Applied Network Science

RESEARCH Open Access

Inflow and outflow centrality: novel centrality metrics inspired by graph convolution

Aram Papazian[1] and Volkhard Helms[1*]

*Correspondence:
Volkhard Helms
volkhard.helms@bioinformatik.uni-saarland.de
[1]Center for Bioinformatics, Saarland University, Saarland Informatics Campus, PO Box 15 11 50, 66041 Saarbrücken, Germany

Abstract
Centrality metrics quantify a node's importance within a network based on a node's connectivity, path position, proximity to other nodes, or influence from neighbors. All of these properties are influenced by the network structure and do not consider a node's features. To overcome this, two novel centrality metrics, termed inflow and outflow centrality, were introduced here. The metrics were derived from the aggregation approach used in graph convolutional networks, which allow for direct incorporation of node features with graph structure. The metrics were contrasted against the unweighted betweenness centrality and four node-weighted centrality metrics, weighted-degree, weighted-closeness, personalized PageRank, and alpha centrality, for an airport, an airplane trade, and a protein-protein interaction network. By emphasizing the contribution of otherwise little connected neighbor nodes, the new metrics prioritize nodes that are crucial to maintain a graph's connectivity.

Keywords Centrality measure, Graph convolution, Weighted network, Node features, Weighted centrality

Introduction

In graph theory and network analysis, centrality metrics are utilized to rank nodes based on network positions, connectivity, and other structural properties, these measures remain invariant under isomorphic transformations. Some relevant applications of centrality include identifying individuals of interest within a social network, key infrastructure nodes in urban networks, and analysis of activity observed in brain networks (Heuvel and Sporns 2013, Saberi et al. 2021). Among the most popular centrality metrics are degree, closeness, betweenness, eigenvector, and PageRank centrality (Newman 2010): degree centrality of a node describes the number of direct links a node has to other nodes in a graph, while closeness centrality entails the average length of the shortest path between a node and all other nodes in a graph, betweenness centrality measures each incidence of a node's presence in the shortest path between two other nodes in the graph, eigenvector centrality computes the influence of a node in a graph, given that connections to high-scoring nodes contribute more to node score than equal or low-scoring nodes, and lastly, PageRank assigns each node an importance score based on

the importance of nodes linking to it, computed iteratively with probability based link-following and random teleportation to ensure convergence.

All aforementioned metrics above were initially defined for unweighted graphs; however, in most applicable real-world scenarios, each node and edge within a graph typically differ from one another in terms of respective features. Namely, in protein-protein interaction networks, each protein, represented by a node, has a unique expression value and each interaction, portrayed as an edge, has a different confidence score or interaction strength. To address this, some studies have extended the centrality metrics to consider edge weights in determining the topological significance of nodes (Opsahl et al. 2010), while still assuming similar features for each node. This approach was implemented by Singh et al., who introduced multiple extensions of the aforementioned centrality metrics to consider distinct node features (Singh et al. 2020).

With the emergence of deep learning and representation learning, various approaches have been developed to tackle salient graph-based challenges, most notably with graph neural networks. The key design component of graph neural networks is the incorporation of message passing, whereby nodes iteratively update their representations by exchanging information with their neighbors (Wu et al. 2022, Scarselli et al. 2008, Micheli 2009). Alternative modifications of this technique have also been introduced. Graph convolutional networks (GCNs) use convolutional layers to propagate information between nodes in a graph; the GCNs aggregation process is based on a weighted average of node features, where weights are determined by graph structure (Kipf 2016). Unlike GCNs, graph attention networks (GANs) aggregate neighbor information in an adaptive manner using an attention mechanism to glean the importance of each neighbor and aggregate accordingly (Veličković et al. 2017). These models can be implemented to propose graph-based predictions, such as drug target identification (Zhai et al. 2023), predicting node labels, and inferring putative edges between two nodes.

In this work, the GCNs aggregation approach was utilized as motivation to introduce new centrality metrics, termed inflow and outflow centrality, that are based on graph structure rather than learned weights. Additionally, these novel centrality metrics allow for direct incorporation of node features. It is demonstrated that our new metrics provide useful insight into node-weighted networks from three different disciplines and that these metrics are not biased towards highly connected nodes. Additionally, these novel metrics have been made available in the following Github repository: https://github.com/AramPapaz/Inflow-and-Outflow-Centrality.

Materials and methods

Formula

The formulas for the new metrics were inspired by the normalization approach applied on the adjacency matrix in a graph convolutional operation, which originates from the spectral formulation based on the normalized graph Laplacian (Chung 1997,Kipf 2016). The normalized Laplacian is:

$$L = I - D^{\frac{-1}{2}} A D^{\frac{-1}{2}}$$

Here, A, D and I represent the adjacency, diagonal degree and identity matrices, respectively. $D^{\frac{-1}{2}} A D^{\frac{-1}{2}}$ is used to prevent high-degree nodes from dominating. GCNs

use this exact normalization but apply it directly to the adjacency matrix (with self-loops) to create the propagation matrix

$$A' = D^{\frac{-1}{2}}(A+I)D^{\frac{-1}{2}}$$

with,

$$A'_{i,j} = \begin{cases} \frac{1}{\sqrt{d_i d_j}}, & if\ j \in N(i) \\ 0, & otherwise \end{cases}$$

where, A' represents the normalized adjacency matrix. The variables d_i and d_j signify the degrees of nodes i and j. $N(i)$ represents the set of node i's neighbors. This procedure modifies the adjacency matrix by considering the degrees of nodes i and j. Node i will have low values if it has many neighbors or if its neighbor node j has many neighbors. This boosts the impact of nodes with few connections relative to highly connected nodes.

Two different flow centrality metrics were derived from the normalized adjacency matrix. First, inflow centrality considers the neighbor features inside the aggregation function. Second, outflow centrality takes into account solely the target node's feature during aggregation. In both formulas, the result of the aggregation is normalized by the square root of the target node's degree, termed as relative normalization, plus the median of all node degrees in the graph, termed as global normalization. The median was selected over the mean, as it is less sensitive to highly connected nodes. The formulas for both centrality metrics are as follows:

$$\text{Inflow metric}: \frac{1}{\sqrt{d_i} + median(G)} \sum_{j \in N(i)} \frac{feature_j}{\sqrt{d_i d_j}}$$

$$\text{Outflow metric}: \frac{feature_i}{\sqrt{d_i} + median(G)} \sum_{j \in N(i)} \frac{1}{\sqrt{d_i d_j}}$$

Here, d_i and d_j stand for the degrees of nodes i and j, respectively. The variable $median(G)$ represents the median node degree of all nodes in the graph, $feature_i$ and $feature_j$ represent the node features of nodes i and j, respectively, and $N(i)$ represents the set of neighbors of node i.

