Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Forman curvature for directed networks

by

R.P. Sreejith, Jürgen Jost, Emil Saucan, and Areejit Samal

Preprint no.: 32 2016

Forman curvature for directed networks

R.P. Sreejith,[1] Jürgen Jost,[2, 3, *] Emil Saucan,[2, 4, †] and Areejit Samal[1, ‡]

[1] *The Institute of Mathematical Sciences, Chennai, India*
[2] *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
[3] *The Santa Fe Institute, Santa Fe, New Mexico, USA*
[4] *Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa, Israel*

A goal in network science is the geometrical characterization of complex networks. In this direction, we have recently introduced the Forman's discretization of Ricci curvature to the realm of undirected networks. Investigation of Forman curvature in diverse model and real-world undirected networks revealed that this measure captures several aspects of the organization of complex undirected networks. However, many important real-world networks are inherently directed in nature, and the Forman curvature for undirected networks is unsuitable for analysis of such directed networks. Hence, we here extend the Forman curvature for undirected networks to the case of directed networks. The simple mathematical formula for the Forman curvature in directed networks elegantly incorporates node weights, edge weights and edge direction. By applying the Forman curvature for directed networks to a variety of model and real-world directed networks, we show that the measure can be used to characterize the structure of complex directed networks. Furthermore, our results also hold in real directed networks which are weighted or spatial in nature. These results in combination with our previous results suggest that the Forman curvature can be readily employed to study the organization of both directed and undirected complex networks.

PACS numbers: 89.75.Hc 89.75.Fb 89.75.-k

I. INTRODUCTION

Complex networks [1–6] are present everywhere in nature and society. Metabolic networks [7] capture interactions among different metabolites and enzymes that are responsible for growth and maintenance of a living organism. Ecological networks [8] capture interactions among different species in the ecosystem. Online networks such as Facebook [9] and Twitter [10] capture relationships between different individuals in the society. Transportation networks [11] capture the movement of traffic across various parts of the world. Network science [1–6] aims to characterize the structure of these ubiquitous complex networks. Towards this goal, a recent focus in network science has been the development of geometry based measures to characterize the structure of complex networks [12–21].

A central concept in geometry is curvature which quantifies the deviation of an object from being flat. In differential geometry, where the notion of curvature originated, there are several types of curvature [22]. Among those, it seems that the concept of Ricci curvature is the most useful for the analysis of graphs or networks. In geometry, Ricci curvature depends on a direction, which then for networks translates into the fact that it should be a measure associated to an edge rather than to a node. Among the several curvature measures [12–21, 23–27] that have been proposed for geometrical characterization of complex networks, two different discretizations of the Ricci curvature, Ollivier-Ricci curvature [13, 25, 27] and Forman-Ricci curvature [28] seem particularly appealing and useful for analyses of complex networks.

While Ollivier-Ricci curvature has already been systematically explored in complex undirected networks [14–17, 19, 20, 29, 30], we [21] recently have introduced Forman's discretization of classical Ricci curvature [28] to the realm of complex undirected networks. Astonishingly, the mathematical formula of the Forman curvature of an edge is enticingly simple which renders the measure suitable for analysis of large-scale networks [21]. Importantly, the mathematical formula of the Forman curvature also elegantly incorporates both edge weights and node weights which also makes the measure suitable for analysis of both unweighted and weighted networks [21]. Since Forman curvature represents a discretization of the classical Ricci curvature which is intrinsically associated with edges of a network, this notion of curvature does not necessitate the technical artifice of extending a measure for the curvature of nodes to the edges [21]. Thus, Forman curvature can be exploited for edge-based analysis of complex networks. Of course, once we have defined the curvature of an edge, we can then also define the curvature of a node by summing or averaging the curvatures of its adjacent edges, somewhat analogous to the concept of scalar curvature in Riemannian geometry

* jost@mis.mpg.de
† emil.saucan@mis.mpg.de
‡ asamal@imsc.res.in

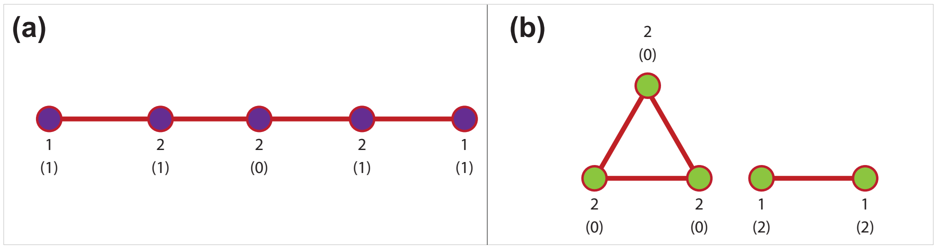

FIG. 1. Two networks with the same degree distribution can have different distributions of Forman curvature. Each of the two networks shown in (a) and (b) have 5 nodes and 4 edges. The degree and Forman curvature is indicated besides each node with Forman curvature in parenthesis.

[22], and we may then define that to be the Forman curvature of a node. We remark that the Forman curvature for an edge is a local measure dependent on weights of adjacent nodes and edges in the network [21]. Nevertheless, two networks with the same degree distribution can have very different distributions of Forman curvature (Figure 1).

Although, we have successfully introduced Forman curvature to undirected networks [21], several important real networks in nature and society are inherently directed in nature. These include the metabolic networks [7], gene regulatory networks [31], signaling networks [32], neural networks [33], the world wide web (WWW) [34], online social networks [10] and transportation networks [35, 36]. However, the two different discretizations of the Ricci curvature, Ollivier-Ricci curvature and Forman-Ricci curvature, have been developed and employed for the analysis of undirected graphs to date [14–17, 19–21, 29, 30]. To enable proper investigations of directed networks, we here extend the concept of Forman curvature to the domain of directed graphs. By investigating a variety of model and real directed networks, we show that our extension of the Forman curvature to directed graphs, which elegantly incorporates both edge weights and node weights, can be utilized to analyze and classify different types of directed networks. Thus, Forman curvature can hereafter be employed to investigate both undirected and directed complex networks.

The remainder of this paper is organized as follows. In section II, we extend the notion of Forman curvature for undirected networks to directed networks. In section III, we list the dataset of directed networks, both model and real, employed to investigate the Forman curvature for directed networks. In section IV, we describe our main results, and in section V, we conclude with a summary and future outlook.

II. MATHEMATICAL DEFINITION

In this section, we extend the notion of Forman curvature for undirected graphs to the setting of directed graphs, and present the simple mathematical formulas that allow the computation of the Forman curvature in directed networks.

