SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2025, Vol. 68, Iss. 9, 192202:1–192202:13
https://doi.org/10.1007/s11432-024-4260-8

Quantifying the weakness of ties with hierarchy-based link centrality

Jianhong MOU[1], Longyun WANG[1], Kang WEN[1], Bitao DAI[1], Suoyi TAN[1], Fredrik LILJEROS[2], Petter HOLME[3,4] & Xin LU[1*]

[1]*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China*
[2]*Department of Sociology, Stockholm University, Stockholm 10691, Sweden*
[3]*Department of Computer Science, Aalto University, Espoo 02150, Finland*
[4]*Center for Computational Social Science, Kobe University, Kobe 657-8501, Japan*

Received 28 August 2024/Revised 18 November 2024/Accepted 11 December 2024/Published online 2 April 2025

Abstract Quantifying the significance of ties in preserving network connectivity is crucial for identifying weak ties, which often serve as bridges between communities, and for detecting community structures. However, accurately characterizing network connectivity and formalizing the relationship between weak ties and communities remain challenging. In this study, we introduce hierarchy-based link centrality (HLC), a novel metric based on the dissimilarity between the original network and its contracted version, where the terminal nodes of links merge and connect to all their neighbors. This dissimilarity is quantified by variations in the network hierarchy, specifically the nodal distance distributions. In addition to the experiments on weak tie identification and link-based network disintegration, we develop a link-based community detection (LCD) approach that focuses on optimal link ranking to elucidate community structures. Experiments across various networks demonstrate that HLC excels in identifying weak ties, achieving a 2.9% higher accuracy than the second-best metric. It also outperforms others in detecting critical link combinations for network disintegration, reducing the average size of the giant connected component by 7.2% compared to the suboptimal counterpart. Furthermore, HLC enhances community detection, achieving optimal partitioning with an average 5.7% improvement in modularity over five other indices. These results highlight the effectiveness of HLC in quantifying weak ties and suggest broad applications for this innovative approach in network analysis.

Keywords weak ties, network hierarchy, hierarchy-based link centrality, link-based community detection, network dissimilarity

Citation Mou J H, Wang L Y, Wen K, et al. Quantifying the weakness of ties with hierarchy-based link centrality. Sci China Inf Sci, 2025, 68(9): 192202, https://doi.org/10.1007/s11432-024-4260-8

1 Introduction

Link importance has been a central focus in network analysis for decades [1,2], encompassing tasks such as identifying backbone structures [3], containing epidemics [4], and dismantling networks [5,6]. Recent studies have highlighted the impact of various link types on networks, particularly emphasizing the role of weak ties [7,8]. Initially, weak ties were seen as channels for disseminating valuable information across groups with longer paths, considering factors like time, intimacy, and reciprocal services between node pairs. More recently, the definition of weak ties has expanded to include inter-community links, often viewed as bridges facilitating the flow of novel information between communities [9–12]. Despite numerous community detection algorithms have been proposed to identify weak ties topologically, inter-community links can sometimes transform into intra-community links in networks with fuzzy or overlapping partitions, where community structures are not well-defined [13–17]. In such cases, quantifying weak ties becomes more compelling. Therefore, accurately assessing link centrality, particularly for links that bridge communities, is crucial.

Numerous metrics have been developed to assess the effect of weak ties in facilitating communication among diverse communities [18–24], such as topology overlap, bridgeness, diffusion importance, and betweenness. They are proposed to quantify weak ties in maintaining local or global connectivity from

* Corresponding author (email: xin.lu.lab@outlook.com)

the perspective of neighborhoods, connecting cliques, information diffusion, and shortest paths across links. In addition, evaluating the significance of links in maintaining global connectivity also involves analyzing the topological dissimilarity between the original network and its altered version after link shrinking, achieved by merging node pairs into a single node connected to their neighbors. Various approaches [25–27] have been introduced to quantify the dissimilarity, with network node dispersion [28], or nodal distance distribution, proving effective in identifying subtle topological shifts. Network node dispersion provides a detailed and compact representation of network topology, inspiring research in diffusion capacity [29], link prediction [30], and community vulnerability [31]. However, the dissimilarity measurement falls short in detailing the effect of links on maintaining high-order neighborhoods among nodes, specifically the hierarchical structure, which is crucial in quantifying weak ties.

To evaluate the effectiveness of topology-based link centrality in identifying weak ties, inter-community links connecting Louvain-based communities are often considered ground-truth weak ties due to Louvain's ability to maximize modularity [10]. Link-based network dismantling [6, 32–34] is further introduced to examine the influence of weak ties on global connectivity, where the sequence of link removal significantly affects network disintegrations. Classification accuracy and the average size of the giant connected component indicate the importance of link centrality in capturing weak ties and maintaining global connectivity. Most importantly, some community detection algorithms [35–38] have been designed based on the relationship between link centrality and community structure, with the Girvan-Newman (GN) method [39] being the most renowned. Its success in detecting communities by recalculating edge betweenness confirms the efficiency of betweenness in separating node pairs and the effectiveness of edge betweenness in quantifying weak ties. However, the GN approach tends to prioritize the sequence of links in separating nodes, which compromises the direct evidence of edge betweenness in capturing the community structure. In addition, the priority of weak ties in separating communities is not sufficiently acknowledged in community detection. For instance, in networks with significant community structures, the distribution of link centrality for weak ties usually shows a substantial separation from that of intra-community links (see Figure A1). An effective threshold helps identify the majority of weak ties [10], facilitating the capture of potential communities and minimizing bias caused by the initial partition. Therefore, a community detection framework based on link ranking is necessary to explore the information about community structure contained in link centrality.

Related studies highlight several unresolved issues: (1) how to quantify the weakness of ties to provide a comprehensive understanding of the role of links in bridging communities, including overlapping or ambiguous ones; and (2) how weak ties can facilitate community detection, or whether communities can be detected solely based on the sequence of links ranked by their weakness. In this study, we introduce a network hierarchy based on high-order neighborhoods among nodes, propose a novel link centrality from the perspective of network dissimilarity, and explore a new community detection method that focuses on identifying weak ties. Specifically, (1) we investigate the significance of network hierarchy in quantifying topological dissimilarity and introduce a hierarchy-based link centrality (HLC) for quantifying the weakness of ties; (2) we propose a link-based community detection (LCD) approach that shifts community detection to link ranking based on their significance in capturing community structures; (3) we utilize hierarchy-based nodal redundancy, defined as the excess of links compared to those necessary to maintain the neighborhood, to refine the community affiliation of nodes. Our study reveals that HLC effectively captures the critical links in bridging communities, demonstrating superior performance compared to conventional topology-based link centralities in terms of identification accuracy, link-based network disintegration, and community detection.

