AKCE Int. J. Graphs Comb., 10, No. 3 (2013), pp. 253-261

PageRank regular digraphs with prime out-degrees*

Josep Conde, Nacho López and Francesc Sebé
Department of Mathematics
Universitat de Lleida
C.Jaume II, 69, E-25001, Spain
e-mail: {*jconde,nlopez,fsebe*}*@matematica.udl.cat*

Abstract

The PageRank algorithm was designed to rank web pages according to their relevance in the web. In this context, web topology is modeled as a directed graph whose vertices correspond to web pages with arcs representing hyperlinks. The algorithm assigns a score to each vertex, representing the relevance of its corresponding page in the web. In this paper we define a *PageRank regular digraph* as a digraph whose vertices have all the same PageRank score. Such digraphs are interesting in the scope of privacy preserving release of digraph data in environments where a dishonest analyst may have previous structural knowledge about the PageRank relevance of some vertices. In this paper we analyze some structural properties of PageRank regular digraphs and provide a characterization when vertices have prime out-degree.

Keywords: Directed Graph; PageRank; Perron vector; Regular digraph.

2000 Mathematics Subject Classification: 05C20; 15A48.

1. Introduction

Relations in a social network can be symmetric (like "be friends" or "be colleagues") or asymmetric (like "be subscribed to" or "trust in"). When relations are asymmetric, a directed graph structure arises where vertices correspond to social network participants and arcs represent the relations among them. This is the case for blogging sites where users can subscribe to other blogs so as to receive a notification each time a new entry is posted. Another example is given in recommendation platforms where users can indicate which other users they do trust (generating a "trust in" relation).

Data managed by social network online platforms is very sensitive due to their personal content. For instance, if a blogging platform released its digraph data, everybody could know the blogs each of its users is subscribed to. In the case of subscriptions to blogs with political, religious or health content some private aspects of bloggers life would be leaked.

When social network (digraph) data have to be released for scientific analysis [7], some privacy preserving measures have to be taken. Vertex identifiers removal is a mandatory

*This work was partly supported by the Spanish Government through project TIN2010-18978.

measure, but it may fall short when a dishonest analyst has some additional background information [10]. For instance, if an analyst had previous information about the number of blogs some participant is subscribed to, the released vertices he/she could correspond to would reduce to those whose out-degree equals the known amount of subscriptions. If just one vertex had such out-degree, the traced user's vertex would be re-identified.

Several techniques have been proposed for privacy preserving release of *undirected* social network data under different assumptions about the previous knowledge an attacker may have. In [4] the authors assume an attacker previously knows the amount of relations (*i.e.* the vertex degree) of some of the network members while in [9] a scenario where an attacker may know the neighborhood subgraph of some vertices is addressed. Other proposals like k-symmetry [8], k-automorphism [11] and k-isomorphism [2] provide solutions against any previous structural knowledge.

The PageRank algorithm [3, 6] takes a directed graph as input and assigns a relevance score to each of its vertices in such a way that vertices linked from relevant vertices tend to receive a higher score. This algorithm is widely employed by search engines as a tool for sorting the results of search queries. Its use in blogging and recommendation platforms is a source of background information an attacker could employ in its attempt to break the privacy of some of the social network participants.

This work addresses the privacy of social network digraph data release in scenarios in which a dishonest analyst is able to obtain additional information about the PageRank relevance of social network participants. In the particular case of a blogging platform, if an analyst knew the identity of the most relevant blogger, it could simply compute the PageRank score of the released anonymous digraph vertices and deduce the highest scored vertex corresponds to the most relevant user. Such an attack is not possible in digraphs with equally PageRank scored vertices. These digraphs will be referred to as *PageRank regular digraphs.*

2. Terminology and notation

A *digraph* $D = (V, A)$ consists of a finite nonempty set V of objects called *vertices* and a set A of ordered pairs of vertices called *arcs*. The *order* of D is the cardinality of its set of vertices V, denoted by $|V|$. If (u, v) is an arc, it is said that u is *adjacent to* v and also that v is *adjacent from* u. The set of vertices that are adjacent from [to] a given vertex v is denoted by $N^+(v)$ [resp. $N^-(v)$] and its cardinality is the *out-degree* of v, $d^+(v)$ [resp. *in-degree* of v, $d^-(v)$]. The maximum out– [resp. in–] degree of D is denoted by $\Delta^+(D)$ [resp. $\Delta^-(D)$]. If $d^+(v) = k$ [resp. $d^-(v) = k$], for all $v \in V$, then D is said to be *out–* [resp. *in–*] *regular* of degree k. The reader is referred to Chartrand and Lesniak [1] for additional concepts on digraphs.

