Knowledge-Based Systems 000 (2017) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Centrality measure in social networks based on linear threshold model

Fabián Riquelme[a,*], Pablo Gonzalez-Cantergiani[b], Xavier Molinero[c], Maria Serna[d]

[a] Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Chile
[b] InstaGIS Inc, San Francisco, CA, USA
[c] Mathematics Department, Universitat Politècnica de Catalunya, Spain
[d] Computer Science Department and Barcelona Graduate School of Mathematics, Universitat Politècnica de Catalunya, Spain

ARTICLE INFO

Article history:
Received 16 May 2017
Revised 23 October 2017
Accepted 24 October 2017
Available online xxx

MSC:
91D30
05C22
68R10

Keywords:
Centrality
Linear threshold model
Independent cascade model
Spread of influence
Social network

ABSTRACT

Centrality and influence spread are two of the most studied concepts in social network analysis. In recent years, centrality measures have attracted the attention of many researchers, generating a large and varied number of new studies about social network analysis and its applications. However, as far as we know, traditional models of influence spread have not yet been exhaustively used to define centrality measures according to the influence criteria. Most of the considered work in this topic is based on the independent cascade model. In this paper we explore the possibilities of the linear threshold model for the definition of centrality measures to be used on weighted and labeled social networks. We propose a new centrality measure to rank the users of the network, the Linear Threshold Rank (LTR), and a centralization measure to determine to what extent the entire network has a centralized structure, the Linear Threshold Centralization (LTC). We appraise the viability of the approach through several case studies. We consider four different social networks to compare our new measures with two centrality measures based on relevance criteria and another centrality measure based on the independent cascade model. Our results show that our measures are useful for ranking actors and networks in a distinguishable way.

1. Introduction

Centrality is one of the most studied concepts in social network analysis and it has been exhaustively studied at least since 1948 [1]. A social network can be represented as a graph, whose nodes are the actors of the network, and the edges are interpersonal ties among the actors [2]. Sometimes, edges have associated weights representing the strength of each interpersonal tie. In this context, centrality measures aim to determine how structurally relevant is an actor within the social network. The most traditional centrality measures, such as *degree*, *closeness*, and *betweenness*, are related with the topology of the graph. In these measures, an actor is considered more central when it has a greater degree, or it is closer to the other actors, or it allows to interconnect the other actors in the network, respectively [3].

In recent years, the massive increment of Internet users has allowed the emergence of varied and complex social networks, which increases the need to create more sophisticated centrality measures based on new relevance classification criteria. Due to the huge size of these networks, in terms of number of nodes and relationships among them, it is necessary that the measures can be efficiently computed. Nowadays, there are centrality measures based on how much information can be dispersed through the nodes of a network [4,5], measures based on power indices of cooperative game theory [6–8], measures based on machine learning and predictive models [9], among others. There are also measures specially created for specific social networks, e.g., for the Twitter network, more than seventy different centrality measures have been created only since 2010 [9].

Two of the most well-known relevance measures are the *PageRank* [10] and the *Katz centrality* [11]. Both measures are variants of the *eigenvector centrality* [12]. Identifying the relevance of users is particularly useful for many applications, such as viral marketing [13], information propagation [14], search strategies [15], expertise recommendation [16], community systems [17], social customer relationship management [18], and percolation theory [19]. Furthermore, centrality measures can be used to identify the most active, popular, or influential users within a network [9].

The spread of influence models the ways in which actors influence each other through their interactions in a social network. The nodes exert their influence through the graph. Once a set of actors adopt a new trend they may influence other actors to also adopt it. This is certainly an intuitive and well-known phenomenon in

* Corresponding author.
E-mail addresses: fabian.riquelme@uv.cl (F. Riquelme), pgonzalez@instagis.com (P. Gonzalez-Cantergiani), xavier.molinero@upc.edu (X. Molinero), mjserna@cs.upc.edu (M. Serna).

https://doi.org/10.1016/j.knosys.2017.10.029
0950-7051/© 2017 Elsevier B.V. All rights reserved.

social network analysis [20]. The most known general models for influence spread are the *linear threshold model* [21] and the *independent cascade model* [21]. The linear threshold model is based on some ideas of collective behavior [22,23]. The independent cascade model was proposed in the context of marketing [24]. Most of the research effort has been devoted to the study of the influence maximization problem, under the linear threshold model and other models [25]. In this problem we attempt to find a set of k key actors that allow maximizing the influence spread among all sets of the same size. Indeed, the influence maximization problem under the linear threshold model is NP-hard [13]. The studies about derived centrality measures are scarce and consider only the independent cascade model [26–28]. Those rankings were proposed, and evaluated, to get good solutions to the influence maximization problem. In this context, the focus lies in the set of the k higher ranked users and the amount of influence that they can exert together.

So far we have mentioned centrality measures to rank the central users of the network. However, although less well known, there are also centralization measures, also known as *hierarchical measures* [29]. These measures aim to determine to what extent the entire network has a centralized structure. The most known centralization measure is the *Freeman centralization*, originally called simply *graph centrality* [3], that measures how central its most central node is in relation to how central all the other nodes are. It is a generic measure, so that each centrality measure can have its own associated centralization measure. Other measures of centralization are the *average clustering coefficient* (ACC) [30] and variations.

In this paper we want to analyze centrality measures based on the linear threshold model. We propose a new centrality measure to rank the users of the network, the *Linear Threshold Rank* (LTR), and a centralization measure associated to the linear threshold model, the *Linear Threshold Centralization* (LTC). The LTR measure can be interpreted as how much an actor can spread his influence within a network, investing resources to be able to convince his immediate neighbors. This distinguishes this influence measure from other classical measures such as the degree centrality. In this measure, an actor with small degree might have a good ranking due to his neighbors. The LTC measure is related to the *k-core*, a notion introduced to study the clustering structure of social networks [31] and to describe the evolution of random graphs [32]. The k-core has also been applied in bioinformatics [33,34] and network visualization [35], and it is a key concept for the *k-shell* decomposition method. It is known that the k-shell predicts the outcome of spreading more reliably than other centrality measures like the degree or the betweenness [36].

We are interested in analyzing whether those new measures differ or not from other centrality measures based on relevance or influence. For doing so we fix our attention in two relevance measures: the PageRank and the Katz centrality. For an influence based centrality we consider a measure naturally derived from the independent cascade model, the *Independent Cascade Rank* (ICR) introduced in [37]. These centrality measures are implemented using different approaches, so we also discuss the computational resources and the accuracy required by each algorithm. Our aim is to compare the different rankings as special purpose centrality measure without having in mind the influence maximization problem as it was done with the independent cascade proposed measures. As centralization measures we consider the average clustering coefficient and the local clustering coefficient.

We evaluate the proposed centrality and centralization measures on four social networks. Two of them are large networks: the Higgs network (directed) and the arXiv network (undirected) [38]. The other two are well known small networks: the Dining-table network (directed) [39,40] and the Dolphins social network (undirected) [41]. We correlate the four centrality measures by using both the Spearman and the Kendall correlation coefficients [42,43]. Table 5 summarizes the results. Each centrality measure provides a different centrality criteria except for the Dolphins social network where LTR, PageRank and Katz centrality tend to be similar. Observe that LTR and ICR do not appear to be correlated in any of the networks. This fact indicates another structural difference among the two models of influence spread. As we will see, LTR measure is a useful measure for ranking actors in a distinguishable way.

The paper is organized as follows. Section 2 briefly describes the related work regarding centrality and centralization measures for general social networks. Next section is devoted to influence graphs, which are social networks where the influence spread is exerted under the linear threshold model. Section 4 contains the main novelty of this paper, which is the definition of the new measures of centrality and centralization. Section 5 shows our experimental setting. We compare all the previous defined measures in four different networks. Finally, the paper ends up presenting our main conclusions and several directions for future work.

2. Preliminaries

In this section we introduce some known centrality measures and give some intuition about how they work. We also explain how to correlate centrality measures. Finally, we introduce centralization measures.

In all what follows, we consider a social network as a graph $G = (V, E)$, where $V(G)$ is the set of actors and $E(G)$ is the set of edges of G. Sometimes we require a weighted graph (G, w), where G is a graph and $w : E(G) \to \mathbb{N}$ is a *weight function* which assigns a weight to every edge. Let us denote $w((i, j)) = w_{ij}$ for any edge $(i, j) \in E(G)$, $n = |V|$, and $m = |E|$.

