Using global diversity and local features to identify influential social network spreaders

Yu-Hsiang Fu
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan, ROC
yuhsiangfu.cs98g@nctu.edu.tw

Chung-Yuan Huang
Department of Computer Science
and Information Engineering
Chang Gung University
Taoyuan, Taiwan, ROC
gscott@mail.cgu.edu.tw

Chuen-Tsai Sun
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan, ROC
ctsun@nctu.edu.tw

Abstract—The identification of influential spreaders of information via social networks can assist in the acceleration or hindrance of information dissemination, in increased product exposure, and in the detection of contagious disease outbreaks. Hub nodes, high betweenness nodes, high closeness nodes, and high *k*-shell nodes have been identified as good initial spreaders. However, researchers have overlooked node diversity within network structures as a means of measuring spreading ability. The two-step framework described in this paper uses a robust and insensitive measure that combines global diversity and local features (e.g., degree centrality) to identify the most influential social network nodes. Preliminary experiment results indicate that the proposed method performs well and maintains stability in single initial spreader scenarios associated with different social network datasets.

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

I. INTRODUCTION

The spreading phenomenon is found in research addressing a wide range of topics; examples include information diffusion via the Internet, viral marketing in business, epidemic disease identification and control, and cascading failures in electrical grids [1][2][3][4][5][6][7][8][9]. Since topological structure plays an important role in network spreading [1][2], identifying the most influential network nodes requires a robust and insensitive measure to spread ideas, information, or diseases as widely as possible. A corresponding strategy to identify spreaders can also be established to accelerate or hinder information dissemination, increase the exposure range of products, detect contagious outbreaks, and support the execution of early intervention strategies [10]. Hence, identifying key spreaders in a network has become an important research issue.

In standard social network analyses, measurements of influence and centrality are categorized as local or global [3], with degree centrality considered a simple yet effective local centrality method for measuring the influence of one or more nodes [7][11]. A node or hub with a high degree of centrality has a large number of connections and high level of network influence. There are two global centralities: betweenness and closeness [3][7][11]. A high level of betweenness indicates that a large number of short paths go through a specific node; a high closeness value indicates a shorter average path between nodes. The most influential nodes in any network are hub, high-betweenness, and high closeness. However, results

from a *k*-shell decomposition analysis indicate that nodes located within the core network layer are capable of spreading throughout a much wider range than nodes located in a network's peripheral layer [1][2].

K-shell values are assigned to all nodes in a network, from core to periphery. Although each node's spreading capability differs, those with similar *k*-shell values are perceived as having equal importance. Researchers have recently proposed new methods for ranking network spreading ability. All-around nodes have been described as having good performance in terms of degree, betweenness, and *k*-shell metrics [3]. An improved method has been proposed for ranking nodes in terms of degree centrality in identical *k*-shell layers to adjust ranking lists [8], and a method referred to as MDD adds ignored degree nodes (also referred to as exhausted links) to the decomposition procedure for purposes of ranking spreaders [6] (see also [3][12] regarding ranking lists).

However, researchers have generally overlooked global node diversity, even though it has a positive correlation with the economic development of communities [13]. The entropy values of locations visited by users has also been shown to be positively correlated with the number of social ties the user has in a social network [14]. The network diversity values of degree, betweenness, and closeness centralities have been combined and applied to create visualizations of social networks [15]. Decision tree algorithms (a classification method used in machine learning in the field of computer science) use entropy to separate data by measuring attributes [16].

The present project, which was inspired by past studies of network diversity, uses entropy to develop a measure and to analyze the spreading capability of a node in a social network. The proposed measure analyzes the number of global layers and local neighborhood nodes that are affected by a node. We assume that *k*-shell decomposition [1][2] can be used as a global analytical method, and that nodes with high degrees of global diversity and centrality can penetrate multiple global layers and influence a large number of neighbors in the local layer of a network.

For purposes of measuring node influence, a two-step framework is proposed to acquire global and local node information within a network. In the first step, global node information is obtained using algorithms such as a community detection algorithm for network community structure [5][17][18], and a *k*-shell decomposition algorithm for core/periphery network layers. Next, entropy is used to evaluate the global diversity of nodes in the network. In step two, local

node information is acquired through the use of various types of local centrality, including degree centrality. Finally, global diversity and local features can be combined to determine node influence. Experiments were conducted to measure spreading ability as the total number of recovered nodes at time t. The spreading capability of the proposed measure and the centralities of the social network analysis were compared using a SIR (susceptible-infective-recovered) model [2][19][20]and a social network dataset [21][22][23][24] with a spreading simulation; a single (top-1) initial spreader scenario [2][25]was also considered as part of our experiments.

