

SCIENTIFIC REPORTS



OPEN

Cycle Based Network Centrality

Xiaoping Zhou^{1,2}, Xun Liang¹, Jichao Zhao³ & Shusen Zhang¹

Paths and cycles are the two pivotal elements in a network. Here, we demonstrate that paths, particularly the shortest ones, are incomplete in information network. However, based on such paths, many network centrality measures are designed. While extensive explorations on paths have been made, modest studies focus on the cycles on measuring network centrality. We study the relationship between the shortest cycle and the shortest path from extensive real-world networks. The results illustrate the incompleteness of the shortest paths on measuring network centrality. Noticing that the shortest cycle is much more robust than the shortest path, we propose two novel cycle-based network centrality measures to address the incompleteness of paths: the shortest cycle closeness centrality (SCC) and the all cycle betweenness centrality (ACC). Notwithstanding we focus on the network centrality problem, our findings on cycles can be applied to explain the incompleteness of paths in applications and could improve the applicability into more scenarios where the paths are employed in network science.

Network science is the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena¹. Currently, network science has become a common paradigm in diverse disciplines as a means of analyzing complex relational data. In social, biological, communication, information, or transportation networks, among others, it is important to know the relative structural prominence of nodes to identify the key elements in the network^{2–4}. Network centrality measure, which illustrates the importance of a node's position in a network, is one of the fundamental problems in network science⁵.

In the past decades, many centrality measures were proposed for better understanding, analyzing, and controlling the spreading dynamics, e.g., degree centrality (DC), closeness centrality (CC)^{6–11}, betweenness centrality (BC)⁶, eigenvector centrality (EC)¹², flow betweenness centrality (FBC)¹³ and so on^{14–18}. Obviously, many closeness measures heavily rely on the shortest paths. Take CC as an example. CC somehow evaluates the efficiency of a node while diffusing information to all the other nodes. Undoubtedly, the shortest path is the most efficient way to deliver information between two nodes. However, the shortest path is not necessarily the quickest way when considering the presence of possible traffic congestion¹⁹. In this scenario, a portion of the “information” may flow through non-shortest paths. To this extent, the shortest path is incomplete.

To address the incompleteness of shortest paths, some betweenness centrality measures are proposed by introducing non-shortest paths. Freeman *et al.* suggested a more sophisticated network centrality measure, teamed as flow betweenness centrality (FBC)¹³. Although FBC utilizes contributions from both shortest paths and non-shortest paths, Newman illustrated that FBC can give counterintuitive results in some cases²⁰. To attack the issues of shortest path and flow betweenness, Newman proposed and offered the random-walk betweenness centrality (RWBC)²⁰. Information is implicitly assumed to spread through random paths in RWBC. However, it is expected that information can be delivered in more efficient paths. Estrada *et al.* defines the communicability²¹ between two nodes and proposes the communicability betweenness centrality (CBC)¹⁸ in complex networks. CBC incorporates all paths between two nodes but penalizing long walks. Since CBC utilizes the “backtracking walks” with loops, it cannot be applied to the scenarios where the loop is not allowed, e.g., the communication routing network. Positively, each of the above solutions has a valid role in accounting for processes that can take place in the network. However, these solutions produce superfluous paths between two nodes in many real-world applications such as communication routing network where information is expected to be transmitted through the optimal paths. Additionally, the above three centrality measures are computationally expensive²², since many unnecessary paths are required to be computed. Another characteristic of the incompleteness of shortest paths in information dissemination is the lack of fault tolerance. In many real-world applications such as communication routing network, factors such as fault tolerance must be considered in case that one of the nodes or links are unavailable due to malfunctions or congestions. This study addresses the incompleteness of the shortest path in CC from a novel perspective by introducing shortest cycles.

¹School of Information, Renmin University of China, Beijing, 100872, China. ²Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China.

³CNPC managers training institute, Beijing, 100096, China. Correspondence and requests for materials should be addressed to X.L. (email: xliang@ruc.edu.cn)

In contrast with path, a cycle consists two loop-free independent paths for any two nodes in the cycle. Thus, the shortest cycle mitigates the incompleteness of the shortest path in CC from two aspects: providing an alternative path and improving the fault tolerance. From this perspective, shortest cycles are more robust than shortest paths in measuring network centrality. Errors, failures and environmental changes may occur at any time in most real-world networks²³, resulting in the dysfunction of shortest paths. It is anticipated that shortest cycles can improve the resilience of the node-ranking results and help identify nodes most central for robust information transmission.

This study employs the power of the shortest cycles and proposes the cycle-based network centrality measures. Firstly, we explore the network structure from the perspective of cycle. The experimental results show that network consists of numerous cycles interweaving in the same network, enabling the possibility to employ the power of cycles in measuring robust network centrality. Secondly, we study the relationship between the shortest path and the shortest cycle. The empirical study results show that the size of shortest cycle is not linear with the length of the shortest path for many nodes. This finding shows that the paths between two nodes are often nonlinear, and that the length of the shortest path cannot characterize the information transmission distance when any node/edge in the shortest path is unavailable. Since the shortest cycle considers the nonlinearity of the paths between two nodes and provides the fault tolerance in information diffusion, it can characterize a more robust information transmission distance than the shortest path. Finally, we propose a novel centrality measure to address the incompleteness of the shortest paths, termed as shortest cycle closeness centrality (SCC). In addition, we also present the all cycle betweenness centrality (ACC) to illustrate that the cycles can also be applied to measure the betweenness centrality. Finally, we give examples that the cycle-based centrality measures can rank nodes more robust in transmitting information to all the other nodes than other path-based network centralities, e.g., CC.

This work discusses the incompleteness of the shortest paths in CC and presents the cycle based network centralities from a different perspectives. Although we focus on network centrality problem, it is anticipated that cycles can be applied to improve the applicability into more scenarios where paths are utilized.

Cycle and Shortest Cycle

Usually, a real-world network is modeled as graph $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ edges, where V and E represent the sets of nodes and edges, respectively. In this study, we only focus on the simple undirected unweighted graph, i.e., $(v, v) \notin E$ for any $v \in V$, or there is no duplicate edge in E .

Cycle and its Computing Method. Paths and cycles are two pivotal elements in a network. Cycles play a role in constructing network architecture something like the status of benzene ring in chemistry. A simple cycle is a sequence of unduplicated consecutive nodes beginning and ending at the same node, with each two consecutive nodes in a sequence adjacent to the next. The size of a simple cycle is the number of edges in the cycle. Without a specification, we only discuss simple cycles in this study.

