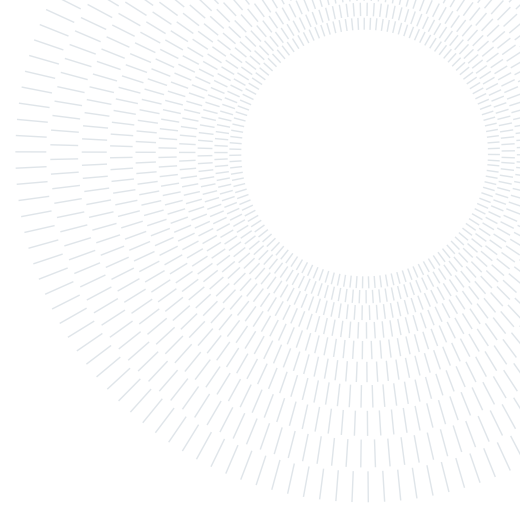




POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE



Two-population test for unlabelled networks

TESI DI LAUREA MAGISTRALE IN
MATHEMATICAL ENGINEERING - INGEGNERIA MATEMATICA

Daniele Mulè, 945667

Advisor:
Simone Vantini

Co-advisors:
Anna Calissano
Aymeric Stamm

Academic year:
2021-2022

Abstract: In this paper, an approach to hypothesis testing for two populations of unlabelled networks is proposed. The Graph Space framework is integrated in order to handle networks that differ in the number of nodes or in node labelling. Permutation tests are considered, along with flexible choices on inter-point test statistics based on intrinsic metrics. Then, distances between networks are computed relying on approximate matching algorithms, in order to find optimal graphs alignments. Simulation studies show effectiveness and superior performance of the method with respect to alternative approaches. Finally, usefulness of the method is shown with an application to football players' passing networks. The proposed approach is available in the R package `nevada`.

Key-words: Network-valued data, Unlabelled networks, Graph space, Hypothesis testing, Permutation test, Football players passing network

1. Introduction

Networks are data structures of particular interest in the framework of Object Oriented Data Analysis. Networks are strongly non-Euclidean data object and, for this reason, making inference on them is a challenging task (Wang and Marron, 2007).

It is even more demanding if the problem is related to study differences in distributions between two populations of networks.

As reported by Lovato et al. (2020), the traditional approach reformulates the problem in a multivariate data analysis, considering a vector of graph summary indices for each network-valued data in the samples. This method flatten the network in a set of features that may not fully represent the graph structure in its complexity. Chen et al. (2020) propose a spectral-based hypothesis testing in populations of networks. This approach considers a test statistic based on the trace of the third order for an adjacency matrix-based representation of the networks. This approach is available only for binary or weighted undirected networks. Ghoshdastidar et al. (2020) propose a minimax testing approach for two-population of inhomogeneous random networks. Another method, based on permutation testing, is proposed by Lovato et al. (2020). It considers test statistics based on inter-point distances and it works for any type of network structure. These cited approaches work for labelled networks, i.e., graphs that share the same exact nodes.

However, in the real world, there exist networks related to the same phenomenon that are not defined on the same set of nodes. Indeed, they may present different number of nodes or inconsistent node labels. These graphs are called unlabelled networks. Making inference on them is not trivial (Calissano et al., 2020) and this also applies to tests for populations of unlabelled networks.

In this work, we introduce a general method, in which we consider a two-sample hypothesis testing approach between two populations of unlabelled networks. Our method is valid for any possible type or structure of networks.

The novelty of our approach is based on the integration of the Graph Space framework (Calissano et al., 2020) in network-valued hypothesis testing, in order to properly handle unlabelled networks. Networks in Graph Space are embedded as equivalence classes, in which two graphs are equivalent if they differ only for a permutation of the nodes labels. A new definition of distance is given in such a space.

As in Lovato et al. (2020), we rely on permutation tests with test statistics based on inter-point distances between networks. We consider intrinsic metrics in the Graph Space and, for this reason, we refer to our method as Intrinsic Approach.

The paper is organized as follows. In Section 2, we present the statistical framework, in order to explain tools used to develop our method. In particular, we give basic definition of network-valued data and we focus on the Graph Space and the permutation test. In Section 3, we introduce our method, i.e., the Intrinsic Approach. Simulation studies are reported in Section 4, aiming to show the effectiveness of our method with respect to alternative approaches. Section 5 consider the application to a real life example, related to football players' passing networks in UEFA EURO 2020 and LaLiga competitions. We report final conclusions and further developments in Section 6. The Intrinsic Approach is implemented in the R (R Core Team, 2020) package `nevada`.

2. Statistical framework

In this section, we formalize the statistical framework. We start from initial definitions of network theory and then we concentrate on the Graph Space and the permutation test.

2.1. Networks and adjacency representation

A network $G = (V, E)$ is a data structure defined by a set V of nodes and a set E of edges that specifies the connections among nodes.

The adjacency representation A of a network with n nodes is an $n \times n$ matrix whose entries a_{ij} , $i, j \in \{1, \dots, n\}$ describe the presence or the absence (for unweighted networks) or the weight (for weighted networks) of the edge (i, j) . Self-loops are indicated along the main diagonal of A ; if a node does not have a self-loop, than its entry on the diagonal is equal to 0. If the network is undirected, than A is symmetric.

2.2. Graph Space

Following Calissano et al. (2020), we report here definitions and results about the Graph Space.

A network with n nodes is fully represented by its $n \times n$ adjacency matrix.

Since we consider two populations of unlabelled networks, they may present unequal number of nodes, say less than or equal to n . As done in Calissano et al. (2020), to represent networks with fixed-size adjacency matrices, we add a number of null nodes without connections to them such that they reach n nodes.

Then, networks are represented in the space $X = \mathbb{R}^{n^2}$ of flattened adjacency matrices.

Since unlabelled networks may have different labels, it is not trivial to spot similarities or differences between them. As cited in Calissano et al. (2020), this problem is softened by implicit or explicit matching of nodes between two networks. Looking for a potential matching among nodes means finding an optimal alignment between networks. In the space X , this is equivalent to find the optimal permutation of networks nodes. Node permutations, that can be grouped in a set T , are possible in X via permutation matrices.

The Graph Space $X_T := X/T$ is the quotient space of the flattened adjacency matrices space in which two networks are equivalent if they differ only by a node relabelling. In other words, $[x] \in X_T$ includes all the networks in X that can be obtained from $x \in X$ by permutations of the nodes labels. An action of T on X is expressed by the following operation:

$$\begin{aligned} T \times X &\rightarrow X, \\ (T, x) &\mapsto Tx \end{aligned}$$

A metric d_X in X implies a metric d_{X_T} in X_T :

$$d_{X_T}([x_1], [x_2]) := \min_{t \in T} d_X(tx_1, x_2) \quad (2.1)$$

Some properties related to the Graph Space can be proven: (i) the Graph Space is a metric space, but not a manifold; (ii) the Graph Space is a geodesic space in which geodesics that connect two points in X_T are not

necessarily unique; (iii) the curvature of a Graph Space is unbounded from above. Proofs of these statements can be found in Calissano et al. (2020).

In the Graph Space it is possible to introduce the concept of mean through the Fréchet mean. Given a sample of m networks $\{[x_1], \dots, [x_m]\} \in X_T$, its Fréchet mean $[\bar{x}] \in X_T$ is defined as:

$$[\bar{x}] := \arg \min_{[x] \in X_T} \sum_{i=1}^m d_{X_T}([x], [x_i])^2$$

Due to the Graph Space properties listed above, it follows that the Fréchet mean is not generally unique in X_T .

2.3. Permutation test

We consider here the same hypothesis testing method used in Lovato et al. (2020).

The permutation test is an exact and consistent non-parametric hypothesis testing procedure. This strategy is computationally heavier with respect to parametric tests but, on the contrary, it does not require tight distributional assumptions or large sample sizes.

Under the null hypothesis of equality in distributions, data observations in the two samples are exchangeable. Then, for each permutation of the observations between the two samples, we can store the corresponding value of the chosen test statistic in the so-called permutation distribution. The distribution of the statistic under the null hypothesis can be approximated by its permutation distribution.

Since considering the overall number of permutations M for the permutation distribution is computationally unfeasible, only a random selection $B < M$ of permutations is evaluated.

Finally, the p-value is evaluated considering definitions by Dwass (1957) and Phipson and Smyth (2010). See Appendix A for more details.

3. Intrinsic Approach

In this section, we introduce the Intrinsic Approach to perform hypothesis testing for unlabelled networks.

3.1. Method formalization

Given two sample of unlabelled networks in the Graph Space with sample sizes m_1 and m_2 , we want to test differences between their distributions.

In particular, consider a sample of independent and identically distributed network random variables $G_{11}, \dots, G_{1m_1} \sim F_1$ and a sample of independent and identically distributed network random variables $G_{21}, \dots, G_{2m_2} \sim F_2$; then, the test can be expressed as:

$$H_0 : F_1 = F_2 \text{ against } H_1 : F_1 \neq F_2 .$$

As Lovato et al. (2020), we use the permutation test described in Section 2.3.

The most natural choice of test statistic would be a distance between the sample Fréchet means. However, as explained in Section 2.2, the sample Fréchet mean may be not unique in the Graph Space, hence a Fréchet mean-based test statistic would lead to ambiguous results. Moreover, differences in distributions may arise in higher-order moments.

Following Lovato et al. (2020), we adopt in our method the IP-StudentFisher statistic, an inter-point statistic based on distances between the pooled observations.

