

Identify influential social network spreaders

Chung-Yuan Huang

Department of Computer Science and Information
Engineering
Chang Gung University
Taoyuan 333, Taiwan
gscott@mail.cgu.edu.tw

Yu-Hsiang Fu, Chuen-Tsai Sun

Department of Computer Science
National Chiao Tung University
Hsinchu 300, Taiwan

Abstract—Identifying the most influential individuals spreading ideas, information, or infectious diseases is a topic receiving significant attention from network researchers, since such identification can assist or hinder information dissemination, product exposure, or contagious disease detection. Hub nodes, high betweenness nodes, high closeness nodes, and high k-shell nodes have been identified as good initial spreaders. However, few efforts have been made to use node diversity within network structures to measure spreading ability. The two-step framework described in this paper uses a robust and reliable measure that combines global diversity and local features to identify the most influential network nodes. Results from a series of Susceptible-Infected-Recovered (SIR) epidemic simulations indicate that our proposed method performs well and stably in single initial spreader scenarios associated with various complex network datasets.

Keywords—network diversity; entropy; social network analysis; k-shell decomposition; epidemic model

I. INTRODUCTION

The network-spreading phenomenon is the focus of studies ranging from information diffusion via online social media sites, to viral marketing, to epidemic disease identification and control, and to cascading failures in electrical power grids and the Internet, among many others [1-9]. Strategies to identify key spreaders are being established to accelerate information dissemination, increase product exposure, detect contagious disease outbreaks, and execute early intervention strategies [10]. Topological structure is a core concept in this network spreading identification process [1-2].

In social network analyses, centrality measures for identifying influential network nodes are broadly categorized as local or global [3, 7, 11]. Degree centrality, defined as the number of nodes that a focal node is connected to, measures node involvement in a network. However, techniques favored by most researchers for measuring network node influence fail to consider the importance of global topological structures. The two most widely used global centrality measures for overcoming these limitations are betweenness and closeness. Betweenness centrality assesses the degree to which a node lies on the shortest path between two other nodes and determines network flow. Closeness centrality is defined as the inverse sum of the shortest distances from a focal node to all other nodes. Influence is tied to the occupation of advantageous network positions. Three basic advantage sources are high

degree, high closeness, and high betweenness. In simple network structures, these advantages tend to vary individually. In complex networks, the potential exists for considerable disjuncture among these position characteristics, meaning that a spreader's location may be advantageous in some ways and disadvantageous in others.

In addition to centrality measures, results from a k-shell decomposition analysis indicate that network nodes located in core layers are capable of spreading throughout a much broader range than nodes located in peripheral layers [1-2]. Although the spreading capability of each node differs, those with similar k-shell values are perceived as having equal importance. A method for ranking the network spreading ability of nodes in terms of degree centrality in identical k-shell layers for purposes of adjusting rank lists has been proposed [8]. To rank spreaders, a method referred to as mixed degree decomposition (MDD) adds otherwise ignored degree nodes to the decomposition process [3, 6, 12]. Still, researchers have shown a tendency to overlook the importance of network topology and node diversity, despite its positive correlation with factors such as community economic development [13]. The entropy values of locations visited by users have also been shown to be positively correlated with the number of social ties those users have in a social network [14]. The combined entropy values of node degree, betweenness, and closeness centralities have been applied to create complex network visualizations [15, 16].

Inspired by past studies of network topology and node diversity, we used the entropy concept to develop a robust and reliable measure for determining the spreading capability of nodes, and for identifying super-spreader nodes in complex networks. The measure can be used to analyze the numbers of global network topological layers and local neighborhood nodes that are affected by specific individual nodes. Our assumption is that k-shell decomposition [1-2] can be used for purposes of global analysis, with nodes having high degrees of global diversity and local centrality being capable of penetrating multiple global layers and influencing large numbers of neighbors in the local layer of a complex network.

To measure node influence, we propose a two-step framework for acquiring global and local node information within complex networks. In the first step, global node information is obtained using algorithms such as a community detection algorithm for complex networks [5, 17-18] and a k-shell decomposition algorithm for core/periphery network layers, after which entropy is used to evaluate the global

diversity of network nodes. In the second step, local node information is acquired through the use of various types of local centrality, including degree centrality. Last, global diversity and local features are combined to determine node influence. In our experiments, spreading ability was measured as the total number of recovered nodes over time. The spreading capability of the proposed measure and the local/global centralities of the social network analysis were compared using a SIR epidemic simulation [2, 19-20] based on various social network types [21-25].

