Leadership groups on Social Network Sites based on personalized PageRank

Francisco Pedroche[a], Francisco Moreno[b], Andrés Gonzalez[b], Andrés Valencia[b]

[a]*Institut de Matemàtica Multidisciplinària*
Universitat Politècnica de València
Camí de Vera s/n. 46022 València. Spain.
{*pedroche@imm.upv.es*}
[b]*Escuela de Sistemas*
Universidad Nacional de Colombia, Sede Medellín
Carrera 80 No 65-223. Medellín. Colombia.
{*fjmoreno, afgonzalezr, afvalenciab@unal.edu.co*}

Abstract

In this paper we present a new framework to identify leaders on an SNS using the Personalized PageRank vector. The methodology is based in the concept of *Leadership group* recently introduced by one of the authors. We show how to analyze the structure of the *Leadership group* as a function of a single parameter. Zachary's network and a Facebook university network are used to illustrate the applicability of the model. As an application we introduce some new concepts such as the probability to be a leader, a classication of networks and the concept of best potential friend.

Keywords: Google matrix, PageRank, link analysis, communication network, social networking, ranking algorithm
2000 MSC: 91D30, 15A21, 60K15, 90B15, 94C15

---

1. Introduction

Some usual centrality measures (degree, betweeness, etc.) can be used to assign importance to users in a Social Network Site (SNS). We use the concept of PageRank [1] since it has proved to be of utility in some fields, apart of being in the core of the searcher Google. For example, PR have been used in biomedical literature retrieval [2], [3]; in this area, a common assumption

Preprint submitted to Special Issue of Math. & Comp. Mod.-MME&HB 2011September 29, 2011

is to combine linkage information and content information. In the field of computational Biology, [4] use PR to classify species in order to analyze which species cause the most damage if removed. In [5] the authors use the PR to classify nodes in water supply networks using real data. Recently, PR have been used to classify tennis players [6].

We present a framework to classify the users of an SNS based on the Personalized PageRank (PPR) vector. PPR is PR when using some prescribed *personalization vector*. PPR was originally introduced to bias PR to personal preferences of the users [7]. See, e.g., [8] for an analytical formulation. [9] uses topics of the queries to bias the PageRank. [10] computes the PPR for some subsets of pages. For computational issues, see [11] for a low-rank approximation technique and [12] for a Monte-Carlo approach. In [13] a technique to preserve the PPR when considering a subgraph is shown.

We are interesting in classifying the nodes of a network considering the direct graph of the network and the features of the nodes. PPR is used to include some features of the nodes. The method presented allows to give an extra of PR to some nodes in a controlled way. In this paper we call a *leader* a node that has higher PR than the others. The fundamentals of the model were presented in [14], where the concept of *Leadership group*, $\mathcal{L}$, was introduced. In this paper we analyze the structure of $\mathcal{L}$ and we show how to characterize $\mathcal{L}$ by means of a single parameter ϵ. We also introduce a new parameter, the frequency in $\mathcal{L}$, that is useful for assigning importance to a node. By means of this we define the *probability to be a leader*. We illustrate the applicability of these new concepts by using two real networks. In section 2 we recall some definitions. We define the concept of $\mathcal{L}$ and analyze it when applied to some toys networks in section 3. In section 4 we present some results in two networks: one is the classical Zachary network (34 nodes) and the other corresponds to a Facebook university network (769 nodes). In section 5 we describe how the model can be used to define the *probability to be a leader*, how to make a classification of networks with ϵ and the concept of best potential friend. In section 6 we give some conclusions and some lines for future work.