Let $T_{IP-Student}$ be the Student statistic and let $T_{IP-Fisher}$ be the Fisher statistic (Lovato et al., 2020) defined as:

$$T_{IP-Student} := \frac{\frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} d(G_{1i}, G_{2j})^2 - (\hat{\sigma}_1^2 + \hat{\sigma}_2^2)}{\frac{\hat{\sigma}_1^2}{m_1} + \frac{\hat{\sigma}_2^2}{m_2}} \text{ and} \quad (3.1)$$

$$T_{IP-Fisher} := \max \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}, \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right),$$

with $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ defined as:

$$\hat{\sigma}_1^2 := \frac{1}{m_1(m_1 - 1)} \sum_{i=1}^{m_1} \sum_{j>i}^{m_1} d(G_{1i}, G_{1j})^2 \text{ and}$$

$$\hat{\sigma}_2^2 := \frac{1}{m_2(m_2 - 1)} \sum_{i=1}^{m_2} \sum_{j>i}^{m_2} d(G_{2i}, G_{2j})^2 .$$

Then, the IP-StudentFisher statistic is defined as the Non-Parametric Combination (Brombin and Salmaso, 2009; Pesarin and Salmaso, 2010) of $T_{IP-Student}$ and $T_{IP-Fisher}$, via the Tippett’s combining function (Tippett, 1937). See Appendix B for more details.

The IP-StudentFisher statistic ensures that the null hypothesis is rejected if even one of the combined statistics rejects it.

This statistic is proven to be consistent under the alternative hypothesis of unequal means and variances by Lovato et al. (2020).

Remark 3.1. *Other inter-point statistics are available in the literature. Lovato et al. (2020), for example, consider the energy statistic (Székely and Rizzo, 2013), a density-based statistic (Biswas and Ghosh, 2014) and the generalized edge-count statistic Chen and Friedman (2017). The choice of focusing on the IP-StudentFisher is motivated by simulations performed in Section 4.*

In (2.1), we explain that a distance in the Graph Space X_T is defined by a distance in the original space X of flattened adjacency matrices. In particular, whenever we need to compute in X_T an intrinsic distance between two networks, we are computing a distance between classes of equivalence. In X , it is the same as computing the distance between the first network and the second one subject to a node relabelling that minimizes this distance. In other words, to compute an intrinsic distance in X_T it is required to find the optimal alignment between two networks with respect to a certain distance in X .

Since our method is based on intrinsic distances in X_T that require alignments between networks, we refer to our method as the Intrinsic Approach.

3.2. Algorithmic implementation

The first step of our procedure require to add null nodes to networks in the two populations that present fewer nodes, as explained in Section 2.2. In this way, we can work with fixed-size adjacency matrices.

Then, in order to evaluate the test statistic, we compute all the inter-point distances between networks in X_T as in (2.1), looking for optimal nodes matching. The exact matching of nodes is practically unfeasible, since it scales rapidly with the number of nodes. Hence, we rely on approximate matching algorithms.

For example, (Qiao and Sussman, 2021) propose the graph matching algorithm with indefinite relaxation of the objective function, counting on the Frank-Wolfe methodology. The barycentre matrix can be selected as initialization of the permutation matrix, as suggested by Qiao and Sussman (2021). This algorithm is based on the Frobenius distance between networks; it is the most natural choice among distances, since it corresponds with the Euclidean distance in X . Given two networks $x, y \in X$ with n nodes, represented respectively with adjacency matrices representations A and B , the Frobenius distance d_X^{FR} is defined as follows:

$$d_X^{FR}(x, y) = \left(\sum_{i=1}^n \sum_{j=1}^n (A_{ij} - B_{ij})^2 \right)^{1/2}$$

Remark 3.2. *Our method may also consider other distances, e.g., Manhattan, Canberra, Jaccard or DeltaCon (Koutra et al., 2016) distances cited in Tantardini et al. (2019). Consequently, it is recommended to use an approximate matching algorithm that allows the use of these metrics.*

After distances computation, the permutation test is performed as described in Section 2.3. We recall that, due to the computational burden, the permutation test consider only B out of M permutations.

Finally, the Dwass-Phipson-Smyth p-value is computed.

We report steps of the Intrinsic Approach in Algorithm 1.

Algorithm 1 Intrinsic Approach

Input: Two samples of unlabelled networks, number of samples permutations B , a distance d_X , a test statistic T

Output: P-value of the test

Add null nodes to networks in the two samples in order to reach the maximum number of nodes for each network;

Compute all the distances between networks in the two samples as in (2.1), via a proper matching algorithm that allows the use of d_X ;

Compute and store the observed value of the statistic T ;

for $b = 1, \dots, B$ **do**

 Randomly permute observations between the two samples;

 Compute and store the value of the statistic T of the permuted samples;

end for

Evaluate the test Dwass-Phipson-Smyth p-value using the stored values of T .

4. Simulation studies

In this section, we perform simulations in order to show: (i) the effectiveness of the Intrinsic Approach, (ii) the proper test statistic to choose when the Frobenius distance is selected and (iii) the reliability of our method with respect to the alternative one, that we call Extrinsic Approach.

For the Intrinsic Approach we pick the Frobenius distance, using the indefinite relaxation approximate matching algorithm as described in Section 3.2. In particular, we run this algorithm choosing the barycentre matrix as initialization of the permutation matrix and we set the maximum number of iterations performed by the Frank-Wolfe algorithm to 20.

The Extrinsic Approach share the same statistical framework related to the Graph Space proposed in Section 3 for the Intrinsic Approach. The difference lies in the distances used to perform the test. Indeed, the competing method employs extrinsic distances in X_T based on maps from X_T to a lower dimensional space that are invariant to node relabelling. For this reason, the Extrinsic Approach does not require any alignment between networks.

In simulations we choose the spectral distance for the Extrinsic Approach. It is an Euclidean distance between networks spectra, that are invariant to permutations on the nodes labels (Tantardini et al., 2019; Lovato et al., 2020).

Remark 4.1. *Tantardini et al. (2019) report other extrinsic distances, e.g., Graphlet-based (Pržulj et al., 2004; Sarajlić et al., 2016; Aparício et al., 2015), NetLSD (Tsitsulin et al., 2018) and Portrait divergence (Bagrow and Bollt, 2019) distances. The choice of a proper distance for the Extrinsic Approach is not trivial and it is one of the drawbacks of this method.*

4.1. Intrinsic Approach effectiveness and test statistics discrimination

In the first simulation, we compare different test statistics. We concentrate on the IP-StudentFisher (Lovato et al., 2020), the energy (Székely and Rizzo, 2013), a density-based (Biswas and Ghosh, 2014) and the generalized edge-count statistic Chen and Friedman (2017).

We aim to show that the Intrinsic Approach is a valid method that is able to recognize differences in distributions of networks, if any. Moreover, we want to illustrate the performance of different statistics if we consider our method with the Frobenius distance.

We replicate the same simulation study led by Lovato et al. (2020). In particular, we generate two samples of $m_1 = m_2 = 20$ unlabelled networks with $n_1 = n_2 = 25$ nodes. We sample undirected networks whose edge weights follow an independent and identically distributed binomial distribution. The generation mechanism is invariant to node relabelling, since edge weights are independent and identically distributed; however, this does not represent a problem for our setting, since we treat these graphs as unlabelled networks in the Graph Space, i.e., as if we do not have any a priori knowledge of correspondence in nodes labels. Let μ_1, σ_1^2 and μ_2, σ_2^2 be the mean and the variance of the distribution of the first and the second sample, respectively. Three scenarios are considered, each one defined by a specific difference in distributions:

- Scenario 1: Location-only differences ($\Delta = |\mu_2 - \mu_1| = 0, 0.125, 0.25, 0.375, 0.5$; $\sigma_1^2 = \sigma_2^2 = 2.5$);
- Scenario 2: Scale-only differences ($\Delta = \sigma_2^2/\sigma_1^2 = 1, 1.05, 1.1, 1.15, 1.2$; $\mu_1 = \mu_2 = 60$);

- Scenario 3: Location and scale differences ($\Delta = |\mu_2 - \mu_1| = 0, 0.1, 0.2, 0.3, 0.4$; $\sigma_2^2/\sigma_1^2 = 1, 1.05, 1.1, 1.15, 1.2$).

In Appendix C, distribution parameters related to the sample generation mechanism in the three scenarios are specified.

The generalized edge-count statistic is computed with a minimal spanning tree of density 5.

We compute for each case in every scenario the Monte-Carlo estimate of the power of the permutation tests, with 100 trials and a significance level $\alpha = 0.05$. Permutation tests consider a number $B = 1000$ of permutations and Dwass-Phipson-Smyth p-value is evaluated.

Simulations are performed considering the Intrinsic Approach with the Frobenius distance and, for the sake of completeness, the Extrinsic Approach with the spectral distance.

