

PAPER • OPEN ACCESS

## The spindle approximation of network epidemiological modeling

To cite this article: Jianhong Mou *et al* 2024 *New J. Phys.* **26** 043027

View the [article online](#) for updates and enhancements.

You may also like

- [Aubry–Mather theory on graphs](#)  
Antonio Siconolfi and Alfonso Sorrentino
- [Random walks with long-range steps generated by functions of Laplacian matrices](#)  
A P Riascos, T M Michelitsch, B A Collet et al.
- [Beyond COVID-19: network science and sustainable exit strategies](#)  
J Bell, G Bianconi, D Butler et al.

**PAPER****The spindle approximation of network epidemiological modeling****Jianhong Mou<sup>1,6</sup>**, **Bitao Dai<sup>1,6</sup>** , **Suoyi Tan<sup>1</sup>** , **Petter Holme<sup>2,3</sup>** , **Sune Lehmann<sup>4</sup>** , **Fredrik Liljeros<sup>5</sup>**  and **Xin Lu<sup>1,\*</sup>** <sup>1</sup> College of Systems Engineering, National University of Defense Technology, Changsha, Hunan 410073, People's Republic of China<sup>2</sup> Department of Computer Science, Aalto University, Espoo 02150, Finland<sup>3</sup> Center for Computational Social Science, Kobe University, Kobe 657-8501, Japan<sup>4</sup> DTU Compute, Technical University of Denmark, Kongens Lyngby 2100, Denmark<sup>5</sup> Department of Sociology, Stockholm University, Stockholm 10691, Sweden<sup>6</sup> These authors contributed equally to this work.

\* Author to whom any correspondence should be addressed.

**E-mail:** [xin.lu.lab@outlook.com](mailto:xin.lu.lab@outlook.com)**Keywords:** network diffusion, propagation approximation, spindle vector, pairwise approximation**OPEN ACCESS****RECEIVED**  
4 February 2024**REVISED**  
27 March 2024**ACCEPTED FOR PUBLICATION**  
18 April 2024**PUBLISHED**  
3 May 2024Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.**Abstract**

Understanding the dynamics of spreading and diffusion on networks is of critical importance for a variety of processes in real life. However, predicting the temporal evolution of diffusion on networks remains challenging as the process is shaped by network topology, spreading non-linearities, and heterogeneous adaptation behavior. In this study, we propose the ‘spindle vector’, a new network topological feature, which shapes nodes according to the distance from the root node. The spindle vector captures the relative order of nodes in diffusion propagation, thus allowing us to approximate the spatiotemporal evolution of diffusion dynamics on networks. The approximation simplifies the detailed connections of node pairs by only focusing on the nodal count within individual layers and the interlayer connections, seeking a compromise between efficiency and complexity. Through experiments on various networks, we show that our method outperforms the state-of-the-art on BA networks with an average improvement of 38.6% on the mean absolute error. Additionally, the predictive accuracy of our method exhibits a notable convergence with the pairwise approximation approach with the increasing presence of quadrangles and pentagons in WS networks. The new metric provides a general and computationally efficient approach to predict network diffusion problems and is of potential for a large range of network applications.

**1. Introduction**

Network diffusion modeling is a priority for predicting information propagation, guiding epidemic responses, etc. During diffusion processes, contagious elements, like infectious diseases, information, etc propagate according to the network topology. As a ubiquitous process in networks, diffusion plays a key role in applications ranging from information propagation [1], epidemic spreading [2, 3], and adaptation to new technologies [4], to changes in behavior [5–8]. From the microscopic perspective, diffusion propagation among the individual nodes is highly affected by their properties and contacts, and possibly whom they are connected to. From the mesoscopic perspective, diffusion on networks depends on the local wiring of the network, characteristics of the carrier, and other environmental or spatial factors [9]. From the global perspective, the large-scale network structure, interventions, and the evolution of the pathogens or information spreading, affect the outcome of the diffusion process. This interplay with factors at various levels makes the analysis of network diffusion highly challenging.

Various models [10–13] have been proposed to characterize such diffusion processes, where the most straightforward and intuitive scenario is epidemic-like models where the disease is transmitted from infectious individuals (I) to the susceptible (S) when they interact with each other. Compartmental models [14–16], such as SIS, SIR, SEIR, etc are typically applied to study epidemic-like diffusion phenomena and

predict dynamics making the simplifying assumption that all individuals interact with each other with the same probability [17–19]. A wave of research has focused on the effects of network structure on spreading dynamics, including degree distributions, clustering, assortativity and communities on epidemic spreading velocity [20], epidemic size [21], and epidemic thresholds [22].

From a theoretical perspective, the heterogeneous mean-field [23] (HMF) approach assumes that nodes with the same degree are statistically equivalent, and the quenched mean-field [24] (QMF) approach describes the full network structure using the adjacency matrix. Although QMF improves predictions of outbreak sizes and thresholds relative to HMF, it overestimates the infection probability of susceptible vertices because of dynamic correlations among the infectious states of vertices, generated by the wave of infectious spreading through the network [25–27]. The weakness of the QMF approach is remedied e.g. by dynamic message-passing [28] (DMP), or the pairwise approximation [14] (PA) approach. Although these approaches provide more accurate epidemic size and threshold predictions, the numerical solution is time-consuming, which is an obstacle to wide application.

The key dynamic characteristic of a vertex on networks for our purposes is its role within the hierarchical structure of the network [29]. The hierarchical structure of a root node describes the immediate neighborhood and high-order neighborhoods [30] up to a maximal distance. The distribution of the number of nodes in the hierarchical structure from a specific node is characterized by a very rapid growth for short distances, followed by an exponential decay [31]. Similar to the  $k$ -shell [32], the hierarchical shell  $l$  of a network is defined as the set of nodes that are at distance  $l$  from a randomly chosen root node. A discriminative and computationally efficient metric based on the hierarchical structure is successfully utilized to measure the heterogeneity of a network in terms of connectivity distances, thus distinguishing and quantifying network dissimilarities [33, 34]. Concentrated on the dissimilarity metric, several researches have already sprung up, such as influential nodes identification [35], diffusion capacity of interconnected networks [36], filtering and compression of weighted networks [37], link prediction in multiplex networks [38], and evaluation of community vulnerability [39]. Obviously, the structure of the hierarchical shell is important for understanding the network's diffusion properties since propagation precisely occurs shell after shell [40], but its application with respect to diffusion prediction has been largely ignored, and whether the diffusion-wise hierarchical shell provides more information on spreading prediction has not been investigated.

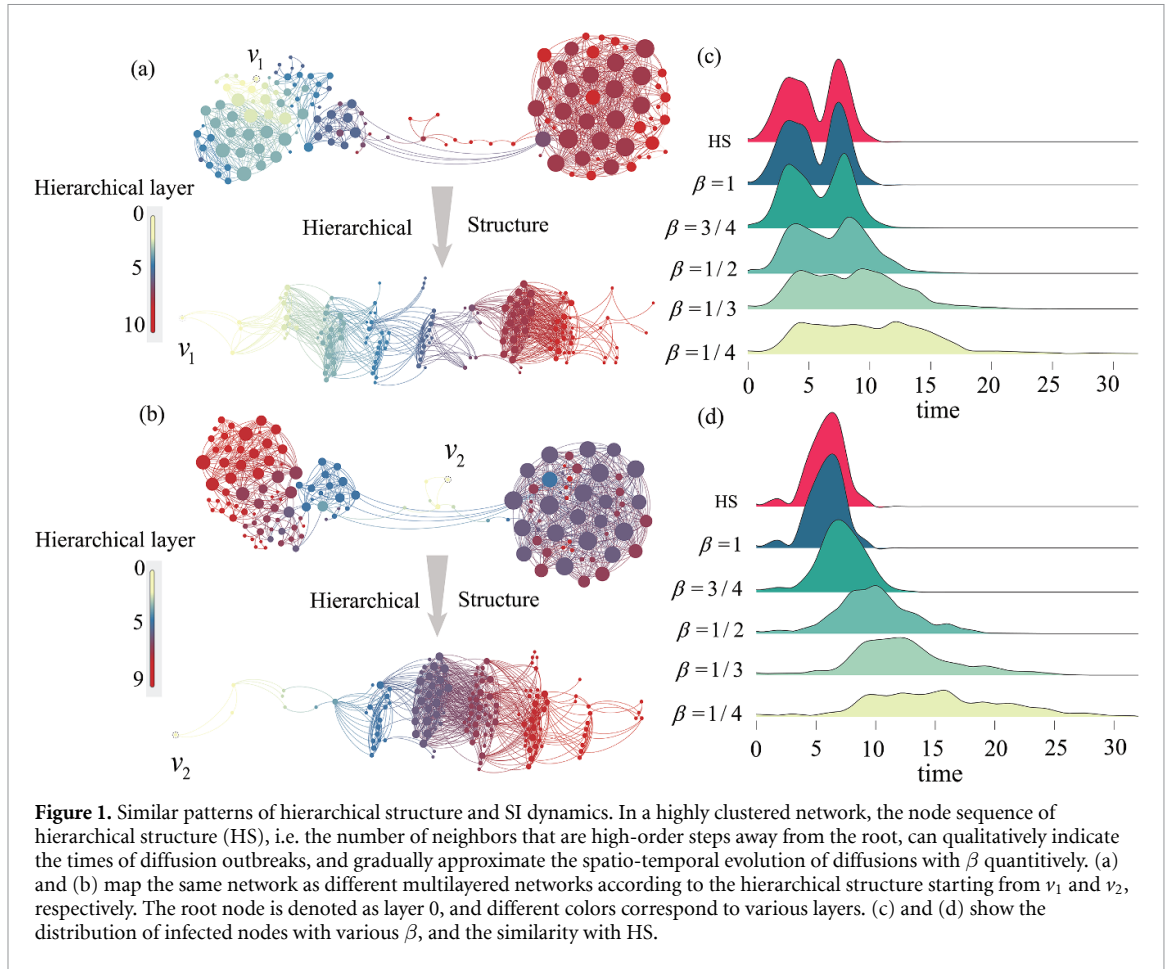
In this work, we will propose a novel theoretical scheme, the spindle approximation (SA) to approximate network diffusion. SA uses the hierarchical structure characterized via the so-called spindle vector and obtained by the breadth-first-search (BFS) process. The BFS process starts at the root node and visits all nodes at the current hierarchical layer before moving on to the nodes at the next hierarchical layer. For the SA, we decompose the epidemic-like diffusion process and find that the diffusion tree is always a possible spanning tree of the hierarchical structure and that the infection sequence of the SI model shows an outbreak pattern which is similar to the spindle vector. These similarities support the idea of basing diffusion modeling on hierarchical structures. We will pursue this idea by asking: First, how the hierarchical structure should be utilized to approximate network spreading on networks? Second, is the approximation better than any other methods for modeling the spreading dynamics?

SA resolves a network as a spindle-shaped multilayered network according to the hierarchical structure and considers effective infections from former layers and the same layer. To reduce the complexity of SA, we simplify it by mean-field theory to an approximation theory called SA\*, which assumes that the number of neighbors is equivalent for nodes in the identical layer and estimated as the average connections. We find that the hierarchical structure enables the outperformance of SA\* compared with QMF, DMP and PA methods on BA networks, and that the less complex SA\* provides more powerful prediction than SA in most WS and BA networks. The successful application of hierarchical structure in infection prediction calls for more analysis of the hierarchical structure.

## 2. Results

### 2.1. Hierarchical structure and network diffusion

At a given time in a diffusion process, the number of nodes in close contact with the infectious nodes is the fundamental predictor for the immediate propagation of the infection. For example, an infection starting from a source  $v$  will first propagate to some of its nearest neighbors, then to those that are two steps away from  $v$ , and so on. In the most extreme scenario, if the infection follows the SI model with an infection probability of  $\beta = 1$ , the spreading becomes the BFS process and is characterized by the hierarchical structure. The number of infected nodes at time  $t$  will be the number of nodes at most  $t$  steps away from  $v$ .



To further exemplify the relation between the hierarchical structure and diffusion evolution, we focus on BFS processes and virus transmission shaped by SI model starting from vertex  $v_1$  and  $v_2$  on the same network (see figures 1(a) and (b)). The hierarchical structure shows a similar pattern with an infection sequence. Specifically, both BFS processes and virus transmission originating from  $v_1$  show two peaks, while those starting from  $v_2$  have only one peak. More importantly, the distributions are more aligned when  $\beta$  is closer to one (see figures 1(c) and (d)).

