CSN Lab 5: Finding and assessing community structure

Clément Herbin & Sara Montese

November 23, 2023

1 Introduction

The goal of this lab is to run and compare different community finding algorithms. We will make use of R with the following packages *igraph* [CN21], *igraphdata* [Cs2] and *clustAnalytics* [Mir23]. We will apply these community detection algorithms to the following networks:

- Karate (*igraphdata*): Social network between members of a university karate club. The edge weights are the number of common activities the club members took part of.
- A synthetic network with scale-free degree distribution, 200 nodes, 800 edges and 4 communities.
- Enron (*igraphdata*): Email dataset made public by the U.S. Department of Justice.
- Immuno (*igraphdata*): Undirected and connected network of interactions in the immunoglobulin protein. It is made up of 1316 vertices representing amino-acids and an edge is drawn between two amino-acids if the shortest distance between their C_alpha atoms is smaller than the threshold value = 8 Angstrom.

To compare the different algorithms, we will evaluate the significance of clusterings produced by these algorithms applied on networks describe above. The evaluation will be a combination of the significance scoring function and the Jaccard similarity with respect to a good reference clustering. We will use these algorithms:

- Louvain (LV): community structure detection by multi-level optimization of modularity.
- Label Propagation (LP): community structure detection based on propagating labels. It works by labeling the vertices with unique labels and then updating the labels by majority voting in the neighborhood of the vertex.
- Walktrap (W): community structure detection via short random walks.
- Edge Betweenness (EB): community structure detection based on edge betweenness (the edge betweenness score of an edge measures the number of shortest paths through it).
- Leading Eigenvector (LE): community structure detection based on the leading eigenvector of the community matrix.

2 Results

This section presents the results of our evaluation on the significance of clusterings produced by different community detection methods applied to four different networks.

2.1 Karate

2.1.1 Clustering Significance Evaluation

Table 1 shows different significance evaluation metrics and the total score (Eq. 1) for each algorithm:

	Louvain	Label Propagation	Walktrap	Edge Betweenness	Leading Eigenvector	Ground Truth
expansion	3.47	3.00	3.47	5.76	3.65	1.29
conductance	0.26	0.23	0.25	0.42	0.29	0.10
norm cut	0.38	0.35	0.38	0.58	0.41	0.19
clustering coef	0.60	0.67	0.61	0.45	0.59	0.53
modularity	0.42	0.39	0.41	0.36	0.41	0.37
Total Score	-0.12	-0.05	-0.20	-0.82	-0.28	**-0.009**

Table 1: Different scoring functions for each algorithm on the Karate graph. Total scores are computed in order to do a comparison between the clusterings and the ground truth.

We find that the best score is the score of the ground truth and the second best score is the label propagation. In Table 2, we can see the local Jaccard similarity between ground truth and the different algorithms:

	GT-1	GT-2
LV-1	0.69	0.00
LV-2	0.31	0.00
LV-3	0.00	0.67
LV-4	0.00	0.33

(a) Louvain

	GT-1	GT-2
LP-1	0.81	0.00
LP-2	0.19	0.00
LP-3	0.00	0.78
LP-4	0.00	0.22

(b) Label Propagation

	GT-1	GT-2
W-2	0.00	0.72
W-3	0.69	0.00
W-1	0.31	0.00
W-4	0.00	0.28

(c) Walktrap

	GT-1	GT-2
EB-1	0.56	0.00
EB-2	0.11	0.10
EB-3	0.31	0.00
EB-4	0.00	0.56
EB-5	0.00	0.22
EB-6	0.00	0.11

(d) Edge Betweenness

	GT-1	GT-2
LE-1	0.62	0.00
LE-3	0.00	0.67
LE-2	0.31	0.00
LE-4	0.06	0.00
LE-5	0.00	0.33

(e) Leading Eigenvector

Table 2: Local Jaccard Similarity between ground truth and different clusterings algorithms on the Karate graph

We notice that the best local indeces are those with the label propagation algorithm, 0.81 and 0.78. These 2 clusters are very similar to the 2 clusters of the ground truth. If we look at the global Jaccard similarity in Fig. 3, we can observe that label propagation is the most similar to the ground truth.

Algorithm	J(Ground Truth, Algorithm)
Louvain	0.56
Label Propagation	0.67
Walktrap	0.59
Edge Betweenness	0.40
Leading Eigenvector	0.53

Table 3: Global Jaccard Similarity between different algorithms and the ground truth community of the Karate graph

2.1.2 Visualizations

Plots of the clusterings for each algorithm are shown in Fig. 1.