The possible number of existing simple cycles in a given social network is a conundrum to researchers. This issue was studied extensively in 1960s^{24–27} and 1970s^{28–31}. Due to its relevance in many scientific areas, computation of simple cycle has attracted both theoretical and practical interests. Recently, a large number of studies on this topic have been reported in many disciplines^{24–33}. However, even the most efficient algorithm cost $O(m + \sum_{o \in C(G)} |o|)$, where $C(G)$ is the set of all the cycles in G , and $|\bullet|$ is the size of cycle³³. Due to the exponentially large numbers of cycles in a given social network, cycle computation requires geometrically large quantities of time to enumerate the cycles. Thus, the computational difficulty is an inescapable obstruction in exploring the properties of cycles in social networks and the existing solutions can only be applied to determine the number of cycles in a network consisting of up to tens of nodes. The computation challenge in large network hinders the applications of cycles.

In this work, the massive data problem on social network was resolved by studying the estimated number of cycles in a social network. We denote the number of cycles with size l ($2 < l < n$) as s_l . In complete graph where any two nodes exist an edge, $s_l = P_l^l/l$ where P is the permutation. Obviously, s_l increases exponentially with size l in complete graph (see Supplementary Fig. 1) and cannot be processed by a regular computer. To investigate the characteristics of cycles with thousands of nodes in social networks, we present a workable method based on the model proposed by Marinari and his colleagues³², which can be resolved jointly by the Legendre transformations³⁴, Bethe approximation³⁵ and Belief Propagation³⁶ (Method).

We study the distribution of cycles in social networks. Figure 2 presents the empirical results on three prestigious synthetic networks, Erdős–Rényi random (ER) networks³⁷, Watts–Strogatz small world (SW) networks³⁸ and Barabási–Albert preferential attachment model (BA) networks³⁹. Each of the experimental networks contains 2000 nodes. The results illustrate that s_l exponentially increased with l when l was small. Additionally, in all these three types of networks, the denser network exhibits a more precipitous increase in s_l in comparison to the other types of networks. Figure 3 demonstrates the cycle distribution in four real-world social networks (for the information of all the datasets see in Supplementary Note 1), and also confirms the exponential growth of s_l along with l . Remarkably, all the experimental results with various datasets of social networks show a surprising consistency in their cycle distribution. The s_l experiences an exponential growth with l when l is small. Moreover, the denser the network (or the more edges in the network), the faster the increase in s_l is.

The results show that a network is consisting of many cycles interweaving in the same network. A network presents more robust in information spreading when an alternative path has been provided. Subsequently, a cycle is more robust than a path between two nodes, because two independent paths exist for any two nodes in a cycle. Numerous simple cycles enable the possibility to employ the power of cycles in measuring robust network centrality.

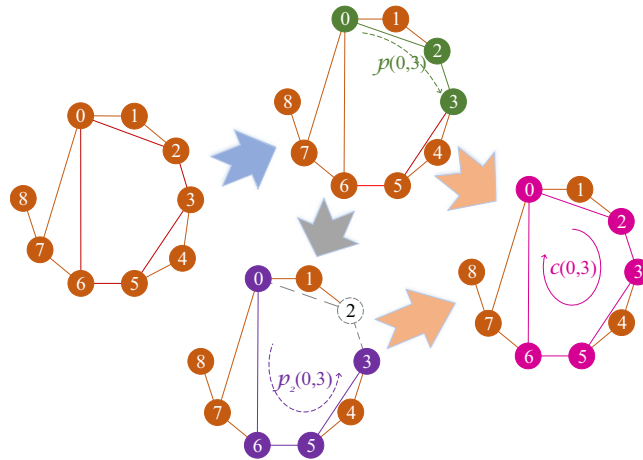


Figure 1. Example of the shortest path and shortest cycle. This figure intuitively presents an example of the shortest path $p(i, j)$, the second shortest path $p_2(i, j)$ and the shortest cycle $c(i, j)$ in graph $G = (V, E)$ for $i, j \in V$, $i \neq j$, where V and E are the sets of all the nodes and edges in G , respectively. Obviously, $p(0, 3) = 0 \rightarrow 2 \rightarrow 3$. When the intermediary of $p(0, 3)$, node 2, is absent from G , it is easy to find that $p_2(0, 3) = 0 \rightarrow 6 \rightarrow 5 \rightarrow 3$. Undoubtedly, $p(0, 3)$ and $p_2(0, 3)$ form $c(0, 3) = 0 \rightarrow 2 \rightarrow 3 \rightarrow 5 \rightarrow 6 \rightarrow 0$.

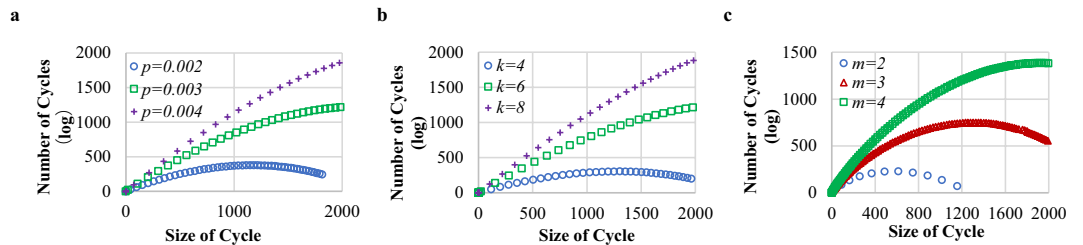


Figure 2. Simple cycle distribution in synthetic networks. The abscissa axes are the size l of simple cycles and the vertical axes are the number of simple cycles s_l in the logarithmic scale. All the synthetic networks have 2000 nodes. The results suggest that the maximum size of simple cycles is no more than 2000. (a) Simple cycle distribution in ER networks. p denotes the probability of an edge existed between two nodes. (b) Simple cycle distribution in SW networks. k represents the number of neighbors of a node, and the probability of rewiring each edge is 0.5. (c) Simple cycle distribution in BA networks. m stands for the number of edges to attach from a new node to existing nodes. For a given l , s_l increases with p in ER networks, with k in SW networks, and with m in BA networks. The curves indicate that the denser the network, the larger s_l has.