## 2.2. Spindle vector

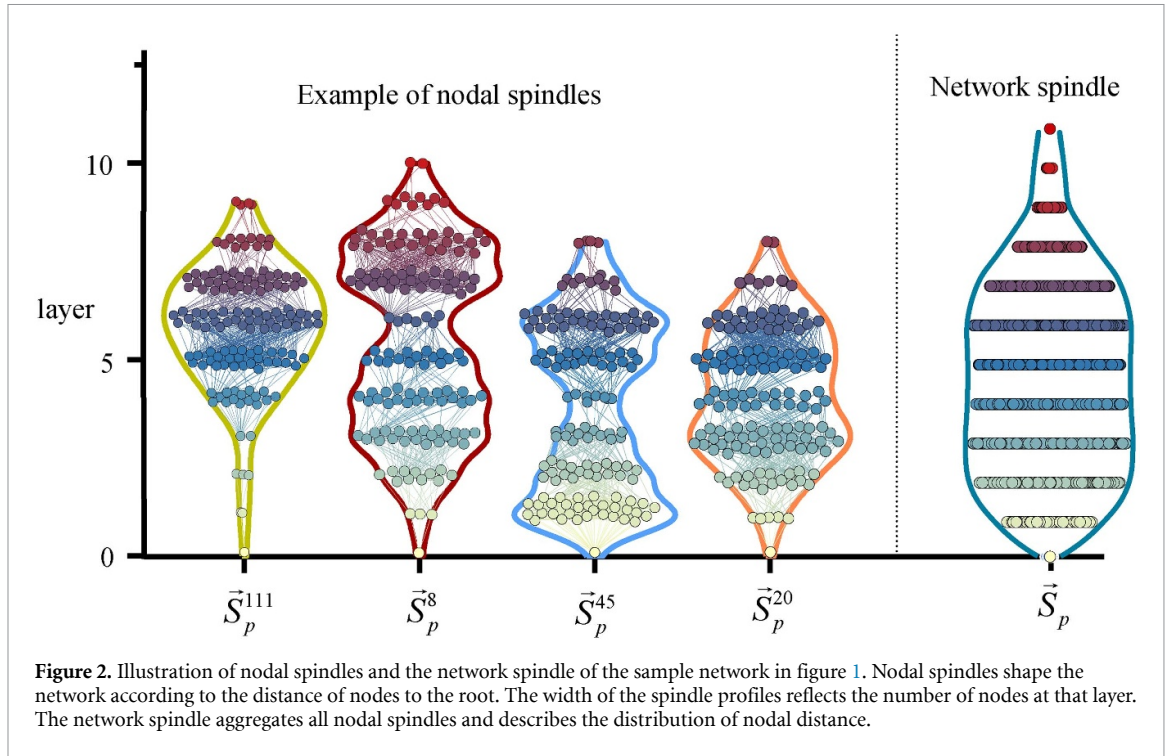
The hierarchical structure characterizes a BFS process and is also critical for describing how diffusion will occur. In most networks, elements in this sequence usually start with small numbers, i.e. the degree of the root vertex, increase until the peak and then reduce to a small number again at the most distant neighbors. We describe this pattern as a ‘nodal spindle’ and name  $\vec{S}_p^i$  as a new metric characterizing the diffusion capability of vertex  $i$  which is expressed as

$$\vec{S}_p^i = (n_0^i, n_1^i, \dots, n_{L_i}^i) \quad (1)$$

where  $n_{L_i}^i$  denotes the fraction of vertices in layer  $L_i$ , and the root vertex is viewed as layer 0.

Different roots may give rise to various characteristics of hierarchical structure, or distinct  $\vec{S}_p^i$ , in the identical network. To integrate the overall hierarchical structure and describe the diffusion capability of networks, we average  $\vec{S}_p^i$  over all nodes and get the ‘network spindle’ which is characterized by  $\vec{S}_p$  (shown in figure 2)

$$\vec{S}_p = \sum \vec{S}_p^i / N = (\vec{S}_p^i(0), \vec{S}_p^i(1), \dots, \vec{S}_p^i(L_{\max})) \quad (2)$$



**Figure 2.** Illustration of nodal spindles and the network spindle of the sample network in figure 1. Nodal spindles shape the network according to the distance of nodes to the root. The width of the spindle profiles reflects the number of nodes at that layer. The network spindle aggregates all nodal spindles and describes the distribution of nodal distance.

where  $\sum \vec{S}_p^i$  is the summation operator of vectors with various lengths, and  $L_{\max}$  is the total number of layers. Specifically,  $\vec{S}_p^i(j)$  denotes the average number of vertices in layer  $j$  over all nodes

$$\vec{S}_p^i(j) = \sum_{i \in \Theta(B_j)} n_j^i / N \quad (3)$$

where  $\Theta(B_j)(j = 0, 1, \dots, L_{\max})$  is the set of nodes from which layer  $j$  can be reached through BFS process.

As a variation of nodal distance distribution, the nodal spindle does not always illustrate unimodal distribution, such as for the center node of a star network. However, the network spindle, aggregating all nodal spindles, always depicts two narrow poles which can be described as a spindle. To unify the metric for nodes and networks, we extend the definition of distance distribution to spindle in the broad sense. More importantly, the network spindle incorporates some intrinsic characteristics of networks, such as the average degree ( $\langle k \rangle$ ), the average distance ( $\Lambda$ ), node betweenness ( $B_v$ ) and edge betweenness ( $B_e$ ) which are proved in appendix D.

**Lemma 1.** The total number of layers  $L_{\max}$  is the diameter of networks  $D$ , i.e.  $L_{\max} = |\vec{S}_p| = D$ .

**Lemma 2.**  $\vec{S}_p^i(0)$  is the reciprocal of  $N$ , i.e.  $\vec{S}_p^i(0) = 1/N$ .

**Lemma 3.**  $\vec{S}_p^i(1)$  is a linear functions of  $\langle k \rangle$ , specifically,  $\vec{S}_p^i(1) = \langle k \rangle / N$ .

**Lemma 4.** The area under the curve of  $\vec{S}_p^i$  approximates to  $\Lambda$  of networks, i.e.  $\sum_{j=0}^{L_{\max}} j \vec{S}_p^i(j) = \Lambda(N-1)/N$

**Lemma 5.** On tree networks with Poisson degree distribution,  $\vec{S}_p^i(2)$  can be expressed as  $\vec{S}_p^i(2) = (\langle k^2 \rangle - \langle k \rangle) / N$ .

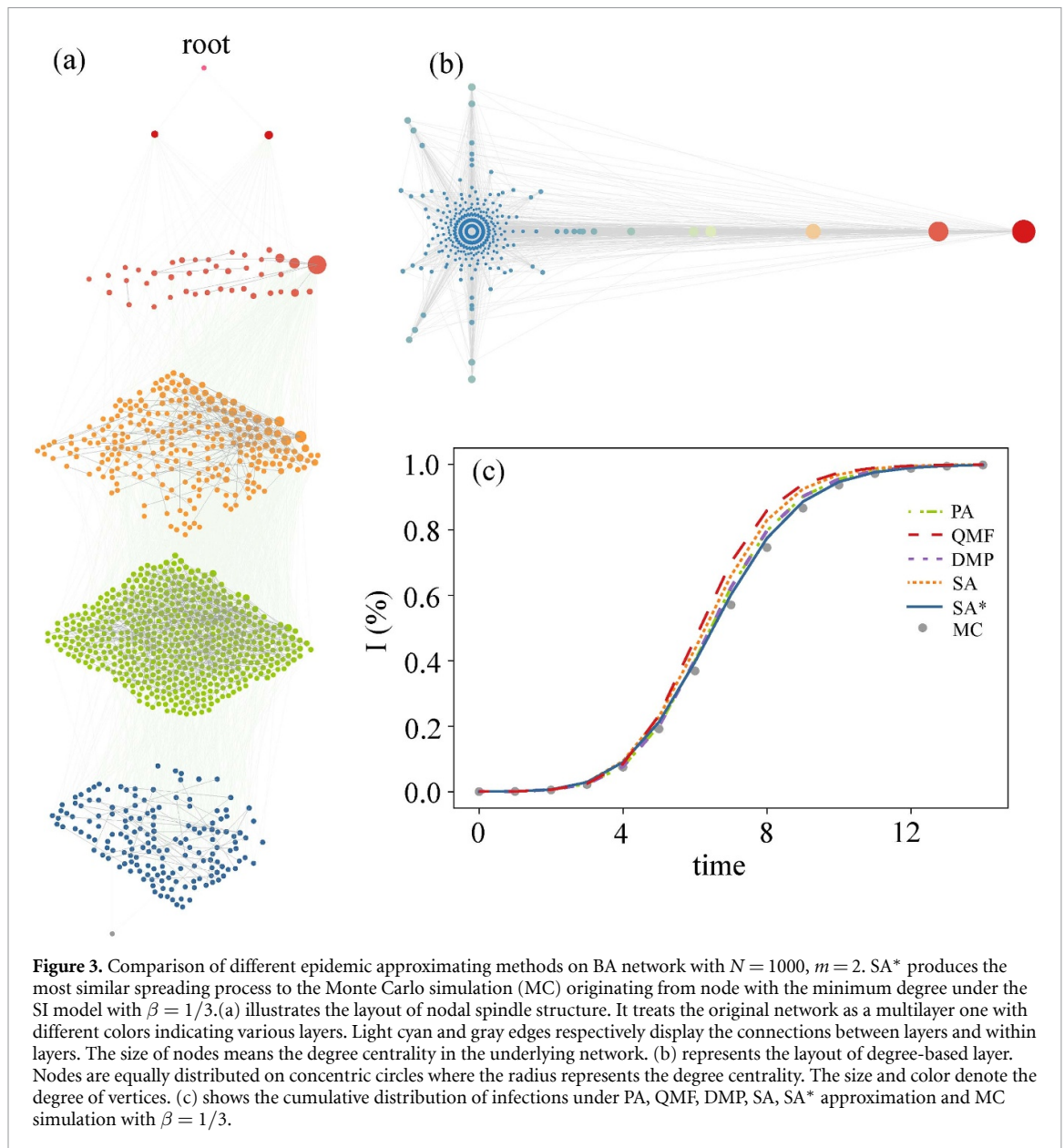
**Lemma 6.** The sum of  $B_v$  and  $B_e$  over all nodes can both be expressed as the function of the area under the curve of  $\vec{S}_p^i$ .

$$\sum_{i \in V} B_v(i) = \frac{N}{N-2} (\Lambda - 1) = \frac{N^2}{(N-2)(N-1)} \sum_{j=0}^{L_{\max}} j \vec{S}_p^i(j) - \frac{N}{N-2} \quad (4)$$

$$\sum_{l \in E} B_e(l) = \Lambda = \frac{N}{N-1} \sum_{j=0}^{L_{\max}} j \vec{S}_p^i(j) \quad (5)$$

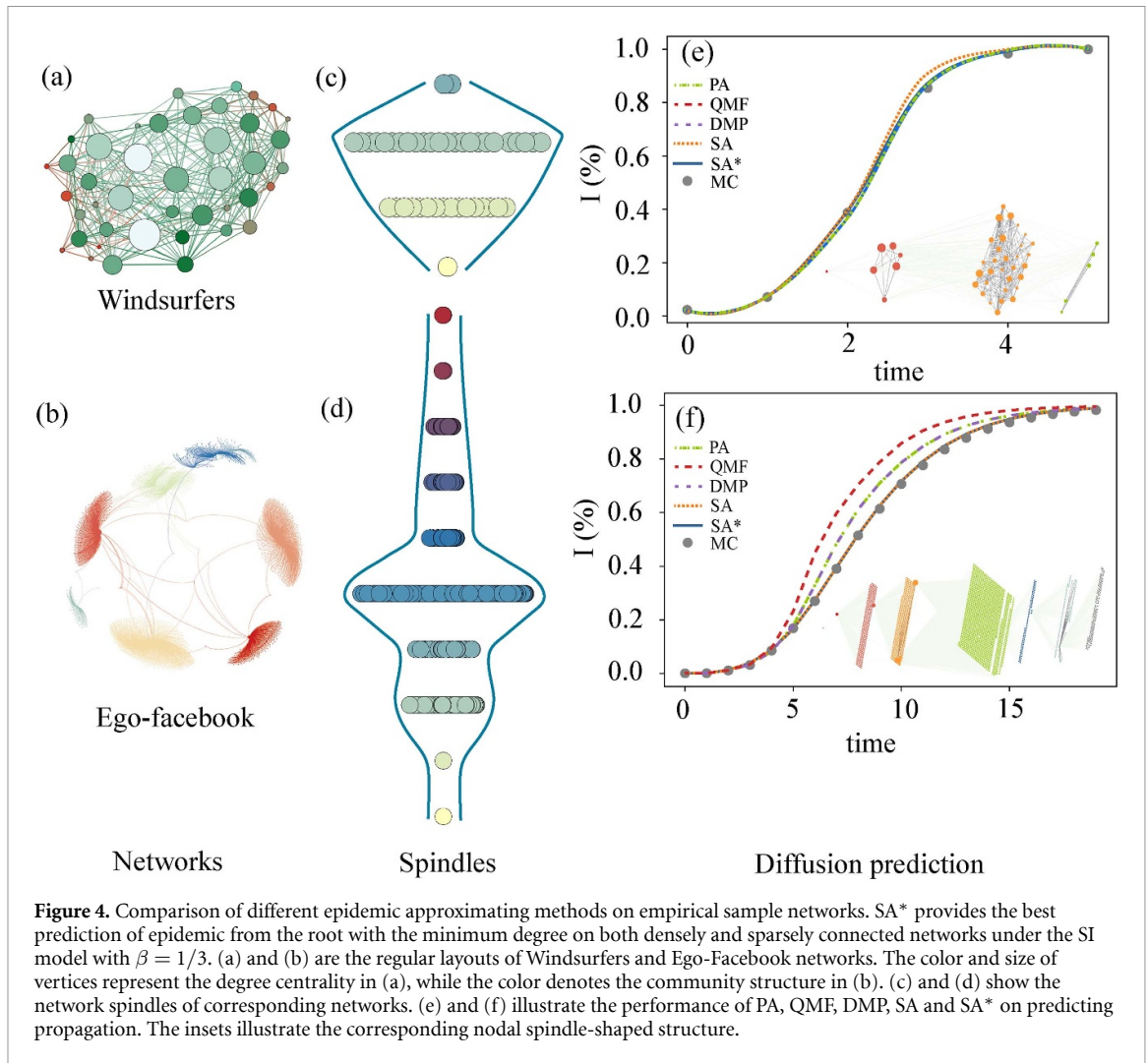
### 2.3. SA and SA\* methods on synthetic networks

The SA methodology is centered on modeling the epidemic propagation from neighboring nodes in the preceding and current layers, utilizing detailed information about node-specific connections within these



layers. The  $SA^*$  variant homogenizes nodes within the same layer and estimates the neighbors in the former and identical layer as the average connections using the mean-field method. Given that the maximal nodal distance within a network is typically significantly smaller than the network's magnitude, SA and  $SA^*$  notably curtail the number of requisite equations to fewer than  $D$ . This reduction stands in stark contrast to the  $N$  equations necessitated in QMF,  $N + 2E$  in DMP, and  $N + 4E$  in PA approach. Further,  $SA^*$  adopts a mesoscopic point for epidemic approximation, centering its focus on the quantification of nodes across each layer and the interlayer as well as intralayer connections. This approach deliberately eschews the intricate mapping of pairwise nodal connections, offering a more generalized yet insightful perspective into epidemic dynamics.

