

Document downloaded from:

<http://hdl.handle.net/10251/53825>

This paper must be cited as:

Pedroche Sánchez, F. (2012). A model to classify users of social networks based on PageRank. *International Journal of Bifurcation and Chaos*. 22(7):1-14.
doi:10.1142/S0218127412501623.



The final publication is available at

<http://www.worldscientific.com/doi/abs/10.1142/S0218127412501623>

Copyright World Scientific Publishing

Additional Information

A MODEL TO CLASSIFY USERS OF SOCIAL NETWORKS BASED ON PAGERANK

FRANCISCO PEDROCHE*

*Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València
Camí de Vera s/n. Valencia, 46022, Spain
pedroche@mat.upv.es*

In this paper we present a model to classify users of Social Networks. In particular, we focus on Social Network Sites. The model is based on the PageRank algorithm. We use the personalization vector to bias the PageRank to some users. We give an explicit expression of the personalization vector that allows the introduction of some typical features of the users of SNSs. We describe the model as a seven step process. We illustrate the applicability of the model with two examples. One example is based on real links of a Facebook network. We also indicate how to take into account real actions of Facebook users to implement the model.

Keywords: Google matrix, PageRank, Link Analysis, Social Networking, Facebook

1. Introduction

Some recent news show the great impact of Social Network Sites nowadays. For example:

- Spanish multi-national company *Telefonica* has recently bought the 90% of the Social Network Site Tuenti for 70 million euros (EL PAIS, August 4, 2010). This SNS was created in 2006 by Zaryn Dentzel.
- Facebook could be the biggest new technology IPO (initial public offering) since Google's launching in 2004 (Financial Times, August 24, 2010). Facebook was created by Mark Zuckerberg in 2004.
- Mobile Social Networking is becoming a daily activity for many people: the number of people using this access on the Internet more than doubled from January 2008 to January 2009 (ComScore.com, March 16, 2009).

The great spread of the online Social Networks in the last years has generated a huge quantity of studies related to this phenomenon. Recently, some researches claimed that a new science was born. This science was called Computational Social Science [Lazer *et al*, 2009]. Some works review the details of this vast field of research from the point of view of *Complex Networks*. See, for example [Boccaletti *et al*, 2006], [Easley & Kleinberg, 2010], [Dorogovtsev, 2010], [Estrada *et al*, 2010] or [Newman, 2010].

Some active research lines associated with SNSs are visualization of networks [Henry *et al*, 2007] (see touchgraph.com or <http://thomas-fletcher.com/friendwheel/> for visualization of Facebook), network formation [Mayer & Puller, 2008], link prediction [Liben-Nowell & Kleinberg, 2003], community analysis [Fortunato, 2010], center-piece subgraphs [Tong & Faloutsos, 2006], networks and matrix functions [Estrada & Higham, 2010], privacy [Maximilien *et al*, 2009], recommender systems [Melville & Sindhvani, 2010], reputation [De Kerchove & Van Dooren, 2009], [Sabater & Sierra, 2005], leadership [Criado *et al*, 2007],

*ETSEE-UPV. Camí de Vera s/n. 46022. Valencia. Spain. pedroche@mat.upv.es

synchronization [Arenas *et al.*, 2008], human behaviour [De Choudhury *et al.*, 2010], tie strength [Gilbert & Karahalios, 2006], small-world experiments [Goel *et al.*, 2009], [Leskovec & Horvitz, 2008], etc.

One of the actual features of the SNSs is that they combine many media and many communication technologies (webmail, images, videos, discussion groups, search engine, etc). Some call SNSs as Integrated Social Networking Sites [Fuchs, 2009]. We also note that Online Networking can be accessed by fixed media or by mobile media. This had led to the terms Online Social Networks (ONS) and Mobile Social Networks (MSN).

The administrators of an SNS have some practical reasons to analyze SNSs. For example, community analysis helps to group users that have similar tastes and therefore can be classified as consumers or reporters of certain type of products; this helps techniques such as personalized advertising and study of new product developments; see, for example, [Lawrence *et al.*, 2010] for marketing strategies in social media. The administrators want users who show up daily, who engage other people, users that help create community.

Focusing on Facebook, some recent research has treated a great amount of topics. In [Traud *et al.*, 2008] five Facebook networks corresponding to five American universities are analyzed. The study uses data from 2005. They consider only ties between students at the same institution. Thereafter they compare the resulting community structure of each network. The authors correlate the belonging to a community with some features of the users such as residence, career, class year and high school. In [Kostakos, 2009] a comparison between presence relationships (i.e., face-to-face relationships or spatial relationships) and online relationships (also called transpatial) is shown. They use Bluetooth technology to track spatial relationships, while using Facebook to track transpatial relationships. The authors show that there exist very similar structural features between online and physical social networks. In [Gjoka *et al.*, 2009] it is shown a method to sample users of Facebook obtaining a uniformly distributed random sample of nodes. They obtain a sample of nearly 1 million users. They obtain that Facebook has a power-law degree distribution¹ separated in two zones, one using an exponent $\gamma = 1.32$, for $1 \leq k \leq 300$, and the other with an exponent $\gamma = 3.38$, for $300 \leq k \leq 5000$. The authors plan to make publicly available this sample. In [Golder *et al.*, 2007] messages sent by 4 million of users of Facebook are analyzed. They conclude that messages are mostly sent to friends, but most friends do not receive messages. They propose that messaging is a more reliable measure of Facebook activity than friends links, since you have to invest more time to send messages than to have friends in an online platform. The number of messages sent in function of the number of users sending that amount of messages is of the form $n^{-\alpha n^\beta}$, i.e., a small number of users send a huge number of messages. In [Lee *et al.*, 2010] a community assignment algorithm valid for overlapping communities is shown. They apply the algorithm to synthetic networks and to the Facebook network given in [Traud *et al.*, 2008]. They claim that the usual Facebook network shows overlapping communities. In [Vasalou *et al.*, 2010] the authors study reasons people argue to use Facebook. They analyze the activity of 423 Facebook users from 5 countries. They found that one of the main reasons to use Facebook is *social searching*: finding out what friends are doing, reconnecting with people, finding and contacting people, etc. They also remark that Facebook help people connect to their offline social networks. *Social browsing* (viewing other people's friends, looking at other profiles, browsing your friends' friends, etc.) occurs at higher level in French and Italian users than in US users. In [Mayer & Puller, 2008] data from 10 public and private US universities are analyzed. The data are from Facebook users, in 2005. The authors focus on the motivations of tie strength [Granovetter, 1973]. Features such same race, similar characteristics of students, same political orientations, make the formation of a friendship more likely. They concluded that very few friendships appear to be only online friendships. A model of social network formation is also presented. In [Gilbert & Karahalios, 2006] the authors analyze seven dimensions related to tie strength: intensity, intimacy, duration, reciprocal services, structural, emotional support and social distance. They enrolled 35 participants to rate the strength of their Facebook friends, obtaining data about 2000 ties. They relate these seven dimensions with 74 Facebook variables. They concluded that the dimensions that most affect the tie strength are intimacy (32.8 %) and intensity (19.7%), follow by duration (16.5 %). As a result, they presented a model