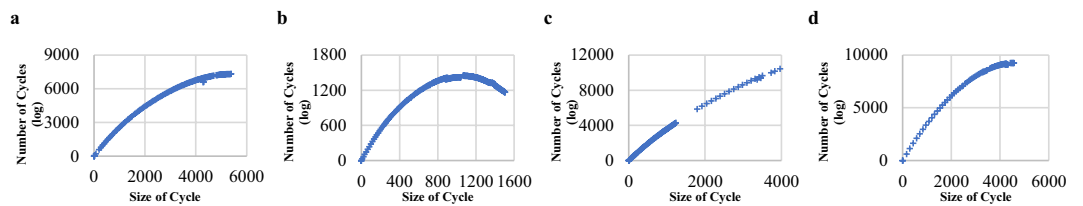


Figure 3. Simple cycle distribution in real-world networks. The abscissa axes are the size l of simple cycles and the vertical axes are the number of simple cycles s_l in the logarithmic scale. (a) Simple cycle distribution in the Sina dataset, which has 5375 nodes and 40224 edges. (b) Simple cycle distribution in the RenRen dataset, which has 1505 nodes and 10019 edges. (c) Simple cycle distribution in the Facebook dataset, which has 3964 nodes and 88159 edges. (d) Simple cycle distribution in the Twitter dataset, which has 4543 nodes and 84212 edges. (Note: the leaf nodes are removed from the original networks).

The Shortest Cycle and the Nonlinearity of Network. Although the shortest path has been widely used in network centrality metrics, it is incomplete for many applications due to its incompleteness in measuring distance or closeness. Take Fig. 1 as an illustration. Both nodes 0 and 3 are the neighbors of node 2. Traditionally, we hold that nodes 0 and 3 have the equal “distance” to node 2. However, what if the information flows in the network? It is anticipated that the information may flow more robust between nodes 0 and 2 than nodes 3 and 2, because the second shortest path between nodes 0 and 2 ($0 \rightarrow 1 \rightarrow 2$) is much shorter than that between nodes 3

and 2 (3 → 5 → 6 → 0 → 2). The shortest path represents the characteristic path between two nodes, and the length of the shortest path describes the characteristic distance of two nodes. In the scenarios where the shortest path has the ability to predict the length of all the other paths (e.g., tree graph), the shortest path can exactly describe the closeness of two nodes. By demonstrating the relationship of the shortest path and the shortest cycle, we show that the network structure in real-world networks is often nonlinear. Thus, the shortest paths can only reveal the characteristic yet partial structure of the networks.

The shortest path between nodes i and j , denoted as $p(i, j)$, is a node sequence with the least nodes or edges starting with i and ending with j without any node or edge reused. The shortest distance between nodes i and j , denoted as $d(i, j)$, is the number of edges in $p(i, j)$. The shortest cycle across nodes i and j , denoted as $c(i, j)$, is the simple cycle across nodes i and j with the least nodes or edges. We denote $l(i, j)$ as the shortest distance measured by the shortest cycle, or the size of $c(i, j)$. Denoting $I_V(i, j)$ and $I_E(i, j)$ as the sets of intermediary nodes and edges in $p(i, j)$, respectively. The second shortest path between nodes i and j , denoted as $p_2(i, j)$, is the shortest path between i and j with $I_V(i, j)$ and $I_E(i, j)$ absent. We represent $d_2(i, j)$ as the size of $p_2(i, j)$. Apparently, $p(i, j)$ and $p_2(i, j)$ form the shortest cycle across nodes i and j , $c(i, j)$. Figure 1 exemplifies a relationship among $p(i, j)$, $p_2(i, j)$ and $c(i, j)$.

Unlike $p(i, j)$, $c(i, j)$ does not exist for some $i, j \in V, i \neq j$ in G . For an instance, node 8 cannot form simple cycle with any other nodes in Fig. 1.

If nodes i and j can form a simple cycle in network G , they are a cyclic node pair, denoted as (i, j) or (j, i) . Reversely, if nodes i and j cannot form a simple cycle, they form an acyclic node pair. The set of all the cyclic and acyclic node pairs in G are written as $H(G)$ and $U(G)$, respectively.

Obviously, the connection relation in an acyclic node pair is unstable. Once some nodes or edges on the shortest path are absent, the connection is broken at an acyclic node pair. Except the node pairs in $U(G)$, we explore the relationship of $d(i, j)$ and $l(i, j)$, or $d_2(i, j)$ and $l(i, j)$.

Due to the numerous number of node pairs in a real-world network, it is labourious to depict all the relationship between $d(i, j)$ and $l(i, j)$. Here we use the characteristic path length and the characteristic cycle size to study the relationship between the shortest path and the shortest cycle.

The characteristic path length of G is defined as

$$D(G) = \frac{2 \times \sum_{i,j \in V, i < j} d(i, j)}{(n - 1) \cdot n} \tag{1}$$

The characteristic cycle size of G is defined as

$$C(G) = \frac{2 \times \sum_{i,j \in V, i < j, (i,j) \in H(G)} l(i, j)}{|H(G)|} \tag{2}$$

and the characteristic cycle size of node i is defined as

$$l(i) = \frac{\sum_{j \in V, i \neq j, (i,j) \in H(G)} l(i, j)}{|H_i(G)|} \tag{3}$$

where $H_i(G)$ denotes the nodes set with $(i, j) \in H(G)$ for $j \in V$.

To mitigate the influence of $U(G)$ in comparing $C(G)$ and $D(G)$, we define the cyclic characteristic path length. Formally, the cyclic characteristic path length of G is defined as

$$D'(G) = \frac{2 \times \sum_{i,j \in V, i < j, (i,j) \in H(G)} d(i, j)}{|H(G)|} \tag{4}$$

Similarly, the characteristic path length of node i is defined as

$$d(i) = \frac{\sum_{j \in V, i \neq j, (i,j) \in H(G)} d(i, j)}{|H_i(G)|} \tag{5}$$

Obviously, $d(i)$ and $l(i)$ represent the average length of $p(i, j)$ and average size of $c(i, j)$ for $j \in V$, respectively. We also define the path incremental coefficient as

$$\eta(G) = \frac{C(G)}{D'(G)} \tag{6}$$

and the path incremental coefficient for node i as

$$\eta(i) = \frac{l(i)}{d(i)} \tag{7}$$

We conducted the experiments on six real-world social networks (for the information of all the datasets see in Supplementary Note 1). In the experiment settings, the acyclic node pairs $U(G)$ were removed.

In our empirical studies, small-world phenomenon³⁸ is observed in all the six networks. Table 1 demonstrates the values of $D(G)$, $D'(G)$, $C(G)$ and $\eta(G)$. No matter what the values of $D'(G)$ or $D(G)$ in the six networks are,

Dataset	# of Nodes	# of Edges	$\mathcal{D}(G)$	$\mathcal{D}'(G)$	$\mathcal{C}(G)$	$\eta(G)$	$\mathcal{L}(G)$ (%)
Dolphin Social Network	62	159	3.357	3.054	6.730	2.204	27.12%
Karate Club	34	78	2.408	2.148	4.924	2.292	29.95%
Twitter	5182	84851	3.026	2.832	5.974	2.109	23.03%
Facebook	4039	88234	3.693	3.340	8.605	2.576	15.95%
Sina Weibo	5375	40224	3.630	3.520	8.084	2.296	7.23%
RenRen	1975	10539	3.261	2.963	6.437	2.173	42.25%

Table 1. Observation of the Real-world Social Networks.