Figure 1: Plots of the ground truth and of different community finding algorithms on the Karate graph

2.2 Synthetic scale-free network

The network of study is composed of 200 nodes, 800 edges and 4 communities. All communities are equally likely, as indicated by the parameters B (matrix with all 1s) and p = (0.25, 0.25, 0.25, 0.25) of the Barabasi-Albert model. By setting the type parameter to *Hajek*, we are specifying that all nodes are equally likely to connect to each other. This leads to a network with no clear community structure. Our aim is to investigate the behaviour of community detection algorithms on homogeneous graphs.

2.2.1 Clustering Significance Evaluation

Table 4 shows different scoring functions metrics and the total score for each algorithm on the synthetic graph.

	Louvain	Label Propagation	Walktrap	Edge Betweenness	Leading Eigenvector	Ground Truth
expansion	4.40	0.00	4.17	5.28	5.02	6.03
conductance	0.55	0.00	0.54	0.70	0.63	0.75
norm cut	0.71	-	0.73	0.86	0.81	1.26
clustering coef	0.17	0.09	0.17	0.18	0.10	0.05
modularity	0.31	0.00	0.26	0.20	0.25	-0.01
Total Score	**-1.44**	-	-1.57	-1.97	-1.83	-2.98

Table 4: Different scoring functions for each algorithm on the synthetic graph. Total scores are computed in order to do a comparison between the various clusterings and the ground truth.

In the case of homogeneous graphs, Label Propagation struggles to accurately propagate labels and detect the underlying communities. We can also notice that the best score is not for the ground truth, but it is for the Louvain algorithm. Walktrap is not far away from the best score. Table 10 shows the local Jaccard similarity between ground truth and the differents algorithms (local Jaccard similarity are listed in the Appendix). Except for the Label Propagation algorithm, which have only 1 cluster, all the other algorithm have 9 or 10 clusters. Obviously, the best local index value is from Label propagation, since there is only 4 local index and so the minimum best value we can get is 0.25. For Louvain and Walktrap, we have some local index value around 0.2. So we can see in the Table 5 that the best global Jaccard similarity is Label Propagation which is therefore not surprising if we refer to the previous table.

Algorithm	J(Ground Truth, Algorithm)
Louvain	0.14
Label Propagation	0.28
Walktrap	0.15
Edge Betweenness	0.07
LeadingEigenvector	0.13

Table 5: Summary of the global Jaccard Similarity between different algorithms and the ground truth communities of the synthetic graph

2.2.2 Visualizations

Plots of the clusterings for each algorithm are shown in Fig. 2.