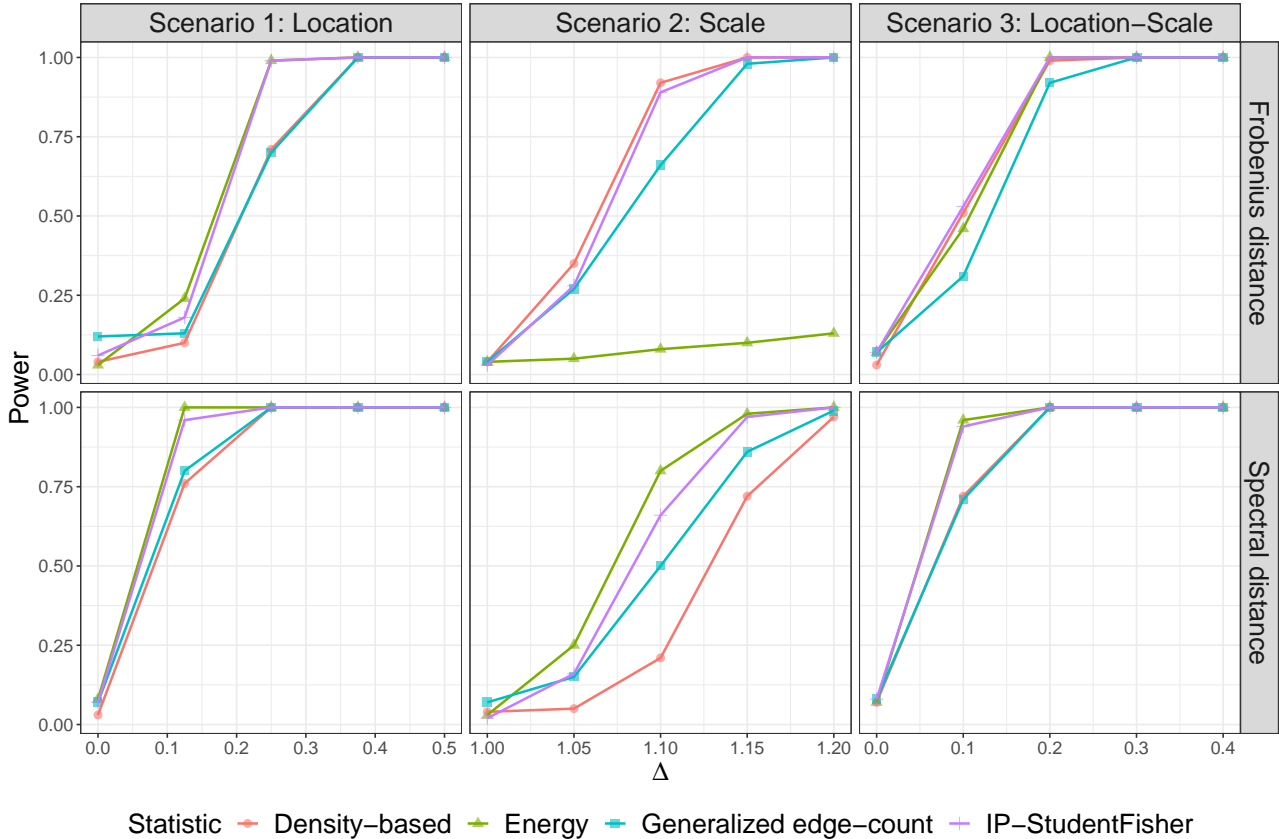


Figure 1: Monte-Carlo estimates of the power of test in different scenario (columns) using different approaches (rows), with various statistics.

In Figure 1, power estimates are reported for each scenario, along with the null rejection rates when $\mu_2 - \mu_1 = 0$ or $\sigma_2^2/\sigma_1^2 = 1$.

Focusing on the Intrinsic Approach, we can notice that our method gives expected results. Both in Scenario 1 and Scenario 3, the power grows coherently with the increasing value of Δ ; this is also valid for Scenario 2, except for the energy statistic that has low power. The IP-StudentFisher statistic performs well in each situation, and it is the best statistic in detecting differences in mean and simultaneously in mean and variance; the density-based statistic has similar performance and it is slightly more powerful than the IP-StudentFisher in finding scale-only differences. The generalized edge-count has lower power in each scenario, meanwhile the energy statistic reaches good performance in location-only and location-scale differences as the IP-StudentFisher but, as reported before, it has very low power in detecting differences in variance.

For the Extrinsic Approach, we notice a loss in power in Scenario 2 along with higher power in the other ones with respect to the Intrinsic Approach. Probably, the latter observation is due to the fact that spectral distance is too sensitive even to small changes in the graph structure, as described in Tantardini et al. (2019). For this method, the greatest power is reached by the energy and the IP-StudentFisher statistics.

4.2. Limitation of the Extrinsic Approach

In the second simulation, we compare the Intrinsic Approach with the Extrinsic Approach. Especially, we show that our method is able to detect actual differences in distribution of the two samples in situations in which the alternative one fails. More precisely, the Intrinsic Approach is not critically restricted to the choice of a specific metric, since intrinsic distances in the Graph Space preserve the original structure of the graph. Instead, the choice of a distance for the Extrinsic Approach is crucial, since it maps the network in a lower dimensional space with resulting loss of information; as a matter of fact, the chosen embedding may not capture important features of the graph structure.

Taking for example the Extrinsic approach with spectral distance, we illustrate how this method is not able to capture differences in networks that share the same spectrum distribution, but with two different sets of eigenvectors, one for the first sample and the other for the second sample.

We generate two samples of undirected unlabelled networks with $n = 8$ nodes, starting from the spectral decomposition of the adjacency matrix, since it is symmetric for an undirected graph. In particular, we fix a set of eigenvectors per sample and we draw independent and identically distributed spectra from a certain multivariate Poisson distribution, for both populations. Then, adjacency matrices are built following the spectral decomposition. Modelling details can be found in Appendix C. We vary the sample sizes, setting $m_1 = m_2 = 4, 8, 16, 32, 64$.

We consider two scenarios:

- Scenario A: false H_0 . For the two samples, eigenvalues are drawn from the same spectrum distribution and eigenvectors are taken from two different fixed orthogonal matrices, one per sample.
- Scenario B: true H_0 . For the two samples, eigenvalues are drawn from the same spectrum distribution and eigenvectors are taken from the same fixed orthogonal matrix.

We compute the Monte-Carlo estimate of the power of the permutation tests, performing 100 trials with a significance level $\alpha = 0.05$. We consider $B = 1000$ permutations for the permutation tests, evaluating the Dwass-Phipson-Smyth p-value.

To make a fair comparison between the Intrinsic Approach with Frobenius distance and the Extrinsic Approach with the spectral distance, we choose the IP-StudentFisher statistic, since it reaches good performance for both methods, as noticed in Section 4.1.

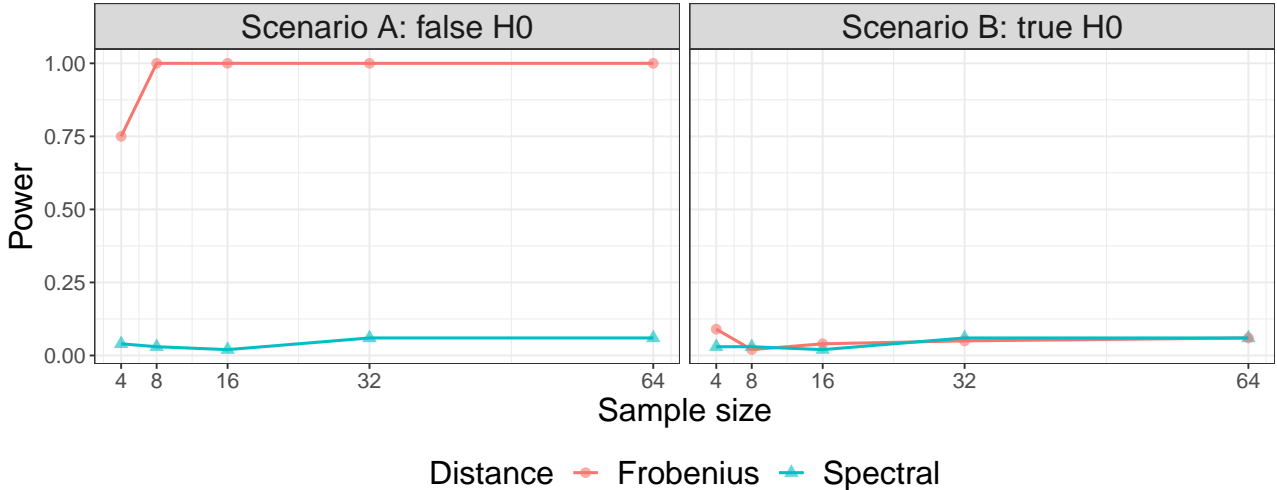


Figure 2: Monte-Carlo estimates of the power of test for Scenario A and the null rejection rate for Scenario B. The test considers the IP-StudentFisher test statistic.

In Figure 2, we report the estimate of the power for Scenario A and the null rejection rate for Scenario B. As expected, the Extrinsic Approach is not able to detect differences between samples under the alternative hypothesis (Scenario A), meanwhile the Intrinsic Approach has very high power, that firmly reaches 1 starting from only 8 networks per sample. Then, both methods succeed in performing low null rejection rate under the null hypothesis (Scenario B).

This confirms that the main limitation of the Extrinsic Approach is the downgrading of datum implied by the choice of an extrinsic distance, whereas the Intrinsic Approach does not significantly suffers from the choice of a particular intrinsic distance.

5. Application to football data

In this section, the Intrinsic Approach is applied on a real example in the field of sports analytics, in order to show the usefulness of our method.

Our aim is to analyse football passing strategies, that can have an important role for a team in the final outcome of a match (see, for example Ievoli et al., 2021). This task can be achieved by the study of the players' passing network, i.e., a weighted graph with nodes representing the players in a team and edges expressing the number of passes between two players in a match (Calissano et al., 2022; Ievoli et al., 2021).

We consider a football database available in StatsBomb Open Data repository (StatsBomb Services Limited, 2022). This database contains a collection of different matches of various competitions and respective in-match events, detailed by a set of variables. In particular, building our code on the scripts provided by Gómez (2020), we organize features of a given match of a specific team to create weighted passing networks with $n = 11$ players, considering edge weights as the number of completed passes between two players from kick-off to first player substitution or red card, if any. This last choice let us focus on passing strategies that consider only the first 11 players on the field, as it is usually done. For simplicity, we consider the undirected version of passing networks. In particular, we focus on the UEFA EURO 2020 and LaLiga competitions.

Football passing networks are perfect examples of unlabelled networks since, even in the same team, there may be different players that take part to games. In this way, the Intrinsic approach can properly handle tests on players' passing networks distributions.

