International Journal of Modern Physics C
Vol. 25, No. 5 (2014) 1440015 (10 pages)

DOI: 10.1142/S0129183114400154

World Scientific
www.worldscientific.com

# Analysis of co-occurrence networks with clique occurrence information

Bin Shen

*Innovation Centre for System Science and Big Data*
*Ningbo Institute of Technology, Zhejiang University*
*Ningbo, 315100, P. R. China*
*tsingbin@zju.edu.cn*

Yixiao Li*

*School of Information*
*Zhejiang University of Finance and Economics*
*Hangzhou, 310018, P. R. China*
*yixiao_li@126.com*

Received 30 June 2013
Accepted 30 September 2013
Published 19 December 2013

Most of co-occurrence networks only record co-occurrence relationships between two entities, and ignore the weights of co-occurrence cliques whose size is bigger than two. However, this ignored information may help us to gain insight into the co-occurrence phenomena of systems. In this paper, we analyze co-occurrence networks with clique occurrence information (CNCI) thoroughly. First, we describe the components of CNCIs and discuss the generation of clique occurrence information. And then, to illustrate the importance and usefulness of clique occurrence information, several metrics, i.e. single occurrence rate, average size of maximal co-occurrence cliques and four types of co-occurrence coefficients etc., are given. Moreover, some applications, such as combining co-occurrence frequency with structure-oriented centrality measures, are also discussed.

*Keywords*: Co-occurrence networks; co-occurrence coefficient; metrics and applications.

PACS Nos.: 89.75.Fb, 89.20.Ff.



## 1. Introduction

Co-occurrence networks have been recognized very important for discovering previously unrevealed insights.[1–3] They have been widely studied in various domains, such as linguistic units analysis,[1] gene associations,[2] oriental medicine,[3] co-authorship analysis,[4,5] promotion-group design and so on. Especially in the incoming big data era, since it is usually hard to find cause-and-effect relationships between

*Corresponding author.

entities, discovering interesting co-occurrence relationships from big data becomes a shining and valuable replaceable option.[6] For example, Toivonen *et al.*[7] studied the similarity network of Finnish emotion concepts, and identified meaningful clusters and specific local network structures from the network. Yang *et al.*[3] reported a kind of symptom-herbal material bipartite network for oriental medicine analysis.

However, in the process of modeling and analysis, most of co-occurrence networks[1–5,7] currently only record co-occurrence relationships between two entities, and two other important types of clique co-occurrence information have not been paid enough attention yet: (i) the weights about co-occurrence of cliques whose size is bigger than two and (ii) the information about maximal co-occurrence cliques. These ignored data may help us to understand the co-occurrence phenomena of systems better. For example, when constructing a coauthorship network, two scientists are connected if they have coauthored one or more papers together.[4,5] Although edge weights can be directly the number of times that a partnership has been repeated for the corresponding two scientists,[8,9] the information of coauthorship of three or more scientists, sole authorship and maximal cliques of coauthorship are still missing in such cases.

Since co-occurrence networks are often derived from bipartite networks,[10] this clique-co-occurrence information missing problem also occurs in the process of bipartite network projection.[10,11] Figures 1(a) and 1(b) show an example of bipartite graph and its one-mode projection in entity set. We can find that the information of co-occurrence of cliques whose sizes are not less than three (such as clique $\{y_1, y_2, y_3\}$), and single occurrences (such as $\{y_6\}$) are discarded after one-mode projection. Maximal co-occurrence cliques (e.g. $\{y_1, y_3, y_4, y_5\}$) with repeat times are not distinguishable after projection. The discarded or undistinguishable information provides clues for group occurrences and will be very informative in various co-occurrence applications.

Motivated by this concern, in this paper, we highlight co-occurrence networks accompanied by clique occurrence information, a technique which allow us to analyse co-occurrence relationships informatively and easily for familiar co-occurrence unipartite networks. Several metrics and applications are illustrated to show the usefulness of clique co-occurrence information.

