IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **2(1)** (2012), pp. 37–42

Computing the minimum number of new friends required in a social network to get the highest PageRank

Francisco Moreno[1,†], Andrés González[1] and Andrés Valencia[1]

[1] *Computer Science Department, Universidad Nacional de Colombia, Sede Medelln Carrera 80 No 65-223*

Abstract. The PageRank is one of the most well-known methods for classifying web pages. We apply this method to classify users in a social network. In a similar way to web pages, users with high PageRanks are probably very popular and influencial. In this paper, we propose a novel algorithm called NewFriends based on the PageRank method. Our algorithm calculates the minimum number of new friends required by a user of a social network to become the user with the highest PageRank in the network. We provide formal mathematical definitions and validate our proposal with some experiments using a subnetwork of Facebook.

Keywords: Social Networks, Leadership, Friends, PageRank Method

† **Corresponding author:** fjmoreno@unal.edu.co
Received: November 20th, 2012
Published: December 17th, 2012

1. Introduction

The PageRank[1] is one of the most well-known methods for classifying web pages. The PageRank of a web page p represents the probability that a web surfer is visiting p after a considerable time of navigation. Web pages with high PageRank are probably very popular and influencial in the web. This method has also been applied for classifying social network users[2]. The PageRank method considers the number of links (incoming and outgoing) of the web pages and the structure of their navigation network (graph). In this method, the web page with the largest number of links (incoming) is not necessarily the web page with the highest PageRank, since the network structure plays a decisive role.

In this paper, we propose a novel algorithm based on PageRank method, to calculate the minimum number of new friends required by a user u of a social network to become the user with the highest PageRank. The idea is to add new users to the social network one at at time. Each new user is connected to u and then we analyzed if u has become the user with the highest PageRank. This analysis can be applied in several domains. For example, in politics it could help to determine the minimum number of electors that a candidate require to become the most popular candidate in a social network. In marketing, it could help to determine the minimum number of customers that a company require to become popular, and in this way attract and reach more customers.

There are different measures to assess the importance of a node in a network (graph)[4] ,[5], [6].

On the other hand, different authors have proposed ways to increase the PageRank of web pages [7], [8], [9].

The paper is organized as follows. In Section 2., we present our proposal for computing the minimum number of friends that a user requires to get the highest PageRank. In Section 3., we present our experimental results with artificial networks and with a real subnetwork obtained from Facebook. In Section 4. we conclude the paper and give some perspectives for future research.

2. Computing the minimum number of friends that a user requires to get the highest PageRank in a social network

Consider a social network with $n = 5$ users represented with a directed graph, see Figure 1. The idea is to classify the nodes (users) of the social network according to their links and structure. To achieve this, we apply the PageRank method[2].

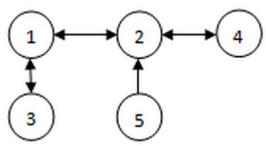

Figure 1: A social network with five users represented with a directed graph

We denote PPR the Personalized PageRank as the PageRank vector computed using a personalization vector and we denote PR_j the PageRank of node i, computed with a personalization vector v_j.

Our goal is to determine the minimum number of new friends that a user requires to become the user with the highest Pagerank in the network.

Formally, let $G = (N, E)$ be the initial graph that represents the network, where N is the set of nodes and E is the set of links of the network. Each link is represented as (n_1, n_2) where $n_1, n_2 \in N$, $n_1 \neq n_2$. Let $\pi_i(G)$ denote the i component of the PPR for some personalization vector v, where $i \in N$. Let $Newfriends(i, m) = (k_1, i), (k_2, i), ..., (k_m, i)$ with $k_y \notin$ N, $1 \leq y \leq m$. Newfriends(i,m) represents the set of new nodes that will be connected to node i: $N' = N \cup k_1, k_2, ..., k_m$, $E' = E \cup Newfriends(i, m)$, and $G'(Newfriends(i, m)) = (N', E')$; where m is the smallest positive integer such that $\pi_i(G'(Newfriends(i, m))) = max(\pi_j(G'(Newfriends(i, m))))$, $j \in N'$. That is, m is the minimum number of new friends that i requires to get the highest PageRank in the network. The algorithm can be seen in [3].

3. Experiments

We used a subnetwork from Facebook with 769 nodes called Caltech [10]. We considered three groups of nodes: the ten nodes with the highest PageRank, ten nodes whose PageRank is in the middle, and the ten nodes with the lowest PageRank. Our results are shown in Tables 1.