Our metrics will capture nodes that act as bridges between weakly and highly connected nodes. Weakly connected neighbor nodes with comparably large features values will contribute strongly to the summation term in the formula. Also, neighboring hub nodes with large feature values (highly populated cities, very wealthy countries, etc...) will contribute favorably.

The values obtained from both metrics are bounded below by 0 and unbounded above, as long as node features take positive values. These values are sensitive to the node features; therefore, they can be used to compare nodes within the same graph, but not between different graphs. Optionally, a final step could be added where all IM and OM values are divided by the largest value obtained for a network. This would normalize IM and OM values to the interval [0,1].

Simple graph models

Two types of graphs were created using the Barabasi Albert and Erdos Renyi algorithms which generate scale-free and random networks, respectively. Each network was set to contain 10 nodes with node characteristics that were sampled from a uniform distribution between 1 and 10. The node features are similar in both networks.

Node-weighted airport network

First, the new metrics were applied to an undirected airport network, where nodes represent airports and an edge exists between two nodes if there exists a direct flight between them. The log2 transformed population of each city was used as a (normalized) node feature. Airline route data was retrieved from the Open Flight Database (2025) and city population was retrieved from the World Cities Dataset (2019), which was last updated in 2019. If a city possesses multiple airports, their connections were concatenated and they were considered as a single node. Both tables are provided in the supplementary materials. This network is composed of 1,306 nodes and 11,079 edges.

Node-weighted airplane trade network

Next, the new metrics were applied to an undirected airplane trade network, where each node represents a country and an edge signifies airplane trade history between two nations. These trades include both newly built and used airplanes. The log2 transformed gross domestic product (GDP) of a country was considered as a (normalized) node feature. Bilateral airplane trade data between countries in 2023 was retrieved from CEPII (Mayer et al. 2023) and country GDP in 2023 was retrieved from the World Bank Group (2025). The tables used for our analysis are provided in the supplementary materials. This network is composed of 163 nodes and 1,161 edges.

Node-weighted breast cancer protein-protein interaction network

An undirected protein-protein interaction network (PPIN) was formed such that each node represents a protein and an edge is constituted between two proteins if there exists an interaction score greater than or equal to 500 in the STRING database (Mering et al. 2003). As a node feature of individual proteins, the mean RNA expression of the genes encoding these proteins in 1218 breast cancer samples from TCGA (Tomczak et al. 2015) was used. These values were increased tenfold to circumvent obtaining negative node features. The preprocessed expression data are provided in supplementary materials. This network contains 15,882 nodes and 453,242 edges.

Node weighted centrality metrics

The metrics proposed in this study were compared against the unweighted betweenness centrality and against the node-weighted degree centrality metric proposed by Singh et al. (Singh et al. 2020):

$$\frac{\sum_{v \in V \setminus \{u\}} (W_v . a_{u,v})}{\sum_{x \in V} W_x}$$

where W_v and W_x represent the features of nodes v and x, respectively. The variable $a_{u,v}$ is equal to 1 if nodes u and v are connected and zero otherwise.

Our metrics were also compared against the node-weighted closeness centrality metric (Singh et al. 2020):

$$\frac{W_u \sum_{v \in V \setminus \{u\}} \frac{W_v}{d_{u,v}+1}}{\sum_{x \in V} W_x}$$

where $d_{u,v}$ represents the geodesic distance, length of the shortest path, between nodes u and v.

As betweenness centrality, the personalized PageRank (Brin and Page 1998) and AlphaCentrality (Bonacich and Lloyd 2001) metrics are well established in the literature, their complete mathematical definitions are not included here for conciseness.

The node features signify the city population, GDP, or RNA expression level, depending on the network model.

Software and libraries

The Python package "*NetworkX*" version 3.6 (Hagberg et al. 2008) was used to generate the simple graph models, and to compute the betweenness and the four node-weighted centrality metrics. Cytoscape version 3.10.3 (Shannon et al. 2003) was employed to visualize each network. The scripts to compute the node-weighted degree centrality (Singh et al. 2020), node-weighted closeness centrality (Singh et al. 2020), inflow, and outflow centrality metrics are made available in the following Github repository: https://github.com/AramPapaz/Inflow-and-Outflow-Centrality. All analyses and computational procedures in this study were performed using Python 3.12.4.

Results and discussion

To illustrate the effect of the GCN normalization: $\sum_{j \in N(i)} \frac{1}{\sqrt{d_i d_j}}$, Fig. 1 presents how the neighbors of node g contribute to its degree normalization term in a toy network.

From Fig. 1, it is evident that node "h" is the most significant contributor to node "g", as node "h" is the only node directly connected to node "g" that shares no connections with other nodes, aside from its single edge linking to node "g". By construction, the new metrics prioritize nodes that have neighbor nodes that are otherwise little connected.

Simple graph models

To help in better understanding the two new metrics, two types of simple graph models were generated, see Fig. 2. These graphs use similar node features which were sampled from a uniform distribution between 1 and 10. In the left panel of Fig. 2, the node height is based on the node feature value (the larger the value, the larger the node height), in the middle panel these are based on the inflow metric, IM, and in the right panel these are based on the outflow metric, OM.

In both networks, nodes 3 and 7 were assigned the largest node features, which are represented by the height of their node boxes in graphs (a) and (d). In the Barabasi Albert network, node 4 was assigned the highest IM value (graphs b) because it is connected to multiple nodes having varying degrees. Additionally, some of its neighbors had relatively large feature values such as nodes 3, 8 and 6. In terms of connectivity, node 3 is similar to node 4, by being connected to nodes having varying degrees. However, the feature values of node 3's neighbors tend to be smaller. Hence, node 3 was assigned

