order to master the subject, although every possible correlation between edges gets much blurred when networks have unlabelled nodes. Thus, we should introduce a metric for which the probability distribution of the edges could be considered independent at all. In Section 2, after defining an opportune space for our networks and a proper distance measure, we will see that such assumption of independence can be stated for certain metrics. In Section 3, we generate simulated data to apply our method and obtain Fréchet Mean and Fréchet Medoid, treating both labelled and unlabelled data, and introduce the specific algorithm for unlabelled networks. In Section 4, the same results are applied to a real dataset of interactions among players of the game *The Resistance.*

2   Method

In this section, a definition of the Graph Space in which our network data live is provided, and is followed by the definition of our proposed metric, along with the description of a possible application. The notion of Graph Space in particular is a simplification of the one enunciated by [Calissano et al., 2020], and it will be applied to a case which is simpler than its actual potential.

2.1 Graph Space

We define a graph as a triple $G = (V, E, a)$, where V is a set of v nodes, E is a set of m edges provided with attributes, and $a : E \to X$ is an attribute map. Since *networks* are represented mathematically as graphs, the terms are considered equivalent. In our framework, we consider each edge to have a single real attribute, and do not contemplate attributes of nodes for simplicity, so that $X = \mathbb{R}^m$. For now, the single attribute of each edge can be considered as a *weight* of the edge, so that the two terms will be treated as equivalent at least until Section 2.2. Moreover, without loss of generality networks are considered undirected and potentially complete, since each absent edge may have 0 as attribute. In this way, we have that $m = \frac{v(v-1)}{2}$.

This formulation allows to represent a graph with a m-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_m)$, and it is safely applied on labelled networks, since we can explicit a bijection between each component of a network $\boldsymbol{x} \in X$ with a pair of nodes $(i; j)$, with $i, j \in \mathbb{N}, 1 \le i < j \le m$, meaning that the specific component is an edge between nodes i and j (we do not allow self-loops again for simplicity). For unlabelled networks, a more complex thought must be made. Each network $\boldsymbol{x}$ represents just one of all $v!$ possible permutations of the nodes.

Let now T be the set of all permutation operators. The most simple operator to construct a permutation is the *swap operator* s_{ij}, which exchanges node i with node j ($i < j$). A generic permutation $t \in T$ can be seen as a collection of (at most $v - 1$) swaps, indeed.

Remark 2.1. One may think that a network $\boldsymbol{x}$ as it is defined may have $m!$ possible permutations, but $\boldsymbol{x}$ is not just an $m-$dimensional vector, since each of its components is associated with a specific pair od nodes. The swap operator s_{hk} exchanges the components of $\boldsymbol{x}$ associated with pairs:

- $(i; h)$ and $(i; k)$, $\forall i < h$;

- $(h; i)$ and $(i; k)$, $\forall h < i < k$;

- $(h; i)$ and $(k; i)$, $\forall i > k$.

In such framework, $\forall \boldsymbol{x} \in X, t \in T$ we call $t\boldsymbol{x} \in X$ a permutation of network $\boldsymbol{x}$. We indicate with X/T the Graph Space which considers each element $\boldsymbol{x} \in X$ along with all its permutations, and it is actually a quotient space of equivalence classes $[\boldsymbol{x}] \in X/T$.

Example 2.1. This example shows in detail a graph $\boldsymbol{x}$ with $v = 5$ nodes ($m = 10$ possible edges) subject to a swap $s_{24} \in T$ between nodes 2 and 4. Both $\boldsymbol{x}$ and the obtained permuted network $s_{24}\boldsymbol{x}$ belong to the equivalence class $[\boldsymbol{x}] \in X/T$.

Edge	(1;2)	(1;3)	(1;4)	(1;5)	(2;3)	(2;4)	(2;5)	(3;4)	(3;5)	(4;5)
$\boldsymbol{x}$	1	0	0	1	1	1	0	0	0	1
	$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$	$\downarrow$		$\downarrow$
$s_{24}\boldsymbol{x}$	0	0	1	1	0	1	1	1	0	0
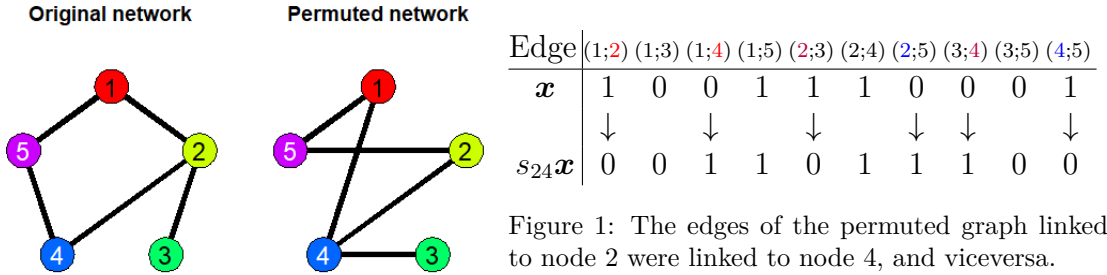
Figure 1: The edges of the permuted graph linked to node 2 were linked to node 4, and viceversa.

Each graph of this work has been represented with the R package *network*, whose documentation is provided by [Butts, 2008].

2.2 Metrics

In order to produce a consistent metric for our data, we have to begin with the probabilistic likelihood of the graphs, since we would like to calculate a mean, or a medoid, to be also the most likely result to occur. An edge i of a network, with $1 \leq i \leq m$, as it is defined, has naturally a probability p_i of being present, so that this behaviour is captured by a Bernoulli distribution. An edge which does occur has on the other hand a value coming from a continuous distribution with probability density f_i, having a set of parameters θ_i. The overall likelihood function for the graph $\boldsymbol{x}$ factorizes then into a probability component L_1, which concerns only the aspect of occurrence, and a conditional component L_2, which concerns only the situation upon the presence of the edges:

$$L_1(p_i, \boldsymbol{x}) = \prod_{i:x_i=0} (1 - p_i(\boldsymbol{x})) \prod_{i:x_i \neq 0} p_i(\boldsymbol{x}); \qquad L_2(\theta_i, \boldsymbol{x}) = \prod_{i:x_i \neq 0} f_i(\theta_i, \boldsymbol{x}). \quad (1)$$

If the respective parameters are independent of each other, these two parts can be treated completely separately [Yau et al., 2002]. Moreover, this formulation works also within a network population in which data are independent and identically distributed.

With such result, we may concentrate (for now) on binary networks, in which an edge can be absent or present with unitary attribute. The assumptions made before don't allow yet to consider even the edges of the same graph independent of one another; indeed, all the network $\boldsymbol{x}$ may influence the nature of probability distributions for each edge in (1).

Let us concentrate on the simplest case, the bivariate one (namely, a graph with two edges), and extract analytically the maximum likelihood estimator: we want to see whether the dependence between occurrences of the edges is relevant.

5

Independent edges: we have a set of n independent and identically distributed graphs $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where the two components x_{i1}, x_{i2} are realizations of the Bernoulli random variables $(X_1, X_2) \sim (Be(p), Be(q))$, $p, q \in [0; 1]$. Since $X_1 \perp X_2$, the maximum likelihood estimators for p and q are:

$$\hat{p}_{MLE\perp} = \frac{\sum\limits_{i=1}^{n} x_{i1}}{n}; \qquad\qquad \hat{q}_{MLE\perp} = \frac{\sum\limits_{i=1}^{n} x_{i2}}{n}. \qquad (2)$$

Because each component can be analysed separately.