2.1. Centrality under relevance criteria

A widely used measure related with relevance criteria is the eigenvector centrality [12], which considers that an actor in the network is important if it is linked from other important actors or if it is highly linked. More formally, consider an adjacency matrix A, so that the elements (a_{ij}) of A take a value 1 if actor or node i is connected to actor j, and 0 otherwise. The *eigenvector centrality* of an actor u, denoted by $\textsc{ev}(u)$, is given by

$$\textsc{ev}(u) = \frac{1}{\lambda} \sum_{v \in V(G)} (a_{uv}) \ \textsc{ev}(v)$$

where λ is a constant called *eigenvalue*.

The eigenvector centrality provides reasonable results only if the graph is highly connected, like in the case of undirected networks with strongly connected components. In real directed networks, we can obtain several vertices with a null eigenvector centrality, so the measure becomes useless. For instance, this is the case for the vertices that can reach strongly connected components but that are not reachable from them.

Nevertheless, the Katz centrality [11] overcome this deficiency of the eigenvector centrality, by giving a small amount of centrality for free, regardless of the position of the actor in the network. The *Katz centrality* of an actor u, denoted by $\textsc{katz}(u)$, is given by

$$\textsc{katz}(u) = \alpha \sum_{v \in V(G)} (a_{vu}) \ \textsc{katz}(v) + \beta$$

where β is a constant which is independent of the network structure, and α is called the *damping factor*, a number between 0 and $\frac{1}{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest eigenvalue of A. Note that when $\alpha = \frac{1}{\lambda_{\max}}$ and $\beta = 0$, if we calculate $\textsc{ev}(u)$ with $\lambda_{\max}$, then $\textsc{katz}(u) = \textsc{ev}(u)$.

Additionally, there is a well-known centrality measure called *PageRank* [10], denoted for an actor u by $\textsc{pr}(u)$. It is given by

$$\textsc{pr}(u) = (1-\alpha) + \alpha \sum_{v \in V(G)} \frac{(a_{vu})\ \textsc{pr}(v)}{\delta^{+}(v)}$$

where $\delta^{+}(v)$ is the out-degree of node v and the damping factor α here is such that $0 < \alpha \leq 1$.

Although the Katz and PageRank centrality measures can be solved in polynomial time, a naive algorithm to solve them in $O(n^3)$ can be infeasible for a real network with millions of nodes. By avoiding the computation of $\lambda_{\max}$, the computational complexity of the Katz centrality can be reduced to $O(n+m)$ [44]. Furthermore, for sparse networks, PageRank can be computed in almost linear time [45].

Both centrality measures solve the problem of division by zero presented in the eigenvector centrality. One way to implement the algorithm is by using multiplication of matrices. However, for large networks this approach may be useless due to high memory consumption. Another way to implement these algorithms is by using mathematical methods like the power method, which may consume too much runtime, or even never converge. Standard computational tools, such as Python's NetworkX library,[1] try to avoid this second problem by adding two additional parameters: a tolerance parameter (`tol`), i.e., the number of significant digits that we want to accept in the results, and the maximum number of iterations (`max_iter`). Thus, the algorithm will stop after `max_iter` iterations, or after an error tolerance of $n \cdot \texttt{tol}$ has been reached, where n is the number of network actors. Sometimes, if the tolerance is too high, the algorithm may not converge, whereas if the number of iterations is too small, the algorithm may end up yielding partial results.

2.2. Centrality under influence criteria

Besides these measures based on the eigenvector centrality, there exist other centrality measures based on influence spread on the independent cascade model [26–28]. Under this model, each actor has a probability to influence the actors he targets. To spread its influence, the actor must be active, and it has only one chance to influence each actor. When an actor achieves to influence another one, this actor becomes active, and the process is repeated for this actor. The whole process ends when there are no active nodes with a new chance to spread its influence. As an example, consider the following measure [37], that we call *Independent Cascade Rank* (ICR):

$$\text{ICR}(u,p) = \frac{|F'(u,p)|}{\max\{|F'(v,p)| \mid v \in V(G)\}}$$

where V is the set of actors within the network, and $F'(u, p)$ is the influence spread process under the independent cascade model, starting from the activation of actor u. This measure considers the same constant probability p to influence every actor. Therefore, each actor has a probability $1-p$ to remain inactive from the influence of a neighbor, a probability of $(1-p)^r$ to remains inactive from the r actors pointing to it, and a total probability $1-(1-p)^r$ to becomes active from at least one of the actors pointing to it. To calculate this measure we use a public available addon for NetworkX.[2]

In the measures based on the independent cascade model, such as ICR, the activated nodes depend on a diffusion probability, so that the same measure can return different rankings for each execution.

2.3. Comparing centrality measures

In order to compare the results of two centrality measures, it is common to use statistic correlation. The most used coefficients to correlate centrality measures are the *Spearman's rank correlation coefficient* (ρ) [42] and the *Kendall Tau rank correlation coefficient* (τ) [43]. Let w and y be two lists of n users each, we have [46]:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)} \quad \text{and} \quad \tau = \frac{n_c - n_d}{0.5n(n-1)}$$

where x_i and y_i are the rankings of the users i in the lists x and y, respectively. Furthermore, n_c is the number of concordant pairs (i, j) (i.e., such that either $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$) and n_d is the number of discordant pairs, i.e., those that are not concordant. The values of both ρ and τ are in the $[-1, 1]$ interval, where 1 means that both measures are equal, 0 that they are completely independent, and -1 that one is the inverse of the other.

The computation of the correlation coefficients has an associated p-value. As usual, we consider the standard 0.05 cutoff as significance level, so that the null hypothesis is rejected when the p-value is lower than 0.05. Briefly speaking, our correlation results will be meaningful only when the p-value is lower than 0.05.

2.4. Centralization measures

For each centrality measure, the associated Freeman centralization measure requires to know the maximum possible sum of differences in centrality for all the possible networks. This can be easy for some centrality measures like degree, betweenness or closeness on undirected networks [3], but can be very difficult for other measures.

In addition, we can state the *average clustering coefficient* (ACC) [30] as a centralization measure, that corresponds to the average of the local clustering coefficients of all the vertices in the network. This measure does not take into account the direction of the edges, but differs whether the edges are weighted or not. It is defined by

$$\text{ACC} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

where C_i is the *local clustering coefficient* of actor i. For unweighted graphs, i.e., when all the edges have a weight equal to 1, C_i is the number of triangles $T(i)$ in which i participates normalized by the maximum possible number of such triangles:

$$C_i = \frac{2T(i)}{\delta(i)(\delta(i)-1)}$$

where $\delta(i)$ is the degree of actor i. For weighted graphs, there are several variations [47], but the NetworkX library uses this one:

$$C_i = \frac{1}{\delta(i)(\delta(i)-1)} \sum_{\substack{j,k \in V(G) \\ j \neq k}} (\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{1/3}$$

where $\hat{w}_{uv}$ is the weight of the edge (u, v), normalized by the maximum weight in the network, denoted by $\max(w)$, i.e., $\hat{w}_{uv} = w_{uv}/\max(w)$. For both formulae, $C_i = 0$ if $\delta(i) < 2$.

3. The linear threshold model for influence spread

A social network can be represented as a graph, whose nodes are the actors of the network, and the edges are interpersonal ties among the actors. Sometimes, edges have associated weights representing the strength of each interpersonal tie. Less common are the labels on nodes, which can be used to represent various network features. In this work, the labels represent the resistance of the actors to be influenced, whereas the weights of the edges are the power of influence exerted by one actor on another. A

[1] https://networkx.github.io/.

[2] https://github.com/hhchen1105/networkx_addon/tree/master/information_propagation.