2. Preliminaries

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be the directed graph representing a Social Network Site. Users are represented by the set of nodes $\mathcal{N} = \{1, 2, \ldots, n\}$ and the set of directed links is $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. The link represented by the pair (i, j) belongs to the set $\mathcal{E}$ if and only if there exists a link pointing from node i to node j. In this paper we assume that each node has at least one outlink; i.e., there are no *dangling nodes*; denoting d_i the number of outlinks of a node i, we assume $d_i \neq 0$ for all $i \in \mathcal{N}$. We use the PageRank vector [1] as the main classification tool. Since there are no *dangling nodes* we can define the row stochastic matrix $P = (p_{ij}) \in \mathbb{R}^{n\times n}$, in the form

$$p_{ij} = \begin{cases} d_i^{-1} & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \qquad 1 \leq i, j \leq n.$$

Let $0 < \alpha < 1$ be the damping factor (that we use as $\alpha = 0.85$). Let $\mathbf{e} \in \mathbb{R}^{n\times 1}$ be the vector of all ones and let $\mathbf{v}$ be the personalization (or teleportation) vector, i.e., $\mathbf{v} = (v_i) \in \mathbb{R}^{n\times 1} : v_i > 0$ for all $i \in \mathcal{N}$ and $\mathbf{v}^T\mathbf{e} = 1$. The Google matrix is defined as $G = \alpha P + (1-\alpha)\mathbf{e}\mathbf{v}^T$, and is an stochastic and primitive (irreducible and aperiodic) matrix [7]. The PageRank vector is defined as the unique left Perron vector of G $\pi^T = \pi^T G$, with $\pi^T\mathbf{e} = 1$. Denoting $\mathbf{e}_i$ the ith column of the identity matrix of order n, the PageRank of a node i is $\pi_i = \pi^T\mathbf{e}_i$. We call basic PageRank, and denote it by *basic PR* to the vector $\pi(\mathbf{e}/n)$. We call *basic leader* a node that is in the top of the *basic PR*.

3. Leadership group

The concept of Leadership group was initially defined in [14] and some details were given in [16]. We here redefine this concept and include some new concepts. The following two definitions constitute the framework that allows classify users using PPR.

Definition 1. *Given a directed graph* $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, *let* $0 < \epsilon \leq \frac{n-1}{n}$ *and let* $\mathbf{v}_i(\epsilon) = [v_{ij}] \in \mathbb{R}^{n\times 1} : v_{ii} = 1-\epsilon$, $v_{ij} = \epsilon/(n-1)$ *if* $i \neq j$. *For each* $i \in \mathcal{N}$, *let* $PR_i = \pi(\mathbf{v}_i)$. *and we denote as* $(PR_i)_j$ *the* j*th entry of* PR_i.

Note that $(PR_i)_j$ represents the value of the PR corresponding to node j when using the personalization vector $\mathbf{v}_i(\epsilon)$.

Definition 2. *Given a directed graph* $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, *and* $0 < \epsilon \leq \frac{n-1}{n}$, *the* Leadership group, $\mathcal{L} \subseteq \mathcal{N}$ *is defined as follows:* $j \in \mathcal{L}$ *if, for some* $i \in \mathcal{N}$ *it holds that*

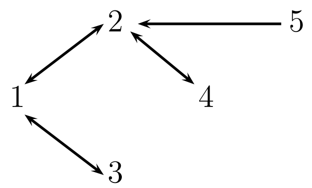
$$(PR_i)_j \geq (PR_i)_k \text{ for all } k \neq j. \tag{1}$$

i.e. for some personalization vector $\mathbf{v}_i(\epsilon)$*, node* j *has the greatest PageRank. The number of different indices* $i \in \mathcal{N}$ *for which* (1) *occurs is called the frequency of node* j *in* $\mathcal{L}$*, and we denote it as* $\nu_{\mathcal{L}}(j)$*.*

It is easy to see that in case of taking $\epsilon = \frac{n-1}{n}$ we obtain that $\pi(\mathbf{v}_i) = \pi(\mathbf{e}/n)$, which is the *basic PR*. Attending to an statistical meaning, the effect of $\mathbf{v}$ in the PageRank is the following. Let us assume that $\mathbf{v} = \mathbf{e}/n$ (which is the usual value). Then, if there is not a link from page i to page j the effect of $\mathbf{v}$ is putting an artificial link (i.e., a teleportation) from page i to page j. The random surfer will follow this artificial link with probability $(1-\alpha)/n$. In the case that we take $\mathbf{v} = \mathbf{e}_i$ what happens is that from nodes that did not go to node i now we have put an artificial link that the surfer will follow with probability $(1-\alpha)$.

