

Triad-based Comparison and Signatures of Directed Networks

Xiaochuan Xu and Gesine Reinert

Department of Statistics, University of Oxford, `reinert@stats.ox.ac.uk`

Abstract. We introduce two methods for comparing directed networks based on triad counts, called *TriadEuclid* and *TriadEMD*. *TriadEuclid* clusters the Euclidean distance between triad counts, whereas *TriadEMD* is an adaptation of *NetEMD* for directed networks. We apply both methods to cluster synthetic networks, a set of web networks including google, twitter, peer-to-peer, amazon, slashdot and citation networks, as well as world trade networks from 1962-2000. Furthermore, we find signature triads and signature orbits for each type of networks in our data, which show the main triad and orbit contributions of the networks when comparing them to the other networks in the respective data set.

Key words: network comparison, triad, signature triad, signature orbit

1 Introduction

Networks as representation of complex interaction dataset have become increasingly common. Friendships, world trade flow and protein-protein interactions can all be described as networks [14]. A key question to be addressed is network comparison, see for example [25, 15, 1, 26]. Subject specific network comparison may provide new insights; for example, comparing protein-protein interaction networks may increase our understanding of evolutionary processes [1].

Network comparison is addressed in different ways. In machine learning, graph kernels are used to obtain classifiers to predict the class membership of networks. There are many applications in computer vision, personalized medicine e.g. [3] and drug discovery e.g. [22]. Other computational algorithms are based on network alignment [15, 9, 11] which can be quite computer-intensive. Instead algorithms which compare counts of small subgraphs have become popular, such as [17] and *NetEMD* [24]. Network comparison based on small subgraphs is motivated by the observation that many real networks contain some characteristic small subgraphs, sometimes called *motifs*, which relate to the function of the network, see [13]. In [13], motifs are introduced as over-or under-represented subgraphs, where the over-or under-representation is usually judged by comparison to an underlying network null model. When there is no suitable underlying network model available, often a configuration model is used as null model. The significance of motifs depends crucially on the underlying null model, see [16].

In [20] a non-parametric method based on subgraph counts of up to 4-node subgraphs is developed, but excluding graphlets with bi-directional edges. Their two main methods are *DGCD129*, including all directed graphlets on 2-4 nodes, and *DGCD13*,

which uses only directed graphlets on 2-3 nodes and is deemed superior to *DGCD13* in [20]. The fact that directed graphlets contain more information than undirected graphlets has been confirmed in [2], where the graphlet degree distribution GDD is generalised to cluster directed networks, using graphlets on up to 5 nodes. The underlying GDD agreement measure for undirected networks is known to suffer from a pronounced dependence on the number of nodes though, see [18], and is not recommended for comparing networks with very different numbers of nodes.

In the social network analysis paradigm, directed triads are often used as unit of information, see for example [7], [23] and [5]. A triad is a directed graph composed of 3 nodes. Taking account of the different directions and configurations, there are 16 types of triads, of which 13 are connected. In [12], these 13 triads in networks are attached a significance score which is obtained by comparison to a configuration model, yielding a vector of length 13 which is called the *Triad Significance Profile* (TSP). The TSP again depends crucially on the underlying null model.

In this paper we propose two different methods for network comparison which are both based on counting the different triads as induced sub-graphs of networks. Both methods are non-parametric; instead of requiring a null model, they crucially depend on the other networks in the set of networks to be compared, as different features may be distinguishing for different comparisons. The first triad method, which we call *TriadEuclid*, uses the Euclidean distance between the count vectors and hierarchical clustering to group networks. The second method, *TriadEMD*, is a generalization of *NetEMD*. For the comparison of undirected networks of different sizes and densities, *NetEMD* outperforms other graphlet-based methods on some test data sets, see [24]. This paper provides an extension to directed networks.

Here, network comparison is used first, to classify networks according to generating mechanism, and second, to recover the time-ordering of networks. We test *TriadEuclid* and *TriadEMD* on three data sets and compare them to *DGCD13* and *DGCD129* from [20] using their default parameter settings.

The first data set is a synthetic data set of networks generated under three different models, with different numbers of nodes and different densities. All four methods cluster these networks flawlessly according to generating mechanism.