Recently, we have introduced the Forman curvature to the realm of undirected networks, and showed that this simple yet powerful notion of curvature has several advantages over other curvature measures for networks [21]. Here, we choose not to elaborate the geometric motivation and physical interpretation of the Forman curvature for networks, and refer the interested reader to Ref. [21]. We remark that the classical Ricci curvature operates directionally along vectors, and in the discrete setting of networks, Forman curvature is associated with the discrete analogue of vectors, namely, edges. Forman curvature for an edge e in the graph is given by the following formula [21]:

$$\mathbf{F}(e) = w_e \left(\frac{w_{v_1}}{w_e} + \frac{w_{v_2}}{w_e} - \sum_{e_{v_1} \sim e,\ e_{v_2} \sim e} \left[\frac{w_{v_1}}{\sqrt{w_e w_{e_{v_1}}}} + \frac{w_{v_2}}{\sqrt{w_e w_{e_{v_2}}}} \right] \right) \tag{1}$$

where

- e denotes the edge under consideration between two nodes v_1 and v_2.
- w_e denotes the weight of the edge e under consideration.
- w_{v_1} and w_{v_2} denote the weights associated with the nodes v_1 and v_2, respectively.
- $e_{v_1} \sim e$ and $e_{v_2} \sim e$ denote the set of edges incident on nodes v_1 and v_2, respectively, after *excluding* the edge e under consideration which connects the two nodes v_1 and v_2.

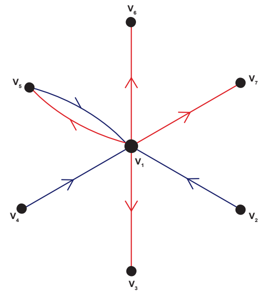

FIG. 2. The set of *incoming* and *outgoing* edges for a node v_1 in an example network.

Using the definition of Forman curvature for edges in the network, one can elegantly define the Forman curvature for a node v in the network as follows:

$$\mathbf{F}(v) = \sum_{e_v \sim v} \mathbf{F}(e_v) \tag{2}$$

where e_v denotes the set of edges incident on the node v. Previously [21], we had used a slightly different definition of the Forman curvature for a node v where the sum of the Forman curvatures for edges incident on node v in the network was normalized by dividing the sum by degree of node v. Subsequently, we have found that the unnormalized definition of the Forman curvature for a node v (Eq. 2) has even better correlation with common network measures such as degree and betweenness centrality in undirected networks. Thus, we decided to use here the unnormalized definition of the Forman curvature for a node in the network.

In the original work [28], Forman had not envisaged his curvature function for directed networks. Notice that in Forman's original work only positive weights were considered, since these weights represent therein generalizations of such natural geometric notions as length, area and volume. In consequence, no directed surfaces (or, more general complexes) can be modeled directly using Forman's original formalism. However, we will show here that it is simple to adapt the Forman curvature for undirected graphs [21] to directed graphs. In this aspect, our work, though restricted solely to networks, also represents a novel theoretical extension. In fact, when we rearrange the terms in the definition of the Forman curvature for an edge e (Eq. 1), to separate the contributions of the two nodes v_1 and v_2 involved,

$$\mathbf{F}(e) = w_e \left(\frac{w_{v_1}}{w_e} - \sum_{e_{v_1} \sim e} \frac{w_{v_1}}{\sqrt{w_e w_{e_{v_1}}}} \right) + w_e \left(\frac{w_{v_2}}{w_e} - \sum_{e_{v_2} \sim e} \frac{w_{v_2}}{\sqrt{w_e w_{e_{v_2}}}} \right) \tag{3}$$

we see that one can naturally define the curvature of a directed edge by only using the term involving its initial node, or alternatively that for its terminal node.

Alternatively, when we start with a node in a directed network, we can distinguish between its incoming and outgoing edges. Given a node v, let us denote the set of *incoming* and *outgoing* edges for a node v by $E_{I,v}$ and $E_{O,v}$, respectively (Figure 2). Then, one can elegantly define the *In Forman curvature* $\mathbf{F}_I(v)$ and the *Out Forman curvature* $\mathbf{F}_O(v)$ as follows:

$$\mathbf{F}_I(v) = \sum_{e \in E_{I,v}} \mathbf{F}(e_v)\,; \tag{4}$$

$$\mathbf{F}_O(v) = \sum_{e \in E_{O,v}} \mathbf{F}(e_v)\,; \tag{5}$$

where the summations are taken over only the incoming and outgoing edges, respectively. Moreover, one can obtain the total amount of flow through a node v as follows:

$$\mathbf{F}_{I/O}(v) = \mathbf{F}_I(v) - \mathbf{F}_O(v)\,. \tag{6}$$

III. DATASETS

In this work, we have analyzed the Forman curvature for directed graphs in both model and real networks. We have considered two generative models for directed networks:

- **Erdös-Rényi (ER) model** [37] is commonly used to generate random graphs. ER model generates an ensemble $G(n, p)$ of graphs where n is the number of nodes and p is the probability that each possible directed edge exists between any pair of nodes in the network.

- **Scale-free model** [38] generates directed graphs with power-law degree distribution for both in-degree and out-degree of nodes. This growing network model implements a preferential attachment scheme wherein new nodes are connected to existing nodes based on in-degree and out-degree distribution. Starting from an initial network, the graph expansion at each discrete time step occurs through addition of new nodes or new edges. The graph expansion is based on three model parameters: α, β and γ. The parameter α gives the probability of adding a new node v with an edge from v to an existing node w where node w is chosen based on in-degree distribution. The parameter γ gives the probability of adding a new node v with an edge to v from an existing node w where node w is chosen based on out-degree distribution. The parameter β gives the probability of adding a new edge from an existing node v to another existing node w, where v and w are chosen independently according to out-degree and in-degree distribution, respectively.

Apart from the above mentioned model networks, we have also considered the following directed and unweighted real networks:

- ***E. coli* TRN** [39, 40] gives the transcriptional regulatory network in the bacterium *Escherichia coli.* In this network of 3072 nodes and 7853 edges, nodes corresponds to genes and directed edges represent control of target gene expression through transcription factors.

- ***B. subtilis* TRN** [40] gives the transcriptional regulatory network in the bacterium *Bacillus subtilis.* In this network of 1594 nodes and 2902 edges, nodes corresponds to genes and directed edges represent control of target gene expression through transcription factors.

- **Human protein interaction network** [41] captures a set of protein protein interactions obtained using a mass spectrometry based approach. In this network of 2239 nodes and 6452 edges, nodes are proteins and directed edges represent interactions between them.

- **Air traffic control** is a directed network which was reconstructed based on the Preferred Routes Database of the US Federal Aviation Administration National Flight Data Center (NFDC). In this network of 1226 nodes and 2613 edges, nodes corresponds to airports or service centers and directed edges represent preferred routes recommended by the NFDC.

