Influence maximization based on the least influential spreaders

Didier A. Vega-Oliveros, Lilian Berton, Alneu de Andrade Lopes and Francisco A. Rodrigues
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo - Campus de São Carlos
13560-970 São Carlos, SP - Brazil
Email: {davo, lberton, alneu, francisco}@icmc.usp.br

Abstract

The emergence of social media increases the need for the recognization of social influence mainly motivated by online advertising, political and health campaigns, recommendation systems, epidemiological study, etc. In spreading processes, it is possible to define the most central or influential vertices according to the network topology and dynamic. On the other hand, the least influential spreaders have been disregarded. This paper aims to maximize the mean of information propagation on the network by recognizing the non influential individuals by making them better spreader. Experimental results confirm that selecting 0.5% of least influential spreaders in three social networks (*google+*, *hamsterster* and *advogato*) and rewiring one connection to some important vertex, increase the propagation over the entire network.

1 Introduction

Nowadays social media, such as blogs, social networks, sharing sites, etc., can quickly spread rumors or information [Castellano *et al.*, 2009; González-Baillón *et al.*, 2011], reaching millions of users in the Global village. Propagation process evolving information or rumors can be both beneficial or destructive in scenarios like stock market, advertising and marketing, adoption of technologies, politics and national security, among other.

Many of the approaches to information or rumor spreading [Moreno *et al.*, 2004; Castellano *et al.*, 2009; González-Baillón *et al.*, 2011; Borge-Holthoefer *et al.*, 2012] concentrate in how ideas are shared among individuals and what are the conditions that allow a large dissemination. For this reason, they are understood as being equivalent to dynamics like epidemics spreading, in the sense that individuals would be psychological 'infected' with some idea or opinion [Daley and Kendall, 1964; Maki and Thompson, 1973; Barrat *et al.*, 2008; Castellano *et al.*, 2009]. The rumor spreading process assumes that the population can be divided into groups with different states, which is generally described as the Susceptible (inactive), Infectious (spreader) and Recovered (stifler) (SIR) model [Daley and Kendall, 1964; Maki and Thompson, 1973].

Some individuals can have a higher social influence than others, according to their position, topological characteristics in the network and dynamic [Kitsak *et al.*, 2010; González-Baillón *et al.*, 2011; Borge-Holthoefer *et al.*, 2012]. A lot of researches have tried to identify influence in social networks [Burt, 1992; Sabidussi, 1996; Kitsak *et al.*, 2010; Borge-Holthoefer *et al.*, 2012]. It was found that the final fraction of informed individuals is not significantly affected by chosen the spreaders by topological features [Moreno *et al.*, 2004]. However, incorporating characteristics of real scenarios as the activity of spreader or apathy with the new information [Borge-Holthoefer *et al.*, 2012], or exposure to multiple sources and individual's thresholds [González-Baillón *et al.*, 2011] lead to the emergence of influential spreaders in the information spreading dynamic.

In the other hand, the least influential individuals of the network are ignored in the literature. Let consider the case of a scientific collaboration network. Some research groups are more prominent than others. If a researcher with few publication or contacts visits a prominent group, by the homophily phenomenon he/she can consequently improve his/her publication capacity. Based on this, we measure the vertices' influence and recognize the least influential spreaders. Thereby, we propose to turn those individuals into better spreaders by connecting to the most central nodes.

The main contributions of this paper are: i) confirm the existence of least influential individuals in the network; ii) characterize the least influential spreaders by analyzing centrality measures, such as degree, betweenness, closeness, pagerank, eigenvector, k-core, clustering coefficient and structural holes; iii) select a proportion of least influential users from *google+*, *hamsterster* and *advogato* social networks and make them better spreaders rewiring one of its connection to a more central individual; iv) analyze the impact of different proportions in the rewiring process and how it increases the mean of information spreading on the network.

The paper is organized as follows: Section 2 describes the centrality measures considered for network characterization and the spreading process (SIR model); Section 3 presents the simulations in three social networks (*google+*, *hamsterster* and *advogato*), recognizing the least influential users, making them potential spreaders and increasing the information capacity of the networks. Finally, Section 4 reports the final remarks.