¹The probability of having k friends is given by $P(k) \sim k^{-\gamma}$.

to predict tie strength. In [Lewis *et al*, 2008] a network dataset based on Facebook users from universities is shown. The population is made of 1640 students². They show that users that connect with one another share some sociodemographic traits. They also show that Facebook is used to maintain or reinforce existing offline relationships. In [Paek *et al*, 2010] an study based on 24 Facebook users is presented. They collect 4989 newsfeed ratings and 4238 friends ratings. One of the goals is to predict tie strength by means of information displayed on, or associated with, newsfeed posts. A significant correlation between newsfeed rating and friend rating is obtained. In [Song *et al*, 2010] properties of message threads are analyzed to gain insight about tie strength. 1783 postings and 2575 comments from 281 Facebook users are analyzed; they present a model to extract social relations from message threads.

In this paper we present a model to classify users of SNSs. The model is based on the PageRank algorithm [Page *et al*, 1999], [Langville & Meyer, 2006]. We use the so-called *personalization vector* that was originally introduced to bias the PageRank to personal preferences of the users. When we first analyzed the effect of the personalization vector in the PageRank vector we introduced the concept of Competitivity groups and Leadership groups; see [Pedroche, 2010a] for details.

PageRank is a well known centrality measure; see, for example [Newman, 2010] for a relation between PageRank and other centrality measures. From the point of view of Markov chains, PageRank can be considered as a stationary state of a Markov chain with restart [Boldi *et al*, 2009], [Tong *et al*, 2006]. There are some analytical expressions that relate the PageRank vector with the personalization vector but they are not of practical use for our purposes: some require the computation of a huge Jordan form [Serra-Capizzano, 2005]; others require the computation of a huge inverse matrix [Boldi *et al*, 2009]. To our knowledge some few works have used the personalization vector as a practical tool [Haveliwala, 2003], [Jeh & Widow, 2002]; neither of these works are intended to classify users in Social Networks in the form that we present. Regarding computational issues, some numerical methods have been proposed to optimize the computation of the PageRank vector; see [Langville & Meyer, 2006], [Avrachenkov *et al*, 2010], [Pedroche, 2007]. In the present stage we are not concern about the computational issues of the presented model; we leave these details as a future work.

The structure of the paper is as follows. In section 2 we recall some definitions. In section 3 we present a model to rank users based on PageRank. In section 4 we apply the model to a small graph of four nodes. The simplicity of the example allows to show some complete computations. This example illustrates how the model works. In section 5 we apply the model to a network of Facebook users. Some capabilities of the model are shown. In section 6 we show how to apply the model to a network of Facebook users when we may access to all the data of the users, i.e., as we were administrators of Facebook. Here we show what features of the users can be implemented in the model. In section 7 we give some conclusions.

2. Definitions

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be the directed graph representing a Social Network Site. Users are represented by the set of nodes $\mathcal{N} = \{1, 2, \dots, n\}$ and the hyperlinks are represented by the set of directed links $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$. The link represented by the pair (i, j) belongs to the set \mathcal{E} if and only if there exists a hyperlink connecting node i to node j .

A link (i, j) is said to be an *outlink* for node i and an *inlink* for node j . We denote d_i the number of outlinks of a node i ; this is usually called the *outdegree* of node i . In an SNS we assume that each node has at least one outlink; i.e., there are no *dangling nodes*. This is a natural assumption: in an SNS each user has, at least, one friend. Therefore we have $d_i \neq 0$ for all $i \in \mathcal{N}$.

We use the PageRank vector [Page *et al*, 1999] 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

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

Let $0 < \alpha < 1$ be the so-called damping factor (that we use as $\alpha = 0.85$). Let $\mathbf{e} \in \mathbb{R}^{n \times 1}$ be the vector

²<http://dvn.iq.harvard.edu/dvn/dv/t3>

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 $\mathbf{v}^T \mathbf{e} = 1$. Then the Google matrix is defined as

$$G = \alpha P + (1 - \alpha) \mathbf{e} \mathbf{v}^T, \quad (2)$$

and is an stochastic and primitive (irreducible and aperiodic) matrix [Langville & Meyer, 2006]. The PageRank vector is defined as the unique left Perron vector of G

$$\pi^T = \pi^T G, \quad (3)$$

with $\pi^T \mathbf{e} = 1$. Denoting \mathbf{e}_i the i th column of the identity matrix of order n , the PageRank of a node i is $\pi_i = \pi^T \mathbf{e}_i$.

Since the PageRank vector depends on the personalization vector, we write $\pi(\mathbf{v})$. The present model prescribes the explicit expression of the personalization vector; see section 3.

We call basic PageRank, and denote it by *basicPR* to the vector $\pi(\mathbf{e}/n)$. We recall the following definitions from [Pedroche, 2010a].

Definition 2.1. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, let $0 < \epsilon < 1$ and let $\mathbf{v}_i = [v_{ij}] \in \mathbb{R}^{n \times 1} : v_{ii} = 1 - \epsilon, v_{ij} = \epsilon/(n - 1)$ if $i \neq j$. For each $i \in \mathcal{N}$, let

$$PR_i = \pi(\mathbf{v}_i). \quad (4)$$

and we denote as $(PR_i)_j$ the j th entry of PR_i .

Definition 2.2. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and $0 < \epsilon < 1$, for each node $j \in \mathcal{N}$ we define the *Competitiveness interval* $S_C(j)$ as

$$S_C(j) = [\min_{i \in \mathcal{N}} (PR_i)_j, \max_{i \in \mathcal{N}} (PR_i)_j]. \quad (5)$$

Definition 2.3. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, and $0 < \epsilon < 1$ we define the *Competitiveness matrix* of the graph, $C = [C_{ji}] \in \mathbb{R}^{n \times 2}$, as follows

$$C_{j,1} = \min_{i \in \mathcal{N}} (PR_i)_j, \quad C_{j,2} = \max_{i \in \mathcal{N}} (PR_i)_j. \quad (6)$$

Definition 2.4. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, and $0 < \epsilon < 1$, a *Competitiveness group* is a subset of \mathcal{N} . Nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ belong to the same *Competitiveness group* if $S_C(i) \cap S_C(j) \neq \emptyset$.