Dependent edges: in this general case, both the edges of a graph can have either 0 or 1 as values, and so the log-likelihood $l(p_{11}, p_{10}, p_{01}|x_{1:n})$ is:

$$l = \sum_{x=1}^{n} x_{i1}x_{i2} \log \frac{p_{11}p_{00}}{p_{10}p_{01}} + \sum_{x=1}^{n} x_{i1} \log \frac{p_{10}}{p_{00}} + \sum_{x=1}^{n} x_{i2} \log \frac{p_{01}}{p_{00}} + n \log p_{00}. \qquad (3)$$

Where p_{ab} is the probability of a network to have the first edge with value a and the second with value b, and $a, b \in \{0, 1\}$. Since naturally $p_{00} = 1 - p_{11} - p_{01} - p_{10}$, calculations lead to these Maximum Likelihood Estimators:

$$\hat{p}_{11} = \frac{\sum\limits_{i=1}^{n} x_{i1}x_{i2}}{n}; \;\; \hat{p}_{10} = \frac{\sum\limits_{i=1}^{n} x_{i1}(1 - x_{i2})}{n}; \;\; \hat{p}_{01} = \frac{\sum\limits_{i=1}^{n}(1 - x_{i1})x_{i2}}{n}. \qquad (4)$$

See Appendix A.1 for detailed calculations. Let now $\hat{p}_{MLE} = \hat{p}_{11} + \hat{p}_{10}$, obtained by relaxing the independence assumption. We can easily prove (both formally and by experiment) that:

$$\mathbb{E}\left[\hat{p}_{MLE\perp} - \hat{p}_{MLE}\right] = 0; \qquad\qquad Var\left[\hat{p}_{MLE\perp} - \hat{p}_{MLE}\right] = 0. \qquad (5)$$

Where $\hat{p}_{MLE\perp}$ is the one defined in equation (2); similarly for $\hat{q}_{MLE} = \hat{p}_{11} + \hat{p}_{01}$. Namely, the maximum likelihood estimator of the probability of an outcome is just its relative frequence. Without much effort, such result can be achieved also with networks having more than two edges, reasoning by induction.

This is a considerably important outcome, because we may work in this environment to consider edges independent without losing information. However, finding an appropriate distance measure (or loss function) to be minimized is not trivial. A proper loss function should not include dependences among edges in its formulation, in order to make the hypothesis of independence irrelevant. Therefore, let us define two candidate loss functions:

Definition 2.1. The Probability of Error (PE) of a binary network $\boldsymbol{x}$ with respect to a population of binary networks $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_n\}$ is defined as:

$$PE(\boldsymbol{x}) = 1 - p_{\boldsymbol{x}}. \qquad (6)$$

Where $p_{\boldsymbol{x}}$ is the relative frequence of network $\boldsymbol{x}$ within the set.

6

Remark 2.2. In this framework, true probabilities (which are impossible to know from a dataset) are estimated through relative frequences, so that when the term *probability* is used, it naturally refers to a relative frequence.

Definition 2.2. The Expected Hamming Distance (EHD) of a binary network $\boldsymbol{x}$ with respect to a sample of binary networks $\{\boldsymbol{x}_1, \dots \boldsymbol{x}_n\}$ is defined as:

$$EHD(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} d_H(\boldsymbol{x}_i, \boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} |x_{ik} - x_k|. \tag{7}$$

Where d_H is the Hamming Distance between two networks (which is a proper metric in the space of binary networks with a specific number of edges). It can be seen that this formulation is indeed an average of the Hamming distances between network $\boldsymbol{x}$ and each one of the graphs in the sample.

Using an opportune example, we can show that these loss functions behave in different ways.

Example 2.2. The tables below show a particular situation in which one of the joint probabilities/frequences (in this case, p_{11}) is greater than the others, but it is not true that $p > 0.5 \wedge q > 0.5$ (in this case, $p = 0.4$). Recall that p, q are the parameters of the marginal Bernoulli distributions, whose estimators are defined in (2). It happens that $PE(\boldsymbol{x})$ is minimized by a different network when we simplify the frequences assuming independence between the two marginal distributions of the edges, whereas $EHD(\boldsymbol{x})$ not only does not change its choice, but preserves even the value of the loss, and so it is a suitable function for our purposes.

True joint distribution

Distr.	$y_2 = 1$	$y_2 = 0$
$y_1 = 1$	**0.4**	0
$y_1 = 0$	0.3	0.3

Assuming independence

Distr.	$y_2 = 1$	$y_2 = 0$
$y_1 = 1$	0.28	0.12
$y_1 = 0$	**0.42**	0.18

$LF(\boldsymbol{x} = (1, 1))$

Loss F.	Joint	Marginal
PE	**0.6**	0.72
EHD	0.9	0.9

$LF(\boldsymbol{x} = (0, 1))$

Loss F.	Joint	Marginal
PE	0.7	*0.58*
EHD	**0.7**	**0.7**

The nature of the resulting network can be deduced from the following example:

Example 2.3. The table below shows a sample of $n = 5$ networks, with $v = 4$ nodes and $m = 6$ edges.

Edge	(1;2)	(1;3)	(1;4)	(2;3)	(2;4)	(3;4)
$\boldsymbol{x}_1$	1	1	1	1	0	0
$\boldsymbol{x}_2$	1	1	1	0	0	1
$\boldsymbol{x}_3$	1	1	1	0	1	1
$\boldsymbol{x}_4$	0	0	1	1	1	1
$\boldsymbol{x}_5$	1	1	0	0	1	1
Candidate mean $\bar{\boldsymbol{x}}$	1	1	1	0	1	1

Since EHD is based on a sum of different occurrences of the edges, the most natural candidate (Fréchet) Mean is simply the one obtained by majority rule for each edge, which indeed minimizes the total number of different edges with respect to $\bar{\boldsymbol{x}}$. Then, $EHD(\bar{\boldsymbol{x}}) \leq EHD(\boldsymbol{x}) \ \forall \boldsymbol{x} \in X$.

At this moment, we should reconnect to the general zero-augmented formulation for the edges, and see if there exist a proper enrichment of the EHD to continuous distributions. Let then $\boldsymbol{x} \in X$ be a network of our defined Graph Space.