2 Definitions

A social network can be represented as a static unweighted network $G = (V, E)$, such as $V = \{v_1, v_2, \ ... \ v_n\}$ is the set of N vertices or actors and $E = \{e_1, e_2, \ ... \ e_m\}$ is the set of M edges or connections. The adjacency matrix A is the mathematical entity that represents the connections of the network. We consider undirected networks, that means i, j $\in V$, $A_{ij} = A_{ji} = 1$ if i and j are connected, or 0 otherwise.

A walk between a pair of vertices $(i,\ j)$ is a consecutive sequence that starts at i and ends in j, so that any vertices are visited more than once. The distance or length of the walk is defined as the number of edges contained in the sequence. Two vertices are neighbors if they are connected in a walk of length 1. The shortest distance between two vertices is known as the shortest path or geodesic path ℓ_{ij}. The shortest paths can be computed by the Dijkstra, BellmanFord algorithms or by a Breadth-first search method [Cormen *et al.*, 2009]. A component is the largest sub-set of vertices from the network in which exist at least one walk between each pair of vertices, but never connect to another component. A connected network has only one component. In the case that i and j belong to different components, it is assumed that $\ell_{ij} = \infty$. For this reason, here we considered the largest component of the network.

2.1 Centrality measures

Several measures have been proposed to describe the importance or centrality of a vertex in the network [Costa *et al.*, 2007]. These centrality measures are defined considering particular definitions of influence [Newman, 2003]. It is assumed that vertices with higher centrality measure are more suitable to influence in the opinion of others. Among some of the possibilities, we have the popularity of a vertex (degree centrality) [Costa *et al.*, 2007], the proximity or how close an individual is to the others (closeness centrality) [Sabidussi, 1996], the trusted vertices in the transmission of information (betweenness centrality) [Freeman, 1977], the proximity of vertices to the network core (k-core) [Seidman, 1983] or even the renowned that an individual has (pagerank centrality) [Brin and Page, 1998]. In the following, we present the centrality measures adopted in this paper.

Degree centrality (*DG*) considers the number of connections or relations of a vertex. The set of vertices connected to a certain vertex i is defined as the neighborhood and the connection degree DG_i represents the size of its neighborhood [Costa *et al.*, 2007]. It means, the higher the degree, the more popular the vertex.

Betweenness centrality (*BE*) is related to the capacity of information transmission of vertices. For a vertex j this measure quantified the number of shortest paths that pass through j between all pair of vertices $(i,\ k)$ [Freeman, 1977], with i, j and k different. The centrality measure expresses how much the vertex j works as bridge, meaning how confidence or trusted is j in the network.

Eigenvector centrality (*EV*) takes into account that vertices with same degree have different levels of importance according to the importance of their neighbors. The eigenvector centrality is the principal eigenvector associated with the greatest eigenvalue of the adjacency matrix A. It describes the importance of the vertices given the quality of its connections [Bonacich, 1972].

Pagerank centrality (*PR*) derives from a Markovian process that follows a random walk navigation through the network. It expresses the importance of the vertices considering the probability of arriving at certain vertex after a large number of steps. The pagerank was initially proposed to rank web pages [Brin and Page, 1998] and the idea is to simulate the behavior of a user that is surfing on the net. The user navigates following the links available at the current page or, she/he can jump to other pages by typing a new URL in the browser. In social networks it can be approached like the more cited or renowned individuals.