Definition 2.5. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, and $0 < \epsilon < 1$, the *Leadership group* is a subset of \mathcal{N} . Node $j \in \mathcal{N}$ belongs to the *Leadership group* if, for some $i \in \mathcal{N}$ it holds that $(PR_i)_j \geq (PR_i)_k$ for all $k \neq j$. i.e. for some personalization vector \mathbf{v}_i node j has the greatest PageRank.

It can be shown, see [Pedroche, 2010b], that in the practice we have $\epsilon \leq (n - 1)/n$. We recall from [Pedroche, 2010b] the following definition.

Definition 2.6. Given $n > 1$ and $\epsilon \leq (n - 1)/n$, let:

$$\phi(\epsilon, n) = \frac{(n - 1)(1 - \epsilon)}{\epsilon} \geq 1 \quad (7)$$

Note that ϕ is a global property of the network. As we see later, this parameter controls the *social mobility* of the nodes. A great value of ϕ means a great social mobility. Note that ϕ depends on the network size while ϵ does not depend on the scale.

3. Description of the model

We now give the fundamental ingredient of the present model. We present a formulation for the personalization vector that agrees in some way with the assumption expressed by [Granovetter, 1973] in the sense that tie strength was a (probably linear) combination of some dimensions. We assume that each component of the personalization vector is a linear combination of some *social competencies* of the nodes.

Definition 3.1. Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, which represents a Social Network, without dangling nodes, we classify the users attending to the PageRank vector $\pi(\mathbf{u}/\|\mathbf{u}\|_1)$, where $\mathbf{u} = [u_i] \in \mathbb{R}^{n \times 1}$ is the *Social Network personalization vector* and is given by:

$$u_i = 1 + (\phi - 1)\mathbf{w}^T \mathbf{s}(i), \quad i \in \mathcal{N} \quad (8)$$

where:

- $\mathbf{w} \in \mathbb{R}^{m \times 1}$, is a nonnegative normalized weight vector, i.e., $\mathbf{w} \geq 0$, $\|\mathbf{w}\|_1 = 1$.
- $\mathbf{s}(i) \in \mathbb{R}^{m \times 1}$ is the *Social Competencies vector* of node i . It is a nonnegative vector, i.e., $\mathbf{s}(i) \geq 0$. But it is not necessarily normalized. The idea is that we share each social competence over the entire net, i.e.: for each competence $j \in \{1, 2, \dots, m\}$ we have $\sum_{i \in \mathcal{N}} s_j(i) = 1$.

Remark 1. Note that when $\phi = 1$ we have $\pi(\mathbf{u}/\|\mathbf{u}\|_1) = \pi(\mathbf{e}/n) = \text{basicPR}$.

Remark 2. It is easy to show that $u_i \in [1, \phi]$; each node has a minimum of 1 in the corresponding entry of \mathbf{u} only by the fact that the node is a member of the Social Network. The maximum value of u_i happens when a node has the maximum of $\mathbf{s}(i)$, i.e.,

$$\max(\mathbf{s}(i)) = \mathbf{e} \longrightarrow \max(u_i) = 1 + (\phi - 1)\mathbf{w}^T \mathbf{e} = \phi \quad (9)$$

Note also that when a node i has $u_i = \phi$, then the rest of the components of the personalization vector are lower (strictly) than ϕ .

Remark 3. Note that in this model arbitrary values of the personalization vector are not allowed. The model only uses the *Social Network personalization vector*, \mathbf{u} , given by Eq. (8).

Table 1. Steps for classifying users using biased PageRank

1)	For some $\epsilon \in (0, \frac{n-1}{n})$ compute the resulting Leadership groups and Competitivity groups.
2)	Decide how many LG and CG you want: select ϵ , compute $\phi(\epsilon, n)$: it controls the social mobility.
3)	Decide how many user competencies you use: select m .
4)	Decide which features of the users are related to each competence: construct $\mathbf{s}(i)$, $\forall i \in \mathcal{N}$.
5)	Decide a global weight for each competence: construct \mathbf{w} .
6)	Construct the Social network personalization vector: \mathbf{u} from Eq. (8).
7)	Compute $\pi(\mathbf{u}/\ \mathbf{u}\)$ which, after normalization, gives the ranking.

The model consists of the following steps (see Table 1). In the first step one has to characterize the network in function of the parameter ϵ . This analysis of the groups in function of ϵ holds when the topology of the network does not change. That is, the behavior of the LG and CG as a function of ϵ is a property of the network, not of the users. The values of ϵ to make this analysis could be some values (say ten) distributed in the interval $(0, \frac{n-1}{n})$. We give two examples below, see also [Pedroche, 2010b] and [Pedroche, 2010c]. In the second step we must decide the social mobility that we shall allow to the users. A greater value of ϕ (i.e., a small value of ϵ) shall allow a high effect of the personalization vector. In step three we must select the number of social competencies that we incorporate in the personalization vector. In the two examples that we show below we recommend to use only three social competencies which we call *popularity*, *recentness* and *activity*. In section 6 we indicate how to relate these competencies to real features of the users of Facebook. In step four we construct the vector of social competencies for each node.

This information is, of course, a private one. In a general way only the managers of the Social Network may develop a formula to take into account some specific features or actions of the users to construct each vector $\mathbf{s}(i)$. Note that the vectors $\mathbf{s}(i)$ may show a dynamic behavior. Therefore the Social Network personalization vector given by Eq. (8) may show a dynamical behavior. In step five one has to decide the weight given to each social competence. This weight is the same for all the nodes. The utility of this vector is that different Social Networks may promote users that are better in different competencies (e.g., *to be active* versus *to be up of date*). In step six one have all the ingredients to compute the Social Network personalization vector given by Eq. (8). Note that this vector is not normalized (i.e., it is not a distribution probability vector). After that, in the final seven step one computes the PageRank vector using $\mathbf{u}/\|\mathbf{u}\|_1$ as personalization vector. The resulting PageRank vector, after normalization, gives the ranking. When the graph does not change one can go to step three to change the inputs of the model obtaining different classifications.

In the following sections we give examples of how to apply this model.

4. A simple example 4×4

In this example we show the ingredients of the model. Let us consider the graph shown in Figure 1. Since this graph is very small we can give most of the details of the computations.

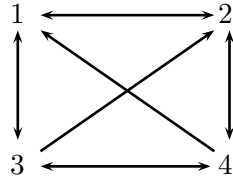


Fig. 1. Example 4×4

The row stochastic matrix, Eq. (1), associated to this graph is:

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} \quad (10)$$

constructing the matrix G from Eq. (2) and computing the PageRang vector from Eq. (3) using the personalization vector $\mathbf{v} = \mathbf{e}/n = \frac{1}{4}[1, 1, 1, 1]^T$ one can obtain that the *basicPR* is:

$$\text{basicPR} = \pi\left(\frac{1}{4}[1, 1, 1, 1]\right) = \begin{bmatrix} 0.281 \\ 0.281 \\ 0.219 \\ 0.219 \end{bmatrix}. \quad (11)$$

We note that in this small SNS -of only four friends- there are two users who are *important* in the sense of having great PageRank, and two other users that are less important. Let us imagine that user number 4 is a very active user. How can we enhance its PageRank? Using the present model we can use an *appropriate* personalization vector in such a way that the corresponding PageRank shows a bigger component for node number 4. This *appropriate* personalization vector is of the form defined in Eq. (8). We see that this personalization vector depends on ϕ , and therefore, on ϵ .