Definition 2.3. The Binary Counterpart of $\boldsymbol{x}$ is a binary network $\boldsymbol{x}^B$, whose components have value:

$$x_k^B = \begin{cases} 1 & \text{if } x_k \neq 0 \\ 0 & \text{if } x_k = 0 \end{cases} \quad \forall k = 1, \ldots, m. \tag{8}$$

The idea behind this definition is to extend the calculation of the Expected Hamming Distance to any networks $\boldsymbol{x} \in X$, by applying its formulation to their binary counterparts $\boldsymbol{x}^B$. Moreover, since EHD is by definition a sum of zeros and ones, we may formulate it within the space of bounded sequences $\ell_2 \supset X$, rather than ℓ_1 (squaring the terms does not affect the result). The latter space might be more natural to define a distance which is basically a Manhattan Distance, but now we would like to get into an Euclidean framework, because we are applying to more general networks whose edges have values coming from continuous distributions. We conclude that the Expected Hamming Distance may be reformulated as follows:

$$EHD(\boldsymbol{x}) = EHD(\boldsymbol{x}^B) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} |x_{ik}^B - x_k^B| = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{m} (x_{ik}^B - x_k^B)^2. \tag{9}$$

In order to add a continuous part to our loss function, it's clear that an Euclidean expression based on the plain difference between the values of correspondent edges could be considered: the strength of such formulation stays in its simplicity and insensitivity to dependences among edges. Therefore, maintaining the quadratic expression we obtain an overall loss function:

Definition 2.4. The Euclidiscrete Distance (EDD) of a network $\boldsymbol{x} \in X$ with respect to a sample of networks $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_n\}$ is a loss function defined as:

$$EDD(\boldsymbol{x}) = \frac{\omega_1}{m} EHD(\boldsymbol{x}) + \frac{\omega_2}{nm} \sum_{i=1}^{n} \sum_{k=1}^{m} \frac{(x_{ik} - x_k)^2}{R_k^2}. \tag{10}$$

Here, ω_1 and ω_2 are (non-negative) weights to be assigned depending on the dataset, and R_k denotes the range of values assumed by edge k.

Remark 2.3. The term *Discrete* in this definition does not refer to the so called *Discrete Distance*, but relates to the discrete range of assumable values from the first piece of the function.

A good practice should be an opportune transformation of the networks in order to have all the values in $[0; 1]$, so that $R_k = 1 \ \forall k = 1, \ldots, m$. In this manner, ω_1 and ω_2 can be chosen so that $\omega_1 + \omega_2 = 1$, in order to have a normalized value for $EDD(\boldsymbol{x})$. If $\omega_2 > 0$, it can be easily proven that the "dissimilarity" underlying the Euclidiscrete Distance, namely:

$$d_{EDD}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{m} (x_k^B - y_k^B)^2 + \sum_{k=1}^{m} (x_k - y_k)^2 \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X, \tag{11}$$

being the sum of a pseudometric with a metric, is therefore a metric (see [Wang and Marron, 2007], for further details). Furthermore, since it sums all the differences between attributes of the edges singularly, this formulation works even supposing a dependence among the edges.

Remark 2.4. Not all the populations of networks, however, might be conceptually coherent with such formulation. Indeed, we should wonder whether, for a specific sample of networks, it exists an effective difference between a (basically) zero-weighted edge and a non-existent edge at all, or not. If the answer is *no*, then the values of the edges can be considered as *weights* and absent edges can be interpreted as zero-weighted edges, so that EDD suits also the conceptual meaning of the dataset to which it is applied. But, if the answer is *yes*, the values of the edges may be considered as *attributes*, with the zeros meaning an effective absence if it. With such consideration, the formula (10) presents a conceptual discrepancy when it has to include differences between existent edges with non-existent counterparts. Hence, we may also call *Weighted* Euclidiscrete Distance (WEDD) the loss function defined in (10), and try to define an *attributed* version of such loss function, which may be applied in the specific case written above. Anyway, in this work we are going to apply the Weighted Euclidiscrete Distance for the calculation of Fréchet Mean and Fréchet Medoid of a population of graphs. See Appendix A.2 for further details.

9

2.3 The Fréchet Mean

In this section, we introduce the concept of Fréchet Mean network with respect to a set of network data, applying the metric introduced in Section 2.2 to labelled and unlabelled samples.

Definition 2.5. The Fréchet Mean $\bar{\boldsymbol{x}} \in X$ with respect to a sample of labelled graphs $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset X$ is the network which minimizes a loss function relative to some distance between a candidate graph $\boldsymbol{x}$ and each of the networks of the sample [Calissano et al., 2020]. Using the expression (10) of the Euclidiscrete Distance, we have:

$$\bar{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in X} EDD(\boldsymbol{x}). \tag{12}$$

The Fréchet Medoid $\widetilde{\boldsymbol{x}}$ is instead found among the network data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset X$, namely:

$$\widetilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}} EDD(\boldsymbol{x}). \tag{13}$$

When it comes to unlabelled networks, a further reasoning is necessary. The Graph Space underlying such type of data is X/T, defined in Section 2.1 and formed by equivalence classes. Thus, the expression for the Fréchet Mean becomes:

$$[\bar{\boldsymbol{x}}] = \arg\min_{[\boldsymbol{x}] \in X/T} EDD([\boldsymbol{x}]). \tag{14}$$

Similarly the formula for the Medoid. In order to legitimate this notation, it remains to give a proper definition of the relative metric $d_{EDD}([\boldsymbol{x}], [\boldsymbol{y}])$, starting from expression (11) and referring again to [Calissano et al., 2020].

Definition 2.6. Let $[\boldsymbol{x}], [\boldsymbol{y}] \in X/T$ be two equivalence classes of unlabelled networks. The metric underlying the Euclidiscrete Distance in this Graph Space is:

$$d_{EDD}([\boldsymbol{x}], [\boldsymbol{y}]) = \min_{t \in T} d_{EDD}(t\boldsymbol{x}, \boldsymbol{y}). \tag{15}$$

Furthermore, if it happens that $d_{EDD}(t\boldsymbol{x}, \boldsymbol{y}) = \min_{t' \in T} d_{EDD}(t'\boldsymbol{x}, \boldsymbol{y})$, we say that $t\boldsymbol{x}$ is in *optimal position* with respect to $\boldsymbol{y}$.

Remark 2.5. Observe that $d_{EDD}(\boldsymbol{x}, \boldsymbol{y})$ is a metric in X, while $d_{EDD}([\boldsymbol{x}], [\boldsymbol{y}])$ is a distance in X/T. Moreover, it should be noticed that the Fréchet Mean is generally not unique, because X/T is a Geodesic Space, namely the shortest *path* connecting two graphs of the Graph Space, called *Geodesic*, may be not unique [Calissano et al., 2020]. This fact is not necessarily a problem for the purposes of this work, but it helps to make the overall formulation more clear.

3 Simulated data

In this section, we calculate Fréchet Mean and Fréchet Medoid from simulated samples of labelled and unlabelled networks, in order to study the effects of the application of the Euclidiscrete Distance (definition 2.4). We are going to analyse the complexity and see what happens when the weights $\omega_1, \omega_2 > 0$ (with $\omega_1 + \omega_2 = 1$) vary, and what happens when the method for labelled networks is used for a population of unlabelled networks (and vice-versa). All the analysis has been performed with R.

All simulated data have edges following a zero-augmented distribution, in which the probability of being zero-valued is independent of its continuous counterpart (see (1)), and the overall distribution is different for each edge. Classical distributions as Beta, Uniform, Exponential, Gamma and Normal with different parameters are utilized and possibly transformed in order to have all the values of the edges in $[0; 1]$. To make a generated population of networks unlabelled, each graph has been permuted in a different way, using the method better described in remark 2.1.