II. BACKGROUND

To represent a social network, let an undirected graph $G = (V, E)$, where V is the node set and E the edge set of the network. Let $n = |V|$ indicate the number of network nodes and $m = |E|$ the number of edges. Network structure is represented as an adjacency matrix $A = \{a_{ij}\}$ and $a_{ij} \in R^n$, where $a_{ij} = 1$ if a link exists between nodes i and j , otherwise $a_{ij} = 0$.

A. Local centrality

Degree centrality is a simple yet effective method for measuring node influence in a network. Let $C_d(i)$ denote the degree centrality of node i . A high degree centrality indicates a large number of connections between a node and its neighbors. $NB_h(i)$ denotes the set of neighbors of node i at a h -hop distance. The degree centrality of node i is therefore defined as [7]

$$C_d(i) = |NB_h(i)| = \sum_{j=1}^n a_{ij} \quad (1)$$

where $|NB_h(i)|$ is the number of neighbors of node i at the h -hop distance; in most cases, $h=1$.

B. Global centrality

Betweenness centrality measures the proportion of the shortest paths going through a node in a network. Let $C_b(i)$ denote the betweenness centrality of node i . A high betweenness value indicates that a node is located along an important communication path. Accordingly, the betweenness centrality of node i is defined as [3][7][11]

$$C_b(i) = \sum_{s \neq t \neq v \in V} \frac{Q_{st}(i)}{Q_{st}} \quad (2)$$

where $Q_{st}(i)$ is the number of shortest paths from node s to node t through node i , and Q_{st} the total number of shortest paths from node s to node t .

Closeness centrality measures the average length of the shortest paths from one node to other nodes. Let $C_l(i)$ denote the closeness centrality of node i . A high closeness centrality value indicates that a node is located in the center of a network, and that the average distance from that node to other nodes is shorter compared to nodes with low closeness centrality. The closeness centrality of node i is defined as [11]

$$C_b(i) = \frac{1}{l_i}, l_i = \frac{1}{n} \cdot \sum_{j=1}^n d_{ij} \quad (3)$$

where l_i is the average length of the shortest paths from node i to the other nodes, and d_{ij} is the distance from node i to node j .

C. K -shell decomposition

The k -shell decomposition [1][2] iteratively assigns a k -shell layer value to every node in a network. During the first step let $k = 1$, and remove all nodes where $C_d(n) = k = 1$. After removal, the degrees of some remaining network nodes may be $k = 1$. Nodes are continuously pruned from the network until there are no $k = 1$ nodes. All removed nodes are assigned a k -shell value of $ks = 1$. The next step entails a similar process: let $k = 2$, prune nodes, and assign a k -shell value of 2 to all removed nodes. This procedure is continued until all network nodes are removed and assigned a k -shell index. This method exposes the significant features of a network—for example, in the case of the Internet, all nodes can be classified as a nucleus, peer-connected component, or isolated component [1].

D. The SIR epidemic model

The SIR epidemiology model shown in Figure 1 [2][19][20] is widely used in multiple fields to study spreading processes within populations (e.g., information, rumors, and disease epidemics). The model consists of three states: susceptible (S), infective (I), and recovered (R). Nodes in the S set are susceptible to information or a disease, nodes in the I set are capable of infecting neighbors, and nodes in the R set are immune and cannot be reinfected. During the initial step, almost all network nodes are in the susceptible set S ; a small number of infected/infective nodes (sometimes only one) act as spreaders. During each time step, the I nodes infect their neighbors at a pre-established infection rate, after which they become recovered nodes at a recovery rate of γ . Let $S(t)$ denote the number of susceptible nodes at time t , $I(t)$ the

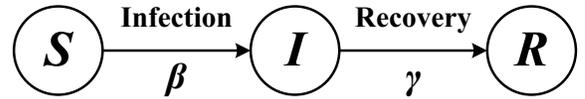


Fig. 1. The SIR (Susceptible-Infected-Recovered) epidemic model.

number of infected nodes at time t , $R(t)$ the number of recovered nodes at time t , and $\rho(t) = R(t)/N$ the proportion of total immunity nodes. The total number of nodes in an SIR model is $S(t) + I(t) + R(t) = n$.