- **Flight connections** [35] is a directed network representing flights between airports of the world. In this network of 2939 nodes and 30501 edges, nodes corresponds to airports and directed edges represent a direct flight from one airport to another airport.

- **Twitter Lists** [10] is a directed network of connections between Twitter users. In this network of 23370 nodes and 33101 edges, nodes correspond to users and directed edges signify that the left user follows the right user on Twitter.

- **Physicians** [42] is a directed network representing innovation among 246 physicians in different towns of Illinois, Peoria, Bloomington, Quincy and Galesburg in USA based on data collected in 1966. In this network of 241 nodes and 1098 edges, nodes corresponds to physicians and directed edges signify that the left physician mentioned that the right physician is his friend or that the left physician turns to the right physician for advice when needed.

Note that we take the weight of nodes and edges equal to 1 while computing the Forman curvature in above mentioned directed and unweighted networks.
In addition, we have also considered the following directed and weighted real networks:

- **US Airport** [36] represents a directed network with positive edge weights which was generated based on flights between US airports in 2010. In this network of 1574 nodes and 28236 edges, nodes correspond to airports and directed edges represent flight connections between airports. The weight of each edge is proportional to the number of flights between the connecting cities.

- **Florida Ecosystem** [43] represents a directed network with positive edge weights which captures the ecological network of carbon exchanges in the Cypress wetlands of South Florida during the dry season. In this network of 128 nodes and 2137 edges, nodes correspond to taxa and directed edges signify that one taxon utilizes another taxon as food source.

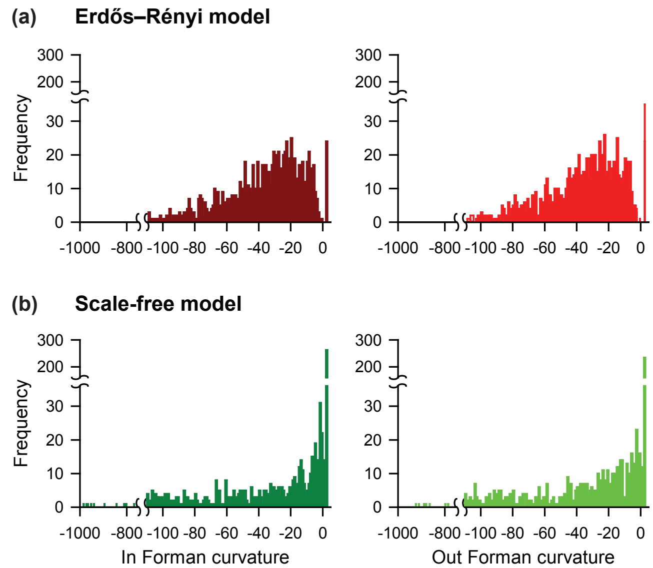

FIG. 3. Distribution of In Forman curvature and Out Forman curvature of nodes in model networks. (a) Erdös-Rényi (ER) model with parameters: number of nodes $n = 1000$ and probability p that two nodes in the graph are directly connected $= 0.00335$. (b) Scale-free model with parameters: number of nodes $n = 1000$, probability of randomly adding a new node connected to an existing node based on the in-degree distribution $\alpha = 0.142$, probability for adding an edge between two existing nodes $\beta = 0.716$, and probability of randomly adding a new node connected to an existing node based on the out-degree distribution $\gamma = 0.142$.

- **Advogato** [44] is a directed network with positive edge weights which is based on an online community platform for developers of free software launched in 1999. In this network of 5155 nodes and 47135 edges, nodes correspond to users of Advogato and directed edges represent trust relationships between users. A trust link is called a certification on Advogato.

- **Adolescent health** [45] is a directed network with positive edge weights which was generated based on a survey in 1994-95 where each student was asked to list his or her 5 best female and 5 best male friends. In this network of 2539 nodes and 12969 edges, nodes correspond to students and directed edges between two students signify that the left student chose the right student as a friend.

Lastly, we have also considered the following directed and spatial real networks:

- ***C. elegans* neural network** [46] represents global neural network of the organism *Caenorhabditis elegans*. In this network of 277 nodes and 2105 edges, nodes are neurons and directed edges give connections between the neurons. Note that this network is both directed and spatial unlike above mentioned real networks. In this network, edge weights correspond to the cartesian distance between the start and end nodes of an edge.

- **Macaque neural network** [46] gives the cortical connectivity network within one hemisphere of Macaque monkey. In this network of 94 nodes and 2390 edges, nodes correspond to cortical regions and directed edges signify links between each region. This network is also a spatial network, and edge weights correspond to the cartesian distance between the start and end nodes of an edge.

Several of the above mentioned real directed networks were downloaded from the KONECT [47] database (Supplementary Table S1). The structure of the considered directed networks was analyzed using common network measures such as maximum degree, minimum degree, average degree, average in-degree, average out-degree, size of the largest weakly connected component and degree assortativity, and these results are contained in Supplementary Table S2.