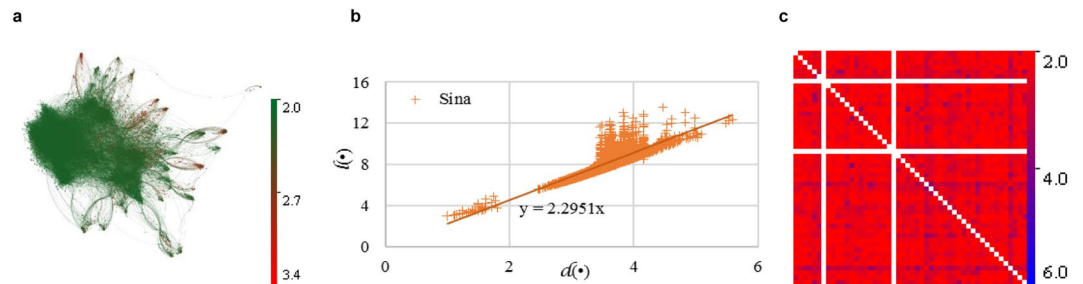


Figure 4. The nonlinearity of network structure from the perspective of the shortest path and shortest cycle in the Sina Weibo dataset. **(a)** The network topology of the Sina Weibo dataset. As the path incremental coefficient $\eta(i)$ increases, the color of the node turns from green to red. Although the path incremental coefficient $\eta(G)$ is around 2.296, $\eta(i)$ varies from 2.048 to 3.403. Moreover, many nodes have an extremely large $\eta(i)$. **(b)** The scatter grams of $l(\bullet)$ and $d(\bullet)$. Obviously, the network witnesses a linear growth of $l(\bullet)$ along with $d(\bullet)$. Strikingly, the slope coefficient values 2.295 and approximately equals to the path incremental coefficient $\eta(G)$. Although $l(\bullet)$ is linear with $d(\bullet)$, a large number of nodes hold $\eta(i) = l(i)/d(i) > \eta(G)$. This phenomenon indicates the shortest path is patchy to measure the distance of two nodes. **(c)** The ratio $\eta(i, j)$ of the size of shortest cycle $l(i, j)$ and length of shortest path $d(i, j)$ in the first 50 nodes. The white color represents an acyclic node pair. The maximal value of $\eta(i, j)$ is 8.0. Although a large portion of the cyclic node pairs hold $\eta(i, j) < 3.0$, $\eta(i, j) > 4.0$ in a large portion of cyclic node pairs.

$\eta(G) \in (2.1, 2.6)$. In particular, in the Twitter dataset, $\eta(G)$ is 2.109, which is the smallest in the six networks. In the Facebook dataset, $\eta(G)$ is 2.576, which is the largest. In all the other four networks, $\eta(G)$ is between 2.2 and 2.3.

In detail, we also explored the relationship of $d(i)$ and $l(i)$ in the six real-world networks. Generally and surprisingly, $d(i)$ increases linearly with $d(i)$ for a large portion of nodes. The slope coefficient of the fit curve of $d(i)$ and $d(i)$ is approximately equal to $\eta(G)$ in all the six networks. However, $d(i)$ is much larger than $\eta(G) \times d(i)$ for many nodes. Figure 4 illustrates an example of the Sina Weibo dataset. Although $\eta(G) = 2.296$, $\eta(i)$ varies from 2.05 to 3.41 in the Sina Weibo dataset. Figure 4a presents the network of the Sina Weibo network. As $\eta(i)$ increases, the color of the node turns from green to red. Obviously, there are plenty of nodes with a pretty high $\eta(i)$. Figure 4b gives the scatter graph of the relationships of $d(\bullet)$ and $l(\bullet)$. Intuitively, $l(\bullet)$ is linear with $d(\bullet)$, with a slope coefficient of 2.295. According to the locations in the scatter graph, the nodes can be directly divided into two categories: the nodes with $d(i) < 2$ and the rest. Obviously, only a few nodes with $d(i) < 2$. It is because that some nodes form a cluster with the nodes, which form cyclic node pairs among them. From most of the nodes, $d(i) > 2.5$. Although we have already remove the acyclic node pairs, we also discovered that a large number of nodes has $\eta(i) > \eta(G)$. In addition, we also observed the values of $\eta(i, j) = l(i, j)/d(i, j)$ for all the node pairs. Figure 4c displays $\eta(i, j)$ of any two nodes in the first 50 nodes. $\eta(i, j)$ varies from 2.0 to 8.0. Obviously, $\eta(i, j)$ is extremely large in a large portion of node pairs. In this study, the phenomenon that the size of the shortest cycle is not linear with the length of the shortest path is termed as nonlinearity. Thus, the length of the shortest path cannot depict the characteristic of the length of the second shortest path in a nonlinear network. Obviously, Sina Weibo network is nonlinear.

The nonlinearity phenomenon was also noticed in all the other five real-world networks (see Supplementary Figs 2 and 7). Although the size of the shortest cycle is linear with the length of the shortest path for a portion of cyclic node pairs, our empirical study results also illustrate that $l(i, j)$ is not linear with $d(i, j)$ for many nodes i and j . Subsequently, the shortest path cannot depict the length of the information flow when the information does not flow through the shortest path, even the second shortest path. From this viewpoint, the shortest path is incomplete to measure the distance of two nodes and illustrate the network centrality. Since the shortest cycle provides the second shortest path besides the shortest path, the shortest cycle is more reasonable than the shortest path to measure the distance of two nodes.

Moreover, the cycle-based closeness can also be applied to the scenarios where factors such as fault tolerance must also be considered in case that one of the nodes or links are unavailable due to malfunctions or congestions,

