Network comparison

Nina Velikajne[a], **Jer Pelhan**[a], **and Jana Štremfelj**[a]

[a]University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, SI-1000 Ljubljana, Slovenia

The manuscript was compiled on May 26, 2023

Network comparison is crucial in various scientific and industrial domains, allowing evaluation and analysis of different aspects of the networks. In this paper, we provide an overview of multiple established methods for network comparison. The isomorphism-based network comparison relies on mapping one network into another and is an NP problem. Furthermore, it is not useful as the output is binary; yes the networks are the same or no, the networks are not the same. For this reason, different network comparison methods evolved. Graph edit distance measures the minimum number of operations required to transform one graph into another, but it is still an exponentially hard problem. Comparison by network fragments involves breaking down the network into smaller components, focusing the comparison on the local structural properties. Comparison by network distances is assessing the dissimilarity based on the structural or topological relations between the compared networks using measures such as Hamming distance, D-measure, Graph Edit Distance etc. In the last part, we present the comparison of networks using individual metrics, such as clustering coefficient, characteristic path length, degree distribution, etc. Single network comparison however provides useful insight into single aspects of similarities of networks, but no single metric can capture the complexity and diversity of network structures, thus we also present the statistical comparison over multiple metrics.

Problem definition, motivation, background, contributions etc. Network comparison is important in various fields of science and industry as it allows us to evaluate and analyze different networks, and systems based on specific criteria. Through the analysis of networks, we gain valuable insights into the intricate interactions within various systems that might otherwise remain unexplored. Several different methods and approaches have been developed in order to compare networks.

Comparing networks is frequently valuable (1, 2), especially when the properties of a specific network are known. This enables knowledge transfer by assessing the similarities or differences between two networks. For example, comparing two virus networks might be useful for finding vaccination (3). Comparing brain networks and their distances correlates with the IQ of a person (4). Furthermore, based on the triad significance profile (TSP) similarity the social networks and the WWW nets may be a part of a super-family. Network comparison is also useful in language analyses. Languages and their text structures from the same linguistic group tend to have similar networks (3). In the field of international economics, one might seek to compare the trade structures of different product categories. Similarly, in transportation, comparing the flight networks of various airlines is essential. Or in social media, examining and comparing the propagation cascades of news across different platforms (5). We can see that there are plenty of practical uses for network comparison.

However, due to computational limitations, comparing large networks directly is not feasible. As a result, researchers have established methods based on heuristics, such as the degree distribution, clustering coefficient, diameter, and relative graphlet frequency distribution. These metrics enable us to identify differences between networks by examining a list of properties. However, proving the similarity between the two networks is far more challenging and demanding. The similarity has to be shown across exponentially numerous properties. Herein, network comparison is a tough challenge and several methods have been established in order to evaluate the similarity (6).

The most primitive method is comparing networks by isomorphism. The method is based on mapping – isomorphism. Due to exponential space growth, the method is impractical for large networks (7). Herein, the researchers have been forced to establish new methods. Some methods focus on calculating specific distances between networks (see (8, 9)). Another group of methods is based on fragment analysis, which allows researchers to compare specific parts of networks (see (3, 6, 10)). And last, comparison by individual (see (11–13) or multiple metrics (see (1, 2)).

In this paper, we tend to describe and explain the theory behind the different established methods for network comparison. We tend to merge some of the most common approaches for network comparison in one place and allow a clear and transparent insight into the selected network comparison methods.

Comparison by isomorphism

In mathematics, isomorphism refers to mapping that preserves the underlying structure between two structures of the same type. Mapping can be reversed by inverse mapping. In terms of network comparison isomorphism is used to compare two networks in order to determine if they are structurally equivalent. It means finding an isomorphism between the nodes of the two networks, such that the adjacency relationships are preserved. Or in other words, two graphs are said to be isomorphic if it is possible to obtain one from the other by changing the node labels without affecting their topology (7). However, a full description of the differences between two large networks is unachievable since it requires solving the graph isomorphism problem (6).

Definition 1 (Graph isomorphism). *Given two graphs, $G1$ and $G2$ respectfully, an isomorphism of the two graphs is a bijection between the vertex sets of $G1$ and $G2$:*

$$f : V(G1) \longrightarrow V(G2),$$

such that any two vertices u and v, where $u, v \in V(G1)$, are adjacent in $G1$ if and only if $f(u)$ and $f(v)$ are adjacent in $G2$ (7).