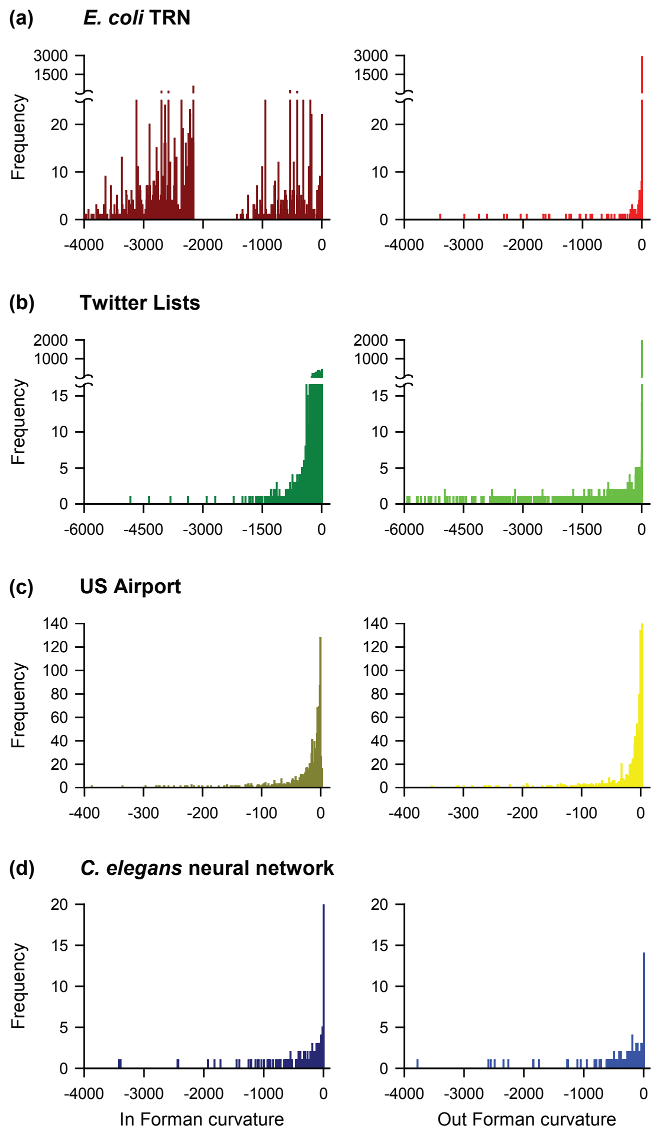

FIG. 4. Distribution of In Forman curvature and Out Forman curvature of nodes in real networks. (a) *E. coli* TRN. (b) Twitter Lists. (c) US Airport. (d) *C. elegans* neural network.

IV. RESULTS AND DISCUSSION

A. Curvature distribution in networks

We have shown that two networks with the same degree distribution can have different distribution of Forman curvatures (Figure 1). In Figure 3, we show the distribution of In Forman and Out Forman curvature of nodes in two models of directed networks with very different degree distribution. It is seen that most nodes in both models of directed networks have negative curvature. One can clearly distinguish the two models of directed networks, ER and scale-free, by the observed nature of distribution for In Forman curvature and Out Forman curvature (Figure 3). In random ER networks, the distribution of both node curvatures is narrow with most nodes having In Forman curvature and Out Forman curvature close to -20. In scale-free networks, the distribution of both node curvatures is broad with several nodes having In Forman curvature and Out Forman curvature less than -100. Also, random ER networks have narrow distribution of edge curvatures while scale-free networks have broader distribution of edge curvatures. These results for Forman curvature in directed networks extend our recent results for undirected networks [21].

In Figure 4, we show the distribution of In Forman curvature and Out Forman curvature of nodes in real directed

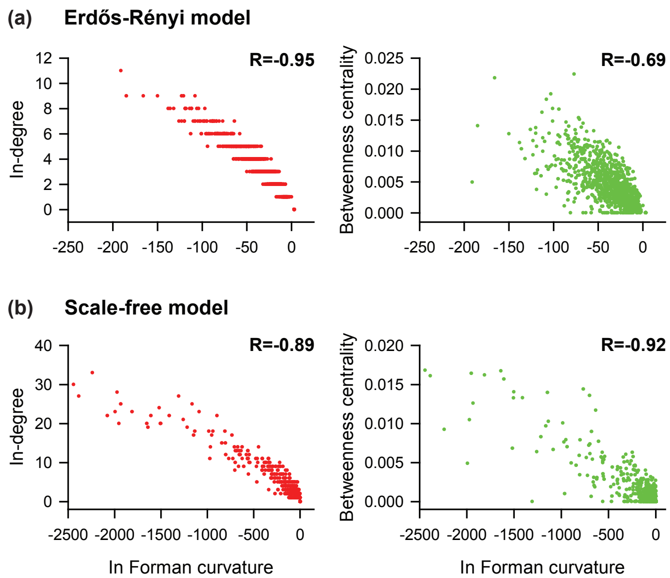

FIG. 5. Correlation between In Forman curvature and in-degree or betweenness centrality of nodes in model networks. (a) Erdös-Rényi (ER) model. (b) Scale-free model. The parameters used to construct graphs from the generative models are same as those mentioned in Figure 3. We also indicate the Pearson correlation coefficient R for each case.

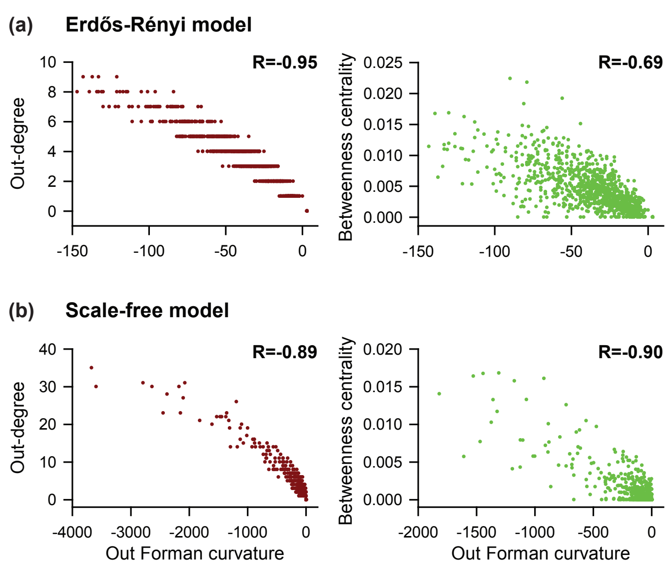

FIG. 6. Correlation between Out Forman curvature and out-degree or betweenness centrality of nodes in model networks. (a) Erdös-Rényi (ER) model. (b) Scale-free model. The parameters used to construct graphs from the generative models are same as those mentioned in Figure 3. We also indicate the Pearson correlation coefficient R for each case.

networks. Similar to models of directed networks, it is seen that most nodes in considered real directed networks have negative curvature. The considered real directed networks have a broad distribution of both In Forman curvature and Out Forman curvature (Figure 4). Also, it is clear that the nature of distribution of In Forman curvature and Out Forman curvature in three considered real directed networks, Twitter Lists, US Airport and *C. elegans* neural network, is similar to that of scale-free directed network. These observations are expected as most real networks have a scale-free architecture with power-law degree distribution [3]. We remark that US Airports is a directed and weighted real network where edge weights are proportional to the number of flights between the connecting cities, while *C. elegans* neural network is a directed and spatial network where edge weights correspond to the cartesian