The second data set is a set of 31 sparse directed social networks from [10]. The networks in the different clusters differ considerably in size and density, and hence here the focus is whether *TriadEuclid* and *TriadEMD* make use of information which is not covered by nodes and degree distributions already. For this task, *TriadEuclid* outperforms *TriadEMD* as well as *DGCD13* and *DGCD129*.

The third case study concerns dense World Trade networks from [6]; here the task is to recover the time ordering of the networks. This data set is analysed first using a single-layer analysis which compares aggregated yearly trade networks, and then using a multi-layer approach which compares the trade networks for different commodities separately. For this task, *TriadEMD* outperforms *TriadEuclid* as well as *DGCD13* and *DGCD129*.

In all cases, both *TriadEuclid* and *TriadEMD* pick some topological structure beyond network size and density. This topological structure is described using triad and orbit counts. Motivated by the search of network motifs without underlying network

models, we also identify the triads and orbits which contribute most to the dissimilarity scores of a network compared to all other networks in the collection of networks which is to be clustered. We call these *signature triads* and *signature orbits*. Thus our notion of signatures depends on the networks which the network in question is compared against.

This paper thus has two main contributions. Firstly, it provides two new methods for network comparison, which outperform the often used *DGCD* methods. Secondly, it adds to the discussion on motifs by introducing signatures, which can be viewed as context-dependent motifs. The context is given by the networks which are involved in the network comparisons.

The paper is structured as follows. Section 2 gives some notation, reviews *NetEMD*, and details the performance measures which are used in this paper. Section 3 introduces *TriadEuclid* and *TriadEMD* as well as the signature calculations. Section 4 gives the results for synthetic networks, Section 5 shows the results for the sparse directed real-world networks, and Section 6 considers the trade networks. The results are discussed in Section 7. More details of the results are found in the Supplementary Material; together with the R code for this paper it is available at www.stats.ox.ac.uk/all-people-gesine-reinert-network-comparison/.

2 Background

NetEMD. In this paper all directed networks $G = (V, E)$ are simple, finite, and subgraph counts are understood to be induced subgraph counts. We compare two networks using *NetEMD*, a pseudo-distance which is based on the Earth Mover’s Distance (EMD, also known as the 1st Wasserstein metric). The EMD between two probability distributions p and q on the real line is given by $EMD(F, G) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx$, where F and G are the cumulative distribution functions of p and q respectively, see for example [19] and [21].

In [24], the EMD is used to construct a score called *NetEMD* which compares two undirected networks through their distributions of network features, such as the degree distribution or the distribution of the number of triangles which a node belongs to. When the aim is to compare two finite networks G and G' , and $p_t(G)$ and $p_t(G')$ are the distributions of feature t on G and G' which we assume to have non-zero variance, then we re-scale p_t and q_t so that they have empirical variance 1; denote the re-scaled distributions by \bar{p} and \bar{q} . The *NetEMD* pseudo-distance between two networks G and G' based on feature t is defined as

$$NetEMD_t(G, G') = \inf_{c \in \mathbb{R}} (EMD(\bar{p}(\cdot + c), \bar{q}(\cdot))). \quad (1)$$

Taking the infimum over all $c \in \mathbb{R}$ ensures that $NetEMD_t(G, G')$ is location-invariant; a proof as well as a calculation example can be found in the Supplementary Material. This property implicitly adjusts for different expectation values for different networks to render the method robust for comparing networks of different size and densities. In [24], two networks G and G' are compared using a set of features t_1, \dots, t_m ; the corresponding *NetEMD* is then the average of $NetEMD_{t_1}(G, G'), \dots, NetEMD_{t_m}(G, G')$. In this paper we adapt *NetEMD* to directed networks and use counts of triads as features; we call the resulting method *TriadEMD*.

Evaluation of the output. We use three different methods to assess the similarity between a partition which is obtained through a clustering method and the true assignment to clusters. The first method is the Adjusted Rand Index (ARI) from [8] which compares two partitions. If the networks are time-ordered, then how well the clustering reflects the time-ordering is assessed using the second method, which is based on two nearest neighbour scores. The third method is a Monte Carlo test to assess whether the method captures topological information beyond size and density of the networks.