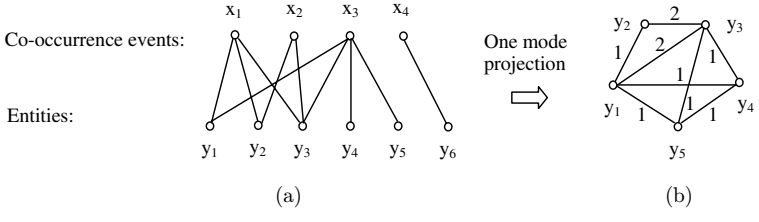



Fig. 1. Illustration of a bipartite network (a) and its one-mode projection in entity set (b). Edge weights are set as the number of co-occurrence in co-occurrence event set.

## 2. Co-occurrence Networks with Clique Occurrence Information

### 2.1. *Description and generation of CNCIs*

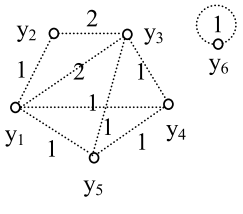
A co-occurrence networks with clique occurrence information (CNCI) includes three correlated parts: a unipartite graph $G$, a set of co-occurrence cliques (whose sizes are no more than 3) with weights (i.e. $B$) and a table containing all maximal co-occurrence cliques with weights (i.e. $C$). Here, only cliques whose sizes are no more than 3 are listed in $B$. This is because the dataset containing all cliques is normally gigantic even for a not very big network. For instance, suppose $n$ entities occur together in a same event, a fully connected network with $n$ vertices can be established. Then the total number of cliques reaches $2^n - 1$, and it will increase exponentially with the increase of $n$. So, it is commonly infeasible to list all cliques with their weights in $B$ for a large network. Instead, considering extensive use of co-occurrence information about cliques whose sizes are 1, 2 and 3, we only provide these cliques with their occurrence weights in $B$.[a] An illustration of CNCI is given in Fig. 2.

The generation of $B$ and $C$ is straightforward. $G$ can be drawn using a modified one-mode projection method, which is the same to the standard one-mode projection[10,11] except that self-loops are drawn and weights of cliques (or maximal cliques) are assigned.[b]

### 2.2. *Data*

For numerical analysis, four representative co-occurrence networks accompanied by clique occurrence information are studied using data drawn from three disparate fields: (I) Actor-Top1K — An actor collaboration CNCI whose vertices are top 1000 among frequent vertices of actor dataset. The actor data is downloaded from the

*G*




*B*

| *clq* | $w_B$ | *clq* | $w_B$ | *clq* | $w_B$ |
|---|---|---|---|---|---|
| $y_1$ | 2 | $y_1,y_3$ | 2 | $y_1,y_2,y_3$ | 1 |
| $y_2$ | 2 | $y_1,y_4$ | 1 | $y_1,y_3,y_4$ | 1 |
| $y_3$ | 3 | $y_1,y_5$ | 1 | $y_1,y_3,y_5$ | 1 |
| $y_4$ | 1 | $y_2,y_3$ | 2 | $y_1,y_4,y_5$ | 1 |
| $y_5$ | 1 | $y_3,y_4$ | 1 | $y_3,y_4,y_5$ | 1 |
| $y_6$ | 1 | $y_4,y_5$ | 1 | | |
| $y_1,y_2$ | 1 | $y_3,y_5$ | 1 | | |

*C*

| *clq* | $w_c$ |
|---|---|
| $y_1,y_2,y_3$ | 1 |
| $y_2,y_3$ | 1 |
| $y_1,y_3,y_4,y_5$ | 1 |
| $y_6$ | 1 |

Fig. 2. An illustration of the CNCI generated from a bipartite graph in Fig. 1(a). Clique weights are set as the numbers of co-occurrence in co-occurrence event set. *clq* means clique, and $w_B$ and $w_C$ represents the corresponding weight, respectively.

[a] Other cliques with their occurrence weights still can be computed with demands.
[b] These weights are not shown in the unipartite graph of Fig. 2 for clarity. But they can be displayed using a proper visualization method in an interactive software.