3. PageRank vector of a digraph

Web search engines are widely employed to locate pages in the web. In particular, the search engine Google uses the PageRank algorithm [3, 6] (together with many other techniques) to sort results pages resulting from search queries. The idea underlying PageRank is that the relevance of a page increases when it is linked from relevant pages. Its input is a directed graph representing web pages (vertices) and the links among them (arcs). Given a directed graph $D = (V, A)$ of order n, the $n \times n$ *normalized link matrix* $\mathbf{P} = (p_{ij})$ is created. Given two vertices $v_i, v_j \in V$, p_{ij} is defined as

$$p_{ij} = \begin{cases} \frac{1}{d^+(v_j)} & \text{if } d^+(v_j) > 0 \text{ and } (v_j, v_i) \in A, \\ 0 & \text{if } d^+(v_j) > 0 \text{ and } (v_j, v_i) \notin A, \\ \frac{1}{n} & \text{if } d^+(v_j) = 0. \end{cases}$$

Hence, the coefficient p_{ij} of $\mathbf{P}$ can be viewed as the probability that a surfer located at vertex v_j jumps to vertex v_i (assuming the next movement is taken uniformly at random among the arcs emanating from v_j). For vertices v_j such that $d^+(v_j) = 0$, p_{ij} is $1/n$ for each i. That is, when the random surfer falls in a vertex with no outgoing arcs, it is able to restart the navigation from any vertex uniformly chosen at random. So as to permit this random restart behaviour when the surfer is at any vertex (with a small probability $1 - \alpha$), matrix $\mathbf{P}(\alpha)$ is created as

$$\mathbf{P}(\alpha) = \alpha\mathbf{P} + (1 - \alpha)\frac{1}{n}\mathbf{J}, \tag{1}$$

where $\mathbf{J}$ denotes the order n all ones square matrix. By construction, $\mathbf{P}(\alpha)$ is a positive matrix [5], hence, $\mathbf{P}(\alpha)$ has a unique positive eigenvalue (its value is 1) on the spectral circle. The *PageRank vector* is defined to be the (positive) eigenvector $\mathcal{P} = (p_0, \ldots, p_{n-1})$ with $\sum_i p_i = 1$ (the Perron vector of $\mathbf{P}(\alpha)$) associated to this eigenvalue. The probability α, known as the *dumping factor*, is usually chosen to be $\alpha = 0.85$. The relevance score assigned by PageRank to vertex v_j is p_j. This value represents the long-run fraction of time the surfer would spend at vertex v_j.

Example 3.1. *Let us consider a small digraph (Figure 1)* $D = (V, A)$, *with* $V = \{v_0, v_1, v_2, v_3, v_4\}$ *and* $A = \{(v_0, v_1), (v_1, v_2), (v_2, v_0), (v_3, v_1), (v_3, v_4), (v_4, v_1)\}$. *Applying the PageRank algorithm to digraph* D, *the following vector is obtained:* $\mathcal{P} = (p_0, p_1, p_2, p_3, p_4) = (0.2919, 0.3272, 0.3081, 0.0300, 0.0428)$. *Hence,* v_1 *is the most relevant vertex while* v_3 *is the least one. In this particular example, value* α *has been chosen to be* 0.85. *Matrix* $\mathbf{P}$ *is as follows,*

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & \frac{1}{2} & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

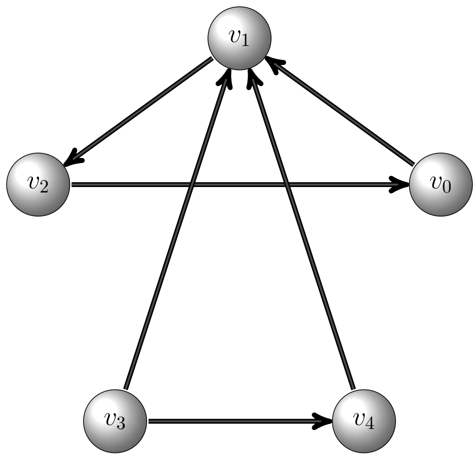


Figure 1: Digraph having v_1 as its most relevant vertex, while v_3 is the least one.