2 Methodology

2.1 Network hierarchy

Different from most conventional definitions, in this study, the nodal hierarchical structure in networks encapsulates the extended neighborhoods up to a specified maximal distance, delineated using a breadth-first search (BFS) method. This framework is essential for articulating network topology from a mesoscopic perspective, abstracting away the details of individual node connections while preserving the overarching hierarchical relationships between nodes. Typically, in many networks, the number of nodes within the hierarchy initially expands, reaches a peak, and then tapers off, resembling a "spindle" shape [40]. In this

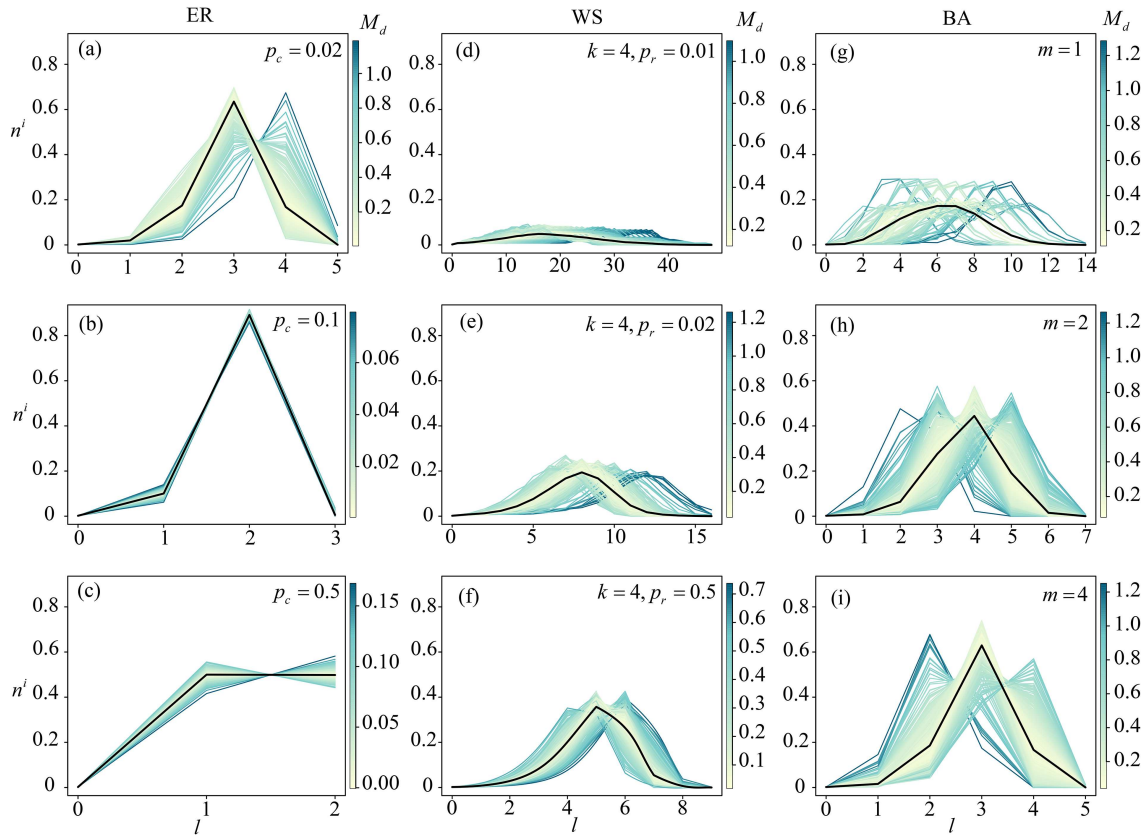

Figure 1 (Color online) Line illustrations of network hierarchies for different synthetic networks with varying parameters. There are obvious graphic distinctions among networks with different generating mechanisms. Each network contains 500 nodes, with the black line representing the average nodal spindle, i.e., $S_p = (\sum_{i=0}^{N} n_0^i/N^2\ , \sum_{i=0}^{N} n_1^i/N^2\ , \ldots, \sum_{i=0}^{N} n_L^i/N^2\ )$. The colors of the lines indicate the Manhattan distance between each nodal spindle and the average one, calculated as $M_d^i = \sum_{j=1}^{L_i} |S_p^i(j) - S_p(j)|$. (a)–(c), (d)–(f), and (g)–(i) respectively depict the spindle matrices of ER, WS, and BA networks with different linking probability p_c, rewiring probability p_r, and number of adding links m.

structure, nodes in layer l represent neighbors at a distance l from the root node. The nodal spindle, $\boldsymbol{S}_i^p$, is described by the fraction of nodes at various distances from a reference node i, formulated as follows:

$$\boldsymbol{S}_p^i = \frac{1}{N}(n_0^i, n_1^i, \ldots, n_{L_i}^i), \tag{1}$$

where $n_{L_i}^i$ denotes the number of nodes in layer L_i, and the root node is viewed as layer 0. The network hierarchy aggregates the hierarchical structure over all nodes and is defined as $\boldsymbol{S}$:

$$\boldsymbol{S} = (\boldsymbol{S}_p^1, \boldsymbol{S}_p^2, \ldots, \boldsymbol{S}_p^N)^{\mathrm{T}} = \frac{1}{N}\begin{bmatrix} n_0^1, & n_1^1, & \ldots & n_{L_1}^1, & 0, & \ldots 0 \\ n_0^2, & n_1^2, & \ldots & n_{L_2-1}^2, & n_{L_2}^2, & \ldots 0 \\ & & \ldots & & & \ldots \\ n_0^i, & n_1^i, & \ldots & n_{L_N-1}^i, & n_{L_N}^i, & \ldots 0 \end{bmatrix}. \tag{2}$$

The network hierarchy explicitly reveals the degree distribution, diameter, and average shortest path length of networks [40]. In addition, synthetic networks generated by different mechanisms exhibit distinct hierarchical patterns, while different configurations of the same network model show clear similarities from the perspective of line and graphic illustrations of the network hierarchy (see Figure 1 and Figure A2).

2.2 Hierarchy-based link centrality

HLC evaluates the dissimilarity between the original and altered networks using the Jensen-Shannon (JS) divergence of their network hierarchies. The altered network is created by merging node pairs into a single node connected to their neighbors (see Figure A3 for an example of the HLC calculation). In the

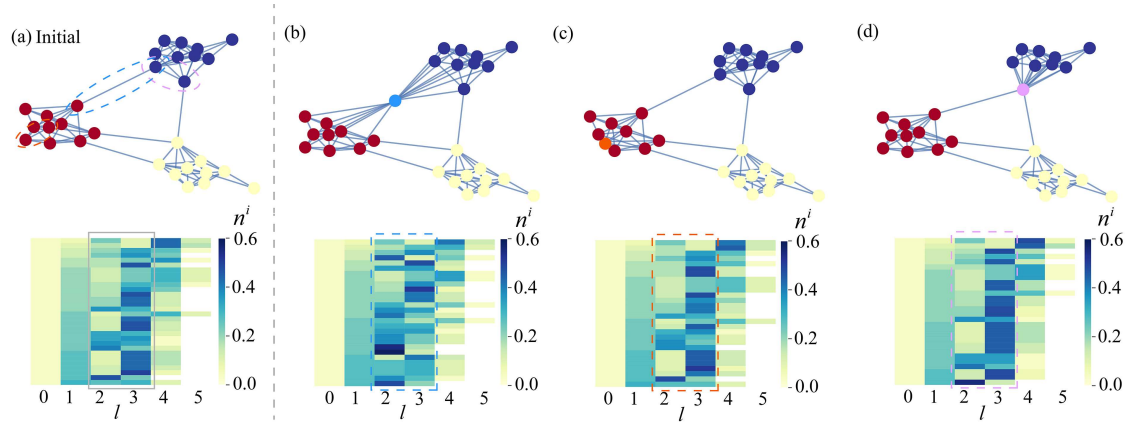

Figure 2 (Color online) Effect of link shrinking on the hierarchical structure of a toy network. The shrinking of links in different positions results in varying degrees of network dissimilarities compared to the original topology. This process typically alters the distribution of layers 2 and 3, as most nodes in this toy network are in the second or third neighborhood of others. (a) The initial network topology and its corresponding network hierarchy; (b)–(d) the new network configurations after merging nodes connected by weak ties (b) and intra-community links (c) and (d). The newly formed nodes are indicated in blue, orange, and purple, respectively. The network hierarchy matrices are arranged in a specific order: the nodal spindle with the largest value of the first layer is listed as the first component of the matrix, and nodal spindles with identical values in the first layer are arranged according to the order of the second layer.