We use a BA network with  $N = 1000$  and  $m = 2$  to first test the performance of our method compared with QMF, DMP and PA, in predicting epidemic propagation originating from a single node with the minimum degree (see figure 3). The ground truth is the average infection originating from the node on a SI process with  $\beta = 1/3$  under Monte Carlo simulation 100 times. The layouts of the nodal spindle structure, i.e. the hierarchical layer, and degree-based layer of the corresponding network are shown in figures 3(a) and (b). The approximations from theoretical methods shown in figure 3(c) indicate that SA and  $SA^*$  perform better than QMF, and that  $SA^*$  is even more powerful than DMP and PA which take the dynamical correlations among the states of the nodes into consideration. Specifically,  $SA^*$  provides the most precise prediction with the mean absolute error (MAE) equaling 0.0113, smaller than those of SA, QMF, DMP, PA methods, i.e.  $MAE_{SA} = 0.0282$ ,  $MAE_{QMF} = 0.0358$ ,  $MAE_{DMP} = 0.0147$ ,  $MAE_{PA} = 0.0147$ .



We repeat the approximation with 100 different source nodes and average the MAE to assess the performance of the proposed method. The synthetic networks include WS networks with different rewiring probability ( $p_r^{WS} \in [0.01, 0.05, 0.1, 0.5, 0.9]$ ) and BA networks with different number of edges for new attaching nodes ( $m \in [2, 3, 4, 5]$ ). Experiments on synthetic networks indicate that SA, SA\* method predicts the temporal evolution of diffusion dynamics on BA networks better than those on WS networks, and that SA\* is more efficient than SA on most BA and WS networks. For the SA method, the MAE of BA networks ranges from 0.0134 to 0.0383 with an average of 0.0234, much smaller than that of WS networks ranging from 0.0126 to 0.2058 with an average of 0.0966. For the SA\* method, the average MAE on BA and WS networks equals 0.0154 and 0.0906, respectively. More importantly, SA\* outperforms QMF, DMP and PA in all BA networks with an average improvement of 38.6% on MAE with 95% confidence interval [36.7%, 40.6%], while SA\* can only produce more precise prediction than QMF on WS networks with high rewiring probability bigger than 0.5 (see appendix A tables A1 and A2). Theoretically, WS networks with high rewiring probability usually induce a low clustering coefficient, reducing the connections within the same layer shaped by the nodal spindle. The few connections decrease the complexity of correlations among the states of neighboring nodes, thus facilitating the propagation prediction by SA\* for its advantage in partially dealing with the correlations which is missing in QMF.

In addition, we also check the performance of SA on scale-free networks with the power law  $\gamma > 3$  using the uncorrelated configuration model [41] with the minimum degree  $m = 4$ . Results illustrate that all approximations generated larger MAE to predict the epidemic propagation process, while SA\* is consistently superior to QMF and inferior to PA and DMP in these networks, which is compatible with the findings on WS networks (see appendix B figure B1).

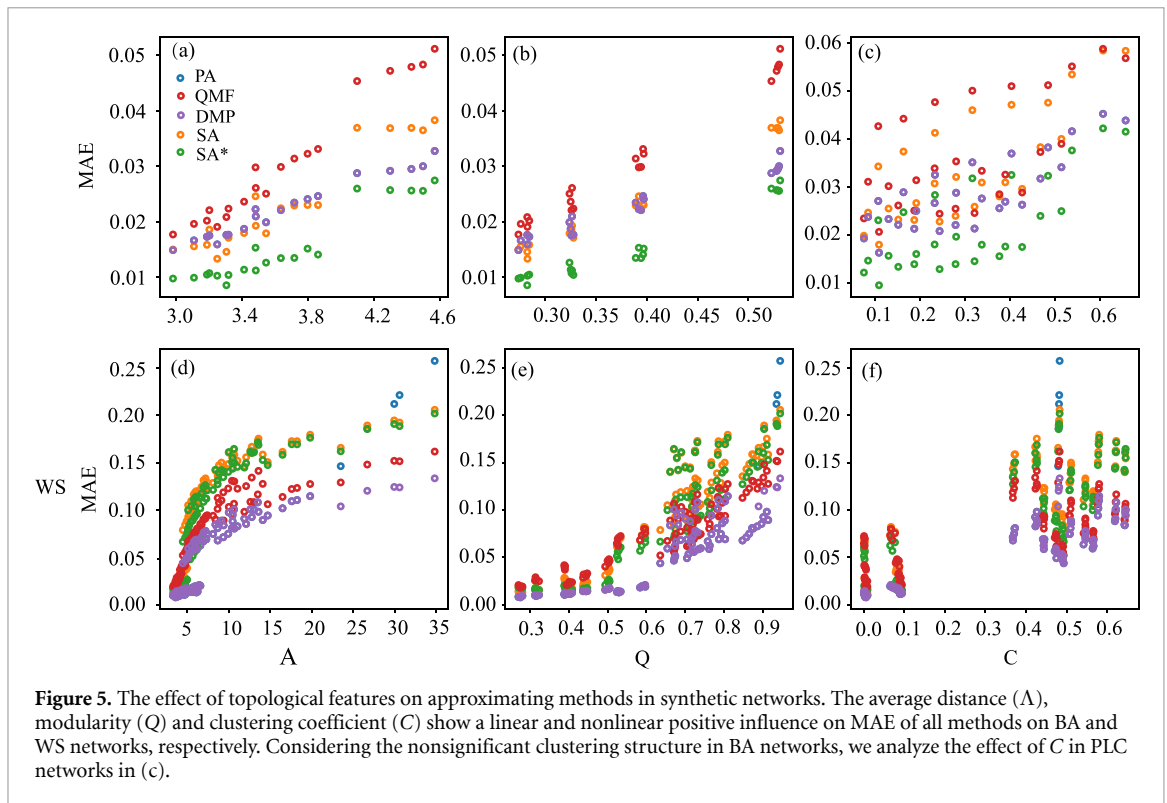
#### 2.4. SA and SA\* methods on empirical networks

Next, we illustrate the performance of SA and SA\* under the SI epidemic process on Windsurfers [42] and Ego-Facebook [43] networks. As shown in figure 4, SA\* provides the best prediction of the epidemic from

**Table 1.** The performance of SA\* compared to QMF, PA in diffusion prediction on empirical networks.

Networks	MAE <sub>PA</sub>	MAE <sub>QMF</sub>	MAE <sub>SA*</sub>	RI <sub>QMF</sub>	RI <sub>PA</sub>
Caenorhabditis	<b>0.017</b>	0.019	0.019	0.042	0.123
Ego-facebook	0.038	0.059	<b>0.033</b>	-0.429	-0.117
Iceland	<b>0.029</b>	0.043	0.039	-0.082	0.357
Interactome_vidal	<b>0.035</b>	0.044	0.046	0.030	0.300
Karate	<b>0.021</b>	0.025	0.023	-0.065	0.123
Metabolic_ana	<b>0.028</b>	0.032	0.029	-0.084	0.060
Metabolic_blo	<b>0.027</b>	0.031	0.028	-0.094	0.071
Ugandan	0.011	0.013	<b>0.011</b>	-0.204	-0.002
Windsurfers	0.013	0.012	<b>0.012</b>	0.024	-0.019
Zebras	<b>0.011</b>	0.012	0.013	0.078	0.199

Note: Relative Improvement  $RI_{QMF} = (MAE_{SA*} - MAE_{QMF})/MAE_{QMF}$ ,  
 $RI_{PA} = (MAE_{SA*} - MAE_{PA})/MAE_{PA}$ .



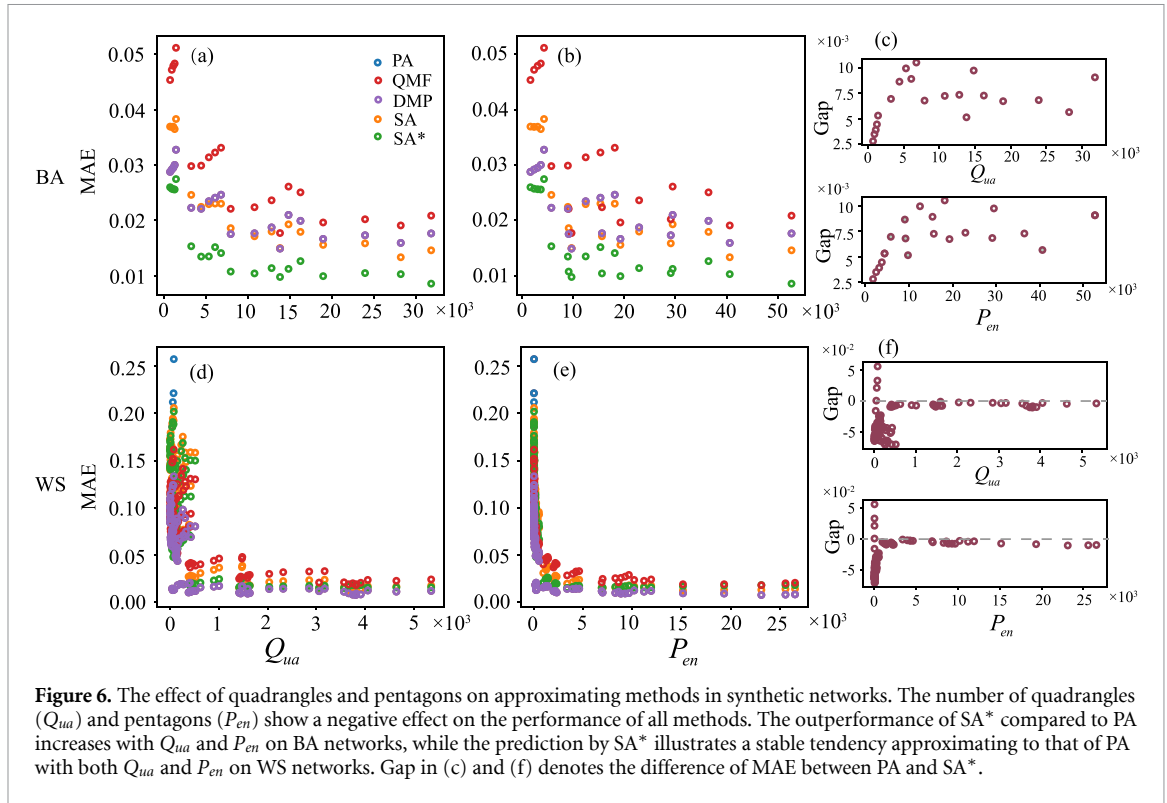
**Figure 5.** The effect of topological features on approximating methods in synthetic networks. The average distance ( $\Lambda$ ), modularity ( $Q$ ) and clustering coefficient ( $C$ ) show a linear and nonlinear positive influence on MAE of all methods on BA and WS networks, respectively. Considering the nonsignificant clustering structure in BA networks, we analyze the effect of  $C$  in PLC networks in (c).

the root with the minimum degree on both densely and sparsely connected networks. We repeat the approximation with 100 different source nodes for empirical networks. Our experiments demonstrate that SA\* outperforms SA in all 10 empirical networks. For SA\* method, MAE ranges from 0.0106 to 0.0455 with an average of 0.0255, smaller than SA method ranging from 0.0181 to 0.0471 with an average of 0.0298 (see appendix A table A3). More importantly, SA\* demonstrates comparable or superior performance to QMF, and shows predictions that are not significantly worse and, in some cases, even superior to PA and DMP. Specifically, SA\* outperforms QMF in 60% of real networks with an average improvement of 7.8% on MAE, while it surpasses PA in 30% of real networks (see table 1 and appendix A table A3).

## 2.5. Sensitivity analysis

To understand the performance of SA and SA\*, we further investigate how performance changes with the network topological features. Specifically, we measure MAE for networks with varying average distance ( $\Lambda$ ), modularity ( $Q$ ) and clustering coefficients ( $C$ ).  $\Lambda$ ,  $Q$  in BA networks and  $C$  in PLC networks (Scale-free networks with adjustable clustering coefficients), all have a linear positive influence on MAE of all approximating methods, while those in WS networks illustrate nonlinear positive effect (see figure 5). Except for triangles described by  $C$ , we find that the number of quadrangles ( $Q_{ua}$ ) and pentagons ( $P_{en}$ ) show a negative effect on the performance of all approximating methods. In addition, the superiority of SA\* increases with the number of quadrangles and pentagons in BA networks, and the epidemic prediction of





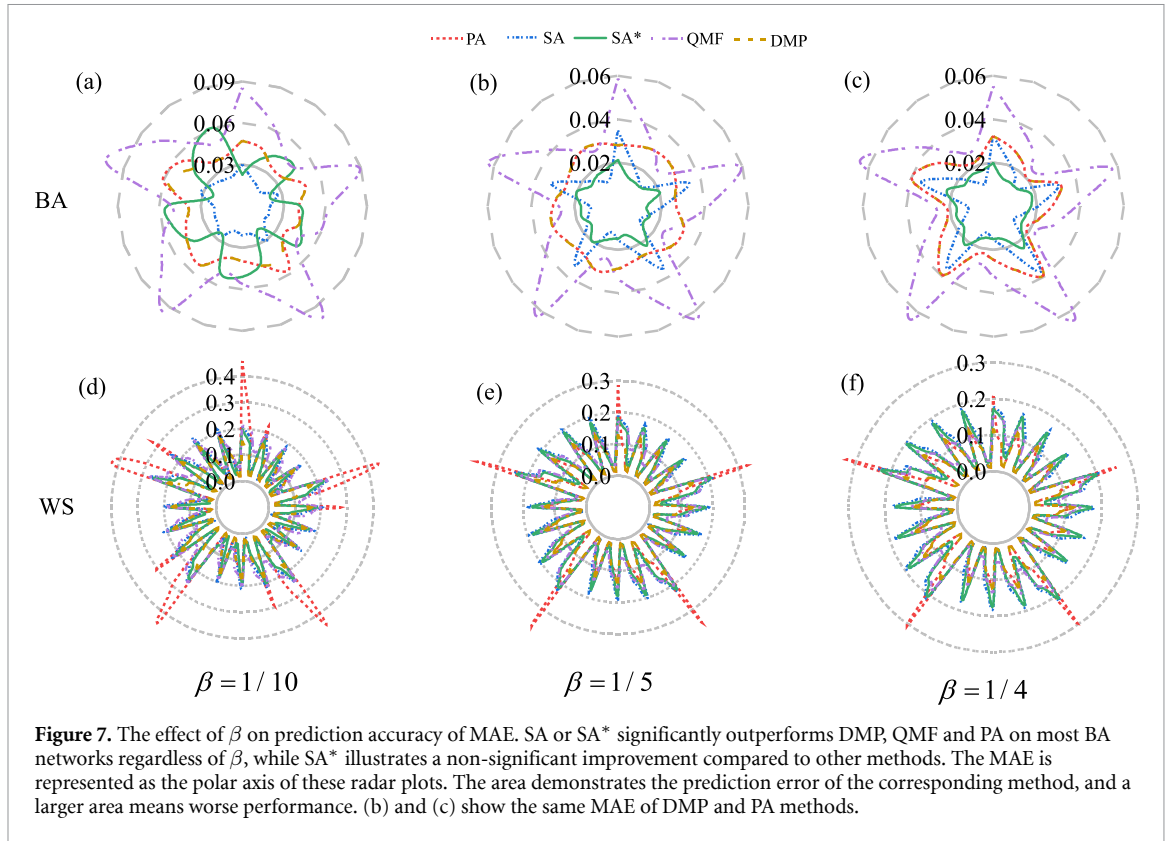
SA\* is nearly identical to that of PA method on WS networks with more than 1000 quadrangles or 3000 pentagons (see figure 6). Large  $Q_{ua}$  and  $P_{en}$  are associated with low clustering coefficient on WS networks (see appendix A table A2), which is in turn explainable to the observation.