To assess how well the time ordering of networks is recovered, we use two quality scores. The first score, a Nearest Neighbour Score (NNscore) which ranges from 0 to 1, is constructed as follows. First we describe it for a time series of items. For a given $n \times n$ dissimilarity matrix, we start with NNscore = 0. For each item in the dissimilarity matrix, we find the item with shortest dissimilarity to the chosen item (ties are broken at random). If the shortest dissimilarity neighbour of an item is one of its two true neighbours, then we count it as correct and add $\frac{1}{n}$ to NNscore (otherwise we count it as false). Our second quality score, the Nearest Two Neighbour Score (N2Nscore) considers the closest two neighbours. Note that in a time series of networks, with time ordering as ground truth, the first and last item only have one neighbour. For a given $n \times n$ dissimilarity matrix, the N2N score is calculated like the NNscore but replacing the second step by finding the two items with shortest dissimilarity to the chosen item (ties are broken at random). For the first and the last item, only find the item with shortest dissimilarity to the chosen item. If the two (or one) shortest dissimilarity neighbours of the selected item are its two true neighbours, then we add $\frac{1}{n}$ to N2NScore. The N2NScore is thus bounded above by the NNscore.

In order to see whether our method captures topological information beyond than density and size of the networks we apply a Monte Carlo Test [4], as follows. We use the ARI as test statistics and test the null hypothesis that an ARI at least as high as the one observed is consistent with the assumption that the underlying network is random (in the sense of an ER network or a configuration model). We simulate many networks under the null hypothesis (which is either an ER graph or a configuration model); the p -value of the Monte Carlo test is the proportion of (simulated+observed) test statistic which are at least as extreme as the observed test statistic. If we can reject the null hypothesis then there is evidence that the network comparison methods pick up information from the network which goes beyond the information used to build the random networks under the null hypothesis.

3 Triad methods, signature triads and signature orbits

TriadEuclid combines the Euclidean distance between the feature vectors with standard hierarchical clustering to cluster networks using features based on the triad census data. The triad censuses are induced sub-network counts of triads. The 16 different types of triads are shown in Figure 1(a). We exclude the disconnected triads 1, 2 and 5.

For a directed network G , we inspect all of its induced sub-networks with three nodes and count the different triads on these sub-networks. On any set of three nodes there will be exactly one triad from the set of 16 triads present. Denote by $n_i(G)$ the number of induced sub-graphs in G which match triad i , for $i \in \mathcal{T} := \{1, \dots, 16\} \setminus$

$\{1, 2, 5\}$. Set $N(G) = \sum_{i \in \mathcal{T}} n_i(G)$ and

$$T(G) = (f_i(G), i \in \mathcal{T}) \quad \text{where} \quad f_i(G) = n_i(G)/N(G). \quad (2)$$

Then T maps the space of finite simple directed networks to the feature space $([0, 1] \cap \mathbb{Q})^{13}$. *TriadEuclid* gives a pseudo-distance between two different networks by calculating the Euclidean distance between their feature vectors. For a finite set of networks $\{G_i, i \in I\}$ which are to be clustered, their pairwise *TriadEuclid* dissimilarities constitute the entries of a dissimilarity matrix D . Then, we cluster the networks $\{G_i, i \in I\}$ based on the dissimilarity matrix D using hierarchical clustering according to a pre-set number of clusters.

Following [24], we further classify the nodes into different orbits according to automorphisms. An orbit is a specific location of a node in a triad. Figure 1(b) shows the 30 different triad orbits. For example, in a bi-directed 2-star, the centre node corresponds to orbit 27, whereas the two nodes which are not connected to each other correspond to orbit 26. They have the same orbit because they could be interchanged without changing the triad.