context of an undirected and unweighted network G with N nodes and E links, for a specific link e_{ij}, the network hierarchy of the unchanged $N-2$ nodes in both the initial and new networks is expressed as

$$\boldsymbol{S} = (\boldsymbol{S}_p^1, \ldots, \boldsymbol{S}_p^{i-1}, \boldsymbol{S}_p^{i+1}, \ldots, \boldsymbol{S}_p^{j-1}, \boldsymbol{S}_p^{j+1}, \ldots, \boldsymbol{S}_p^N)^{\mathrm{T}}, \tag{3}$$

$$\tilde{\boldsymbol{S}} = (\tilde{\boldsymbol{S}}_p^1, \ldots, \tilde{\boldsymbol{S}}_p^{i-1}, \tilde{\boldsymbol{S}}_p^{i+1}, \ldots, \tilde{\boldsymbol{S}}_p^{j-1}, \tilde{\boldsymbol{S}}_p^{j+1}, \ldots, \tilde{S}_p^N)^{\mathrm{T}}. \tag{4}$$

The HLC of link e_{ij} is calculated as the sum of the JS divergence values for the $N-2$ nodes, with each node represented by its nodal spindle divergence:

$$H_{\mathrm{lc}}(e_{ij}) = \sum_{k \in N \setminus i,j} \mathrm{JS}(\boldsymbol{S}_p^k, \tilde{\boldsymbol{S}}_p^k)/(\mathrm{CN}(i,j)+1)^2, \tag{5}$$

where $\mathrm{CN}(i,j)$ represents the number of common neighbors of nodes i and j, $\mathrm{JS}(\boldsymbol{S}_p^k, \tilde{\boldsymbol{S}}_p^k)$ is the JS divergence of nodal spindle of node k:

$$\mathrm{JS}(\boldsymbol{S}_p^k, \tilde{\boldsymbol{S}}_p^k) = \frac{1}{2}\mathrm{KL}\left(\boldsymbol{S}_p^k(x) \middle\| \frac{\boldsymbol{S}_p^k(x) + \tilde{\boldsymbol{S}}_p^k(x)}{2}\right) + \frac{1}{2}\mathrm{KL}\left(\tilde{\boldsymbol{S}}_p^k(x) \middle\| \frac{\boldsymbol{S}_p^k(x) + \tilde{\boldsymbol{S}}_p^k(x)}{2}\right), \tag{6}$$

and $\mathrm{KL}(P,Q)$ is the KL divergence of the distributions of P and Q:

$$\mathrm{KL}(P,Q) = \sum P(x) \log \frac{P(x)}{Q(x)}. \tag{7}$$

In the toy network shown in Figure 2, merging nodes connected by weak ties significantly alters the network's topology compared to merging nodes within communities, as evidenced by changes in the network hierarchy. Noticeably, intra-community links connecting nodes directly accessible by other communities, i.e., the peripheral nodes of communities, also play a crucial role in maintaining global connectivity (see Figure 2(d)).

2.3 Link-based community detection model

The weakness of ties is considered a characteristic of links, quantifying the extent or probability of connecting communities. Ties with significant weakness are referred to as weak ties. Typically, nodes connected by weak ties are located at the peripheries of communities, while nodes indirectly connected to

these peripheries through several intra-community links are considered inner nodes of the corresponding community. This observation leads to the development of the link-based community detection (LCD) algorithm, which utilizes the information embedded in link centrality to uncover community structures. The general form of our algorithm is outlined below, with the corresponding pseudocode presented in Table A1.

Periphery identification. The LCD algorithm categorizes links into intra-community links L_{intra} and weak ties L_{inter} based on a threshold k of the link centrality metric M, as demonstrated in Figures 3(a) and (b). This categorization helps identify the peripheries of communities through nodes linked by weak ties. The LCD algorithm begins the detection process by initially focusing on several candidate communities defined by peripheral nodes. Subsequently, an agglomerative phase follows, merging peripheral nodes once they are connected by intra-community links, to reduce the number of communities before assigning inner nodes. For example, Figure 3(c) shows that 3 weak ties in a simplified network delineate 4 boundary nodes. However, nodes v_G and v_H are connected by an intra-community link, indicating their membership in the same cluster.

Inner node allocation. Nodes often connect to multiple communities through intra-community links, which makes it difficult to allocate these nodes and blurs community detection. Assuming that communities are composed of inner nodes with the most intra-community links or the lowest average link centrality helps address this dilemma. If node i has more intra-community links within community C_1 than C_2, it is allocated to C_1. In cases where the intra-community links to C_1 and C_2 are equivalent, node i is assigned to the community with the lower average link centrality, as illustrated in Figure 3(d). Notably, the LCD algorithm avoids the recalibration commonly required in divisive community detection methods, maximizing the relevance of weak ties in identifying community peripheries. The focus instead shifts to the rank of links and the threshold.

Partition fine-tuning. Considering the changing affiliation of nodes to communities during the step-by-step allocation process, we evaluate the partitioning of each node based on its redundancy to each community after the agglomerative process (see Figure 3(e)). Following this, the optimal partition is considered to consist of several communities containing nodes with the highest redundancy (see Figure 3(f)). The redundancy of node i in community C, defined as r_i^c, represents the excess connections compared to those necessary to maintain the hierarchical relationship between the first-order and second-order neighbors of i within community C, expressed as

$$r_i^c = \frac{e_i^c}{n_i^c} k_i^c, \tag{8}$$

where k_i^c and n_i^c respectively represent the number of first-order and second-order neighbors of node i within community C, and e_i^c denotes the number of links connecting them. Node redundancy is primarily determined by the connections to community C, represented as k_i^c. In addition, we consider the characteristics of the community, specifically the second-order node redundancy quantified by $\frac{e_i^c}{n_i^c}$, when multiple communities exhibit the same k_i^c. Furthermore, we can quantify high-order node redundancy by considering the excess connections between higher layers necessary to maintain the hierarchy structure. Most importantly, node i is allocated to a community with the maximum r_i^c.

2.4 Competing link importance metrics

This section introduces several link centralities that evaluate the significance of links in separating communities.

Edge betweenness [41] (BET) quantifies an edge's significance based on the count of shortest paths traversing it:

$$B_{\text{et}}(e) = \sum_{s \neq t \in V} \frac{n_{st}^e}{n_{st}}, \tag{9}$$

where n_{st} represents the number of shortest paths between nodes v_s and v_t, while n_{st}^e denotes the number of those paths that traverse link e. Edges with high betweenness are crucial for maintaining communication and facilitating the spread of information across the network.