Closeness centrality (*CL*) is the average of the shortest paths from each vertex to the rest of the network [Sabidussi, 1996]. Formally, the closeness centrality is the inverse of the average of the shortest paths from i to all the vertices, i.e., $CL_i = N/\sum_{j \neq i} \ell_{ij}$. Thereby, vertices that are closer to the others have higher closeness centrality. For the sake of calculating the centrality for disconnected networks, the average is considered by component.

k-core centrality (*KC*) describes the topology of the network in terms of sub network decomposition in cores. The core of k order (H_k) is the set of vertices where for each vertex i in H_k its degree $k_i \geq k$. It means, k is the maximum core that i can belongs, with $Kc(i) = k$ and H_k been the largest set of vertices with this property [Seidman, 1983]. The main core is the set of vertices with the largest k-core value from the network, and these vertices are the most central. Vertices with low values of k-core are commonly located at the periphery of the network. Not necessary all the high-degree vertices have higher k-core values. For instance, hubs located in the periphery would have small values of k-core, or vertices with larger k-core value could have not so large degree [Kitsak *et al.*, 2010].

Clustering coefficient (*CT*) or transitivity is a common property in real-world networks. In social networks it means that if A has one friend that is also friend of B, there is a strong tendency that A being also a friend of B. In topology terms it is the presence of triangles (cycles of order three) in the network. The clustering coefficient [Watts and Strogatz, 1998] for a vertex CT_i is defined as the number of triangles centered on i over the maximum number of possible connections for i. CT_i has value 1 if all neighbors of i are interconnected.

Structural Holes (*HO*) considers all the vertices as an ego network, where connections no related with each vertex have not direct effect. The key factor is the redundancy that each vertex has in its neighborhood, evaluating if its position and connections brings some opportunities. It means that the success of an individual i within a social network or organization is related to access local bridges (trusted people) [Burt, 1992]. If removed i from the network, a structural hole will happen in the local neighborhood. These individuals are important to the connectivity of local regions.

The network can be characterized by average the centrality measures of all vertices i.e., $\langle DG \rangle \ = \frac{1}{N} \sum_{i=1}^{n} \ DG_i$, and is similar for the other measures.

Table 1: Structural properties of the complex networks

Network	N	$\langle DG\rangle$	$\langle CT\rangle$	$\langle HO\rangle$	$\langle BE\rangle$	$\langle CL\rangle$	$\langle EV\rangle$	$\langle KC\rangle$	$\langle PR\rangle$
hamsterster	2000	16.1	0.5399	0.2898	2.588×10^3	1.43×10^{-4}	0.011	9.287	5.00×10^{-4}
advogato	5054	15.6	0.2525	0.3700	5.747×10^3	6.19×10^{-5}	6.82×10^{-3}	8.137	1.98×10^{-4}
google+	23613	3.32	0.1742	0.8112	3.581×10^4	1.10×10^{-5}	2.30×10^{-3}	1.669	4.21×10^{-5}

2.2 Spreading process

Spreading is a pervasive process in society and several models have been developed in order to understand the propagation of ideas or opinions through social networks [Castellano *et al.*, 2009]. In classical rumor spreading models the ignorant or inactive (S) are those who remain unaware of the information, the spreaders (I) are those who disseminate the ideas, and the stifler (R) are those who know the information but lose the interest in spreading it. All vertices have the same probability β for convincing their neighbors and probability γ for stopping to be active as propagator.

The Maki-Thompson (MT) [Maki and Thompson, 1973] rumor approach was employed to model the spreading process. In the MT whenever an active spreader i contacts a vertex j that is inactive, the latter will become active with a fixed probability β. Otherwise, when j knows about the rumor, it means j is a spreader or a stifler, the vertex i will turn into a stifler with probability γ. The behavior when the spreader stops to propagate is understood because the information is too much known (contacting spreaders) or without novelty (contacting stifler). The general rules of contact can be represented as:

$$\begin{cases} I_i \, + \, S_j \, \xrightarrow{\beta} \, I_j \, , \\ I_i \, + \, R_j \, \xrightarrow{\gamma} \, R_i \, , \\ I_i \, + \, I_j \, \xrightarrow{\gamma} \, R_i \, , \end{cases} \tag{1}$$

where i and j are neighbors and the operator "+" means the contact between them.