All the next permutation tests follow the Intrinsic Approach and are performed considering the Frobenius distance, the IP-StudentFisher statistic and $B = 1000$ permutations; moreover, we use the indefinite relaxation approximate matching algorithm with Frank-Wolfe methodology (barycentre as initialization matrix, maximum number of iterations = 20). We consider a significance level $\alpha = 0.05$.

In order too visualize results, for each competition, we plot the matrix p of p-values related to pairwise tests between populations of passing networks under consideration. Then, we perform a hierarchical agglomerative clustering (average linkage, dissimilarity matrix = $-\log_2(p)$). Finally, to qualitatively visualize differences in distributions, we compute Fréchet means of players' passing networks via the Align All and Compute algorithm, proposed by Calissano et al. (2020) (tolerance = 0.001, maximum number of iterations = 200).

5.1. UEFA EURO 2020: competing teams

We consider here two-sample tests in which we compare two by two passing strategy distributions between all the 24 competing teams in UEFA EURO 2020.

We assign to each team a passing network for every game it plays.

We report the teams in game and, in brackets, the number of their available matches (i.e., sample size): Turkey, Slovakia, North Macedonia, Russia, Finland, Hungary, Scotland, Poland (3); Germany, Portugal, Croatia, Sweden, France, Austria, Netherlands, Wales (4); Belgium, Ukraine, Switzerland, Czech Republic (5); Spain, Denmark (6); Italy, England (7).

We can notice that teams have low sample sizes; this does not represent a problem, since the Intrinsic Approach adopts permutation tests that allow a small number of observations in samples, as explained in Section 2.3.

In Figure 3, we plot the matrix p of p-values related to tests between pairs of teams. As we can notice, teams can be grouped in 4 classes of passing networks. Inside each cluster, there is no statistical evidence to reject the null hypothesis of equal passing networks distributions. Qualitatively, Fréchet means in Figure 4 give us interesting insights about structures in clusters. Indeed, we can observe that, in broad terms, teams can be divided into 4 categories based on high (red), medium-high (blue), medium-low (yellow) and low (green) number of passes during the match. It is notable to mention that some teams in the yellow cluster, such as Poland and Turkey, also present high p-values in tests with networks in the blue cluster.

For the sake of completeness, we report in Appendix D the same analysis using the Extrinsic Approach.

5.2. LaLiga: Barcelona managers

We now consider LaLiga competition, comparing the passing strategies adopted by different Barcelona managers from October 2004 to May 2021.

We consider the 8 managers in this period (the available matches in brackets), i.e.: Frank Rijkaard (78), Pep Guardiola (136), Tito Vilanova (32), Gerardo Martino (31), Luis Enrique (105), Ernesto Valverde (84), Quique Setién (19) and Ronald Koeman (35).

In Figure 5, we report the p-values matrix p of tests highlighting differences in distribution of the Barcelona managers' passing networks and in Figure 6 we plot their Fréchet means.

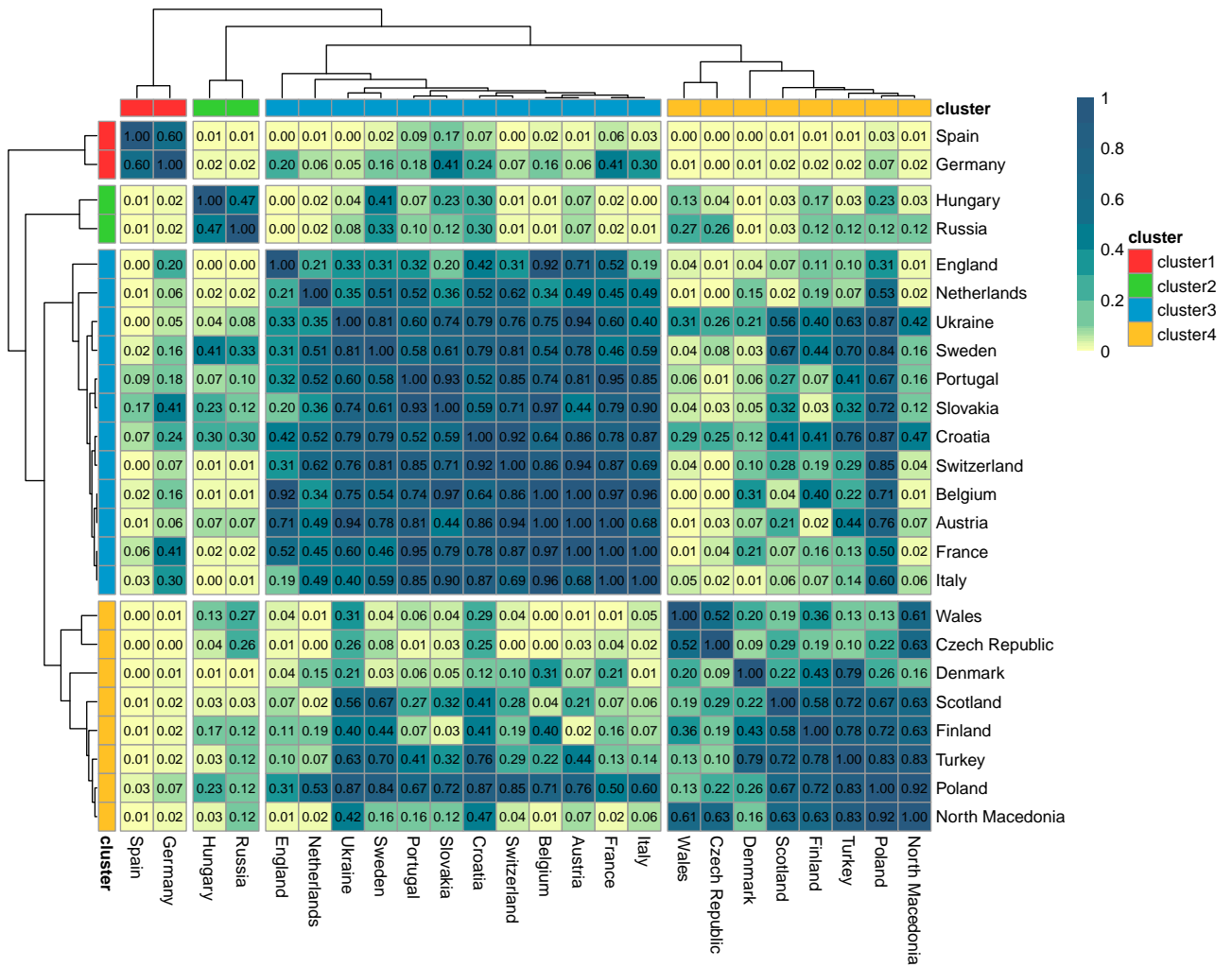


Figure 3: P-values matrix p of permutation tests in which we compare team by team players' passing networks distributions. Dendrograms represent the result of the hierarchical agglomerative clustering (Cophenetic Correlation Coefficient = 0.6936247). We cut dendrograms in order to show 4 clusters. Data source: StatsBomb Open Data.

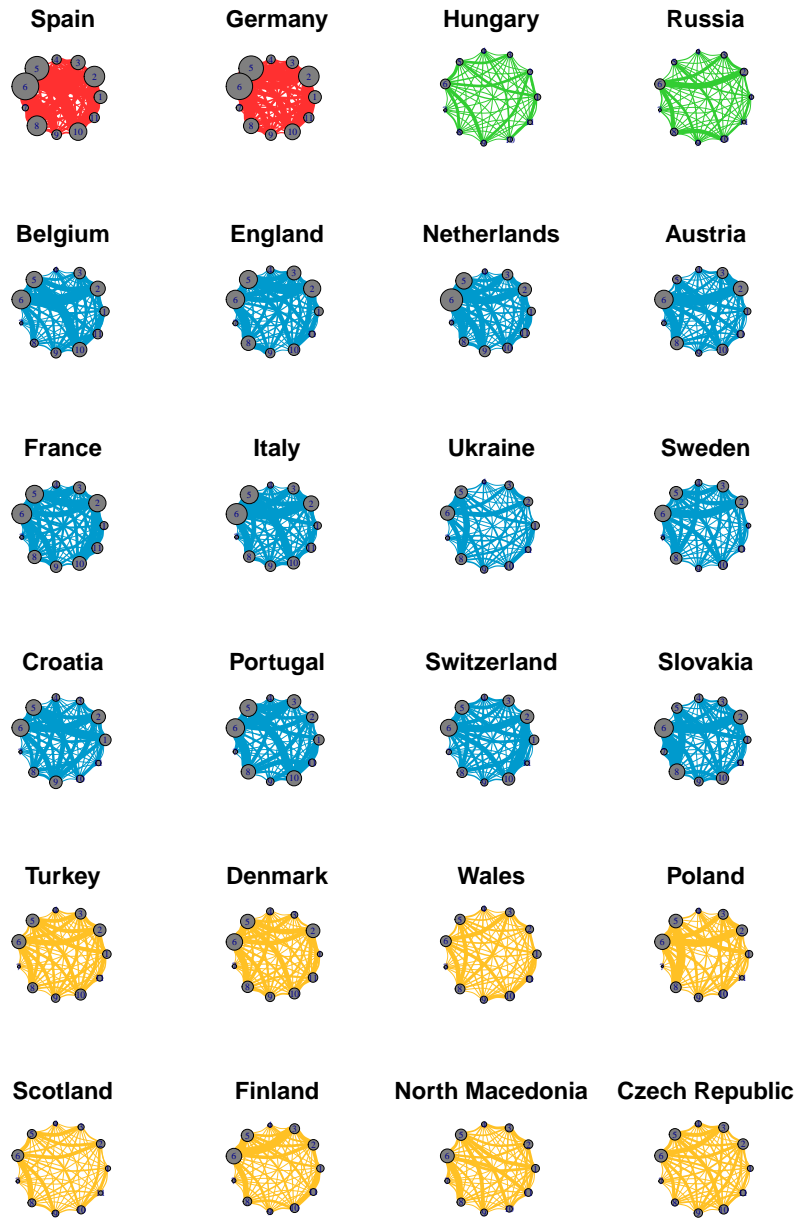


Figure 4: Fréchet means of the players' passing networks for each team. Colours represent the membership to one of the 4 clusters represented in Figure 3, node size refers to node strength (the larger the strength, the larger the size) and edge width to its weight (the wider the edge, the larger its weight). Data source: StatsBomb Open Data.