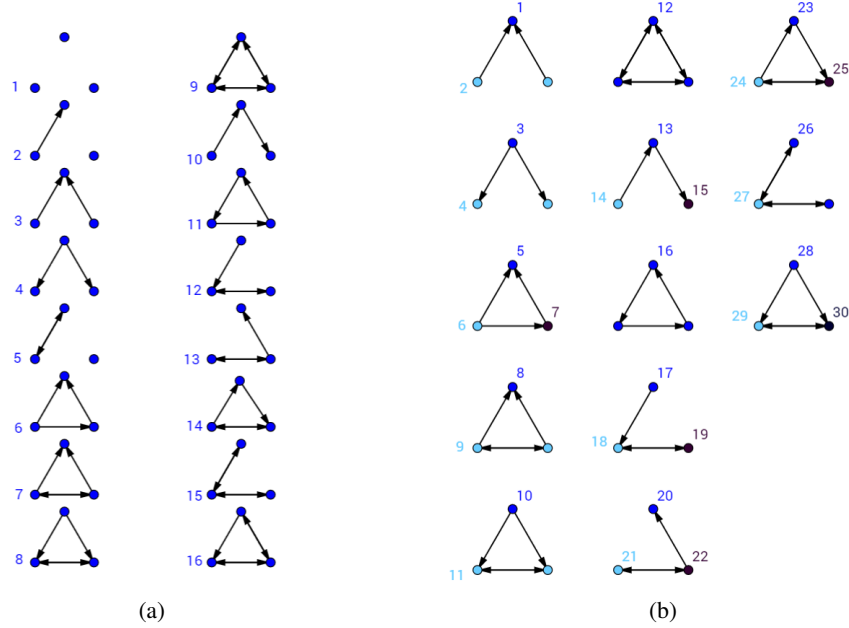


Fig. 1: (a) 16 different triads and their labels. (b) 30 triad orbits. The different shades in each triad represent different automorphism orbits, numbered from 1 to 30.

Analogous to the degree distribution, *TriadEMD* is based on degree distributions of orbits in triads. For each node in the graph, we count the *orbit degree* of the node. The orbit degree k_i of orbit i of a node is the number of appearances of orbit i that the

node is involved in. The orbit i degree distribution $P(k_i)$ of a network is then defined to be the fraction of nodes in the network with orbit i degree k_i . A worked example can be found in the Supplementary Material. Given two networks, we calculate *NetEMD* between their triad orbit degree distributions of the 30 different orbits and take their average as their dissimilarity *TriadEMD*.

To calculate *signature triads*, we disentangle *TriadEuclid* as follows. For each network $G_i, i \in I$ in the collection of interest, we record the index $e_{i,j}$ in \mathcal{T} of the triad (or triads) which give the highest contribution to the $TriadEMD(G_i, G_j)$ -score, for each $j \in I \setminus \{i\}$. The index is given a sign; it is positive when the network G_i over-represents the corresponding triad, and negative when it under-represents the corresponding triad. We then choose the triad index for G_i with the largest frequency of appearance among $e_{i,j}, j \in I \setminus \{i\}$. If there is a tie, then all indices which achieve the highest frequency are reported. We call the corresponding triad the *signature triad* for network G_i . Similarly, disentangling *TriadEMD* gives *signature orbits* for networks. The signature triads and signature orbits are thus calculated relative to the other networks in the set of networks to be clustered, and are hence cluster-dependent.

4 Case Study I: Synthetic Networks

To test our methods we use the R package *igraph* to generate 27 synthetic directed network which are versions of the Barabási–Albert (BA) models, the Erdős–Rényi (ER) model, and the 2-D geometric (GEO) model. For each model, we simulated three networks with different sizes (500 and 1000 nodes) and densities, with 3 replicas each. The detailed parameter settings for the directed networks are;

for BA: 1 edge or 2 edges added for every new node;

for ER: edge probability 0.1 or 0.2 for 500 nodes, $0.1 * 499/999$ for 1000 nodes;

for GEO: radius r such that $\pi r^2 = \frac{1}{2} \times (\text{number of nodes}) \times (\text{ER edge probability})$.

The BA models start with a single edge and are naturally directed networks with a single direction per edge. For the ER networks, each directional edge has probability p of appearing, and bi-directional edges are possible. For the GEO networks, first an undirected network is generated, and then two directed edges are created for each undirected edge, one in each direction. For all of these networks, replicas are given subscripts indicating their parameter settings.