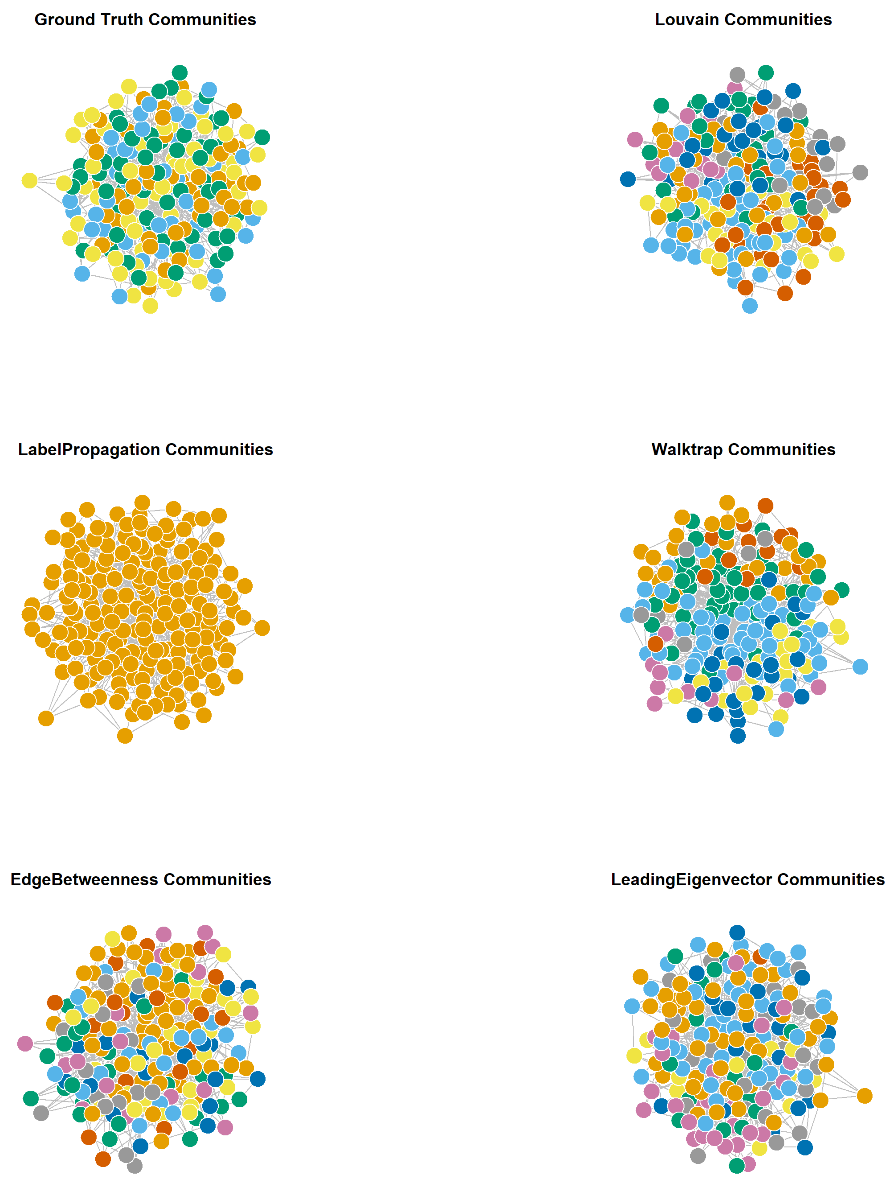

Figure 2: Plots of the ground truth and of different community finding algorithms on the synthetic graph

2.3 ENRON network

Before applying the algorithm to this network, we traced it, and we notice that there were a lot of self loops as well as vertices of degree 0. So we decided to remove these self loops and these vertices (2 vertices to be exact).

2.3.1 Clustering Significance Evaluation

We can see in Table 6, the score of different scoring functions and the total score for each algorithm:

	Louvain	Label Propagation	Walktrap	Edge Betweenness	Leading Eigenvector
expansion	240.33	323.11	272.50	680.75	310.46
conductance	0.21	0.30	0.24	0.60	0.26
norm cut	0.25	0.34	0.28	0.70	0.31
clustering coef	1.02	1.05	1.04	1.00	1.06
modularity	0.30	0.25	0.30	0.23	0.28
Total Score	**0.17**	-0.08	0.10	-0.83	0.06

Table 6: Different scoring functions for each algorithm on the ENRON graph. Total scores are computed to find the best ranking clustering.

The best score is for the Louvain clustering, followed by Walktrap and Leading Eigenvector. Since we don't have a ground truth, we will choose Louvain as a good reference clustering since it has the best score. From Table 11 to 14, we can see the local Jaccard similarity between Louvain and the others algorithms (see Appendix). Leading Eigenvector have a good local value, and also have only 1 more cluster than Louvain. Label Propagation and Walktrap have also good values but they present a higher number of clusters. Table 5 shows the global Jaccard similarity.

Algorithm	J(Louvain, Algorithm)
Label Propagation	0.64
Walktrap	0.65
Edge Betweenness	0.31
Leading Eigenvector	0.73

Table 7: Global Jaccard Similarity between different algorithms and the reference clustering of the ENRON graph.

The most similar algorithm to Louvain is Leading Eigenvector, followed by Walktrap and Label Propagation.

2.3.2 Visualizations

Plots of the clusterings for each algorithm are shown in Fig. 3.