Table 1. The basic topological features and research results of five example networks. $N$ is the number of co-occurrence events. $V$ and $L$ are the number of vertices and edges, respectively. $M_1$, $M_2$ and $M_3$ are the number of co-occurrence cliques whose sizes are 1, 2 and 3, respectively. Here, $M_1$ is exactly $V$ and $M_2$ is equal to $L$. The number of cliques in $B$ is the sum of $M_1$, $M_2$ and $M_3$. $M_C$ is the total number of maximal co-occurrence cliques in $C$. $R$ is the average rate of single occurrence. $S$ and $S_w$ are the average size and the weighted average size of maximal co-occurrence cliques, respectively. CC is the clustering coefficient. $CC_1$, $CC_2$, $CC_3$ and $CC_4$ are type I, type II, type III and type IV co-occurrence coefficients, respectively.

| | Actor-Top1K | HEP-PH-9 | HEP-TH-2K | BMS-POS |
|---|---|---|---|---|
| $N$ | 46 464 | 3 240 | 2 983 | 515 597 |
| $V(M_1)$ | 1 000 | 14 600 | 9 689 | 1 657 |
| $L(M_2)$ | 132 662 | 531 031 | 481 701 | 379 387 |
| $M_3$ | 1 102 101 | 14 639 955 | 13 199 742 | 22 005 568 |
| $M_C$ | 24 167 | 3 216 | 2 965 | 320 285 |
| $R$ | 0.165 | 0.009 | 0.010 | 0.011 |
| $S$ | 4.51 | 13.20 | 17.08 | 9.30 |
| $S_w$ | 2.91 | 13.11 | 16.99 | 6.53 |
| CC | 0.614 | 0.457 | 0.300 | 0.644 |
| $CC_1$ | 0.109 | 0.941 | 0.790 | 0.348 |
| $CC_2$ | 0.075 | 0.937 | 0.769 | 0.147 |
| $CC_3$ | 0.067 | 0.430 | 0.237 | 0.224 |
| $CC_4$ | 0.040 | 0.423 | 0.222 | 0.074 |

website of Barabási Lab (www.barabasilab.com). (II) HEP-PH-9 — A HEP-PH (high energy physics phenomenology) co-citation CNCI drawn from the dataset of "hep-ph part 9," which is one part of the dataset of HEP-PH papers in the e-print arXiv running from 1993 to 2003. The data was originally released by KDD Cup 2003 (www.sigkdd.org). (III) HEP-TH-2K — A HEP-TH (high energy physics theory) co-citation CNCI drawn from hep-th 2000 dataset, which is all papers in the HEP-TH portion of the e-print arXiv in the year 2000. The data was also downloaded from the website of KDD cup 2003. (IV) BMS-POS — A product co-purchasing CNCI containing 1 657 products. The data is available at the website of KDD Cup 2000 (www.sigkdd.org).

In Table 1, we give a summary of some of the basic topological features and experiment results for the networks studied here. The theories for these new metrics are discussed in detail in Sec. 3.

## 3. Some Metrics Based on Clique Occurrence Information

Currently, most of metrics for evaluating topology properties of complex networks focus on statistical quantities between two vertices, and few work has been done towards developing clique-occurrence-based statistical metrics. Actually, clique occurrence information and the topology of the connections can be incorporated into

various metrics naturally for measuring complex networks. In the following, we give some examples.

### 3.1. *Single occurrences of vertices*

Single occurrence rate reflects the possibility of single occurrence for each entity. For each vertex $y_j (j \in 1, 2, \ldots, m)$, its corresponding single occurrence rate $r_{y_j}$ is written as below.

$$r(y_j) = \frac{w_C(y_j)}{w_B(y_j)}, \tag{1}$$

where $w_C(y_j)$ and $w_B(y_j)$ are the corresponding weight of $y_j$ in $C$ and $B$, respectively. Thus, the average rate of single occurrence is defined as below.

$$R = \frac{\sum_j r(y_j)}{m}, \tag{2}$$

where $m$ is the total number of vertices.