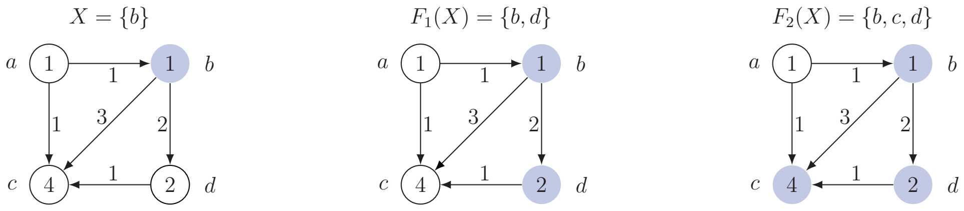


Fig. 1. Spread of influence (highlighted nodes) from the initial activation $X=\{b\}$.

weighted, labeled social network can be defined more formally as an influence graph [48]. In all what follows we consider the linear threshold model for influence spread.

Definition 1. An *influence graph* is a tuple (G, w, f), where $G=(V,E)$ is a digraph formed by a set of actors V and a set of directed edges E; $w: E\rightarrow\mathbb{N}$ is a *weight function* which assigns a weight to every edge, and $f: V\rightarrow\mathbb{N}$ is a *labeling function* that quantifies how easily influenceable each actor is. An actor $i\in V$ exerts influence over another actor $j\in V$ if and only if $(i, j)\in E$.

Note that an undirected graph can be seen as a symmetric digraph, so this measure also applies for that kind of graphs.

Given an influence graph (G, w, f) and an initial activation set $X\subseteq V$, consider the following iterative activation process. Let $F_t(X)\subseteq V$ be the set of nodes activated at some iteration t. Initially, at step $t=0$, only the nodes in X are activated, that is $F_0(X)=X$. At the next $t+1$ iteration, a node $i\in V$ will be activated if and only if

$$\sum_{j\in F_t(X)} w_{ji} \geq f(i).$$

The process stops when no additional activation occurs. In other words, a node i is activated when the weights' sum of the activated nodes connected to i is greater or equal to its resistance to be influenced.

Definition 2. Let (G, w, f) be an influence graph, the *spread of influence* of X is

$$F(X)=\bigcup_{t=0}^{k} F_t(X)=F_0(X)\cup\ldots\cup F_k(X)$$

where $k=\min\{t\in\mathbb{N}\mid F_t(X)=F_{t+1}(X)\}\leq n$. The t-value of $F_t(X)$ denotes the current *spread level* of X.

Note that for any initial activation set X, the influence spread $F(X)$ can be computed in polynomial time [48].

Example 1. Fig. 1 illustrates the spread of influence $F(X)$ in an influence graph from the initial activation $X=\{b\}$. In the first step we obtain $F_1(X)=\{b,d\}$ and in the second step (the last one), $F_2(X)=\{b,c,d\}$.

4. New measures of centrality and centralization

In this section we define the new influence measures based on the influence spread following the linear threshold model.

4.1. Linear threshold centrality

The linear threshold model, instead of the independent cascade model, does not depend on any chance, but on the capacity of influence of each actor, and their resistance to being influenced. We introduce a centrality measure to rank the users of the network as follows.

Definition 3. Let (G, w, f) be an influence graph, with $G=(V,E)$, $n=|V|$, and $i\in V$ an actor. The *Linear Threshold Rank* of i, denoted by $\textsc{ltr}(i)$, is given by

$$\textsc{ltr}(i)=\frac{|F(\{i\}\cup\text{neighbors}(i))|}{n}$$

where $\text{neighbors}(i)=\{j\in V\mid (i,j)\in E\vee(j,i)\in E\}$.

In our definition, G is a directed graph, but the neighbors of an actor i are the actors that are connected to i by an edge in any direction. This allow us to increase the initial activation. Observe that, those actors with small out-degree would not be able to spread their influence through the network, and thus the obtained measure would be similar to the degree centrality. Hence, we take as the initial activation the set $F_0(X)=\{i\}\cup\text{neighbors}(i)$. As $F(X)$ can be computed in polynomial time, $\textsc{ltr}(i)$ is polynomial time computable. Furthermore, as the ICR measure, and instead of PageRank and Katz centrality, LTR always converges.

Note that we could also consider other criteria regarding the neighbors of the actor i, e.g., $\text{neighbors}(i)=\{j\in V\mid (i,j)\in E\wedge(j,i)\in E\}$, or $\text{neighbors}(i)=\{j\in V\mid (i,j)\in E\}$, or $\text{neighbors}(i)=\{j\in V\mid (j,i)\in E\}$. Those sets could also be filtered with respect to the weights, in order to regulate the role of the central actor in the influence spread. For instance, we could consider in the initial activation only those neighbors j for which w_{ij} meets a given lower bound. Under the latter consideration, the influence spread $F(\{i\})$ of actor i could be more correlated with the LTR measure. However, if the restrictions to include neighbors is too high, the measure runs the risk of becoming similar to the degree centrality, which is a local measure that only considers a tiny portion of the entire network. We left such options as future work.

The LTR measure can be interpreted as how much an actor i can spread his influence within a network, investing resources outside the formal system to be able to convince his immediate neighbors, regardless of the edges directions. From a positive point of view, this resources investment can represent the capacity of the actor i to manage his contacts in the network. From a negative or questionable point of view, it could represent his ability to bribe.

In some sense, this measure considers the existence of two different networks. A formal network of known interpersonal ties, and an informal network, with relationships that actors can use in the formal network. This is a very common reality, to which we are constantly exposed. For example, on the Twitter network, thousands of users can interact in several ways to spread an event. The news will be viralized through the network, depending on the influence capacity of each actor interested in the news. Although the most influential users will generate a greater impact on the network, each user is free to comment about the event if she or he wishes. In this formal network, users cannot directly interfere with the decisions made by other users. However, the organizers of the event could use their "external" network of contacts (friendships, media, etc.) to help them spread the event. Thus, the initial activation set will commonly not be formed by only one actor, but by several ones, related in a way not evident to the formal network.

Note that the increase in the LTR measure does not depend so much on the degree of the nodes in the initial activation, as on the amount of influence the actors are able to exert together as a coalition. This distinguishes this influence measure from other classical

Please cite this article as: F. Riquelme et al., Centrality measure in social networks based on linear threshold model, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.029

Table 1
Passes during a football match.

	First period											Second period										
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}
P_1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
P_2	6	0	18	0	0	20	0	0	0	0	0	3	0	9	0	0	10	0	0	0	0	0
P_3	6	10	0	0	8	12	0	11	0	0	0	2	5	0	0	4	6	0	5	0	0	0
P_4	2	6	0	0	0	6	4	2	0	0	0	1	2	0	0	0	2	1	1	0	0	0
P_5	2	0	6	0	0	14	0	7	0	5	0	0	0	2	0	0	7	0	3	0	2	0
P_6	0	0	1	2	1	0	0	3	2	3	2	0	0	3	6	3	0	0	9	6	11	7
P_7	0	1	0	0	0	2	0	0	2	2	2	0	3	0	3	0	6	0	0	8	9	7
P_8	0	0	1	0	0	1	0	0	3	3	2	0	0	3	0	3	4	0	0	9	11	7
P_9	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	1	2	2	0	6	4
P_{10}	0	0	0	0	0	1	0	1	2	0	2	0	3	0	0	0	5	0	6	9	0	9
P_{11}	0	0	0	0	1	0	0	1	1	3	0	0	0	2	0	3	0	0	6	6	10	0
Received passes	16	18	27	3	10	56	4	25	10	19	9	6	13	19	9	13	41	3	32	38	50	34
q (75%)	12	14	20	2	8	42	3	19	8	14	7	5	10	14	7	10	31	2	24	29	38	26

Table 2
During the first (left) and the second (right) period of a football match, given a player with its neighbors, it indicates the new influential players after each step.