To illustrate the definition of $\mathcal{L}$ and the effect of the parameter ϵ we show now the following examples using toy graphs.

3.1. Examples



$$P = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1: A toy graph and its corresponding row stochastic matrix

Node	PR_1	PR_2	PR_3	PR_4	PR_5	$\nu_{\mathcal{L}}(i)$
1	**0.38**	0.29	**0.34**	0.26	0.26	2
2	0.30	**0.39**	0.27	**0.35**	**0.35**	3
3	0.17	0.13	0.25	0.12	0.12	0
4	0.14	0.18	0.13	0.25	0.16	0
5	0.01	0.01	0.01	0.01	0.11	0

Table 1: PR_i, and $\nu_{\mathcal{L}}(i)$ for the graph of Fig. 1, with $\epsilon = 0.3$.

For the graph shown in Fig 1, computing the *basic PR* we obtain $basicPR = \pi(\frac{1}{5}[1,1,1,1,1]) = \begin{bmatrix} 0.31 & 0.33 & 0.16 & 0.17 & 0.03 \end{bmatrix}^T$. In Table 1 we show the resulting PR_i giving by Definition 1 with $\epsilon = 0.3$. Note that in this example changing $\mathbf{v}$ we allow to change the ranking. The ranking, giving by the *basic*

PR is $\{node2, node1, node4, node3, node5\}$ while using $\mathbf{v} = \mathbf{v}_1$ the ranking is giving by PR_1, which is $\{node1, node2, node3, node4, node5\}$. In this example we have $\mathcal{L} = \{1, 2\}$ and $\nu_{\mathcal{L}}(1) = 2$ and $\nu_{\mathcal{L}}(2) = 3$. Note also that, giving a node, the maximum value of PPR for that node is obtained when computing PR_i; this is in accordance with the statistical meaning of the personalization vector commented above. We say that the $\mathbf{v}_i$ is useful to give an extra of PR to node i. Note also that there are nodes such as nodes 3, 4 or 5 that do not win even though we give the maximum of extra PR for the ϵ considered. Computing the same experiment for some values of ϵ we find that for $\epsilon \leq 0.68$ we have that $\mathcal{L} = \{1, 2\}$ while for $\epsilon \geq 0.69$ we have that $\mathcal{L} = \{1\}$. Therefore, we conclude that in this graph the ranking, the structure of $\mathcal{L}$ and the values of $\nu_{\mathcal{L}}(i)$ depend on ϵ. In this example the *basic leader* ($node2$) verifies that $\nu_{\mathcal{L}}(basic\, leader) > \nu_{\mathcal{L}}(rest\, of\, nodes)\ \forall \epsilon$. We shall see later that in some networks this inequality does not hold. Therefore $\nu_{\mathcal{L}}(i)$ can be used as a centrality measure different from the *basic PR*.

In Fig. 2 a) we show a linear graph of three nodes. In this case an easy computation shows that $\mathcal{L} = \{2\}$, for all ϵ. Therefore we have $\nu_{\mathcal{L}}(2) = 3, \forall \epsilon$. In this graph we always have one leader. By means of the personalization vector we can promote node 2 or node 3 to second position, as much. In this example the parameter ϵ is useful to make bigger or lower the relative importance of the nodes, but does not allow to change $\mathcal{L}$.

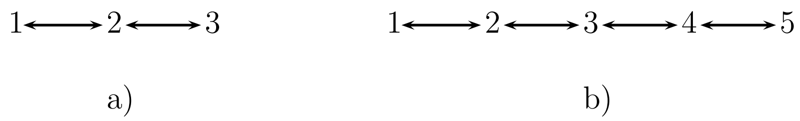


Figure 2: In a) node 2 is the leader $\forall \epsilon$. In b) node 3 can be a leader when $\epsilon \leq 0.8$.