### 3.2. *Sizes of maximal co-occurrence cliques*

Average size of maximal co-occurrence cliques (i.e. $S$) and weighted average size of maximal co-occurrence cliques (i.e. $S_w$) reflect the scale of co-occurrence in a complex network, and can be written as below.

$$S = \frac{\sum_{\forall \phi} \text{size}(\phi)}{num_C}, \tag{3}$$

$$S_w = \frac{\sum_{\forall \phi} w_C(\phi) \times \text{size}(\phi)}{\sum_{\forall \phi} w_C(\phi)}, \tag{4}$$

where function size($\phi$) returns the size of maximal co-occurrence clique $\phi$, $w_C(\phi)$ is the weight of $\phi$ in $C$ and $num_C$ is the size of $C$.

Based on the sizes of maximal co-occurrence cliques, the distribution of these sizes can be obtained easily.

### 3.3. *Co-occurrence coefficients*

The concept of clustering coefficient has already been proposed to capture the probability that two people will be acquainted, if they have another acquaintance in common.[4,12,13] The global clustering coefficient is defined as the number of closed triplets (or 3 $\times$triangles) over the total number of triplets (both open and closed).[c] Similarly, we propose four types of co-occurrence coefficients to reflect the probability that three entities will occur together under different conditions as discussed below.

[c] Wikipedia, http://en.wikipedia.org/wiki/Clustering_coefficient (2013).

#### 3.3.1. *Type I and type II co-occurrence coefficients*

In a CNCI, both type I and type II co-occurrence coefficients reflect the probability that "triangles"[d] of vertices occur together. The difference is that the former metric only considers whether vertices in a "triangle" occur together or not, while the latter also take the frequency of co-occurrences into consideration. The formal definitions of these two metrics are discussed as below.

Suppose $e1$, $e2$ and $e3$ are three edges of a "triangle" (i.e. $\delta$). Type I co-occurrence coefficient (i.e. $\mathrm{CC}_1$) and type II co-occurrence coefficient (i.e. $\mathrm{CC}_2$) for $\delta$ are defined as below.

$$\mathrm{CC}_1(\delta) = \begin{cases} 1 & w_B(\delta) \text{ is nonzero,} \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$\mathrm{CC}_2(\delta) = \frac{w_B(\delta)}{\min(w_B(e1), w_B(e2), w_B(e3))}, \tag{6}$$

where $w_B()$ returns the corresponding weight in $B$ for the parameter and function min() returns the minimal one of its three parameters. Obviously, both $\mathrm{CC}_1(\delta)$ and $\mathrm{CC}_2(\delta)$ satisfy $[0, 1]$.

And then, the global co-occurrence coefficients of type I and type II can be obtained using the following formulas.

$$\mathrm{CC}_1 = \frac{\sum_{\forall\delta} \mathrm{CC}_1(\delta)}{N_\delta}, \tag{7}$$

$$\mathrm{CC}_2 = \frac{\sum_{\forall\delta} \mathrm{CC}_2(\delta)}{N_\delta}, \tag{8}$$

where $N_\delta$ is the number of "triangles" in the CNCI. It is easy to infer that $\mathrm{CC}_1 \in [0, 1]$ and $\mathrm{CC}_2 \in [0, 1]$.

For instance, there are five "triangles" in the CNCI as illustrated in Fig. 2. For "triangle" $\delta = \{y_1, y_2, y_3\}$, $\mathrm{CC}_1(\delta)$ is 1 and $\mathrm{CC}_2(\delta)$ is computed as $1/\min\{1, 2, 2\} = 1$. Similarly, $\mathrm{CC}_1(\delta)$ and $\mathrm{CC}_2(\delta)$ for the other "triangles" also can be obtained. Thus, we have the global $\mathrm{CC}_1$ and $\mathrm{CC}_2$, both of which are 1.