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 [7]. 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

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

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 [3][7][11]. 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

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

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 [11]. 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

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

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 is widely used in multiple fields to study spreading processes within populations (e.g., information, rumors, and disease epidemics) [2][19][20]. 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 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 1 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
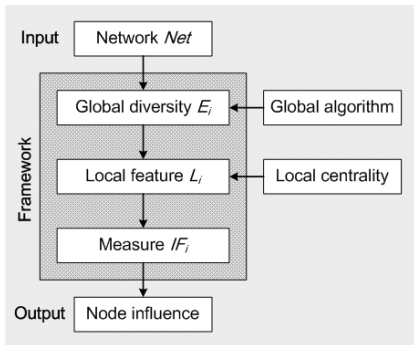
Figure 1. The proposed two-step framework for computing the influence of network nodes.

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. The k-core entropy (also referred to as k-shell entropy) of node i is defined as

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

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

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

where $X_i = \{1, 2, ..., ks_{max}\}$ 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

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

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.

The global diversity E_i of node i makes sure the neighbors' k-shell values which are much more diverse in global layers, and the local feature L_i of node i maintains the most neighbors which can be infected in spreading process. Therefore, the E_i and L_i are considered at the same time to maximize spreading ability of node i in a network. The identified nodes are expected to connect to hub nodes in different k-shells of

network. Finally, E_i and L_i are combined to denote the final IF_i influence of node i, defined as

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

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 collaboration, communication and traditional social. The collaboration network datasets include two collaborating types which are scientific and musician collaboration. The first, scientific collaboration networks are collected form the arXiv website, and research domains cover astro physics (ca_astroph), condense matter physics (ca_condmat), general relativity and quantum cosmology (ca_grqc), high energy physics – phenomenology (ca_hepph) and high energy physics – theory (ca_hepth) [21]. The period of collected data is from January 1993 to April 2003 (124 months). The networks are undirected graph. The nodes are represent as authors, and the edges are relationships of authors who are co-authored in a same paper. The second, musician collaboration network [24], is musical collaboration in Jazz music (jazz_musicians). The nodes are represent as musicians, and the edges are relationships of musicians who are co-recorded.

The communication network datasets are e-mail network and e-mail communication network [21]. These two networks are directed graphs. According to the experimental requirement, directed graphs are transformed to undirected graphs. The first, e-mail network, is collected from an EU research institution (email_contacts). The period of collected data is from October 2003 to May 2005 (18 months). The nodes are email addresses, and the edges are relationships of e-mail addresses which sent at least one message to others. The second, e-mail communication network, is collected from Enron Corporation (email_enron). The nodes and edges are the same as above dataset. The traditional social network dataset are collected from websites of [22] and [23]. The networks include celegansneural, dolphins, lesmis, netscience and polblogs networks which are often used in the fields of complex network and social network analysis.

The measures used in spreading experiment were degree, betweenness, and closeness centralities; k-shell decomposition; neighbor's core (also known as coreness) [27]; PageRank [28]; and our proposed method. These measures are used to quantify importance and spreading ability of nodes in a network. 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_{thd} 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 I

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

Network	N	E	$<c>$	k_{max}	$<k>$	ks_{max}	$<ks>$	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].

entered the recovered set R immediately after infecting its neighbors.

Preliminary experimental results and detailed data are shown in Table 2 and Figure 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%:

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

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. 2, 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

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

*, spreading dynamic result shown in Figure 4; **Bold**, measurement result is inside the leading network group; Subscript, rank of network in the measurement.

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, to clarify how the proposed method is affected by network structure, and to verify whether exists a linear or trade-off relationship between global diversity and local feature. 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.Aa

[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," *PhysicaA: 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.

Figure 2. Spreading dynamic results for different networks. The colored measurements are betweenness (gray), closeness (orange), degree (yellow), k-core (blue), proposed method (cyan), neighbor-core (magenta) and PageRank (red).

[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://lisgi1.engr.ccny.cuny.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.