In terms of Monte Carlo (MC) implementation, let consider a constant population of N vertices in all time steps. Each vertex can be only in one state, that is $I_i(t) = 1$ *if* $i \in I$, otherwise $I_i(t) = 0$, and $S_i(t) + I_i(t) + R_i(i) = 1$. The macroscopic fraction of ignorant ($\psi(t)$), spreaders ($\phi(t)$) and stifler ($\varphi(t)$) over time is calculated as $\psi(t) = \frac{1}{N}\sum_{i=1}^{N} S_i(t)$, that is similar to the other states and always fulfill $\psi(t) + \phi(t) + \varphi(t) = 1$. We assume that infection and recovering do not occur during the same discrete time window or step. Also the case when a spreader randomly contacts one neighbor per unit time, the contact process, was adopted.

The initial setup for the propagation is $\psi(0) = 1 - 1/N$, $\phi(0) = 1/N$ and $\varphi(0) = 0$. Each simulation begins with a uniformly selection of vertices as defined in the initial setup. At each time step, all spreaders uniformly select one of its neighbors and try to infect it with probability β, or stop the propagating with probability γ. The simulations run until the end of the propagation process is reached, when $\phi_\infty = 0$.

Different theoretical models have been proposed for modeling the rumor dynamics in networks [Moreno *et al.*, 2004; Barrat *et al.*, 2008; Castellano *et al.*, 2009; Borge-Holthoefer *et al.*, 2012]. These analytical models make assumptions about the network structure such as the degree correlation or distribution, compartments or class of vertices with same probabilities, homogeneous mixing or mean field theory. Notwithstanding all of them claim that their numerical solutions agree with the MC simulations, so we adopt this approach.

3 Maximization of least influential spreaders

We analyze the spreading capacity in the microscopic and macroscopic scales of the propagation. For each vertex $i \in V$, we calculate the final fraction of stifler φ_∞^i. This quantity represents how long the rumor propagates in the network starting in i. Each φ_∞^i were average over 90 realizations. For the macroscopic propagation scale, the φ_∞^V represents the mean of spreading capacity for all vertices, it means $\varphi_\infty^V = (\sum_{i \in V} \varphi_\infty^i)/N$.

3.1 Dataset

We employ the following social networks: the *hamsterster* [Kunegis, 2014], an undirected and unweighted network based on the website data from HAMSTERSTER.COM. The vertices are the users of the system and the edges represent a relationship among users. It consists of friend and family relationship between the users of the website. The *advogato* [kon, 2014], an online community platform for developers of free software launched in 1999. Vertices are users of advogato and the directed edges represent trust relationships. Finally, the *google+* [Gpl, 2014] contains Google Plus user-user links. The directed edge denotes that one user has the other in his circles, but we assume the network as undirected. We always consider the main component for these networks.

The topological characteristics of these networks are summarized in Table 1, with the corresponding averages of the centralities: degree $\langle DG\rangle$, clustering coefficient $\langle CT\rangle$, structural holes $\langle HO\rangle$, betweenness $\langle BE\rangle$, closeness $\langle CL\rangle$, eigenvector $\langle EV\rangle$, k-core $\langle KC\rangle$ and pagerank $\langle PR\rangle$. The *google+* is more an egocentric network ($\langle HO\rangle$) than others, where vertices are very close ($\langle CL\rangle$) and most of them have few connections. The *hamsterster* is a more sparse network with more triangles ($\langle CT\rangle$) and connections between users. And, *advogato* is a more dense network and in the middle term of the previous.

3.2 Propagation analysis

For evaluating the impact of the least influential users in the networks, we employ the z-score normalization over the spreading capacities φ_∞^i and sort the values in ascending order. The z-score indicates how many standard deviations an element is according to the mean. The z-score of a raw value

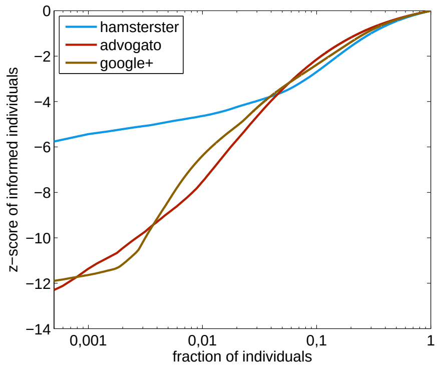


Figure 1: Comparative z-score of the propagation capacity in function of the proportion of least influential spreaders

x is $z = \frac{x-\mu}{\sigma}$, where: μ is the mean and σ is the standard deviation of the population. The impact is shown in Figure 1, where considering 10% of least influential spreaders, the mean decreases around 3 standard deviations. This decreasing behavior continues successively. The less influential spreaders impact the overall mean of spreading capacity.