4. Definition and properties of PageRank regular digraphs

This section begins with a formal definition of the *PageRank regularity* property for digraphs.

Definition 4.1. *A digraph D is defined to be PageRank regular when its PageRank vector $\mathcal{P}$ is regular, that is, all the vertices of D have the same PageRank score.*

By definition $\mathcal{P} = (p_0, \dots, p_{n-1})$ satisfies $\sum_i p_i = 1$, so that, when $\mathcal{P}$ is regular it turns out to be $\mathcal{P} = (\frac{1}{n}, \cdots, \frac{1}{n}) = \frac{1}{n}\mathbf{j}$, with $\mathbf{j}$ denoting the all ones length n vector. Since the espectral radius of $\mathbf{P}(\alpha)$ is 1 (its transposed matrix is stochastic [5]), PageRank regular digraphs satisfy that $\big(1, \frac{1}{n}\mathbf{j}\big)$ is an eigenpair for $\mathbf{P}(\alpha)$. The next proposition shows that the normalized link matrix $\mathbf{P}$ of a PageRank regular digraph satisfies the same property.

Proposition 4.2. *Let $0 < \alpha \le 1$ and let $\mathbf{P}$ be the normalized link matrix of an order n digraph. Then the following assertions are equivalent,*

- $\big(1, \frac{1}{n}\mathbf{j}\big)$ *is an eigenpair for* $\mathbf{P}$.
- $\big(1, \frac{1}{n}\mathbf{j}\big)$ *is an eigenpair for* $\mathbf{P}(\alpha)$, *being* $\frac{1}{n}\mathbf{j}$ *the Perron vector of* $\mathbf{P}(\alpha)$.

Proof. It is easy to see that $\big(1, \frac{1}{n}\mathbf{j}\big)$ is an eigenpair for $\frac{1}{n}\mathbf{J}$ since, for each i,

$$\left[\frac{1}{n}\mathbf{J} \cdot \frac{1}{n}\mathbf{j}\right]_i = \frac{1}{n^2}\,[J \cdot j]_i = \frac{1}{n^2}\sum_{k=0}^{n-1} J_{ik} \cdot j_k = \frac{1}{n^2}\sum_{k=0}^{n-1} 1 = \frac{1}{n} = \left[\frac{1}{n}\mathbf{j}\right]_i .$$

Let us assume $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for $\mathbf{P}$. Then,

$$\mathbf{P}(\alpha)\cdot\frac{1}{n}\mathbf{j} = \left(\alpha\mathbf{P} + (1-\alpha)\frac{1}{n}J\right)\cdot\frac{1}{n}\mathbf{j} = \alpha\mathbf{P}\cdot\frac{1}{n}\mathbf{j} + (1-\alpha)\frac{1}{n}J\cdot\frac{1}{n}\mathbf{j} = \alpha\frac{1}{n}\mathbf{j} + (1-\alpha)\frac{1}{n}\mathbf{j} = \frac{1}{n}\mathbf{j},$$

hence $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for $\mathbf{P}(\alpha)$. Since $\mathbf{P}(\alpha)$ is a positive matrix and $\frac{1}{n}\mathbf{j}$ is a positive eigenvector of $\mathbf{P}(\alpha)$, this must be the Perron vector of $\mathbf{P}(\alpha)$.

Conversely, assuming $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for $\mathbf{P}(\alpha)$, then,

$$\mathbf{P}\cdot\frac{1}{n}\mathbf{j} = \left(\frac{1}{\alpha}\mathbf{P}(\alpha) - \frac{(1-\alpha)}{\alpha}\frac{1}{n}J\right)\cdot\frac{1}{n}\mathbf{j} = \frac{1}{\alpha n}\mathbf{j} - \frac{(1-\alpha)}{\alpha n}\mathbf{j} = \frac{1}{n}\mathbf{j},$$

that is, $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for $\mathbf{P}$. □

The previous proposition permits to conclude that the value α taken by the PageRank algorithm does not influence the PageRank regularity of a digraph. Another important consequence is that a digraph D is PageRank regular if and only if $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for its normalized link matrix $\mathbf{P}$.