	First period			Second period			
	$F_0 = \{i\} \cup \{\text{neighbors}(i)\}$	F_1	F_2	$F_0 = \{i\} \cup \{\text{neighbors}(i)\}$	F_1	F_2	F_3
P_1	$\{P_1, P_2, P_3, P_4, P_5, P_{10}\}$	$\{P_6, P_7, P_8\}$	$\{P_9, P_{11}\}$	$\{P_1, P_2, P_3, P_4\}$	–	–	–
P_2	$\{P_1, P_2, P_3, P_4, P_6, P_7\}$	$\{P_5\}$	$\{P_8\}$	$\{P_1, P_2, P_3, P_4, P_6, P_7, P_{10}\}$	–	–	–
P_3	$\{P_1, P_2, P_3, P_5, P_6, P_8\}$	$\{P_4\}$	$\{P_7\}$	$\{P_1, P_2, P_3, P_5, P_6, P_8, P_{11}\}$	–	–	–
P_4	$\{P_1, P_2, P_4, P_6, P_7, P_8\}$	$\{P_5\}$	–	$\{P_1, P_2, P_4, P_6, P_7, P_8\}$	$\{P_3\}$	$\{P_5\}$	–
P_5	$\{P_1, P_3, P_5, P_6, P_8, P_{10}, P_{11}\}$	$\{P_4, P_7, P_9\}$	$\{P_2\}$	$\{P_3, P_5, P_6, P_8, P_{10}, P_{11}\}$	$\{P_9\}$	$\{P_7\}$	$\{P_4\}$
P_6	$\{P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_1\}$	–	$\{P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_1\}$	–	–
P_7	$\{P_2, P_4, P_6, P_7, P_9, P_{10}, P_{11}\}$	–	–	$\{P_2, P_4, P_6, P_7, P_9, P_{10}, P_{11}\}$	$\{P_3, P_8\}$	$\{P_1, P_5\}$	
P_8	$\{P_3, P_4, P_5, P_6, P_8, P_9, P_{10}, P_{11}\}$	$\{P_2, P_7\}$	$\{P_1\}$	$\{P_3, P_4, P_5, P_6, P_8, P_9, P_{10}, P_{11}\}$	$\{P_2, P_7\}$	$\{P_1\}$	–
P_9	$\{P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_4\}$	–	$\{P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_4\}$	–	–
P_{10}	$\{P_1, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_4\}$	–	$\{P_3, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_2, P_4\}$	$\{P_1\}$	–
P_{11}	$\{P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_4\}$	–	$\{P_3, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$	$\{P_2, P_4\}$	$\{P_1\}$	–

measures such as the degree centrality. In this measure, an actor with a little contribution in the influence spread may have a good ranking due to his neighbors. This is common in the spread of influence phenomenon, as well as in many centrality measures that consider the behavior of the entire network. For example, measures based on eigenvector centrality, such as PageRank, where an actor can be quite influential only because of its direct connection to other influential actors.

Example 2. Let $P_1, P_2, P_3, \ldots, P_{11}$ be the eleven players of a football team. P_1 is the goalkeeper, P_2 and P_3 are defenders, P_4, P_5, P_6, P_7 and P_8 are midfielders, and P_9, P_{10} and P_{11} are strikers. Table 1 lists the passes during the first and the second period of a football match. The first column indicates who kicks the ball, and the other columns indicate who receives the ball (for instance, in the first period, player P_2 kicks the ball 6 times to player P_1, 18 times to player $P3$ and 20 times to player $P6$). Note that we consider the quota q of each player as the 75% of total received passes, i.e., a player is *happy enough* (with some of their colleagues) when he/she has obtained the 75% of the total received passes: See the last row.

From Table 1, we can compute the influence of each player according to the LTR measure. Given a player i with its neighbors, i.e., $F_0 = \{i\} \cup \{\text{neighbors}(i)\}$, Table 2 shows the new influential players after each step F_1, F_2 and F_3 during the first period and the second period of the football match. On the one hand, during the first period of the football match the propagation from players P_1, P_5, P_6 and P_8 (together with their neighbors) extends to all of the other players. That is, the goalkeeper (player P_1) and three midfielders (players P_5, P_6 and P_8) are more influential than the other players according our criteria. On the other hand, during the second period there are more passes among attacking players (forwards) than during the first period. Now players P_6, P_7, P_8, P_{10} and P_{11} (together with their neighbors) influence to all the other players. That is, three midfielders (players P_6, P_7 and P_8) and two strikers (players P_{10} and P_{11}) are more influential than the other players.

4.2. Linear threshold centralization

Using the linear threshold model, we define a novel centralization measure to determine how centralized the entire network is. The *k-core* of a graph is the maximal subgraph such that every vertex has degree at least k. The *k-shell* is the subgraph of nodes in the k-core but not in the $(k+1)$-core. The *main core* is the core with the largest degree.

Definition 4. Let (G, w, f) be an influence graph, with $G = (V, E)$ and $n = |V|$. The *Linear Threshold Centralization* of G, denoted by $\text{LTC}(G)$, is given by

$$\text{LTC}(G) = \frac{|F(\hat{C}(G))|}{n}$$

where $\hat{C}(G) = \{i \in V \mid i \text{ belongs to the main core of } G\}$.

The justification for this measure is quite natural. As the actors outside the k-shell have a degree smaller than the actors inside of it, the first ones are more able to be influenced by the second ones.

Example 3. We have applied such measure to four networks. Fig. 2 summarizes such results emulating the usual k-shell decomposition visualization, i.e., the nodes are distributed in such a way that a node has larger degree as is closer to the center. Notice that the graph (a) is more concentrated in the center, so the initial activation is a large set that is more able to spread its influence over the distant nodes. Hence, if the nodes in the center are capable to spread its influence over the distant nodes, this would be a good

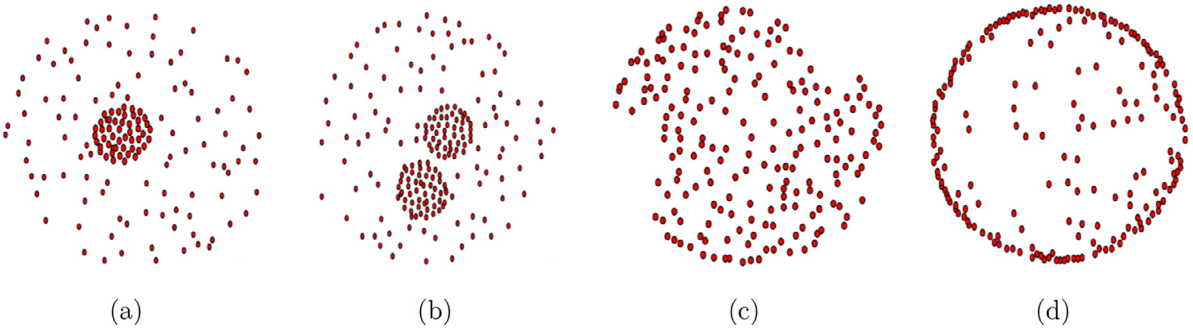


Fig. 2. Four networks with different centralizations.

example of a graph with a high Linear Threshold Centralization. The graph (b) presents two large cores instead of one. Although it has a high centralization, there are nodes of the surface that could be accessible only from one of the two cores, and therefore our measure can return lower values than in the first graph. Finally, as the main cores of both graphs (c) and (d) are lower, the initial activations may not be enough to spread their influence to the peripheral nodes, resulting in lower values for the centralization measure.

5. Experiments and results

In this section, we present the results of different influence measures on four different networks:

- Higgs network: a large, directed network.
- arXiv network: a large, undirected network.
- Dining-table network: a small, directed network.
- Dolphins social network: a small, undirected network.

We consider four centrality measures, namely the PageRank, Katz centrality, ICR, and our new Linear Threshold Rank (LTR). We also consider LTC and ACC as centralization measures.

The first two network datasets are provided by the SNAP's Stanford Large Network Dataset Collection [38].[3,4] The third one is available on Pajek datasets,[5] and the last one on the UCI Network Data Repository.[6] The experiments were programmed in the programming language Python 3. For the work with graphs we used the NetworkX library. All experiments were run on a machine HP ProLiant DL380p server with two Xeon(R) E5-2650 CPU.

Every network is represented as an influence graph (G, w, f). For each actor $i \in V$, we set a label $f(i) = \lfloor \bar{w}/2 \rfloor + 1$, where $\bar{w} = \sum_{(j,i) \in E(G)} w_{ji}$, so that an actor becomes active if either it belongs to the initial activation, or the active nodes pointing to it sum more than half of the total influence power pointing to it.

5.1. Higgs network

The first dataset contains all the tweets related with the Higgs boson experiment that were re-tweeted between 1st and 7th July 2012. The collection was initially used to study the information spreading processes on Twitter [49]. This dataset allows to generate an influence graph with $n = |V| = 256,491$ actors and $|E| = 328,132$ directed edges. A directed edge $(i, j) \in E$ represents an actor i retweeting an actor j, so we say that i exerts a certain influence over j. The weight of the edge, w_{ij}, represents how many times actor j retweeted actor i.