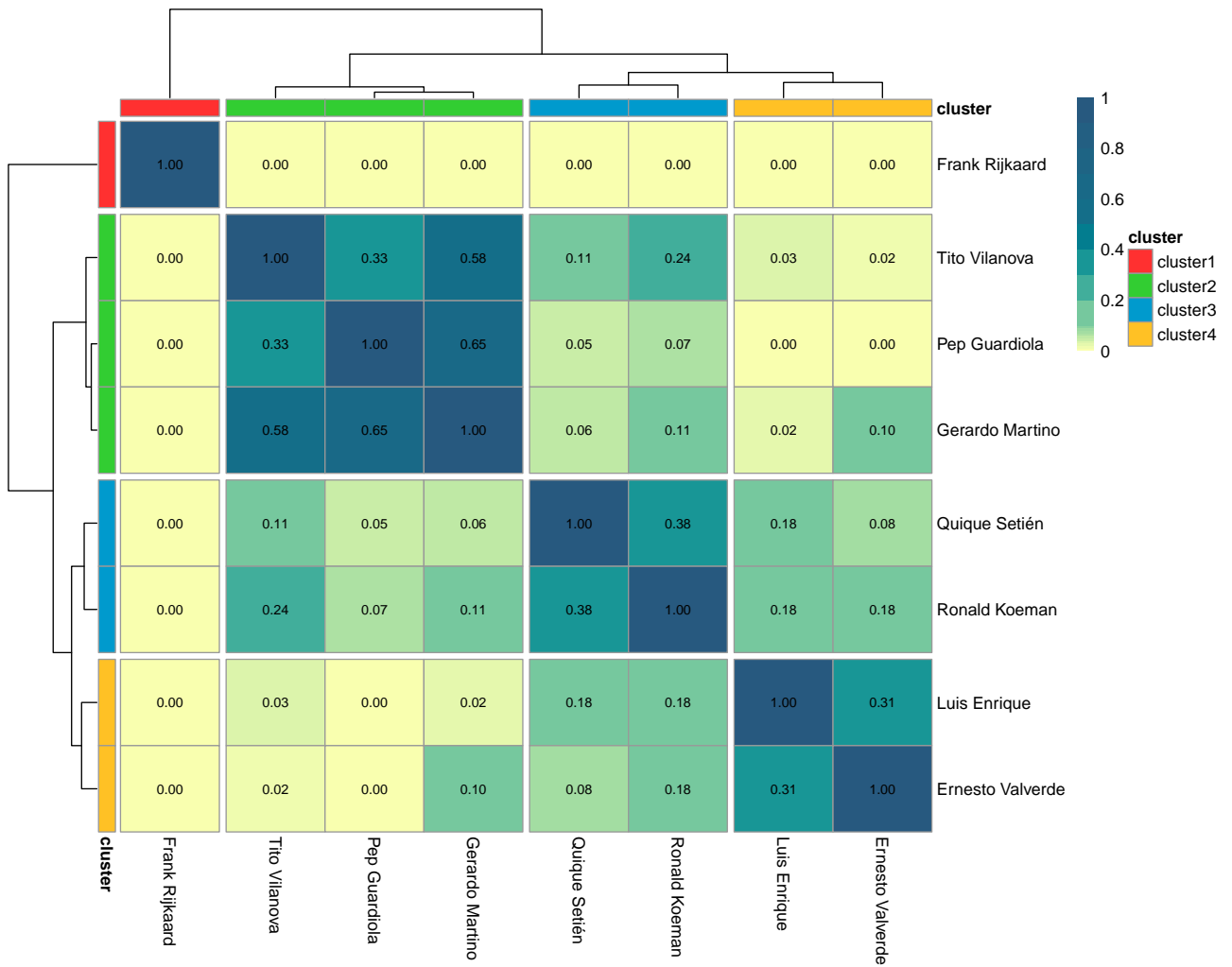


Figure 5: P-values matrix p of permutation tests in which we compare Barcelona managers' passing networks distributions. Dendrograms represent the result of the hierarchical agglomerative clustering (Cophenetic Correlation Coefficient = 0.9001367). We cut dendrograms in order to show 4 clusters. Data source: StatsBomb Open Data.

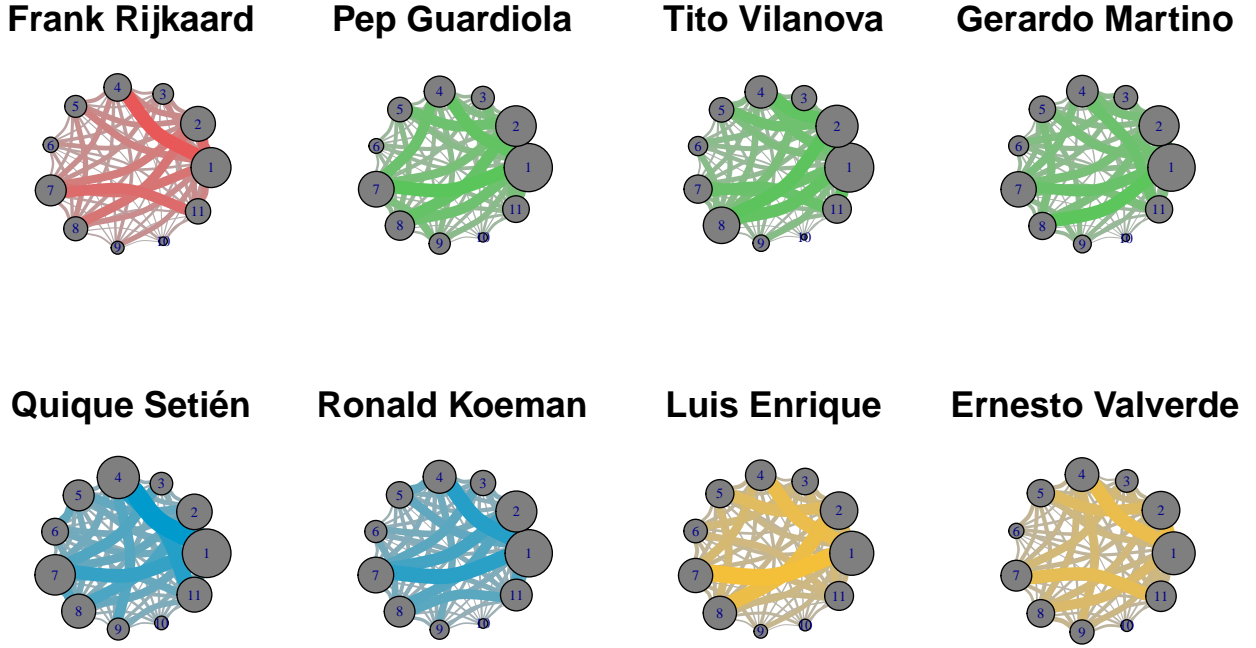


Figure 6: Fréchet means of the players' passing networks for each Barcelona manager. Colours represent the membership to one of the 4 clusters represented in Figure 5, node size refers to node strength (the larger the strength, the larger the size) and edge width and colour saturation to its weight (the wider and more vivid the edge, the larger its weight). Data source: StatsBomb Open Data.

The matrix p points out that Barcelona managers can be divided in 4 groups composed by Rijkaard (red), Vilanova, Guardiola and Martino (green), Setién and Koeman (blue) and Enrique and Valverde (yellow). We can notice that Rijkaard implemented a different passing strategy with respect to all the other managers (small p -values); qualitatively looking at Rijkaard's Fréchet mean, it is probably because large number of passes are restricted only to few players in comparison to other managers'. Observing the green Fréchet means, we can deduce that managers in this cluster adopted a passing strategy that favours large number of passes between all the players. This latter comment is applicable in a more modest way to the other managers, i.e., Setién, Koeman, Enrique and Valverde that, looking at p , can be grouped as an unique class; indeed, tests involving blue and yellow managers' passing networks result in p -values that are greater than the significance level $\alpha = 0.05$. For the sake of completeness, we report in Appendix D the same analysis using the Extrinsic Approach.

6. Discussion

The Intrinsic Approach is a new method that allows hypothesis testing between two populations of unlabelled networks. Its innovation lies in the integration of Graph Space theory in testing framework.

The strength of our method is supported by a great flexibility in the choice of test statistics and intrinsic distances.

In contrast to the alternative method, i.e., the Extrinsic Approach, the Intrinsic Approach does not downgrade the network to a lower dimensional object and, in this way, our method is not tied to the choice of extrinsic distances that may not consider crucial information about the original structure of the graph. Since the information about networks is preserved, with the Intrinsic Approach we can reach reliable results.

Moreover, the non-parametric permutation tests do not need critical distributional assumptions, nor particular requirements on the sample sizes.

The main limitation of this approach derive directly from drawbacks of the Graph Space and it is related to the fact that approximate matching algorithms slow down the entire method.

However, new algorithms may be studied and developed in order to improve this aspect and to support other intrinsic distances besides the Frobenius one. Moreover, further developments can be done providing quantitative tools to properly interpret results from testing in populations of unlabelled networks.