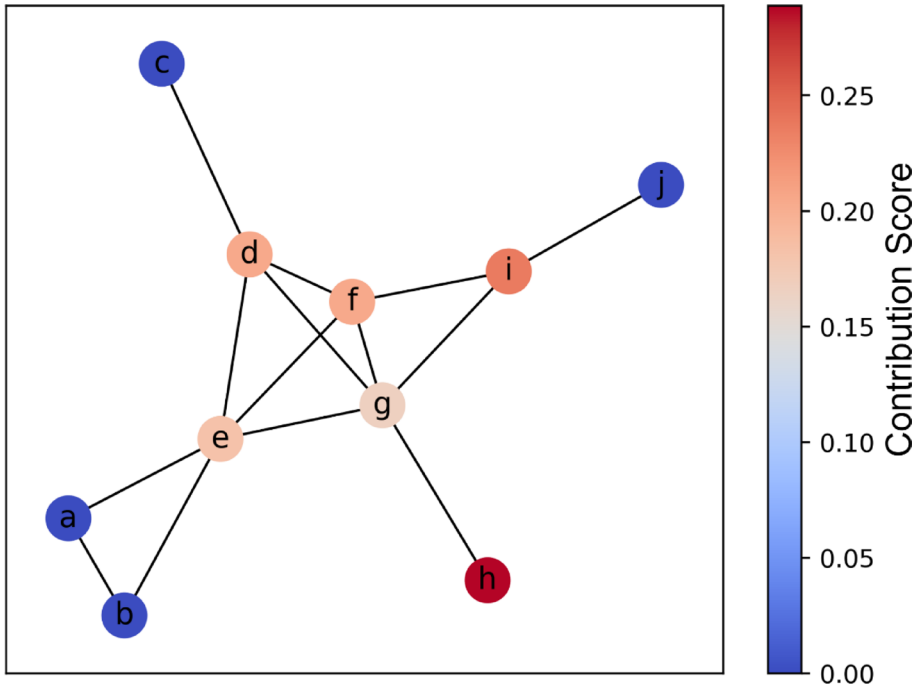

Fig. 1 Toy network in which nodes are colored on the basis of neighbor contribution scores to node "g". Higher values correspond to higher contributions, as per $\frac{1}{\sqrt{d_i d_j}}$

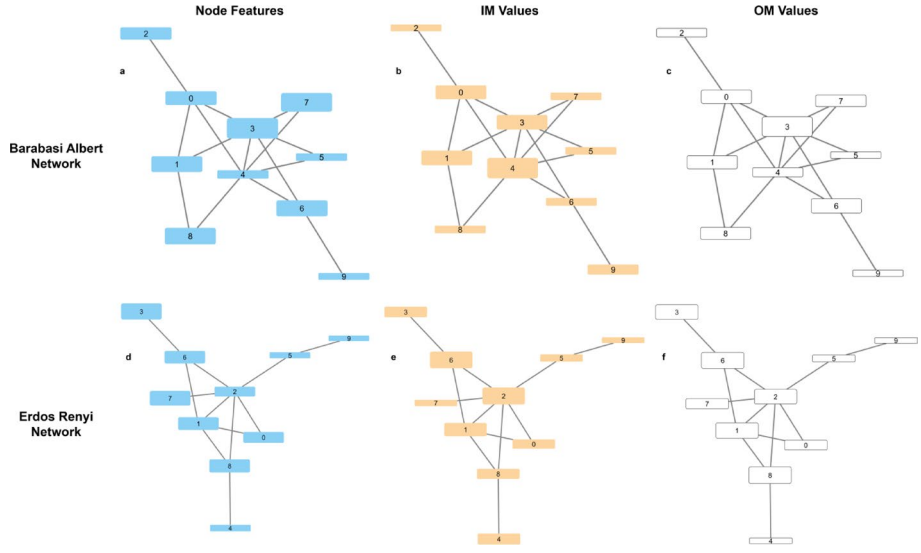

Fig. 2 Simple graph models generated based either (top) by the Barabasi Albert, or (bottom) by the Erdos Renyi models. The node heights reflect the magnitudes of the node's feature value (left panel), inflow value IM (middle panel), and outflow value OM (right panel), respectively. The larger the value, the larger the node

a smaller IM value than node 4. In the case of OM (graph c), the opposite was seen, whereby node 3 was assigned a larger OM value than node 4. The reason for this is that OM relies on the node's own feature value rather than those of its neighbors. Node 6 was ranked second based on OM since it had a large feature value as well (4th highest), and it is connected to both highly and weakly connected nodes. Thereby, it connects node 9 to major hub nodes 4 and 3.

For the Erdos Renyi network, node 2 had the highest IM value (graph e) and it is connected to both weakly and highly connected nodes. Notably, two of its neighbors, nodes 7 and 8, have the second and third highest feature values and are overall weakly connected which provide strong feature flow towards node 2. Node 6 had the second highest IM value; this is mainly because it is the only node connected to node 3 which has the largest feature value. Concerning the OM values (graph f), nodes 8 and 6 were ranked first and second, respectively. Both share similar connectivity patterns, whereby they are connected exclusively to a node and to hub nodes 1 and 2. However, node 8 had a larger feature value than node 6 which caused it to be ranked higher.

Overall, this analysis suggests that nodes having high IM values tend to be connected to nodes having varying degrees. This metric prioritizes nodes that are connected to weakly connected nodes with rather high feature values or to nodes having very large feature values that are well-connected. The name inflow stems from the fact it is reliant on the features of its neighbors that pass through it. Similarly, highly ranked OM nodes are also connected to nodes having varying degrees, but in this metric the node itself should have a relatively large feature value. Therefore, it passes its features to both weakly and highly connected nodes.

The same graph models but with different node feature values can be found in the supplementary Figures S1 and S2.

As a check of the effectiveness of IM and OM, the metrics were also applied to the well-known Zachary's karate club network (Girvan and Newman 2002) that was retrieved using the *karate_club*_graph function of (Hagberg et al. 2008). As this network lacks continuous node features, all node features were set to 1 which renders IM and OM to be the same. Our metric was able to correctly identify the known top 3 nodes being node 33 (president of the club, ranked first), node 0 (sensei, ranked second), and node 32 (vice president of the club, ranked third).

Airport network

The airport network had a mean shortest path length of the largest connected component of 3.468 and a mean clustering coefficient of 0.52, which are characteristic values of a densely connected network. Additionally, the betweenness, node-weighted degree, node-weighted closeness, personalized PageRank, and alpha centralities were computed for each node (city). PageRank and alpha centralities were computed with default parameters (alpha of 0.85 and 0.1, respectively) using the *NetworkX* package (Hagberg et al. 2008). The results were compared against the inflow and outflow centrality metrics and are illustrated in Table 1. For brevity, the following abbreviations are used: Betweenness Centrality: BC, Weighted-Degree Centrality: WD, Weighted-Closeness Centrality:

Table 1 Top 5 most central nodes in the airport network. Each node represents a City with City population as node feature for IM, OM, WD, WC, PR and AC. Cities are listed according to the order of decreasing centrality values

BC	WD	WC	PR	AC	IM	OM
London	London	London	London	Mumbai	Bogota	Bogota
Paris	Paris	Paris	Paris	Delhi	Buenos Airos	Buenos Airos
Moscow	Istanbul	Frankfurt	Moscow	Stockholm	Campinas	Moscow
Istanbul	Frankfurt	Amsterdam	Istanbul	Beijing	Moscow	Sydney
Beijing	Amsterdam	New York	Atlanta	Oslo	Mumbai	Denver

WC, Personalized PageRank: PR, Alpha Centrality: AC, Inflow Metric: IM, and Outflow Metric: OM.

Compared to the other metrics, the IM and OM metrics provided dissimilar results: both inflow and outflow metrics reported Bogota, Colombia to be the most relevant and not London, which had the most connections. Node degree and node feature distributions are visualized in Fig. 3a and b.