Before using Eq. (8) to construct the personalization vector, we must analyze the effect of the parameter ϵ ; this is the step 1) of the model, as shown in Table 1. This parameter is a global one. It is a property that we assign to the entire network. It can be seen as a parameter that controls the *social mobility* of the

nodes. Depending on ϵ we allow the nodes to move from a situation of low PageRank to a situation of high PageRank by using the personalization vector. To study to what extent we can enhance the importance of a node using the personalization vector we use the personalization vectors of the form defined in Definition 2.1. We see that these personalization vectors depend on a parameter ϵ . The greater ϵ is, the lower is the enhancement that we are giving to node i when using \mathbf{v}_i .

When analyzing the behavior of the PageRank of the nodes in function of ϵ one sees that some nodes compete in the same zone of the PageRank. This is the reason for defining the Competitiveness groups in Definition 2.4. Another fact that arises is that only some nodes can win for some values of ϵ . This fact leads to the definition of Leadership group in Definition 2.5. Some details about Competitiveness groups and Leadership groups have been studied in [Pedroche, 2010c] and [Pedroche, 2010b]. We now analyze these concepts for this example 4×4 .

Table 2. Competitiveness groups and Leadership group for the Example 4×4

ϵ	Competitiveness groups	Leadership group
0.01	{1, 2, 3, 4}	{1, 2, 3, 4}
0.1	{1, 2, 3, 4}	{1, 2, 3, 4}
0.3	{1, 2, 3, 4}	{1, 2, 3, 4}
0.37	{1, 2, 3, 4}	{1, 2, 3, 4}
0.38	{1, 2}, {3, 4}	{1, 2}
0.4	{1, 2}, {3, 4}	{1, 2}
0.6	{1, 2}, {3, 4}	{1, 2}
0.75	{1, 2}, {3, 4}	{1, 2}

In Table 2, we give some values of the Competitiveness groups and Leadership groups for some values of ϵ when using the small graph of Figure 1. From Table 2 we see that for $\epsilon \leq 0.37$ the network behaves as a *democratic network*: any node can be the leader (all the nodes belong to the Leadership group) and all the nodes compete in the same Competitiveness group. When $\epsilon \geq 0.38$ we see that the network splits in two Competitiveness groups, one composed of the most important people -the VIP zone- formed by nodes 1 and 2, and the second division group in where compete nodes 3 and 4. In this case the leaders are the VIP group, {1, 2}.

Table 3. Competitiveness matrix for Example 4×4 .

ϵ	Competitiveness Matrix
0.37	$\begin{bmatrix} 0.2644 & 0.3243 \\ 0.2644 & 0.3243 \\ 0.1935 & 0.2652 \\ 0.1935 & 0.2652 \end{bmatrix}$
0.38	$\begin{bmatrix} 0.2648 & 0.3232 \\ 0.2648 & 0.3232 \\ 0.1942 & 0.2640 \\ 0.1942 & 0.2640 \end{bmatrix}$

To understand the idea of Competitiveness group and Leadership group we show the Competitiveness

matrix (see Definition 2.3) corresponding to two values of ϵ in Table 3. When $\epsilon = 0.37$ there is an interval $(0.2644, 0.2652)$ in which any node can have its PageRank. There is therefore only one Competitivity group. For $\epsilon = 0.38$ we see that there is a gap in the values of the possible PageRanks and the nodes split in two groups; in one group they compete to have PageRank in the range $(0.2648, 0.3232)$. In the second group, nodes 3 and 4 compete in the range $(0.1942, 0.2640)$.

Table 4. PR_i for $\epsilon = 0.37$, Example 4×4 .

PR_1	PR_2	PR_3	PR_4
$\begin{bmatrix} 0.3243 \\ 0.2710 \\ 0.2112 \\ 0.1935 \end{bmatrix}$	$\begin{bmatrix} 0.2710 \\ 0.3243 \\ 0.1935 \\ 0.2112 \end{bmatrix}$	$\begin{bmatrix} 0.2644 \\ 0.2644 \\ 0.2652 \\ 0.2060 \end{bmatrix}$	$\begin{bmatrix} 0.2644 \\ 0.2644 \\ 0.2060 \\ 0.2652 \end{bmatrix}$

In Table 4 we see that for $\epsilon = 0.37$ any node can be the winner -when using the corresponding \mathbf{v}_i given by Definition 2.1. In Table 5 we see that for $\epsilon = 0.38$ the only nodes that are winners are node 1 (when using $\mathbf{v}_1, \mathbf{v}_3$, or \mathbf{v}_4) and node 2 (when using $\mathbf{v}_2, \mathbf{v}_3$, or \mathbf{v}_4).

Table 5. PR_i for $\epsilon = 0.38$, example 4×4 .

PR_1	PR_2	PR_3	PR_4
$\begin{bmatrix} 0.3232 \\ 0.2713 \\ 0.2114 \\ 0.1942 \end{bmatrix}$	$\begin{bmatrix} 0.2713 \\ 0.3232 \\ 0.1942 \\ 0.2114 \end{bmatrix}$	$\begin{bmatrix} 0.2648 \\ 0.2648 \\ 0.2640 \\ 0.2064 \end{bmatrix}$	$\begin{bmatrix} 0.2648 \\ 0.2648 \\ 0.2064 \\ 0.2640 \end{bmatrix}$

Once that we have analyzed the parameter ϵ for this small network we proceed to step 2) of the model, consisting in the election of the parameter ϵ . To that end we must answer the following question: What kind of network we want? If we want a network where everyone can be a leader we should choose $\epsilon \leq 0.37$. If we want a network in which only some nodes can be the winners we should choose $\epsilon \geq 0.38$. In the following we give examples of both situations using the graph 4×4 . Our aim is to illustrate the effect of $\mathbf{w}, \mathbf{s}(i)$ and ϵ . To that end we shall assume certain values for the vectors \mathbf{w} and $\mathbf{s}(i)$.

We now go to step 3) of the model. We have seen in definition 3.1 that we consider m social competencies for each node of the graph. Let us take, for convenience, $m = 3$. Then we are going to consider three social competencies for each node, i.e.,

$$\mathbf{s}(i) = [s_1(i), s_2(i), s_3(i)]^T, \quad i \in \{1, 2, 3, 4\} \quad (12)$$

These competencies could measure the following:

- $s_1(i)$ measures the popularity of node i (e.g., it can be related to the number of friends of node i).
- $s_2(i)$ measures the recentness of node i .
- $s_3(i)$ measures the activity of node i .