Let us observe that, by construction, each column of $\mathbf{P}$ sums 1, that is, for each j, $\sum_i p_{ij} = 1$. Let us also remark that the non-zero elements of a given column j have all the same value $1/d^+(v_j)$. Next, we characterize PageRank regular digraphs in terms of the coefficients of $\mathbf{P}$.

Proposition 4.3. *Let D be a digraph and let $\mathbf{P}$ be its normalized link matrix. Then, D is a PageRank regular digraph if and only if each row of $\mathbf{P}$ sums 1, that is, $\sum_j p_{ij} = 1$, for each i.*

Proof. $\left(1, \frac{1}{n}\mathbf{j}\right)$ is an eigenpair for $\mathbf{P}$ if and only if $\mathbf{P}\frac{1}{n}\mathbf{j} = \frac{1}{n}\mathbf{j}$. That is, for each i, $\frac{1}{n}\sum_j p_{ij} = \frac{1}{n}$, which is equivalent to $\sum_j p_{ij} = 1$. □

Sumarizing, the normalized link matrix $\mathbf{P}$ of a PageRank regular digraph satisfies:

- All the elements in the diagonal of $\mathbf{P}$ are zero, that is $p_{ii} = 0$ (we consider digraphs without self-loops).
- Given $p_{ij} = \dfrac{1}{k} \Rightarrow \exists\, i_1, \cdots, i_k$ such that $p_{i'j} = \begin{cases} \frac{1}{k} & \text{if } i' \in \{i_1, \cdots, i_k\}, \\ 0 & \text{otherwise.} \end{cases}$
- Each row of $\mathbf{P}$ sums 1, that is, $\sum_j p_{ij} = 1$.

The above conditions provide nice properties of PageRank regular digraphs in graphical terms. For instance, the following proposition states that the maximum out-degree in a PageRank regular digraph D cannot exceed its maximum in-degree.

Proposition 4.4. *Let D be a PageRank regular digraph. Then $\Delta^-(D) \leq \Delta^+(D)$. The equality holds if and only if D is an out-regular digraph.*

Proof. First of all, we observe that the degrees of D can be writted in terms of the coefficients of $\mathbf{P}$:

- $d^+(v_j) = \begin{cases} \frac{1}{p_{ij}} & \text{if } p_{ij} \neq 0, \frac{1}{n}, \text{ for some } i; \\ 0 & \text{if } p_{ij} = \frac{1}{n}, \text{ for each } i. \end{cases}$
- $d^-(v_i) = \sum_{j=0}^{n-1} d^+(v_j) \cdot p_{ij}$.

Let v_i be a vertex of D. Then,

$$d^-(v_i) = \sum_j d^+(v_j) p_{ij} \leq \sum_j \Delta^+(D) p_{ij} = \Delta^+(D) \sum_j p_{ij} = \Delta^+(D).$$

The previous inequality holds for all the vertices of D, including those satisfying $d^-(v_i) = \Delta^-(D)$. Hence, $\Delta^-(D) \leq \Delta^+(D)$. Equality holds when $d^+(v_j) = \Delta^+(D)$ for each $v_j \in V$, that is, when D is an out-regular digraph. □

Next, a characterization of PageRank regularity for out-degree regular digraphs is provided.

Proposition 4.5. *Let D be an out-regular digraph of degree k. D is PageRank regular if and only if D is an in-regular digraph of degree k.*

Proof. If $k = 0$, D does not have any arc and its normalized link matrix $\mathbf{P}$ is $\frac{1}{n}J$. Each row of this matrix sums 1 so that by Proposition 4.3, D is PageRank regular.