A. Dwass-Phipson-Smyth p-value

We report here the definition of p-value by Dwass (1957) and Phipson and Smyth (2010).

$$\begin{aligned}
 p(T) &= \frac{1}{M+1} \sum_{i=0}^M F\left(b(T); B, \frac{i+1}{M+1}\right) \\
 &\simeq \frac{b(T)+1}{B+1} - \int_0^{0.5/(M+1)} F(b(T); B, t) dt,
 \end{aligned}
 \tag{A.1}$$

where M is the overall number of permutations, $B < M$ is the number of randomly-selected permutations, T is the chosen test statistic, F is the cumulative probability function of the binomial distribution and $b(T)$ is the number of values of the permutation distribution greater than the observed value of T .

In practical applications, the integral approximation is used if $M > 10000$, otherwise the exact computation is considered.

B. IP-StudentFisher statistic

We report here the procedure explained in Lovato et al. (2020) to define the IP-StudentFisher statistic.

Consider a sample of independent and identically distributed network random variables $G_{11}, \dots, G_{1m_1} \sim F_1$ and a sample of independent and identically distributed network random variables $G_{21}, \dots, G_{2m_2} \sim F_2$. Lovato et al. (2020) introduce the IP-Student and the IP-Fisher statistics defined as in (3.1).

The two statistics in (3.1) can be combined using Non-Parametric Combination (Brombin and Salmaso, 2009; Pesarin and Salmaso, 2010). In particular, randomly choosing B permutations, the observed values of the Student and the Fisher statistics are computed along with the values of the permutation distribution; then the observed values and the permutation distributions are concatenated in two vectors of dimension $B+1$, one for each statistic. The values of these two vectors are then replaced by their rank, divided by $B+1$ and sorted in decreasing order. Finally, the last two vectors are combined in a unique vector by the Tippett's combining function $\psi(x, y) = 1 - \min(x, y)$ (Tippett, 1937). The combined vector stores the observed value and the permutation values of the IP-StudentFisher statistic. This statistic ensures that the null hypothesis is rejected if even one of the combined statistics rejects it.

C. Distribution parameters in simulation studies

We report in Table 1 the parameters of the i.i.d. Binomial distributions ($\text{Binomial}(n_i, p_i), i \in \{1, 2\}$) used to compute the power of test in the three scenarios analysed in Section 4.1. We consider the same exact parameters reported in Lovato et al. (2020).

Table 1: Parameters of i.i.d. Binomial distributions used in the three scenarios: location-only, scale-only and location-scale differences.

	n_1	p_1	n_2	p_2	μ_1	μ_2	σ_1^2	σ_2^2	$ \mu_2 - \mu_1 $	σ_2^2/σ_1^2
Location-only	10	0.5	10	0.5	5	5	2.5	2.5	0	1
	10	0.50625	10	0.49375	5.0625	4.9375	2.5	2.5	0.125	1
	10	0.5125	10	0.4875	5.125	4.875	2.5	2.5	0.25	1
	10	0.51875	10	0.48125	5.1875	4.8125	2.5	2.5	0.375	1
	10	0.525	10	0.475	5.25	4.75	2.5	2.5	0.5	1
Scale-only	300	0.2	300	0.2	60	60	48	48	0	1
	300	0.2	375	0.16	60	60	48	50.4	0	1.05
	300	0.2	500	0.12	60	60	48	52.8	0	1.1
	300	0.2	750	0.08	60	60	48	55.2	0	1.15
	300	0.2	1500	0.04	60	60	48	57.6	0	1.2
Location-scale	20	0.1	20	0.1	2	2	1.8	1.8	0	1
	20	0.1	21	0.1	2	2.1	1.8	1.89	0.1	1.05
	20	0.1	22	0.1	2	2.2	1.8	1.98	0.2	1.1
	20	0.1	23	0.1	2	2.3	1.8	2.07	0.3	1.15
	20	0.1	24	0.1	2	2.4	1.8	2.16	0.4	1.2

In Table 2, we report parameters of the multivariate Poisson distribution ($\text{Poisson}(\boldsymbol{\lambda}_i), i \in \{1, 2\}$) used to draw i.i.d spectra, along with the two fixed sets of eigenvectors in the two scenarios analysed in Section 4.2 .

Table 2: Parameters of the multivariate Poisson distribution for the i.i.d. spectra and eigenvectors used in the two scenarios: false H_0 and true H_0 .

	λ_1	λ_2	Eigenvectors sample 1								Eigenvectors sample 2							
False H_0	30	30	0.22	0.24	-0.22	-0.07	0.16	-0.08	0.21	0.88	0.37	-0.19	-0.46	-0.44	-0.56	0.32	0.10	0.02
	9	9	0.35	-0.03	0.24	0.43	-0.20	0.74	-0.14	0.16	0.23	0.60	-0.50	0.10	0.22	-0.24	0.43	-0.20
	8	8	0.18	-0.15	0.39	-0.20	0.03	-0.29	-0.78	0.23	0.49	-0.21	0.24	0.67	-0.17	0.16	0.37	0.11
	5	5	0.44	0.09	0.04	0.43	-0.53	-0.56	0.15	-0.08	0.26	0.22	0.60	-0.31	-0.26	-0.20	0.09	-0.55
	4	4	0.74	0.18	-0.10	-0.48	0.20	0.11	0.07	-0.34	0.58	-0.18	0.13	-0.39	0.57	-0.13	-0.036	0.34
	4	4	0.14	-0.35	-0.76	0.32	0.19	-0.03	-0.36	-0.05	0.24	0.54	0.08	0.11	0.08	0.62	-0.49	0.02
	3	3	0.19	-0.59	0.36	0.22	0.53	-0.15	0.37	0.03	0.15	0.29	0.02	0.10	-0.46	-0.55	-0.33	0.52
	3	3	0.04	-0.64	-0.12	-0.45	-0.55	0.11	0.17	0.17	0.29	-0.32	-0.31	0.27	0.08	-0.27	-0.55	-0.51
True H_0	30	30	0.22	0.24	-0.22	-0.07	0.16	-0.08	0.21	0.88	0.22	0.24	-0.22	-0.07	0.16	-0.08	0.21	0.88
	9	9	0.35	-0.03	0.24	0.43	-0.20	0.74	-0.14	0.16	0.35	-0.03	0.24	0.43	-0.20	0.74	-0.14	0.16
	8	8	0.18	-0.15	0.39	-0.20	0.03	-0.29	-0.78	0.23	0.18	-0.15	0.39	-0.20	0.03	-0.29	-0.78	0.23
	5	5	0.44	0.09	0.04	0.43	-0.53	-0.56	0.15	-0.08	0.44	0.09	0.04	0.43	-0.53	-0.56	0.15	-0.08
	4	4	0.74	0.18	-0.10	-0.48	0.20	0.11	0.07	-0.34	0.74	0.18	-0.10	-0.48	0.20	0.11	0.07	-0.34
	4	4	0.14	-0.35	-0.76	0.32	0.19	-0.03	-0.36	-0.05	0.14	-0.35	-0.76	0.32	0.19	-0.03	-0.36	-0.05
	3	3	0.19	-0.59	0.36	0.22	0.53	-0.15	0.37	0.03	0.19	-0.59	0.36	0.22	0.53	-0.15	0.37	0.03
	3	3	0.04	-0.64	-0.12	-0.45	-0.55	0.11	0.17	0.17	0.04	-0.64	-0.12	-0.45	-0.55	0.11	0.17	0.17

D. Extrinsic Approach to real application

We report here the same analysis performed in Section 5, using the Extrinsic Approach with the spectral distance.

For the UEFA EURO 2020 competition, we can make similar observations with respect to the Intrinsic Approach. From Figure 7 and Figure 8, we can qualitatively interpret the membership to a cluster being based on the number of passes. Moreover, we can observe that the medium-high cluster (blue) is bigger, since it contains Turkey, Poland and Scotland that in the Intrinsic Approach are inside the medium-low cluster (yellow).