Bridgeness [42] (BRI) highlights an edge's significance in enhancing the connectivity of a network:

$$B_{\text{ri}}(e) = \frac{\sqrt{S_x S_y}}{S_e}, \tag{10}$$

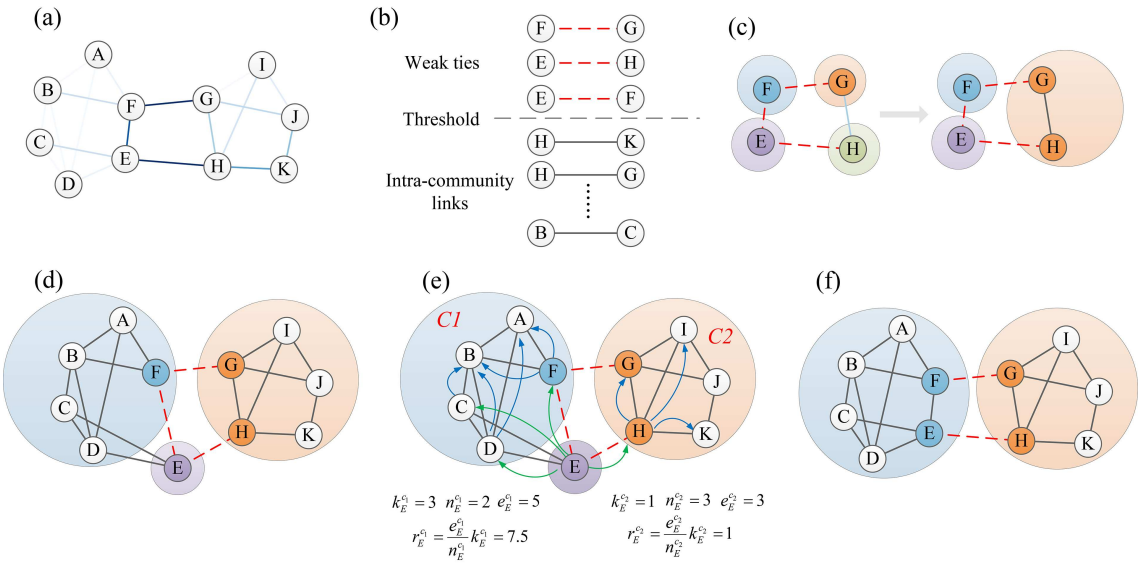

Figure 3 (Color online) Link-based community detection framework. The LCD framework views the network partition as a process of link ranking and supports the evaluation of link centrality in capturing community structure. (a) A toy network comprised of 19 links, with different link centralities represented by various colors; (b) weak ties (red dashed line) and intra-community links (black lines) separated by the threshold k; (c) the initial clusters separated by weak ties; (d) the allocation of inner nodes connected to peripheries via the intra-community links; (e) partition fine-tuning based on nodal redundancy, using node E as an example (green edges represent connections between node v_E and its first-order neighbors, while blue edges connect the first-order and second-order neighbors); (f) the final communities of the toy network.

where v_x and v_y are the two nodes of the edge e, S_x and S_y respectively represent the size of the maximal cliques containing nodes v_x and v_y, and S_E is the size of the maximal clique containing edge e. This metric prioritizes an edge's role in integrating disparate network segments over its involvement in local cliques which are the most basic forms of local connectivity within a network.

Degree product [43] (DP) quantifies the connectivity strength between pairs of nodes, defined as

$$D_{\mathrm{p}}(e) = d_x d_y, \tag{11}$$

where nodes v_x and v_y are connected by link e, with degree d_x and d_y, respectively. This metric performs robustly in assortative networks, where high-degree nodes are likely to connect with each other. Conversely, in disassortative networks, where high-degree nodes tend to connect with low-degree nodes, the quantification of edge significance is unreliable.

Topological overlap [19] (TO) describes the probability that the neighbors of two endpoints of an edge are the same:

$$T_{\mathrm{o}}(e) = \frac{c_{ij}}{(d_i - 1) + (d_j - 1) - c_{ij}}, \tag{12}$$

where c_{ij} represents the number of common neighbors of nodes v_i and v_j, while the denominator represents the number of union neighbors. Edges with high topological overlap are typically located within communities.

Diffusion important [22] (DI) quantifies the importance of links through disease spread to other parts of networks from endpoints in both directions:

$$D_{\mathrm{i}}(e) = \frac{n_{i\to j} + n_{j\to i}}{2}, \tag{13}$$

where $n_{i\to j}$ is the number of links from node v_j connecting outside the nearest neighbors of node v_i. The index is inherently influenced by the degree of the node to some extent: an edge comprising one high-degree node and one low-degree node may exhibit an inflated value of edge significance relative to its actual impact, particularly when the edge is situated on the periphery of the network.

The counterpart link centralities primarily focus on either the local structure of edges or the global connectivity through shortest paths. In contrast, HLC integrates both aspects by utilizing JS divergence

for $N-2$ nodes, along with the local information derived from the number of common neighbors. In addition, edge betweenness has been shown to correlate with the network hierarchy [41,44]:

$$\sum_{e\in E} B_{\mathrm{et}}(e) = \frac{N}{N-2}\left(\frac{N}{N-1}\sum_{j=0}^{L_{\max}} j\boldsymbol{S}_p(j) - 1\right), \tag{14}$$

where $\boldsymbol{S}_p = \frac{1}{N}\sum \boldsymbol{S}_p^i$, $\sum \boldsymbol{S}_p^i$ is the summation operator for vectors of various lengths and $L_{\max}$ is the total number of layers. Compared to the integrated information of shortest paths captured by edge betweenness, HLC offers a more nuanced representation using JS divergence, thereby providing a finer resolution in capturing the influence of edges in maintaining global connectivity.

3 Experimental setting

3.1 Data description

Lancichinetti-Fortunato-Radicchi (LFR) networks. The LFR benchmark [45] excels at creating networks with predetermined community structures, degree distributions, and mixing patterns. By forming communities of diverse sizes and varying levels of inter-community linkage, LFR networks effectively simulate the complex connectivity observed in real-world networks. This capability makes them ideal for evaluating community detection algorithms across varied network architectures. In this investigation, we synthesize 3 LFR networks with distinct community structures, characterized by $\delta = 0.05, 0.15, 0.25$ (see Table A2).

Empirical networks. We introduce 6 empirical networks, excluding self-loops, directional aspects, and edge weights. Our analysis concentrates solely on the giant connected components of these networks. The topological characteristics of the empirical networks are summarized in Table A3.

3.2 Weak ties identification

To evaluate the efficacy of HLC in identifying weak ties, we analyze the identification accuracy within the top l links of the ordered sequence, increasing by increments of 10 (i.e., $\mathrm{d}l = 10$). The area under the accuracy distribution curve is used to assess the performance of HLC:

$$\sigma = \int G(l)/L\mathrm{d}l, \tag{15}$$

where $G(l)$ denotes the number of weak ties within the top l links, and L is the total number of weak ties.

3.3 Link-based network disintegration

In addition to identifying weak ties, we assess the performance of HLC in identifying critical links for network connectivity through experiments on link-based network disintegration. In a highly clustered network, the simultaneous removal of weak ties will inevitably fragment the network into several isolated communities. In this section, we further concentrate on combinations of links that most significantly disrupt the network. We use the proportion of nodes in the giant connected component (g) as a proxy measure of network connectivity. The average size of the giant connected component, r, quantifies the extent of link-based network disintegration:

$$r = \sum_{l=0}^{L} g(l)/N, \tag{16}$$

where $g(l)$ represents the size of the giant connected component after removing l links. During network disintegration, links are removed sequentially in descending order of link centrality, with the maximum ratio of disintegrated nodes set to $\sqrt{N}$.