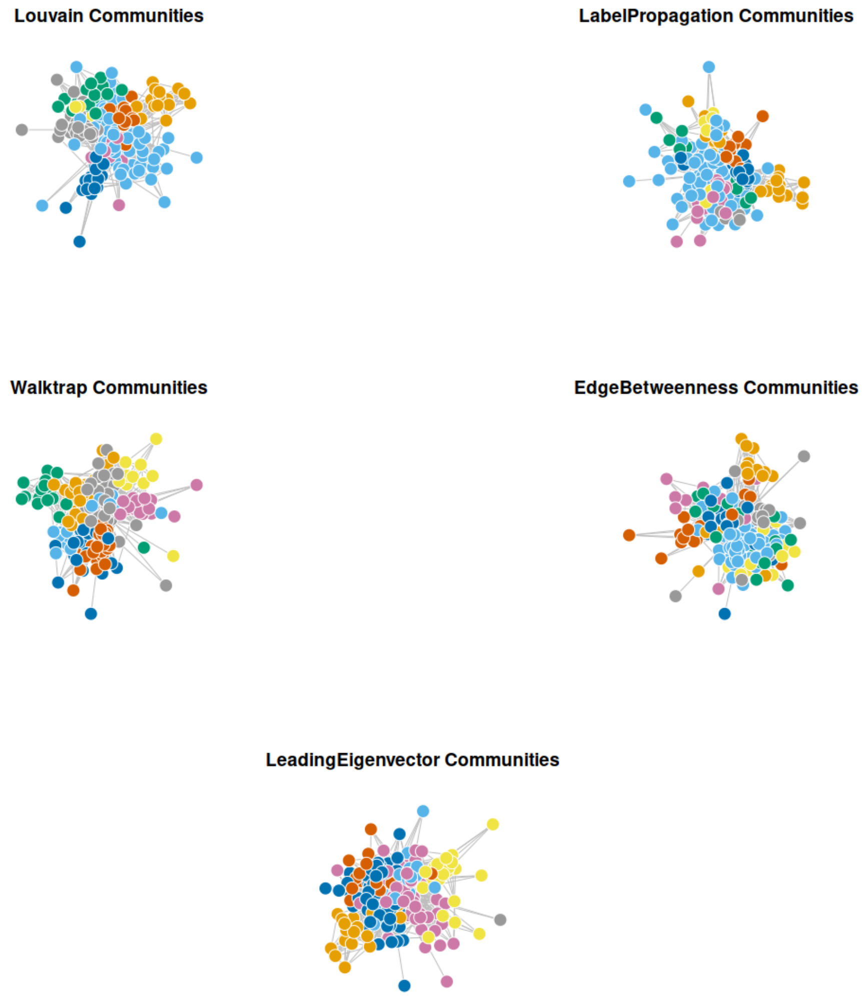

Figure 3: Plots of different community finding algorithms on the ENRON graph

2.4 IMMUNO network

2.4.1 Clustering Significance Evaluation

Table 8 shows different scoring functions and the total score for each algorithm:

	Louvain	Label Propagation	Walktrap	Edge Betweenness	Leading Eigenvector
expansion	0.43	1.90	0.42	0.44	0.51
conductance	0.04	0.20	0.04	0.05	0.05
norm cut	0.05	0.21	0.05	0.05	0.06
clustering coef	0.51	0.59	0.51	0.50	0.50
modularity	0.87	0.77	0.86	0.87	0.86
Total Score	**1.159**	0.66	1.14	1.150	1.12

Table 8: Different scoring functions for each algorithm on the IMMUNO graph. The maximum total score is computed to find the best ranking clustering.

The algorithm with the best score is Louvain, followed by Edge Betweenness and Walktrap. In the next table, we can see the global Jaccard similarity between different algorithms and the reference clustering, here Louvain:

Algorithm	J(Louvain, Algorithm)
Label Propagation	0.34
Walktrap	0.86
Edge Betweenness	0.95
Leading Eigenvector	0.89

Table 9: Global Jaccard Similarity between different algorithms and the reference clustering communities of the IMMUNO graph

We see that, Edge Betweenness, Walktrap and Leading Eigenvector are similar to Louvain clustering.

2.4.2 Visualizations

Plots of the clusterings for each algorithm are shown in Fig. 4.