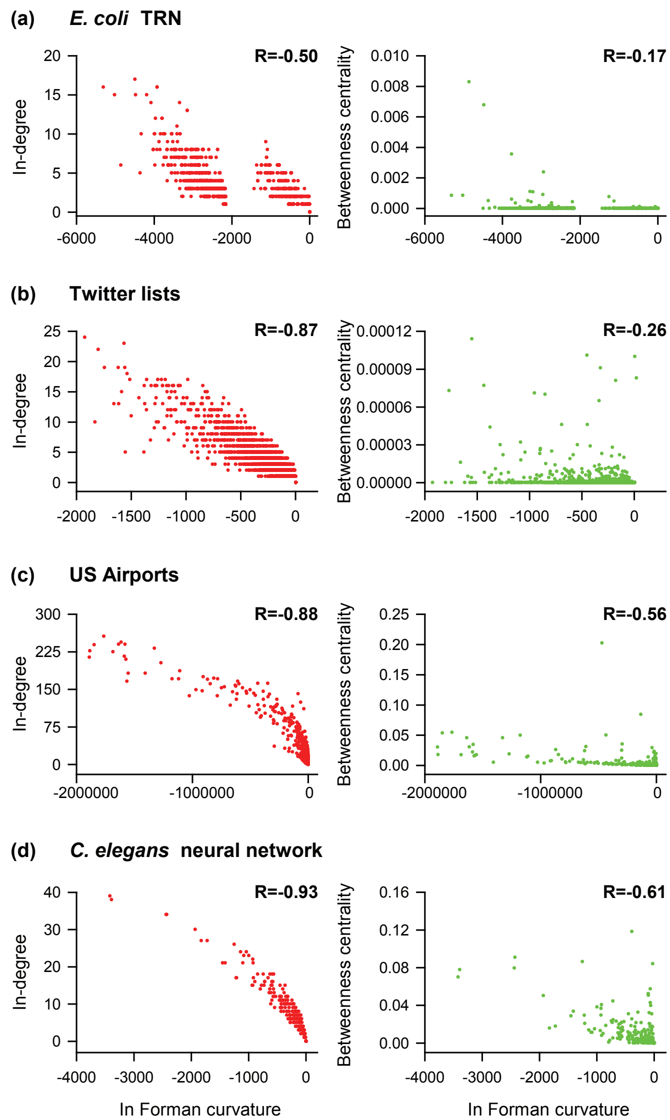

FIG. 7. Correlation between In Forman curvature and in-degree or betweenness centrality of nodes in real networks. (a) *E. coli* TRN. (b) Twitter Lists. (c) US Airport. (d) *C. elegans* neural network. We also indicate the Pearson correlation coefficient R for each case.

distance between the start and end nodes of an edge.

Interestingly, the observed distribution of In Forman curvature is very different from that of Out Forman curvature in the *E. coli* TRN (Figure 4(a)). Bacterial transcriptional regulatory networks (TRNs) display an inherently hierarchical architecture [40, 48–50] where few transcriptional factors at the top of the hierarchy have no incoming links (i.e., their in-degree is zero), while a large number of target genes at the bottom of the hierarchy have no outgoing links (i.e., their out-degree is zero). Moreover, the out-degree distribution of *E. coli* TRN is broad with many global transcription factors regulating several target genes, while the in-degree distribution is narrow due to constraints on size of promoter regions in the bacterial genome [51]. This asymmetry, but not the hierarchical structure by itself, seems to explain the observed difference in the distribution of In Forman curvature and Out Forman curvature in the *E. coli* TRN (Figure 4(a)).

Recall that Ricci curvature controls the growth of volumes in the classical Riemannian setting [22, 52]. Thus, spaces with negative Ricci curvature have exponential type growth while those with positive Ricci curvature have a finite diameter, and the result also holds for the Forman's discretization of the Ricci curvature [28]. Thus, our observation that most nodes and edges in considered directed networks have a negative curvature suggests that these networks have potential of infinite growth.

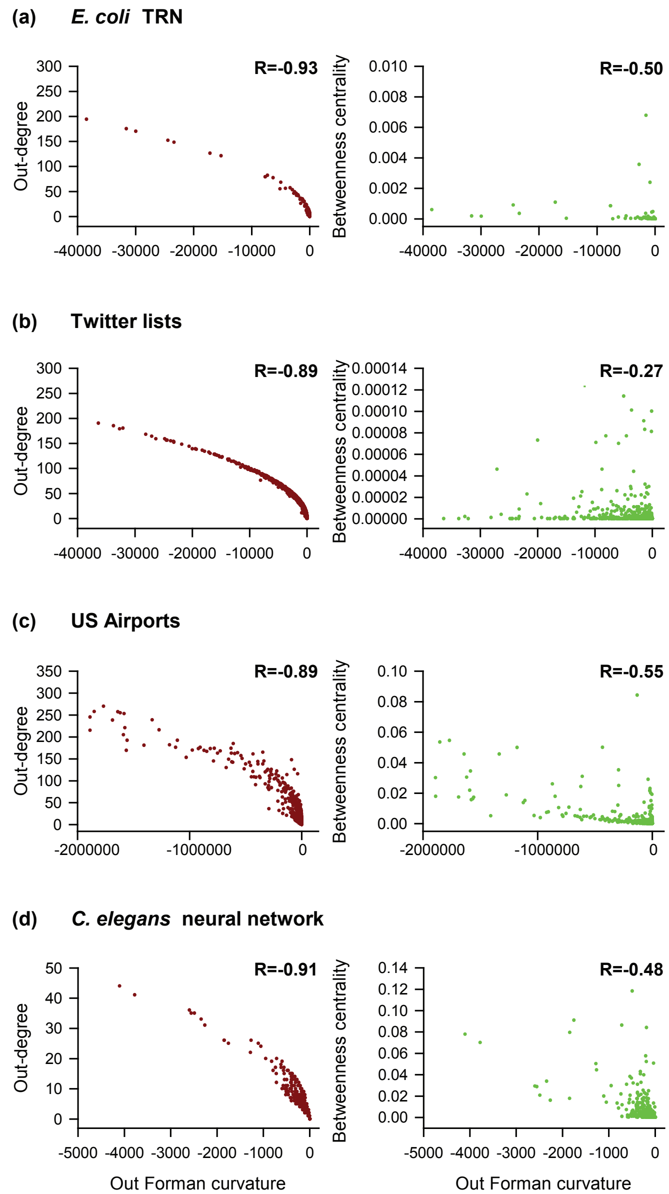

FIG. 8. Correlation between Out Forman curvature and out-degree or betweenness centrality of nodes in real networks. (a) *E. coli* TRN. (b) Twitter Lists. (c) US Airport. (d) *C. elegans* neural network. We also indicate the Pearson correlation coefficient R for each case.