Theoretically, considering the relationship  $\Lambda \approx \sum_{l=1}^{L_{\max}} l \bar{S}_p(l)$ , large  $\Lambda$  usually produces a flat and long distribution of  $\bar{S}_p$  in WS networks and induces the cumulation of prediction error. Moreover, the nonlinear positive relationship between  $Q$ ,  $C$  and  $\Lambda$  in WS networks explains the effect of  $Q$  and  $C$  on MAE. In BA networks, the small  $C$  described by  $C \sim (ln t)^2 / t$  and short  $\Lambda$  induced by  $\Lambda \sim ln N / ln ln N$  respectively witness sparse connections within the same layer and reduce the cumulation of error, thus supporting the better prediction. In addition, quadrangles and pentagons illustrate the weak dynamic correlation among neighbors for the lack of connections within the same layer under the projection of spindle structure.

In addition, to understand the effect of infection rate on diffusion prediction, we investigate how the performance of the proposed methods changes along with  $\beta$ . Generally, SA is more powerful than SA\* when  $\beta = 1/10$ , and SA or SA\* significantly outperforms DMP, QMF and PA on most BA networks regardless of  $\beta$ . For WS networks, SA\* illustrates a non-significant improvement compared to other methods, except for slight superiority to PA in several cases (see figure 7).

### 3. Conclusions and outlook

Diffusion on networks propagates through the direct interaction among vertices, of which the temporal evolution is determined by the distance of vertices from the source. Interestingly, the spindle vector is capable of quantifying the hierarchical structure from a source and describing the feature of the propagation chain, thus providing an interesting view of the diffusion dynamics. Based on this new network feature, the proposed SA and SA\* methods demonstrate their superiority over the state-of-the-art in predicting the temporal evolution of spreading dynamics on many synthetic and empirical networks. This work extends the existing literature on diffusion prediction in two ways. First, we demonstrate that the hierarchical structure of networks exhibits the intrinsic diffusion pattern, and define nodal spindle vector and network spindle vector to quantify the diffusion ability of vertices and networks. Second, we propose a novel approximation method based on the spindle vector and successfully seek a compromise between efficiency and complexity. Notably, the SA\* approach is only based on the count of nodes in each BFS layer and the connection between different layers, ignoring the whole network information. More importantly, it outperforms the state-of-the-art on BA networks in diffusion prediction, especially the PA method which is based on details about the evolution of the pair node states.



Although we have tried to access the characteristics of hierarchical structure embedded in the nodal or network spindle, more thorough analyses are required to reveal the correlation and high-order topology described by the spindle vector. In line with the statistical analysis, the application of hierarchical structure on network dynamics will support a brand-new view of mesoscopic structure on network functionality. In addition, the extension of the hierarchical structure to high-order networks modeled by hypergraph and simplicial complex may also be beneficial. The study is limited to SI epidemic-like processes on several synthetic and empirical networks, supporting the significant outperformance on propagation prediction of SA\* on BA networks. However, understanding the propagation dynamics from the perspective of hierarchical structure may require the construction of hybrid diffusion models. Analyzing the epidemic thresholds of SIS and SIR model based on the hierarchical structure will be the work at the corner.

## 4. Data and methods

### 4.1. Dataset

**Synthetic networks.** We choose generative network models that are common and reflect several topological properties of real-world networks (summarized in appendix C table C1).

*Small-world networks (WS)* [44]. WS model explains the coexistence of a high clustering coefficient and a short distance (small-world behavior) by rewiring links with probability  $p_r^{WS}$  in a ring lattice with non-overlap connections to  $K$  nearest neighbors for each node. This construction allows us to tune the graph between a symmetric state with long distances ( $p_r^{WS} = 0$ ) and a disordered state with short distances ( $p_r^{WS} = 1$ ).

*Scale-free networks (BA)* [45]. BA is a model that generates random scale-free networks using a preferential attachment mechanism. It aims to explain the existence of highly heterogeneous degree distribution in real networks. It generates a graph by attaching new nodes along with  $m$  edges which are attached to existing nodes in proportion to their degree.

*Scale-free networks with adjustable clustering coefficients (PLC)* [46]. PLC model can diversify average clustering in scale-free BA networks with an extra step that the new node will be linked to a random neighbor of the connected with probability  $p_{\Delta}^{BA}$  to form triangles. The model takes  $(m-1)p_{\Delta}^{BA}$  as the parameter to control clusters. This construction improves BA networks in the sense that it enables a higher average clustering to be attained if desired.

**Real-world networks.** We use 10 empirical networks with different characteristics to testify the performance of the proposed method. The self-loops, directionality, and edge weights are all removed, and we only consider the giant connected components of the networks. The topological features of these empirical networks are summarized in appendix C table C2. Note that  $N, M, D, \langle k \rangle, \Lambda, C, Q$  respectively denote the number of vertices and links, the diameter, the average degree, the distance, average clustering coefficient and modularity.

#### 4.2. SA on network diffusion

We let  $G = (V, E)$  denote an undirected and unweighted graph. Here  $V$  is the vertex set and  $E$  is the set of edges (unordered pairs of vertices). Furthermore,  $N$  denotes the number of nodes and  $M$  the number of edges. We consider the spread of epidemics on  $G$  described by SI model with a single source  $j$ . Since we are interested in the relationship between nodal spindle vector  $\vec{S}_p^j$  and the fraction of infection  $I^j$ , we turn to the probability of a randomly chosen individual in layer  $i$  being infected at time  $t$ . The proposed method separates nodes according to nodal spindle layers from  $j$ , which can be regarded as a standard metapopulation model [47] where vertices belonging to the same layer are gathered as a subpopulation with heterogeneous connections and subpopulations are connected in the spindle structure.

A susceptible vertex  $k$  in layer  $i$  remains susceptible until time  $t$  if  $t$  is numerically smaller than  $i$ , which means that no infection has been traversed to layer  $i$  since the transmission occurs once each time along the spindle. Supposing that neighbors of  $k$  are independent, the probability of getting infected by its  $d_k(t)$  infectious neighbors at time  $t$  is thus

$$\phi_k(t) = 1 - (1 - \beta)^{d_k(t)}. \quad (6)$$

Extending the definition to the case of spindle layers, infectious neighbors of  $k$  at time  $t$  are comprised of those in the preceding layer  $d_k^-(t)$ , the same layer  $d_k^o(t)$ , and the following layer  $d_k^+(t)$ , yielding

$$d_k(t) = d_k^-(t) + d_k^o(t) + d_k^+(t). \quad (7)$$

Connections between layer  $i$  and  $i - 1$  will be traversed twice when we focus on layer  $i$  and  $i - 1$  respectively, causing repeated infections that could be compensated by ignoring  $d_k^+(t)$  for each vertex. Therefore, the number of infected individuals in layer  $i$  at time  $t$ ,  $F_i^j(t)$ , is determined by connections to infection in the preceding layer  $i - 1$  and the identical layer  $i$  at time  $t - 1$ . Since the contagious item cannot reach vertices in layer  $i$  when  $i > t$ ,  $F_i^j(t)$  evolves like

$$\frac{dF_i^j(t)}{dt} = \sum_{k \in \Theta_i} (1 - \rho_k^t) \left[ 1 - (1 - \beta)^{d_k^-(t) + d_k^o(t)} \right], i \leq t \quad (8)$$

where  $\Theta_i$  is the assembly of vertices in layer  $i$  and  $1 - \rho_k^t$  is the probability of  $k$  being in the susceptible state at time  $t$ .

Furthermore, we assume that the infected probability of susceptible nodes at time  $t$  within the same layer  $i$  are identical and equal to the fraction of infectious vertices in layer  $i$ :

$$\rho_k^t \approx p_i^t = F_i^j(t) / n_i^j \left( n_i^j \in \vec{S}_p^j \right). \quad (9)$$

Infectious neighbors of vertex  $k$  can be respectively estimated as

$$d_k^-(t) \approx k_k^- p_{i-1}^t \quad (10)$$

$$d_k^o(t) \approx k_k^o p_i^t \quad (11)$$

where  $k_k^-, k_k^o$  correspond to the in-layer and mid-layer degree of vertex  $k$ . So far, the SA method can be summarized as

$$\frac{dF_i^j(t)}{dt} = \sum_{k \in \Theta_i} (1 - p_i^t) \left[ 1 - (1 - \beta)^{k_k^- p_{i-1}^t + k_k^o p_i^t} \right], i \leq t \quad (12)$$

and the total infections at time  $t$  is

$$I_j^{SA}(t) = \sum_{i=0}^{L_{\max}} F_i^j(t) \quad (13)$$

where  $L_{\max}$  is the total number of layers.

### 4.3. The simplified SA\* on network diffusion

The prediction of SA method requires information about the in-layer and mid-layer degree of each node, which may induce the complexity of the method. Therefore, we simplify it by mean-field theory to an approximation theory called SA\*. Specifically, we assume that the in-layer and mid-layer degree of nodes within the same layer  $i$  are identical and equal to the average connections of each node:

$$k_k^- = E_i^- / n_i, k_k^o = E_i^o / n_i, k \in \Theta_i \quad (14)$$

where  $E_i^-$ ,  $E_i^o$  represent connections between layer  $i - 1$  and  $i$ , and those within layer  $i$ . Therefore, SA\* can be expressed as

$$\frac{dF_i^j(t)}{dt} = n_i (1 - p_i^t) \left[ 1 - (1 - \beta)^{(E_i^- p_{i-1}^t + E_i^o p_i^t) / n_i} \right], i \leq t \quad (15)$$

Here,  $n_i$  is the number of nodes in layer  $i$  originating from node  $j$ ,  $(1 - p_i^t)$  is the probability of a node, in layer  $i$ , being in the susceptible state at time  $t$ , and  $1 - (1 - \beta)^{(E_i^- p_{i-1}^t + E_i^o p_i^t) / n_i}$  is the probability that a node will be infected by its neighbors, where  $(E_i^- p_{i-1}^t + E_i^o p_i^t) / n_i$  represents the infected neighbors. The total infections are denoted as:

$$I_j^{SA^*}(t) = \sum_{i=0}^{L_{\max}} F_i^j(t) \quad (16)$$

### 4.4. Discrete version of QMF method on network diffusion

As for the SI epidemic dynamics approximated by QMF, an infected node tries to transmit the disease to its neighbors with probability  $\beta$  per unit time. This forms a Markov chain where the probability of a node being infected depends only on the last time step. Specifically, the susceptible node  $i$  is infected at time  $t$  by at least a neighbor with probability  $(1 - q_i(t))(1 - p_i(t))$ , then the discrete-time version of the evolution of the probability of infection of any node  $i$  reads

$$p_i(t+1) = (1 - q_i(t))(1 - p_i(t)) + p_i(t) \quad (17)$$

where  $p_i(t)$  is the probability that node  $i$  is infected at time  $t$ ,  $q_i(t)$  is the probability of node  $i$  not being infected by any neighbor

$$q_i(t) = \prod_{j=1}^N (1 - \beta a_{ji} p_j(t)) \quad (18)$$

where  $a_{ji}$  represents the adjacent relationship between node  $j$  and  $i$ . If they are connected,  $a_{ji} = 1$ , otherwise,  $a_{ji} = 0$ . The fraction of infection at time  $t$  can be expressed as

$$I_j^{QMF}(t) = \sum_{k=1}^N p_k(t) / N \quad (19)$$

### 4.5. Discrete version of DMP method on network diffusion

In the DMP approach, a node designated as being in a 'cavity' state is precluded from transmitting an infection to its adjacent nodes, while still remaining susceptible to infection from them. This framework inherently incorporates dynamic correlations among the states of neighboring nodes. The temporal evolution of the probability that a given node  $i$  remains susceptible at time  $t$ , denoted as  $p_S^i(t)$ , is articulated as

$$p_S^i(t) = p_S^i(0) \prod_{j \in \Phi_i} \theta^{j \rightarrow i}(t). \quad (20)$$

Here,  $\theta^{j \rightarrow i}(t)$  symbolizes the probability that the disease has not traversed through the edge  $j \rightarrow i$  up to time  $t$ ;  $\Phi_i$  signifies the set neighbors of node  $i$ . The updating rule for  $\theta^{j \rightarrow i}(t)$  is given by

$$\theta^{j \rightarrow i}(t) = \theta^{j \rightarrow i}(t-1) - \beta \phi^{j \rightarrow i}(t-1) \quad (21)$$

where  $\phi^{j \rightarrow i}(t)$  is the probability that the disease has not been transmitted through the edge  $j \rightarrow i$  up to time  $t$  and node  $j$  is infected at time  $t$ . The update mechanism for  $\phi^{j \rightarrow i}(t)$  is as follows

$$\phi^{j \rightarrow i}(t) = (1 - \beta) \phi^{j \rightarrow i}(t-1) + (p_S^j(t-1) - p_S^j(t)). \quad (22)$$

In this expression,  $p_S^{i \setminus k}(t)$  represents the probability that node  $i$  remains susceptible upon disregarding any infection originating from its neighbor node  $k$ . The right-hand side of (22) delineates two distinct scenarios: the first term describes the situation where node  $j$  infects node  $i$  with a rate  $\beta$  at time  $t$  if it is infected at time  $t - 1$ ; the second term reflects the probability of node  $j$  being infected given it was susceptible at time  $t - 1$ .