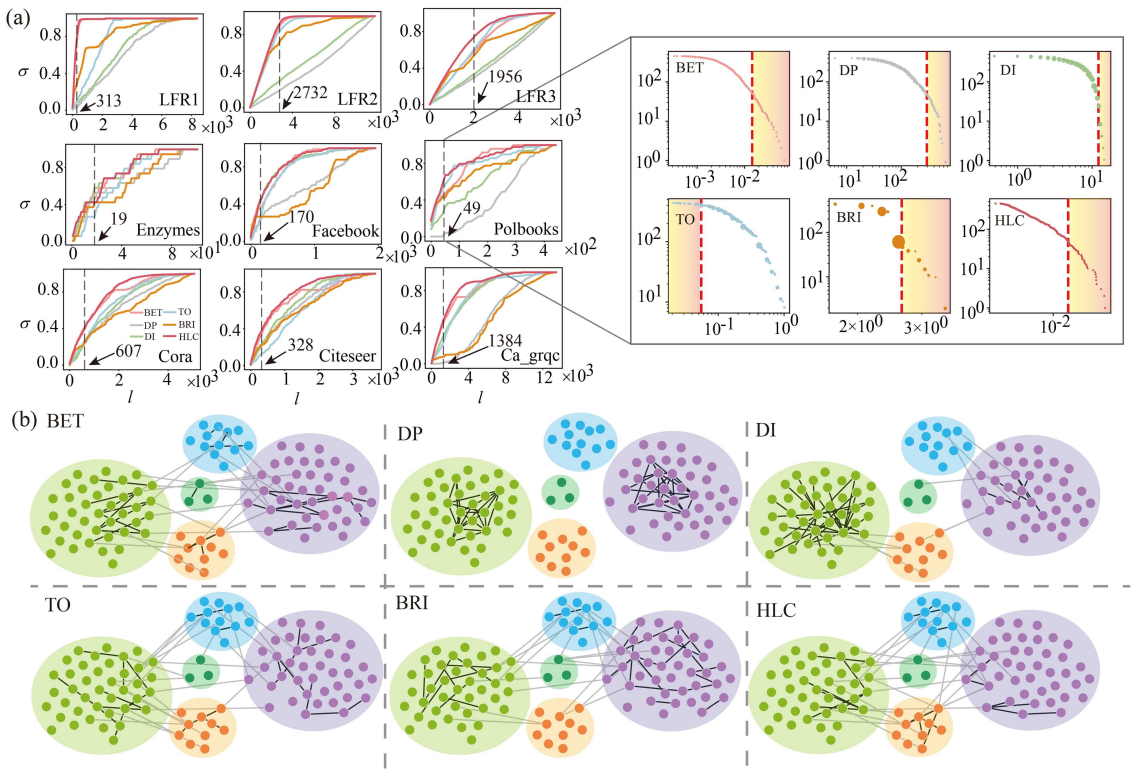

Figure 4 (Color online) Weak tie identification based on link centrality. HLC outperforms its counterparts across all networks. (a) The performance of various link centralities in capturing weak ties across 3 synthetic and 6 empirical networks. l and σ denote the number of links and the identification accuracy, respectively. The vertical dashed lines indicate the number of weak ties detected by the Louvain method. The inset shows the cumulative distribution of link centralities, with the size of each point representing the number of links with identical centrality. The shaded areas indicate l_m links, ordered by their corresponding link centralities, predicted as weak ties. (b) The spatial distribution of weak ties identified by link centralities in the Polbooks network, where nodes of the same color are clustered into communities. Black and gray links represent intra-community links and weak ties, respectively.

3.4 Link-based community detection

The performance of LCD is highly dependent on the threshold k. A significant gap in the distribution of link centrality assists in identifying the optimal threshold for effective community detection. However, in many networks, the community structure is not pronounced, leading to an indistinct separation in link centrality distribution. Therefore, to determine the most effective partition, we perform an exhaustive search across all potential thresholds, aiming to maximize modularity.

4 Results

4.1 Weak ties identification based on the link importance metrics

HLC outperforms its counterparts in identification accuracy across all networks, with average improvements of approximately 3.2%, 50%, 23.6%, 10.7%, and 26.8% compared to BET, DP, DI, TO, and BRI, respectively (see Table A4). In addition, HLC captures more weak ties in most instances, particularly when $l = l_w$, where l_w represents the number of weak ties detected by the Louvain method [46] (indicated by the vertical dashed lines in Figure 4(a)). For example, when examining the effectiveness of link centrality in identifying weak ties within the Polbooks network, HLC demonstrates a remarkable power law distribution and a clear separation, facilitating the detection of weak ties, as shown in Figure 4(b) when $l = l_w = 49$. Although topological overlap centrality also reveals a similar heterogeneous distribution, HLC is more monotonous, offering higher resolution in distinguishing links.

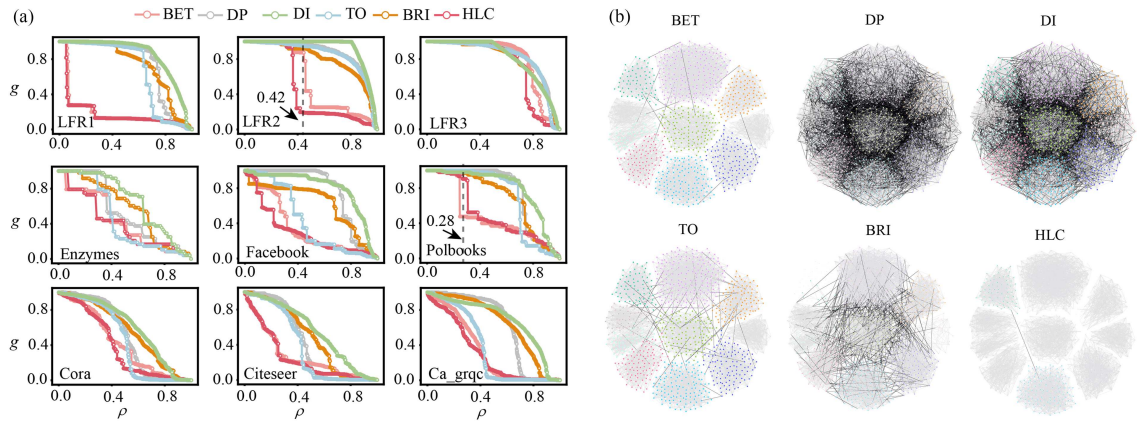

Figure 5 (Color online) Link-based network disintegration using link centralities. Network disintegration based on HLC typically results in a rapid collapse of the giant connected component. (a) The process of link-based network dismantling; (b) the disintegration results for the LFR2 network with $\rho = 0.42$. Nodes of the same color are clustered, and the largest subset of interconnected nodes represents the giant connected component. Black and gray edges denote weak ties and intra-community links, respectively.