#### 3.3.2. *Type III and type IV co-occurrence coefficients*

In comparison with type I and type II co-occurrence coefficients, type III and type IV co-occurrence coefficients focus on analyzing connected triplets of vertices,[e] which are connected by two (open triplet) or three (closed triplet) edges. Both of them reflect the probability that connected triplets of vertices occur together. The difference between type III and type IV is the same to that between type I and type II. That is to say, type III co-occurrence coefficient is only concerned about whether connected

[d] A triplet of vertices is called a triangle, if three vertices are connected with each other by three edges.
[e] Clustering coefficient is also based on analyzing connected triplets of vertices.

triplets of vertices occur together, while type IV co-occurrence coefficient also utilizes the frequency of co-occurrences to estimate that probability.

Suppose a connected triplet (i.e. $\tau$) has two edges (i.e. "$y_1y_2$" and "$y_2y_3$") and this triplet is also marked as "$y_1$-$y_2$-$y_3$", where $y_2$ is the connecting vertex. First, we define type III co-occurrence coefficient (i.e. $\mathrm{CC}_3$) and type IV co-occurrence coefficient ($\mathrm{CC}_4$) for $\tau$ as below.

$$\mathrm{CC}_3(\tau) = \begin{cases} 1 & w_B(\tau) \text{ is nonzero,} \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

$$\mathrm{CC}_4(\tau) = \frac{w_B(\tau)}{\min(w_B(y_1y_2), w_B(y_2y_3))}, \tag{10}$$

where $w_B()$ is the same function as the above, function min() returns the smaller of the two parameters. Obviously, $\mathrm{CC}_3(\tau)$ and $\mathrm{CC}_4(\tau)$ is in the range of $[0, 1]$.

Then, based on $\mathrm{CC}_3(\tau)$ and $\mathrm{CC}_4(\tau)$, the global $\mathrm{CC}_3$ and $\mathrm{CC}_4$ can be written as follows.

$$\mathrm{CC}_3 = \frac{\sum_{\forall \tau} \mathrm{CC}_3(\tau)}{N_\tau}, \tag{11}$$

$$\mathrm{CC}_4 = \frac{\sum_{\forall \tau} \mathrm{CC}_4(\tau)}{N_\tau}, \tag{12}$$

where $N_\tau$ is the total number of connected triplets in the CNCI. It is easy to infer that $\mathrm{CC}_3 \in [0, 1]$ and $\mathrm{CC}_4 \in [0, 1]$.

It is necessary to mention that one "triangle" has three connected triplets from the views of different connecting vertices. Still take the CNCI illustrated in Fig. 2 as an example. There are 19 connected triplets, which include four open triplets and 15 closed ones. For "$y_1$-$y_3$-$y_2$," because the occurrence number of clique $\{y_1, y_3, y_2\}$ is nonzero, we have that its $\mathrm{CC}_3$ equals 1 and its $\mathrm{CC}_4$ is $1/\min(2, 2) = 0.5$. Then, based on values of $\mathrm{CC}_3$ and $\mathrm{CC}_4$ of these 19 connected triplets, the global $\mathrm{CC}_3$ and $\mathrm{CC}_4$ are 0.789 and 0.763, respectively.

## 4. Applications

Clique occurrence information provides additional useful information for co-occurrence networks, and can be applied in various applications. Some examples are given below.

### 4.1. *Combination of frequency metric and structure-oriented centrality measures*

Currently, much work has been done towards finding interesting vertices (or groups) in a complex network, such as leaders,[14] influential spreaders,[15] prominent groups[16] and so on. However, most of them has not taken co-occurrence frequency into consideration. In comparison with other commonly used metrics for measuring the

importance of a vertex (or a clique) in complex networks, co-occurrence frequency is a nonstructure-oriented measurement. So, it is very interesting to find meaningful vertices (or cliques) satisfying both co-occurrence frequency interest measure and popular structure-oriented centrality measures (such as degree centrality, betweenness and closeness etc.).

In the following, we give two illustrations for such type of combinations.