For example, graphs $G1$ and $G2$ in Figure 1 are not isomorphic since the unique vertex of degree 5 in $G1$ is adjacent to a vertex of degree 2 which does not hold for $G2$ (14).

According to the Definiton 1 and the shown example, we can see that solving the problem requires exploring all possible mappings

NV wrote comparison by isomorphism and network fragments. JS wrote comparison by distances and edit distances. JP wrote comparison by individual and multiple metrics.

[1]To whom correspondence should be addressed. E-mail: js7437@student.uni-lj.si.

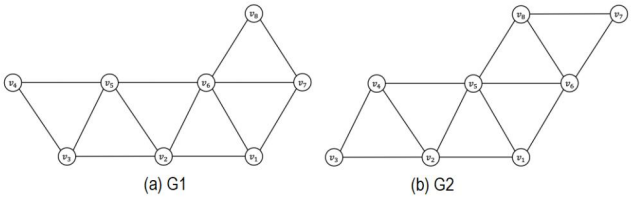

Fig. 1. Pair of non-isomorphic graphs from (14).

between the vertices of the two graphs and checking if the adjacency relationships are preserved. The search space grows exponentially with the size of the input graphs. Graph isomorphism is highly computationally complex. It requires exploring a large search space without any efficient known algorithm. Herein, graph isomorphism is an NP-complete problem (7).

Comparison using graph edit distance

Graph edit distance (GED) has been used as a similarity measure for representing the distances between attributed graphs. It defines the similarity of two graphs by the minimum amount of distortion, needed to transform one graph into the other. Unlike other measures, GED does not require any restrictions and can be applied to any type of graphs. Its exponential computational complexity in terms of the number of graph vertices is the main drawback of graph edit distance. The computation of GED is a NP-hard problem and would take unacceptable time of computation on large graphs (15). By a finite sequence of graph edit operations any graph can be transformed to another one and GED is defined by the least-cost edit operation sequence. For each operation a cost function is defined and the cost for this edit operation sequence is sum of costs for all operations in the sequence. The sequence of graph edit operations and its cost needed for transforming a graph is not unique, but the least cost is exclusive. Then sequence with the least cost is requested and its cost is the GED between the two graphs. The key issues are how to define the similarity of components in graphs and to determine costs of edit operations. Graph edit distance for attributed graphs is computed according to the attributes which depend on the algorithm. For the non-attributed graphs which only have information of connectivity structure, GED algorithms require conversion of graphs to strings and computation of edit distance for strings (16).

Comparison by network fragments

Comparison by fragments refers to analyzing and comparing networks by breaking them into smaller components or fragments. This allows a more focused examination of specific parts of the network and enables comparison based on the characteristics and behavior of such fragments. This approach is especially useful when comparing large networks, where other analyses (e.g., comparing by isomorphism) of the entire network may be infeasible and computationally expensive.

One such method was introduced by N. Pržulj (6) in 2007. The method is based on graphlets and their degree distributions. Graphlets are defined as the set of 30 non-isomorphic undirected graphs with at most 5 nodes. The degree distribution measures, for each value of k, the number of nodes of degree k. The degree distribution is then calculated for each graphlet, resulting in graphlet degree distribution (GDD). GDD measures the local structural properties of a network.

Definition 2 (GDD agreement). *Let G be a network, and d_G^j the j-th GDD. We scale $d_G^j(k)$ as*

$$S_G^j(k) = \frac{d_G^j(k)}{k}.$$

Next, we normalize the distribution as:

$$N_G^j(k) = \frac{S_G^j(k)}{\sum_{k=1}^{\infty} S_G^j(k)}.$$

Let H be another network, and N_H^j its normalized distribution. The distance between G and H is then defined as:

$$D^j(G,H) = \left(\sum_{k=1}^{\infty} [N_G^j(k) - N_H^j(k)]^2 \right)^{\frac{1}{2}}.$$

To get the j-th GDD agreement the distance is reversed as $A^j(G,H) = 1 - D^j(G,H)$. And the final agreement between the two networks is either the arithmetic or geometric mean of $A^j(G,H)$ over all graphlets j (6).