4.2 Link-based network disintegration

As shown in Table A5, HLC surpasses nearly all counterparts across all networks, with the exception of betweenness on the Polbooks network. Specifically, the average size of the giant connected component after HLC-based network dismantling decreases by approximately 7.3%, 40.8%, 47%, 29.3%, and 41% compared to BET, DP, DI, TO, and BRI, respectively.

The dramatic collapse of g in synthetic networks is typically caused by the combined removal of weak ties (see Figure 5(a)). Furthermore, HLC is slightly more effective than betweenness in ranking links in terms of disrupting network connectivity; in other words, HLC leads to an earlier network collapse in all cases except for the Polbooks network. In the case of LFR2 networks with a significant community structure (see Figure 5(b)), HLC significantly aids in network disintegration after removing 42% of the links, leaving only two communities connected, which is drastically fewer than those based on other indices. Conversely, the combination of weak ties and intra-community links may facilitate disintegration in networks with less pronounced community structures, as link-based network dismantling is a combinatorial optimization problem. For example, in the Polbooks network (as illustrated in Figure A4), the circled intra-community link plays a crucial role in maintaining global connectivity after removing 28% of the links in the descending order of HLC. This particular link is identified as a weak tie and is therefore removed in the betweenness approach, resulting in significant disruption to the network.

4.3 Community detection

Since identifying weak ties facilitates the separation of peripheral and inner nodes within communities, we examine the performance of HLC in community detection using the LCD model in this section. Results show that HLC produces superior partitions compared to other centrality measures in terms of modularity across both LFR and empirical networks. Specifically, the LCD model based on HLC achieves the optimal partition, with an average modularity improvement of approximately 1.3%, 7.5%, 5.7%, 11.5%, and 2.2% compared to BET, DP, DI, TO, and BRI, respectively (see Table 1 and Figure 6). Furthermore, the relationship between the threshold and modularity in the LCD model for the LFR1 network, as illustrated in Figure A5, indicates that both very small and very large thresholds reduce modularity because they respectively result in either an excessive number of communities or too few communities. The stability of thresholds in partitioning by the LCD model based on HLC and BET indicates their effectiveness in separating weak ties from intra-community links. Additionally, the longer duration of HLC in seeking the maximum modularity compared to BET indicates HLC's higher resolution in identifying weak ties, consistent with our findings on identification accuracy.

In addition to comparing the community structure information captured by various link centralities, we also evaluate the efficiency of the LCD method in detecting communities against the Greedy, LPA, Louvain, Infomap, GN, and Walktrap methods (see Table A6). The LCD method produces exact partitions on LFR networks with significant community structures and generates suboptimal partitions, inferior to

Table 1 Modularity of LCD based on various centralities on networks. The best results are in bold. The second best results are underlined.

Network	BET	DP	DI	TO	BRI	HLC
LFR1	**0.863**	0.845	0.849	**0.863**	**0.863**	**0.863**
LFR2	**0.644**	0.627	0.639	0.412	**0.644**	**0.644**
LFR3	0.512	0.387	0.409	0.428	0.501	**0.553**
Enzymes	0.591	0.583	0.596	0.587	0.584	**0.604**
Facebook	0.693	0.690	0.673	0.681	0.680	**0.694**
Polbooks	**0.526**	**0.526**	**0.526**	**0.526**	0.522	**0.526**
Cora	0.744	0.684	0.714	0.730	0.739	**0.745**
Citeseer	0.781	0.771	0.756	0.789	0.790	**0.794**
Ca_grqc	**0.791**	0.763	0.790	0.716	0.773	0.790

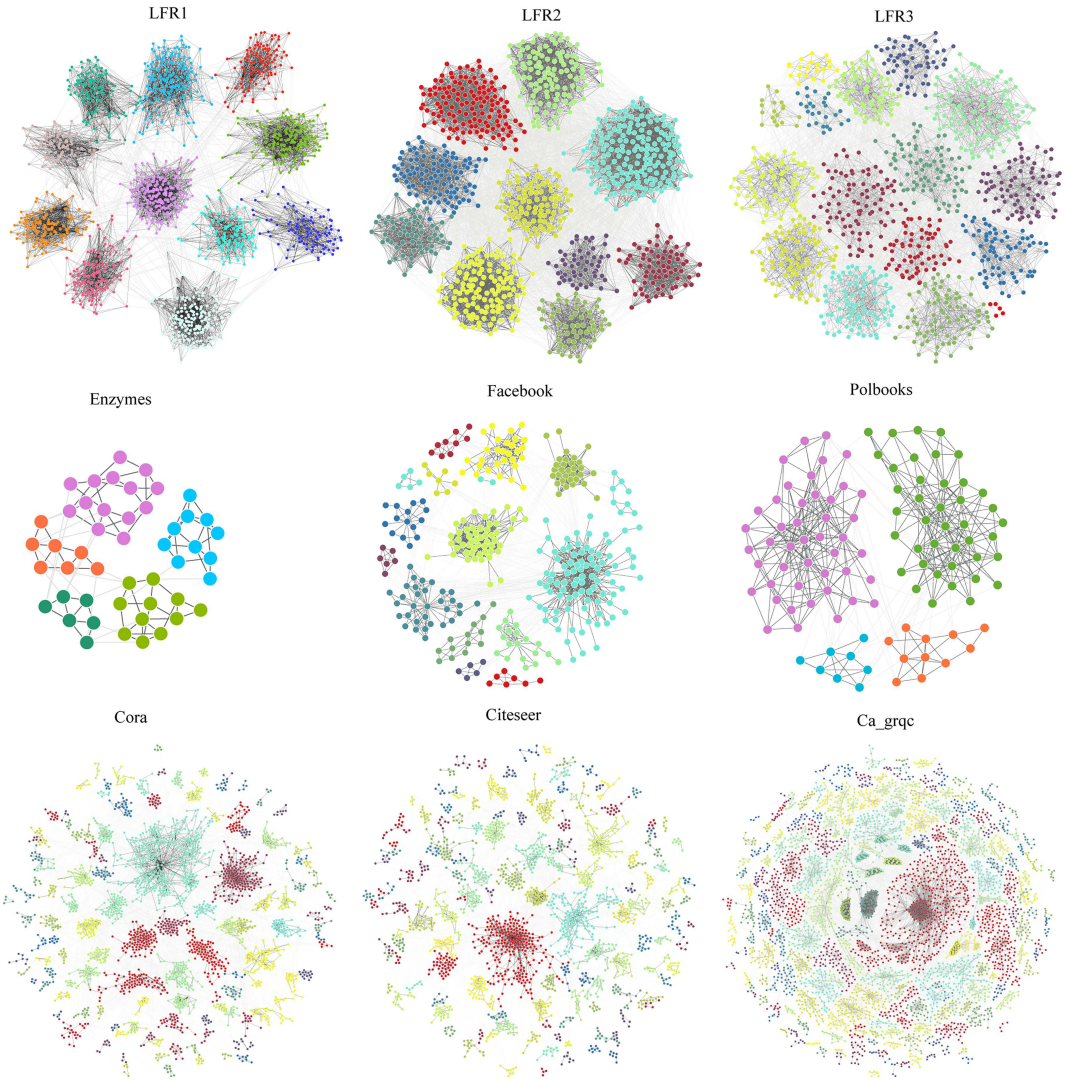

Figure 6 (Color online) Communities detected by LCD based on HLC. Nodes of the same color are clustered within the identical community. Black and gray edges respectively represent the intra-community and weak ties.