B. Curvature and common network measures

In-degree of a node gives the number of edges incident to the node while out-degree of a node gives the number of edges originating from the node in a directed network. We show in Figure 5 the correlation between in-degree and In Forman curvature of nodes, and in Figure 6 the correlation between out-degree and Out Forman curvature of nodes in the two models of directed networks. We find a high negative correlation between In Forman curvature and in-degree of a node in the two models of directed networks (Figure 5). A similar high negative correlation is obtained between Out Forman curvature and out-degree of a node in the two models of directed networks (Figure 6). These results are expected because the In Forman curvature (respectively, the Out Forman curvature) of a node is computed based on Forman curvature of edges incident on (respectively, originating from) the node, and the Forman curvature of an edge in the network is dependent on its neighboring edges. In particular, in the unweighted case, the curvature becomes more negative the higher the degree of the corresponding node. Also, these results for In Forman curvature and Out

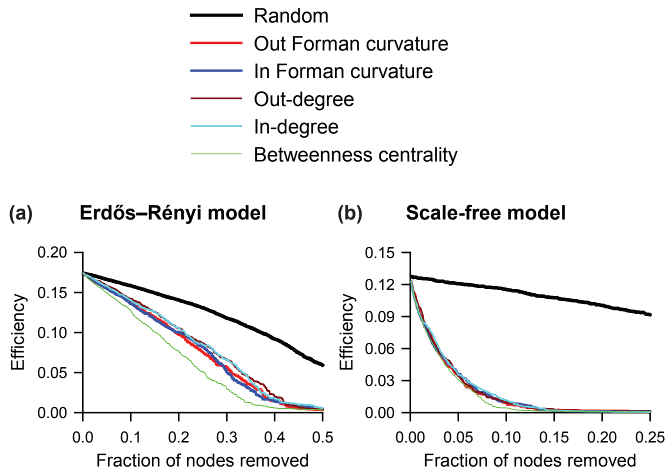

FIG. 9. Communication efficiency as a function of the fraction of nodes removed in models of directed networks. (a) Erdös-Rényi (ER) model. (b) Scale-free model. In this figure, the order in which the nodes are removed is based on the following criteria: Random order, Increasing order of In Forman curvature, Increasing order of Out Forman curvature, Decreasing order of in-degree, Decreasing order of out-degree, and Decreasing order of betweenness centrality. The parameters used to construct graphs from these generative models are same as those mentioned in Figure 3.

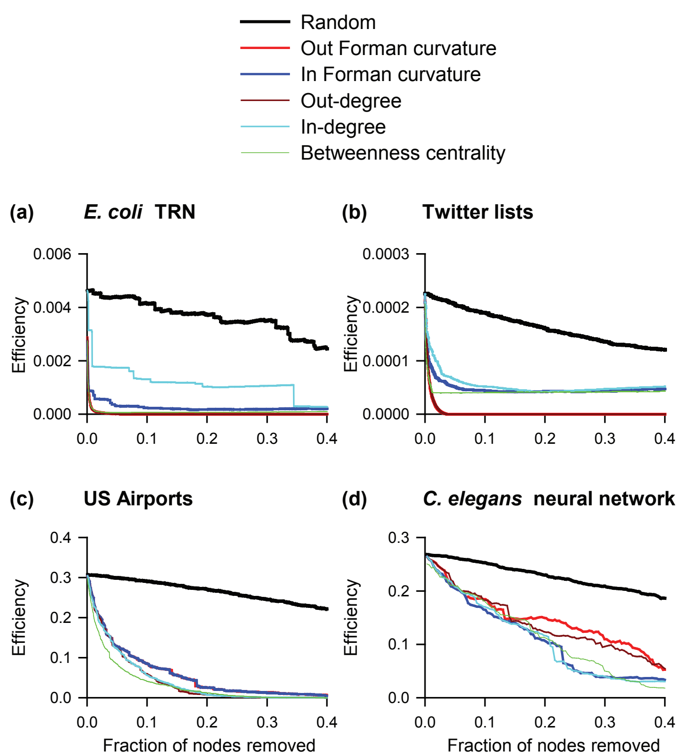

FIG. 10. Communication efficiency as a function of the fraction of nodes removed in real directed networks. (a) *E. coli* TRN. (b) Twitter Lists. (c) US Airport. (d) *C. elegans* neural network. In this figure, the order in which the nodes are removed is based on the following criteria: Random order, Increasing order of In Forman curvature, Increasing order of Out Forman curvature, Decreasing order of in-degree, Decreasing order of out-degree, and Decreasing order of betweenness centrality.

Forman curvature in the two models of directed networks reinforce our recent results for Forman curvature in models of undirected networks [21].

Figure 7 shows the correlation between in-degree and In Forman curvature of nodes, and Figure 8 shows the correlation between out-degree and Out Forman curvature of nodes in real directed networks. Similar to models of directed networks, we find a high negative correlation between In Forman curvature and in-degree of a node (Figure 7), and between Out Forman curvature and out-degree of a node (Figure 8), in considered real directed networks. Supplementary Table S3 summarizes the results of these comparative analyses between In Forman curvature, Out Forman curvature, in-degree and out-degree of nodes in considered directed networks.

Betweenness centrality [4, 53] of a node gives the fraction of shortest paths between all pairs of nodes in the network that pass through that node. This global measure is an indicator of the centrality of a node in the network, and a node with high betweenness centrality is a bottleneck for flows in the network. Interestingly, high negative correlation is obtained between In Forman curvature and betweenness centrality of nodes (Figure 5), and between Out Forman curvature and betweenness centrality of nodes (Figure 6), in the two models of directed networks. Moreover, the magnitude of negative correlation between In Forman curvature and betweenness centrality of nodes is higher than the positive correlation between in-degree and betweenness centrality of nodes in the two models of directed networks (Supplementary Table S3). Also, the magnitude of negative correlation between Out Forman curvature and betweenness centrality of nodes is higher than the positive correlation between out-degree and betweenness centrality of nodes in the two models of directed networks (Supplementary Table S3).

In the considered real directed networks, moderate to high negative correlation is obtained between In Forman curvature and betweenness centrality of nodes (Figure 7), and between Out Forman curvature and betweenness centrality of nodes (Figure 8). Similar to models of directed networks, the magnitude of negative correlation between In Forman curvature and betweenness centrality of nodes is higher than the positive correlation between in-degree and betweenness centrality of nodes in most of the considered real directed networks (Supplementary Table S3). Also, the magnitude of negative correlation between Out Forman curvature and betweenness centrality of nodes is higher than the positive correlation between out-degree and betweenness centrality of nodes in most of the considered real directed networks (Supplementary Table S3). Thus, we find that there is a high negative correlation between the global measure, betweenness centrality, and the local measures, In Forman curvature and Out Forman curvature, in model and real directed networks.