3.1 Labelled networks

With our formulation, labelled networks allow the assumption of independence among the edges, thus it is easy to calculate the mean network, because such task can be performed one edge at a time. Set indeed a sample of one-edged networks $\{x_1, \ldots, x_n\} \subset X = \mathbb{R}$ ($m = 1$). $EDD(x)$ becomes a real function

$$EDD(x) = \omega_1 EHD(x) + \frac{\omega_2}{n} \sum_{i=1}^{n} \frac{(x_i - x)^2}{R^2}, \tag{16}$$

where, given the number of positive edges k, the Expected Hamming Distance has value

$$EHD(x) = \begin{cases} \dfrac{n-k}{n} & \text{if } x \neq 0 \\ \dfrac{k}{n} & \text{if } x = 0 \end{cases} \tag{17}$$

and the second part of $EDD(x)$ has a parabolic behaviour: we observe that it achieves a minimum at the average of all the edges. Thus, the only possible values for the point of minimum of $EDD(x)$ is either 0 or the average of the one-edged network in the sample, including the zero-valued ones. This result can be similarly extended to network populations with any number of edges.
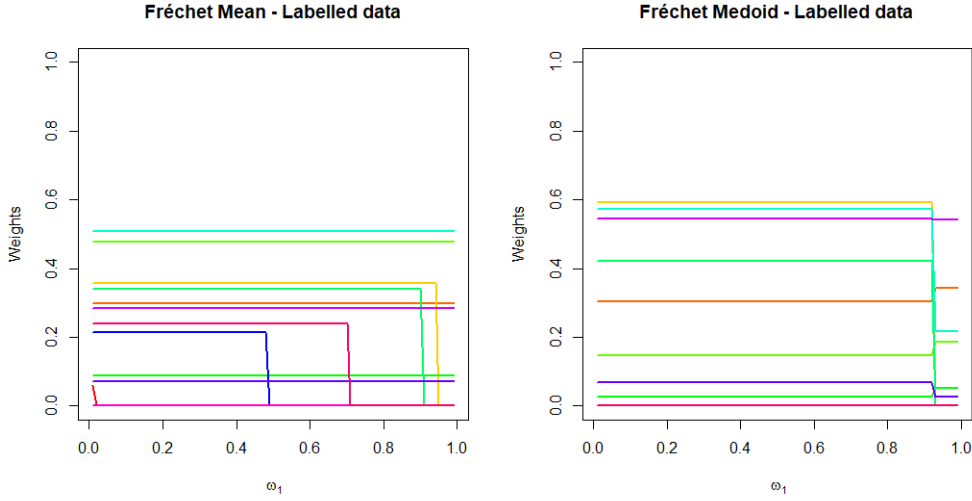
Figure 2: Fréchet Means (left) and Medoids (right) obtained for every value of ω_1, ω_2, from a simulated population of $n = 1000$ networks, with $v = 6$ nodes, $m = 15$ edges. Each line represents one edge of the extracted network. As noted before, each edge of the Fréchet Mean assumes as weight just either the average of the sample or zero, and as ω_1 increases, the number of positive edges reduces from 11 to 6 (out of 15).

Extracting the Fréchet Mean presents a complexity of $O(nm) = O(nv^2)$. We note that as ω_1 increases, some of the edge values drop to 0. Thus it can be said that higher values of ω_1 make the result be more conservative, and this recalls the notion of *Lasso Regularization*, since the part of $EDD(\boldsymbol{x})$ carried by ω_1 lives naturally in ℓ_1. For what concerns the Fréchet Medoid, the outcome is not the same, because possibilities are restricted to the effective networks of the sample, and the overall meaning is less organic. The process calculating the Medoid has a time complexity of $O(n^2m)$, since n controls also the number of alternatives to be considered.

To see that such result does not apply to a single networks, an extended analysis to more populations has been done, and all the obtained Fréchet Means has been collected into boxplots to visualize their distribution (Figure 3).

It might be now interesting to figure out what happens if the Euclidiscrete Distance on the labelled Graph Space X is applied to populations of unlabelled graphs. The process scans every network one "labelled" edge at a time (edges are considered in a specific order), but now all the values are shuffled among the columns of the dataset. Computing the Fréchet Mean, the averages of the edges become then very close, and a definitely poor result is obtained (see Figure 4). Also the computation of the Fréchet Medoid brings a far low significance, at least conceptually.
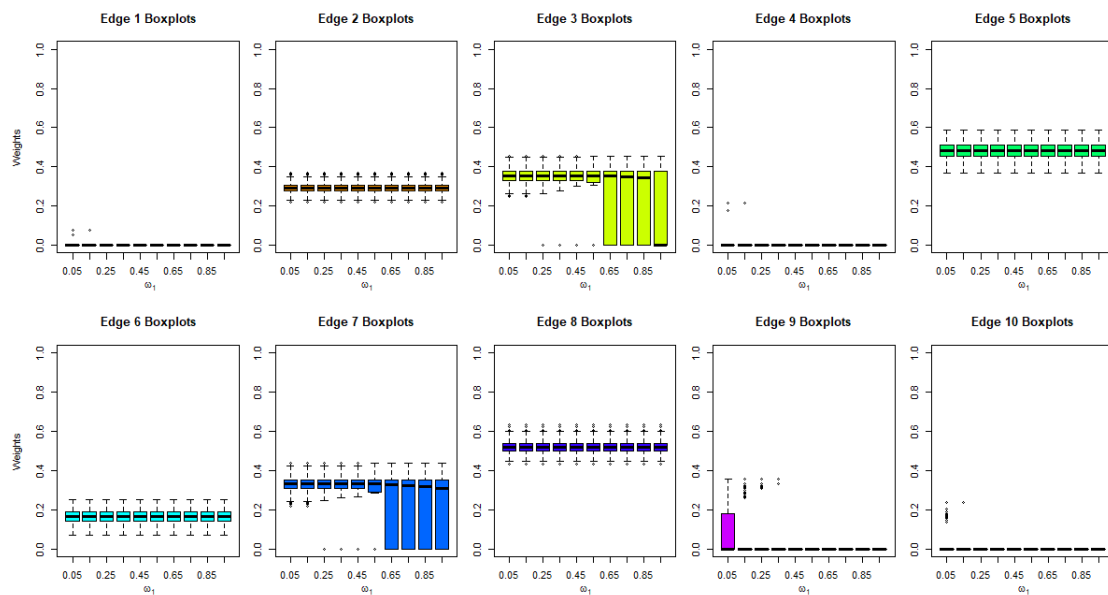
Figure 3: Boxplots of Fréchet Means from 500 different simulated datasets with the same distributions, with $m = 10$ edges and various choices of ω_1, ω_2. As ω_1 increases, we observe again a partial drop to 0 of the values of some edges, in particular edges 3, 7 and 9.

We conclude by saying that populations of unlabelled networks should be handled using concepts based only on the Graph Space of equivalence classes X/T, which considers the permutations of the edges.
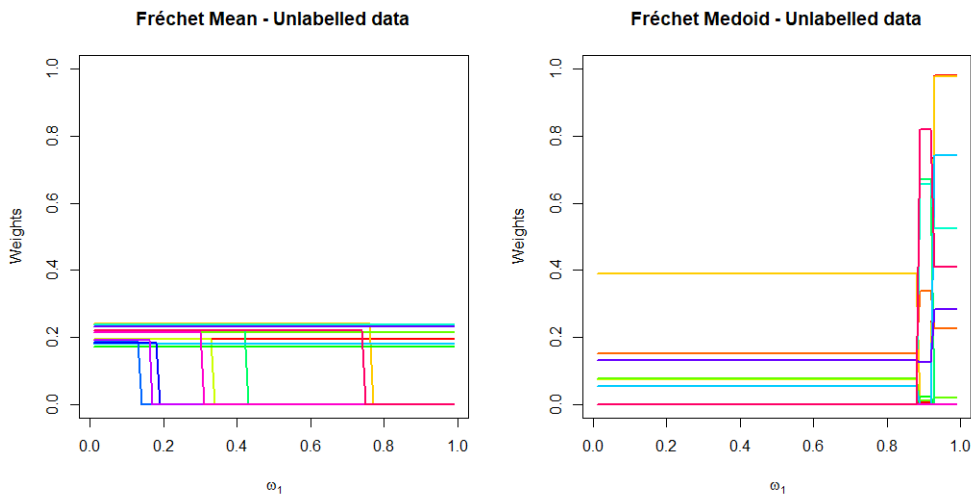

Figure 4: Fréchet Mean and Medoid of a population of unlabelled networks, treated as if they were labelled.