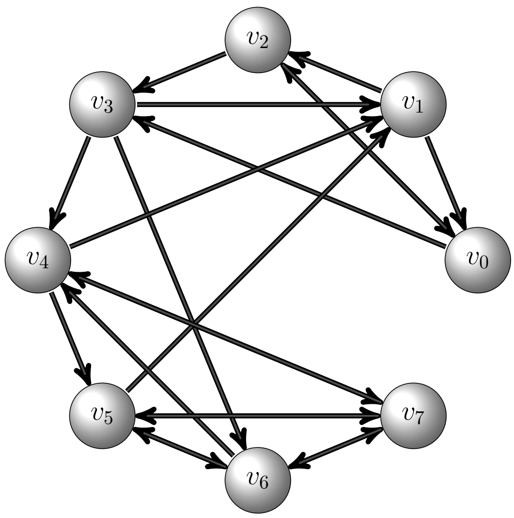
For $k > 0$, each vertex v_i of D satisfies that,

$$d^-(v_i) = \sum_j d^+(v_j) \cdot p_{ij} = \sum_j k \cdot p_{ij} = k \sum_j p_{ij}.$$

Hence, $\sum_j p_{ij} = 1$ if and only if $k = d^-(v_i)$ for each i. □

5. PageRank regular digraphs with prime out-degrees

In this section we present a 'local' characterization of PageRank regularity for digraphs whose vertices have prime out-degree. For example, the out-degree of the vertices of the digraph depicted in Figure 2 is either 2 or 3. It is easy to check that each row of its normalized link matrix sums 1 so that it is a PageRank regular digraph. In this particular example, we can see that the non-null elements of each row of $\mathbf{P}$ are equal. This property means that all the in-neighbors of a vertex v of D have the same out-degree which coincides with the in-degree of v. Next, we prove this is a necessary and sufficient condition.



$$\mathbf{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Figure 2: A PageRank regular digraph with prime out-degrees (left) and its normalized link matrix (right).

Lemma 5.1. *Let* $\{q_1,\ldots,q_s\}$ *be a set of different prime numbers and let* $\{n_1,\ldots,n_s\}$ *be a set of non-negative integers. The following assertions are equivalent,*

- $\sum\limits_{i=1}^{s} n_i \frac{1}{q_i} = 1$
- *There exists* $k \in \{1,\ldots,s\}$ *such that* $n_k = q_k$ *and* $n_j = 0$ *for each* $j \neq k$ *.*

Proof. If there exists k such that $n_k = q_k$ and $n_j = 0$ if $j \neq k$, then,

$$\sum_{i=1}^{s} n_i \frac{1}{q_i} = \frac{n_k}{q_k} = 1.$$

Conversely, let $\sum_{i=1}^{s} n_i \frac{1}{q_i} = 1$. Without loss of generality, we assume that $n_1 \neq 0$. Multiplying both sides of the equality by $q_1 q_2 \cdots q_s$ we obtain:

$$n_1(q_2 q_3 \cdots q_s) + n_2(q_1 q_3 \cdots q_s) + \ldots + n_s(q_1 q_2 \cdots q_{s-1}) = q_1 q_2 \cdots q_s,$$

that is,

$$n_1(q_2 q_3 \cdots q_s) = q_1 q_2 \cdots q_s - (n_2(q_1 q_3 \cdots q_s) + \ldots + n_s(q_1 q_2 \cdots q_{s-1})).$$

We observe that q_1 divides every term in the right side of the equality. As a consequence q_1 must also divide $n_1(q_2 q_3 \cdots q_s)$ and since each $q_i \neq q_1$ is prime, we derive that q_1 divides n_1, that is, $n_1 = m_1 q_1$, with $m_1 > 0$.

1. If $m_1 > 1$ then $n_1 \frac{1}{q_1} > 1$ which implies that $\sum_{i=1}^{s} n_i \frac{1}{q_i} > 1$ (contradiction).

2. If $m_1 = 1$ then $n_1 = q_1$ and $n_1 \frac{1}{q_1} = 1$. Hence $\sum_{i=1}^{s} n_i \frac{1}{q_i} = 1$ if and only if $n_2 = \ldots = n_s = 0$.

□

Next theorem provides a characterization of PageRank regularity for digraphs whose vertices have prime out-degree.