Per the distribution visualized in Figs. 3a and b, it is evident that there exists a significant gap between the degrees of both cities, yet the populations are remarkably similar. To seek further insight into local topologies, the node degrees of Bogota's neighbors were compared against those of London and other cities; see Fig. 4a. Additionally, Fig. 4b displays the connectivity of Bogota and its direct neighbors, where node size directly correlates to the degree of the respective node, signifying that a larger node size implies a higher degree; see Fig. 4b.

Figure 4a highlights that there exists considerable variance in the degrees of the neighbors that Bogota, Moscow, and Sydney possess. In contrast, the neighbors of London, Paris, and Frankfurt are rather well connected. This continues to prove evident in Fig. 4b, where Bogota's neighbors on the left feature smaller node sizes, representing weakly connected nodes, compared to those on the right. Therefore, the relevance of Bogota emerges: the city serves as a "gateway" for more weakly connected cities to transit to major hub cities, such as Paris and Frankfurt. In fact, the El Dorado International Airport in Bogota, Colombia is globally considered as the Hub of the Americas (Bogotá 2026). It serves as an important link for passengers and cargo connecting South America with North America, Europe, and beyond. It handles a substantial volume of both international and domestic flights (Bogotá 2026). This may also apply to Moscow, which provides transportation for large populations in rural areas. For Sydney, as it was ranked highly by OM, it provides the dense population of Sydney with direct transportation to rural areas in Australia and to large densely populated cities like Beijing and Delhi.

Furthermore, when Bogota was removed from the airport network, the number of connected components increased from 7 to 22 reflecting the important role of Bogota for the connectivity of the graph. For comparison, when London was removed, the number of connected components increased only to 10. Upon removing Frankfurt, another hub, the number of connected components stayed at 7. To study this more

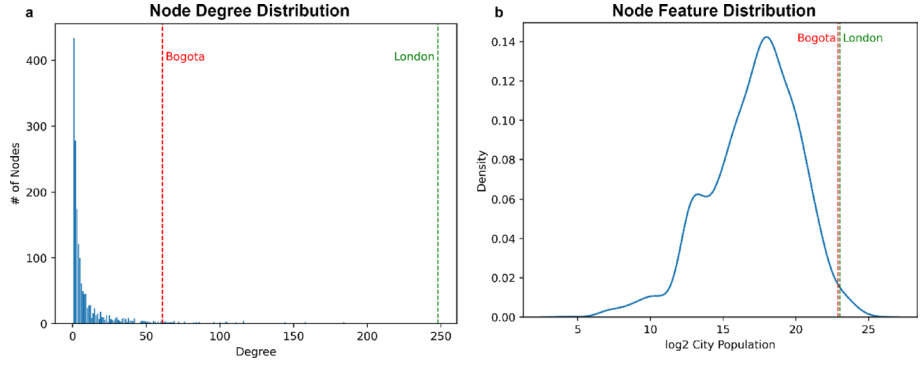

Fig. 3 **a** Node degree distribution of the airport network. Red and green dashed lines show the node degrees of Bogota and London, respectively. **b** Node feature distribution of the airport network. Red and green dashed lines show the log2 transformed city populations of Bogota and London, respectively

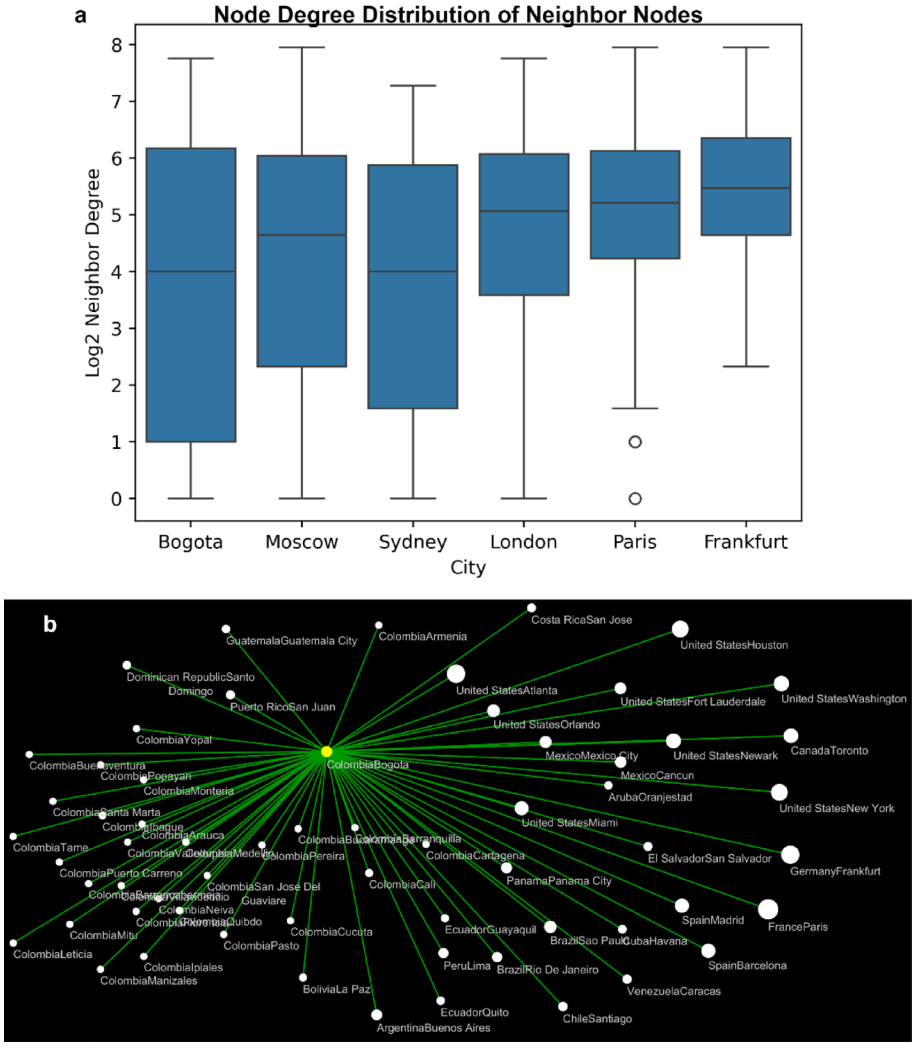

Fig. 4 **a** Degree distribution of the neighbors of Bogota, Moscow, Sydney, London, Paris, and Frankfurt. **b** Visualization of Bogota's direct neighbors. The larger the node size, the higher its degree. Bogota is highlighted in yellow

systematically, we investigated the number of connected components after removing any of the top 20 ranked cities based on either BC, WD, WC, PR and AC (one at a time). This distribution was compared against the effect of removing any of the top 20 ranked cities based on IM and OM (Fig. 5). Evidently, the increase in the number of connected components is significantly higher when removing the top 20 ranked IM or OM nodes compared to when removing any of the top 20 nodes based on the other metrics (Mann-Whitney U test p-value < 0.001).