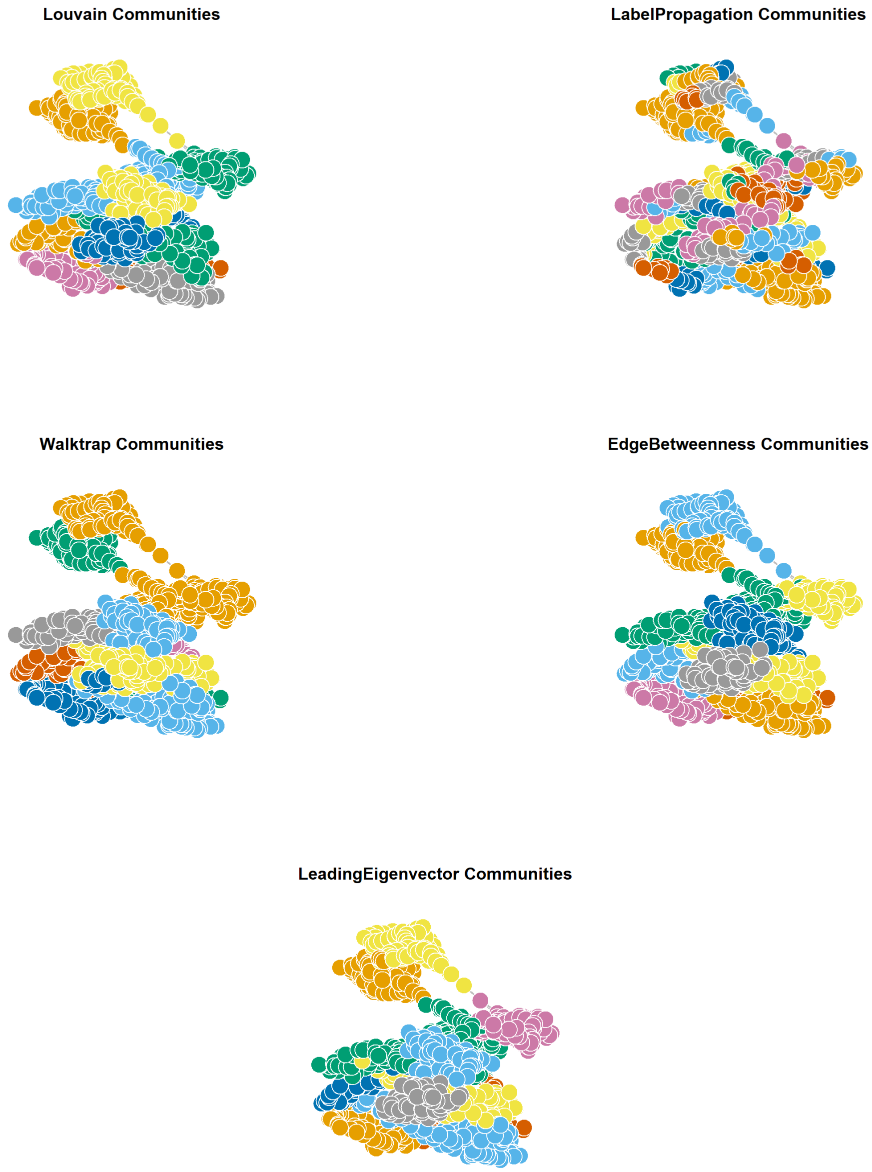

Figure 4: Plots of different community finding algorithms on the IMMUNO graph

3 Discussion

For the first two networks, to interpret the results, we can compare the scores of each algorithm with the ground truth and see which one is closest to it. The higher the score, the better the algorithm performs.
For Karate network, Table 1 the Label Propagation algorithm has the second highest total score, which means it is the most similar to the ground truth partition. It has the best score for the clustering coefficient, but in general all clusterings show a quite high value for this metric, meaning that there is a high degree of local connectivity. This confirms the origin of the dataset itself, since social networks often exhibit high clustering coefficients due to the presence of communities. The modularity of the graph, which is greater than 0 for all clusterings, suggests that the network has more edges within communities than one would expect in a random network. What mostly makes the ground truth stand out from other algorithms is the expansion. The lower value of expansion indicates that the communities of the clustering are well-connected internally (the majority of edges from vertices within the community are directed towards other vertices within the same community). This suggests a strong and

isolated communities. As confirmed from the visualizations 1, Edge Betweenness clustering has the highest value of expansion. Furthermore, the Edge Betweenness algorithm provides the lowest total score, which means it is the worst performer among the five algorithms. It has the worst scores for expansion, conductance, norm cut, and clustering coefficient. It also has a low modularity score, which means it may not find meaningful communities at all. Therefore, based on the results, the best algorithm for community detection is the Label Propagation algorithm, followed by the Louvain algorithm. Edge Betweenness algorithm is not recommended for this task. By looking at the global Jaccard similarity in Table 3, Label Propagation communities are the most similar to the ground truth.

For the Synthetic scale-free network, the best total score in Table 1 is held by the Louvain clustering, followed by the Walktrap clustering. We notice that the ground truth have the lowest total score, which is surprising at first. But as we said in the results section, we are specifying that all nodes are equally likely to connect to each other, which can explain why others algorithms may have better score than the ground truth. Louvain have the highest modularity, is not far from having the highest clustering coefficient and also has the lowest normalized cut. By looking at the global Jaccard similarity in Table 5, we notice that the detected communities are not similar to the ground truth. This is understandable, given the nature of the network. Communities in homogeneous graphs may not have well-defined boundaries, making it difficult for algorithms to separate nodes into distinct groups.