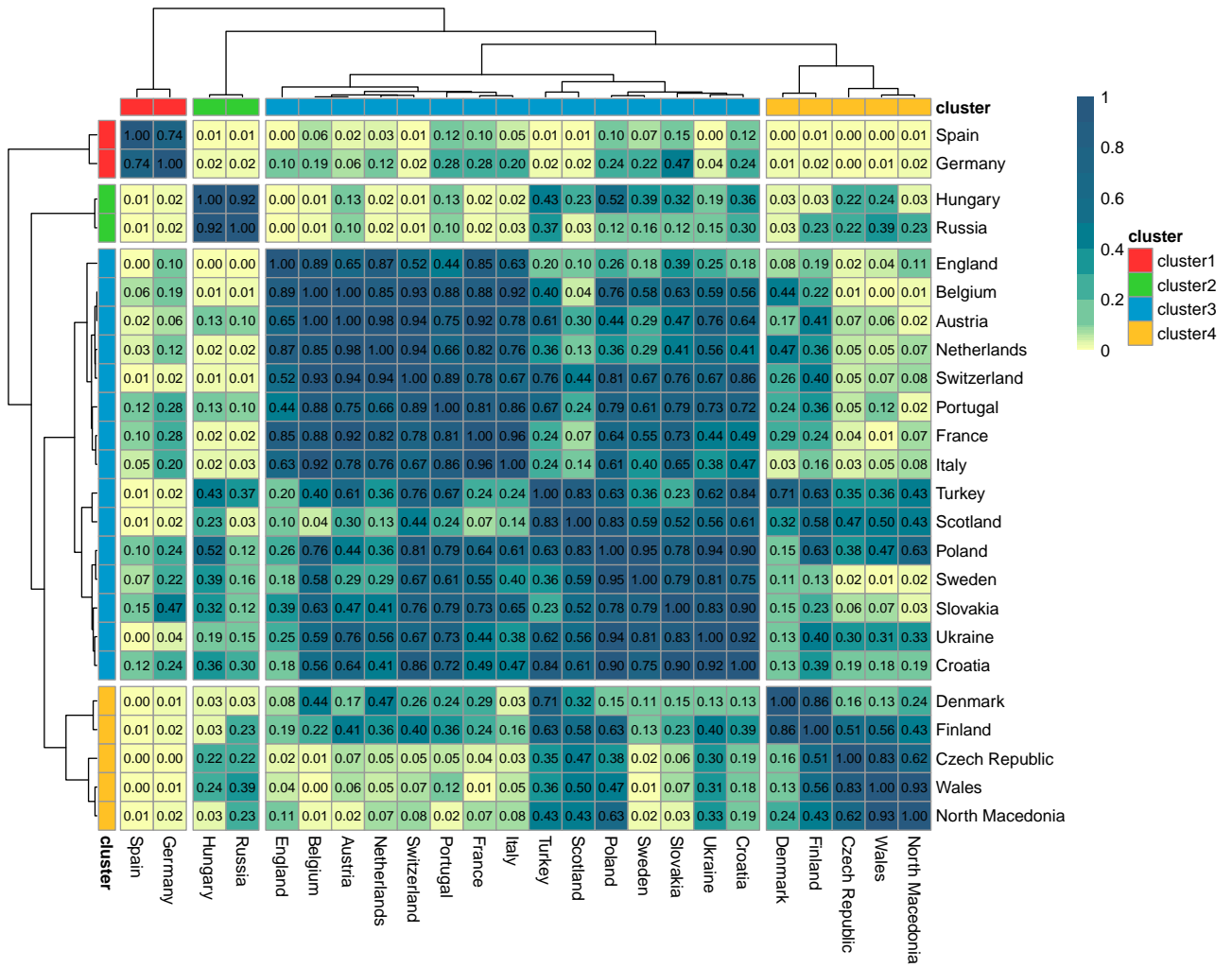


Figure 7: P-values matrix p of permutation tests in which we compare team by team players' passing networks distributions. Dendrograms represent the result of a hierarchical agglomerative clustering (Cophenetic Correlation Coefficient = 0.7167526). We cut dendrograms in order to show 4 clusters. Extrinsic Approach is considered. Data source: StatsBomb Open Data.

For LaLiga competition, we can make some observations with respect to the Intrinsic Approach. From Figure 9, it is clear to visualize two different clusters, one composed by only Rijkaard's passing networks (red) and the other one from all the remaining managers' (green, blue and yellow). It seems that the Extrinsic Approach is not able to detect qualitative differences inside the larger cluster, since p-values are almost all greater than the significance level $\alpha = 0.05$. These differences should be based on distribution of passes and can be qualitatively deduced from Figure 10, as done in an analogous way for the Intrinsic approach in Section 5.2.

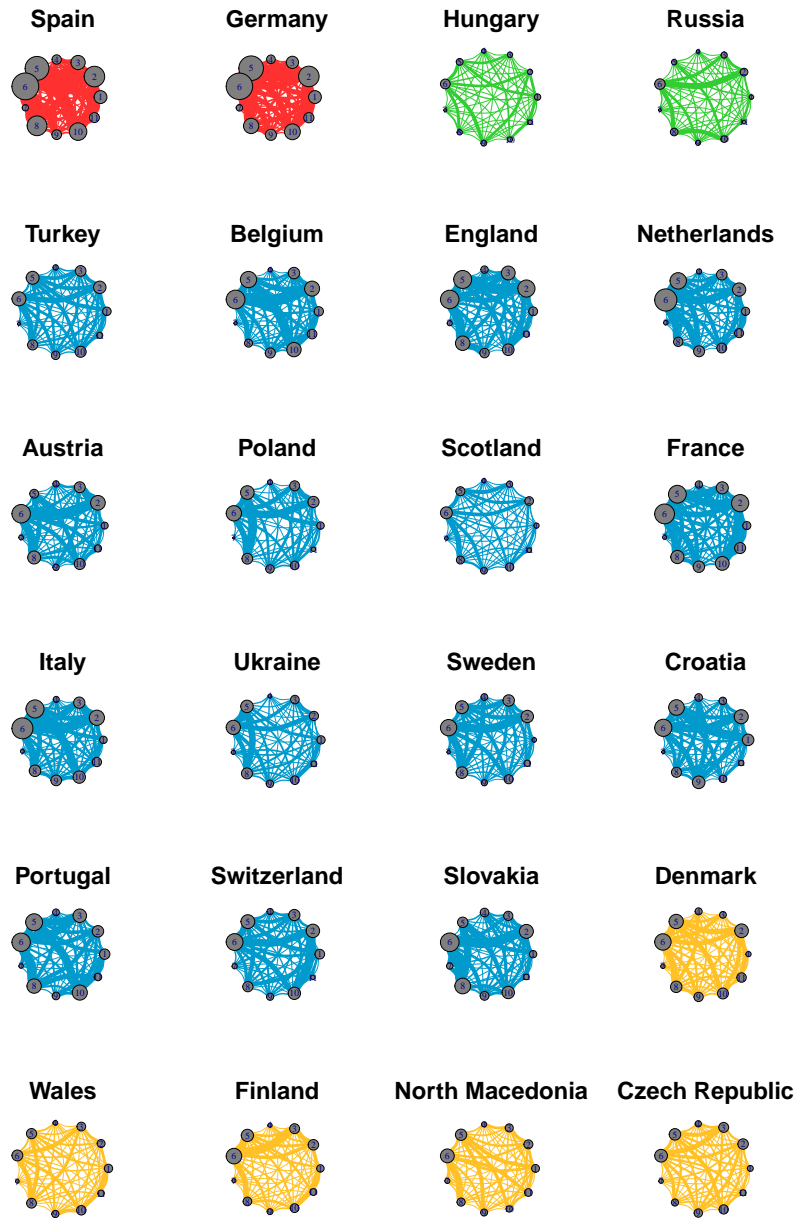


Figure 8: Fréchet means of the players' passing networks for each team. Colours represent the membership to one of the 4 clusters represented in Figure 7, node size refers to node strength (the larger the strength, the larger the size) and edge width to its weight (the wider the edge, the larger its weight). Extrinsic Approach is considered. Data source: StatsBomb Open Data.

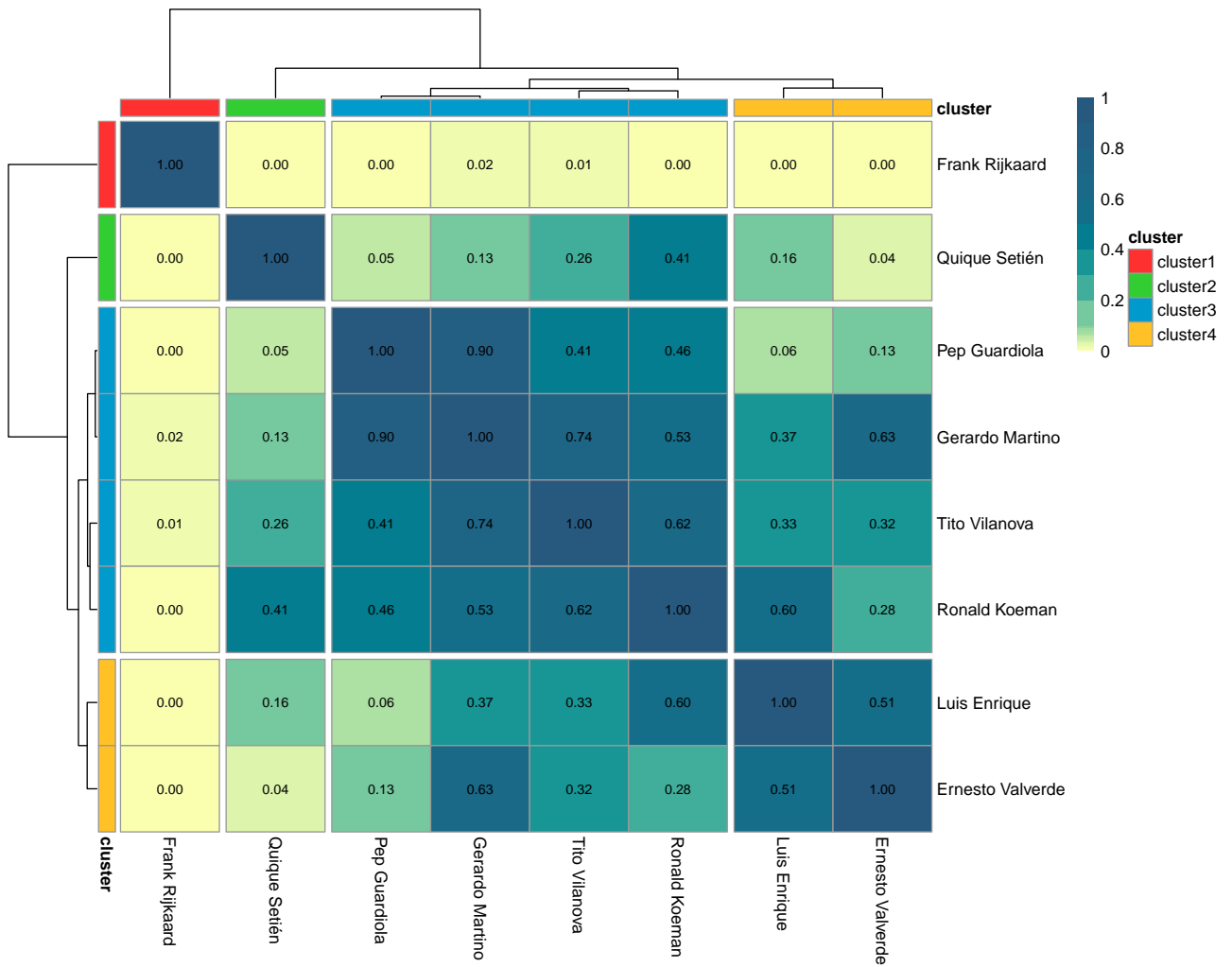
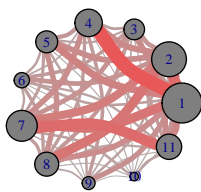
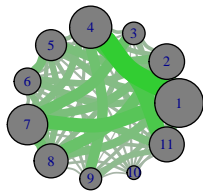


Figure 9: P-values matrix p of permutation tests in which we compare Barcelona managers' passing networks distributions. Dendrograms represent the result of a hierarchical agglomerative clustering (Cophenetic Correlation Coefficient = 0.9376636). We cut dendrograms in order to show 4 clusters. Extrinsic Approach is considered. Data source: StatsBomb Open Data.