We are not interested now in the actual features that we use to construct this values of $s_j(i)$. Let us assume that we have computed in some prescribed form these values; this is step 4) of the model. These values are given in Table 6.

In Table 7 we show the ranking with the present model when taking $\mathbf{w}^T = [1, 1, 30]/\|[1, 1, 30]\|$, which means that we are giving more weight to the feature ‘‘activity’’ than to the rest of features; we are in the

Table 6. Assumed vectors $\mathbf{s}(i)$ for Example 4×4 .

Feature	$\mathbf{s}(1)$	$\mathbf{s}(2)$	$\mathbf{s}(3)$	$\mathbf{s}(4)$
Popularity	0.1	0.1	0.5	0.3
Recentness	0.3	0.1	0.5	0.1
Activity	0.1	0.1	0.1	0.7

Table 7. Example 4×4 . Node ranking with $\mathbf{w}^T = [1, 1, 30]/\|[1, 1, 30]\|$, $\epsilon = 0.01$ and $\mathbf{s}(i)$ given in Table 6.

$\mathbf{u}/\ \mathbf{u}\ $	$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
$\begin{bmatrix} 0.1082 \\ 0.1020 \\ 0.1267 \\ 0.6632 \end{bmatrix}$	$\begin{bmatrix} 0.2623 \\ 0.2617 \\ 0.2068 \\ 0.2693 \end{bmatrix}$	$\begin{matrix} 4 \\ 1 \\ 2 \\ 3 \end{matrix}$

Table 8. Example 4×4 . Node ranking with $\mathbf{w}^T = [1, 1, 30]/\|[1, 1, 30]\|$, $\epsilon = 0.6$ and $\mathbf{s}(i)$ given in Table 6.

$\mathbf{u}/\ \mathbf{u}\ $	$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
$\begin{bmatrix} 0.2213 \\ 0.2200 \\ 0.2250 \\ 0.3338 \end{bmatrix}$	$\begin{bmatrix} 0.2772 \\ 0.2771 \\ 0.2165 \\ 0.2292 \end{bmatrix}$	$\begin{matrix} 1 \\ 2 \\ 4 \\ 3 \end{matrix}$

step 5) of the model. We take $\epsilon = 0.01$ which means that we allow a great social mobility, and $\mathbf{s}(i)$ given in Table 6. The personalization vector \mathbf{u} is computed according Eq. (8); this is step 6) of the model. In Table 8 we show the results of the step 7) of the model: the ranking obtained computing the PageRank with the constructed personalization vector. Note that since node 4 is a very active user and we are promoting these kind of users we have that node 4 is the winner. Compare these results with the ones shown in Table 8. In this table, we take $\epsilon = 0.6$ which means that we do not allow a great social mobility and node 4 does not reach to the top despite being a very active user.

Table 9. Assumed vectors $\mathbf{s}(i)$

feature	$\mathbf{s}(1)$	$\mathbf{s}(2)$	$\mathbf{s}(3)$	$\mathbf{s}(4)$
Popularity	0.1	0.1	0.7	0.1
Recentness	0.1	0.1	0.5	0.3
Activity	0.1	0.1	0.4	0.4

Let us now assume that the $\mathbf{s}(i)$ vectors are those giving in Table 9. In Table 10 and Table 11 we show the effect of varying vector \mathbf{w} . In both cases we are using $\epsilon = 0.01$, which means great social mobility. Note that a slight variation in the weight vector results in a different classification of the nodes. When giving a *sufficient* weight to the dimension “popularity” we have that node 3 is the winner, since it is a very popular node as it is shown in the vector $\mathbf{s}(3)$ in Table 9.

Table 10. Example 4×4 . Node ranking with $\mathbf{w}^T = [3, 1, 1]/\|[3, 1, 1]\|$, $\epsilon = 0.01$ and $\mathbf{s}(i)$ given in Table 9.

$\mathbf{u}/\ \mathbf{u}\ $	$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
$\begin{bmatrix} 0.1020 \\ 0.1020 \\ 0.5953 \\ 0.2007 \end{bmatrix}$	$\begin{bmatrix} 0.2616 \\ 0.2616 \\ 0.2615 \\ 0.2154 \end{bmatrix}$	1 2 3 4

Table 11. Example 4×4 . Node ranking with $\mathbf{w}^T = [4, 1, 1]/\|[4, 1, 1]\|$, $\epsilon = 0.01$ and $\mathbf{s}(i)$ given in Table 9.

$\mathbf{u}/\ \mathbf{u}\ $	$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
$\begin{bmatrix} 0.1020 \\ 0.1020 \\ 0.6118 \\ 0.1842 \end{bmatrix}$	$\begin{bmatrix} 0.2616 \\ 0.2616 \\ 0.2634 \\ 0.2134 \end{bmatrix}$	3 2 1 4

With this small example 4×4 we have shown some of the capabilities of the model. We bear in mind that the actual application of the present model is aimed to greater SNSs. In the following section we address the problem of apply the model to a real network from an SNS.

5. Example using a network of Facebook

In this section we apply the proposed model to a graph corresponding to a real social network. We use the data of a Facebook network from a date in September of 2005. This network corresponds to users of Facebook from the California Institute of Technology; See [Traud *et al*, 2008] for details³. The adjacency matrix is of order $n = 769$.

In Table 12 we show the ranking of the first 20 nodes when using the basicPR. We could ask the following question. Let us assume that node 19 is a very active user. Can we promote it to the first position with the present model?. As we have seen in the example 4×4 , to answer this question we need first to know the effect of ϵ in the classification of the nodes; step 1) of the model.

In Table 13 we show the results for the cardinal of the Leadership group as a function of ϵ . Note that for $\epsilon \leq 0.7$ the network has high social mobility: every node can be the winner. When $\epsilon \geq 0.9$ there are some nodes that cannot win. Our computations show that for $\epsilon = 0.9$ the nodes that do not belong to the Leadership group are those showed in Table 14. In this table we see, for example, that for $\epsilon = 0.9$ node 406 cannot be a winner. To show the capabilities of the proposed model we shall answer the following question: Can node 406 be in the second position in the ranking?

Once that we have certain insight about the effect of ϵ in the structure of the Leadership group, we need to construct the vectors \mathbf{w} and $\mathbf{s}(i)$ to apply the present model.

As in the small example 4×4 we assume here that only $m = 3$ social competencies are to be considered. Therefore, we take the vector $\mathbf{s}(i)$ as in Eq. (12), for $i \in \{1, 2, \dots, 769\}$. We can also assign the same meaning to each competence $s_j(i)$ as in the example 4×4 .