We aim to recognize these least influential spreaders in order to improve their influence. Following, we considered the centrality measures and analyzed the topological influence over the propagation and data clustering to find patterns on these individuals.

3.3 Topological analysis

Z-score was calculated for each centrality measure and network (Figure 2 a-c). The fraction of individuals was sorted in ascending order for each centrality. Vertices with lowest eigenvector centrality always led the lowest z-score values.

Jaccard coefficient was employed in order to measure the proportion of least influential spreaders contained in each group of vertices with lower centrality measure (Figure 2 d-f). The Jaccard coefficient compares the similarity and diversity of sample sets. It measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets: $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$. Here, for each centrality measure A represents the proportion of vertices with lowest values and B the fraction of least influential vertices. We define the least influential users as the vertices that achieved propagation values $\varphi^i_\infty \leq \frac{\varphi^V_\infty}{2}$. Selecting less than 10% of vertices with lowest eigenvector centrality, it was obtained around half of the less influential spreaders.

The topological properties considering 1% of vertices with least influential spreading and lowest EV value were analyzed. We verify if these vertices also have low values in other centralities. For each centrality measure we normalize the values by its own average. Therefore, it is possible to compare the networks and the maximum values achieved (Figure 2 g-i). We observe that the topological properties of individuals with lowest EV values are a subset of the values presented for the least influential spreaders group. Unlike expected, vertices with BE, DG or PR above the mean may be also less influential spreaders.

Table 2: k-means clustering result for the *google+* network

Measures	Cluster 1 (10%)	Cluster 2 (67%)	Cluster 3 (23%)
DG	1.029 ± 1.001	1.128 ± 0.386	10.553 ± 72.435
BE	0.00 ± 0.0004	0.00 ± 0.0005	0.0023 ± 0.023
CL	0.5 ± 0.00	0.5 ± 0.00	0.501 ± 0.019
PR	0.0007 ± 0.0007	0.0007 ± 0.0002	0.0043 ± 0.0299
EV	**0.00 ± 0.00**	0.0021 ± 0.0037	0.0117 ± 0.0202
KC	0.083 ± 0	0.094 ± 0.032	0.292 ± 0.178
CT	0.00 ± 0.00	0.00 ± 0.00	0.746 ± 0.312
HO	**0.991 ± 0.0291**	0.942 ± 0.166	0.360 ± 0.145
Propagation	**2612.274 ± 332.095**	3112.858 ± 56.965	3135.407 ± 37.310

Table 3: Propagation improvements for the real networks: 1st place are in bold and 2nd place are underlined

Network	Rewired (%)	DG	BE	EV	PR
hamsterster	5	1.0817	**1.0932**	1.0810	1.0844
	2.5	1.0804	**1.0817**	1.0794	1.0783
	0.5	1.0493	**1.0510**	1.0479	1.0489
advogato	5	**1.0600**	1.0594	1.0599	1.0594
	2.5	1.0529	1.0556	1.0549	**1.0557**
	0.5	1.0450	1.0439	**1.0453**	1.0447
google+	5	1.0124	**1.0158**	1.0050	1.0112
	2.5	1.0143	1.0102	1.0103	**1.0171**
	0.5	1.0241	1.0190	1.0187	**1.0266**