By definition, both the ICR and the LTR measures always converge. To compute the ICR measure we use a probability $p = 0.1$. To compute the PageRank we used the damping factor $\alpha = 0.85$ set by default, and the algorithm converged before the 100 iterations set also by default. To compute the Katz centrality we also used the values $\alpha = 0.1$ and $\beta = 1.0$ set by default. However, instead of the PageRank, with the standard values of tolerance and even one million of iterations, the Katz algorithm diverged. Therefore, to allow convergence using a feasible amount of time and memory resources, we decided to reduce the tolerance, from the standard six significant digits to only two.

Before comparing the results of the different influence measures, let us focus on the LTR measure. Table 3 shows how many actors reached a number of spread levels equal to 0, 1, 2, and so on. Note that almost the 90% of the actors do not reach the fourth spread level. However, the curve formed by these two variables does not have a strictly monotonous decrease. In fact, the most common actors are those who reach exactly two spread levels, and there is no actors who reach exactly eleven spread levels. Moreover, the maximum spread level of the actors correlates well with the LTR measure ($\rho = 0.89$ and $\tau = 0.77$). This means that for this network, the nodes with many neighbors are scarce, because if not, these nodes could have a high LTR score just by reaching low spread levels.

Now, we can continue with a more comparative analysis. Given a complex network, it is important that the centrality measures rank the actors adequately. On one hand, this means that the measures generate distinguishable classes of actors with different values. On the other hand, the differences in their values should be large enough so that they do not lend themselves to confusion. Taking this into account, the results of the four influence measures are illustrated in Fig. 3(a), where the abscissa axis represents the different nodes numbered from 0 to $n - 1$, and the ordinate axis represents the values obtained by each measure.

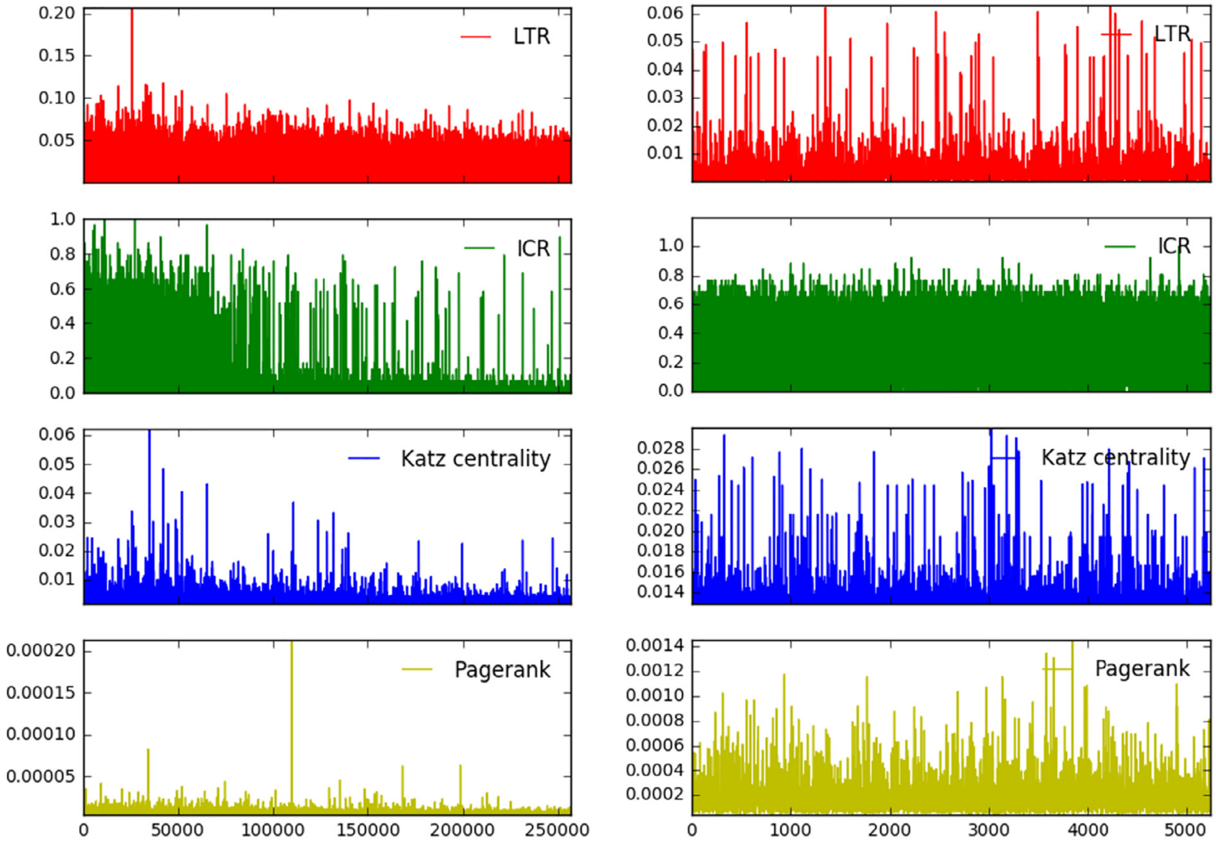
At first sight, Fig. 3(a) shows that the values with less variation are those of PageRank. Indeed, the PageRank presents the lowest standard deviation, but also the largest number of different values. Table 4 shows the standard deviation and the number of different values for the measures. There we include another three traditional centrality measures (closeness, betweenness, and degree) [3], which are based on the topology of the network, in order to broaden the perspective of analysis. Note that the ICR and the LTR measures present the highest standard deviation. However, the ICR measure returns only 30 different values, while the LTR measure has the second largest number of different values, after PageRank. Note that PageRank presents the largest number of different values, but also the lowest standard deviation. This means that this measure qualifies the actors with very similar values, so that if we would use fewer decimals for our calculations, their rankings would become equivalent. The degree measure, as expected, provides the worst results after ICR.

[3] http://snap.stanford.edu/data/higgs-twitter.html.
[4] http://snap.stanford.edu/data/ca-GrQc.html.
[5] http://vlado.fmf.uni-lj.si/pub/networks/data/esna/dining.htm.
[6] https://networkdata.ics.uci.edu/data.php?id=6.

Please cite this article as: F. Riquelme et al., Centrality measure in social networks based on linear threshold model, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.029

Table 3
Number of actors that, starting by the initial activation X formed by he/she together with their neighbors as Definition 3, reach a maximum spread level t, i.e., such that $F_t(X) = F_n(X)$.

max t	#actors	max t	#actors	max t	#actors	max t	#actors
0	42,129	3	51,118	6	211	9	8
1	63,189	4	11,516	7	171	10	4
2	73,260	5	14,861	8	22	12	2



(a) Higgs network's results

(b) arXiv network's results

Fig. 3. Results for the different centrality measures considered in the Higgs and the arXiv networks. The abscissa axis represents the different nodes numbered from 0 to $n1$, and the ordinate axis represents the values obtained by each measure.

Table 4
Standard deviation and number of different values for centrality measures on the different networks.

network	Higgs		arXiv		Dining-table		Dolphins	
measure	σ	#diff	σ	#diff	σ	#diff	σ	#diff
LTR	0.009133	7000	0.005950	169	0.293	13	0.259	34
ICR	0.035022	30	0.282791	22	0.276	6	0.257	10
Katz	0.000572	774	0.001613	4111	0.000	4	0.047	60
PageRank	0.000001	33,956	0.000132	3469	0.038	25	0.008	60
closeness	0.003662	4331	0.056596	2850	0.140	19	0.052	43
betweenness	0.000003	5015	0.001975	1564	0.036	21	0.051	54
degree	0.000152	375	0.001511	65	0.066	6	0.048	12

In addition, we have correlated the four influence measures by using both the Spearman and the Kendall coefficient correlations. Table 5a summarizes the results. All the correlations are low and significant, i.e., with a p-value lower than 0.05. The higher correlation was obtained by the Spearman coefficient between the LTR and the Katz measure, and it is only around 0.5. This means that each measure provides a different influence criteria, as expected. Furthermore, there are several negative correlations. However, the values are closer to 0 than −1. Thus we cannot say that these measures tend to rank the actors in the opposite way, rather this is an indication that they are not correlated.