Table 1: Signature Triads of Synthetic Networks

Network	Signature Triads No.	Signature Orbits No.
all BA	3	6, 12, 19
ER on 1000 nodes	10	1, 4, 16
ER on 500 nodes	-15	11, 13, 19, 26
GEO	9, 15	4, 12, 19

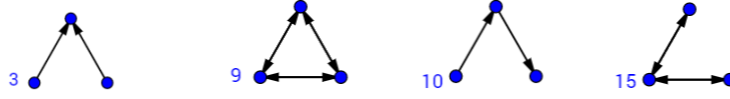


Fig. 2: The signature triads for Case Study I

Both *TriadEuclid* and *TriadEMD* classify all of the networks correctly according to generating model, so all the Adjusted Rand Indices and the nearest neighbour scores are 1. Table 1 shows the labels of the signature triads of different types of networks, and Figure 2 shows these triads. All of the BA networks have as signature triad Triad 3, two directed edges both pointing to the same node, and this triad is over-represented. That Triad 3 is the overwhelming triad structure in the BA method is not surprising, as BA networks are close to directed trees. In all of the ER models on 500 nodes, Triad 15 is under-represented - two nodes connected by bi-directional edges to a third node, without an edge between the two nodes. In the ER models on 1000 nodes, the directed path triad 10 is over-represented. The GEO models do not show quite as clear a pattern: triads 9 (3 bi-directional edges) and 15 (2 bi-directional edges) are the only ones appearing as signature triads, both are over-represented, but there appears to be no density or size dependence on which of the two patterns is the signature pattern. These signature patterns for GEO are plausible as in the generated GEO networks, all edges are bidirectional, whereas in the BA networks all edges except the initial edge are uni-directional; in the ER networks edges can be bi-directional or uni-directional.

The signature orbits for the BA networks are 6 (two edges emanating from a central node with the other two nodes linked by a directed edge), 12, and 19. These orbits do not sit on the signature triad 3 but instead pick out parts of other triads. Orbits 6, 12 and 19 pick up on the lack of triangles in BA networks which have a reciprocal edge. The signatures are unusual compared to ER and GEO networks as constructed, but may not appear unusual when compared to other networks. The complete results can be found in the Supplementary Material.

5 Case Study II: Clustering of Different Types of Networks

For this case study, we selected 31 networks of 6 different clusters from [10] - the directed social networks with less than 10 M nodes, as well as the co-purchasing networks and the internet peer-to-peer networks. The selected networks are google networks, twitter networks, amazon networks, a citation network, p2p-Gnutella networks and Slashdot networks. The edge density regions for the google networks and the twitter networks overlap, and the edge density of the citation network is in the range of that of the p2p-Gnutella networks. Detailed information of the networks can be found in the Supplementary Material. The networks vary considerably in the number of nodes and density; here the research question is whether *TriadEuclid* and *TriadEMD* detect information beyond the number of nodes and the density. Note that we assume as ground truth that networks from the same source belong to the same cluster. For comparison, *DGCD13* achieves an ARI of 0.38, and *DGCD129* achieves an ARI of 0.45.

Results for *TriadEuclid*. We map the networks to the feature space using the map T in (2) to obtain the dissimilarity matrix, and assign the networks to 6 clusters using hierarchical clustering. The ARI between the true partition and the clustering result is 0.70. To assess whether such a large ARI could be obtained just based on size and densities of the networks, we carry out Monte Carlo tests. First we generate 49 groups of 31 random networks by the ER model, which clone the sizes and densities of the original group. The p -value of the Monte Carlo test is 0.02, hence the null hypothesis is rejected under a significance level of 5%, indicating that *TriadEuclid* captures information beyond the number of nodes and the densities of the networks. Indeed this Monte Carlo test rejects any ARI value which is larger than 0.2 at the 5% level.

In the second test we generate 49 groups of 31 random networks which clone the degree distributions of the original group and are generated by the configuration model. Again, the p -value is 0.02, hence the null hypothesis is still rejected under a significance level of 5%. In contrast, an ARI of 0.45 as obtained for *DGCD129* has a p -value of 0.22, so that the hypothesis that *DGCD129* picks up no more information from the networks than is provided by the degree distribution cannot be rejected.

We obtain the signature triads by comparing the pairwise Euclidean distance between the feature vectors; the details are found in the Supplementary Material. Amazon and Cit-Hep have signature triad 3, a hierarchical structure with two reciprocal edges. The google networks tend to have signature triad 4, a 2-star with both edges pointing outward. However, Slashdot – a community of technology-related users, has triad 15 as signature triad.