3.4 Cluster analysis

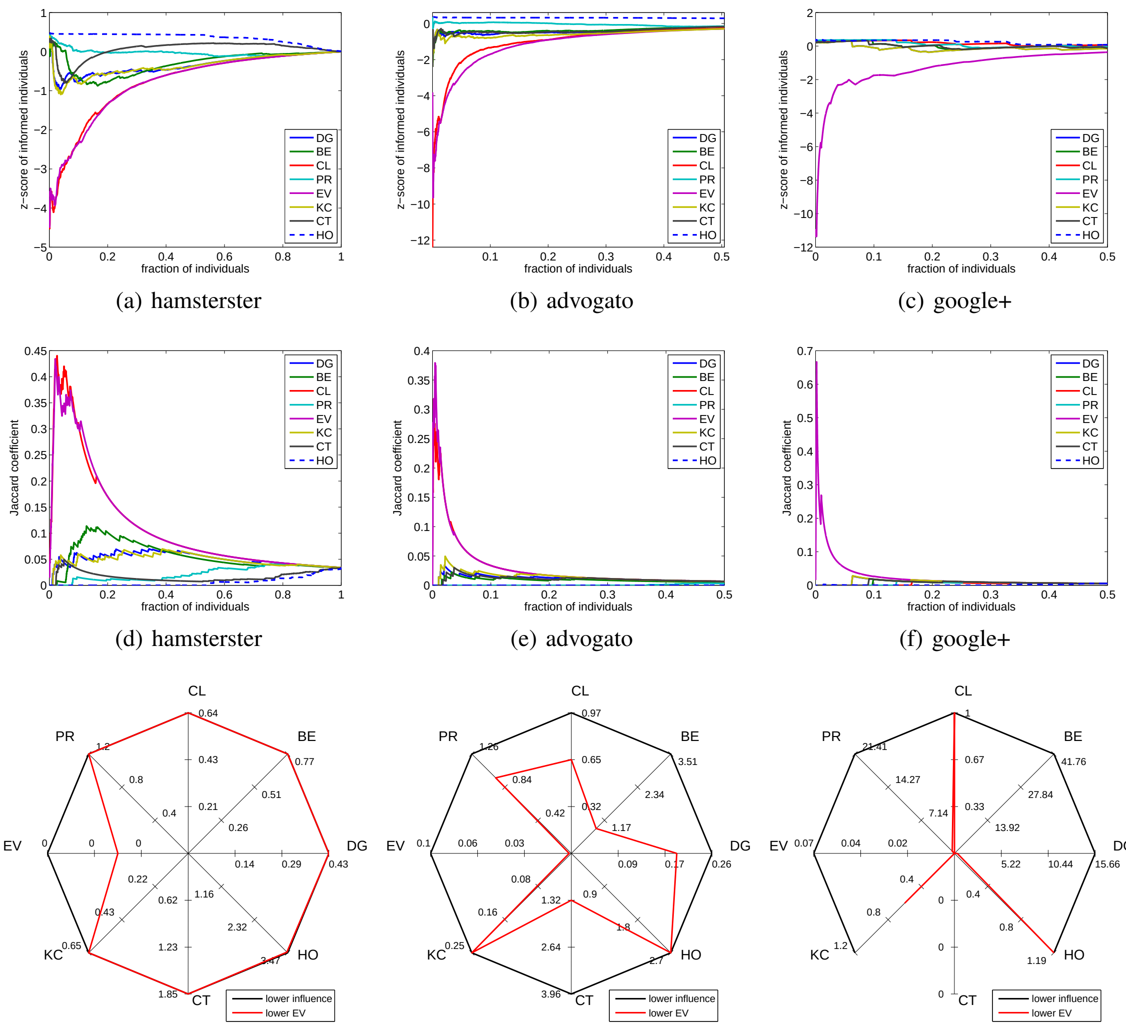
We perform a clustering analysis in order to discover the group of least influential spreaders according to the topological characteristics of the vertices. As objects in the same group are more similar to each other than those in other groups, we aim to identify common properties of the elements that belong to the group with lowest propagation rates.

We apply the popular k-means clustering [Macqueen, 1967] that partitions data into k mutually exclusive clusters. Each cluster is defined by its members and by its centroid that represents the point to which the sum of distances from all objects in the cluster is minimized. We used the Weka[1] implementation with the standard configuration.