13

3.2   Unlabelled networks

In order to compute the Fréchet Mean (or Medoid) of a sample of unlabelled networks, we have just noticed that the metric $d_{EDD}([\boldsymbol{x}], [\boldsymbol{y}])$ (with $[\boldsymbol{x}], [\boldsymbol{y}] \in X/T$) has to be used. But the definition of such metric (see Definition 2.6) requires however the usage of the correspondent metric in X. To go across this link, we should explore all the permutations $t \in T$ and choose the best one before the calculation of our statistic (at this point, just in the Euclidean space X). As a general strategy, we now present the *Align All and Compute* (AAC) algorithm, which consists of iteratively selecting an optimal representative network from every equivalence class belonging to X/T [Calissano et al., 2020]. Indeed, the distance $d_{EDD}([\boldsymbol{x}], [\boldsymbol{y}])$ corresponds exactly to its counterpart in X after posing the first network in optimal position with respect to the other. We now define the AAC algorithm which, for each iteration, is based on first putting observations in optimal position with respect to the current mean/medoid estimate, then re-estimating the mean/medoid based on the aligned observations, until some convergence is achieved. The *Align All* segment of the process pertains to a more general subject called *Graph Matching*, inherent in particular to unlabelled networks.

Algorithm 1: Align All and Compute for the Fréchet Mean/Medoid

Data: $\{[\boldsymbol{x}_1], \ldots, [\boldsymbol{x}_n]\} \in X/T$ observations in X.
Result: An estimate of the Fréchet Mean/Medoid of $\{[\boldsymbol{x}_1], \ldots, [\boldsymbol{x}_n]\}$.
Initialization: Select randomly a permutation $\widetilde{\boldsymbol{x}}_i \in [\boldsymbol{x}_i], \forall i \in 1 : n$, and randomly one of the permuted data $\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{x}}_i \in \{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n\}$;
Set $s = 1$; (supposing a normalization of the distance)
Fix a threshold $\varepsilon > 0$;
While $s > \varepsilon$
    Put every observation in optimal position w.r.t. $\widetilde{\boldsymbol{x}}$, obtaining an aligned set of representatives $\{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n\} \subset X$;
    Compute the Fréchet Mean (or Medoid) $\bar{\boldsymbol{x}}$ in X of $\{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n\}$, minimizing $EDD(\widetilde{\boldsymbol{x}})$;
    Put $s = d_{EDD}(\widetilde{\boldsymbol{x}}, \bar{\boldsymbol{x}})$;
    Set $\widetilde{\boldsymbol{x}} = \bar{\boldsymbol{x}}$;
Return $[\bar{\boldsymbol{x}}]$.

When we apply the AAC for the Fréchet Medoid, convergence might not be reached because of the possibility of a cyclic solution (though significant). On the other hand, it can be shown that the algorithm for the mean theoretically converges in finite time [Calissano et al., 2020], but it might be a long time due to the definitely high $O(nv!)$ time complexity. A threshold ε may be applied to control at least the number of iterations.

Note also that this algorithm in practice relies on inexact Graph matching, as such process is generally NP complete [Calissano et al., 2020]. Indeed, a matching might also separate the values of the edges more than the reality (as if it were the first component of a Principal Component Analysis). Above all, the major drawback of this algorithm, in common with most of Graph Matching processes, is a high sensitivity to initial conditions (Figure 5). There might be indeed many local minima of the loss function, reached in very few iterations, with suboptimal values, .
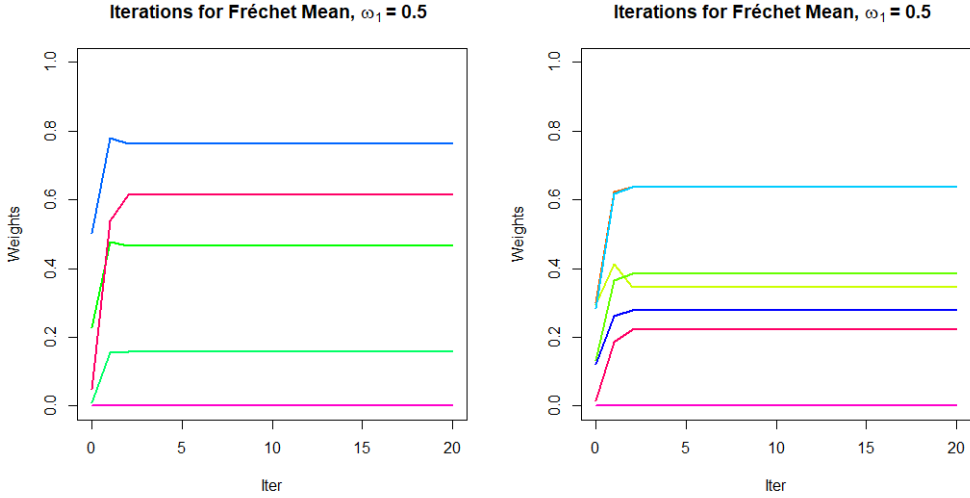
Figure 5: Fréchet Means extracted from the same unlabelled network population ($n = 1000, v = 6$), with two different initial conditions and $\omega_1, \omega_2 = 0.5$. Each line represents the value of an edge. Observe that the algorithm converges in very few iterations, but the resulting graph has always the same positive edges of the initial network. Such phenomenon does not always happen, but it has a probability much higher than it should.

Due to this fact, choosing the same starting network with different values of ω_1, ω_2 does not affect the number of zero-valued edges of the solution, on the contrary compared to the Fréchet Mean of labelled networks populations. Eventually, the two main obstacles of this procedure are time complexity and sensitivity to initial conditions.

To overcome the first, greedier Graph Matching techniques might be considered, such as more advanced metrics based on node centrality [Dwivedi and Singh, 2020], Munkres Algorithm [Riesen and Bunke, 2009], and the use of random walks [Gori et al., 2005].

About the latter issue, a multi-start process should be taken into account, even if this entails a greater time complexity. Different initial graphs may be chosen with respect to what edges are non-zero, so that ideally there should be 2^m starting points. However, such choice would be really expensive, hence the safest procedure

should be picking a fixed number of samplings in the dataset and using them as initial networks.