III. THE PROPOSED MEASURE

The two-step framework shown in Figure 2 is proposed as a means of obtaining global and local node information in a network. In step one, global algorithms (e.g., community detection, graph clustering, k -shell decomposition) are used to analyze the global features of nodes in a network. Results are used to compute global node diversity. In step two, the local centrality (e.g., degree centrality) is used to measure local node

features. Last, global diversity and local features are combined to determine the final influence of nodes in the network.

In step one, the k -shell decomposition method was used as an example of network decomposition to obtain global information on nodes in the network. The k -shell values of nodes were obtained to calculate global diversity as a Shannon entropy [26], which was then used to describe how many network layers are affected by a node. For example, a maximum entropy of $\log_2 ks_{max}$ represents a case in which a node is capable of connecting equally with all layers of a network, and a minimum entropy of 0 represents a case in which all connections of a node are to the same layer of a network. As shown in Figure 3, the k -core entropy (also referred to as k -shell entropy) of node i is defined as

$$E_i(X_i) = - \sum_{j=1}^{ks_{max}} p_i(x_j) \cdot \log_2 p_i(x_j) \quad (4)$$

$$p_i(x_j) = \frac{|x_j|}{\sum_{j=1}^{ks_{max}} |x_j|} \quad (5)$$

$$\hat{E}_i(X_i) = \frac{E_i(X_i)}{\log_2 ks_{max}} \quad (6)$$

where $X_i = \{1, 2, \dots, ks_{max}\}$ is the k -shell value of the neighbors of node i , $p_i(x_j)$ the probability of the x_j -core layer of neighbors, $|x_j|$ the number of nodes in the x_j -core layer of the network, and $\hat{E}_i(X_i)$ the normalized k -core entropy for the required case.

In step two, the node's degree centrality is used to analyze the value of local features in the network; the degree centrality of neighbors is also considered. A high influence value indicates that a node and its neighbors have high degree centrality, indicating that the node is capable of reaching the widest possible local range. The local feature of node i is

defined as

$$L_i(i) = \log_2 \left(\sum_{j \in NB_{h=1}(i)} C_d(j) \right) \quad (7)$$

where $C_d(j)$ is the degree centrality of neighbor j , and $NB_{h=1}(i)$ is the neighbor set of node i at a h -hop distance. $L_i(i)$ can be extended to become a "neighbor's neighbor" version, which means that all neighbors of node i with a 2-hop distance are considered.

Finally, global diversity E_i and local feature L_i are combined to denote the final IF_i influence of node i , defined as

$$IF_i = E_i \cdot L_i \quad (8)$$

IV. PRELIMINARY EXPERIMENTAL RESULTS AND DISCUSSION

Results from an analysis of a giant connected component (GCC) and basic network attributes are presented in Table 1. The network dataset classifications used in this study were traditional social, scientific collaboration, and e-mail communication. Measures were degree, betweenness, and closeness centralities; k -shell decomposition; neighbor's core (also known as coreness) [27]; PageRank [28]; and our proposed method. Spreading experiment and SIR model parameters were as follows: 1,000 simulations for each network dataset, with each simulation consisting of 50 time steps, and with the top-1 node for each measure as the initial spreader. The β infection rates of the SIR model used in the experiment are shown in Table 1. According to previous studies, a large infection rate makes no difference in terms of spreading measures [2]. To assign a suitable infection rate for each network dataset, infection rates were determined by comparing the theoretical epidemic threshold β_{tha} to the number used in referenced studies [27]. The recovery rate was always set at $\gamma = 1$, meaning that every node in the infected set

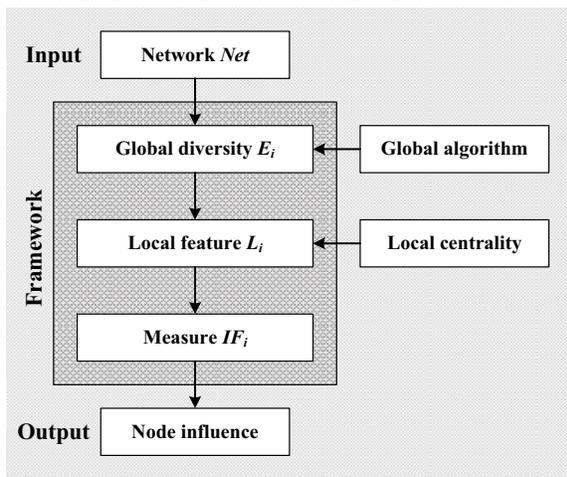


Fig. 2. The proposed two-step framework for computing the influence of network nodes.