For simplicity, to show how the model works, in order to construct the vectors $\mathbf{s}(i)$ we make the following assumptions:

³Data available at <http://people.maths.ox.ac.uk/porterm/data/facebook5.zip>

Table 12. Node ranking of the first 20 nodes using the basicPR for the network of Facebook at Caltech.

basic PR	Node ranking
0.0067	623
0.0056	207
0.0056	563
0.0050	60
0.0048	405
0.0046	88
0.0043	82
0.0043	411
0.0043	95
0.0043	648
0.0043	9
0.0041	457
0.0041	176
0.0040	72
0.0040	201
0.0039	124
0.0038	279
0.0038	51
0.0038	96
0.0037	19

Table 13. Cardinal of the Leadership group as a function of ϵ for the network of Facebook at Caltech.

ϵ	Cardinal of the Leadership group
0.50	769
0.70	769
0.90	757
0.95	734
0.97	98

Table 14. Nodes that do not belong to the Leadership group for $\epsilon = 0.9$. Network of Facebook at Caltech.

Node:	406	476	494	615	620	661	664	671	710	717	755	769
-------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- The competencies $s_1(i)$ and $s_2(i)$ are homogeneously distributed over all the nodes, i.e.:

$$s_1(i) = s_2(i) = 1/769, \quad i \in \{1, 2, \dots, 769\} \tag{13}$$

- The competence $s_3(i)$ follows a power-law, of the form (see [Clauset *et al.*, 2009]):

$$y(x) = \frac{x^{-\gamma}}{\sum_{n=0}^{\infty} (n + x_{min})^{-\gamma}} \quad (14)$$

with $x_{min} = 1$, and $x \in \{1, 2, \dots, 769\}$. Note that this function is normalized, i.e., $\sum_{x=1}^{\infty} y(x) = 1$.

5.1. Promoting node 19

We assume that node 19 has a very high activity. Since we want $s_3(19) > s_3(j) \forall j \neq 19$, we take:

$$s_3(i) = \begin{cases} y(i + 751) & \text{for } i \in \{1, 2, \dots, 18\} \\ y(i - 18) & \text{for } i \in \{19, 20, \dots, 769\} \end{cases} \quad (15)$$

Table 15. Node ranking, of the first 20 nodes, with $\mathbf{w}^T = [1, 1, 1]/3$, $\epsilon = 0.7$ and $\mathbf{s}(i)$ given by Eq. (13) and Eq. (15) with $\gamma = 2$.

$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
0.0129	19
0.0066	623
0.0057	563
0.0054	207
0.0049	60
0.0047	405
0.0045	88
0.0044	82
0.0043	95
0.0043	411
0.0043	9
0.0043	648
0.0042	176
0.0041	457
0.0040	72
0.0039	124
0.0039	201
0.0039	51
0.0037	20
0.0037	279

In Table 15 we show the node ranking, for the first 20 nodes, with $\mathbf{w}^T = [1, 1, 1]/3$, $\epsilon = 0.7$ and $\mathbf{s}(i)$ given by Eq. (13) and Eq. (15) with $\gamma = 2$. The personalization vector is computed according Definition 3.1. Comparing Table 12 and Table 15 we see that the high value of $s_3(19)$ is enough to make it the winner. Note that since we have use Eq. (15) this means that node 20 has also a great value of s_3 . This is the reason why this node has climbed positions and now it appears in the *top 20*.

5.2. Promoting node 406

Let us imagine that node 406 has the maximum of activity. As before we assume that this competence follows a power law. Since we want $s_3(406) > s_3(j) \forall j \neq 406$, we can take:

$$s_3(i) = \begin{cases} y(i + 364) & \text{for } i \in \{1, 2, \dots, 405\} \\ y(i - 405) & \text{for } i \in \{406, 407, \dots, 769\} \end{cases} \quad (16)$$

If we use the vector $\mathbf{w}^T = [1, 1, 1]/3$ it means that we give the same weight to all the three competencies for the entire network. Our model gives that node 406 is in position 31. Therefore it is outside the top twenty. Nevertheless when we put $\mathbf{w}^T = [1, 1, 3]/5$ we obtain that node 406 arrives at the third position in the ranking, i.e., when we give more weight to the activity, node 406 can be promoted. Note also the movement of node 5 from Table 16 to Table 17. An analysis of the adjacency matrix shows that node 5 is connected with node 406. As a consequence when promoting node 406 it results that node 5 is also promoted.

A deeper analysis shows that putting more weight to w_3 one can achieve that node 406 arrives, at most, at the second position. In fact, with $\mathbf{w}^T = [1, 1, 5]/7$ node 406 arrives at the second position. One can also show that node 406 does not arrive to be the first. In fact, taking $\mathbf{w}^T = [1, 1, 10000]/10002$ node 406 is in second position.

All in all we see that node 406 can be very active node, and we can promote active people like it. But, with this value of ϵ , i.e., with the present policy of social mobility, certain nodes do not reach high positions; To be in a high position one needs a good linkage strategy (as measures by PageRank) and good competence skills (as measured by the personalization vector).

Table 16. Node ranking, of the first 20 nodes, with $\mathbf{w}^T = [1, 1, 1]/3$, $\epsilon = 0.9$ and $\mathbf{s}(i)$ given by Eq. (13) and Eq. (16) with $\gamma = 2$.

$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
0.0066	623
0.0056	207
0.0056	563
0.0053	5
0.0050	60
0.0048	405
0.0045	88
0.0044	411
0.0043	82
0.0042	9
0.0042	648
0.0042	95
0.0041	457
0.0041	176
0.0040	72
0.0040	201
0.0040	124
0.0038	279
0.0038	51
0.0038	96

In this section we have seen the applicability of the model to a real network. We have use invented values of the vectors \mathbf{w} and $\mathbf{s}(i)$. In the next section we give hints on how to compute in a practical way these vectors when one has complete access to all the information of the users of the network.

6. Guidelines to apply the present model to Facebook when having all the data

We have developed a general framework in which m competencies of the users can be considered to obtain a ranking. Nevertheless, as in the previous examples, we think that is enough to consider three dimensions: popularity, recentness and activity. In this section we indicate which Facebook variables could be used to

Table 17. Node ranking with $\mathbf{w}^T = [1, 1, 3]/5$, $\epsilon = 0.9$ and $\mathbf{s}(i)$ given by Eq. (13) and Eq. (16) with $\gamma = 2$.