A Spearman correlation was computed for the rankings of the airport network provided by the IM and OM metrics and any of the other metrics. Notably, PR had the highest spearman correlation with both the OM and IM ranking, with correlations of 0.84 and 0.82, respectively. This is mainly due to a somewhat similar normalization approach in the summation term of PageRank, where instead of normalizing with $\sqrt{d_i d_j}$, the values are normalized by the outgoing edges of the neighbor node. However, PR ranked Bogota and Buenos Aires only as 24th and 25th, respectively. The Pearson correlations

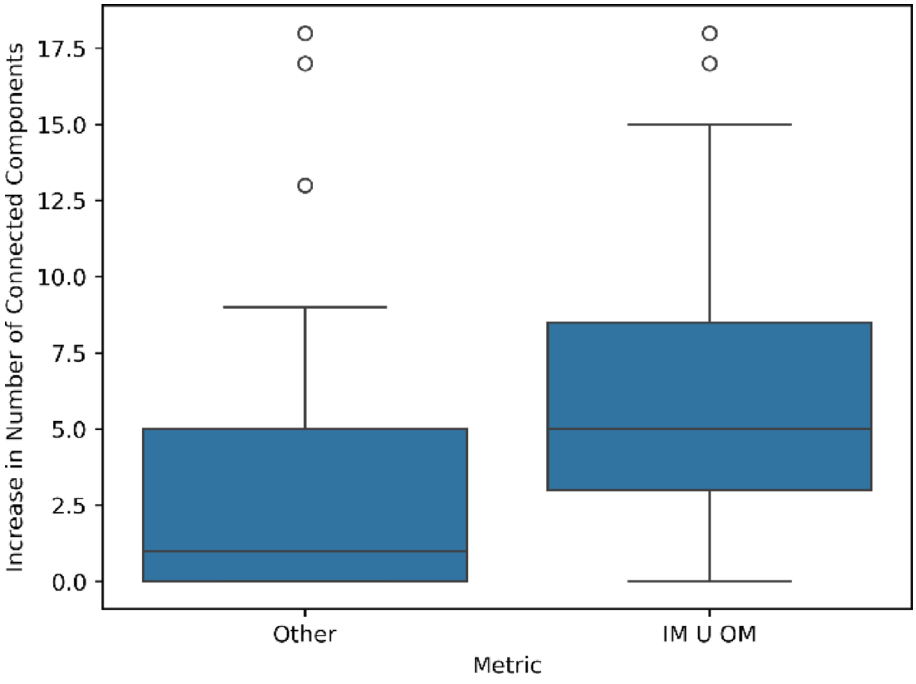

Fig. 5 Distribution of the number of connected components after removing any of the top 20 nodes based on BC, WD, WC, PR and AC ("Other") versus removing any of the top 20 nodes based on IM and OM

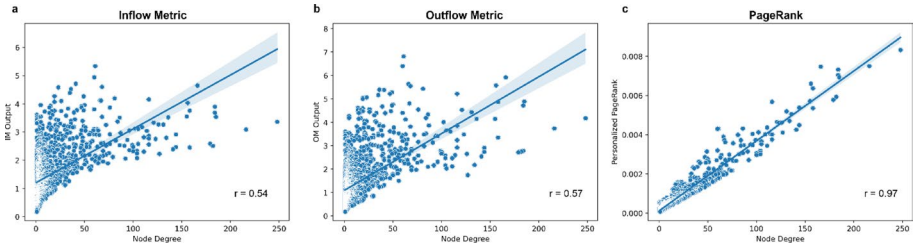

Fig. 6 Pearson correlations of IM (**a**), OM (**b**), and PR (**c**) values with node degrees

of the new metrics with node degrees were 0.54 in IM and 0.57 in OM, Fig. 6a and b, whereas PR had a much higher correlation with node degree of 0.97, Fig. 6c. In the supplementary materials, analysis is shown how the damping factor of PR affects its correlation with our metrics.

Airplane trade network

The constructed airplane trade network is a connected graph of countries with nodes having as feature their GDP, and edges representing airplane trade relationships. Its mean shortest path length was computed to 2.24 and it possessed a mean clustering coefficient of 0.52, which is representative of a rather dense network. Table 2 shows the centrality metrics for each country in this trade network.

In Table 2, the IM and OM metrics gave similar results as BC and PR. Together, these were the only metrics that ranked South Africa among the top 5 central nodes in the network. South Africa appears to act similarly to Bogota as it is connected to major hub nodes like USA and to very little connected nodes like Seychelles.

Table 2 Most central nodes in the airplane trade network. Each node represents a country, with its country's GDP as node feature. Countries are listed according to the order of decreasing centrality values

BC	WD	WC	PR	AC	IM	OM
USA	USA	USA	USA	Botswana	USA	USA
France	France	France	France	Zimbabwe	South Africa	China
South Africa	Germany	Germany	Germany	Honduras	France	South Africa
Germany	Switzerland	Switzerland	South Africa	Congo Dem. Rep.	China	France
China	Netherlands	Netherlands	Switzerland	Mozambique	Germany	Germany

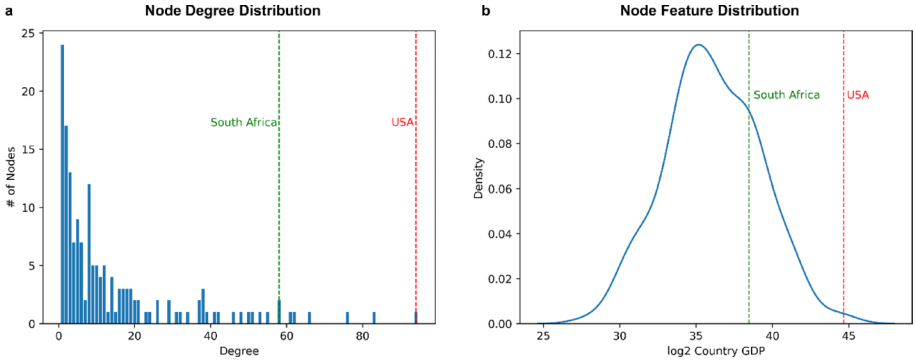

Fig. 7 **a** Node degree distribution of the airplane trade network. Red and green dashed lines mark the node degrees of USA and South Africa, respectively. **b** Node feature distribution of the trade network. Red and green dashed lines signify the log2 of the national GDP of USA and South Africa, respectively

The airplane trade network possesses a similar degree distribution as the airport network; see Figs. 3a and 7a. USA is considered the wealthiest and most well-connected node in this network, see Fig. 7b. To further analyze the significance of South Africa, Fig. 8 compares the node degrees of the neighbors of some highly ranked countries.