Regarding execution time, both PageRank and Katz centrality returned all the results in 7.6145 and 7.7375 s, respectively. The PageRank was computed with a tolerance of six significant digits, and the Katz centrality with a tolerance of just two significant digits. For a tolerance of three significant digits, the Katz measure always diverged, or it was computing by weeks without providing any output. The execution time of both the LTR and the ICR mea-

Table 5
Correlation coefficients for the measures applied to the (a) Higgs, (b) arXiv, (c) Dining-table, and (d) Dolphins network. Lower triangular is for Spearman coefficient (ρ) and upper triangular is for Kendall coefficient (τ). Values with a p-value greater than 0.05 are strikethrough.

(a)					(b)				
$\rho \setminus \tau$	LTR	ICR	Katz	PageRank	$\rho \setminus \tau$	LTR	ICR	Katz	PageRank
LTR	1	−0.175	0.380	−0.335	LTR	1	0.317	~~−0.010~~	~~0.006~~
ICR	−0.214	1	−0.239	−0.229	ICR	0.456	1	~~0.006~~	~~0.014~~
Katz	0.502	−0.271	1	0.289	Katz	~~−0.014~~	~~0.008~~	1	~~0.009~~
PageRank	−0.299	−0.279	0.388	1	PageRank	~~0.009~~	~~0.020~~	~~0.013~~	1
(c)					(d)				
$\rho \setminus \tau$	LTR	ICR	Katz	PageRank	$\rho \setminus \tau$	LTR	ICR	Katz	PageRank
LTR	1	0.436	~~0.117~~	−0.499	LTR	1	~~0.170~~	0.784	0.688
ICR	0.543	1	~~0.257~~	~~−0.259~~	ICR	~~0.224~~	1	~~0.159~~	0.210
Katz	~~0.129~~	~~0.297~~	1	~~0.044~~	Katz	0.929	~~0.222~~	1	0.735
PageRank	−0.655	~~−0.348~~	~~0.046~~	1	PageRank	0.853	0.258	0.910	1

Table 6
Spread of influence from the initial activation formed by the nodes in the main core, i.e., $\hat{C}(G)$.

t	$F_t(\hat{C}(G))$	t	$F_t(\hat{C}(G))$	t	$F_t(\hat{C}(G))$	t	$F_t(\hat{C}(G))$
0	26,828	4	67,325	8	67,701	12	67,983
1	58,064	5	67,382	9	67,931	13	67,986
2	65,637	6	67,403	10	67,966	14	67,986
3	67,086	7	67,408	11	67,973	$\cdots$	$\cdots$

sures were in the order of hours, running with 16 parallel processes. Despite of the differences in time, we remark that both measures always converge and return exact values. Execution time could be reduced by using more parallelism.

Finally, regarding the centralization measures, the main core returns a subgraph with 57 nodes, with a largest degree equal to 12. The influence spread of $\hat{C}(G)$, which has 57 nodes, is shown in Table 6. As $|F_{13}(\hat{C}(G))| = |F(\hat{C}(G))| = 67,986$ and $n = 256,491$, we conclude that the Linear Threshold Centralization of the network is $\text{LTC}(G) = 67,986/256,491 = 0.265$. Moreover, the network's ACC is 0.0156. Since both values are closer to 0 than 1, we can say that this network is fairly decentralized.

5.2. arXiv network

The second dataset contains scientific collaborations between authors papers submitted to arXiv's General Relativity and Quantum Cosmology category.[7] The collection was initially used to study graph evolution [50]. This influence graph has $n = 5242$ actors and 14,496 undirected edges, so that for each edge $(i, j) \in E(G)$, $w_{ij} = 1$. An edge (i, j) represents that author i co-authored a paper with author j.

For this network we also computed the ICR measure with a probability $p = 0.1$. Here the PageRank also converges with the parameters by default. We try to use the power method for Katz calculation, but failed to converge using the default parameters, so we use the multiplication of matrices approach setting the α parameter to 0.01. Note that we cannot use this approach for the Higgs network due to the high memory consumption. The results of the centrality measures are illustrated in Fig. 3(b). In this case, the correlations obtained between the measures are even lower than in the previous case, although most of the results are not significant due to p-values greater than 0.05. From Table 5b we can only conclude that LTR and ICR have a small positive correlation. The highest correlation was obtained with Spearman coefficient between LTR and ICR, reaching a value 0.456. Hence, for this network the measures also provide different influence criteria.

The standard deviation and the number of different values are shown in Table 4. The PageRank and the Katz centrality provide the largest number of different values. However, as in the previous case, both measures have the smallest standard deviations among the considered influence measures. This means that with these measures we obtain different rankings, but also similar values. Again, ICR is the measure with the highest standard deviation but, at the same time, the one that rank worse the different actors. In contrast, LTR behaves as a more balanced measure, with a high standard deviation compared to the others, and returning almost eight times more values than LTR. Interestingly, in this case of undirected graph, the closeness returns good results, although they are not related to influence criteria.

Regarding execution time, PageRank goes back to have a good performance, taking just 4.25 s of computation with the power method. In this case, the Katz centrality was computed without the power method, so it returned its ideal results in 679 s, i.e., a little more than 11 min. As in the previous network, here the ICR measure took longer than the previous ones. It took 8.31 s on average per node, with a maximum of 27.39 s. Since it was computed in a parallel process with 16 threads, it took 2723.90 s in total, i.e., around 45 min. Remarkably, the LTR measure had a better performance even than the Katz centrality, taking only 23.83 s in total, i.e., just 0.0727 s on average per node.

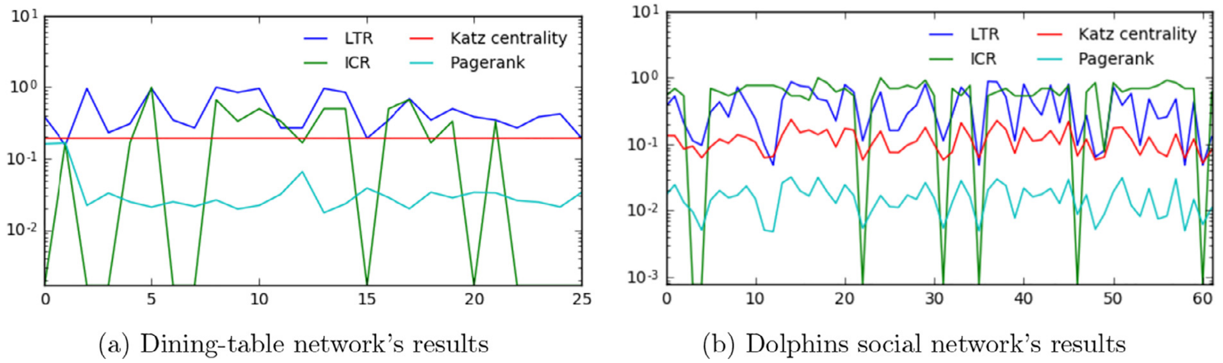
Finally, regarding the centralization measures, the ACC is 0.530, which means that the average clustering coefficient is much larger than the obtained in the Higgs network. Under this approach, we could say that the arXiv network is more centralized than the previous one. However, our new centralization measure provides a different conclusion. In this case, the main core returns a subgraph with 44 nodes, that correspond to the 0.8% of the nodes in the network. This is much more than the 0.022% obtained in the Higgs network. Even more, here the largest degree is 43, rather than 12, as in the previous study case. However, the influence spread of the main core takes 7 steps and it is just $|F(\hat{C}(G))| = 722$, in such a way that the Linear Threshold Centralization is $\text{LTC}(G) = 722/5242 = 0.138$. This value is lower than the Higgs network value. This means that the Higgs network has a main core smaller than the one of the arXiv network, but better connected. Notice also that ACC and LTC provide different centralization criteria.

5.3. Dining-table partners and Dolphins social network

The last two case studies are small networks, so they can be analyzed together.

The Dining-table partners network [39,40] is a directed influence graph with 26 vertices and 52 arcs. The vertices represent girls living in one cottage at a New York State Training School.