#### 4.1.1. *Frequent vertices with high degree centrality*

For a vertex, its frequency, degree[17] and weighted degree[18] are correlated but different. The frequency registers the occurrence times, while the degree reflects the number of neighboring vertices, and the weighted degree is the sum of the weights attached to the edges connected to the vertex. For example, in Fig. 2, the values of frequency, degree and weighted degree for vertex $y_1$ are 2, 4 and 5, respectively. Since these three measures are indicators of the importance of a vertex from the correlated aspects, we attempt to use three tuning parameters (i.e. $\alpha, \beta$ and $\gamma$) to combine them together. Thus, we have the following definition.

For each vertex $y_j$, its combined importance measure is given as below.

$$\mathrm{Cob}(y_j) = \alpha \times f(y_i) + \beta \times k(y_i) + \gamma \times w(y_i), \tag{13}$$

where $f(y_i), k(y_i)$ and $w(y_i)$ are the frequency, the degree and the weighted degree of $y_j$ respectively, tuning parameters $\alpha, \beta, \gamma \in [0, 1]$ and $\alpha + \beta + \gamma = 1$. If $\mathrm{Cob}(y_j)$ is not less than the threshold $\delta_{\mathrm{Cob}}$, we call that $y_j$ is a frequent vertex with high degree centrality.

#### 4.1.2. *Frequent co-occurrence cliques with high clique betweenness*

The combination of occurrence frequency with centrality measures for cliques will enable researchers to answer such questions as "which cliques not only have a high group betweenness, but also occur frequently in the complex networks research community?" Since occurrence frequency and group betweenness[19,20] describe the importance of cliques from different aspects, we do not use a combined measure to combine them directly but set two minimal thresholds for them.

Given the threshold of occurrence frequency $\delta_f$ and the threshold of group betweenness $\delta_{\mathrm{gb}}$, for any clique $\phi$, if its occurrence frequency $f(\phi)$ and group betweenness $\mathrm{gb}(\phi)$ are not less than $\delta_f$ and $\delta_{\mathrm{gb}}$ respectively, $\phi$ is called a frequent co-occurrence clique with high clique betweenness.

### 4.2. *Applying association rules to complex networks*

The method of association rules[21] is one of the well-researched techniques in the field of data mining. An association rule has the form "$X \Rightarrow Y$," which means that the occurrence of item set $X$ (the antecedent) implies the presence of item set $Y$ (the consequent). The strength of rules is commonly characterized by two measures: the support and the confidence. The support of a rule is defined as co-occurrence

times that both $X$ and $Y$ occur. That is to say $\text{supp}(\text{“}X \Rightarrow Y\text{”}) = \text{supp}(X \cup Y)$. The confidence of a rule is an estimate of the probability $P(Y|X)$, which equals $\text{supp}(X \cup Y)/\text{supp}(X)$.

From the above description, we can find that association rules actually are a kind of connections between the antecedent and the consequent. If each item is regarded as a vertex, these connections are not only established between two vertices, but also between two groups of vertices. So, associations rule can be a useful tool for analyzing connections between groups of vertices in complex network studies. Potential applications may exist in personal recommendation, missing links prediction and discovering interesting co-occurrence relationships etc.

For $X \cup Y$ whose size is no more than 3, since $w_B(X \cup Y)$ and $w_B(X)$ registered in $B$ of the CNCI are exactly $\text{supp}(X \cup Y)$ and $\text{supp}(X)$, association rules can be produced directly using generated clique co-occurrence information. Still take the illustration in Fig. 2 as an example. If we want to generate an association rule between $y_2$ and $y_3$, because $w_B(\{y_2, y_3\}) = 2$ and $w_B(\{y_2\}) = 2$, we have “$y_2 \Rightarrow y_3$ [2, 100%],” where 2 and 100% are the support and the confidence of this rule, respectively. If the size of $X \cup Y$ is bigger than 3, mature association rule mining methods[21] can be adopted to generate connections between two groups of vertices.