South Africa was found to have the widest variation in the degrees of its neighbor nodes. Compared to the Airport network, the airplane trade network is a connected graph that has a single connected component. When South Africa was removed from this network, the number of connected components increased to 5, but when USA was removed the number of connected components increased only to 2.

Breast cancer protein-protein interaction network (PPIN)

The largest connected component of the breast cancer PPIN had a mean shortest path length of 3.22 and a mean clustering coefficient of 0.32. Table 3 organizes the metrics computed for each gene, assigned to a node, in the breast cancer PPIN.

Both the IM and OM metrics consider TRIM28 to be most relevant. The only other metrics in the table to propose this were BC and PR. Most of the other metrics ranked TP53 among the top five highest, which was ranked 11th in IM and OM. TP53, transformation-related protein 53, is a regulatory transcription factor; protein TP53 is considered a tumor suppressor gene due to its role in conserving genome stability by preventing mutations and uncontrolled cell division, inducing apoptosis to cells with damaged genomes (Toufektchan and Toledo 2018). TP53 serves as a highly significant factor in the development of carcinogenesis, as it is the most frequently mutated gene in human cancer (Surget et al. 2013). Moreover, TRIM28 is a protein involved in transcriptional regulation and plays a role in mediating cell growth and differentiation;

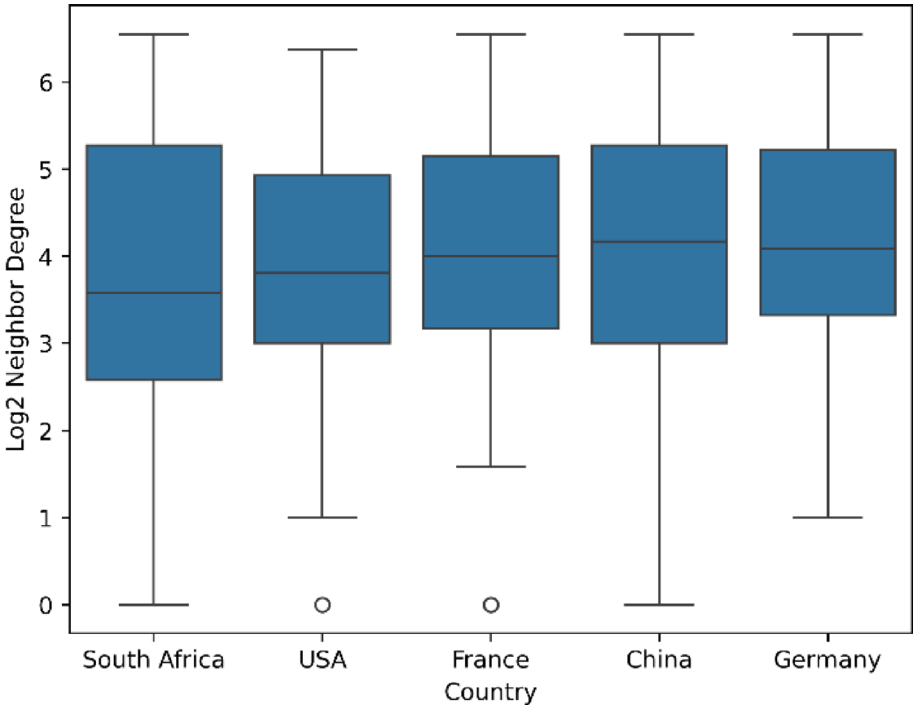

Fig. 8 Degree distribution of the neighbors of South Africa, USA, France, China, and Germany

Table 3 Top 5 most central genes and proteins in the breast cancer network. Each node represents a gene, with a gene's average expression level serving as a node feature

BC	WD	WC	PR	AC	IM	OM
TRIM28	TP53	TP53	TRIM28	IFI16	TRIM28	TRIM28
TP53	AKT1	ACTB	TP53	VWF	GNB1	GNG13
ACTB	RPS27A	AKT1	PRKACA	CYP11B2	GNAL	ARRB1
SRC	ACTB	GAPDH	PRKACB	MFSD9	PRKACB	GNB1
EGFR	SRC	EGFR	PRKACG	CD40	PRKACG	PRKACA

additionally, TRIM28, tripartite motif-containing 28, has been shown to promote breast cancer metastasis (Wei et al. 2016). Likewise, ranking second in the IM metric in Table 3, GNB1, guanine nucleotide-binding protein subunit beta-1, has also been shown to be involved in breast cancer proliferation and metastasis (Zou et al. 2024).

TP53 accrued the most connections in the breast cancer network, with a total of 1447, whereas TRIM28 had 502 connections. The node degree distribution of the breast cancer network, portrayed in Fig. 9a, is similar to the airport network distribution featured in Fig. 3a. To further validate the significance of TRIM28, the node degrees of its neighbors were compared against those of TP53; see Fig. 8.

Figure 10a shows a similar pattern to the data trend visualized in Fig. 4a in that the highly connected node representing TP53 is connected with neighbors that are also well connected. In contrast, the neighbors of GNB1 and TRIM28 vary from well connected to poorly connected in terms of their respective neighbor topologies; this is more evident in TRIM28. Therefore, TRIM28 exhibits the capability to directly reach genes that are more difficult to interact with, unlike TP53. Additionally, Fig. 10b highlights that TRIM28's neighbors tend to be expressed at a slightly higher level, using a t-test with a p-value < 0.01. Therefore, it proves crucial to investigate common direct interactors

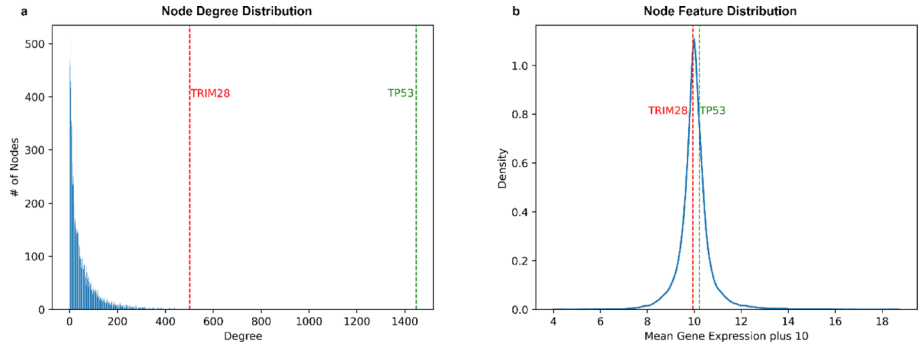

Fig. 9 (**a**) Node degree distribution of the breast cancer network. Red and green dashed lines label the node degrees of TRIM28 and TP53, respectively. (**b**) Node feature distribution of the breast cancer network. Red and green dashed lines show the mean gene expression of TRIM28 and TP53, respectively