[7] https://arxiv.org/archive/gr-qc.



(a) Dining-table network's results

(b) Dolphins social network's results

Fig. 4. Results for the different centrality measures considered in the Dining-table and the Dolphins networks. The abscissa axis represents the different nodes numbered from 0 to n1, and the ordinate axis represents the values obtained by each measure.

Each girl was asked about who prefers as dining-table partner in first and second place. Therefore, each edge (i, j) represents girl i preferring girl j as dining-table partner. Every node has an out-degree equals 2: edges with weight 1 denote the first option of the girl, and edges with weight 2 denote her second option. This network can be easily modified as an influence graph [8]. We just replace each arc (i, j) by (j, i), assuming an influence from girl j to girl i. Further, a girl has more influence over another one if that other has chosen her in the first place rather than in the second place. Hence, the weights equals 1 are replaced in the corresponding inverted arc by 2, and the weights equals 2 are replaced by 1. Thus, now every node has an in-degree equals 2.

The Dolphin Social Network is an undirected network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand [41]. In this case, all the edges have a weight equal to 1.

In both cases, the PageRank and Katz centrality converge for the parameters by default. As before, the ICR measure was computed with a probability $p = 0.1$. The results were obtained for each measure in less than one second. The results are illustrated in Fig. 4. The standard deviation and the number of different values are shown in Table 4. For the first network, the Katz centrality is by far the most useless. Indeed, it returns only four different values, and it has a standard deviation of 7.216450e−17, which is almost zero. On the other hand, the pattern is repeated with respect to large networks: The LTR and ICR measures have the highest standard deviations, and of both, LTR always returns the largest number of different values.

The correlation results are shown in Table 5c and d, respectively. For the Dining-table network, the correlations are still low, although we remark an inverse correlation of −0.655 for the Spearman coefficient between LTR and PageRank. On the other hand, the Dolphins social network is the only study case that presents high correlation results. Considering only the significant results, we can note a high correlation between LTR and Katz centrality, between LTR and PageRank, and between PageRank and Katz centrality. Moreover, the Spearman coefficient presents higher correlation results than the ones given by the Kendall coefficient. This means that for this small, undirected network, LTR, PageRank and Katz centrality tend to be similar.

Finally, regarding the centralization measures, the ACC measure is 0.118 for the Dinning-table network, and 0.259 for the dolphins network. However, the LTC measure is 1 for both networks, because the main core is big enough (20 and 36 nodes, respectively) to spread its influence to all the remaining actors, just by considering its neighbors.

6. Conclusions and future work

In this paper we have focused on comparing a centrality measure derived in a natural way from the linear threshold model with generic centrality measures. Given a social network represented as an influence graph, we introduce the Linear Threshold Rank (LTR), a new centrality measure based on the linear threshold model, which is defined for each actor as the number of nodes that can be spread when he/she forms an initial activation with his/her neighbors.

We compare this measure with three different centrality measures based on influence criteria (the Katz centrality, the PageRank, and the Independent Cascade Rank), in four real case studies: two large networks (one directed and one undirected) and two small networks (one directed and one undirected). The larger network has 255,491 actors and 328,132 relationships among them. For the large networks, and even for the small, directed one, we show that the correlation among these measures is low, which means that they provide different influence criteria and can be used to obtain different results. In general, Table 5 shows that the absolute value of the correlation results tends to be higher for the Spearman coefficient than for the Kendall coefficient. Moreover, the new LTR measure, together with the Independent Cascade Rank (ICR), present the highest standard deviations. For these two measures, LTR returns a larger number of different values. This proves that the new LTR measure is a useful measure for ranking actors in a distinguishable way.

Besides the centrality measure, in this paper we also introduce a centralization measure called Linear Threshold Centralization (LTC), that corresponds to the number of actors that can be influenced by the ones that belong to the main core of the network. We proved that this measure is useful for networks with a large number of actors. The measure was also compared with the known average clustering coefficient (ACC). We conclude that both measures provide different centralization criteria, and can be used to provide different information about the network. For future work, it would be interesting trying to define the Freeman centralization for some centrality measures based on the influence criteria, in order to have additional centralization measures to be compared with LTC.

Regarding execution time, we checked experimentally that both the LTR and the ICR measures, although polynomial, do not seem to be suitable for computations in real time. However, despite centrality measures like Katz and PageRank, LTR and ICR always converge, and can be easily paralleled in order to decrease its execution times.

As the other measures use the tolerance as the maximum level of influence (k-value in Definition 2) the maximum level of the LTR measure could be bounded upperly, in order to obtain a significant performance improvements. In this line, it is also interesting to define other new measures and to study their behavior. For instance, we could define the *kth Linear Threshold Rank* as

$$\text{LTR}_k(i) = \frac{|F_k(\{i\} \cup \text{neighbors}(i))|}{n}$$

for any $0 \le k \le n$. Note that $\text{LTR}_n(i) = \text{LTR}(i)$, where n is the number of players. It leaves open other new interested measures based on linear threshold model to be studied.

There are several lines for future research. The first one is to compare the LTR with other rankings based in the same mechanism of spread of influence. As we mention before the definition of neighborhood of a vertex used to define the initial activation set should be contrasted with other options. Besides of edge directions, we could consider neighbors at a certain distance. Other direction is to relax the tolerance on the level of influence. A throughout study will shed light on which of the two parameters has a highest impact in centrality.

A second line of research concerns the study of the suitability of LTR with respect to the maximum influence spread problem. In this context LTR and other measures as the proposed before should be compared with independent cascade based rankings to see which one provides the best estimator for the spread of influence. In such context we plan to use the above measures together with the proposed in [27–29] to perform a study on the impact of the first k ranked nodes as measure of centrality. Those results should also be compared to the LTC measure proposed in this paper.

Finally, we want to mention the difficulty in finding networks with edge weights and node labels. Therefore it would be of interest to find procedures that allow the labeling of the networks suitably for a process of influence spread. In this sense methods like the ones proposed in [51–54] might be adapted to the influence context.

Acknowledgments

Fabián Riquelme was partially funded by project PMI USA1204. Xavier Molinero is partially funded by grant MTM2015-66818-P from the Spanish Economy and Competitiveness Ministry. Maria Serna is partially funded by Ministerio de Economía y Competitividad under grants TIN2013-46181-C2-1-R (COMMAS) and MDM-2014-044 (BGSMath), and Generalitat de Catalunya, Agència de Gestió d'Ajuts Universitaris i de Recerca, under project 2014 SGR 1034 (ALBCOM).