For ENRON network, if we look at the results in Table 6, we see that the Louvain algorithm has the highest total score, followed by Walktrap. Louvain have the lowest value for expansion, normalized cut and conductance and the highest for modularity but not the highest for clustering coefficient. Edge betweenness shows higher values for expansion, conductance and normalized cut, so the community are not very well-connected internally, clusters are less cohesive and the algorithm is bad for minimizing the dissimilarity between vertices within a cluster and maximizing the dissimilarity between vertices across different clusters. To return to the best algorithm, when we look at the best score, we see that the best algorithm for community detection is the Louvain, but Walktrap and Leading Eigenvector have also good results. Edge Betweenness has the worst results and is not recommended for this task. If we look at the global Jaccard similarity, Label Propagation, Walktrap and Leading Eigenvector clusters are similar to the cluster of the reference clustering, which is Louvain.
For IMMUNO network, based on the results in Table 8, the Louvain algorithm has the highest total score, with Edge Betweenness just behind of 0.009. Label propagation clustering shows higher values of expansion, conductance and normalized cut, compared to the other clusterings. For the expansion, it means that it performs worse at creating communities that don't overlap much with each others. For the conductance, suggesting that the clusters in this particular algorithm are less cohesive. The visualization in Fig. 4 confirms the above comment. The higher normalized cut may indicate that the partition produced by that algorithm is less desirable in terms of minimizing the number of edges between clusters relative to the size of the clusters. It is worth nothing that in each clustering the modularity ranges around 0.80, which suggests that the partition of the graph into communities is meaningful for every algorithm. Therefore, based on the results, the best algorithm for community detection for this network is the Louvain algorithm, but Walktrap, Edge Betweenness and Leading Eigenvector have similar results. Label propagation algorithm is not recommended for this task. With reference to the Global Jaccard Similarity, the results in 9 confirm the similarity between the clusterings of Walktrap, Edge Betweenness and Leading Eigenvector and the reference clustering (Louvain), suggesting that the detected communities are very similar.

One difficulty we encountered is the assurance of reproducibility of experiments. To ensure reproducibility, we set a seed before running the ensemble of clustering algorithms. This ensures that the random number generator produces the same sequence of numbers each time the code is run, leading to reproducible results (since most algorithms include randomness). Another idea to address this problem could be to run several experiments and provide the obtained results along with their margin of variability, but this required more computations.

4 Methods

To evaluate the significance of clusterings produced by different community detection methods, for each algorithm we select a combination of significance scoring functions. Particularly, for each algorithm i we define a new scoring function, $Score_i$, to maximize:

$$Score_i = 2modularity_i + clustering\_coefficient_i - \frac{expansion_i}{max(expansion)} - normalized\_cut_i - conductance_i \quad (1)$$

To quantity the quality of the community structure, we choose a linear combination of quality criteria based on internal connectivity, external connectivity, mixture of internal and external connectivity, and on a network model. Namely, in Eq. 1 we consider:

- Modularity: is a measure that evaluates the quality of a partition of a network into communities. It quantifies the density of edges within communities compared to the density of edges between communities. This score is multiplied by a factor of 2 since we want to give more importance to it, compared to other metrics. It has to be maximized.
- Clustering Coefficient: it measures the degree to which nodes in a graph tend to cluster together. It has to be maximized.
- Expansion: is a measure that evaluates how well a community is separated from the rest of the network. It has to be minimized. To ensure equal significance alongside other metrics, we have normalized its value over the maximum value of expansion of all the algorithms.
- Normalized Cut: is a criterion used for partitioning a graph into clusters. It aims to minimize the number of edges between clusters while normalizing for the sizes of the clusters. It has to be minimized.
- Conductance: is a measure that assesses the quality of a partition by considering the ratio of the number of edges leaving the community to the total number of edges incident to the community. It has to be minimized.

To enhance the decision quality regarding the reference network, we initially attempted to rewire the edges while maintaining a constant degree distribution. This was done to create a comparable network with uniformly distributed edges. The stability of a cluster under minor network perturbations was assessed using the *evaluate_significance_r* function. However, due to the extensive computational time required, we decided to forgo the edge rewiring process and instead utilized the *evaluate_significance* function. This approach allowed us to efficiently determine the significance of our network clusters.

References

[CN21] Gabor Csardi and Tamas Nepusz. *igraph: Network Analysis and Visualization*, 2021. R package version 1.2.6.