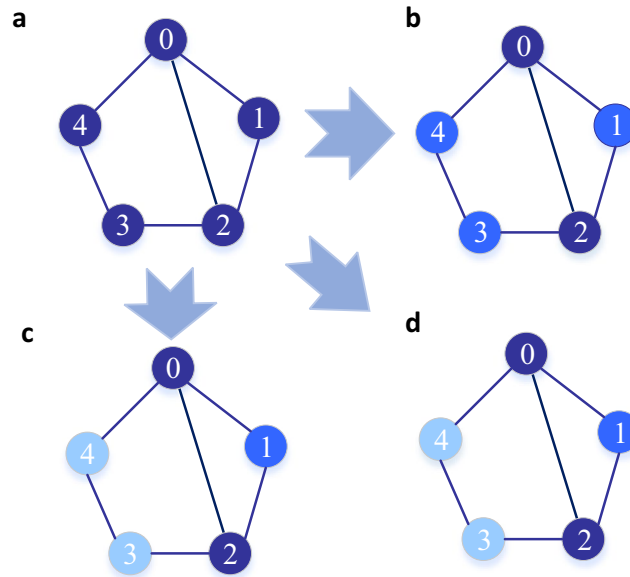


Figure 5. Comparisons of the closeness centrality (CC), the shortest cycle closeness centrality (SCC) and the all cycle closeness centrality (ACC). (a) A network has five nodes and six edges, on which the comparisons of CC, SCC and ACC are made in the paper. (b) The centrality ranking by CC. Obviously, nodes 0 and 2 have the same and top centrality, while nodes 1, 3 and 4 score the next. Intuitively, node 1 should score higher than nodes 3 and 4 in centrality since node 1 is closer to both nodes 0 and 2. (c) The centrality ranking by SCC. Nodes 0 and 2 have the most centrality in the network, node 1 is the next, and nodes 3 and 4 hit the last. This is consistent with the expectation. (d) The centrality ranking by ACC. nodes 0 and 2 have the most centrality in the network, node 1 is the next, and nodes 3 and 4 hit the last. The result is the same as SCC and consistent with the expectation.

e.g., communication routing network. One example is the distances between node 0 and nodes 3/8. Although both nodes 3 and 8 can reach node 0 by one intermediary node, we probably think that node 3 plays a much more important role for node 0 than node 8. Because node 3 has another path to node 0, and also bridges node 0 with other nodes, e.g., node 4. Contrarily, the relationship between nodes 0 and 8 depends on node 7. Once node 7 is inactive or unavailable due to malfunctions or congestions, the relationship between 0 and 8 is broken.

Cycle-Based Network Centralities. In this study, we employ the power of the cycles in the networks and propose two novel network centrality measures: the shortest cycle closeness centrality (SCC) and all cycle betweenness centrality (ACC).

Here we take CC as an instance to illustrate the incompleteness of the shortest paths in measuring network centrality. CC relies on the length of shortest paths from a node to all other nodes in the network and is defined as the inverse total length. The CC of node i , denoted as $C_2(i)$, is defined as

$$C_2(i) = \frac{1}{\sum_{j \in V} d(i, j)}. \quad (8)$$

Clearly, a trivial case is $d(i, i) = 0$ for $i \in V$.

We illustrate the insufficiency of the shortest paths in CC intuitively in a smaller network in Fig. 5.

Example 1 (Example of CC). When CC is utilized in the network in Fig. 5, we have

$$C_2(0) = [d(0, 1) + d(0, 2) + d(0, 3) + d(0, 4)]^{-1} = 1/5,$$

$$C_2(1) = [d(1, 0) + d(1, 2) + d(1, 3) + d(1, 4)]^{-1} = 1/6,$$

$$C_2(2) = [d(2, 0) + d(2, 1) + d(2, 3) + d(2, 4)]^{-1} = 1/5,$$

$$C_2(3) = [d(3, 0) + d(3, 1) + d(3, 2) + d(3, 4)]^{-1} = 1/6,$$

$$C_2(4) = [d(4, 0) + d(4, 1) + d(4, 2) + d(4, 3)]^{-1} = 1/6.$$

As a result, nodes 0 and 2 have the same centrality, and nodes 1, 3 and 4 have a lower, but identical, centrality in the network. Intuitively, node 1 is more cohesive to the network than nodes 3 and 4, because nodes 0, 1 and 2

form a cluster. Moreover, more distances are required to transmit information by removing any edge of nodes 3 or 4 than removing any edge of node 1. From this perspective, we hold that node 1 is more robust than nodes 3 and 4 when diffusing information to all the other nodes, and that the shortest path is incomplete in measuring network centrality.

The SCC of node i , denoted as $C_3(i)$, is defined as

$$C_3(i) = \frac{1}{\sum_{j \in V} l(i, j)}. \quad (9)$$

Generally, $l(i, j) = \infty$ when $(i, j) \in U(G)$. In this scenario, acyclic node pair (i, j) would $C_3(i) = C_3(j) = 0$. To mitigate this problem, we have $l(i, j) = n + d(i, j)$, $(i, j) \in U(G)$.

Before conducting SCC on real-world networks, we demonstrate the power of SCC on a 5-node network, as showed in Fig. 1.

Example 2 (Example of SCC). When SCC is employed in the network in Fig. 1, we have

$$C_3(0) = [l(0, 1) + l(0, 2) + l(0, 3) + l(0, 4)]^{-1} = 1/14,$$

$$C_3(1) = [l(1, 0) + l(1, 2) + l(1, 3) + l(1, 4)]^{-1} = 1/16,$$

$$C_3(2) = [l(2, 0) + l(2, 1) + l(2, 3) + l(2, 4)]^{-1} = 1/14,$$

$$C_3(3) = [l(3, 0) + l(3, 1) + l(3, 2) + l(3, 4)]^{-1} = 1/17,$$

$$C_3(4) = [l(4, 0) + l(4, 1) + l(4, 2) + l(4, 3)]^{-1} = 1/17.$$

Consequently, we can hold that nodes 0 and 2 have the most centrality in the network, node 1 is the next, and nodes 3 and 4 hit the last. The result is consistent with our expectation. As a result, we distinguish the ranking of node 1 with nodes 3 and 4 in the network. Obviously, SCC is more suitable to measure the network centrality than, i.e. CC, in the example network.

Further, we extend the shortest cycle to all the simple cycles in the networks and show that the cycles can also be applicable in measuring betweenness centrality. We denote S_k as the set of simple cycles with size k and s_k as the sizes of S_k . Also, $S_k(i)$ represents the set of simple cycles going through node i and $s_k(i)$ is the size of $S_k(i)$. In real-world networks, s_k increases exponentially with k . Subsequently, the cycles with larger s_k result in less consequence to the nodes lying on the cycles.

The ACC of node i , denoted as $C_A(i)$, is defined as

$$C_A(i) = \sum_{k=3}^n \alpha^{k-2} s_k(i), \quad (10)$$

where $\alpha \in (0, 1)$ is an attenuation factor.