The result for the *google+* network is shown in Table 2. As we set the parameter $k = 3$ the k-means generates three clusters. We are mainly interested in the cluster 1 because has the individuals with the lowest propagation mean. By comparing the mean of the measures among clusters 1, 2 and 3, we observe that the values of EV ≈ 0 and HO ≈ 1 are detachable in comparison to other clusters.

When performed the cluster analysis in *hamsterster* and *advogato*, the more significant measures were CL < 0.5 and also EV ≈ 0. In all the case, individuals with eigenvector values close to zero are in the cluster with lowest spreading capacity. If we aim to find a cluster with non influential spreaders in the network, we can run the k-means algorithm until some group partition achieve an eigenvector mean $\langle EV \rangle < 1/n$. This pattern confirms the results of the topological analysis. Hence, we generalize that the lower the eigenvector value, the lower the spreading capacity of vertices.

[1] http://www.cs.waikato.ac.nz/ml/weka

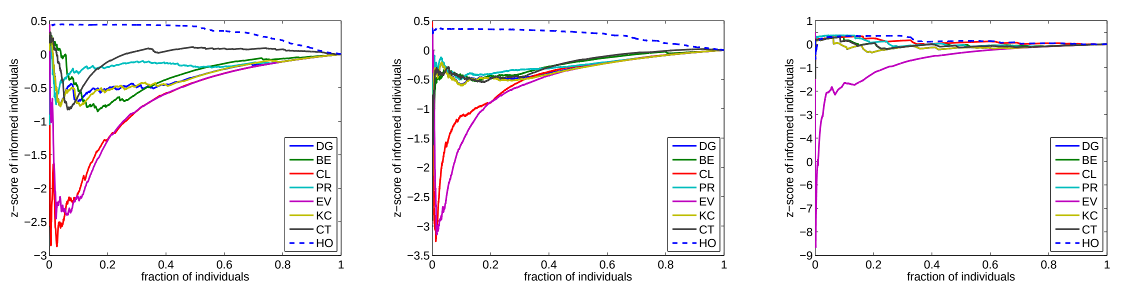


(a) hamsterster (b) advogato (c) google+

(d) hamsterster (e) advogato (f) google+

(g) hamsterster (h) advogato (i) google+

Figure 2: Propagation analysis for the real social networks *hamsterster*, *advogato* and *google+*: (a-c) the impact of vertices with low centrality in the propagation; (d-f) the similarity between vertices with lower centrality and the less influential spreaders; (g-i) the topological characteristic of the least influential spreaders and vertices with lowest EV values. For *hamsterster* and *advogato* were considered $\beta = 0.3$ and $\gamma = 0.1$, and for *google+* $\beta = 0.5$ and $\gamma = 0.1$, in all simulations



(a) hamsterster (b) advogato (c) google+

Figure 3: Propagation analysis for the real social networks when maximizing the social influence rewiring 0.5% of least influential spreaders considering vertices with highest DG centrality: the impact of low centrality values in the propagation mean

3.5 Influence maximization

For improving the spreading capacity of the network we selected $5\%, 2.5\%$ and 0.5% of the vertices with lower EV. For each vertex only one of its edges was randomly selected and rewired to an influential vertex of the network. We randomly consider vertices with highest DG, BE, EV and PR centrality and exchange only one edge of each influencer. For each measure and proportion of lowest EV vertices to be rewired, we have the resulting spreading capacity (φ_{∞}^{V}) normalized by the mean of the non maximized case (Table 3). When connecting the least influential spreaders to vertices with highest DG centrality leads to 6 second places and 1 first place results. Hence, the visual results selecting 0.5% of lowest EV connecting to vertices with highest DG centrality are shown in Figure 3. The overall mean of propagation was increased for each network. The z-score also increases for the other measures in the entire network and not only for the 0.5% of the selected vertices (Figure 3). For the best cases the improvements achieved an overall mean larger than the highest spreading value of the non maximized case.