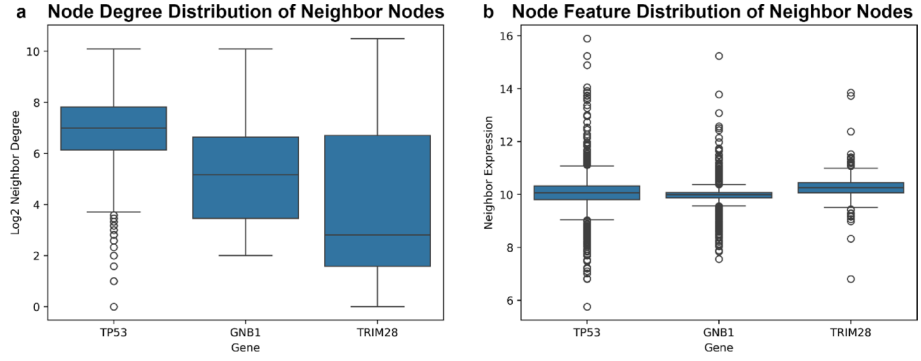

Fig. 10 Comparison between the degree (**a**) and expression (**b**) distribution of the neighbors of TP53, GNB1, and TRIM28

for such genes. A Spearman correlation was computed between the IM and OM metrics and all other metrics: similar to the airport network, PR yielded the highest Spearman correlation with both OM and IM, with correlations of 0.91 and 0.92, respectively. Interestingly IM and OM did not consider TP53 as a top 10 central gene compared to almost all other metrics. When we removed TP53 from the network the number of connected components only increased from 6 to 7. However, when TRIM28 was removed, the number of connected components increased to 64. Next, the metric ranks of known breast cancer drug targets were analyzed: in this dataset, 79 breast cancer drug targets were retrieved from CUSABIO (Fig. 11) (CUSABIO 2025).

IM and OM assigned lower ranks (further away from the top positions) to the drug target dataset, compared to the other metrics. The only exception was AC, which possesses a much lower median ranking. It is important to assign these drug targets lower rankings in order to indicate that knocking out such targets by blocking their action with a small-molecule inhibitor may prove safe and may not induce deleterious side effects on the patient, when compared to knocking out paramount hub genes, such as TP53 or TRIM28.

Computational complexity

The computational complexity to calculate all node degrees in a graph is $O(V+E)$ where V and E are the total number of vertices and edges, respectively. Computing the median of an unsorted array of node degrees will take $O(VlogV)$. Using the

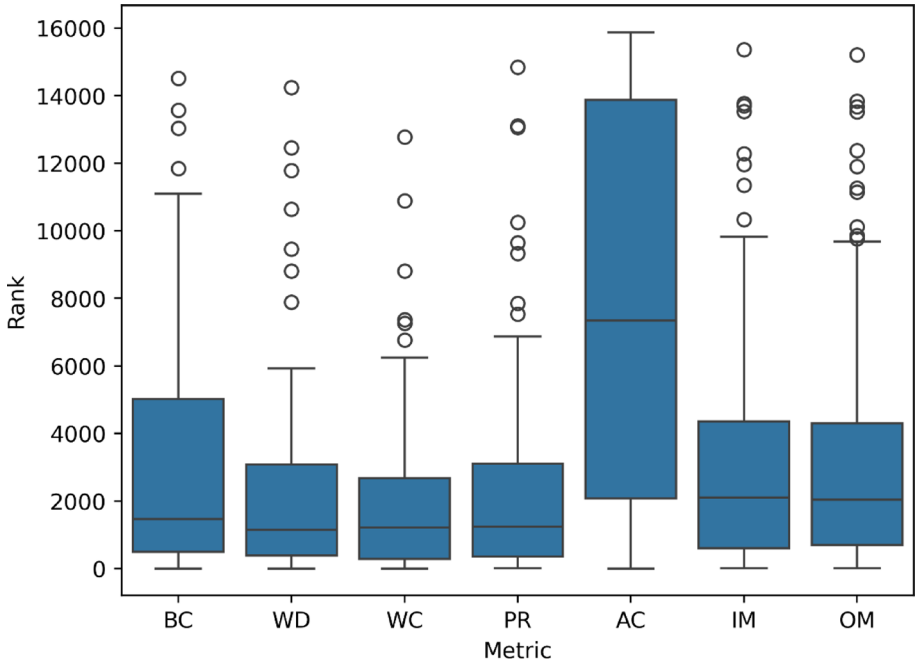

Fig. 11 Distributions of the rankings for the dataset of 79 breast cancer drug targets in each metric

computed degrees and median, it will take $O(V^2)$ to calculate IM and OM. Our implementation at Github contains an outer loop over all vertices and an inner loop over all neighbors of a specific vertex. For a dense network this yields a complexity $O\left(V^2\right)$ where $E \approx V^2$. Therefore, the total time complexity to compute the new metrics will be $O(V + E + VlogV + V^2)$ which can be simplified to $O\left(V^2\right)$.

For comparison, weighted-degree centrality has a computational complexity of O(E + V), weighted-closeness centrality has a computational complexity of $O\left(|V|^2 * \middle| E \right|)$ based on (Then et al. 2017), personalized pagerank has a computational complexity of $O\left(\frac{E}{\alpha}\log\frac{1}{\epsilon}\right)$ based on (Chen et al. 2023), where α is the damping factor and ϵ a precision parameter, and alpha centrality has a computational complexity of $O\left(V^3\right)$ based on (Lin et al. 2021).

To obtain empirical run times, four Erdos Renyi graphs were generated using the *erdos_renyo_graph* function from (Hagberg et al. 2008) having 1000, 2500, 5000 and 10,000 nodes, respectively. Node features for each network were sampled from a uniform distribution between 1 and 10. For each metric the default parameters were used. The time it takes to compute each metric on each network was measured in seconds using the *repeat* function from the python package *timeit.* For each metric, its outcome was computed ten times on each network and its median time ± standard deviation are reported in Table 4. These run times do not include one-time graph preprocessing costs such as graph construction and data structure initialization.

All experiments were conducted on a Linux server equipped with two Intel Xeon Silver 4310 CPUs (12 physical cores per socket, 24 cores total, 48 hardware threads) with a base frequency of 2.10 GHz and 251 GB RAM, running Ubuntu 24.04.2. All implementations were executed in *Python* 3.12.4 using *NetworkX* 3.6 in a single-threaded setting. The code for these experiments can be found in the Github repository.