Results for *TriadEMD*. First a dissimilarity matrix is obtained using *TriadEMD*. Then, as for *TriadEuclid*, we apply hierarchical clustering to cluster the networks. The ARI is 0.34, which is lower not only than that of *TriadEuclid*, but also of the *DGCD* methods. For this case study with a clear separation between the network types, *TriadEuclid* outperforms the *DGCD* methods, which in turn outperform *TriadEMD*.

6 Case Study III: Analysis of World Trade Networks

In this case study we compare networks in an example with fairly high density, namely 39 world trade networks, one per year between the years 1962 to 2000, from [6]; details are given in the Supplementary Material. There are 507 commodities which are common across the years, and we include only the trade of these commodities. Our aim is to recover the time ordering according to the dissimilarity between years.

We employ two ways to compute the dissimilarity between two years, a single-layer analysis and a multi-layer analysis. In the single-layer analysis we combine the networks of different commodities into one network and compute the dissimilarity based on the whole network of all the commodities; in the multi-layer analysis each commodity trade network is considered separately.

First we carry out the single-layer analysis. Note that the ARI does not apply as there are no pre-defined clusters which could be recovered. Instead the focus is on recovering the time ordering of the networks. For comparison, the NNscore for *DGCD13* is 0.61, and the NNscore for *DGCD129* is 0.64, whereas the N2Nscore for *DGCD13* is 0.20, and the N2Nscore for *DGCD129* is 0.25.

Results for *TriadEuclid*. The NNscore and the N2Nscore of *TriadEuclid* are 0.58 and 0.23. The N2Nscore is considerably lower than the NNscore, but higher than that of *DGCD13*. We also generate 49 random groups of 39 networks according to the ER model and 49 random groups of 39 networks according to the configuration model for the Monte Carlo Test. The null hypothesis is that our method does not give higher scores than on random networks. All p -values of the tests are 0.02 and hence we reject the null hypothesis at 5% level for both models. We conclude that this data set, *TriadEuclid* captures more information than only network size and degree distributions.

The network comparison result for *TriadEuclid* is visualised using a heatmap; different dissimilarity values are given different colours. For a good temporal ordering we would see diagonal lines. The heatmap can also help identify sudden changes in the network. Figure 3 is the heatmap of the dissimilarity matrix. Years that are close to each other are more similar than years that are far away, which is what we expected to see. There is a sudden change from 1983 to 1984. Indeed, the way the dataset was built changed in 1983, and that is a plausible explanation for the sudden change. Also, some small changes observed from the heatmap can be related to some historical events. For instance, the Soviet Union dissolved in 1991, and the World Trade Organization (WTO) was established in 1995.

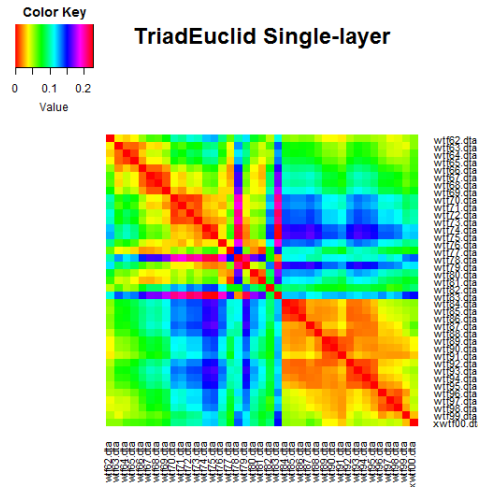


Fig. 3: Heatmap of the Single-layer World Trade Network for *TriadEuclid*

Results for *TriadEMD*. The NNscore and the N2Nscore of *TriadEMD* are 0.76 and 0.30, thus outperforming *TriadEuclid* as well as the *DGCD* methods. We also carry out the Monte Carlo test here, using the same random networks as generated before. The null hypothesis is that our method does not give higher scores than on random networks. The p -values for all tests are 0.02 and hence again we reject the null hypothesis at 5%

level for both models. Again we conclude that for this data set, *TriadEMD* captures more information than only network size and degree distributions.

Similar to the heatmap obtained by *TriadEuclid*, the heatmap in Figure 4 reflects the time evolution of the World Trade Network. The sudden change from 1983 to 1984 and the changes in 1991 and 1995 are even clearer here.