By excluding node  $k$  from  $\Phi_i$  in (20), the following is obtained

$$p_S^{i \setminus k}(t) = p_S^i(0) \prod_{j \in \Phi_i \setminus k} \theta^{j \rightarrow i}(t). \tag{23}$$

To complete the recursion updating rules, the initial conditions are set as:

$$\theta^{j \rightarrow i}(0) = 1, \phi^{j \rightarrow i}(0) = \delta_j^0. \tag{24}$$

Here,  $\delta_j^0$  indicates the initial status of node  $j$ . If it is infectious,  $\delta_j^0 = 1$ , otherwise,  $\delta_j^0 = 0$ . The fraction of infection at time  $t$  within the DMP framework is computed as

$$I_j^{\text{DMP}}(t) = \sum_{k=1}^N (1 - p_S^k(t)). \tag{25}$$

#### 4.6. Discrete version of PA method on network diffusion

In the PA approach, the focal point is the intricate interplay of joint and conditional probabilities associated with each link, formulated through epidemic link equations. For a given link connecting nodes  $i$  and  $j$ , the probability of node  $i$  remaining in a susceptible state is expressed as

$$P_i^S = P_{ij}^{SS} + P_{ij}^{SI}. \tag{26}$$

Here,  $P_{ij}^{SS}$  represents the joint probability that both node  $i$  and node  $j$  are susceptible, while  $P_{ij}^{SI}$  signifies the joint probability of node  $i$  being susceptible and node  $j$  being infected. In a similar fashion, the probability of node  $i$  being infected is deduced as  $P_i^I = P_{ij}^{II} + P_{ij}^{SI}$ . Integrating these restrictions, the equations for each node  $i$  can be succinctly formulated as

$$P_i^I(t+1) = (1 - P_i^I(t))(1 - q_i(t)) + P_i^I(t). \tag{27}$$

In this context,  $q_i(t)$  delineates the probability of node  $i$  evading infection through any pairwise interaction with its neighbors, defined as

$$q_i(t) = \prod_{j \in \Phi_i} \left( 1 - \beta \frac{P_{ij}^{SI}(t)}{P_i^S(t)} \right). \tag{28}$$

To fully capture the dynamics of the system,  $L$  additional equations are required, each corresponding to a specific link. These equations account for the probability of a link connecting two nodes in the infected state II, transitioning from one of the four potential states: SS, SI, IS, II. Thus, we have

$$P_{ij}^{II}(t+1) = P_{ij}^{SS}(t)(1 - q_{ij}(t))(1 - q_{ji}(t)) \tag{29a}$$

$$+ P_{ij}^{SI}(t)(1 - (1 - \beta)q_{ij}(t)) \tag{29b}$$

$$+ P_{ij}^{IS}(t)(1 - (1 - \beta)q_{ji}(t)) \tag{29c}$$

$$+ P_{ij}^{II}(t). \tag{29d}$$

Here,  $q_{ij}(t)$  specifies the probability of node  $i$  not being infected by any neighboring node other than  $j$  through a link

$$q_{ij}(t) = \prod_{r \in \Phi_i, r \neq j} \left( 1 - \beta \frac{P_{ir}^{SI}(t)}{P_i^S(t)} \right). \tag{30}$$

Finally, the fraction of infection at time  $t$ , as predicted by the PA approach, is determined by

$$I_j^{\text{PA}}(t) = \sum_{k=1}^N P_i^k(t). \tag{31}$$

#### 4.7. Evaluation metrics

The prediction error of epidemic propagation originating from node  $j$  through the SA method can be quantified by the MAE as follows

$$\text{MAE}_{\text{SA}}^j = \frac{1}{m} \sum_{t=1}^m |I_j^{\text{SA}}(t) - I_j^{\text{MC}}(t)| \quad (32)$$

where  $I_j^{\text{MC}}(t)$  is the fraction of infection at time  $t$  under Monte Carlo simulation,  $m$  is the maximum length of  $I_j^{\text{SA}}$  and  $I_j^{\text{MC}}$ . In this work, we focus on the overall performance of approximating methods on different types of networks by averaging the MAE over several infection sources

$$\text{MAE}_{\text{SA}} = \frac{1}{s} \sum_{j=1}^s \text{MAE}_{\text{SA}}^j \quad (33)$$

where  $s$  represents the number of experiments on networks, i.e. the number of independently chosen infection sources. The calculation of 32 and 33 also applies to QMF, DMP, PA and SA\* when different approximation methods are used, respectively.

#### Data availability statement

The 10 empirical networks are extracted from links <https://networks.skewed.de> and <https://graph-tool.skewed.de>. No new data were created or analysed in this study.

#### Acknowledgments

We thank Dr Fei Gao for the helpful discussion. This study was funded by the National Nature Science Foundation of China (72025405, 72088101, 72001211), the Innovation Team Project of Colleges in Guangdong Province (2020KCXTD040), and the Hunan Science and Technology Plan Project (2020TP1013, 2020JJ4673). P H was supported by JSPS KAKENHI Grant Number JP 21H04595.

#### Author contribution

J M and X L designed the experiments. B D, S T collected and analyzed the data. J M, S T and X L interpreted the results. J M, B D, S T, and X L wrote the manuscript. X L, P H, S L, and F L edited and revised the manuscript.

#### Conflict of interest

All authors declare no competing interests.

## Appendix A. Tables for the performance of SA and SA\* on synthetic and empirical networks

**Table A1.** Performance of SA, SA\* and other methods on BA networks.

$N$	$m$	$M$	$C$	$\Lambda$	$Q$	$Q_{in}$	$P_{em}$	MAE <sub>PA</sub> (Mean95%CI)	MAE <sub>DMP</sub> Mean(95%CI)	MAE <sub>QMF</sub> Mean(95%CI)	MAE <sub>SA</sub> Mean(95%CI)	MAE <sub>SA*</sub> Mean(95%CI)
1000	2	1996	0.032	4.096	0.523	729	1735	0.029 (0.026,0.032)	0.029 (0.026,0.032)	0.045 (0.042,0.049)	0.037 (0.034,0.04)	<b>0.026</b> (0.023,0.029)
1000	3	2991	0.036	3.481	0.392	3241	5811	0.022 (0.019,0.025)	0.022 (0.019,0.025)	0.030 (0.027,0.033)	0.025 (0.021,0.028)	<b>0.015</b> (0.013,0.018)
1000	4	3984	0.033	3.200	0.326	7937	9196	0.018 (0.015,0.02)	0.018 (0.015,0.02)	0.022 (0.02,0.024)	0.019 (0.016,0.021)	<b>0.011</b> (0.009,0.013)
1000	5	4975	0.038	2.980	0.274	13 804	9754	0.015 (0.013,0.017)	0.015 (0.013,0.017)	0.018 (0.016,0.02)	0.015 (0.013,0.017)	<b>0.010</b> (0.008,0.011)
1500	2	2996	0.019	4.297	0.528	930	2446	0.029 (0.026,0.032)	0.029 (0.026,0.032)	0.047 (0.044,0.05)	0.037 (0.034,0.04)	<b>0.026</b> (0.023,0.029)
1500	3	4491	0.027	3.634	0.394	4425	9039	0.022 (0.02,0.025)	0.022 (0.02,0.025)	0.030 (0.027,0.033)	0.022 (0.02,0.025)	<b>0.013</b> (0.011,0.016)
1500	4	5984	0.027	3.317	0.328	10 768	15 687	0.018 (0.016,0.02)	0.018 (0.016,0.02)	0.022 (0.02,0.025)	0.017 (0.015,0.019)	<b>0.010</b> (0.009,0.012)
1500	5	7475	0.032	3.107	0.276	18 933	19 282	0.017 (0.015,0.019)	0.017 (0.015,0.019)	0.020 (0.017,0.022)	0.016 (0.013,0.018)	<b>0.010</b> (0.008,0.012)
2000	2	3996	0.014	4.426	0.529	1111	3147	0.030 (0.027,0.032)	0.030 (0.027,0.032)	0.048 (0.045,0.051)	0.037 (0.034,0.04)	<b>0.026</b> (0.023,0.028)
2000	3	5991	0.021	3.716	0.389	5338	12 515	0.023 (0.021,0.026)	0.023 (0.021,0.026)	0.031 (0.029,0.034)	0.023 (0.02,0.026)	<b>0.014</b> (0.012,0.015)
2000	4	7984	0.021	3.411	0.325	12 801	22 935	0.019 (0.017,0.021)	0.019 (0.017,0.021)	0.024 (0.021,0.026)	0.018 (0.016,0.021)	<b>0.011</b> (0.01,0.013)
2000	5	9975	0.027	3.186	0.284	23 917	29 113	0.017 (0.015,0.019)	0.017 (0.015,0.019)	0.020 (0.018,0.022)	0.016 (0.014,0.018)	<b>0.010</b> (0.009,0.012)
2500	2	4996	0.014	4.496	0.530	1274	3757	0.030 (0.027,0.033)	0.030 (0.027,0.033)	0.048 (0.045,0.051)	0.037 (0.033,0.04)	<b>0.026</b> (0.023,0.029)
2500	3	7491	0.019	3.796	0.397	6078	15 388	0.024 (0.021,0.027)	0.024 (0.021,0.027)	0.032 (0.029,0.035)	0.023 (0.02,0.026)	<b>0.015</b> (0.013,0.017)
2500	4	9984	0.018	3.482	0.327	14 815	29 478	0.021 (0.019,0.023)	0.021 (0.019,0.023)	0.026 (0.024,0.029)	0.019 (0.017,0.022)	<b>0.011</b> (0.01,0.013)
2500	5	12 475	0.023	3.248	0.283	28 164	40 615	0.016 (0.014,0.018)	0.016 (0.014,0.018)	0.019 (0.017,0.021)	0.013 (0.011,0.016)	<b>0.010</b> (0.009,0.012)
3000	2	5996	0.012	4.567	0.531	1432	4382	0.033 (0.03,0.036)	0.033 (0.03,0.036)	0.051 (0.048,0.054)	0.038 (0.035,0.042)	<b>0.027</b> (0.025,0.031)
3000	3	8991	0.016	3.860	0.396	6789	18 186	0.025 (0.022,0.027)	0.025 (0.022,0.027)	0.033 (0.03,0.036)	0.023 (0.02,0.026)	<b>0.014</b> (0.012,0.016)
3000	4	11 984	0.017	3.544	0.324	16 237	36 473	0.020 (0.017,0.023)	0.020 (0.017,0.023)	0.025 (0.022,0.028)	0.018 (0.015,0.021)	<b>0.013</b> (0.011,0.015)
3000	5	14 975	0.021	3.304	0.282	31 779	52 665	0.018 (0.016,0.019)	0.018 (0.016,0.019)	0.021 (0.019,0.023)	0.015 (0.013,0.016)	<b>0.009</b> (0.008,0.01)

Note: The 95% Confidence Interval (CI) on the mean is calculated through the 100 infection sources.