Aparício *et al.* (10) introduce a similar approach, only with changed distance calculation. Their reversed distance is defined as:

$$GDA^j(G,H) = 1 - \frac{1}{\sqrt{2}} \left(\sum_{k=1}^{\infty} [N_G^j(k) - N_H^j(k)]^2 \right)^{\frac{1}{2}}.$$

The final agreement is then calculated as the arithmetic mean over all $GDA^j(G,H)$ similar as in the (6). They modified the metric to only consider orbits that appear in at least one of the networks. Modifying the metrics was crucial, since the original agreement from (6) of two networks is increased even if the orbit frequency is zero in both networks. For bigger graphlets, it is likely that many of the possible orbits do not appear in either network. Herein, this may increase the original defined agreement.

Another approach was presented by Milo *et al.* (3). They developed an approach for comparing network fragments based on the significance profile (SP). In order to calculate the SP of a network, the network is compared to an ensemble of randomized networks with the same degree sequence.

Definition 3 (Significance profile (SP)). *Let G be a graph. For each subgraph i, statistical significance is described by the Z-score:*

$$Z_i = \frac{n_i^{real} - \overline{n}_i^{rand}}{std(n_i^{rand})},$$

where n_i^{real} is the number of times the subgraph appears in the network and $\overline{n}_i^{rand}$ and $std(n_i^{rand})$ are the mean and standard deviation of its appearances in the randomized network ensemble. The SP vector is the vector of Z scores normalized to length 1 as follows:

$$SP_i = \frac{Z_i}{\sqrt{\sum_i Z_i^2}}.$$

SP can be used in directed connected triads – triad significance profile (TSP). By calculating the correlation coefficient matrix of TSP for different networks, we can compare directed networks (3).

The described approach is also suitable for undirected networks. Undirected networks only have V and triangle types of triads. Herein, the authors analyse the connected tetrads. Instead of using Z-scores Δ_i is calculated as:

$$\Delta_i = \frac{n_i^{real} - \overline{n}_i^{rand}}{n_i^{real} + \overline{n}_i^{rand} + \epsilon},$$

where ϵ (i.e., $\epsilon = 4$) ensures that $|\Delta|$ is not too large if the subgraph is not frequent in both real and random networks. Subgrah ratio profile (SRP) is then the vector of Δ_i normalised to length of 1 as in Definiton 3 (3).

Comparison by network distances

Identifying and quantifying dissimilarities between graphs is a challenging problem, where the similarity or dissimilarity is determined on the structural or topological characteristics of graphs. There are several distance metrics that can be used for comparison, but especially for large networks it is important, that the comparison is not limited to extract only partial information and that it is not computationally too demanding.

A. D-measure. Schieber et al. (8) introduced D-measure on a simple networks example as seen in Figure 2. All three networks N1, N2 and N3 consists out of 9 nodes and 9 links. N1 presents one connected component, N2 has just one isolated node and N3 consists of 3 connected components. A good measure should give a low distance value between N1 and N2 and the highest of all pairs to N1 – N3. Between N2 and N3 the value should still be quite high, but lower than between N1 – N3. In Table 1 are the dissimilarity distances. The Hamming distance – H and the graph edit distance – GED do not find topological differences, returning same value for all three network examples.

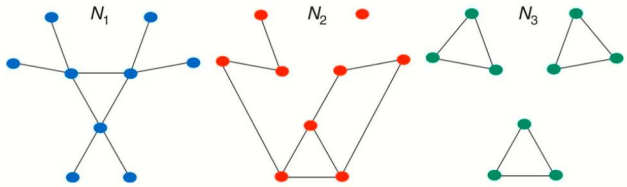

Fig. 2. Different networks with the same number of nodes and links from (8).

Table 1. Comparison between dissimilarity distances (H–Hamming distance; GED–graph edit distance; D–dissimilarity) from (8).

Networks	H	GED	D
(N1, N2)	12	6	0.252
(N1, N3)	12	6	0.565
(N2, N3)	12	6	0.473

We can start defining the D-measure with Network Node Dispersion – NND, which is a measure of the heterogeneity of a graph G in terms of connectivity distances. The network qualifies as heterogeneous when it possesses high diversity of node-distance patterns, therefore, high NND. The vectors, containing PDF values, are used, for precise comparison of networks.