## 5. Conclusions and Discussions

Clique occurrence information is quite helpful for gaining insight into the co-occurrence phenomena of systems. However, it is often ignored in analyzing co-occurrence networks. In this study, we highlight CNCI and focus on how to exploit clique occurrence information.

A CNCI is composed of three components: a unipartite graph, a table containing co-occurrence cliques (whose sizes are no more than 3) with weights and a table of maximal co-occurrence cliques with weights. In order to illustrate the usefulness of clique occurrence information, we propose several metrics: single occurrence rate, average rate of single occurrence, weighted and unweighted average size of maximal co-occurrence cliques and four types of co-occurrence coefficients. These metrics combine clique occurrence information and the topology of the connections naturally to evaluating statistical properties of complex networks. Furthermore, Several new applications are also developed based on clique occurrence information, e.g. combining frequency and structure-oriented centrality measures together to discover interesting vertices or cliques.

It is necessary to mention that these proposed metrics and applications are only several examples for the exploitation of clique occurrence information. Actually, along this research direction, more excellent work can be carried out. Since frequency measure has been widely adopted and many frequency-based useful techniques have been developed in the field of data mining, this work is also an attempt to build a bridge between complex networks and data mining methods. More work can be done

towards the direction of applying data mining methods to solve the emerging and cutting-edge complex network problems.

## Acknowledgments

This research was partially supported by Scientific Research Foundation for the Returned Overseas Chinese Scholars (Ministry of Human Resources and Social Security of China, 2013), Zhejiang Provincial Natural Science Foundation of China (No. Y1110960; No. LQ13F030004), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No. 13YJC630084) and the NSFC (No. 11347201).



## References

1. A. Özgür, B. Cetin and H. Bingol, *Int. J. Mod. Phys. C* **19**, 689 (2008).
2. P. J. Kim and N. D. Price, *PLoS Comput. Biol.* **7**, e1002340 (2011).
3. D. H. Yang *et al.*, *PLoS ONE* **8**, e59241 (2013).
4. M. E. J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
5. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA*. **98**, 404 (2001).
6. V. Mayer-Schonberger and K. Cukier, *Big Data: A Revolution That Will Transform how We Live, Work, and Think*, Chap. 4 (Houghton Mifflin Harcourt Press, New York, 2013).
7. R. Toivonen, M. Kivelä, J. Saramäki, M. Viinikainen, M. Vanhatalo and M. Sams, *PLoS ONE* **7**, e28883 (2012).
8. M. Li, Y. Fan, J. Chen, L. Gao, Z. Di and J. Wu, *Physica A* **350**, 643 (2005).
9. J. J. Ramasco and S. A. Morris, *Phys. Rev. E* **73**, 016122 (2006).
10. T. Zhou, J. Ren, M. Medo and Y. C. Zhang, *Phys. Rev. E* **76**, 046115 (2007).
11. J. Ohkubo, K. Tanaka and T. Horiguchi, *Phys. Rev. E* **72**, 036120 (2005).
12. M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
13. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
14. L. Lü, Y. C. Zhang, C. H. Yeung and T. Zhou, *PLoS ONE* **6**, e21202 (2011).
15. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley and H. A. Makse, *Nat. Phys.* **6**, 888 (2010).
16. R. Puzis, Y. Elovici and S. Dolev, *AI Commun. — Network Analysis in Natural Sciences and Engineering* **20**, 287 (2007).
17. L. C. Freeman, *Soc. Netw.* **1**, 215 (1978).
18. A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, *Proc. Natl. Acad. Sci. USA* **101**, 3747 (2004).
19. M. G. Everett and S. P. Borgatti, *J. Math. Sociol.* **23**, 181 (1999).
20. E. D. Kolaczyk, D. B. Chua and M. Barthélemy, *Soc. Netw.* **31**, 190 (2009).
21. R. Agrawal, T. Imielinski and A. Swami, *Proc. ACM-SIGMOD Int. Conf. Management of Data*, eds. P. Buneman and S. Jajodia (ACM Press, USA, 1993), p. 207.