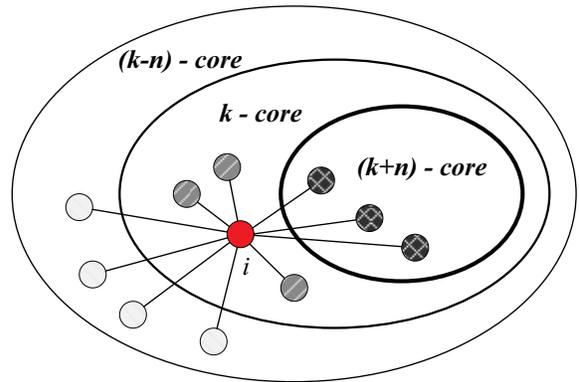


Fig. 3. An illustration of k -core entropy.

TABLE I. DATA FROM AN ANALYSIS OF A GIANT CONNECTED COMPONENT (GCC) AND NETWORK ATTRIBUTES.

Network	N	E	$\langle c \rangle$	k_{max}	$\langle k \rangle$	k_{smax}	$\langle ks \rangle$	H	r	β_{thd}	β
ca_astroph	17903	196972	0.63	504	22.00	56	13.11	2.99	0.20	0.02	0.02
ca_condmat	21363	91286	0.64	279	8.55	25	5.12	2.63	0.13	0.04	0.05
ca_grqc	4158	13422	0.56	81	6.46	43	4.58	2.79	0.64	0.06	0.15
ca_hepph	11204	117619	0.62	491	21.00	238	15.93	6.23	0.63	0.01	0.05
ca_hepth	8638	24806	0.48	65	5.74	31	3.41	2.26	0.24	0.08	0.12
jazz_musicians	198	2742	0.62	100	27.70	29	17.27	1.40	0.02	0.03	0.04
email_contacts	12625	20362	0.11	576	3.23	23	1.65	34.25	-0.39	0.01	0.05
email_enron	33696	180811	0.51	1383	10.73	43	5.73	13.27	-0.12	0.01	0.05
celegansneural	297	2148	0.29	134	14.46	10	7.98	1.80	-0.16	0.04	0.06
dolphins	62	159	0.26	12	5.13	4	3.16	1.33	-0.04	0.15	0.15
lesmis	77	254	0.57	36	6.60	9	4.73	1.83	-0.17	0.08	0.08
netscience	379	914	0.74	34	4.82	8	3.47	1.66	-0.08	0.12	0.20
polblogs	1222	16714	0.32	351	27.36	36	14.82	2.97	-0.22	0.01	0.02

$H = \langle k^2 \rangle / \langle k \rangle^2$, degree heterogeneity [29].

$\beta_{thd} = \langle k \rangle / \langle k^2 \rangle$, theoretical epidemic threshold [30].

I entered the recovered set R immediately after infecting its neighbors.

Preliminary experimental results and detailed data are shown in Figure 4 and Table 2. We found that the leading group LG can be defined as the spreading result of measures that are larger than the maximum result minus an inaccuracy factor of 1%:

$$LG = \left\{ m \mid p_m(t) \geq (p_{max}(t) - err * p_{max}(t)), m \in M \text{ and } err \in [0, 1] \right\}, \quad (9)$$

where M is the set of measures used in the experiment, $p_{max}(t)$ the maximum result at time t , err the inaccuracy rate (0.01), and time step $t = 50$.

The number of recovered nodes $\rho(t)$ was used to measure and rank the spreading capability of various measures. The leading group can help determine how stable a measure is for identifying the influence of nodes in different networks. The measures inside the leading group had approximately the same spreading capability. The average rank in Table 2 was used to place the expected rank in different networks; a measure with a lower average rank was viewed as having better discrimination in terms of identifying good spreaders.

According to the inside leading group number (an indicator of measure stability), the proposed method performed well in terms of identifying the most influential nodes in different networks. Based on our preliminary experimental results, the proposed method is capable of identifying nodes that serve as good spreaders with global diversity in a network. Not only was our proposed method within the leading group, it also had a better ranking compared to other measures within the leading group. The identified influence spreaders were capable of reaching large numbers of network nodes through their diverse global connections and of affecting all network layers. The degree centrality of a node and its neighbors can be used to

maintain the number of contacted nodes in the local layer of a network. However, important differences were noted among measures. For example, the closeness measure performed well in the top-1 position of the ca_hepth and email_enron networks (Fig. 4, Table 2), but not in the ca_grqc, jazz_musician, or netscience networks. Since the characteristic the measure wanted to capture did not exist in those networks, the most influential spreaders could not be identified.