Definition 4 (Node-distance distribution). *The distance distribution in each node i, $P_i = p_i(j)$, with $p_i(j)$ being the fraction of nodes, that are connected to node i at distance j. The set of N node-distance distributions, $P_1 ... P_N$, contains detailed information of the topology of the network, in a compact way. A network with N nodes, the set of N distance distributions $P_1 ... P_N$, is normalized by log(d+1), where d is the network's diameter. The Network Node Dispersion – NND, the Jensen–Shannon divergence – $\mathcal{J}$ and the average of the N distributions – μ_j are defined as follows:*

$$NND(G) = \frac{\mathcal{J}(P_1, ..., P_N)}{log(d+1)}$$

$$\mathcal{J}(P_1, ..., P_N) = \frac{1}{N} \sum_{ij} p_i(j) log(\frac{p_i(j)}{\mu_j})$$

$$\mu_j = \frac{\sum_{n=1}^{N} p_i(j)}{N}$$

However, most k-regular networks (graphs in which all nodes have degree k) possess NND=0. To define a general dissimilarity measure, it is important to properly discriminate them. That is why it is also taken into account the difference between the graphs averaged node-distance distributions (network's distance distribution), μ_G and μ'_G, and the comparison between the α–centrality values of the graphs and their complements, computed through the Jensen–Shannon divergence ($\mathcal{J}$).

Definition 5 (Dissimilarity measure). *The dissimilarity measure is defined as follows:*

$$D(G, G') = w_1 \sqrt{\frac{\mathcal{J}(\mu_G, \mu_{G'})}{log2}} + w_2 \left| \sqrt{NND(G)} - \sqrt{NND(G')} \right|$$

$$+ \frac{w_3}{2} \left(\sqrt{\frac{\mathcal{J}(P_{\alpha G}, P_{\alpha G'})}{log2}} + \sqrt{\frac{\mathcal{J}(P_{\alpha G^c}, P_{\alpha G^{c'}})}{log2}} \right)$$

where N is the size of G and M the size of G'. G^c indicates the complement of G, $w_1, w_2 and w_3$ are the arbitrary weights such that $w_1 + w_2 + w_3 = 1$.

As the NND is always less than 1 and $\mathcal{J}(P_G, P_{G'})/log2 \leq 1$ then, $0 \leq D(G, G') < 1$. D-measure captures global and local graphs dissimilarities, The first part compares averaged connectivity node's patterns, corresponding to the graph distance distribution. The second term analyses the heterogeneity of the nodes and the third part considers the centrality of each node, taken into account each node's direct and indirect connectivity span.

D(G, G') equals to zero, only if G and G' have the same graphs distance distribution, the same NND and the same α–centrality vector. However, there is no guarantee that D returns a non-zero value for all non-isomorphic networks.

B. Network portraits. Network portraits were introduced by Bagrow et al. in 2008 (9) to visualize and encode structural properties of a given network. The network portrait $B_{l,k}$ is the array with (l,k) elements and includes the number of nodes who have k nodes at distance l. The limitation for l is $0 \leq l \leq d$ and for k $0 \leq k \leq N-1$, where distance is taken as the shortest path length and d is the graph's diameter. The portrait remains identical, regardless how the nodes are ordered or labeled. The matrix encodes many structural graph features. The zeroth row stores the number of nodes N in the graph $B_{0,k} = N\delta_{k,1}$ and the first row captures the degree distribution $B_{1,k} = NP(k)$, as the neighbors are at the distance l=1. Second row captures the distribution of next-nearest neighbors etc. M is the number of edges and $\sum_{k=0}^{N} kB_{1,k} = 2M$. The diameter d of the graph is $d = max l | B_{l,k} > 0; for k > 0$. The number of shortest paths of length l is $\frac{1}{2} \sum_{k=0}^{N} kB_{l,k}$. One of most important properties of portraits is that they are graph invariant.

Definition 6 (Graph invariant). *A graph invariant is a property of a graph that is invariant under graph isomorphism, i.e., it is a function f such that f(G)=f(H) whenever G and H are isomorphic graphs.*

Instead of directly comparing graphs G and G', the portraits of the graphs B and B' can be computed, on which the comparison can be made. For each portrait B, the matrix C consisting of row-wise cumulative distributions of B, is computed $C_{l,k} = \sum_{j=0}^{k} B_{l,j} / \sum_{j=0}^{N} B_{l,j}$.