In Fig. 2 b) we show a linear graph of five nodes. Now the situation is complectly different. For $\epsilon \leq 0.7$ we have that $\mathcal{L} = \{2, 3, 4\}$, with $\nu_{\mathcal{L}}(2) = \nu_{\mathcal{L}}(4) = 2$ and $\nu_{\mathcal{L}}(3) = 1$. For $\epsilon \geq 0.8$ we have, surprisingly, that $\mathcal{L} = \{2, 4\}$, with $\nu_{\mathcal{L}}(2) = \nu_{\mathcal{L}}(4) = 3$. This happens since for the ranking with PR_3, with $\epsilon \geq 0.8$, *both* nodes 2 and 4 are winners. Therefore this example shows that $\sum_{i \in \mathcal{N}} \nu_{\mathcal{L}}(i) \geq n$, since there are situations when we have two (or more) leaders at the same time (*cohabitation*). Note also that in this graph even though we give all the extra PR to node 3, with $\epsilon \geq 0.8$, this extra PR goes to benefit node 2 and node 4.

With this examples we have seen that ϵ is a interesting parameter, in this model, to characterize networks: some networks have only one leader, some can have two leaders, in some networks any user can be in second position, etc. These features could be very useful for the managers of SNSs.

4. Experiments in real networks

4.1. Zachary's karate club [17]

It is a network of 34 nodes representing the members of a university-based karate club. The club split in two factions. The leaders of the two resulting communities were node 1 and node 34. Using the *basic PR* one can classify the users of the network. In Table 2 only the classification of the first six nodes is shown. We see that the *basic PR* detect well the two main leaders of the network.

Node	34	1	33	3	2	32
basic PR	0.1009	0.0970	0.0717	0.0571	0.0529	0.0372

Table 2: Classification of the first six nodes using the *basic PR*

ϵ	card($\mathcal{L}$)	$\mathcal{L}$
0.001	33	$\mathcal{N} - \{12\}$
0.01	28	$\mathcal{N} - \{12, 15, 16, 19, 21, 23\}$
0.05	27	$\mathcal{N} - \{12, 15, 16, 19, 21, 23, 27\}$
0.1	24	$\mathcal{N} - \{12, 13, 15, 16, 18, 19, 21, 22, 23, 27\}$
0.3	23	$\mathcal{N} - \{10, 12, 13, 15, 16, 18, 19, 21, 22, 23, 27\}$
0.4	18	$\mathcal{N} - \{5, 10, 11, 12, 13, 15, 16, 18, 19, 20, 21, 22, 23, 27, 29, 30\}$
0.6	5	$\{1, 2, 3, 33, 34\}$
0.7	4	$\{1, 3, 33, 34\}$
0.8	2	$\{1, 34\}$
0.9	2	$\{1, 34\}$
0.9706	1	$\{34\}$

Table 3: Cardinal of the Leadership group for Zachary's karate club for some values of ϵ, and the corresponding nodes that form $\mathcal{L}$.

In Table 3 we show how the cardinal of $\mathcal{L}$ depends on ϵ. For example, using $\epsilon = 0.01$ we have that all the nodes are in $\mathcal{L}$ except nodes $\{12, 15, 16, 19, 21, 23\}$. We remark here that ϵ is a practical tool for the managers of an SNS to decide how they want to treat their SNS. For example they can use $\epsilon = 0.001$ if they want a nearly democratic network in which anyone (except node 12) can be the leader. Using this method one can enhance the importance of node 12 using the personalization vector but we cannot make 12 the winner. In Table 4 we show $\nu_{\mathcal{L}}(i, \epsilon)$ for some nodes and for some values of ϵ. Note, for example, that node 33 can be a leader for values of ϵ ranging between 0.001 and 0.7. Let us select $\epsilon = 0.4$ to analyze this network: there are two leaders (node 1 and node 34) which are more important than the rest, since $\nu_{\mathcal{L}}(1) = 8$, for $\epsilon = 0.4$. Here we have $\nu_{\mathcal{L}}(\textit{basic leader}) < \nu_{\mathcal{L}}(1)$. If we use $\nu_{\mathcal{L}}$ as a centrality measure we have that for this ϵ node 1 is more important than node 34. This is a feature that *basic PR* does not predict.