Ultimately, we may wonder what happens when the AAC algorithm is applied to a population of already labelled networks. The best outcome should be obtaining the same Fréchet Mean/Medoid network it would result from the analysis carried out with labelled networks (see Section 3.1). Nevertheless, the Graph Space of equivalence classes X/T is structurally larger, meaning that the loss function of the Fréchet Mean/Medoid obtained with AAC is lesser or equal than the standard one. If it is strictly lesser, for sure some permutations of the initial dataset have occurred: we call such phenomenon *over-matching*, or hyper-alignment.
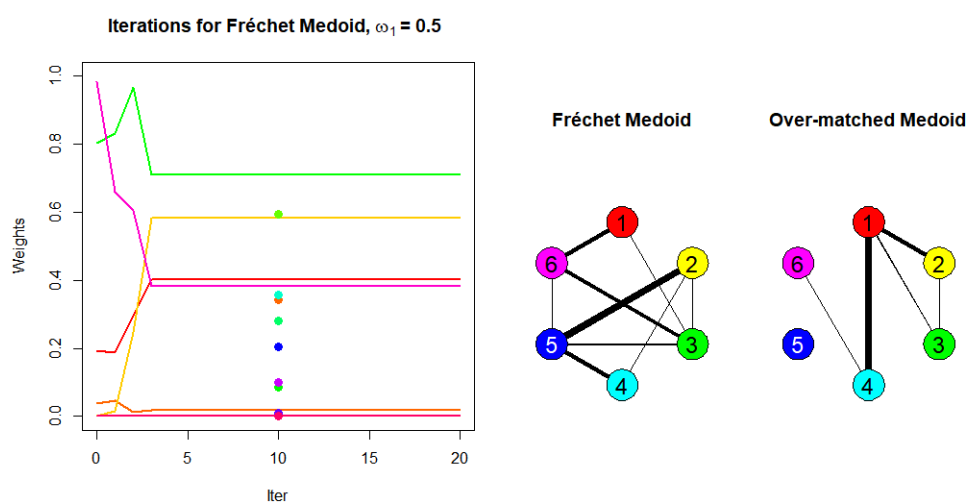

Figure 6: On the left, the values of the edges of the Fréchet Medoid obtained with AAC algorithm (lines) compared with the ones obtained with the more proper approach with labelled networks (points). On the right, the two represented networks, which are very different from each other: this entails that over-matching is definitely compromising. A population of $n = 1000$ networks with $v = 6$ nodes has been used.

4   Case study

We now analyse the introduced calculations on a real dataset, provided and already handled by [Bai et al., 2019], [Kumar et al., 2021] and based on a set of board game matches. The game is called *The Resistance*, and consists of a group of missions the players have to face, divided in two teams: the Resistance, aiming to complete successfully the missions, and the Spies, who secretly try to make the missions fail (only Spies know how the teams are divided). Each match corresponds to a different set of data, in which each network represents an instant. Nodes correspond to the players (from 5 to 8), and each edge tracks interactions and

16

glances between two specific players. For simplicity, we constructed an undirected version of the graphs. In this work, we apply our metric extract first the Fréchet Mean and Medoid from the labelled dataset of a single game, and then trying to analyse more games together, in order to obtain a mean network for all the games: the latter comes to be a naturally unlabelled framework.
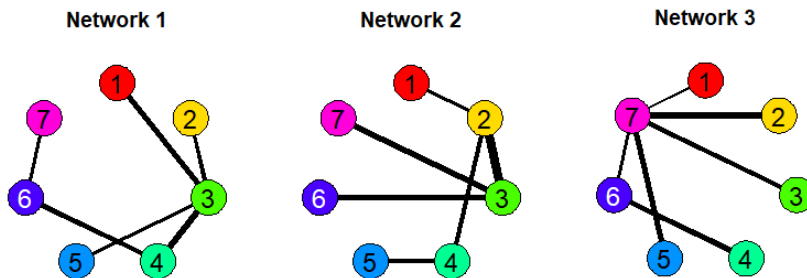

Figure 7: Sample networks from a *The Resistance* game. From the interactions we can deduce the player whose turn was in progress at each instant.

4.1 Single game analysis

In this section, the computation of Fréchet Mean and Medoid has been carried out on a population of $n = 7323$ networks, representing the interactions for each instant of one game among $v = 7$ players. Each node is bound to a fixed player, so that the dataset is labelled. The computed statistics have the purpose to be single networks which may explain the overall flow of the game. From Figure 8 we can see that player 3 is the one who interacts more with the others, during the game. With lower values of ω_1 the extracted network contains more positive-valued edges (as said in Section 3.1), including some triangles: for what concerns the game dynamics, players involved in triangles are more likely to be the three *Spies* of the game.
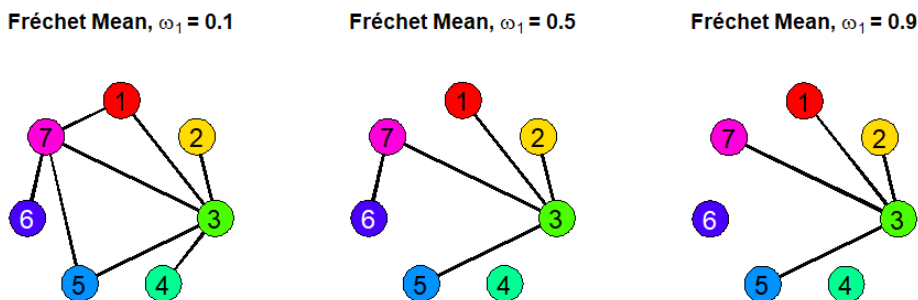

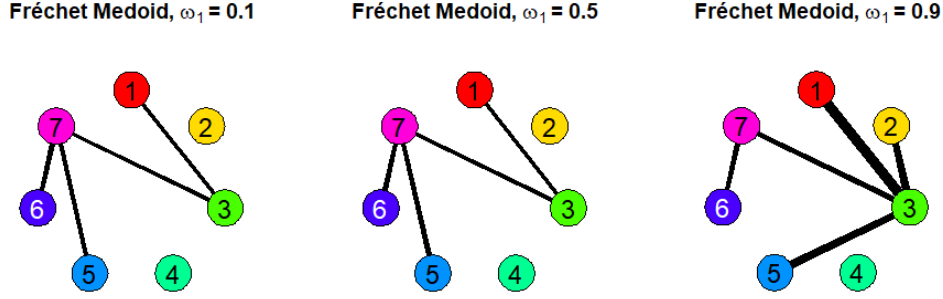Figure 8: Fréchet Means of the game with different values of ω_1, ω_2.

17

Figure 9: Fréchet Medoids of the game with different values of ω_1, ω_2.

Also the Fréchet Medoid has been computed (Figure 9), and from it we observe that player 7 had noticeable interactions with the others as well, but player 3 had the strongest interplays altogether (as we can see from tougher edges).

4.2 Multiple games analysis

We selected $g = 23$ games with $v = 7$ players, where each game embodies a different population of networks representing instants. Observe that each game is played by different people, thus if we compute the Fréchet Mean or Medoid of all the games (as done in Section 4.1), the set of obtained graphs constitutes a naturally unlabelled population of networks. We then applied the AAC algorithm to compute the Fréchet Mean, introduced in Section 3.2. On the other hand, due to the low number of games, a Fréchet Medoid would be hardly significant. In this framework we therefore apply our implementation on two levels, since the final result is a mean of means, or a mean of medoids. For coherence, we apply the same weights of the Euclidiscrete Distance for both levels: $\omega_1 = 0.1$, $\omega_2 = 0.9$. A low value of ω_1 allows indeed to obtain a resulting network with more positive-valued edges, which is advisable for datasets like the one taken into consideration, having rather sparse networks. Aware of the dependence of the AAC algorithm on initial conditions and taking advantage of the low value of g, we calculated the Fréchet Mean starting from each one of the statistics obtained after the first step of computation (namely, the Fréchet Means, or Medoids, of each game), in order to compare all the outcomes, and see the most common.