the Louvain method, on 3 empirical networks. The indistinguishable nature of weak ties connecting small communities from the intra-community links within larger communities in the sparsely clustered Citeseer, Cora, and Ca_grqc networks may account for the inferiority of the LCD compared to the Louvain method. The superior performance of the GN method on the Citeseer, Cora, and Ca_grqc networks compared to LCD is attributed to its recalculation of link betweenness, albeit a highly time-consuming process. Furthermore, the performance of the LCD method in community detection, measured by normalized mutual

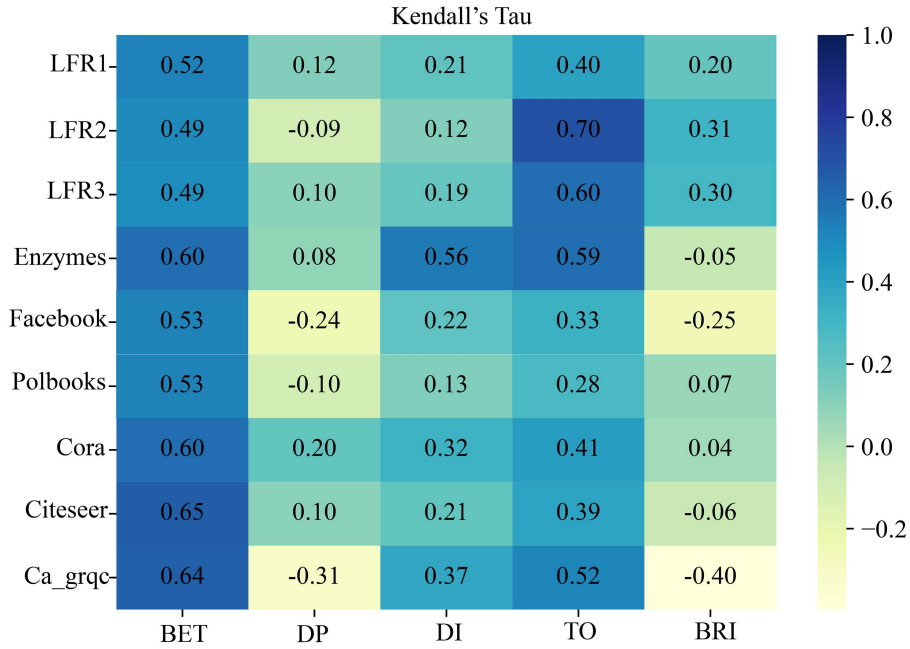

Figure 7 (Color online) Kendall's Tau correlation matrix between HLC and other indices over nine networks.

information (NMI), demonstrates its effectiveness compared to other methods (see Table A7). Similarly, the comparison of HLC with other link centralities in terms of NMI further supports the efficacy of both LCD and HLC (see Table A8).

4.4 Correlation analysis

To explore the relationship between HLC and other centralities, we measure the correlation between pairs of indices using Kendall's Tau (τ). The correlations between HLC and the other five indices across 9 networks are displayed in Figure 7. It is evident that the correlations among HLC, BET, and TO are higher than other indices, indicating that HLC incorporates characteristics of both betweenness and topological overlap. This is attributed to HLC's ability to describe network dissimilarity and its link-based shrinking procedure, which facilitates the quantification of global connectivity through local topological variations. The low or negative correlations with DP, DI, and BRI suggest that weak ties have a strong correlation with global connectivity.

4.5 Time complexity analysis

In networks composed of n nodes and m edges, the HLC metric quantifies the JS divergence between hierarchical metrics of $n-2$ nodes before and after link shrinking. The computational effort required to map the hierarchical structure of a node, specifically through the BFS process, is $O(n+m)$. Consequently, the overall computational cost of calculating HLC is $O(n(n+m))$, as it involves the comprehensive analysis of hierarchical metrics across the network. Link centralities that concentrate on local structures incur a time cost that is either linear or quadratic to the average node degree, often resulting in a limited representation of global connectivity. In contrast, both edge betweenness and HLC provide a comprehensive description of global connectivity, albeit at a higher computational cost of $O(n(n+m))$. The time required to calculate edge betweenness can be reduced to $O(nm)$ through the application of the Brandes algorithm.

The LCD algorithm, following a specified link sequence, operates through two primary stages: identifying community peripheries and assigning inner nodes to communities. It iterates over the link sequence to determine the optimal threshold. During the community detection phase, after distinguishing between inter-community and intra-community links, the algorithm iterates through all links to merge dispersed communities, resulting in a time complexity of $O(m)$. Subsequently, examining each node for redundancy adds a complexity of $O(n)$. In the optimization phase, the centrality of each link is tested as a potential threshold to maximize modularity, introducing a complexity of $O(m)$. Since community merging and

node examination are performed in line for each threshold, the overall time complexity of the LCD algorithm is approximately $O(m(m+n))$. Comparatively, the time complexity of the LPA [47] and Louvain methods is linear to m, thereby outperforming others across all networks. LCD requires less time than another link-based community detection method, namely GN. In sparse networks, i.e., $m \sim n$, LCD performs better than Walktrap [48] and is slightly inferior to the Greedy [49] and Infomap [50] methods (see Table A9).

5 Conclusion

To address the impact of link centrality related to community structures on overall network topology, we introduce the concept of a "nodal spindle" as a tool for mapping the network's hierarchical layout and providing a detailed description of the network hierarchy. In this context, we propose the HLC metric to quantify the weakness of ties, recognizing the greater disparity introduced by shrinking weak ties compared to intra-community ones. Building on the intriguing connection between the significance of community-related links and community architecture, we unveil an advanced community detection framework anchored in link rank analysis, called LCD. This method revolutionizes community detection by seeking the optimal link ranks for assessing the connectivity strength among distinct communities. LCD initially separates community boundaries connected by weak ties and subsequently allocates inner nodes to these peripheries via intra-community connections. To rectify potential mispartitions caused by inner node allocation, where newly added nodes may alter the affiliation of existing nodes, we introduce a refinement strategy that leverages node redundancy within each community. In addition to LCD, experiments on weak tie identification and link-based network disintegration are conducted to demonstrate the effectiveness of HLC over traditional metrics.

Although HLC outperforms five other indices in quantifying the weakness of ties, its time complexity may limit its widespread application to large-scale networks. The variance in weak tie identification by the HLC index between synthetic and empirical networks underscores the need for further exploration into the significance of community structures from the perspective of link centrality. Understanding the influence of overlapping communities on the properties of weak ties could enhance our comprehension of the relationship between link centrality and complex community structures within networks. Furthermore, the detailed information regarding the high-order neighborhood that underpins HLC supports its applications in various fields, including information propagation within social networks, resilience assessments of transportation routes, and evaluations of interactions among neurons in brain networks. Additionally, the hierarchical structure of networks necessitates a more comprehensive analysis and broader application, given its success in predicting diffusion patterns, quantifying network dissimilarity, and more.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 72025405, 72421002, 92467302, 72088101, 72301285, 72474223), Natural Science Foundation of Hunan Province (Grant Nos. 2023JJ40685, 2024RC3133), National Key R&D Program of China (Grant No. 2022YFB3102600), and National University of Defense Technology Cornerstone Project (Grant No. JS24-04).