[Cs2] Gábor Csárdi. *igraphdata: A Collection of Network Data Sets for the 'igraph' Package*, 2022. R package version 1.0.1.

[Mir23] Martí Renedo Mirambell. *clustAnalytics: Cluster Evaluation on Graphs*, 2023. R package version 0.5.4.

5 Appendix

	GT-2	GT-4	GT-1	GT-3
LV-1	0.12	0.07	0.09	0.03
LV-2	0.07	0.08	0.14	0.22
LV-3	0.08	0.03	0.15	0.09
LV-4	0.05	0.10	0.10	0.08
LV-5	0.16	0.12	0.05	0.05
LV-6	0.12	0.13	0.07	0.08
LV-7	0.05	0.08	0.02	0.06
LV-8	0.03	0.09	0.07	0.09
LV-9	0.05	0.00	0.07	0.06

(a) Louvain

	GT-2	GT-4	GT-1	GT-3
LP-1	0.23	0.21	0.27	0.28

(b) Label Propagation

	GT-2	GT-4	GT-1	GT-3
W-2	0.06	0.03	0.09	0.12
W-3	0.15	0.11	0.14	0.16
W-9	0.11	0.22	0.13	0.11
W-5	0.08	0.07	0.01	0.13
W-1	0.06	0.08	0.16	0.05
W-4	0.05	0.02	0.08	0.05
W-10	0.08	0.04	0.03	0.03
W-6	0.06	0.02	0.02	0.02
W-8	0.04	0.00	0.03	0.05
W-7	0.00	0.04	0.02	0.03

(c) Walktrap

	GT-2	GT-4	GT-1	GT-3
EB-1	0.11	0.15	0.14	0.16
EB-2	0.02	0.00	0.03	0.05
EB-3	0.02	0.04	0.02	0.03
EB-4	0.02	0.07	0.02	0.03
EB-5	0.02	0.00	0.09	0.02
EB-6	0.00	0.02	0.00	0.03
EB-7	0.00	0.00	0.02	0.05
EB-8	0.02	0.05	0.02	0.00
EB-9	0.00	0.00	0.02	0.02
EB-10	0.00	0.05	0.00	0.00

(d) Edge Betweenness

	GT-2	GT-4	GT-1	GT-3
LE-1	0.15	0.08	0.07	0.07
LE-2	0.10	0.09	0.08	0.14
LE-8	0.05	0.05	0.09	0.07
LE-10	0.09	0.02	0.05	0.03
LE-3	0.05	0.17	0.04	0.06
LE-7	0.00	0.00	0.00	0.02
LE-9	0.06	0.06	0.11	0.15
LE-5	0.09	0.06	0.14	0.06
LE-4	0.06	0.11	0.11	0.09
LE-6	0.10	0.05	0.07	0.08

(e) Leading Eigenvector

Table 10: Local Jaccard Similarity between ground truth and different clustering algorithms on the synthetic graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11
LP-1	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-2	0.01	0.56	0.13	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00
LP-3	0.00	0.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-5	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00
LP-6	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.00
LP-7	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.00	0.00
LP-8	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
LP-10	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00
LP-11	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-12	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-13	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-14	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-15	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
LP-17	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-18	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00
LP-19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
LP-20	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00
LP-21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table 11: Local Jaccard Similarity between Louvain and Label Propagation clusterings on the ENRON graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11
W-1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-4	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-2	0.00	0.56	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.05
W-3	0.00	0.12	0.38	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
W-8	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.86	0.00	0.00	0.00
W-6	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
W-7	0.00	0.01	0.00	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00
W-11	0.00	0.01	0.00	0.50	0.00	0.16	0.00	0.00	0.00	0.00	0.00
W-5	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-10	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-9	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.00	0.00
W-12	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-13	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
W-15	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-16	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
W-18	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00

Table 12: Local Jaccard Similarity between Louvain and Walktrap on the ENRON graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11
EB-1	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-2	0.00	0.21	0.08	0.04	0.02	0.02	0.00	0.35	0.00	0.00	0.00
EB-3	0.00	0.01	0.26	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.08
EB-4	0.00	0.00	0.25	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
EB-5	0.00	0.12	0.00	0.05	0.06	0.00	0.22	0.00	0.00	0.00	0.00
EB-6	0.00	0.01	0.00	0.00	0.65	0.00	0.04	0.00	0.00	0.00	0.00
EB-7	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-8	0.03	0.05	0.02	0.08	0.00	0.46	0.03	0.00	0.00	0.00	0.00
EB-9	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
EB-10	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-11	0.00	0.11	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
EB-12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00
EB-13	0.00	0.03	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
EB-14	0.19	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
EB-15	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-16	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-17	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-18	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33
EB-19	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
EB-20	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00
EB-21	0.00	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-22	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-23	0.00	0.02	0.00	0.00	0.00	0.05	0.07	0.00	0.00	0.00	0.00
EB-24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
EB-25	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-26	0.00	0.00	0.00	0.00	0.05	0.00	0.08	0.00	0.00	0.00	0.00
EB-27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
EB-28	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
EB-30	0.00	0.00	0.04	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
EB-31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00