$node \backslash \epsilon$	0.001	0.01	0.05	0.1	0.3	0.4	0.6	0.7	0.8	0.9	0.9706
1	2	2	2	5	5	8	14	15	14	13	0
7	1	1	1	1	1	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
18	1	1	1	0	0	0	0	0	0	0	0
23	1	0	0	0	0	0	0	0	0	0	0
33	1	1	1	1	1	1	1	1	0	0	0
34	1	6	7	7	8	10	17	17	20	21	34

Table 4: Some values of $\nu_{\mathcal{L}}(i, \epsilon)$ for Zachary's karate club.

4.2. Facebook network from Caltech-2005

Here we use a network of 769 nodes that corresponds to users of Facebook from the California Institute of Technology; see [18] for details[1]. In Table 5 we show the classification of the first 8 nodes using the *basic PR*.

In Table 6 we show the cardinal of the Leadership group for some values of ϵ. We see, for example, that for $\epsilon = 0.97$, we have 98 nodes that can be the leader. Note that for $\epsilon \leq 0.7$ any node can be the leader. A deeper analysis is shown in Table 7 for two selected values of ϵ. In this table we show the

[1]Data available at $http://people.maths.ox.ac.uk/porterm/data/facebook5.zip$

Node	623	207	563	60	405	88	82	411
basic PR	0.0067	0.0056	0.0056	0.0050	0.0048	0.0046	0.0043	0.0043

Table 5: Classification of the first 8 nodes from *basic PR* for the network of Facebook.

ϵ	0.5	0.7	0.9	0.95	0.97
card($\mathcal{L}$)	769	769	757	734	98

Table 6: Cardinal of $\mathcal{L}$ for the network of Facebook at Caltech.

$\epsilon =$ 0.9	Node	3	5	9	60	88	192	207	360	455	461
	$\nu_{\mathcal{L}}$	2	2	2	2	2	2	2	2	2	2
$\epsilon =$ 0.97	Node	623	207	201	563	3	9	60	88	754	5
	$\nu_{\mathcal{L}}$	659	5	3	3	2	2	2	2	2	1

Table 7: Values of $\nu_{\mathcal{L}}(i, \epsilon)$ for the network of Facebook at Caltech.

first 10 nodes that contribute to $\mathcal{L}$ for $\epsilon = 0.9$ and $\epsilon = 0.97$. We have noted before that the parameter ϵ controls who may be a leader. In this table we also see this feature. When using $\epsilon = 0.9$ not all the nodes contribute to the Leadership group, while there are some nodes with $\nu_{\mathcal{L}}(i) = 2$. Note that for $\epsilon = 0.97$ the situation has changed dramatically and we have an absolute leader (node 623, which corresponds with that computed with the *basic PR*). Note that $\nu_{\mathcal{L}}(i, \epsilon)$ can be used to define communities: for $\epsilon = 0.97$ in this example we could define 6 communities in the network. 5 communities inside $\mathcal{L}$- those corresponding to $\nu_{\mathcal{L}} \in \{659, 5, 3, 2, 1\}$- and the other community formed by the nodes outside the Leadership group ($\nu_{\mathcal{L}}(i) = 0$). We also note that with the analysis of Table 7 we have new leaders that were no identified with the *basic PR*. Therefore this experiment shows again that the presented technique improves the classification computed using *only* the *basic PR*.

5. Other applications of the model

In the above sections we have seen that the framework defined in section 2 can be useful to classify users in SNSs. The experiments have shown that we can introduce some more useful concepts. In this section we describe three applications of the presented model.

5.1. Probability to be a leader

With ϵ we have the opportunity to change the classification given by the *basic PR*. For a given ϵ we could define the probability of node i to be a leader as $\nu_{\mathcal{L}}(j)$, but since we can have the phenomenon of the cohabitation it is formally better to do the following.

Definition 3. *Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, and $0 < \epsilon \leq \frac{n-1}{n}$, let $\mathcal{L}$ and $\nu_{\mathcal{L}}(j)$ given by Definition 2. We define the probability of node $i \in \mathcal{N}$ to be a leader as $P(i,\epsilon) = \frac{\nu_{\mathcal{L}}(i,\epsilon)}{\sum_{j\in\mathcal{N}} \nu_{\mathcal{L}}(j,\epsilon)}$.*

For example, in the Zachary network we have that using our model node 12 can never win; i.e. $\nu_{\mathcal{L}}(j = 12) = 0$, $\forall \epsilon$. Therefore $P(12,\epsilon) = 0$, $\forall \epsilon$. Values of $P(i,\epsilon)$ can be obtained from Table 4. For example $P(34, \epsilon = 0.1) = 7/34$ and $P(34, \epsilon = 0.9706) = 1$.