For example, the node with ranking 10, which currently has 157 friends, requires 13 new friends to become the node with the highest PagerRank in the network; the node with ranking 388, which currently has 39 friends, requires 23 new friends; and the node with ranking 769, which has only 1 friend, requires only 13 to become the node with the highest PagerRank in the network. In the results it is noteworthy that the minimum number of friends required by the nodes with the lowest PageRank to get the highest PageRank is lower than the minimum number of friends required by the nodes whose PageRank is in the middle of the network. The lowest PageRank nodes only require thirteen new friends to get highest PageRank (this number corresponds to the 5% of the links of the node with the highest PageRank which has 248 links), whereas the nodes whose PageRank is in the middle of the network required twenty three new friends. In [11] it is explained that this behaviour is not unusual in the PageRank method, and in [12] the authors have shown that social networks are very sensitive to small changes. The explanation may be due to the following.

We must consider that the new nodes that are added to the network are only connected to the node that wish to get the highest PageRank (we call these new nodes "beginners"). The results show that beginners nodes strongly affect those nodes of the network with low PageRank, increasing their PageRank. Consider a node w with few friends in the network, e.g., a node w with a single friend u. When w is connected to a beginner node b, the probability of finding the surfer of the network in w will be high because

Table 1: Results for the ten nodes with the highest, middle, and lowest PageRank

Node ranking	**Minimum number of friends required by the node to get the highest PageRank**	**Number of links of the node in the network (before applying the NewFriends algorithm)**
1	0	248
2	6	194
3	6	203
4	9	171
5	11	184
6	11	152
7	13	156
8	13	172
9	13	156
10	13	157
379	24	41
380	23	38
381	22	29
382	23	40
383	24	40
384	23	35
385	23	39
386	23	38
387	24	41
388	23	39
760	13	1
761	13	1
762	13	1
763	13	1
764	13	1
765	13	1
766	13	1
767	13	1
768	13	1
769	13	1

when the surfer arrives at w from b, the surfer will have only two choices to go from w (to u or to b). On the other hand, if the surfer arrives at a node with several friends, the surfer will have several choices, and the probability of finding the surfer in w will be low.

4. Conclusions

In this paper, we analyzed how the PageRank of a node of a network is affected when it is connected to new nodes of the network. We proposed a formal definition and an algorithm to determine the number of new friends (beginners nodes) that a node requires in order to become the node with the highest PageRank in the network. Our method has applications in fields such as marketing, politics, sales, and entertainment, among others; where their users want to gain visibility and be leaders of the network. The results showed that in the PageRank method not necessarily the nodes with the largest number of friends are the ones with the highest PageRank. Indeed, the results showed that nodes of the network with few friends only require to be connected to a few beginner nodes to get the highest PageRank in the network.

As future work we plan to develop our proposal with other centrality measures. In particular, we expect to modify our method considering weighted networks, i.e., networks where every link is associated with a weight. For example, the weight of a link may indicate the influence of a node on another. Finally, we would like to determine the best potential friends for a node w. That is, which are the nodes of the network that if connected to w, they will increase the most the PageRank of w.

References

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

[2] Pedroche F. Ranking nodes in Social Network Sites using biased PageRank. Instituto de Matemática Multidisciplinaria, Universidad Politécnica de Valencia 2010; E-46022 Valencia.

[3] http://www.unalmed.edu.co/~fjmoreno/newfriends.txt

[4] Freeman LC. Centrality in social networks conceptual clarification. Social Networks 1978-1979: 215–239.

[5] Molina JL, Muñoz J, Losego P. Red y realidad: aproximación análisis de Redes científicas. VII Congreso Nacional de Psicología Social Oviedo, Universitat Autónoma de Barcelona 2000.

[6] Bonacich P. Factoring and weighting approaches to clique identification. Journal of Mathematical Sociology 1972.

[7] Avrachenkov K, Litvak N. The Effect of New Links on Google PageRank. INRIA, Rapport de recherché 2004.

[8] Sydow M. Can one out-link change your PageRank? Advances in Web Intelligence. Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2005; pp. 408–414.

[9] `http://stason.org/articles/money/seo/google/12_things_to_do_to_improve_your_site_google_page_rank.html`.

[10] Traud AL, Kelsic ED, Mucha PJ, Porter MA. Community structure in online collegiate social networks. SIAM 2012; 343-361.

[11] Bonacich P and Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23(3) (2001) pp.191–201.

[12] Pedroche, F., Criado, C., García, E., and Romance, M. Matrix growth models based on centrality measures: a first analysis. International Journal of Complex Systems in Science. (2011) vol.1(2), pp.124-128..