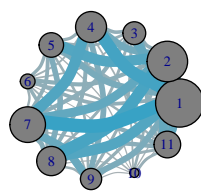
Frank Rijkaard



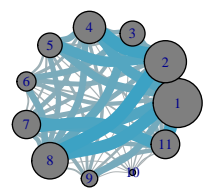
Quique Setién



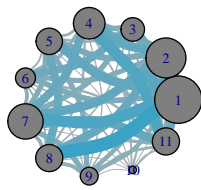
Pep Guardiola



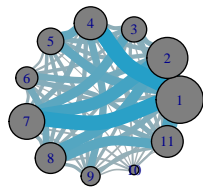
Tito Vilanova



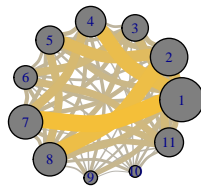
Gerardo Martino



Ronald Koeman



Luis Enrique



Ernesto Valverde

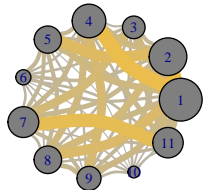


Figure 10: Fréchet means of the players' passing networks for each Barcelona manager. Colours represent the membership to one of the 4 clusters represented in Figure 9, node size refers to node strength (the larger the strength, the larger the size) and edge width and colour saturation to its weight (the wider and more vivid the edge, the larger its weight). Extrinsic Approach is considered. Data source: StatsBomb Open Data.

References

- H. Wang, J. S. Marron, Object oriented data analysis: Sets of trees, *The Annals of Statistics* 35 (2007). URL: <https://doi.org/10.1214/2F009053607000000217>. doi:10.1214/009053607000000217.
- I. Lovato, A. Pini, A. Stamm, S. Vantini, Model-free two-sample test for network-valued data, *Computational Statistics & Data Analysis* 144 (2020) 106896. URL: <https://www.sciencedirect.com/science/article/pii/S0167947319302518>. doi:<https://doi.org/10.1016/j.csda.2019.106896>.
- L. Chen, N. Josephs, L. Lin, J. Zhou, E. D. Kolaczyk, A spectral-based framework for hypothesis testing in populations of networks, 2020. URL: <https://arxiv.org/abs/2011.12416>. doi:10.48550/ARXIV.2011.12416.
- D. Ghoshdastidar, M. Gutzeit, A. Carpentier, U. von Luxburg, Two-sample hypothesis testing for inhomogeneous random graphs, *The Annals of Statistics* 48 (2020) 2208 – 2229. URL: <https://doi.org/10.1214/19-AOS1884>. doi:10.1214/19-AOS1884.
- A. Calissano, A. Feragen, S. Vantini, Populations of unlabeled networks: Graph space geometry and geodesic principal components, MOX Report (2020).
- R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. URL: <https://www.R-project.org/>.
- M. Dwass, Modified Randomization Tests for Nonparametric Hypotheses, *The Annals of Mathematical Statistics* 28 (1957) 181 – 187. URL: <https://doi.org/10.1214/aoms/1177707045>. doi:10.1214/aoms/1177707045.
- B. Phipson, G. K. Smyth, Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn, *Statistical Applications in Genetics and Molecular Biology* 9 (2010). URL: <https://doi.org/10.2202/1544-6115.1585>. doi:doi:10.2202/1544-6115.1585.
- C. Brombin, L. Salmaso, Multi-aspect permutation tests in shape analysis with small sample size, *Computational Statistics & Data Analysis* 53 (2009) 3921–3931. URL: <https://www.sciencedirect.com/science/article/pii/S0167947309001741>. doi:<https://doi.org/10.1016/j.csda.2009.05.010>.
- F. Pesarin, L. Salmaso, *Permutation Tests for Complex Data*. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, 2010.
- L. H. C. Tippett, *The methods of statistics : an introduction mainly for experimentalists / by L. H. C. Tippett, 2. ed. revised and enlarged ed., Williams and Norgate, London, 1937.*
- G. J. Székely, M. L. Rizzo, Energy statistics: A class of statistics based on distances, *Journal of Statistical Planning and Inference* 143 (2013) 1249–1272. URL: <https://www.sciencedirect.com/science/article/pii/S0378375813000633>. doi:<https://doi.org/10.1016/j.jspi.2013.03.018>.
- M. Biswas, A. K. Ghosh, A nonparametric two-sample test applicable to high dimensional data, *Journal of Multivariate Analysis* 123 (2014) 160–171. URL: <https://www.sciencedirect.com/science/article/pii/S0047259X13001966>. doi:<https://doi.org/10.1016/j.jmva.2013.09.004>.
- H. Chen, J. H. Friedman, A new graph-based two-sample test for multivariate and object data, *Journal of the American Statistical Association* 112 (2017) 397–409. URL: <https://doi.org/10.1080/01621459.2016.1147356>. doi:10.1080/01621459.2016.1147356. arXiv:<https://doi.org/10.1080/01621459.2016.1147356>.
- Z. Qiao, D. Sussman, *iGraphMatch: an R Package for the Analysis of Graph Matching*, 2021. URL: <https://arxiv.org/abs/2112.09212>. doi:10.48550/ARXIV.2112.09212.
- D. Koutra, N. Shah, J. T. Vogelstein, B. Gallagher, C. Faloutsos, Deltacon: Principled massive-graph similarity function with attribution, *ACM Trans. Knowl. Discov. Data* 10 (2016). URL: <https://doi.org/10.1145/2824443>. doi:10.1145/2824443.
- M. Tantardini, F. Ieva, L. Tajoli, C. Piccardi, Comparing methods for comparing networks, *Scientific Reports* 9 (2019). doi:10.1038/s41598-019-53708-y.
- N. Pržulj, D. G. Corneil, I. Jurisica, Modeling interactome: scale-free or geometric?, *Bioinformatics* 20 (2004) 3508–3515. URL: <https://doi.org/10.1093/bioinformatics/bth436>. doi:10.1093/bioinformatics/bth436.

- A. Sarajlić, N. Malod-Dognin, O. Yaveroglu, N. Przulj, Graphlet-based characterization of directed networks, *Scientific Reports* 6 (2016) 35098. doi:10.1038/srep35098.
- D. Aparício, P. Ribeiro, F. Silva, Network comparison using directed graphlets, arXiv preprint arXiv:1511.01964 (2015).
- A. Tsitsulin, D. Mottin, P. Karras, A. Bronstein, E. Müller, NetLSD: Hearing the Shape of a Graph, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2347–2356.
- J. P. Bagrow, E. M. Bollt, An information-theoretic, all-scales approach to comparing networks, *Applied Network Science* 4 (2019) 1–15.
- R. Ievoli, A. Gardini, L. Palazzo, The role of passing network indicators in modeling football outcomes: an application using bayesian hierarchical models, *AStA Advances in Statistical Analysis* (2021) 1–23.
- A. Calissano, A. Feragen, S. Vantini, Graph-valued regression: Prediction of unlabelled networks in a non-euclidean graph space, *Journal of Multivariate Analysis* 190 (2022) 104950. URL: <https://www.sciencedirect.com/science/article/pii/S0047259X22000021>. doi:<https://doi.org/10.1016/j.jmva.2022.104950>.
- StatsBomb Services Limited, StatsBomb Open Data, 2022. URL: <https://github.com/statsbomb/open-data>.
- I. Gómez, Football passing networks using R, 2020. URL: <https://www.datofutbol.cl/passing-networks-r/> and <https://github.com/Dato-Futbol/passing-networks>, [Online; accessed 16-June-2022].

Abstract in lingua italiana

In questo articolo, si propone un approccio al test d'ipotesi per reti non etichettate. Per trattare reti che differiscono nel numero o nelle etichette dei nodi, si considera la teoria sullo Spazio dei Grafi. Si adottano test di permutazione, garantendo flessibilità nella scelta di statistiche test "inter-point" basate su metriche intrinseche. Dunque, si valutano le distanze tra reti, affidandosi ad algoritmi di matching approssimato, in modo da ottenere allineamenti ottimali dei grafi. Le simulazioni rivelano efficacia e prestazioni superiori del metodo rispetto ad approcci alternativi. Infine, si mostra l'utilità del metodo nell'applicazione a reti di passaggio tra calciatori. L'approccio proposto è disponibile nel pacchetto R *nevada*.

Parole chiave: Reti, Reti non etichettate, Spazio dei Grafi, Test d'ipotesi, Test di permutazione, Rete passaggi tra calciatori

Acknowledgements

I would like to acknowledge my brother and my parents, that inestimably supported me in every situation. Special thanks to my grandmother, Isabella, that encouraged me to improve myself.

I am very grateful to my advisor Simone Vantini and co-advisors Anna Calissano and Aymeric Stamm, for their constructive feedbacks and invaluable support.

Lastly, I would like to thank everyone that accompanied me in my personal growth path.