5.2. Classification of networks

The parameter ϵ of the model can be used to classify networks.

Definition 4. *Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a directed graph of a network and let $\mathcal{L}$ be given by Definition 2. We say that the network is a* democratic network *if there exists ϵ_C such that for $\epsilon \leq \epsilon_C$ it holds that $i \in \mathcal{L}$, $\forall i \in \mathcal{N}$. In other case we say the network is an* oligarchic network *of order $p = card\mathcal{L} < n$. When $p = 1$ we say the network is an* absolute monarchy.

For example, network in Fig. 2 a) is an absolute monarchy, and network in Fig. 2 b) is an oligarchic network of order 2 for all ϵ. Zachary network is an oligarchic network of order 33 for ϵ =0.001. Facebook network at Caltech is a democratic network for $\epsilon \leq 0.7$.

5.3. PPR and linking strategies

We define the Best Potential Friend of a node, BPF, to be the node that when linked to i, provides the highest increase in the PR of i. That is:

Definition 5. *Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be the initial graph. Let $\pi_i(\mathcal{G})$ denote the i component of the PPR for some $\mathbf{v}$. Given $i \in \mathcal{N}$, let: $\mathcal{P}(i) = \{j \in \mathcal{N} : i \neq j, (j,i) \notin \mathcal{E}\}$, the set of nodes that are not linked to i. $\widehat{\mathcal{E}}(i,j) = \mathcal{E} \cup \{(j,i)\}$, with $j \in \mathcal{P}(i)$, the initial set of edges plus a new edge from j to i. And let $\widehat{\mathcal{G}}(i,j) = (\mathcal{N}, \widehat{\mathcal{E}}(i,j))$. Then we say that $k \in \mathcal{N}$ is a $BPF(i)$ if the following conditions hold: 1) $\pi_i(\widehat{\mathcal{G}}(i,k)) = max_{j\in\mathcal{P}(i)} \pi_i(\widehat{\mathcal{G}}(i,j))$. 2) $\pi_i(\widehat{\mathcal{G}}(i,k)) \geq \pi_i(\mathcal{G})$.*

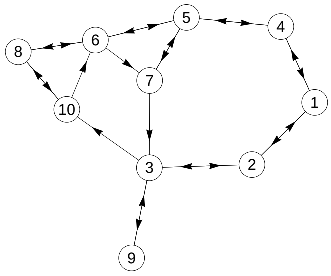


Figure 3: Node 8 is the basic leader

Consider the graph in Fig, 3. The *basic leader* is node 8. An easy computation shows that node 9 cannot win for any ϵ. For $\epsilon \leq 0.84$ there are more than one leader. Focusing on node 6, we obtain that for $\epsilon \leq 0.68$ node 6 can be a leader. Note that $\mathcal{P}(6) = \{1, 2, 3, 4, 7, 9, 10\}$ and $\widehat{\mathcal{G}}(6, 1)$ is the initial graph plus the new link that goes from 1 to 6. An easy computation, taking $\mathbf{v} = \mathbf{e}/n$, shows that in the initial graph we have $\pi_6(\mathcal{G}) = 0.1071$. The results corresponding to the resulting graphs when adding a new link to node 6 are shown in Table 8. From this table we conclude that $BPF(6) = 3$. When using $\mathbf{v} = \mathbf{v}_6(\epsilon = 0.5)$ then node 6 is the winner. Which is the BPF of node 8? We have: $\mathcal{P}(8) = \mathcal{G} - \{6, 10\}$. An easy computation shows that, when using $\mathbf{v} = \mathbf{v}_6(\epsilon = 0.5)$ the BPF of node 8 is node 5. It fact in this new graph node 8 is again the winner.

j	1	2	3	4	7	9	10
$\pi_6(\widehat{\mathcal{G}}(6, j))$	0.1343	0.1351	0.1355	0.1315	0.1324	0.1301	0.1294

Table 8: Computations for BPF(6) for the graph of Fig.3 using *basic PR*.