$PR(\mathbf{u}/\ \mathbf{u}\)$	Node ranking
0.0073	5
0.0066	623
0.0057	406
0.0056	207
0.0055	563
0.0050	60
0.0048	405
0.0044	88
0.0044	411
0.0042	9
0.0042	82
0.0042	648
0.0042	95
0.0042	457
0.0041	176
0.0040	124
0.0039	201
0.0039	72
0.0038	51
0.0038	279

measure these three dimensions. We take into account some results -see, e.g., [Gilbert & Karahalios, 2006], [Vasalou *et al*, 2010] and [Paek *et al*, 2010]- that relate some Facebook variables either with the importance of users or with the importance of ties. Note that we are assuming that we know all the features of all the users; i.e., we assume we access all the private information administrators manage about their users and their relationships.

6.1. *Popularity*

The dimension that we have called *popularity* can be related to the following variables of Facebook⁴:

- Number of friends. We think it is reasonable to use this number as a sign of popularity. This variable has been already proposed and justified as a measure of tie strength in a Facebook network; see [Gilbert & Karahalios, 2006].
- Inbox Messages. The number of messages that a user receives can be a measure of its popularity. This variable is also used and analyzed in [Gilbert & Karahalios, 2006].
- Duration of membership. The antiquity in a relationship is also a measure of a tie strength and we also think it can be used as a sign of popularity. It is logic to assume that old users are well known users and thus popular. This dimension was proposed by [Granovetter, 1973] and analyzed in Facebook by [Gilbert & Karahalios, 2006].
- Reputation. We think reputation (as a measure to trust someone) must affect the popularity of a user. We think that good (bad) reputation must increase (decrease) the popularity of a user. In a Facebook framework we can use as a measure of the reputation of a user the following variables:

- (1) As *good reputation* we can use the number of *likes* that the rest of users have put in some action of the user. This action can be a wall post, a comment in other's wall, a comment in a thread or a comment

⁴We are referring to the variables that at present time the users encounter on the SNS Facebook. Note that years ago users of Facebook used to have friends; nowadays users *like* other users.

in a photograph. Any action susceptible of being voted by other users.

- (2) As *bad reputation* we can consider the number of users that have blocked a user and the number of *flags* of the type *abusive* that other users have denounced.

The reputation of a user can also be related with the concept of positive and negative links; see [Leskovec *et al*, 2010] and the references therein.

Taking into account these features we could formulate an expression (e.g. using a weighted mean) to compute the component $s_1(i)$ for each node i .

6.2. Recentness

This dimension tell us whether a user is up to date or not. It informs whether a user shows up frequently. We remark that nowadays Facebook informs about the *recent activity* of our friends. We propose to measure the *recentness* using the following Facebook variables:

- The inverse of the number of days since the last action. This variable was analyzed in [Gilbert & Karahalios, 2006].
- The number of recent *likes* in any action.
- The number of recent comments.
- The number of recent friends.
- The number of recent sent messages.
- The number of recent *Facebook applications*.

Note that we must update the computations frequently in our model if we want to include some of these variables. Recentness in a relationship, as related with tie strength, was analyzed in [Lin *et al*, 1978].

Considering these features we could formulate an expression to compute the component $s_2(i)$ for each node i .

6.3. Activity

This dimension can be related to many Facebook variables. In fact, it can be computed as the number of *actions* of any type. For example, in [Golder *et al*, 2007] it is shown that messaging is a good measure of Facebook activity. In particular, we propose to use:

- The number of words in comments (on wall, friends statuts, photographs, etc). The wall words was used and analyzed in [Gilbert & Karahalios, 2006] as a measure of activity.
- Thread depth. This can be a measure of the quality of the threads, of the intensity.
- The number of sent messages.
- The number of likes that the user has post
- The number of *shares* that the user has produced.
- The number of *Facebook applications* that the user uses.

This dimension is related with the dimension called *intensity* that [Granovetter, 1973] related with tie strength. Considering these features we could formulate an expression to compute the component $s_3(i)$ for each node i .

All in all we have shown some ideas on how to construct the vector $\mathbf{s}(i)$ when using the model to rank Facebook users.

Regarding the vector \mathbf{w} , it can be constructed following the preferences of the managers. These preferences may change from time to time; i.e. maybe the managers want to search for active people in some instances but maybe they want to search for popular people in other circumstances. We have shown that the present model allow managers to classify users attending to different preferences.

7. Conclusions

A new model to classify users of Social Networks has been shown. We have focused on Social Network Sites. The model is based on the PageRank algorithm. The model uses an explicit expression of the

personalization vector that allows the introduction of some features of the users of SNS. We assume that each component of the personalization vector is a linear combination of some social competencies. We have presented the model as a seven step process. We have shown that the model allows control the social mobility of the nodes depending on their social competencies and on the structure of the network. The model allows a dynamical interpretation. We have shown how to use the model with two examples; one of them uses links of a Facebook network. In the examples we use three social competencies that we denote as popularity, recentness and activity. We also indicate how to take into account real actions of Facebook users to implement the model. We leave theoretical analysis and numerical recommendations to implement the model in large SNSs as future work.

Acknowledgments

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

References

- Arenas, A., Díaz-Guilera, Kurths, J., Moreno, Y. & Zhou, C. [2008] “Synchronization in complex networks,” *Physics Reports*, **469**, 93–153.
- Avrachenkov, K., Litvak, N., Nemirovsky, D., Smirnova, E. & Sokol, M. [2010] “Monte Carlo Methods for Top-k Personalized PageRank Lists and Name Disambiguation,” *Rapport de recherche n 7367, INRIA. arXiv:1008.3775*.
- Boccaletti, S. Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. [2006] “Complex networks: Structure and dynamics,” *Physics Reports*, **424**, 175–308.
- Boldi, P., Santini, M., & Vigna, M. [2009] “PageRank: Functional Dependencies,” *ACM Transactions on Information Systems*, **27**, Article 19, 1–23.
- Clauset, A., Cosma Rohilla Shalizi, C. R. & Newman, M. E. J. [2009] “Power-law distributions in empirical data,” *SIAM Review*, **51**, 661–703.
- Criado, R., Flores, J., Gonzalez-Vasco, M.I. & Pello, J. [2007] “Choosing a leader on a complex network,” *Journal of Computational and Applied Mathematics*, **204** 10–17.
- De Choudhury, M., Mason, W. A., Jake M. Hofman, J. M., & Watts, D. J. [2010] “Inferring Relevant Social Networks from Interpersonal Communication,” *Proc. of the 19th international conference on World Wide Web*.
- De Kerchove, C., & Van Dooren, P. [2009] “Reputation systems and nonnegativity”. *Lectures Notes in Control And Information Sciences*, **389**, 3–16.
- Dorogovtsev, S. N. [2010] *Lectures on Complex networks*, (Oxford University Press).
- Easley, D. & Kleinberg, J. [2010] *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, (Cambridge University Press, New York).
- Estrada, E., Fox, M., Higham, D. J. & Oppo, G-L (Eds.). [2010] *Network Science. Complexity in Nature and Technology*, (Springer)
- Estrada, E.; Higham, D. J. [2010] “Network Properties Revealed Through Matrix Functions”, *SIAM Review* **52**, 696–714.
- Fortunato, S. [2010] “Community detection in graphs,”. *Physics Reports* **486**, 75–174.
- Fuchs, C. [2009] “Social Networking Sites and the Surveillance Society. A Critical Case Study of the Usage of studiVZ, Facebook, and MySpace by Students in Salzburg in the Context of Electronic Surveillance.” (Published by Research Group Unified Theory of Information. Salzburg and Vienna, Austria).
- Sharad Goel, S., Muhamad, R & Watts, D. [2009] “Social Search in Small-World Experiments,” *Proc. of the 18th international conference on World Wide Web*.
- Golder, S., Wilkinson, D. M. & Huberman, B. A. [2007] “Rhythms of social interaction: messaging within a massive online network”, In *Steinfeld, Pentland, Ackerman, and Contractor (eds.), Communities and Technologies 2007: Proc. of the Third Communities and Technologies Conference, Michigan State University, London: Springer*. 41–66.