4 Final remarks

This paper explored the least influential spreaders of a network considering centrality measures and analyzing topological and data clustering approaches. The results indicate that vertices with eigenvector value $EV_i < 1/N$ have little influence in the network. We selected $5\%, 2.5\%$ and 0.5% of these individuals and rewired only one edge to an influential vertex. This approach improves propagation capacity of those vertices and the mean of the propagation in three social networks considered: *google+*, *hamsterster* and *advogato*. More efficient targeted actions can be performed to improve the diffusion on the network. Selecting a few non influential users and presenting them to one influencer, they can change their action, become better spreaders and improve the overall diffusion capacity. For example, the promotion of interchange of students to prominent institutions, or famous/popular people giving talks in common places or to specific groups of individuals, can produce a considerable impact in the diffusion of ideas and benefits. The motivation or incentive for the most important vertices may be monetary, ideological causes, increase publications and impact, among others. This is a work in progress that shows promising experimental results. New paths of study can be developed by analyzing the least influential spreaders and dynamics.

Acknowledgments

This research was supported by National Council for Scientific and Technological Development (CNPq) grant: 140688/2013-7 and Sao Paulo Research Foundation (FAPESP) grant: 2011/21880-3.

References

[Barrat *et al.*, 2008] Alain Barrat, MarseilleMarc Barthélemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[Bonacich, 1972] Phillip Bonacich. Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.

[Borge-Holthoefer *et al.*, 2012] Javier Borge-Holthoefer, Sandro Meloni, Bruno Gonçalves, and Yamir Moreno. Emergence of Influential Spreaders in Modified Rumor Models. *Journal of Statistical Physics*, 151(1-2):383–393, September 2012.

[Brin and Page, 1998] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, V:107–117, 1998.

[Burt, 1992] Ronald S. Burt. *Structural holes: The social structure of competition*. Harvard University Press, Cambridge, MA, 1992.

[Castellano *et al.*, 2009] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical Physics of Social Dynamics. *Reviews of Modern Physics*, 81(2):591–646, may 2009.

[Cormen *et al.*, 2009] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 3 edition, 2009.

[Costa *et al.*, 2007] L.D.F. Costa, F.A. Rodrigues, G Travieso, and P.R. Villas Boas. Characterization of complex networks: A survey of measurements. *Adv. in Physics*, 56:167–242, 2007.

[Daley and Kendall, 1964] D. J. Daley and D. G. Kendall. Epidemics and Rumours. *Nature*, 204(4963):1118–1118, 1964.

[Freeman, 1977] L C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[González-Baillón *et al.*, 2011] Sandra González-Baillón, Borge-Holthoefer, Javier Rivero, and Yamir Moreno. The dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, 1:197:1–7, 2011.

[Gpl, 2014] Google+ network dataset – KONECT, dec 2014.

[Kitsak *et al.*, 2010] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, August 2010.

[kon, 2014] Advogato network dataset – KONECT, October 2014.

[Kunegis, 2014] J. Kunegis. Hamsterster full network dataset-KONECT, jan 2014.

[Macqueen, 1967] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symp. on Math. Stats. and Prob.*, pages 281–297, 1967.

[Maki and Thompson, 1973] D. P. Maki and M Thompson. *Mathematical Models and Applications, with Emphasis on the Social, Life, and Management Sciences*. Prentice-Hall, 1973.

[Moreno *et al.*, 2004] Yamir Moreno, Maziar Nekovee, and Amalio F. Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E*, 69(6):066130, jun 2004.

[Newman, 2003] M E J Newman. The structure and function of complex networks. *SIAM Review*, 45:1–8, 2003.

[Sabidussi, 1996] G Sabidussi. The centrality index of a graph. *Psychometrika*, 31:581–603, 1996.

[Seidman, 1983] S Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.

[Watts and Strogatz, 1998] Duncan J Watts and Steven H Strogatz. Collective dynamics of s̀mall-world' networks. *Nature*, 393(6684):440–442, 1998.