**Table A2.** Performance of SA, SA\* and other methods on WS networks.

$N$	$m$	$p$	$M$	$C$	$\Lambda$	$Q$	$Q_{ua}$	$P_{en}$	MAE <sub>PA</sub> Mean(95%CI)	MAE <sub>DMP</sub> Mean(95%CI)	MAE <sub>QMF</sub> Mean(95%CI)	MAE <sub>SA</sub> Mean(95%CI)	MAE <sub>SA*</sub> Mean(95%CI)
1000	4	0.01	2000	0.479	23.483	0.889	30	4	0.146 (0.135,0.157)	0.104 (0.101,0.107)	0.129 (0.126,0.132)	0.166 (0.162,0.17)	0.162 (0.158,0.166)
1000	4	0.05	2000	0.423	11.039	0.886	103	3	0.083 (0.081,0.085)	0.083 (0.081,0.085)	0.123 (0.12,0.125)	0.152 (0.149,0.155)	0.146 (0.143,0.148)
1000	4	0.1	2000	0.368	8.659	0.848	181	22	0.068 (0.066,0.07)	0.068 (0.066,0.07)	0.113 (0.111,0.115)	0.136 (0.133,0.138)	0.127 (0.125,0.129)
1000	4	0.5	2000	0.071	5.6	0.584	154	133	0.018 (0.016,0.02)	0.018 (0.016,0.02)	0.068 (0.066,0.07)	0.068 (0.066,0.07)	0.057 (0.055,0.059)
1000	4	0.9	2000	0.002	5.335	0.527	64	214	0.013 (0.012,0.015)	0.013 (0.012,0.015)	0.062 (0.06,0.065)	0.058 (0.056,0.06)	0.048 (0.046,0.05)
1000	6	0.01	3000	0.577	14.684	0.767	9	0	0.094 (0.092,0.096)	0.094 (0.092,0.096)	0.106 (0.104,0.108)	0.151 (0.148,0.153)	0.147 (0.145,0.15)
1000	6	0.05	3000	0.507	7.402	0.768	48	22	0.075 (0.073,0.076)	0.075 (0.073,0.076)	0.094 (0.093,0.096)	0.129 (0.127,0.131)	0.121 (0.119,0.123)
1000	6	0.1	3000	0.444	6.163	0.762	119	49	0.059 (0.057,0.061)	0.059 (0.057,0.061)	0.08 (0.079,0.082)	0.107 (0.105,0.109)	0.096 (0.094,0.098)
1000	6	0.5	3000	0.088	4.283	0.494	622	949	0.016 (0.014,0.018)	0.016 (0.014,0.018)	0.04 (0.038,0.042)	0.031 (0.029,0.033)	0.021 (0.02,0.022)
1000	6	0.9	3000	0.006	4.104	0.39	411	1988	0.012 (0.011,0.013)	0.012 (0.011,0.013)	0.037 (0.035,0.039)	0.024 (0.023,0.025)	0.018 (0.017,0.018)
1000	8	0.01	4000	0.619	10.452	0.713	0	0	0.088 (0.086,0.09)	0.088 (0.086,0.09)	0.096 (0.094,0.098)	0.145 (0.143,0.147)	0.144 (0.141,0.146)
1000	8	0.05	4000	0.542	5.912	0.717	18	74	0.064 (0.062,0.065)	0.064 (0.062,0.065)	0.075 (0.074,0.077)	0.111 (0.109,0.113)	0.104 (0.102,0.106)
1000	8	0.1	4000	0.475	5.069	0.705	133	169	0.049 (0.048,0.051)	0.049 (0.048,0.051)	0.062 (0.06,0.063)	0.089 (0.087,0.09)	0.078 (0.076,0.08)
1000	8	0.5	4000	0.091	3.706	0.437	1592	3374	0.014 (0.013,0.016)	0.014 (0.013,0.016)	0.029 (0.027,0.03)	0.02 (0.019,0.021)	0.015 (0.015,0.016)
1000	8	0.9	4000	0.009	3.571	0.323	1416	6990	0.01 (0.009,0.012)	0.01 (0.009,0.012)	0.025 (0.023,0.026)	0.015 (0.014,0.016)	0.016 (0.015,0.016)
1000	10	0.01	5000	0.643	8.385	0.682	0	2	0.085 (0.083,0.087)	0.085 (0.083,0.087)	0.09 (0.088,0.092)	0.142 (0.139,0.144)	0.143 (0.14,0.145)
1000	10	0.05	5000	0.564	5.09	0.654	33	226	0.059 (0.058,0.061)	0.059 (0.058,0.061)	0.067 (0.066,0.069)	0.105 (0.103,0.107)	0.1 (0.098,0.102)
1000	10	0.1	5000	0.493	4.452	0.636	152	528	0.044 (0.043,0.045)	0.044 (0.043,0.045)	0.052 (0.051,0.053)	0.079 (0.077,0.081)	0.067 (0.065,0.068)
1000	10	0.5	5000	0.094	3.384	0.394	3053	6893	0.012 (0.011,0.013)	0.012 (0.011,0.013)	0.021 (0.02,0.022)	0.015 (0.014,0.016)	0.016 (0.015,0.016)
1000	10	0.9	5000	0.009	3.27	0.279	3659	15162	0.01 (0.009,0.011)	0.01 (0.009,0.011)	0.019 (0.018,0.021)	0.013 (0.012,0.013)	0.017 (0.017,0.017)
1500	4	0.01	3000	0.48	26.681	0.904	44	3	0.186 (0.175,0.197)	0.12 (0.118,0.122)	0.148 (0.146,0.15)	0.189 (0.186,0.192)	0.185 (0.182,0.189)
1500	4	0.05	3000	0.427	12.081	0.897	150	8	0.085 (0.083,0.088)	0.085 (0.083,0.088)	0.125 (0.123,0.128)	0.156 (0.153,0.159)	0.149 (0.146,0.152)
1500	4	0.1	3000	0.367	9.327	0.858	271	19	0.071 (0.069,0.073)	0.071 (0.069,0.073)	0.119 (0.117,0.121)	0.143 (0.141,0.146)	0.135 (0.132,0.137)
1500	4	0.5	3000	0.065	5.933	0.588	214	191	0.02 (0.018,0.022)	0.02 (0.018,0.022)	0.073 (0.071,0.075)	0.073 (0.071,0.076)	0.063 (0.06,0.065)
1500	4	0.9	3000	0.001	5.638	0.524	84	259	0.013 (0.012,0.015)	0.013 (0.012,0.015)	0.065 (0.063,0.067)	0.061 (0.059,0.064)	0.051 (0.049,0.054)
1500	6	0.01	4500	0.578	16.472	0.784	5	14	0.102 (0.099,0.104)	0.102 (0.099,0.104)	0.114 (0.111,0.117)	0.161 (0.158,0.165)	0.158 (0.155,0.162)
1500	6	0.05	4500	0.508	8.095	0.792	68	28	0.075 (0.073,0.077)	0.075 (0.073,0.077)	0.094 (0.092,0.096)	0.129 (0.126,0.131)	0.121 (0.118,0.124)
1500	6	0.1	4500	0.444	6.558	0.763	167	41	0.062 (0.061,0.064)	0.062 (0.061,0.064)	0.085 (0.084,0.087)	0.113 (0.111,0.115)	0.102 (0.1,0.104)
1500	6	0.5	4500	0.079	4.529	0.502	896	1248	0.017 (0.015,0.018)	0.017 (0.015,0.018)	0.044 (0.042,0.046)	0.034 (0.033,0.036)	0.023 (0.022,0.024)
1500	6	0.9	4500	0.005	4.349	0.389	520	2107	0.012 (0.011,0.014)	0.012 (0.011,0.014)	0.039 (0.037,0.041)	0.026 (0.025,0.028)	0.019 (0.018,0.02)
1500	8	0.01	6000	0.62	11.868	0.7	6	18	0.091 (0.089,0.093)	0.091 (0.089,0.093)	0.098 (0.096,0.1)	0.146 (0.144,0.149)	0.145 (0.142,0.147)
1500	8	0.05	6000	0.547	6.453	0.705	36	42	0.069 (0.067,0.071)	0.069 (0.067,0.071)	0.081 (0.08,0.083)	0.12 (0.118,0.122)	0.114 (0.111,0.116)
1500	8	0.1	6000	0.476	5.421	0.723	102	152	0.056 (0.054,0.057)	0.056 (0.054,0.057)	0.069 (0.068,0.071)	0.099 (0.097,0.1)	0.088 (0.085,0.09)
1500	8	0.5	6000	0.087	3.924	0.444	2032	3940	0.014 (0.013,0.016)	0.014 (0.013,0.016)	0.03 (0.029,0.032)	0.021 (0.02,0.023)	0.016 (0.016,0.017)

(Continued.)



Table A2. (Continued.)

<i>N</i>	<i>m</i>	<i>p</i>	<i>M</i>	<i>C</i>	<i>Λ</i>	<i>Q</i>	<i>Q<sub>ua</sub></i>	<i>P<sub>en</sub></i>	MAE <sub>PA</sub> Mean(95%CI)	MAE <sub>DMP</sub> Mean(95%CI)	MAE <sub>QMF</sub> Mean(95%CI)	MAE <sub>SA</sub> Mean(95%CI)	MAE <sub>SA*</sub> Mean(95%CI)
1500	8	0.9	6000	0.007	3.771	0.313	1423	8278	0.01 (0.009,0.011)	0.01 (0.009,0.011)	0.026 (0.024,0.027)	0.016 (0.015,0.017)	0.016 (0.016,0.017)
1500	10	0.01	7500	0.643	9.369	0.661	5	21	0.084 (0.082,0.086)	0.084 (0.082,0.086)	0.089 (0.087,0.091)	0.14 (0.137,0.143)	0.141 (0.138,0.143)
1500	10	0.05	7500	0.567	5.459	0.673	18	136	0.062 (0.06,0.063)	0.062 (0.06,0.063)	0.07 (0.069,0.072)	0.109 (0.108,0.111)	0.104 (0.102,0.106)
1500	10	0.1	7500	0.491	4.725	0.671	122	411	0.047 (0.045,0.048)	0.047 (0.045,0.048)	0.056 (0.054,0.057)	0.084 (0.083,0.086)	0.071 (0.069,0.072)
1500	10	0.5	7500	0.09	3.57	0.398	3564	8605	0.011 (0.01,0.012)	0.011 (0.01,0.012)	0.021 (0.02,0.023)	0.016 (0.015,0.017)	0.015 (0.015,0.016)
1500	10	0.9	7500	0.007	3.455	0.28	3815	19 350	0.009 (0.008,0.01)	0.009 (0.008,0.01)	0.019 (0.018,0.021)	0.013 (0.013,0.014)	0.018 (0.018,0.018)
2000	4	0.01	4000	0.481	29.928	0.935	56	3	0.212 (0.202,0.222)	0.124 (0.121,0.127)	0.152 (0.149,0.155)	0.195 (0.19,0.199)	0.191 (0.186,0.195)
2000	4	0.05	4000	0.423	12.761	0.902	229	21	0.093 (0.09,0.095)	0.093 (0.09,0.095)	0.134 (0.132,0.136)	0.167 (0.164,0.169)	0.16 (0.157,0.163)
2000	4	0.1	4000	0.371	9.924	0.865	326	32	0.075 (0.073,0.077)	0.075 (0.073,0.077)	0.123 (0.121,0.125)	0.149 (0.146,0.151)	0.14 (0.138,0.143)
2000	4	0.5	4000	0.077	6.228	0.597	274	160	0.019 (0.017,0.021)	0.019 (0.017,0.021)	0.076 (0.074,0.077)	0.077 (0.075,0.079)	0.066 (0.064,0.068)
2000	4	0.9	4000	0.003	5.882	0.526	40	203	0.015 (0.013,0.016)	0.015 (0.013,0.016)	0.07 (0.068,0.072)	0.066 (0.064,0.068)	0.056 (0.054,0.058)
2000	6	0.01	6000	0.579	17.576	0.803	5	0	0.109 (0.107,0.111)	0.109 (0.107,0.111)	0.122 (0.12,0.124)	0.173 (0.17,0.175)	0.169 (0.167,0.172)
2000	6	0.05	6000	0.51	8.646	0.782	91	18	0.081 (0.079,0.082)	0.081 (0.079,0.082)	0.101 (0.099,0.103)	0.139 (0.137,0.141)	0.131 (0.129,0.133)
2000	6	0.1	6000	0.445	6.947	0.758	216	60	0.067 (0.065,0.069)	0.067 (0.065,0.069)	0.09 (0.089,0.092)	0.119 (0.118,0.121)	0.109 (0.107,0.11)
2000	6	0.5	6000	0.087	4.721	0.505	1005	1181	0.017 (0.016,0.019)	0.017 (0.016,0.019)	0.046 (0.045,0.048)	0.037 (0.035,0.039)	0.025 (0.023,0.026)
2000	6	0.9	6000	0.003	4.512	0.39	466	2174	0.012 (0.011,0.013)	0.012 (0.011,0.013)	0.041 (0.039,0.043)	0.028 (0.026,0.029)	0.02 (0.019,0.021)
2000	8	0.01	8000	0.621	12.595	0.725	0	4	0.101 (0.099,0.103)	0.101 (0.099,0.103)	0.109 (0.107,0.11)	0.163 (0.161,0.165)	0.161 (0.159,0.164)
2000	8	0.05	8000	0.546	6.718	0.72	22	33	0.072 (0.07,0.073)	0.072 (0.07,0.073)	0.085 (0.083,0.086)	0.125 (0.123,0.127)	0.117 (0.115,0.119)
2000	8	0.1	8000	0.471	5.628	0.717	145	192	0.057 (0.055,0.058)	0.057 (0.055,0.058)	0.071 (0.07,0.072)	0.101 (0.099,0.102)	0.089 (0.087,0.091)
2000	8	0.5	8000	0.086	4.079	0.438	2318	4278	0.014 (0.013,0.016)	0.014 (0.013,0.016)	0.032 (0.031,0.033)	0.023 (0.022,0.024)	0.017 (0.017,0.018)
2000	8	0.9	8000	0.004	3.913	0.316	1589	8848	0.009 (0.008,0.01)	0.009 (0.008,0.01)	0.026 (0.024,0.027)	0.016 (0.015,0.017)	0.017 (0.017,0.017)
2000	10	0.01	10 000	0.644	9.994	0.69	4	0	0.097 (0.096,0.099)	0.098 (0.096,0.099)	0.103 (0.102,0.105)	0.16 (0.158,0.163)	0.161 (0.159,0.164)
2000	10	0.05	10 000	0.565	5.76	0.676	13	157	0.066 (0.065,0.068)	0.066 (0.065,0.068)	0.075 (0.074,0.077)	0.116 (0.114,0.118)	0.111 (0.109,0.113)
2000	10	0.1	10 000	0.488	4.915	0.673	107	385	0.049 (0.048,0.051)	0.049 (0.048,0.051)	0.059 (0.057,0.06)	0.088 (0.087,0.09)	0.074 (0.072,0.076)
2000	10	0.5	10 000	0.091	3.702	0.408	4051	10 198	0.013 (0.012,0.014)	0.013 (0.012,0.014)	0.024 (0.022,0.025)	0.017 (0.017,0.018)	0.016 (0.015,0.016)
2000	10	0.9	10 000	0.006	3.574	0.275	3807	23 139	0.008 (0.007,0.009)	0.008 (0.007,0.009)	0.018 (0.017,0.02)	0.013 (0.013,0.014)	0.018 (0.018,0.019)
2500	4	0.01	5000	0.482	30.573	0.938	71	0	0.221 (0.209,0.234)	0.124 (0.122,0.126)	0.151 (0.149,0.154)	0.192 (0.189,0.196)	0.189 (0.185,0.192)
2500	4	0.05	5000	0.427	13.573	0.911	260	18	0.098 (0.096,0.1)	0.098 (0.096,0.1)	0.141 (0.139,0.143)	0.175 (0.173,0.178)	0.169 (0.166,0.171)
2500	4	0.1	5000	0.373	10.382	0.873	417	46	0.081 (0.079,0.083)	0.081 (0.079,0.083)	0.131 (0.129,0.133)	0.159 (0.157,0.161)	0.15 (0.149,0.152)
2500	4	0.5	5000	0.067	6.428	0.597	406	216	0.021 (0.019,0.023)	0.021 (0.019,0.023)	0.08 (0.078,0.082)	0.082 (0.08,0.084)	0.07 (0.068,0.073)
2500	4	0.9	5000	0.001	6.086	0.529	52	224	0.013 (0.011,0.015)	0.013 (0.011,0.015)	0.071 (0.069,0.073)	0.068 (0.066,0.07)	0.058 (0.056,0.06)
2500	6	0.01	7500	0.579	18.221	0.786	8	0	0.111 (0.109,0.112)	0.111 (0.109,0.112)	0.123 (0.121,0.125)	0.173 (0.17,0.175)	0.169 (0.167,0.172)
2500	6	0.05	7500	0.511	8.809	0.796	94	14	0.086 (0.084,0.088)	0.086 (0.084,0.088)	0.108 (0.106,0.109)	0.147 (0.145,0.149)	0.139 (0.137,0.141)
2500	6	0.1	7500	0.445	7.216	0.787	299	39	0.069 (0.067,0.071)	0.069 (0.067,0.071)	0.093 (0.091,0.095)	0.123 (0.121,0.125)	0.112 (0.11,0.114)

(Continued.)