Pagerank [54, 55] is an algorithm for directed networks which was originally developed to rank websites by the Google search engine. Pagerank can be employed with any directed network to measure importance of different nodes in the network. The algorithm counts the number and quality of incoming edges to a node to estimate the importance of a node in the network. Thus, we find as expected a high correlation between in-degree and pagerank of a node in the considered model and real directed networks (Supplementary Table S4). Interestingly, we find a high negative correlation between In Forman curvature and pagerank of a node in the considered model and real directed networks (Supplementary Table S4). We also find a moderate to high negative correlation between Out Forman curvature and pagerank in most of the considered model and real directed networks (Supplementary Table S4). Importantly, the magnitude of correlation between In Forman curvature and pagerank of a node in some of the weighted real directed networks is higher than the corresponding correlation between in-degree and pagerank of a node. These results emphasize that Forman curvature can be utilized to estimate the importance of nodes in complex networks.

C. Curvature and robustness of directed networks

We next investigated the effect of removing nodes based on increasing order of their In Forman curvature and Out Forman curvature on the large-scale connectivity of directed networks. Communication efficiency [56] is a measure that captures how efficiently the information can be exchanged across the network, and the measure can be used to quantify a network's resistance to failure in face of small perturbations. Figure 9 shows the communication efficiency in two models of directed networks as a function of the fraction of nodes removed. Here, the order of removing nodes is based on the following criteria: (a) Random order, (b) Increasing order of In Forman curvature (i.e, starting from the node with most negative In Forman curvature), and (c) Increasing order of Out Forman curvature (i.e, starting from the node with most negative Out Forman curvature). It is clear that targeted removal of nodes with highly negative In Forman curvature or highly negative Out Forman curvature leads to faster disintegration compared to random removal of nodes in the two models of directed networks (Figure 9). Figure 10 shows communication efficiency in real directed networks as a function of the fraction of nodes removed. Similar to models of directed networks, we find that targeted removal of nodes with highly negative In Forman curvature or highly negative Out Forman curvature leads to faster disintegration compared to random removal of nodes in real directed networks.

Previous work [2–4, 7, 57–59] has shown that model and real networks are vulnerable to targeted removal of nodes with high degree or high betweenness centrality. We have also compared the effect of removing nodes based on increasing order of their In Forman curvature or Out Forman curvature against removing nodes based on decreasing order of their in-degree, out-degree or betweenness centrality on the large-scale connectivity of networks (Figures 9 and 10). In random ER directed networks, we find that the removal of nodes based on decreasing order of betweenness centrality leads to slightly faster distintegration compared to removal of nodes based on increasing order of In Forman curvature or Out Forman curvature (Figure 9). In scale-free directed networks, we find that the removal of nodes based on decreasing order of betweenness centrality has an effect similar to removal of nodes based on increasing order of In Forman curvature or Out Forman curvature (Figure 9). However, in several of the considered real directed networks, we find that removal of nodes based on increasing order of In Forman curvature or Out Forman curvature leads to faster distintegration compared to removal of nodes based on in-degree, out-degree or betweenness centrality (Figure 10). Our results highlight that nodes with highly negative In Forman curvature or highly negative Out Forman curvature are important for maintaining the large-scale connectivity of directed networks.

V. CONCLUSIONS

In this paper, we have extended our recent adaptation of the Forman curvature for undirected networks to directed networks. The mathematical expression for the In Forman curvature (Equation 4) and the Out Forman curvature (Equation 5) for directed networks elegantly incorporates the node weights, edge weights and edge direction in networks. The distribution of In Forman curvature and Out Forman curvature is narrow in random directed networks, while the distribution is broad in scale-free directed networks. In most real directed networks, the distribution of In Forman curvature and Out Forman curvature is also broad like scale-free networks. These results highlight that the distribution of Forman curvature can also be employed to distinguish and classify different types of directed networks. We next investigated the correlation between In Forman curvature or Out Forman curvature with common network measures such as in-degree, out-degree and betweenness centrality in considered directed networks. We found a high negative correlation between In Forman curvature and in-degree and between Out Forman curvature and out-degree in considered directed networks. We also find a significant negative correlation between In Forman curvature or Out Forman curvature and betweenness centrality in considered directed networks. By investigating the effect of removing nodes based on their In Forman curvature or Out Forman curvature on the communication efficiency of networks, we show that both model and real directed networks are vulnerable to targeted removal of nodes with highly negative In Forman curvature or highly negative Out Forman curvature. Moreover, we have shown that the above results hold also for real directed networks which are weighted or spatial in nature. Also, the results reported here for directed networks mirror our recent results [21] for undirected networks. Hence, Forman curvature can hereafter be employed to study the structure of both directed and undirected complex networks.

Recently, two different discretizations of the classical Ricci curvature, Ollivier-Ricci curvature [13, 25, 27] and Forman-Ricci curvature [28] have been introduced into graph theory. Whereas Ollivier's curvature has already been systematically investigated and also applied for the study of complex networks [14–17, 19, 20, 29, 30], in [21], we have initiated the investigation of the Forman curvature of empirical networks. While the computation of Forman's curvature in undirected networks is extremely simple, the computation of the more established Ollivier's curvature in undirected networks necessitates solving a linear programming problem associated with optimal mass transport on networks. Hence, the computation of Ollivier's curvature unlike Forman's curvature may not easily scale to extremely large networks such as the world wide web (WWW). Importantly, we have extended here the Forman curvature for undirected networks to the domain of directed networks in a very natural manner, while such an extension of Ollivier's curvature to directed networks is still awaited. Based on our results and the simplicity of the mathematical formulas associated with Forman curvature, we expect this novel curvature measure to be widely used for geometrical characterization of networks.

Previously, the clustering coefficient [1] has been used as a reference measure to quantify curvature in complex undirected networks [12, 60]. The clustering coefficient was originally designed for undirected graphs but the concept has been subsequently extended to directed graphs [61]. In our earlier analyses of undirected networks [21], we had found no or weak correlation between Forman curvature and clustering coefficient in the considered model and real undirected networks. Moreover, Forman curvature in undirected networks as opposed to clustering coefficient was found to have a significant correlation with degree and centrality measures [21]. A natural future direction will be the investigation of the level of association between directed clustering coefficient [61] and In Forman curvature or Out Forman curvature in directed networks. In this direction, let us conclude with the following concrete suggestion. The definition of the Forman curvature is coupled with two fitting Laplacians. One should try to employ these Laplacians associated with the Forman curvature for the denoising of directed biological networks such as transcriptional regulatory networks and signalling networks.

ACKNOWLEDGMENTS

ES and AS thank the Max Planck Institute for Mathematics in the Sciences, Leipzig, for their warm hospitality. AS would also like to thank Anand Pathak for discussions, and acknowledges support from Max Planck-India Mobility Grant, IMSc PRISM project (XII Plan), Ramanujan fellowship (SB/S2/RJN-006/2014) and Department of Science and Technology (DST) start-up project (YSS/2015/000060).