References

[1] A. Bavelas, A mathematical model for group structures, Hum. Organ. 7 (3) (1948) 16–30.
[2] J. Sun, J. Tang, A survey of models and algorithms for social influence analysis, in: C.C. Aggarwal (Ed.), Social Network Data Analytics, Springer, 2011, pp. 177–214.
[3] L. Freeman, Centrality in social networks: conceptual clarification, Soc. Netw. 1 (3) (1979) 215–239.
[4] L. Freeman, S. Borgatti, D. White, Centrality in valued graphs: a measure of between-ness based on network flow, Soc. Netw. 13 (2) (1991) 141–154.
[5] D. Gómez, J.R. Figueira, A. Eusébio, Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems, Eur. J. Oper. Res. 226 (2) (2013) 354–365.
[6] R. Narayanam, Y. Narahari, A Shapley value-based approach to discover influential nodes in social networks, IEEE Trans. Autom. Sci. Eng. 8 (1) (2011) 130–147.
[7] T.P. Michalak, K.V. Aadithya, P.L. Szczepanski, B. Ravindran, N.R. Jennings, Efficient computation of the Shapley value for game-theoretic network centrality, J. Artif. Intell. Res. (JAIR) 46 (2013) 607–650.
[8] X. Molinero, F. Riquelme, M.J. Serna, Power indices of influence games and new centrality measures for agent societies and social networks, in: C. Ramos, P. Novais, C.E. Nihan, J.M.C. Rodríguez (Eds.), Ambient Intelligence - Software and Applications - 5th International Symposium on Ambient Intelligence, ISAmI 2014, Salamanca, Spain, June 4–6, 2014, Advances in Intelligent Systems and Computing, 291, Springer, 2014, pp. 23–30.
[9] F. Riquelme, P. Gonzalez-Cantergiani, Measuring user influence on Twitter: a survey, Inf. Process. Manage. 52 (5) (2016) 949–975.
[10] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford Digital Library, 1999.
[11] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43.
[12] P. Bonacich, Factoring and weighting approaches to clique identification, J. Math. Sociol. 2 (1972) 113–120.
[13] P.M. Domingos, M. Richardson, Mining the network value of customers, in: D. Lee, M. Schkolnick, F.J. Provost, R. Srikant (Eds.), Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 26–29, 2001, ACM, 2001, pp. 57–66.
[14] D. Gruhl, D. Liben-Nowell, R.V. Guha, A. Tomkins, Information diffusion through blogspace, SIGKDD Explor. 6 (2) (2004) 43–52.
[15] L.A. Adamic, E. Adar, How to search a social network, Soc. Netw. 27 (3) (2005) 187–203.
[16] X. Song, B.L. Tseng, C. Lin, M. Sun, Personalized recommendation driven by information flow, in: E.N. Efthimiadis, S.T. Dumais, D. Hawking, K. Järvelin (Eds.), SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6–11, 2006, ACM, 2006, pp. 509–516.
[17] X. Zhang, J. Zhu, Q. Wang, H. Zhao, Identifying influential nodes in complex networks with community structure, Knowl.-Based Syst. 42 (2013) 74–84.
[18] J. Li, W. Peng, T. Li, T. Sun, Q. Li, J. Xu, Social network user influence sense–making and dynamics prediction, Expert Syst. Appl. 41 (11) (2014) 5115–5124.
[19] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (2015) 64–68.
[20] D.A. Easley, J.M. Kleinberg, Networks, Crowds, and Markets - Reasoning About a Highly Connected World, Cambridge University Press, 2010.
[21] D. Kempe, J.M. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, Theory Comput. 11 (2015) 105–147.
[22] M. Granovetter, Threshold models of collective behavior, Am. J. Soc. 83 (6) (1978) 1420–1443.
[23] T. Schelling, Micromotives and Macrobehavior, Norton, 1978.
[24] J. Goldenberg, B. Libai, E. Muller, Using Complex Systems Analysis to Advance Marketing Theory Development, Technical Report, Academy of Marketing Science Review, 2001.
[25] W. Liu, K. Yue, H. Wu, J. Li, D. Liu, D. Tang, Containment of competitive influence spread in social networks, Knowl.-Based Syst. 109 (2016) 266–275.
[26] S.K. Kundu, C.A. Murthy, S.K. Pal, A new centrality measure for influence maximization in social networks, in: S.O. Kuznetsov, D.P. Mandal, M.K. Kundu, S.K. Pal (Eds.), Pattern Recognition and Machine Intelligence - 4th International Conference, PReMI 2011, Moscow, Russia, June 27, - July 1, 2011. Proceedings, Lecture Notes in Computer Science, 6744, Springer, 2011, pp. 242–247.
[27] S.K. Pal, S.K. Kundu, C.A. Murthy, Centrality measures, upper bound, and influence maximization in large scale directed social networks, Fundam. Inform. 130 (3) (2014) 317–342.
[28] I. Gaye, G. Mendy, S. Ouya, D. Seck, New centrality measure in social networks based on independent cascade (IC) model, in: I. Awan, M. Younas, M. Mecella (Eds.), 3rd International Conference on Future Internet of Things and Cloud, FiCloud 2015, Rome, Italy, August 24–26, 2015, IEEE Computer Society, 2015, pp. 675–680.
[29] C. Lozares, P. López-Roldán, M. Bolibar, D. Muntanyola, The structure of global centrality measures, Int. J. Soc. Res. Methodol. 18 (2) (2015) 209–226.
[30] D. Watts, S. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[31] S.B. Seidman, Network structure and minimum degree, Soc. Netw. 5 (3) (1983) 269–287.
[32] B. Bollobás, The evolution of random graphs, Trans. Am. Math. Soc. 286 (1) (1984) 257–274.
[33] G.D. Bader, C.W.V. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinform. 4 (2003) 2.
[34] M. Altaf-Ul-Amin, K. Nishikata, T. Koma, T. Miyasato, Y. Shinbo, M. Arifuzzaman, C. Wada, M. Maeda, T. Oshima, H. Mori, S. Kanaya, Prediction of protein functions based on k-cores of protein-protein interaction networks and amino acid sequences, Genome Inform. 14 (2003) 498–499.
[35] M. Gaertler, M. Patrignani, Dynamic analysis of the autonomous system graph, in: IPS 2004, International Workshop on Inter-domain Performance and Simulation, 2004, pp. 13–24.
[36] M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Stanley, H. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (2010) 888–893.
[37] D. Kempe, J.M. Kleinberg, É. Tardos, Influential nodes in a diffusion model for social networks, in: L. Caires, G.F. Italiano, L. Monteiro, C. Palamidessi, M. Yung (Eds.), Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11–15, 2005, Proceedings, Lecture Notes in Computer Science, 3580, Springer, 2005, pp. 1127–1138.
[38] J. Leskovec, A. Krevl, SNAP datasets: stanford large network dataset collection, 2017, (http://snap.stanford.edu/data).
[39] J.L. Moreno, The Sociometry Reader, The Free Press, 1960.
[40] W. de Nooy, A. Mrvar, V. Batagelj, Exploratory social network analysis with Pajek, Cambridge University Press, 2004.
[41] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations - can geographic isolation explain this unique trait? Behav. Ecol. Sociobiol. 54 (2003) 396–405.

Please cite this article as: F. Riquelme et al., Centrality measure in social networks based on linear threshold model, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.029

[42] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1904) 88–103.

[43] M. Kendall, A new measure of rank correlation, Biometrika 30 (1/2) (1938) 81–93.

[44] K.C. Foster, S.Q. Muth, J.J. Potterat, R.B. Rothenberg, A faster katz status score algorithm, Comput. Math. Organ.Theory 7 (4) (2001) 275–285.

[45] A.N. Langville, C.D. Meyer, Survey: deeper inside Pagerank, Internet Math. 1 (3) (2003) 335–380.

[46] S. Ye, S. Wu, Measuring message propagation and social influence on Twitter.com, Int. J. Commun. Netw.Distrib. Syst. 11 (1) (2013) 59–76.

[47] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, J. Kertész, Generalizations of the clustering coefficient to weighted complex networks, Phys. Rev. E 75 (027105) (2007) 1–4.

[48] X. Molinero, F. Riquelme, M.J. Serna, Cooperation through social influence, Eur. J. Oper. Res. 242 (3) (2015) 960–974.

[49] M. De Domenico, A. Lima, P. Mougel, M. Musolesi, The anatomy of a scientific rumor, Sci. Rep. 3 (2980) (2013) 1–9.

[50] J. Leskovec, J.M. Kleinberg, C. Faloutsos, Graph evolution: densification and shrinking diameters, TKDD 1 (1) (2007) 2.

[51] B.Q. Truong, A. Sun, S.S. Bhowmick, Content is still king: the effect of neighbor voting schemes on tag relevance for social image retrieval, in: H.H. Ip, Y. Rui (Eds.), International Conference on Multimedia Retrieval, ICMR '12, Hong Kong, China, June 5–8, 2012, ACM, 2012, p. 9.

[52] D.J. Crandall, D. Cosley, D.P. Huttenlocher, J.M. Kleinberg, S. Suri, Feedback effects between similarity and social influence in online communities, in: Y. Li, B. Liu, S. Sarawagi (Eds.), Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24–27, 2008, ACM, 2008, pp. 160–168.

[53] Y. Zhou, L. Liu, Social influence based clustering of heterogeneous information networks, in: I.S. Dhillon, Y. Koren, R. Ghani, T.E. Senator, P. Bradley, R. Parekh, J. He, R.L. Grossman, R. Uthurusamy (Eds.), The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11–14, 2013, ACM, 2013, pp. 338–346.

[54] Z. Li, J. Tang, Weakly supervised deep matrix factorization for social image understanding, IEEE Trans. Image Process. 26 (1) (2017) 276–288.

Please cite this article as: F. Riquelme et al., Centrality measure in social networks based on linear threshold model, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.10.029