Although the proposed method underscores the robustness and stability of identifying the influence nodes of different networks, we acknowledge two research limitations. First, in cases of global node diversity and lower node degree centrality, the spreading capability of nodes is constrained and dependent on the degree centrality of its neighbors. The influence of a node is limited to the local layer of a network when the degree centrality of its neighbors is lower. The spreading range is also limited when a node's connected neighbors are located in the network's peripheral layer. However, the spreading range of nodes may be wide when the node's neighbors are located near the hub and within the core network layers, and when information and ideas can still be spread to infect a large number of nodes throughout the network.

Second, maximum k -shell values are lower and network sizes considerably smaller in the absence of global diversity in a network. For example, as shown in Table 2, the nodes in the dolphins network could not be identified. The spreading ability of nodes identified by our proposed method declines to the degree centrality (ignoring the first term), and the influence of nodes is limited to the local layers of networks. In the absence of global diversity, equation (8) becomes $IF_i \approx L_i$, which favors the local layers of networks (i.e., degree centrality). The spreading ranges of nodes were also limited to local network layers when nodes were located in peripheral layers or inside local and dense clusters. However, broad spreading ranges were observed for nodes located in the network's core layers [2]. In addition, the $\hat{E}_i(X_i)$ normalized global diversity values

produced by our proposed method were similar to participation coefficients reported by Teitelbaum et al. [31], and the high global diversity of nodes that we observed were similar in terms of role with connector hubs and kinless hubs, both of which have distinct participation coefficients.

V. CONCLUSION

In this paper we described our proposal for a two-step framework for calculating the influence of network nodes. In step one, a global algorithm is used to analyze node global information, with the entropy concept from information theory being introduced to measure node global diversity. Affected global network layers can be identified using k-core entropy. In step two, the degree centrality of nodes and their neighbors are considered simultaneously in order to maintain the number of affected neighbors in the local layer of a network. In the final step, global diversity and local features are combined to determine the influence of nodes in the network. Our preliminary experimental results indicate that the proposed method performs well and maintains stability in the leading groups of different networks. In other words, the proposed method is capable of identifying the most influential nodes as initial spreaders that disseminate information, ideas, or diseases in different networks.

Our plans are to add considerable detail to our analysis and to clarify how the proposed method is affected by network structure. For example, global algorithms such as community detection algorithms can be used to analyze and obtain global information on community network structures, and to determine how factors such as position and node role [31] affect the degree to which spreaders distribute information or diseases throughout a network. We also plan to study strategies associated with multiple initial spreaders in networks. Since overlapping infected areas for selected spreaders must be

minimized [2], a multiple initial spreader scenario may either accelerate or hinder spreading within a network.

REFERENCES

- [1] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 11150–11154, 2007.
- [2] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, Nov. 2010.
- [3] B. Hou, Y. Yao, and D. Liao, "Identifying all-around nodes for spreading dynamics in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 15, pp. 4012–4017, Aug. 2012.
- [4] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1777–1787, Feb. 2012.
- [5] X. Zhang, J. Zhu, Q. Wang, and H. Zhao, "Identifying influential nodes in complex networks with community structure," *Knowledge-Based Systems*, vol. 42, pp. 74–84, Apr. 2013.
- [6] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, Jun. 2013.
- [7] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 403, pp. 130–147, Jun. 2014.
- [8] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 18, pp. 4154–4159, Sep. 2013.
- [9] B. Doerr, M. Fouz, and T. Friedrich, "Why Rumors Spread So Quickly in Social Networks," *Communications of the ACM*, vol. 55, no. 6, pp. 70–75, Jun. 2012.
- [10] N. A. Christakis and J. H. Fowler, "Social Network Sensors for Early Detection of Contagious Outbreaks," *PLoS ONE*, vol. 5, no. 9, p. e12948, Sep. 2010.

TABLE II. SIMULATION RESULTS FROM SPREADING EXPERIMENTS USING DIFFERENT NETWORKS.