Figure 10 shows the network structures resulted with more occurrences, both for Fréchet Mean of means and Mean of medoids. Observe that nodes have no labels, because of the unlabelled nature of the analysed dataset. For the mean of medoids, 10 out of 23 results represent the same network structure, similar to the ones obtained from the analysis of a single game. We deduce that there is often a player interacting the most with the others, and one doing the least (the disconnected node). Moreover, the only link not including the most interacting player could be a

link between *Spies*. The behaviour of Fréchet Mean on means is similar, although slightly less consistent for what concerns the frequences.
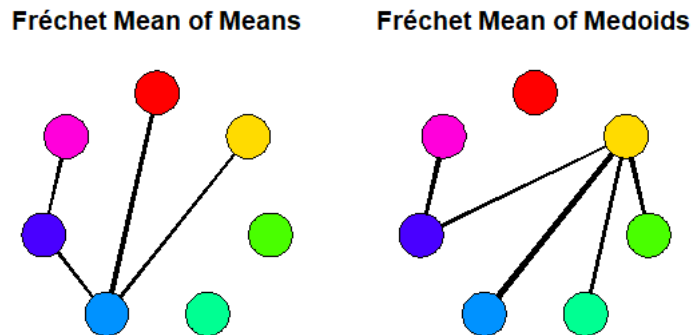

Figure 10: Fréchet Mean of means and Mean of medoids for multiple games, with $\omega_1 = 0.1, \omega_2 = 0.9$.

5   Conclusions

In this work, we first analysed the overall already present network literature, to find existent formulations concerning the calculation of particular statistics, coming from a sample of networks. The target objects were graphs having weighted edge attributes which could have value zero with positive probability. Thus, we fixed an opportune Graph Space both for labelled and unlabelled networks (obtaining respectively X and X/T) and observed the underlying probability distributions of the values, in order to make the proper assumptions for our purpose. We finally introduced our new metric according to the assumptions, and defined the Fréchet Mean and the Fréchet Medoid network in order to see how such metric works. This application in particular involves an *Align All and Compute* algorithm, when made on populations of unlabelled networks. The resulting metric is particularly versatile, due to the free choice of the weights ω_1, ω_2 depending on the more or less conservative selection of the edges.

5.1   Further developments

We may enrich the defined Graph Space, without losing consistency with respect to the assumptions, by adding nodal attributes and making the graphs directed. The only issue which may figure out from such addition regards the possibility of zero-valued *nodes*, but a third likelihood component inherent to the nodes should overcome the problem (look at the expression (1) of the two components). The

rules of swaps between nodes should change accordingly (see Section 2.1). Furthermore, an extension to discrete probability distribution might be considered, even if this may change the nature of the Graph Space. Graphs having edges both with continuous and discrete probability distributions (conditioned to the fact of being positive-valued) would require particular attention, especially in an unlabelled framework (which may generate some contradictions).

The most important development for our application is anyway regression: indeed, the formulation of this work may be interpreted as a null regression model, having only the intercept to compute. A careful study on geodesics [Calissano et al., 2020] and network structures should be carried out for this scope.

References

P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering. A multivariate tobit analysis of highway accident-injury-severity rates. *Accident Analysis & Prevention*, 45:110–119, 2012.

C. Bai, S. Kumar, J. Leskovec, M. Metzger, J. Nunamaker, and V. Subrahmanian. Predicting the visual focus of attention in multi-person discussion videos. In *IJCAI 2019*. International Joint Conferences on Artificial Intelligence, 2019.

I. I. Bojinov, N. S. Pillai, and D. B. Rubin. Diagnosing missing always at random in multivariate data. *Biometrika*, 107(1):246–253, 2020.

C. T. Butts. network: a package for managing relational data in r. *Journal of statistical software*, 24(1):1–36, 2008.

A. Calissano, A. Feragen, and S. Vantini. Populations of unlabelled networks: Graph space geometry and geodesic principal components. *MOX Report*, 2020.

B. Dai, S. Ding, and G. Wahba. Multivariate bernoulli distribution. *Bernoulli*, 19 (4):1465–1483, 2013.

S. P. Dwivedi and R. S. Singh. Error-tolerant approximate graph matching utilizing node centrality information. *Pattern Recognition Letters*, 133:313–319, 2020.

M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310, 1948.

G. Geenens. An essay on copula modelling for discrete random vectors; or how to pour new wine into old bottles. *arXiv preprint arXiv:1901.08741*, 2019.

M. Gori, M. Maggini, and L. Sarti. Exact and approximate graph matching using random walks. *IEEE transactions on pattern analysis and machine intelligence*, 27(7):1100–1111, 2005.

R. A. Johnson, D. W. Wichern, et al. *Applied multivariate statistical analysis*, volume 6. Pearson London, UK:, 2014.

S. Kumar, C. Bai, V. Subrahmanian, and J. Leskovec. Deception detection in group video conversations using dynamic interaction networks. In *ICWSM 2021*. International AAAI Conference on Web and Social Media, 2021.

K. Riesen and H. Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision computing*, 27(7):950–959, 2009.

J. van der Pol. Introduction to network modeling using exponential random graph models (ergm): theory and an application using r-project. *Computational Economics*, 54(3):845–875, 2019.

H. Wang and J. Marron. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.

W. J. Wright, K. M. Irvine, J. M. Warren, and J. K. Barnett. Statistical design and analysis for plant cover studies with multiple sources of observation errors. *Methods in Ecology and Evolution*, 8(12):1832–1841, 2017.

K. K. Yau, A. H. Lee, and A. S. Ng. Theory & methods: A zero-augmented gamma mixed model for longitudinal data with many zeros. *Australian & New Zealand Journal of Statistics*, 44(2):177–183, 2002.

A Appendix

A.1 Calculations of the Multivariate Bernoulli Maximum Likelihood Estimators

Let us write the expression (3) of the log-likelihood $l(p_{11}, p_{10}, p_{01}|x_{1:n})$, keeping in mind that $p_{00} = 1 - p_{11} - p_{01} - p_{10}$:

$$l(p_{11}, p_{10}, p_{01}|x_{1:n}) = \sum_{x=1}^{n} x_{i1}x_{i2} \log \frac{p_{11}(1-p_{11}-p_{01}-p_{10})}{p_{10}p_{01}} +$$

$$+\sum_{x=1}^{n} x_{i1} \log \frac{p_{10}}{1-p_{11}-p_{01}-p_{10}} + \sum_{x=1}^{n} x_{i2} \log \frac{p_{01}}{1-p_{11}-p_{01}-p_{10}} + n\log(1-p_{11}-p_{01}-p_{10}).$$

$$(18)$$

To obtain $\hat{p}_{11}$, we derive (18) with respect to p_{11}, and set the derivative equal to zero:

$$\frac{\partial l}{\partial p_{11}} = \frac{(1-2p_{11}-p_{01}-p_{10})\sum_{i=1}^{n} x_{i1}x_{i2}}{p_{11}(1-p_{11}-p_{01}-p_{10})} + \frac{\sum_{i=1}^{n}(x_{i1}+x_{12})-n}{1-p_{11}-p_{01}-p_{10}} = 0;$$

$$(1-2p_{11}-p_{01}-p_{10})\sum_{i=1}^{n} x_{i1}x_{i2} + p_{11}\sum_{i=1}^{n}(x_{i1}+x_{12}) = np_{11};$$

$$p_{11}\left[2\sum_{i=1}^{n} x_{i1}x_{i2} - \sum_{i=1}^{n}(x_{i1}+x_{12}) + n\right] = (1-p_{01}-p_{10})\sum_{i=1}^{n} x_{i1}x_{i2};$$

$$\hat{p}_{11} = \frac{(1-p_{10}-p_{01})\sum_{i=1}^{n} x_{i1}x_{i2}}{n - \sum_{i=1}^{n}(x_{i1}+x_{i2}) + 2\sum_{i=1}^{n} x_{i1}x_{i2}}. \tag{19}$$

