ARTICLE IN PRESS
JID: KNOSYS [m5G;October 30, 2017;12:59]

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/

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

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 $\textsc{katz}(u) = \textsc{ev}(u)$.

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