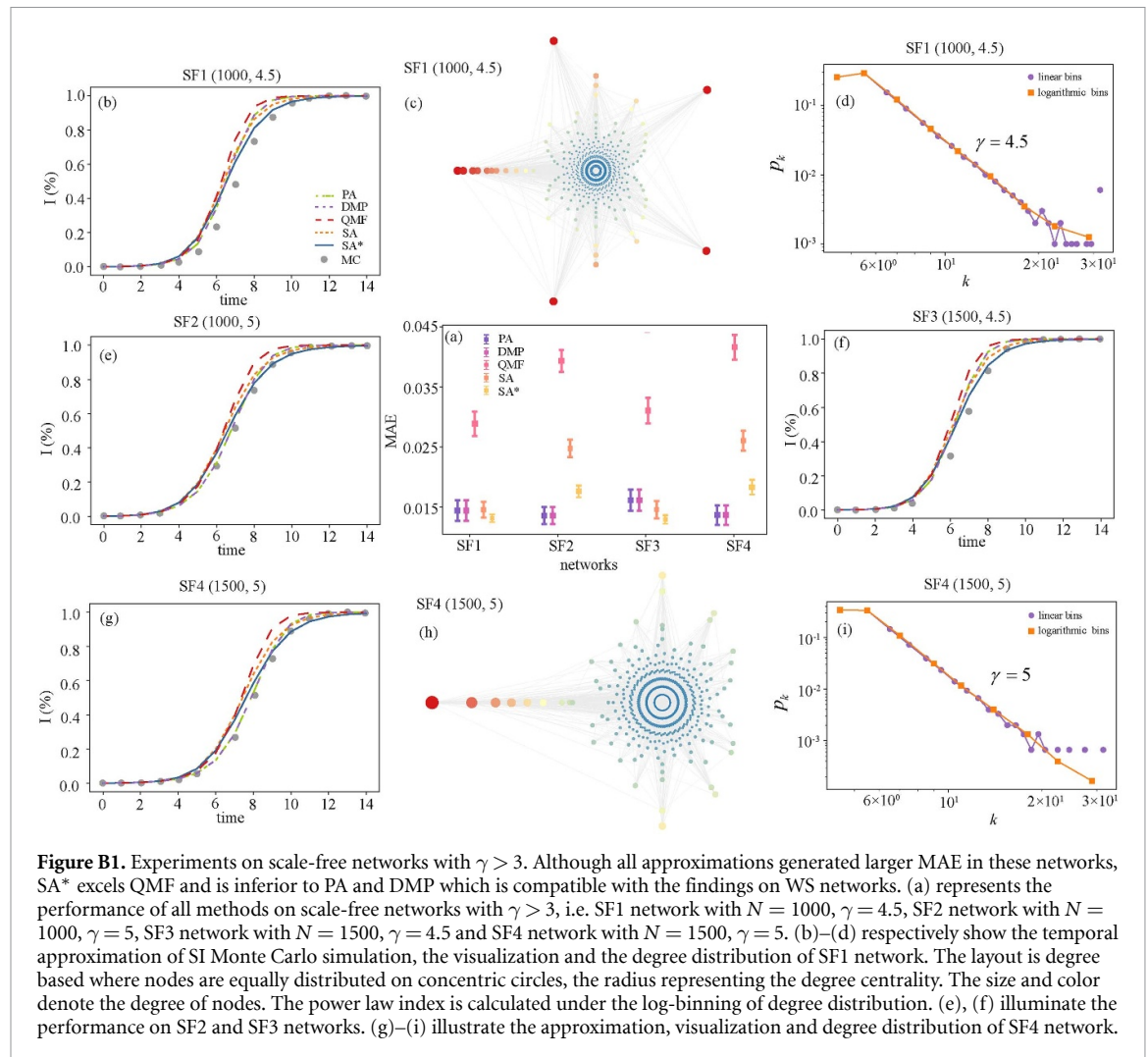
Table A2. (Continued.)

2500	6	0.5	7500	0.08	4.86	0.5	1473	1440	0.017	0.017	0.046	0.037	0.025
									(0.015,0.018)	(0.015,0.018)	(0.045,0.048)	(0.036,0.039)	(0.024,0.026)
2500	6	0.9	7500	0.002	4.644	0.389	379	2314	0.01	0.01	0.04	0.027	0.02
									(0.009,0.012)	(0.009,0.012)	(0.039,0.042)	(0.026,0.028)	(0.019,0.021)
2500	8	0.01	10 000	0.622	13.026	0.726	0	0	0.1	0.1	0.108	0.162	0.161
									(0.099,0.102)	(0.099,0.102)	(0.107,0.11)	(0.16,0.164)	(0.159,0.163)
2500	8	0.05	10 000	0.546	6.949	0.732	36	60	0.076	0.076	0.09	0.131	0.124
									(0.075,0.078)	(0.075,0.078)	(0.088,0.091)	(0.129,0.133)	(0.122,0.126)
2500	8	0.1	10 000	0.474	5.836	0.735	123	205	0.059	0.059	0.073	0.104	0.092
									(0.057,0.06)	(0.057,0.06)	(0.072,0.075)	(0.102,0.105)	(0.09,0.094)
2500	8	0.5	10 000	0.085	4.202	0.446	2851	4283	0.015	0.015	0.033	0.024	0.018
									(0.013,0.016)	(0.013,0.016)	(0.031,0.034)	(0.022,0.025)	(0.017,0.018)
2500	8	0.9	10 000	0.003	4.026	0.315	1447	9278	0.009	0.009	0.027	0.017	0.017
									(0.008,0.011)	(0.008,0.011)	(0.026,0.029)	(0.016,0.018)	(0.017,0.017)
2500	10	0.01	12 500	0.645	10.439	0.678	0	0	0.094	0.094	0.1	0.155	0.155
									(0.093,0.096)	(0.093,0.096)	(0.098,0.102)	(0.152,0.157)	(0.153,0.158)
2500	10	0.05	12 500	0.566	5.951	0.688	68	134	0.068	0.068	0.077	0.119	0.113
									(0.067,0.07)	(0.067,0.07)	(0.076,0.079)	(0.117,0.12)	(0.112,0.115)
2500	10	0.1	12 500	0.492	5.106	0.674	131	521	0.052	0.052	0.061	0.093	0.08
									(0.05,0.053)	(0.05,0.053)	(0.06,0.063)	(0.091,0.095)	(0.078,0.082)
2500	10	0.5	12 500	0.091	3.805	0.404	4637	11 201	0.011	0.011	0.023	0.017	0.016
									(0.01,0.013)	(0.01,0.013)	(0.022,0.024)	(0.016,0.018)	(0.016,0.016)
2500	10	0.9	12 500	0.004	3.665	0.271	3744	25 593	0.008	0.008	0.02	0.014	0.018
									(0.007,0.009)	(0.007,0.009)	(0.019,0.021)	(0.014,0.015)	(0.018,0.019)
3000	4	0.01	6000	0.483	34.807	0.945	76	7	0.257	0.133	0.162	0.206	0.202
									(0.249,0.266)	(0.131,0.136)	(0.159,0.164)	(0.203,0.209)	(0.199,0.205)
3000	4	0.05	6000	0.427	14.085	0.914	307	25	0.089	0.089	0.128	0.159	0.153
									(0.087,0.092)	(0.087,0.092)	(0.124,0.131)	(0.154,0.163)	(0.149,0.157)
3000	4	0.1	6000	0.372	10.66	0.875	514	41	0.08	0.08	0.13	0.158	0.15
									(0.078,0.082)	(0.078,0.082)	(0.128,0.132)	(0.156,0.161)	(0.148,0.152)
3000	4	0.5	6000	0.067	6.57	0.594	385	224	0.02	0.02	0.08	0.082	0.071
									(0.018,0.022)	(0.018,0.022)	(0.078,0.082)	(0.08,0.084)	(0.069,0.074)
3000	4	0.9	6000	0.001	6.238	0.534	79	288	0.014	0.014	0.073	0.071	0.061
									(0.012,0.016)	(0.012,0.016)	(0.07,0.075)	(0.068,0.073)	(0.058,0.063)
3000	6	0.01	9000	0.58	19.799	0.81	11	0	0.115	0.115	0.127	0.18	0.176
									(0.113,0.117)	(0.113,0.117)	(0.125,0.129)	(0.177,0.182)	(0.174,0.179)
3000	6	0.05	9000	0.511	9.114	0.802	131	37	0.089	0.089	0.11	0.15	0.142
									(0.087,0.091)	(0.087,0.091)	(0.108,0.112)	(0.148,0.153)	(0.14,0.145)
3000	6	0.1	9000	0.442	7.289	0.803	425	44	0.069	0.069	0.094	0.123	0.112
									(0.068,0.071)	(0.068,0.071)	(0.092,0.095)	(0.122,0.125)	(0.11,0.114)
3000	6	0.5	9000	0.082	4.978	0.504	1479	1402	0.017	0.017	0.048	0.039	0.026
									(0.016,0.019)	(0.016,0.019)	(0.047,0.05)	(0.037,0.04)	(0.025,0.028)
3000	6	0.9	9000	0.002	4.754	0.388	420	2180	0.011	0.011	0.042	0.028	0.02
									(0.01,0.012)	(0.01,0.012)	(0.04,0.044)	(0.027,0.03)	(0.02,0.021)
3000	8	0.01	12 000	0.622	13.49	0.731	0	4	0.108	0.108	0.117	0.173	0.171
									(0.107,0.11)	(0.107,0.11)	(0.115,0.118)	(0.171,0.175)	(0.169,0.174)
3000	8	0.05	12 000	0.546	7.184	0.726	68	36	0.078	0.078	0.091	0.133	0.126
									(0.077,0.079)	(0.077,0.079)	(0.09,0.093)	(0.132,0.135)	(0.124,0.128)
3000	8	0.1	12 000	0.474	5.982	0.74	104	160	0.06	0.06	0.075	0.106	0.094
									(0.058,0.061)	(0.058,0.061)	(0.073,0.076)	(0.104,0.108)	(0.092,0.096)
3000	8	0.5	12 000	0.087	4.301	0.448	3167	4601	0.015	0.015	0.033	0.024	0.018
									(0.013,0.016)	(0.013,0.016)	(0.032,0.035)	(0.023,0.025)	(0.017,0.018)
3000	8	0.9	12 000	0.003	4.124	0.314	1633	9660	0.01	0.01	0.029	0.018	0.018
									(0.009,0.011)	(0.009,0.011)	(0.027,0.031)	(0.017,0.019)	(0.017,0.018)
3000	10	0.01	15 000	0.645	10.623	0.671	0	10	0.101	0.101	0.107	0.164	0.165
									(0.099,0.102)	(0.099,0.102)	(0.105,0.108)	(0.162,0.166)	(0.162,0.167)
3000	10	0.05	15 000	0.566	6.092	0.716	10	150	0.069	0.069	0.078	0.12	0.114
									(0.067,0.07)	(0.067,0.07)	(0.077,0.079)	(0.119,0.121)	(0.113,0.116)
3000	10	0.1	15 000	0.489	5.217	0.709	131	498	0.054	0.054	0.064	0.096	0.083
									(0.052,0.055)	(0.052,0.055)	(0.062,0.065)	(0.094,0.097)	(0.082,0.085)
3000	10	0.5	15 000	0.089	3.893	0.401	5350	11 925	0.012	0.012	0.024	0.018	0.016
									(0.011,0.013)	(0.011,0.013)	(0.023,0.026)	(0.017,0.019)	(0.016,0.016)
3000	10	0.9	15 000	0.004	3.744	0.27	3907	26 563	0.008	0.008	0.021	0.014	0.018
									(0.007,0.01)	(0.007,0.01)	(0.019,0.022)	(0.014,0.015)	(0.018,0.018)

**Table A3.** Performance of SA, SA\* and other methods on empirical networks.

Networks	MAE <sub>PA</sub> Mean(95%CI)	MAE <sub>DMP</sub> Mean(95%CI)	MAE <sub>QMF</sub> Mean(95%CI)	MAE <sub>SA</sub> Mean(95%CI)	MAE <sub>SA*</sub> Mean(95%CI)
Caenorhabditis	<b>0.017</b> (0.014,0.021)	0.017 (0.014,0.021)	0.019 (0.015,0.022)	0.022 (0.018,0.025)	0.019 (0.016,0.022)
Ego-facebook	0.038 (0.034,0.042)	0.038 (0.034,0.042)	0.059 (0.054,0.064)	0.034 (0.03,0.038)	<b>0.033</b> (0.029,0.038)
Iceland	<b>0.029</b> (0.019,0.039)	0.029 (0.019,0.039)	0.043 (0.031,0.054)	0.046 (0.035,0.056)	0.039 (0.029,0.05)
Interactome_vidal	0.035 (0.03,0.04)	<b>0.035</b> (0.03,0.04)	0.044 (0.039,0.05)	0.047 (0.042,0.053)	0.046 (0.04,0.051)
Karate	<b>0.021</b> (0.012,0.029)	0.022 (0.013,0.03)	0.025 (0.015,0.034)	0.033 (0.023,0.044)	0.023 (0.015,0.032)
Metabolic_ana	<b>0.028</b> (0.023,0.032)	0.028 (0.023,0.032)	0.032 (0.027,0.037)	0.03 (0.026,0.034)	0.029 (0.025,0.033)
Metabolic_blo	<b>0.027</b> (0.023,0.03)	0.027 (0.023,0.03)	0.031 (0.028,0.035)	0.03 (0.027,0.034)	0.028 (0.025,0.032)
Ugandan	0.011 (0.009,0.013)	0.011 (0.009,0.013)	0.013 (0.011,0.015)	0.018 (0.016,0.02)	<b>0.011</b> (0.009,0.013)
Windsurfers	0.013 (0.008,0.018)	0.013 (0.008,0.018)	0.012 (0.007,0.017)	0.018 (0.012,0.025)	<b>0.012</b> (0.007,0.018)
Zebras	<b>0.011</b> (0.003,0.019)	0.011 (0.003,0.019)	0.012 (0.004,0.02)	0.018 (0.012,0.025)	0.013 (0.009,0.017)

**Appendix B. Experiments on scale-free networks with  $\gamma > 3$**



### Appendix C. Tables for statistics of synthetic and empirical networks

**Table C1.** Overview of synthetic network models.

Networks	Size ( $N$ )	Generating parameters
WS	$N \in (1000, 3000, 500)$	$K \in (4, 10, 2), p_r^{WS} \in [0.01, 0.05, 0.1, 0.5, 0.9]$
BA	$N \in (1000, 3000, 500)$	$m \in (2, 5, 1)$
PLC	1000	$m \in (2, 4, 1), p_{\Delta}^{BA} \in (0.1, 0.9, 0.1)$

Note: [\*] denotes the list of elements, (a,b,c) represents the range from a to b with increment c.