- Gilbert, E. & Karahalios, K. [2009] "Predicting tie strength with social media," Proc. Conf. on Human Factors in Computing Systems (CHI09).
- Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. [2009] "Unbiased Sampling of Facebook," *arXiv:0906.0060*.
- Granovetter, M. S. [1973] "The Strength of Weak Ties," *American Journal of Sociology* **78** 1360–1380
- Haveliwala, T. H. [2003] "Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search", *IEEE Transactions on knowledge and data engineering*, **15**, 784–796.
- Henry, N. & Fekete, J.-D. [2007] "MatLink: Enhanced Matrix Visualization for Analyzing Social Networks," *Lecture Notes in Computer Science* **4663**, 288–302
- Jeh, G. & Widow, J. [2002] "Scaling personalized web search," *Technical Report, Stanford University*.
- Kostakos, V. [2009] "An empirical study of spatial and transpatial social networks using Bluetooth and Facebook," *it arXiv:0910.4292v1*.
- Langville, A. N. & Meyer, C. D. [2006] *Google's Pagerank and Beyond: The Science of Search Engine Rankings*, (Princeton University Press).
- Lawrence, R., Melville, P., Perlich, C., Sindhvani, V., Meliksetian, S., Hsueh, P.-Y., & Liu, Y. [2010] "Social Media Analytics," *ORMS Today*, **37**, Number 1, February, 2010.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. [2009] "Computational Social Science," *Science* **323**, 721.
- Lee, C., Reid, F., McDaid, A. & Hurleyar, N. [2010] "Detecting highly overlapping community structure by greedy clique expansion," *arXiv:1002.1827v2*.
- Leskovec, J., Huttenlocher, D., & Kleinberg, J. [2010] "Predicting positive and negative links in online social networks," *Proc. of the 19th int. conf. on World Wide Web*.
- Leskovec, J. & Horvitz, E. [2008] "Planetary-Scale Views on a Large Instant-Messaging Network," *Proc. of the 17th int. conf. on World Wide Web*.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. & Christakis, N. [2008] "Tastes, ties, and time: A new social network dataset using facebook.com", *Social Networks*, **30**, 330–342.
- Liben-Nowell, D. & Kleinberg, J. [2003] "The Link Prediction Problem for Social Networks", *Proc. 12th International Conference on Information and Knowledge Management (CIKM)*, 556–559.
- Lin, N., Dayton, P. W., & Greenwald, P. [1978] "Analyzing the Instrumental Use of Relations in the Context of Social Structure," *Sociological Methods & Research*, **7**, 149–166.
- Mayer, A. and Puller, S. L. [2008] "The old boy (and girl) network: Social network formation on university campuses," *Journal of Public Economics*, **92**, 329–347.
- Maximilien, E. M., Grandison, T., Sun, T., Richardson, D., Guo, S. & Liu, K. [2009] "Privacy-as-a-Service: Models, Algorithms, and Results on the Facebook Platform," *Proc. of W2SP 2009: Web 2.0 Security and Privacy*.
- Melville, P. & Sindhvani, V. [2010] "Recommender Systems," *Encyclopedia of Machine Learning*, (Claude Sammut and Geoffrey Webb (Eds), Springer).
- Newman, M. E. J. [2010] *Networks. An introduction*, (Oxford University Press).
- Paek, T., Gamon, M., Counts, S., Chickering, D. M. & Dhesi, A. [2010] "Predicting the Importance of Newsfeed Posts and Social Network Friends," *Proc. Twenty-Fourth AAAI Conf. on Artificial Intelligence (AAAI-10)*. 1419–1424.
- Page, L., Brin, S., Motwani, R., and Winograd, T. [1999] "The PageRank Citation Ranking: Bringing Order to the Web", *Stanford Digital Library Technologies Project*.
- Pedroche, F. [2010 a] "Competitivity groups on Social Network Sites," *Mathematical and Computer Modelling* **52**, 1052–1057.
- Pedroche, F. [2010 b] "Analysis of Competitivity Groups on Social Network Sites using PageRank," *Proc. Seventh Int. Conf. on Eng. Comp. Tech., Civil-Comp Press*, paper 28.
- Pedroche, F. [2010 c] "Leadership Groups on Social Network Sites using Biased PageRank," *AIMS' Journals*, submitted.
- Pedroche, F. [2007] "Methods of calculating the PageRank vector. (Spanish)," *Bol. Soc. Esp. Mat. Apl.*, **39**, 7–30.

- Serra-Capizzano, S. [2005] “Jordan Canonical Form of the Google Matrix: A Potential Contribution to the PageRank Computation,” *SIAM Journal on Matrix Analysis and Applications*, **27**, 305–312.
- Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. [2008] “Community structure in online collegiate social networks,” *submitted to SIAM Review*, preprint, arXiv0809.0690.
- Sabater, J., & Sierra, C. [2005] “Review on computational trust and reputation models,” *Artificial Intelligence Review*, **24**, 33-60
- Song, M., Lee, W. & Kim, J. [2010] “Extraction and Visualization of Implicit Social Relations on Social Networking Services,” *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 1425–1430.
- Tong, H., Faloutsos, C., & Pan, J-Y. [2006] “Fast Random Walk with Restart and its Applications,” *IEEE Int. Conf on Data Mining, ICDM 2006*.
- Tong, H., & Faloutsos C., [2006] “Center-Piece subgraphs: problem definition and fast solutions,” *12th int. conf. on Knowledge Discovery and Data Mining, KDD-2006*.
- Vasalou, A., Joinson, A. N., & Courvoisier, D. [2010] “Cultural differences, experience with social networks and the nature of *true commitment* in Facebook,” *Int. J. Human-Computer Studies* **68**, 719-728.