Theorem 5.2. *Let $D = (V, A)$ be a digraph such that the out-degree of each of its vertices is a prime number. D is PageRank regular if and only if each arc $(u, v) \in A$ satisfies the following conditions:*

(a) $d^-(v) = d^+(u)$,

(b) $d^+(u') = d^+(u)$ *for each* $u' \in N^-(v)$.

Proof. Let $V = \{v_0, \ldots, v_{n-1}\}$, $(v_j, v_i) \in A$ and $q = d^-(v_i)$. If both (a) and (b) hold then, $p_{ij} = \frac{1}{q}$ for each $v_j \in N^-(v_i)$ and $p_{ij} = 0$ otherwise. Since $|N^-(v_i)| = q$,

$$\sum_j p_{ij} = \sum_{v_j \in N^-(v_i)} \frac{1}{q} = q\frac{1}{q} = 1,$$

and D is PageRank regular by Proposition 4.3.

Conversely, let $Q = \{q_1, q_2, \cdots, q_s\}$ be a set of prime numbers so that each vertex v of D satisfies that $d^+(v) \in Q$. Let $N^-_{q_\ell}(v_i)$ be the set of in-neighbors of $v_i \in V$ whose out-degree is q_ℓ. Note that $\{N^-_{q_\ell}(v_i)\}_{q_\ell \in Q}$ is a partition of $N^-(v_i)$. Let $n_\ell = |N^-_{q_\ell}(v_i)|$. Since we are assuming D is PageRank regular, the i-th row of $\mathbf{P}$ satisfies,

$$1 = \sum_j p_{ij} = \sum_\ell n_\ell \frac{1}{q_\ell}.$$

By Lemma 5.1, there exists a unique k such that $n_k = q_k$ and $n_\ell = 0$ if $\ell \neq k$. Hence, $N^-(v_i) = N^-_{q_k}(v_i)$ and $|N^-(v_i)| = q_k$, that is, vertex v_i has exactly q_k in-neighbors whose out-degree is q_k and (a) and (b) do hold. □

5. Conclusion

In this work, the concept of *PageRank regular* digraph has been introduced and defined. Next, PageRank regularity has been analyzed in terms of the normalized link matrix of the digraph. After presenting some related lemmas and propositions, a characterization has been given for PageRank regular digraphs whose vertices have all prime out-degree. Our future research on this topic will be focused on the characterization of PageRank regularity without the prime out-degree constraint.

References

[1] G. Chartrand, L. Lesniak, *Graphs Digraphs*, Fourth Edition. CRC Press, Boca Raton (2004).

[2] J. Cheng, A.W-C. Fu and J. Liu, *K -Isomorphism: privacy preserving network publication against structural attacks*, Proc. of SIGMOD'10 (2010), 459–470.

[3] A.N. Langville and C.D. Meyer, Deeper inside pagerank, *Internet Mathematics*, **1**(3) (2004), 335–380.

[4] K. Liu and E. Terzi, *Towards identity anonymization on graphs*, Proc. of SIGMOD'08, (2008), 93–106.

[5] C.D. Meyer, *Matrix analysis and applied linear algebra*, SIAM: Society for Industrial and Applied Mathematics (2001).

[6] L. Page, S. Brin, R. Motwani and T. Winograd, *The pagerank citation ranking: bringing order to the web*, Technical Report, Stanford InfoLab (1998).

[7] J. Scott, *Social network analysis handbook*, Sage Publications Inc. (2000).

[8] W. Wu, Y. Xiao, W. Wang, Z. He and Z. Wang, *K -Symmetry model for identity anonymization in social networks*, Proc. of EDBT'2010 (2010), 111–122.

[9] B. Zhou and J. Pei, *Preserving privacy in social networks against neighborhood attacks*, Proc. of ICDE'08 (2008) 506–515.

[10] B. Zhou, J. Pei and W-S. Luk, A brief survey on anonymization techniques for privacy preserving publishing of social network data, *ACM SIGKDD Explorations Newsletter*, **10**(2) (2008), 12–22.

[11] L. Zou, L. Chen and M.T. Öszu, *K -Automorphism: a general framework for privacy preserving network publication*, Proc. of VLDB'09 (2009), 946–957.