The metric-like graph comparison is performed by row-wise Kolmogorov-Smirnov test statistic K_l between corresponding rows l in C and C' $K_l = max|C_{l,k} - C'_{l,k}|$. If the two graphs have different diameters, the portrait for the smaller diameter graph can be expanded to the same size as the larger diameter graph by defining empty shells l<d as $B_{l,k} = N\delta_{o,k}$. Lastly, the aggregation of the test statistics for all pairs of rows is done, using a weighted average to define the similarity between G and G', $\Delta(G, G') = \Delta(B, B') = \frac{\sum_l \alpha_l K_l}{\sum_l \alpha_l}$, where α_l is a weight chosen to increase the impact of the lower, more heavily occupied shells $\alpha_l = \sum_{k>0} B_{l,k} + \sum_{k>0} B'_{l,k}$.

In the article from 2019 (17) the authors introduced an improved network comparison based on portraits. The rows of B matrix may be interpreted as probability distributions, where

$P(k|l) = \frac{kB_{l,k}}{\sum_c n_c^2}$ is the probability of choosing the node, that has k nodes at a distance l and n_c is the number of nodes within connected component c, the $\sum_c n_c^2$ runs over the number of connected components, and the n_c satisfy $\sum_c n_c = N$.That leads to an immediate comparison per row for two portraits: $KL(P(k|l)||Q(k|l)) = \sum_{l=0}^{max(d,d')} P(k|l) log_2 \frac{P(k|l)}{Q(k|l)}$, where KL(p||q) is the Kullback-Liebler (KL) divergence between two distributions p and q, and Q is defined as $Q(k|l) = \frac{kB'_{l,k}}{\sum_c' n_c'^2}$. The definition also holds for directed and weighted networks. The network portrait is a powerful summary of the topological features of the graph, such as the number of nodes and edges, the distribution of the next-nearest neighbours, the degree distribution, and the number of shortest paths of length l that can be recovered from B.

Definition 7 (Network Portrait Divergence). *The Network Portrait Divergence $D_{JS}(G, G')$ between two graphs G and ' is the Jensen-Shannon divergence as follows, $D_{JS}(G, G') = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$ where $M = \frac{1}{2}(P + Q)$ is the mixture distribution of P and Q.*

The Network Portrait Divergence $0 \leq D_{JS} \leq 1$ provides a single value to quantify the dissimilarity of the two networks by means of their distance distributions, with smaller D_{JS} for more similar networks and larger D_{JS} for less similar networks. Unlike the KL divergence, D_{JS} is symmetric, $D_{JS}(G, G') = D_{JS}(G', G)$ and $\sqrt{D_{JS}}$ is a metric.

Network Portrait Divergence is relatively computationally efficient, unlike graph edit distance measures, due to the fast that it is based on a graph invariant and expensive optimizations such as "node matching" are not needed.

Direct comparison of individual metrics

One of the most common approaches when measuring the similarity between networks is to use a global metric on both networks and compare how similar the result for both is. In order to compare the network structures, several metrics have been developed to measure various properties of the network. For starters the obvious metrics for measuring the similarity of networks are number of nodes and number of links in the network, largest connected component, average and maximal degree, and diameter of the network, *i.e.* longest shortest path between any two nodes in the network.

Definition 8 (Characteristic Path Length L(p)). *The characteristic path length is a metric that measures the typical separation between two vertices in the graph. The equation of the characteristic path length is as follows:*

$$L(p) = \frac{1}{N(N-1)} \sum_{i \neq j} d(i, j)$$

where N is the total number of vertices in the network, d(i, j) is the shortest path between vertices i and j, and $\sum_{i,j}$ sums over all pairs of vertices in the network.

The characteristic path length L(p) is a metric used by Watts & Strogatz (11). It measures the average shortest path over all pairs of vertices, and is therefore a measure of a global property. The L(p) is an indicator of how well the network is interconnedted. Less interconnedted network reflects a larger L(p), *e.g.* a chain network. This property reflects the efficacy of the (information) transmission through the network.

Definition 9 (Clustering Coefficient C(p)). *The clustering coefficient is a metric that measures the cliquishness of a typical neighborhood. The equation of the clustering coefficient is as follows:*

$$C(p) = 3 * (\#\ of\ triangles)/(\#\ of\ triplets)$$

where number triangles equals to number of closed triplets in the graph, and number of triplets equals to both number of closed and open triplets

Clustering coefficient C(p) is another metric used by Watts & Strogatz (11) and it reflects the degree of connectivity between the neighbouring nodes. Low clustering coefficient suggests sparsely connected neighbouring nodes.