Similarly for $\hat{p}_{10}$, we derive the log-likelihood with respect to p_{10}:

$$\frac{\partial l}{\partial p_{10}} = \frac{(-1+p_{11}+p_{01})\sum_{i=1}^{n} x_{i1}x_{i2}}{p_{10}(1-p_{11}-p_{01}-p_{10})} + \frac{\sum_{i=1}^{n} x_{i1}(1-p_{11}-p_{01})}{p_{10}(1-p_{11}-p_{01}-p_{10})} + \frac{\sum_{i=1}^{n} x_{i2}-n}{1-p_{11}-p_{01}-p_{10}} = 0;$$

$$\sum_{i=1}^{n} x_{i1}x_{i2}(-1+p_{11}+p_{01}) + \sum_{i=1}^{n} x_{i1}(1-p_{11}-p_{01}) = \left(n - \sum_{i=1}^{n} x_{i2}\right)p_{10};$$

$$\hat{p}_{10} = \frac{(1-p_{11}-p_{01})\sum_{i=1}^{n} x_{i1}(1-x_{i2})}{n - \sum_{i=1}^{n} x_{i2}}. \tag{20}$$

For symmetry reasons, we also get:

$$\hat{p}_{01} = \frac{(1-p_{11}-p_{10})\sum_{i=1}^{n}(1-x_{i1})x_{i2}}{n - \sum_{i=1}^{n} x_{i1}}. \tag{21}$$

The expressions (19), (20) and (21) contain all the parameters p_{11}, p_{10} and p_{01}: to find the estimators, we put these expressions in a linear system with their respective estimators $\hat{p}_{11}, \hat{p}_{10}$ and $\hat{p}_{01}$:

$$\begin{cases} \hat{p}_{11} = \dfrac{(1-\hat{p}_{10}-\hat{p}_{01})\sum_{i=1}^{n} x_{i1}x_{i2}}{n - \sum_{i=1}^{n}(x_{i1}+x_{i2}) + 2\sum_{i=1}^{n} x_{i1}x_{i2}}; \\[3ex] \hat{p}_{10} = \dfrac{(1-\hat{p}_{11}-\hat{p}_{01})\sum_{i=1}^{n} x_{i1}(1-x_{i2})}{n - \sum_{i=1}^{n} x_{i2}}; \\[3ex] \hat{p}_{01} = \dfrac{(1-\hat{p}_{11}-\hat{p}_{10})\sum_{i=1}^{n}(1-x_{i1})x_{i2}}{n - \sum_{i=1}^{n} x_{i1}}. \end{cases}$$

It can be easily checked that the solution of such system is:

$$\hat{p}_{11} = \frac{\sum_{i=1}^{n} x_{i1}x_{i2}}{n}; \quad \hat{p}_{10} = \frac{\sum_{i=1}^{n} x_{i1}(1-x_{i2})}{n}; \quad \hat{p}_{01} = \frac{\sum_{i=1}^{n}(1-x_{i1})x_{i2}}{n}. \tag{22}$$

And we conclude.

A.2 The Attributed Euclidiscrete Distance

Definition A.1. The Attributed Euclidiscrete Distance (AEDD) of a network $\boldsymbol{x} \in X$ with respect to a set of networks $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_n\}$ is defined as:

$$AEDD(\boldsymbol{x}) = \frac{\omega_1}{m} EHD(\boldsymbol{x}) + \frac{\omega_2}{n} \sum_{i=1}^{n} \sum_{k:x_{ik}^B x_k^B = 1} \frac{(x_{ik} - x_k)^2}{c_i R_k^2}. \tag{23}$$

Where c_i denotes the number of common (present) edges between $\boldsymbol{x}$ and $\boldsymbol{x}_i$, namely an edge k is in common if and only if $x_k^B = 1 \wedge x_{ik}^B = 1$.

This definition is a different reformulation of the one explained by [Wang and Marron, 2007]. However, compared to the Weighted Euclidiscrete Distance, this one presents several drawbacks: its meaning is lesser interpretable, its slightly more computationally expensive, it has an ambiguous behaviour for data with $c_i = 0$ (there is no division by 0, considering the fact that there would be no terms to sum), it is quite pessimistic in order to estimate the Fréchet Mean, and - last but not least - it does not allow any assumption of independence among edges. Moreover, since it does not use all the information, it is neither a metric nor a pseudometric.
Indeed, a counterexample in which the triangular inequality does not hold could be easily provided. Regarding Definition A.1, we have that the expression for the underlying dissimilarity measure is:

$$d_{AEDD}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\omega_1}{m} \sum_{k=1}^{m} (x_k^B - y_k^B)^2 + \frac{\omega_2}{c} \sum_{k:x_k^B y_k^B = 1} (x_k - y_k)^2 \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X, \tag{24}$$

Where $c \leq m$ is the number of common positive-valued edges between the two networks. Notice the use of the term *dissimilarity*, due to the fact that such measure is not proven to be a metric. As counterexample, consider the following three networks with $v = 4$ nodes, along with the just defined dissimilarities between them:

Edge	(1;2)	(1;3)	(1;4)	(2;3)	(2;4)	(3;4)
$\boldsymbol{x}_1$	1	1	1	0	0	0
$\boldsymbol{x}_2$	0	0	0	1	1	1
$\boldsymbol{x}_3$	0.5	0.5	0.5	0	0	0

$$d_{AEDD}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{6\omega_1}{6} + 0\omega_2 = \omega_1;$$

$$d_{AEDD}(\boldsymbol{x}_2, \boldsymbol{x}_3) = \frac{6\omega_1}{6} + 0\omega_2 = \omega_1;$$

$$d_{AEDD}(\boldsymbol{x}_1, \boldsymbol{x}_3) = 0\omega_1 + \frac{0.75\omega_2}{3} = 0.25\omega_2.$$

If we choose the positive weights such that $\omega_1 < 0.125\omega_2$, we obtain that:

$$d_{AEDD}(\boldsymbol{x}_1, \boldsymbol{x}_2) + d_{AEDD}(\boldsymbol{x}_2, \boldsymbol{x}_3) = 2\omega_1 < 0.25\omega_2 = d_{AEDD}(\boldsymbol{x}_1, \boldsymbol{x}_3).$$

Namely, the triangular inequality does not hold for d_{AEDD}, which hence is not a metric for certain values of ω_1, ω_2.

We conclude that an idea of Attributed Euclidiscrete Distance should be perpetuated in order to gain interpretative completeness about the subject, but it would require a greater effort in terms of overall costs.