**Table C2.** Basic statistics of empirical networks.

Networks	$N$	$M$	$D$	$\langle k \rangle$	$\Lambda$	$C$	$Q$
Caenorhabditis	453	2025	7	8.94	2.66	0.65	0.447
Ego-Facebook	2888	2981	9	2.06	3.87	0.8	0.809
Iceland	75	114	6	3.04	3.16	0.61	0.574
Interactome_vidal	2783	6007	13	4.32	4.84	0.07	0.615
Karate	34	77	5	4.53	2.42	0.54	0.381
Metabolic_ana	1314	3552	10	5.41	3.23	0.16	0.496
Metabolic_blo	813	2103	8	5.17	3.32	0.13	0.511
Ugandan	369	1753	5	9.5	2.8	0.08	0.300
Windsurfers	43	336	3	15.63	1.67	0.65	0.258
Zebras	23	105	4	9.13	1.86	0.85	0.203

### Appendix D. Proofs for Lemmas of spindle vector

**Lemma 1.** The total number of layers  $L_{\max}$  is the diameter of networks  $D$ , i.e.  $L_{\max} = |\vec{S}_p| = D$ .

**Proof 1.** The diameter of networks,  $D$ , is denoted as the maximum distance between any pairs of nodes, i.e.  $D = \max_{ij} d_{ij}$ . The total number of layers depends on the longest BFS spinning tree, characterizing the length of network spindle vector, i.e.  $L_{\max} = |\vec{S}_p| = \max_{ij} d_{ij}$ . Therefore,  $L_{\max} = |\vec{S}_p| = D$ .  $\square$

**Lemma 2.**  $\vec{S}_p(0)$  is the reciprocal of  $N$ , i.e.  $\vec{S}_p(0) = 1/N$ .

**Proof 2.** Each root node  $i$  is viewed as layer 0 in  $\vec{S}_p^i$ , then  $n_0^i = 1/N$ . According to the definition of  $\vec{S}_p(j)$ ,  $\vec{S}_p(0) = \sum_{i \in \Theta(B_0)} \frac{n_0^i}{N} = 1/N$  because layer 0 is reachable to all vertices.  $\square$

**Lemma 3.**  $\vec{S}_p(1)$  is a linear functions of  $\langle k \rangle$ , specifically,  $\vec{S}_p(1) = \langle k \rangle / N$ .

**Proof 3.** The number of vertices in layer 1 is the degree of each root, then  $n_1^i = k_i/N$ . According to the definition of  $\vec{S}_p(j)$ ,  $\vec{S}_p(1) = \sum_{i \in \Theta(B_1)} \frac{n_1^i}{N} = \frac{\sum_{i \in \Theta(B_1)} k_i}{N^2} = \frac{\langle k \rangle}{N}$ .  $\square$

**Lemma 4.** The area under the curve of  $\vec{S}_p$  approximates to  $\Lambda$  of networks, i.e.  $\sum_{j=0}^{L_{\max}} j \vec{S}_p(j) = \frac{N(N-1)}{N^2} \Lambda$ .

**Proof 4.** Firstly,  $Nn_j^i$  means the number of vertices up to a distance of  $j$  away from the root node  $i$ , then  $N \sum_{i \in \Theta(B_j)} n_j^i$  indicates the number of pairs of vertices which are far apart up to a distance of  $j$ . Secondly,  $\sum_{j=0}^{L_{\max}} (jN \sum_{i \in \Theta(B_j)} n_j^i)$  corresponds to the sum of the distance over all connections.  $\Lambda$  is defined as the average distance over all connections, i.e.  $\Lambda = \frac{\sum_{j=0}^{L_{\max}} (jN \sum_{i \in \Theta(B_j)} n_j^i)}{N(N-1)}$ . Finally, for the discrete distribution, the area under the curve is the integration of  $\vec{S}_p$ . According to the definition of  $\vec{S}_p(j)$ ,  $\Lambda = \frac{N^2 \sum_{j=0}^{L_{\max}} j \vec{S}_p(j)}{N(N-1)}$ , or  $\sum_{j=0}^{L_{\max}} j \vec{S}_p(j) = \frac{N(N-1)}{N^2} \Lambda$ .  $\square$

**Lemma 5.** On tree networks with Poisson degree distribution,  $\vec{S}_p(2)$  can be expressed as  $\vec{S}_p(2) = \frac{\langle k^2 \rangle - \langle k \rangle}{N}$ .

**Proof 5.** In [48], authors denote generating functions for the degree distribution and the excess degree distribution of a network respectively, and explicitly give the generating function for the number of neighbors at any distance  $d$ . Especially, if the degree distribution of the tree network were Poisson with mean  $c$ , the mean number of second neighbors equals the minus between second moment and first moment of degree. We extend the claim in spindle vector, and get  $\vec{S}_p(2) = \frac{\langle k^2 \rangle - \langle k \rangle}{N}$ .  $\square$

**Lemma 6.** The sum of  $B_v$  and  $B_e$  over all nodes can both be expressed as the function of the area under the curve of  $\vec{S}_p$ .

$$\sum_{i \in V} B_v(i) = \frac{N}{N-2} (\Lambda - 1) = \frac{N^2}{(N-2)(N-1)} \sum_{j=0}^{L_{\max}} j \vec{S}_p(j) - \frac{N}{N-2} \quad (\text{D.1})$$

$$\sum_{l \in E} B_e(l) = \Lambda = \frac{N^2}{N(N-1)} \sum_{j=0}^{L_{\max}} j \vec{S}_p(j) \quad (\text{D.2})$$

**Proof 6.** In [49], it is strictly proven that the sum of the node betweenness approximately equals  $\frac{N(\Lambda-1)}{(N-2)}$ , and that the sum of the link betweenness approximately equals  $\Lambda$ . Integrated with the approximation of  $\Lambda$  by the area under the curve of  $\vec{S}_p$ , we get lemma 6.  $\square$

## ORCID iDs

Bitao Dai  <https://orcid.org/0000-0002-5346-6352>

Suoyi Tan  <https://orcid.org/0000-0002-9108-2229>

Petter Holme  <https://orcid.org/0000-0003-2156-1096>

Sune Lehmann  <https://orcid.org/0000-0001-6099-2345>

Fredrik Liljeros  <https://orcid.org/0000-0001-8642-6624>

Xin Lu  <https://orcid.org/0000-0002-3547-6493>

## References

- [1] Zhou B, Pei S, Muchnik L, Meng X, Xu X, Sela A, Havlin S and Stanley H E 2020 *Nat. Hum. Behav.* **4** 1–10
- [2] Shupeng G, Xiangfeng D, Lin W, Nicola P and Zhen W 2023 *IEEE Trans. Cybern.* **53** 1–13
- [3] Yang R, Wang B H, Ren J, Bai W J, Shi Z W, Wang W X and Zhou T 2007 *Phys. Lett. A* **364** 189–93
- [4] Lockyer J 2010 *J. Contin. Educ. Health Prof.* **17** 62–64
- [5] Damon C 2010 *Science* **329** 1194–7
- [6] Zhen W, Marko J, Lei S, Joung-Hun L, Yoh I and Stefano B 2018 *Nat. Commun.* **9** 1–7
- [7] Zhen W et al 2020 *Proc. Natl Acad. Sci. USA* **117** 17650–5
- [8] Zhen W, Marko J, Rui-Wu W, Lei S, Yoh I, Yamir M and Jurgen K 2017 *Sci. Adv.* **3** e1601444
- [9] Md S and Saeed K 2021 A survey on modelling of infectious disease spread and control on social contact networks (arXiv:2102.02768)
- [10] Krapivsky P, Redner S and Volovik D 2011 *J. Stat. Mech.* **12** 12003
- [11] Watts D J 2002 *Proc. Natl Acad. Sci. USA* **99** 5766–71
- [12] Nematzadeh A, Ferrara E, Flammini A and Ahn Y Y 2014 *Phys. Rev. Lett.* **113** 088701
- [13] Van Mieghem P, Vespignani A, Pastor-Satorras R, Castellano C 2015 *Rev. Mod. Phys.* **87** 925–79
- [14] Eames K and Keeling M J 2002 *Proc. Natl Acad. Sci. USA* **99** 13330–5
- [15] Biswas M, Paiva L T and Pinho M 2014 *Math. Biosci. Eng.* **11** 761–84
- [16] Korobeinikov A 2004 *Math. Med. Biol.* **21** 75–83
- [17] Kermack W O and McKendrick A G 1927 *Proc. R. Soc. A* **115** 700–21
- [18] Cohen J E 1992 *JAMA* **268** 3381
- [19] Godfrey K 1983 *Compartmental Models and Their Application* (Harcourt Brace Jovanovich)
- [20] Barthelemy M, Barrat A, Pastor-Satorras R and Vespignani A 2003 *Phys. Rev. Lett.* **92** 178701
- [21] Newman M E J 2001 *Phys. Rev. E* **64** 016132
- [22] Ferreira S C, Castellano C and Pastor-Satorras R 2012 *Phys. Rev. E* **86** 041125
- [23] Pastor-Satorras R and Vespignani A 2000 *Phys. Rev. Lett.* **86** 3200–3
- [24] Claudio C and Romualdo P S 2010 *Phys. Rev. Lett.* **105** 218701
- [25] Altarelli F, Braunstein A, Dall Asta L, Wakeling J R and Zecchina R 2014 *Phys. Rev. X* **4** 021024
- [26] Radicchi F and Castellano C 2016 *Phys. Rev. E* **93** 030302
- [27] Cai C R, Wu Z X, Chen M Z Q, Holme P and Guan J Y 2016 *Phys. Rev. Lett.* **116** 258301
- [28] Karrer B and Newman M 2010 *Phys. Rev. E* **82** 016101
- [29] Costa L D F and Andrade R F S 2007 *New J. Phys.* **9** 311–311
- [30] Shang Y 2023 *J. Complex Netw.* **11** 509–12
- [31] Kalisky T, Cohen R, Mokryn O, Dolev D, Shavitt Y and Havlin S 2006 *Phys. Rev. E* **74** 066108
- [32] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 *Nat. Phys.* **6** 888–93
- [33] Schieber T A, Carpi L, Diaz Guilera A, Pardalos P M, Masoller C and Ravetti M G 2017 *Nat. Commun.* **8** 13928
- [34] Wang Z, Zhan X X, Liu C and Zhang Z K 2022 *iScience* **25** 104446
- [35] Bitao D, Shuo Q, Suoyi T, Chuchu L, Jianhong M, Hongzhong D, Fredrik L and Xin L 2023 *J. Comput. Sci.* **69** 102030
- [36] Laura C, Carpi L C, Pardalos P M, Masoller C, Diaz-Guilera A and Ravetti M G 2023 *Nat. Commun.* **14** 2217
- [37] Jiang Y, Li M and Di Z 2022 *Europhys. Lett.* **139** 42003
- [38] Tofighy S, Charkari N M and Ghaderi F 2022 *Physica A* **606** 128043
- [39] Chen G, Zhou S, Li M and Zhang H 2022 *Physica A* **606** 128079
- [40] Shao J, Buldyrev S V, Braunstein L A, Havlin S and Stanley H E 2009 *Phys. Rev. E* **80** 036105
- [41] Catanzaro B M and Pastor-Satorras R 2005 *Phys. Rev. E* **71** 027103
- [42] Freeman L C, Freeman S C and Michaelson A G 1988 *J. Soc. Biol. Syst.* **11** 415–25
- [43] Mcauley J J and Leskovec J 2012 Learning to discover social circles in ego networks 25th *Int. Conf. on Neural Information Processing Systems (NIPS)* vol 1 pp 539–47

- [44] Watts D J and Strogatz S H 1998 *Nature* **393** 440–2
- [45] Barabási A and Albert R 1999 *Science* **286** 509–12
- [46] Holme P and Kim B J 2002 *Phys. Rev. E* **65** 026107
- [47] Wang N N, Wang Y J, Qiu S H and Di Z R 2022 *Commun. Nonlinear Sci. Numer. Simul.* **109** 106260
- [48] Newman M 2010 *Networks: An Introduction* (Oxford University Press)
- [49] Wang H, Wang C, Shu N, Ma C and Yang F 2020 The betweenness identities and their applications *IEEE 20th Int. Conf. on Communication Technology (ICCT)*