Definition 10 (Degree distribution P(k)). *Network node degree distribution is defined as the probability that a randomly selected node in the graph has degree k, i.e., number of edges incident to it. Formally it is defined as follows:*

$$P(k) = \frac{n_k}{n}, \quad [1]$$

where n_k is the number of nodes with degree k in the graph and n is the total number of nodes in the graph.

The degree distribution (12) of a network reveals the network connectivity properties. It provides information about the prevelance of nodes with low or high degrees, as well as heterogenity of the node degrees. The degree distribution in a random network (ER) follows a Poisson distribution, while in a scale-free network, the degree distribution follows a power-law distribution, *i.e.*, $P(k) \sim k^{-\gamma}$. For two networks to be similar, they need to have the same degree distribution, with similar γ coefficients.

Definition 11 (Degree distribution P(k)). *Network node degree distribution is defined as the probability that a randomly selected node in the graph has degree k, i.e., number of edges incident to it. Formally it is defined as follows:*

$$P(k) = \frac{n_k}{n}, \quad [2]$$

where n_k is the number of nodes with degree k in the graph and n is the total number of nodes in the graph.

The degree distribution (12) of a network reveals the network connectivity properties. It provides information about the prevelance of nodes with low or high degrees, as well as heterogenity of the node degrees. The degree distribution in a random network (ER) follows a Poisson distribution, while in a scale-free network, the degree distribution follows a power-law distribution, *i.e.*, $P(k) \sim k^{-\gamma}$. For two networks to be similar, they need to have the same degree distribution, with similar γ coefficients.

Definition 12 (Assortativity and Dissasortativity). *Assortativity and disassortaty is a single measure that describes the tendency of nodes in a network to connect to other nodes of the network with similar or different characteristics, respectively.*

Network can be degree assortative, if nodes with similar degrees connect to each other. Conversely, in a degree dissasortative network (13), nodes with different degrees tend to be connected to each other, *i.e.*, the difference between the degrees of average nodes is substantial.

Definition 13 (Pearson Degree Mixing Coefficient).

$$r = \sum_{kk'} \frac{kk'(e_{kk'} - q_k q'_k)}{\sum_k (k^2 - q_k) - (\sum_k k q_k)^2} \quad [3]$$

where q_k is the neighbour excess distribution, that is $\frac{(k+1)p_{k+1}}{<k>}$, k and k' are the degrees of nodes. The Pearson correlation coefficient r ranges from -1 to 1, where values close to 1 indicate degree assortativity and values close to -1 indicate degree disassortativity. A value of aproximately 0 indicates no correlation.

Definition 14 (Strongly Connected Component). *Strongly connected component in a directed graph is a maximal subset of network vertices, in which each vertex can be reached from every other vertex in this subset by following a directed path.*

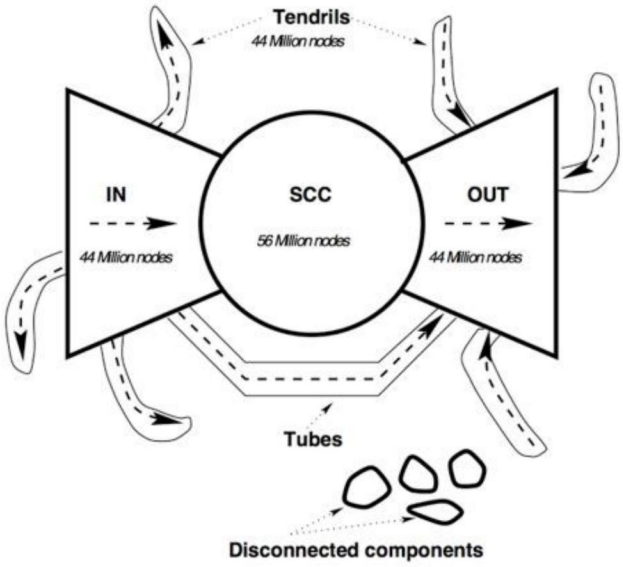

Fig. 3. Bow-tie network decomposition from (18).