Example 3 (Example of ACC). In the network in Fig. 1, there are three simple cycles in the network: $0 \rightarrow 1 \rightarrow 2 \rightarrow 0$, $0 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 0$ and $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 0$. Subsequently, $s_3 = s_4 = s_5 = 1$. Without a generality, $\alpha = 0.5$. We have

$$C_A(0) = 0.5s_3(0) + 0.25s_4(0) + 0.125s_5(0) = 0.875,$$

$$C_A(1) = 0.5s_3(1) + 0.25s_4(1) + 0.125s_5(1) = 0.625,$$

$$C_A(2) = 0.5s_3(2) + 0.25s_4(2) + 0.125s_5(2) = 0.875,$$

$$C_A(3) = 0.5s_3(3) + 0.25s_4(3) + 0.125s_5(3) = 0.375,$$

$$C_A(4) = 0.5s_3(4) + 0.25s_4(4) + 0.125s_5(4) = 0.375.$$

As a result, we argue that nodes 0 and 2 have the most centrality in the network, followed by node 1, and nodes 3 and 4 locates the most marginal places in the network. Clearly, the ranking result is the same as that in SCC.

We also evaluated the scores of nodes in the example network by RWBC and CBC. As illustrated in Table 2, nodes 0 and 2 have the same score and rank the top 2 in both RWBC and CBC. However, both centrality measures fail to identify the importance of node 1.

The example network explicitly shows that both SCC and ACC are more suitable for ranking nodes than CC in information network. It is because that the shortest cycle provides an independent alternative path for the shortest path. Subsequently, SCC enhances the fault tolerance by employing the second shortest path when transmitting

Node	Network Centrality Measure				
	CC	RWBC	CBC	SCC	ACC
0	0.200	0.424	0.515	0.071	0.875
1	0.167	0.152	0.226	0.063	0.625
2	0.200	0.424	0.515	0.071	0.875
3	0.167	0.242	0.233	0.059	0.375
4	0.167	0.242	0.233	0.059	0.375

Table 2. Scores Calculated using CC, RWBC, CBC, SCC and ACC for the Example Network.

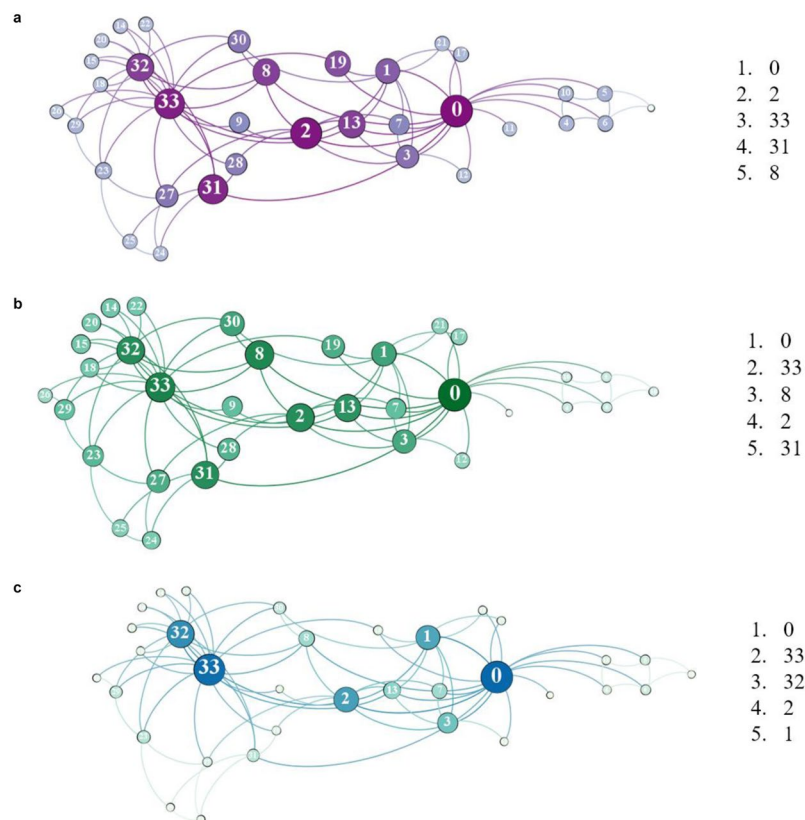


Figure 6. Network centralities of the Karate Club dataset by applying (a) CC, (b) SCC and (c) ACC. In the experiment of ACC, $\alpha = 0.1$. The top five rank nodes are listed on the right of the figures. The Karate Club network naturally has two classes, and nodes 0 and 33 are two leaders of the two classes. Obviously, these two leaders rank the top two in both SCC and ACC, while nodes 33 ranks the third in CC.

information. Similarly, ACC also employs the power of cycles. By introducing cycles, both SCC and ACC are capable of identifying the nodes most central for robust information transmission in information network.

We also demonstrate the power of both SCC and ACC in real-world network. Figure 6 shows the network centralities in the Karate Club network by applying CC, SCC and ACC. The Karate Club network naturally contains two classes, the Office (node 33) and Mr. Hi (node 0). In Fig. 6, nodes on the right of nodes 2 and 8 (including nodes 2 and 8) belong to the Mr. Hi class, and the rest belong to the Office class. The top rank 5 nodes are listed on the right side of the figures. In the experiment of ACC, $\alpha = 0.1$. Although the two leaders of the two classes are contained in the top 5 nodes, one of the two leaders, node 33, ranks the third in CC. Table 3 also presents the top rank 5 nodes evaluated by RWBC and CBC. Interestingly, nodes 0 and 33 rank the top two in both RWBC, CBC, SCC and ACC. That is, RWBC, CBC, SCC and ACC have the ability to identify the two leaders directly for the Karate Club. This is due to the fact that non-shortest paths are considered in RWBC, CBC, SCC and ACC. Moreover, all the five centrality measures accepted the importance of node 2, because node 2 has many links to the nodes in both classes and plays the role of delivering information across the two classes. It is noticeable that nodes 8 and 31 are listed in the top rank 5 nodes in SCC while nodes 32 and 1 are listed in RWBC and CBC instead. Intuitively, nodes 32 and 1 are of paramount importance for the nodes inside their classes. Contrarily, nodes 8 and 31 are valid backups of node 2. Once node 2 is unavailable due to information congestion or malfunctions, nodes

Top Rank 5 Nodes	Network Centrality Measure				
	CC	RWBC	CBC	SCC	ACC
1	0	0	0	0	0
2	2	33	33	33	33
3	33	2	2	8	32
4	31	32	32	2	2
5	8	1	1	31	1

Table 3. Top Rank 5 Nodes Computed by CC, RWBC, CBC, SCC and ACC in the Karate Club Network.

8 and 31 ensure the information transmission across the two classes and between node pairs. To this extent, we hold that SCC has the ability to rank the nodes most central for robust information transmission.

Acceptably, both SCC and ACC cannot compute the network centrality for the leaf nodes, which only have one neighbor. This limitation does not constrain the application of both SCC and ACC. Because the leaf nodes are not prominent nodes, and rank the last in all the network centrality metrics. In addition, ACC can be upgraded to be harmonic in a (not necessarily connected) graph by applying the idea of harmonic centrality⁴⁰.