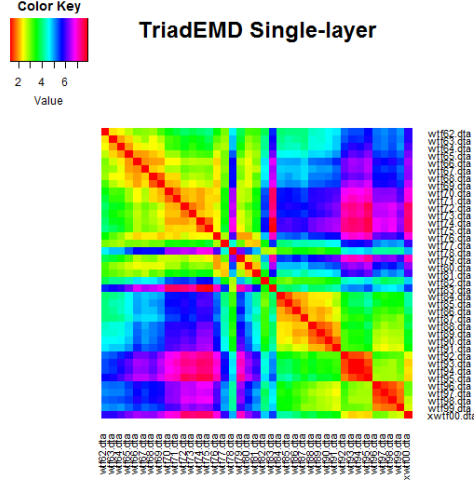


Fig. 4: Heatmap of the Single-layer World Trade Network using *TriadEMD*

Result of the multi-layer analysis. Finally we also carried out a multi-layer analysis of the world trade networks. Recall that in each year, there are 507 commodities, corresponding to 507 trade networks. We assume that the networks have different layers and each layer represents the network of one commodity. We compare the networks of each commodity to get the sub-dissimilarities for each commodity or layer. Then, we sum up the sub-dissimilarities as the dissimilarity measure between the two years. Due to computation complexities, we do not conduct the Monte Carlo test here, nor carry out the comparison to the *DGCD* scores.

The NNscore of *TriadEuclid* is 0.94, and its N2Nscore is 0.53. Both scores are considerably higher than those in the single-layer analysis. Due to computation complexities, we do not conduct the Monte Carlo test here. The heatmap for the multi-layer analysis is in the Supplementary Material. The time ordering are clearer in the multi-layer analysis than in the single-layer analysis. The sudden change from 1983 to 1984 is even more obvious.

The NNscore of *TriadEMD* is 0.97 and the N2Nscore of *TriadEMD* is 0.82. As is also the case for *TriadEuclid*, both scores are considerably higher than those in the single-layer analysis. They are also higher than the *TriadEuclid* scores in the multi-layer analysis. The heatmap in the Supplementary Material shows clear diagonal lines.

7 Conclusion

Network comparison has many valuable applications. In this paper, we focus on comparison of directed networks. We develop two triad-based methods, *TriadEuclid* and *TriadEMD*. *TriadEuclid* uses the Euclidean distance between feature vectors of triad frequencies to measure dissimilarity. *TriadEMD* further divides the triads into 30 orbits and uses *NetEMD* to measure the dissimilarity.

We use ARI, NNscore and N2Nscore to measure the performance of our methods, and we compare these values to those obtained by the *DGCD* methods. All four methods give perfect clustering on the synthetic data set (Case Study I); *TriadEuclid* performs better than the *DGCD* scores and *TriadEMD* in Case Study II, while *TriadEMD* performs better than the *DGCD* scores and *TriadEuclid* in Case Study III. We also apply a Monte Carlo test to test whether our methods can ignore the influence of size and density and find evidence that the best-performing methods use information which goes beyond degree distributions. Moreover, we are able to relate changes in the heatmap of the dissimilarity matrix of the World Trade networks to some historical events. For the World Trade networks, a multi-layer analysis resulted in higher classification scores than the single-layer analysis. It would be interesting to carry out a fine-scale analysis to identify and interpret the layers which contribute most to the classification.

In addition, we find the signature triad and signature orbit of each type of network by looking at the largest component in the dissimilarity matrix. These signatures can be viewed as a non-parametric analog of network motifs; instead of depending on an underlying model, they depend on the other networks in the comparison. In future work, an in-depth analysis of such signatures for different network comparisons may reveal more insights into the building blocks of complex networks.

Acknowledgements. This work was partially supported by the Alan Turing Institute. GR acknowledges the COST Action CA15109. The authors would like to thank Luis Ospina-Forero and Martin O'Reilly for helpful discussions, and Andrew Elliott for helpful discussions as well as computing support. They would also like to thank the anonymous referees for many helpful comments.