---

[1] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[2] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Albert and A. L. Barabási, Reviews of Modern Physics **74**, 47 (2002).
[4] M. Newman, *Networks: an introduction* (Oxford University Press, 2010).
[5] S. Fortunato, Physics Reports **486**, 75 (2010).
[6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, 2013).
[7] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, Nature **407**, 651 (2000).
[8] R. Sole and M. Montoya, Proceedings of the Royal Society of London B: Biological Sciences **268**, 2039 (2001).
[9] N. B. Ellison, C. Steinfield, and C. Lampe, Journal of Computer-Mediated Communication **12**, 1143 (2007).
[10] J. J. McAuley and J. Leskovec, in *NIPS*, Vol. 2012 (2012) pp. 548–56.
[11] L. Šubelj and M. Bajec, The European Physical Journal B **81**, 353 (2011).
[12] J. Eckmann and E. Moses, Proceedings of the National Academy of Sciences **99**, 5825 (2002).
[13] Y. Ollivier, Journal of Functional Analysis **256**, 810 (2009).
[14] Y. Lin and S. Yau, Math. Res. Lett **17**, 343 (2010).
[15] Y. Lin, L. Lu, S. Yau, *et al.*, Tohoku Mathematical Journal **63**, 605 (2011).
[16] F. Bauer, J. Jost, and S. Liu, Mathematical research letters **19**, 1185 (2012).
[17] S. Jost, J.and Liu, Discrete & Computational Geometry **51**, 300 (2014).
[18] Z. Wu, G. Menichetti, C. Rahmede, and G. Bianconi, Scientific Reports **5** (2015).
[19] C. Ni, Y. Lin, J. Gao, X. D. Gu, and E. Saucan, in *IEEE Conference on Computer Communications (INFOCOM)* (IEEE, 2015) pp. 2758–2766.
[20] R. Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum, Scientific Reports **5** (2015).
[21] R. Sreejith, K. Mohanraj, J. Jost, E. Saucan, and A. Samal, to appear in JSTAT (arXiv:1603.00386) (2016).
[22] J. Jost, *Riemannian Geometry and Geometric Analysis* (Springer, 2011).
[23] Y. Shavitt and T. Tankel, in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 1 (IEEE, 2004).
[24] E. Saucan and E. Appleboim, in *Pattern Recognition and Image Analysis* (Springer, 2005) pp. 405–412.
[25] Y. Ollivier, Probabilistic approach to geometry **57**, 343 (2010).
[26] O. Narayan and I. Saniee, Physical Review E **84**, 066108 (2011).
[27] Y. Ollivier, Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the 50th Séminaire de Mathématiques Supérieures (SMS), Montréal, 2011 , 197 (2013).
[28] R. Forman, Discrete and Computational Geometry **29**, 323 (2003).
[29] B. Loisel and P. Romon, Axioms **3**, 119 (2014).
[30] R. Sandhu, T. Georgiou, and A. Tannenbaum, arXiv:1505.05182.
[31] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[32] A. Ma'ayan, S. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. Eungdamrong, G. Weng, P. Ram, J. Rice, A. Kershenbaum, G. Stolovitzky, R. Blitzer, and R. Iyengar, Science **309**, 1078 (2005).
[33] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, Trends in Cognitive Sciences **8**, 418 (2004).
[34] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Computer Networks **33**, 309 (2000).
[35] T. Opsahl, F. Agneessens, and J. Skvoretz, Social Networks **32**, 245 (2010).
[36] T. Opsahl, (2011).
[37] P. Erdös and A. Rényi, Bull. Inst. Internat. Statist **38**, 343 (1961).
[38] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, in *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms* (Society for Industrial and Applied Mathematics, 2003) pp. 132–139.
[39] H. Salgado, M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muniz-Rascado, J. Garcia-Sotelo, V. Weiss, H. Solano-Lira, I. Martinez-Flores, A. Medina-Rivera, G. Salgado-Osorio, S. Alquicira-Hernandez, K. Alquicira-Hernandez, A. Lopez-Fuentes, L. Porron-Sotelo, A. Huerta, C. Bonavides-Martnez, Y. Balderas-Martinez, L. Pannier, M. Olvera, A. Labastida, V. Jimenez-Jacinto, L. Vega-Alvarado, V. del Moral-Chavez, A. Hernandez-Alvarez, E. Morett, and J. Collado-Vides, Nucleic Acids Research **41**, D203 (2013).
[40] S. Kumar, M. Vendruscolo, A. Singh, D. Kumar, and A. Samal, Molecular BioSystems **11**, 930 (2015).
[41] R. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu,

M. Abu-Farha, J.-P. Lambert, H. Duewel, I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.-L. Adams, M. Moran, G. Morin, T. Topaloglou, and D. Figeys, Molecular Systems Biology **3** (2007).
[42] J. Coleman, E. Katz, and H. Menzel, Sociometry **20**, 253 (1957).
[43] R. E. Ulanowicz and D. L. DeAngelis, FY97: The Florida Bay Ecosystem , 20688 (1998).
[44] P. Massa, M. Salvetti, and D. Tomasoni, in *Dependable, Autonomic and Secure Computing.* (IEEE, 2009) pp. 658–663.
[45] J. Moody, Social Networks **23**, 261 (2001).
[46] M. Kaiser and C. Hilgetag, PLoS Comput Biol **2**, e95 (2006).
[47] J. Kunegis, in *Proceedings of the 22nd international conference on World Wide Web companion* (2013) pp. 1343–1350.
[48] G. Balazsi, A.-L. Barabási, and Z. Oltvai, Proceedings of the National Academy of Sciences **102**, 7841 (2005).
[49] H. Yu and M. Gerstein, Proceedings of the National Academy of Sciences **103**, 14724 (2006).
[50] A. Samal and S. Jain, BMC Systems Biology **2**, 1 (2008).
[51] A. L. Barabási and Z. N. Oltvai, Nature Reviews Genetics **5**, 101 (2004).
[52] M. Berger, Notices of the AMS **47**.
[53] L. C. Freeman, Sociometry , 35 (1977).
[54] L. Page, S. Brin, R. Motwani, and T. Winograd, (1999).
[55] A. Langville and C. Meyer, SIAM review **47**, 135 (2005).
[56] V. Latora and M. Marchiori, Physical Review Letters **87**, 198701 (2001).
[57] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).
[58] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang, Journal of Biomedicine and Biotechnology **2005**, 96 (2005).
[59] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, PLoS Computational Biology **3**, e59 (2007).
[60] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Vol. 319 (Springer, 1999).
[61] G. Fagiolo, Physical Review E **76**, 026107 (2007).