Conclusions

The shortest path is incomplete in measuring network centrality in information network, because edges and nodes may be unavailable due to malfunctions or congestions. Many solutions have introduced non-shortest paths to address the incompleteness of shortest paths, e.g., FBC, RWBC, CBC. These centrality measures produce many unnecessary paths, resulting in expensive computation time. Moreover, RWBC assumes to spread information in a random path, while CBC utilize the “backtracking walks” with loops. Both cannot be applied to scenarios where information is expected to transfer in optimal paths without loop, e.g., communication routing network. The shortest cycle consists of two independent yet most efficient paths to deliver information between two nodes. Thus, the shortest cycle has the ability to address the incompleteness of the shortest path, improve fault tolerance for information spreading and identify the nodes most central for robust information transmission.

This study addresses the incompleteness of shortest paths in measuring network centrality by introducing shortest cycles. We illustrated that numerous simple cycles exist in social networks, which enables the possibility to employ the cycles to measure the network centrality. Then, we presented the relationship between the shortest path and the shortest cycle in social networks. We found the nonlinearity of the real-world network, which shows that the sizes of the shortest cycles does not increase linearly with the lengths of the shortest path for many node pairs. We demonstrated that these node pairs impose the incompleteness of the shortest path when applying to applications. Noticing that the shortest cycle provides two independent paths, we proposed a novel network centrality measure SCC based on shortest cycles. Apparently, SCC enhances the fault tolerance during information spread. By extending the shortest cycle to all the simple cycles, another novel network centrality ACC is further presented. Finally, empirical studies on example networks illustrate that both SCC and ACC are capable of ranking nodes more robust in transmitting information to all the other nodes. SCC is a closeness centrality based on the shortest cycle, while ACC is a betweenness centrality based on all the cycles. Subsequently, the cycles can be applied to measure both the closeness centrality and the betweenness centrality.

Materials and Methods

Shortest Path and CC. To compute CC, the algorithms for computing the shortest path in an undirected network $G = (V, E)$ are the key issue. For $i, j \in V, i \neq j$, either the *Dijkstra* algorithm⁴¹ or *Depth-first search* (DFS) algorithm⁴² can be employed to calculate $p(i, j)$. In this study, we utilized the DFS algorithm.

Shortest Cycle and SCC. The shortest cycles can be computed by one more *Depth-first search* algorithm after removing the internal nodes/edges from the original network, or solutions of *K-set-cycle problem*⁴³, or schemes on minimum cost⁴⁴. The SCC can be computed directly using shortest cycles. It is easy to deduce that the time complexity is $O(mn)$ for SCC to evaluate all the scores of nodes in a network with n nodes and m edges.

All Cycles and ACC. For a given l , we define $\sigma(\varepsilon) = \ln s_l/n$ as the entropy of l , where $\varepsilon = l/n \in [0, 1]$. Thus, we have $s_l = e^{n\sigma(\varepsilon)}$ and then we convert the measure of s_l into the estimate of $\sigma(\varepsilon)$.

To estimate $\sigma(\varepsilon)$, a probability law $p(c, u) = u^{|c|}/\Psi(u)$ is introduced, where c is a simple cycle and $|c|$ is the size of c . $\Psi(u)$ is the normalization factor, thus $\Psi(u) = \sum_c u^{|c|} = \sum_{l=3}^n 3s_l u^l$. Applying the saddle point method, so that we have

$$\ln \Psi(u) = n \times \max_{\varepsilon} [\sigma(\varepsilon) + \varepsilon \ln u]. \quad (11)$$

Then, $\sigma(\varepsilon)$ can be expressed in terms of Ψ by inverting the standard Legendre transformations³⁴. Finally, an estimate of Ψ can then be obtained by using the Bethe approximation of the corresponding statistical mechanics model³⁵.

In computation, the well-known correspondence between minimization of the Bethe free energy and iterations of the Belief Propagation equations³⁶ were utilized. For any edge (i, j) , in the $(t + 1)$ th iteration, the value of (i, j) is computed as

$$v^{(t+1)}(i, j) = \frac{u \sum_{k \in \Gamma(i) \setminus j} v^{(t)}(k, i)}{1 + \frac{1}{4} u^2 \left[\left(\sum_{k \in \Gamma(i) \setminus j} v^{(t)}(k, i) \right)^2 - \sum_{k \in \Gamma(i) \setminus j} v^{(t)}(k, i)^2 \right]} \quad (12)$$

where $\Gamma(i)$ is defined as the neighbors of node i , and $\Gamma(i) \setminus j$ is the set of the neighbors of node i except node j . When v converges, the value at each edge is an indication of the probability that the edge is present in cycle c . With v , we obtain:

$$l = \sum_{(i,j) \in E} \frac{uv(i, j)v(j, i)}{1 + uv(i, j)v(j, i)} \quad (13)$$

$$\sigma(\varepsilon) = \frac{1}{n} \sum_{i \in V} \ln \left[1 + \frac{1}{4} u^2 \left[\left(\sum_{k \in \Gamma(i)} v(k, i) \right)^2 - \sum_{k \in \Gamma(i)} v(k, i)^2 \right] \right] - \frac{1}{n} \sum_{i,j \in V} \ln[1 + uv(i, j)v(j, i)] - \varepsilon \ln u \quad (14)$$

$$s_l = \exp \left\{ \begin{aligned} & \sum_{i \in V} \ln \left[1 + \frac{1}{2} u^2 \sum_{m \neq n \in \Gamma(i)} v(m, i)v(n, i) \right] \\ & - \sum_{i \in V} \ln \left[1 + u \sum_{j \in \Gamma(i)} (1 + uv(i, j)v(j, i)) \right] - \varepsilon n \ln u \end{aligned} \right\} \quad (15)$$

The marginal for node i can be computed by

$$g_l(i) = \frac{u^2 \sum_{j \in \Gamma(i)} v(i, j) \sum_{j \in \Gamma(i)} v(j, i)}{1 + u^2 \sum_{j \in \Gamma(i)} v(i, j) \sum_{j \in \Gamma(i)} v(j, i)} \quad (16)$$

Finally, we have

$$s_l(i) = \frac{l g_l(i) s_l}{\sum_{j \in V} l g_l(j)} \quad (17)$$

By applying the Belief Propagation method, we can estimate the number of cycles in $O(nm^2)$ time. The polynomial time enables the possibility of measuring the properties of cycles in social networks. The effectiveness of the method has been proved³².

Data availability. All relevant data are available at <http://www.github.com/lukefchou/bp>.