Definition 15 (Bow-tie Decomposition). *Bowtie network decomposition decomposes the network to following subsets:*

- ***In-component:*** *The subset of nodes that can reach SCC.*
- ***Out-component:*** *The subset of nodes that can reached from the nodes in the SCC.*
- ***SCC:*** *The strongly connected component of the network.*
- ***Tendril:*** *The subsets of nodes that can be reached the In-component or that can reach the Out-component.*
- ***Tubes:*** *The subsets of nodes that creates a passage from the In-component to Out-component without touching the SCC.*
- ***Disconnected components:*** *The subsets of nodes that do not connect to none of listed subsets of nodes.*

The bow-tie decomposition was introduced by Broder (18) and is visualized in Figure 3

Statistical comparison over multiple metrics

Single metric comparisons can provide useful insight into certain aspects of a network structure. However, no single metric can fully capture the complexity, diversity and similarity of network structures. Solely relying on a single metric can lead to wrong conclusions about the network similarities or dissimilarities. Šubelj *et al.* (1, 2) propose a statistical comparison of networks over 21 individual graph statistics which are not independant on each other. At first, metrics are externally studentized, *i.e.*, statistically significant inconsistencies in individual graph statistics are identified over all comparable networks. Pairwise independence of the selected statistics is verified using Fisher z-transformation and z-tests. The Friedman rank test is used to confirm significant inconsistencies between bibliographic databases, and the Nemenyi post-hoc test is applied to identify databases with no statistically significant differences. The results are presented using mean ranks and critical difference diagrams. Ranking is a preferred method over simple summation while comparing networks over multiple statistics because it ensures a fair and unbiased comparison. When different networks are compared over multiple statistics, it is likely that some networks will be more similar on some statistics while others will be more similar on other statistics.

Discussion

The paper addresses the challenge of effectively comparing graphs, specifically in the context of larger networks. It presents various approaches and discusses the associated challenges related to graph comparison and comparison metrics. The paper includes network comparison methods like isomorphism and graph edit distance, which are known as NP-problems, as well as more efficient approaches such as comparison by fragments and distances. Additionally, the paper explores the comparison of individual metrics and multiple metrics. To enhance understanding, relevant examples are provided, and comprehensive definitions are included to provide a better overview of the comparison process. The authors allocated the work equitably with the assistance of an online random wheel, thereby asserting equal contribution to the paper.

1. Lovro Šubelj, Dalibor Fiala, and Marko Bajec. Network-based statistical comparison of citation topology of bibliographic databases. *Scientific reports*, 4(1):1–10, 2014.
2. Lovro Šubelj, Marko Bajec, Biljana Mileva Boshkoska, Andrej Kastrin, and Zoran Levnajić. Quantifying the consistency of scientific databases. *PloS one*, 10(5):e0127390, 2015.
3. Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303 (5663):1538–1542, 2004.
4. Martijn P Van Den Heuvel, Cornelis J Stam, René S Kahn, and Hilleke E Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29(23):7619–7624, 2009.
5. Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. Comparing methods for comparing networks. *Scientific Reports*, 9(1):17557, Nov 2019. . URL https://doi.org/10.1038/s41598-019-53708-y.
6. Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007.
7. Ronald C Read and Derek G Corneil. The graph isomorphism disease. *Journal of graph theory*, 1(4):339–363, 1977.
8. Tiago Schieber, Laura Carpi, Albert Diaz-Guilera, Panos Pardalos, Cristina Masoller, and Martin Ravetti. Quantification of network structural dissimilarities. *Nature Communications*, 8: 13928, 01 2017. .
9. James P Bagrow, Erik M Bollt, Joseph D Skufca, and Daniel Ben-Avraham. Portraits of complex networks. *Europhysics letters*, 81(6):68004, 2008.
10. David Aparício, Pedro Ribeiro, and Fernando Silva. Network comparison using directed graphlets. *arXiv preprint arXiv:1511.01964*, 2015.
11. Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.
12. Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
13. Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.
14. Furqan Aziz, Mian Saeed Akbar, Muhammad Jawad, Abdul Haseeb Malik, M Irfan Uddin, and Georgios V Gkoutos. Graph characterisation using graphlet-based entropies. *Pattern Recognition Letters*, 147:100–107, 2021.
15. Zhiping Zeng, Anthony KH Tung, Jianyong Wang, Jianhua Feng, and Lizhu Zhou. Comparing stars: On approximating graph edit distance. *networks*, 5(6):15, 2009.
16. Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. A survey of graph edit distance. *Pattern Analysis and applications*, 13:113–129, 2010.
17. James P Bagrow and Erik M Bollt. An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, 4(1):1–15, 2019.
18. Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1-6):309–320, 2000.