Supporting information Figures A1–A5 and Tables A1–A9. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Platig J, Ott E, Girvan M. Robustness of network measures to link errors. Phys Rev E, 2013, 88: 062812
2 Wegge D, Vandebosch H, Eggermont S, et al. The strong, the weak, and the unbalanced. Soc Sci Comput Rev, 2015, 33: 315–342
3 Grady D, Thiemann C, Brockmann D. Robust classification of salient links in complex networks. Nat Commun, 2012, 3: 864
4 Matamalas J T, Arenas A, Gómez S. Effective approach to epidemic containment using link equations in complex networks. Sci Adv, 2018, 4: eaau4212
5 Artime O, Grassia M, de Domenico M, et al. Robustness and resilience of complex networks. Nat Rev Phys, 2024, 6: 114–131
6 Grassia M, Mangioni G. Edge dismantling with geometric reinforcement learning. In: Proceedings of the 15th Conference on Complex Networks, Switzerland, 2024. 195–202
7 Granovetter M S. The strength of weak ties. Am J Soc, 1973, 78: 1360–1380
8 Rajkumar K, Saint-Jacques G, Bojinov I, et al. A causal test of the strength of weak ties. Science, 2022, 377: 1304–1310
9 Zhou M Y, Wang F, Chen Z, et al. Weak link prediction based on hyper latent distance in complex network. Expert Syst Appl, 2024, 238: 121843
10 Meo P D, Ferrara E, Fiumara G, et al. On Facebook, most ties are weak. Commun ACM, 2014, 57: 78–84
11 Sanz-Cruzado J, Castells P. Enhancing structural diversity in social networks by recommending weak ties. In: Proceedings of the 12th ACM Conference on Recommender Systems, New York, 2018. 233–241
12 Chen K J, Zhang P, Yang Z, et al. iBridge: inferring bridge links that diffuse information across communities. Knowledge-Based Syst, 2020, 192: 105249
13 Fortunato S, Newman M E J. 20 years of network community detection. Nat Phys, 2022, 18: 848–850

14 Kirkley A, Newman M E J. Representative community divisions of networks. Commun Phys, 2022, 5: 40
15 He Z, Wei X, Chen W, et al. On the statistical significance of a community structure. IEEE Trans Knowl Data Eng, 2023, 35: 2887–2900
16 Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 2005, 435: 814–818
17 Xiao J, Wang Y J, Xu X K. Fuzzy community detection based on elite symbiotic organisms search and node neighborhood information. IEEE Trans Fuzzy Syst, 2022, 30: 2500–2514
18 Qian Y, Li Y, Zhang M, et al. Quantifying edge significance on maintaining global connectivity. Sci Rep, 2017, 7: 45380
19 Onnela J P, Saramäki J, Hyvönen J, et al. Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA, 2007, 104: 7332–7336
20 de Meo P, Ferrara E, Fiumara G, et al. A novel measure of edge centrality in social networks. Knowledge-Based Syst, 2012, 30: 136–150
21 Huszti É, Dávid B, Vajda K. Strong tie, weak tie and in-betweens: a continuous measure of tie strength based on contact diary datasets. Procedia-Soc Behav Sci, 2013, 79: 38–61
22 Liu Y, Tang M, Zhou T, et al. Improving the accuracy of the k-shell method by removing redundant links: from a perspective of spreading dynamics. Sci Rep, 2015, 5: 13172
23 Huang C Y, Chin W C B, Fu Y H, et al. Beyond bond links in complex networks:Local bridges, global bridges and silk links. Phys A-Stat Mech Its Appl, 2019, 536: 121027
24 Zhao N, Li J, Wang J, et al. Identifying significant edges via neighborhood information. Phys A-Stat Mech Its Appl, 2020, 548: 123877
25 Tantardini M, Ieva F, Tajoli L, et al. Comparing methods for comparing networks. Sci Rep, 2019, 9: 17557
26 Chen D, Shi D D, Qin M, et al. Complex network comparison based on communicability sequence entropy. Phys Rev E, 2018, 98: 012319
27 Lacasa L, Stramaglia S, Marinazzo D. Beyond pairwise network similarity: exploring mediation and suppression between networks. Commun Phys, 2021, 4: 136
28 Schieber T A, Carpi L, Díaz-Guilera A, et al. Quantification of network structural dissimilarities. Nat Commun, 2017, 8: 13928
29 Schieber T A, Carpi L C, Pardalos P M, et al. Diffusion capacity of single and interconnected networks. Nat Commun, 2023, 14: 2217
30 Tofighy S, Charkari N M, Ghaderi F. Link prediction in multiplex networks using intralayer probabilistic distance and interlayer co-evolving factors. Phys A-Stat Mech Its Appl, 2022, 606: 128043
31 Chen G, Zhou S, Li M, et al. Evaluation of community vulnerability based on communicability and structural dissimilarity. Phys A-Stat Mech Its Appl, 2022, 606: 128079
32 Peng P, Fan T, Ren X L, et al. Unveiling explosive vulnerability of networks through edge collective behavior. 2023. ArXiv:2310.06407
33 Braunstein A, DallAsta L, Semerjian G, et al. Network dismantling. Proc Natl Acad Sci USA, 2016, 113: 12368–12373
34 Wu L, Ren X L. Optimal bond percolation in networks by a fast-decycling framework. In: Proceedings of International Conference on Complex Networks and Their Applications, Palermo, 2023. 509–519
35 Fortunato S, Hric D. Community detection in networks: a user guide. Phys Rep, 2016, 659: 1–44
36 Shi D, Shang F, Chen B, et al. Local dominance unveils clusters in networks. Commun Phys, 2024, 7: 170
37 Javed M A, Younis M S, Latif S, et al. Community detection in networks: a multidisciplinary review. J Netw Comput Appl, 2018, 108: 87–111
38 Zhang X, Zhou K, Pan H, et al. A network reduction-based multiobjective evolutionary algorithm for community detection in large-scale complex networks. IEEE Trans Cybern, 2020, 50: 703–716
39 Newman M E J, Girvan M. Finding and evaluating community structure in networks. Phys Rev E, 2004, 69: 026113
40 Mou J, Dai B, Tan S, et al. The spindle approximation of network epidemiological modeling. New J Phys, 2024, 26: 043027
41 Brandes U. A faster algorithm for betweenness centrality. J Math Sociol, 2001, 25: 163–177
42 Cheng X Q, Ren F X, Shen H W, et al. Bridgeness: a local index on edge significance in maintaining global connectivity. J Stat Mech, 2010, 2010: P10011
43 Barrat A, Barthélemy M, Pastor-Satorras R, et al. The architecture of complex weighted networks. Proc Natl Acad Sci USA, 2004, 101: 3747–3752
44 Wang H, Wang C, Shu N, et al. The betweenness identities and their applications. In: Proceedings of IEEE 20th International Conference on Communication Technology (ICCT), Nanning, 2020. 1065–1069
45 Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Phys Rev E, 2008, 78: 046110
46 Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks. J Stat Mech, 2008, 2008: P10008
47 Raghavan U N, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E, 2007, 76: 036106
48 Pons P, Latapy M. Computing communities in large networks using random walks. In: Proceedings of International Symposium on Computer and Information Sciences, Istanbul, 2005. 284–293
49 Clauset A, Newman M E J, Moore C. Finding community structure in very large networks. Phys Rev E, 2004, 70: 6111
50 Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA, 2008, 105: 1118–1123