References

1. National Research Council. *Network Science* (National Academies Press, USA, 2006).
2. Saito, K., Kimura, M., Ohara, K. & Motoda, H. Super mediator—A new centrality measure of node importance for information diffusion over social network. *Information Sciences* **329**, 985–1000 (2016).
3. Gangal, V., Narwekar, A., Ravindran, B. & Narayanam, R. Trust and distrust across coalitions: Shapley value based centrality measures for signed networks. *In Proc. Thirtieth AAAI Conf. on Artificial Intelligence*, 4210–4211 (2016).
4. Schieber, T. A. *et al.* Quantification of network structural dissimilarities. *Nature Communications* **8**, 13928 (2017).
5. Ghoshal, G. & Barabási, A. L. Ranking stability and super-stable nodes in complex networks. *Nature communications* **2**, 394 (2011).
6. Beauchamp, M. A. An improved index of centrality. *Behavioral Science* **10**(2), 161–163 (1965).
7. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**(4), 581–603 (1966).
8. Freeman, L. C. Centrality in social networks conceptual clarification. *Social Networks* **1**(3), 215–239 (1978).
9. Newman, M. E. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E* **64**(1), 016132 (2001).
10. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.* **101**(11), 3747–3752 (2004).
11. Brandes, U. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* **25**(2), 163–177 (2001).
12. Bonacich, P. & Lloyd, P. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks* **23**(3), 191–201 (2001).
13. Freeman, L. C., Borgatti, S. P. & White, D. R. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks* **13**(2), 141–154 (1991).
14. Du, Y. *et al.* A new closeness centrality measure via effective distance in complex networks. *Chaos: An Interdisciplinary J. Nonlinear Science* **25**(3), 033112 (2015).
15. Gómez, D., Figueira, J. R. & Eusébio, A. Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems. *European J. Operational Research* **226**(2), 354–365 (2013).
16. Schoch, D. & Brandes, U. Re-conceptualizing centrality in social networks. *European J. Applied Mathematics* **27**(6), 971–985 (2016).
17. Martin, T., Zhang, X. & Newman, M. E. Localization and centrality in networks. *Physical Review E* **90**(5-1), 052808 (2014).
18. Estrada, E., Higham, D. J. & Hatano, N. Communicability betweenness in complex networks. *Physica A* **388**(5), 764–774 (2009).
19. Yan, G., Zhou, T., Hu, B., Fu, Z. Q. & Wang, B. H. Efficient routing on complex networks. *Physical Review E* **73**(2), 046108 (2006).
20. Newman, M. E. A measure of betweenness centrality based on random walks. *Social networks* **27**(1), 39–54 (2005).
21. Estrada, E. & Hatano, N. Communicability in complex networks. *Physical Review E* **77**(3), 036111 (2008).
22. REN, X. & LV, L. Review of ranking nodes in complex networks. *Chinese Science Bulletin* **59**(13), 1175–1197 (2014).

23. Gao, J., Barzel, B. & Barabási, A. L. Universal resilience patterns in complex networks. *Nature* **530**(7590), 307 (2016).
24. Dantzig, G. B., Blattner, W. O. & Rao, M. R. Finding a cycle in a graph with minimum cost to time ratio with application to a ship routing problem. *Int. Symp. on Theory of Graphs*. 77–83 (1967).
25. Kamae, T. A systematic method of finding all directed circuits and enumerating all directed paths. *IEEE Trans. Circuit Theory*. **14**(2), 166–171 (1967).
26. Gottlieb, C. G. & Corneil, D. G. Algorithms for finding a fundamental set of cycles for an undirected linear graph. *Commun. ACM*. **10**(12), 780–783 (1967).
27. Welch, J. T. A mechanical analysis of the cyclic structure of undirected linear graphs. *J. ACM*. **13**(2), 205–210 (1966).
28. Tiernan, J. C. An efficient search algorithm to find the elementary circuits of a graph. *Commun. ACM*. **13**(12), 722–726 (1970).
29. Tarjan, R. Enumeration of the elementary circuits of a directed graph. *SIAM J. Comput.* **2**(3), 211–216 (1973).
30. Johnson, D. B. Finding all the elementary circuits of a directed graph. *SIAM J. Comput.* **4**, 77–84 (1975).
31. Mateti, P. & Deo, N. On algorithms for enumerating all circuits of a graph. *SIAM J. Comput.* **5**(5), 90–99 (1976).
32. Marinari, E., Monasson, R. & Semerjian, G. An algorithm for counting circuits: application to real world and random graphs. *Europhys. Lett.* **73**(3), 301–307 (2005).
33. Birmelé, E. *et al.* Optimal listing of cycles and st-paths in undirected graphs. *Proc. the 24th Ann. ACM-SIAM Sym. on Discrete Algorithms*. 1884–1896 (2013).
34. Tulczyjew, W. M. The legendre transformation. *Ann. Inst. Henri Poincaré*. **27**, 101–114 (1977).
35. Katsura, S. & Takizawa, M. Bethe lattice and the Bethe approximation. *Prog. Theor. Phys.* **51**, 82–98 (1974).
36. Yedidia, J. S., Freeman, W. T. & Weiss, Y. Generalized belief propagation. *NIPS*. **13**, 689–695 (2000).
37. Erdos, P. & Renyi, A. On random graphs I. *Publ. Math.* **6**, 290–297 (1959).
38. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature*. **393**(6684), 440–442 (1998).
39. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science*. **286**(5439), 509–512 (1999).
40. Rochat, Y. Closeness Centrality Extended To Unconnected Graphs: The Harmonic Centrality Index. *ASNA*, (2009).
41. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik* **1**(1), 269–271 (1959).
42. Tarjan, R. Depth-first search and linear graph algorithms. *SIAM Journal on Computing* **1**(2), 146–160 (1972).
43. Björklund, A., Husfeldt, T. & Taslamán, N. Shortest cycle through specified elements. *Proc. 23rd annual ACM-SIAM sym. on Discrete Algorithms*, 1747–1753 (2012).
44. Suurballe, J. W. & Tarjan, R. E. A quick method for finding shortest pairs of disjoint paths. *Networks* **14**(2), 325–336 (1984).

Acknowledgements

This work was supported by the Natural Science Foundation of China under grant nos. 71531012 and 71601013, the Beijing Natural Science Foundation under Grant Nos 4172032 and 4174087, the Scientific Research Project of Beijing Educational Committee under grand no. SQKM201710016002, and the Youth Talent Project of BUCEA. We thank Bingyu Song and Xiaoyuan Huang for the language improvement.

Author Contributions

Xiaoping Zhou and Xun Liang designed the study, conducted the experiments and wrote the paper. Jichao Zhao and Shusen Zhang analyzed the data.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30249-4>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018