References

1. ALI, W., RITO, T., REINERT, G., SUN, F. & DEANE, C. M. (2014) Alignment-free protein interaction network comparison. *Bioinformatics*, **30**(17), i430–i437.
2. APARICIO, D., RIBEIRO, P. & SILVA, F. (2017) Extending the Applicability of Graphlets to Directed Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, **14**(6), 1302–1315.
3. BORGWARDT, K. M., KRIEGER, H.-P., VISHWANATHAN, S. & SCHRAUDOLPH, N. N. (2007) Graph kernels for disease outcome prediction from protein-protein interaction networks. in *Pacific Symposium on Biocomputing*, vol. 12, pp. 4–15.
4. DWASS, M. (1957) Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics*, pp. 181–187.
5. FAUST, K. (2010) A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, **32**(3), 221–233.

6. FEENSTRA, R. C., LIPSEY, R. E., DENG, H., MA, A. C. & MO, H. (2005) World trade flows: 1962-2000. Discussion paper, National Bureau of Economic Research.
7. HOLLAND, P. W. & LEINHARDT, S. (1976) Local structure in social networks. *Sociological methodology*, **7**, 1–45.
8. HUBERT, L. & ARABIE, P. (1985) Comparing partitions. *Journal of Classification*, **2**(1), 193–218.
9. KUCHARIEV, O. & PRŽULJ, N. (2011) Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, **27**(10), 1390–1396.
10. LESKOVEC, J. & KREVL, A. (2014) SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>. Accessed on May 21, 2017.
11. MAMANO, N. & HAYES, W. B. (2017) SANA: Simulated Annealing far outperforms many other search algorithms for biological network alignment. *Bioinformatics*.
12. MILO, R., ITZKOVITZ, S., KASHTAN, N., LEVITT, R., SHEN-ORR, S., AYZENSHAT, I., SHEFFER, M. & ALON, U. (2004) Superfamilies of evolved and designed networks. *Science*, **303**(5663), 1538–1542.
13. MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. & ALON, U. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**(5594), 824–827.
14. NEWMAN, M. (2010) *Networks: An Introduction*. Oxford University Press.
15. NEYSHABUR, B., KHADEM, A., HASHEMIFAR, S. & ARAB, S. S. (2013) NETAL: a new graph-based method for global alignment of protein–protein interaction networks. *Bioinformatics*, **29**(13), 1654–1662.
16. PICARD, F., DAUDIN, J.-J., KOSKAS, M., SCHBATH, S. & ROBIN, S. (2008) Assessing the exceptionality of network motifs. *Journal of Computational Biology*, **15**(1), 1–20.
17. PRŽULJ, N. (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **23**(2), e177–e183.
18. RITO, T., WANG, Z., DEANE, C. M. & REINERT, G. (2010) How threshold behaviour affects the use of subgraphs for network comparison. *Bioinformatics*, **26**(18), i611–i617.
19. RUBNER, Y., TOMASI, C. & GUIBAS, L. J. (1998) A metric for distributions with applications to image databases. in *Sixth International Conference on Computer Vision*, pp. 59–66. IEEE.
20. SARAJLIĆ, A., MALOD-DOGNIN, N., YAVEROĞLU, Ö. N. & PRŽULJ, N. (2016) Graphlet-based characterization of directed networks. *Scientific reports*, **6**, 35098.
21. VILLANI, C. (2008) *Optimal transport: old and new*, vol. 338. Springer Science & Business Media.
22. WALE, N., WATSON, I. A. & KARYPIS, G. (2008) Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowledge and Information Systems*, **14**(3), 347–375.
23. WASSERMAN, S. & FAUST, K. (1994) *Social network analysis: Methods and applications*, vol. 8. Cambridge University Press.
24. WEGNER, A. E., OSPINA-FORERO, L., GAUNT, R. E., DEANE, C. M. & REINERT, G. (2017) Identifying networks with common organizational principles. *Journal of Complex Networks*.
25. WILSON, R. C. & ZHU, P. (2008) A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, **41**(9), 2833–2841.
26. YAVEROĞLU, Ö. N., MALOD-DOGNIN, N., DAVIS, D., LEVNAJIC, Z., JANJIC, V., KARAPANDZA, R., STOJMIROVIC, A. & PRŽULJ, N. (2014) Revealing the hidden language of complex networks. *Scientific reports*, **4**, 4547.