Network (GCC)	$\rho(t)$ and $t=50$						
	Degree	Betweenness	Closeness	k-core	Neighbor-core	PageRank	Proposed
ca_astroph	0.1521 ₃	0.1499 ₅	0.1521 ₃	0.1205 ₇	0.1522 ₂	0.1523 ₁	0.1480 ₆
ca_condmat*	0.0486 ₅	0.0487 ₄	0.0480 ₆	0.0278 ₇	0.0520 ₂	0.0488 ₃	0.0527 ₁
ca_grqc*	0.1471 ₂	0.1178 ₅	0.1169 ₆	0.1456 ₄	0.1477 ₁	0.0761 ₇	0.1464 ₃
ca_hepph	0.1953 ₁	0.1940 ₇	0.1952 ₂	0.1951 ₃	0.1950 ₅	0.1951 ₃	0.1943 ₆
ca_hepth*	0.1131 ₃	0.1063 ₅	0.1369 ₁	0.0658 ₇	0.0690 ₆	0.1119 ₄	0.1359 ₂
jazz_musicians	0.3037 ₅	0.3050 ₄	0.3014 ₆	0.2191 ₇	0.3149 ₁	0.3055 ₃	0.3139 ₂
email_contacts*	0.0487 ₆	0.0533 ₅	0.0538 ₂	0.0535 ₄	0.0538 ₂	0.0136 ₇	0.0541 ₁
email_enron	0.1011 ₅	0.1009 ₆	0.1620 ₁	0.1618 ₄	0.1620 ₁	0.1003 ₇	0.1619 ₃
celegansneural	0.1939 ₃	0.1919 ₅	0.1911 ₆	0.0644 ₇	0.1926 ₄	0.2011 ₁	0.1975 ₂
dolphins	0.1107 ₁	0.0754 ₆	0.0795 ₅	0.0702 ₇	0.1019 ₃	0.1089 ₂	0.0897 ₄
lesmis	0.0885 ₄	0.0890 ₃	0.0893 ₂	0.0792 ₆	0.0753 ₇	0.0878 ₅	0.0905 ₁
netscience*	0.0796 ₄	0.0483 ₅	0.0473 ₆	0.0816 ₁	0.0807 ₃	0.0472 ₇	0.0814 ₂
polblogs*	0.1340 ₃	0.0990 ₅	0.1339 ₄	0.0871 ₇	0.1347 ₂	0.0989 ₆	0.1354 ₁
Inside leading group number:	9	7	9	5	11	7	12
Average rank:	3.4615	5.0	3.8461	5.4615	3.0	4.6923	2.6153

*, network result shown in Figure 4.