Table 13: Local Jaccard Similarity between Louvain and Edge Betweenness on the ENRON graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11
LE-1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
LE-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
LE-6	0.00	0.05	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.33
LE-8	0.00	0.05	0.00	0.00	0.03	0.00	0.65	0.00	0.00	0.00	0.00
LE-7	0.00	0.01	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.00
LE-5	0.00	0.03	0.00	0.00	0.81	0.00	0.00	0.02	0.00	0.00	0.00
LE-11	0.00	0.04	0.05	0.13	0.00	0.02	0.00	0.62	0.00	0.00	0.00
LE-4	0.00	0.01	0.80	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
LE-2	0.00	0.71	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
LE-3	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-12	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 14: Local Jaccard Similarity between Louvain and Edge Leading Eigenvector on the ENRON graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11	LV-12	LV-13
LP-1	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-2	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-3	0.01	0.27	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-4	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-5	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-6	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-7	0.00	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-8	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-9	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-10	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-11	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-12	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-13	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-14	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-15	0.00	0.00	0.25	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-16	0.00	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-17	0.00	0.00	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-18	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-19	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
LP-20	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-21	0.00	0.00	0.00	0.00	0.02	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-22	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-23	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-24	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-25	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LP-26	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00
LP-27	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00
LP-28	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00
LP-29	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
LP-30	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00
LP-31	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.38
LP-32	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00
LP-33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.00	0.00	0.02	0.00	0.00
LP-34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00
LP-35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00
LP-36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.00	0.00
LP-37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.14	0.00	0.00	0.00
LP-38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00
LP-39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00
LP-40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00
LP-41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
LP-42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.00	0.00
LP-43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00
LP-44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00
LP-45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00
LP-46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
LP-47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.01	0.00	0.00
LP-48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.00	0.00	0.00
LP-49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
LP-50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00
LP-51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
LP-52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.00
LP-53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00
LP-54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00
LP-55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00
LP-56	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.32
LP-57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20
LP-58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11	LV-12	LV-13
W-11	0.00	0.54	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
W-9	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81
W-3	0.00	0.00	0.00	0.00	0.00	0.01	0.84	0.00	0.00	0.00	0.00	0.00	0.09
W-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00
W-10	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.97	0.00	0.00	0.00
W-8	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.03	0.00	0.05	0.00	0.00
W-2	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.00

Table 16: Local Jaccard Similarity between Louvain and Walktrap on the IMMUNO graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11	LV-12	LV-13
EB-1	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-2	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-3	0.00	0.92	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-4	0.00	0.01	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-5	0.00	0.01	0.00	0.00	0.76	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00
EB-6	0.00	0.00	0.00	0.00	0.02	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EB-7	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00
EB-8	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.92
EB-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.02	0.00	0.00
EB-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.01	0.00	0.00	0.00
EB-11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.01	0.00	0.00
EB-12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00
EB-13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00

Table 17: Local Jaccard Similarity between Louvain and Edge Betweenness on the IMMUNO graph

	LV-1	LV-2	LV-3	LV-4	LV-5	LV-6	LV-7	LV-8	LV-9	LV-10	LV-11	LV-12	LV-13
LE-1	0.96	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.06	0.00	0.03	0.00	0.00
LE-7	0.00	0.88	0.02	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-4	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-6	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.77	0.06	0.00	0.00	0.00
LE-9	0.00	0.00	0.00	0.00	0.06	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-8	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LE-2	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.01	0.90
LE-5	0.00	0.00	0.00	0.00	0.00	0.01	0.89	0.00	0.00	0.00	0.00	0.00	0.00
LE-11	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00
LE-12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.83	0.00	0.00	0.00
LE-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.90	0.00	0.00

Table 18: Local Jaccard Similarity between Louvain and Leading Eigenvector on the IMMUNO graph