6. Conclusions

We have presented a theoretical framework that allows classifying users in SNSs. This method allows to modify the usual classification made with the PR with the standard personalization vector; i.e., the *basic PR*. Our model uses PPR and therefore shares the computational and feasibility features of the PR method. We have shown how to analyze the Leadership group using a single parameter ϵ. We have introduced the measure $\nu_L(\epsilon)$ that can be used as a centrality measure. $\nu_L(\epsilon)$ can also be used to distinguish communities. We also have introduced the concept of probability to be a leader. In particular, we have seen that the probability that a node wins could be greater than the probability assigned to the *basic leader* to win. We have remarked the differences between the *basic PR* and the PPR as computed in our model. We have applied the method to some toy graphs and to two real networks. We also have introduced a classification of networks based in our methodology. We have introduced the concept of best potential friend which is related with the optimal linking problem. In this line we address our future research.

Acknowledgments

This work is supported by Spanish DGI grant MTM2010-18674.

References

[1] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Stanford Digital Library Technologies Project, 1999.

[2] J. Lin, PageRank without hyperlinks: Reranking with PubMed related article networks for biomedical text retrieval, BMC Bioinformatics, 9 (270) (2008) 1–12.

[3] X. Yin, X., J. X. Huang, Z. Liet, Mining and modeling linkage information from citation context for improving biomedical literature retrieval. Information Processing and Management 47 (1) (2011) 53–67.

[4] S. Allesina, M. Pascual. Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions?, PLoS Computational Biology, (5) 9 (2009) e1000494.

[5] M. Herrera, J. Gutirrez-Prez, J. Izquierdo, R. Prez-Garca, Ajustes en el modelo PageRank de Google para el estudio de la importancia relativa de los nodos de la red de abastecimiento. in: Proceedings of X Iberoamerican Seminar SERE, (2011).

[6] F. Radicchi, Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. PLoS ONE, 6(2) (2011) e17249.

[7] A. N. Langville, C. D. Meyer. Google's Pagerank and Beyond: The Science of Search Engine Rankings, Princeton University Press, 2006.

[8] T. Haveliwala, S. Kamvar, G. Jeh. An Analytical Comparison of Approaches to Personalizing PageRank, Technical Report. Stanford. 2003.

[9] T. H. Haveliwala, Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search, IEEE Transactions on knowledge and data engineering, vol. 15, No. 4. 2003.

[10] G. Jeh, J. Widow, Scaling personalized web search, Technical Report, Standford University, 2002.

[11] H. Tong,C. Faloutsos, J-Y. Pan, Fast Random Walk with Restart and Its Applications, In ICDM '06 Proceedings of the Sixth International Conference on Data Mining IEEE Computer Society Washington, DC, USA, 2006.

[12] K. Avrachenkov, N. Litvak, D. A. Nemirovsky, E. Smirnova, M. Sokol, Monte Carlo Methods for Top-k Personalized PageRank Lists and Name Disambiguation, Research Report N RR-7367, INRIA, 2010.

[13] A. Vattani, D. Chakrabarti, M. Gurevich, Preserving Personalized Pagerank in Subgraphs, in: Proceedings of the 28th Int. Conf. on Machine Learning, Bellevue, WA, USA, 2011.

[14] F. Pedroche, Competitivity Groups on Social Network Sites, Mathematical and Computer Modelling, 52 (2010), p. 1052-1057.

[15] P. G. Constantine, D. F. Gleich, Random Alpha PageRank, Internet Math 6 (2) (2009) 189–236.

[16] F. Pedroche, Analysis of Competitivity Groups on Social Network Sites using PageRank, in: Proc. Seventh Int. Conf. on Eng. Comp. Tech., Civil-Comp Press, paper 28, 2010.

[17] W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research, 33 (1977), 452–473.

[18] A. L. Traud, E. D. Kelsic, P. J. Mucha, M. A. Porter, Community structure in online collegiate social networks, submitted to SIAM Review, preprint, arXiv.0809.0690.