Table 4 Median runtimes (± standard deviation 10 runs) for each centrality metric on four Erdos Renyi graphs of increasing size

# of Nodes	IM & OM	PR	BC	AC	WD	WC
1,000	0.004 ± 0.0001	0.005 ± 0.006	2.18 ± 0.05	0.03 ± 0.004	0.002 ± 0.00	0.5 ± 0.02
2,500	0.02 ± 0.0002	0.02 ± 0.0007	21 ± 0.007	0.18 ± 0.02	0.008 ± 0.00	3 ± 0.006
5,000	0.11 ± 0.0009	0.11 ± 0.001	168 ± 0.56	0.58 ± 0.001	0.03 ± 0.003	12 ± 0.02
10,000	0.53 ± 0.001	0.49 ± 0.003	1423 ± 1.31	2.5 ± 0.019	0.12 ± 0.004	52 ± 0.01

Conclusion

Ultimately, two new centrality metrics were introduced in this paper, termed inflow and outflow centrality, inspired by the aggregation approach implemented in graph convolution networks. The two novel metrics were compared against betweenness and four weighted centralities in three diverse network models. As metrics, inflow and outflow centrality prioritize nodes not only based on the attributes of their connections or features, but holistically incorporating contribution from the connectivity and features of surrounding neighbors. This ability provides new insights on detecting key nodes and more subtle patterns in neighbor contributions, as demonstrated in the three aforementioned networks. By emphasizing the contribution of otherwise little connected neighbor nodes, the new metrics prioritize nodes that are crucial to maintain a graph's connectivity. Inflow and outflow centrality metrics can be computed on any node-weighted network using the following Github repository: https://github.com/AramPapaz/Inflow-and-Outflow-Centrality.

Abbreviations

GCN Graph convolutional network
GAN Graph attention network
GDP Gross domestic product
PPIN Protein protein interaction network
PR Personalized PageRank
AC Alpha centrality
BC Betweenness centrality
WD Weighted-degree centrality
WC Weighted-closeness centrality
IM Inflow metric
OM Outflow metric
USA United States of America

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-026-00782-7.

Supplementary Material 1

Acknowledgements
We would like to thank Anna Elizabeth Schmitz for editing the manuscript.

Author contributions
Aram Papazian: conceptualization, methodology, data analysis, and writing the draft for the manuscript. Volkhard Helms: conceptualization, discussion of results, and editing of the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This study was funded internally by Saarland University.

Data availability
The datasets generated and/or analysed during the current study are available in the supplementary material and the following github repository https://github.com/AramPapaz/Inflow-and-Outflow-Centrality.

Declarations

Competing interests
The authors declare no competing interests.

Received: 29 August 2025 / Accepted: 10 February 2026
Published online: 18 February 2026

References

Bogotá Airport Guide: Navigating El Dorado International (2026). www.remitly.com. Accessed 20 Jan 2026
Bogotá is indeed The Hub of the Americas | Aviation Infographics (2026) www.semana.com (in Spanish). Accessed 20 Jan 2026
Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Soc Networks 23(3):191–201
Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30(1–7):107–117
Chen Z, Guo X, Zhou B, Yang D, Skiena S (2023) Accelerating personalized PageRank vector computation. In: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp 262–273
Chung FR (1997) Spectral graph theory (Vol. 92). American Mathematical Soc.
CUSABIO (2025) Verified breast cancer drug targets. https://www.cusabio.com/drug-targets/breast-cancer.html?srsltid=AfmBOorR6dTElFNciBR-3q4UuQUU_4eloL8SsIpdvj7t7a8oi4EgPfsj. Accessed Jul 2025
den Van Heuvel MP, Sporns O (2013) Network hubs in the human brain. Trends Cogn Sci 17(12):683–696
Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826
Hagberg A, Swart PJ, Schult DA (2008) Exploring network structure, dynamics, and function using networkx (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Laboratory (LANL), Los Alamos
Kipf TN (2016) Semi-supervised classification with graph convolutional networks. Preprint at https://arxiv.org/abs/1609.02907
Lin M, Li W, Song LJ, Nguyen CT, Wang X, Lu S (2021) Sake: estimating Katz centrality based on sampling for large-scale social networks. ACM Trans Knowl Discovery Data (TKDD) 15(4):1–21
Mayer T, Santoni G, Vicard V (2023) The CEPII trade and production database. CEPII
Mering CV, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 31(1):258–261
Micheli A (2009) Neural network for graphs: a contextual constructive approach. IEEE Trans Neural Netw 20(3):498–511
Newman ME (2010) Networks: an introduction
OpenFlights Airports Database (2025) Airport and airline data under Open Database License. https://openflights.org/data. Accessed Jul 2025
Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32(3):245–251
Saberi M, Khosrowabadi R, Khatibi A, Misic B, Jafari G (2021) Topological impact of negative links on the stability of resting-state brain network. Sci Rep 11(1):2176
Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2008) The graph neural network model. IEEE Trans Neural Netw 20(1):61–80
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504
Singh A, Singh RR, Iyengar SRS (2020) Node-weighted centrality: a new way of centrality hybridization. Comput Soc Netw 7(1):6
Surget S, Khoury MP, Bourdon JC (2013) Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. OncoTargets Ther. https://doi.org/10.2147/OTT.S53876
Then M, Günnemann S, Kemper A, Neumann T (2017) Efficient batched distance and centrality computation in unweighted and weighted graphs. Datenbanksysteme für Business, technologie und web (BTW 2017). Gesellschaft für Informatik, Bonn, pp 247–266
Tomczak K, Czerwińska P, Wiznerowicz M (2015) Review the Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia 2015(1):68–77
Toufektchan E, Toledo F (2018) The guardian of the genome revisited: p53 downregulates genes required for telomere maintenance, DNA repair, and centromere structure. Cancers 10(5):135
Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. Preprint at https://arxiv.org/abs/1710.10903
Wei C, Cheng J, Zhou B, Zhu L, Khan MA, He T et al (2016) Tripartite motif containing 28 (TRIM28) promotes breast cancer metastasis by stabilizing TWIST1 protein. Sci Rep 6(1):29822
Word Bank Group (2025) Country gross domestic product per year. https://www.worldbank.org/ext/en/home. Accessed Jul 2025
World Cities Dataset (2019) City population data. https://www.kaggle.com/datasets/viswanathanc/world-cities-datasets?resource=download. Accessed Jul 2025
Wu L, Cui P, Pei J, Zhao L, Guo X (2022) Graph neural networks: foundation, frontiers and applications. In Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 4840–4841
Zhai H, Hou H, Luo J, Liu X, Wu Z, Wang J (2023) DGDTA: dynamic graph attention network for predicting drug–target binding affinity. BMC Bioinformatics 24(1):367
Zou H, Chen P, Li Z, Yan T, Cui D, Gong L et al (2024) Identification of GNB1 as a downstream effector of the circRNA-0133711/miR-145-5p axis involved in breast cancer proliferation and metastasis. Oncologie 26(5):753–769

Publisher's note