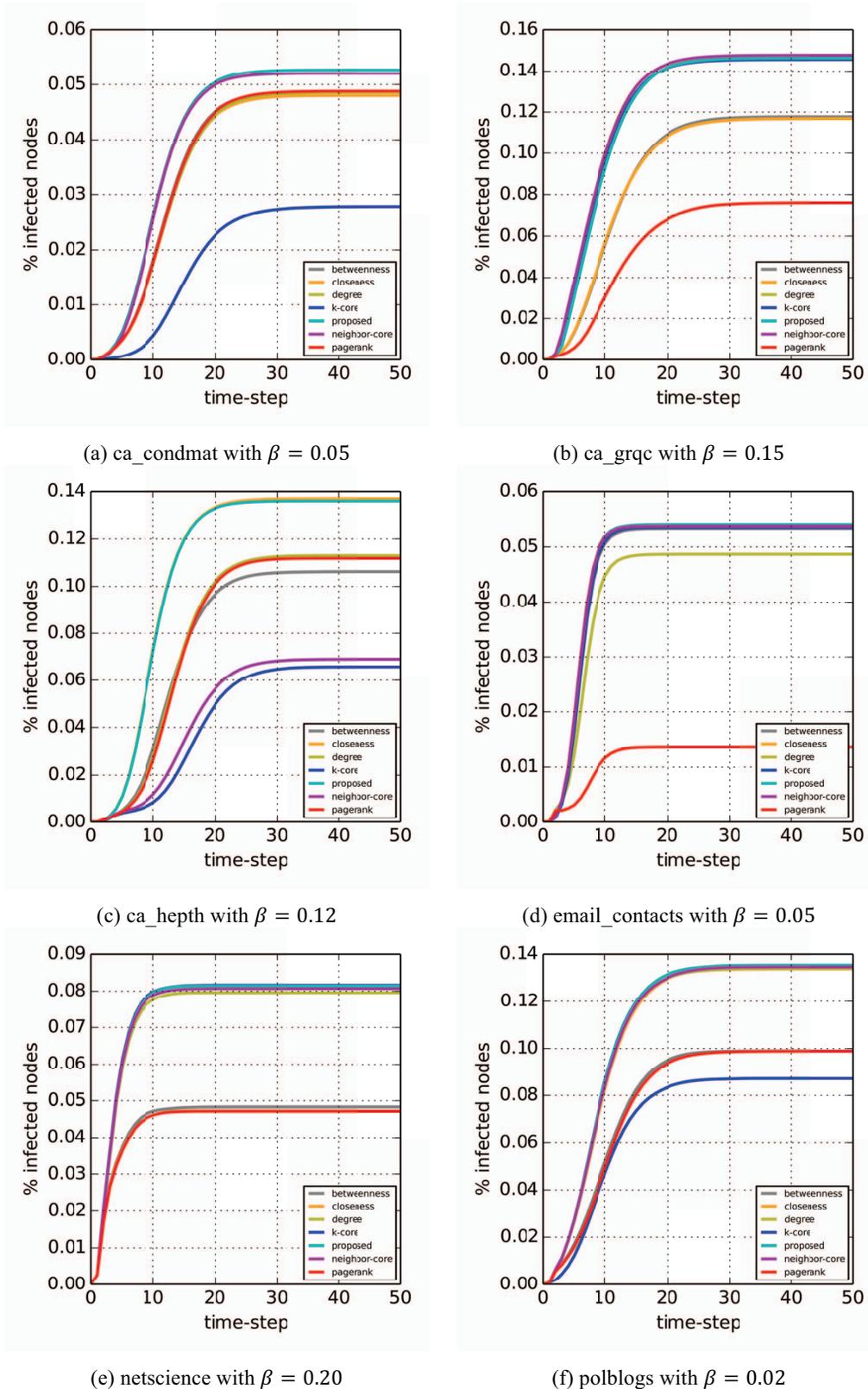


Fig. 4. Spreading dynamic results for different networks.

- [11] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.
- [12] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, Jun. 1938.
- [13] N. Eagle, M. Macy, and R. Claxton, "Network Diversity and Economic Development," *Science*, vol. 328, no. 5981, pp. 1029–1031, May 2010.
- [14] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 119–128.
- [15] E. Serin and S. Balcisoy, "Entropy Based Sensitivity Analysis and Visualization of Social Networks," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2012, pp. 1099–1104.
- [16] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2010.
- [17] M. Rosvall and C. T. Bergstrom, "An information-theoretic framework for resolving community structure in complex networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7327–7331, 2007.
- [18] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
- [19] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics and endemic states in complex networks," *Physical Review E*, vol. 63, no. 6, p. 066117, May 2001.
- [20] C.-Y. Huang, C.-T. Sun, C.-Y. Cheng, and Y.-S. Tsai, "Resource Limitations, Transmission Costs and Critical Thresholds in Scale-free Networks," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2*, Richland, SC, 2008, pp. 1121–1128.
- [21] Stanford large network dataset collection, <http://snap.stanford.edu>.
- [22] Network datasets collected by Mark Newman, <http://www-personal.umich.edu/~mejn/netdata/>.
- [23] Network datasets collected by Hernán Alejandro Makse, http://lisgil.engr.ccnycunyc.edu/~makse/soft_data.html.
- [24] Jazz musicians network dataset, <http://www.infochimps.com/datasets/jazz-musicians-network>.
- [25] P. Basaras, D. Katsaros, and L. Tassiulas, "Detecting Influential Spreaders in Complex, Dynamic Networks," *Computer*, vol. 46, no. 4, pp. 24–29, Apr. 2013.
- [26] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [27] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 549–559, Feb. 2014.
- [28] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web.," Technical Report, Stanford InfoLab, 1999.
- [29] H.-B. Hu and X.-F. Wang, "Unified index to quantifying heterogeneity of complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 14, pp. 3769–3780, Jun. 2008.
- [30] C. Castellano and R. Pastor-Satorras, "Thresholds for Epidemic Spreading in Networks," *Physical Review Letters*, vol. 105, no. 21, p. 218701, Nov. 2010.
- [31] T. Teitelbaum, P. Balenzuela, P. Cano, and J. M. Buldú, "Community structures and role detection in music networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, no. 4, p. 043105, 2008.