Thresholds as Mechanisms for Weighting Influence in the Linear Threshold Rank*

Maria J. Blesa[0000−0001−8246−9926], Alejandro Dominguez-Besserer, and Maria Serna[0000−0001−9729−8648]

Universitat Politècnica de Catalunya – BarcelonaTech
08034 Barcelona, Spain
`{maria.j.blesa, maria.serna}@upc.edu`

Abstract. Social networks are the natural space for the spreading of information and influence and have become a media themselves. Several models capturing that diffusion process have been proposed, most of them based on the Independent Cascade (IC) model or on the Linear Threshold (LT) model. The IC model is probabilistic while the LT model relies on the knowledge of an actor to be convinced, reflected in an associated individual threshold. Although the LT-based models contemplate an individual threshold for each actor in the network, the existing studies so far have always considered a threshold of 0.5 equal in all actors (i.e., a simple majority activation criterion).

Our main objective in this work is to start the study on how the dissemination of information on networks behaves when we consider other options for setting those thresholds and how many network actors end up being influenced by this dissemination. For doing so, we consider a recently introduced centrality measure based on the LT model, the Forward Linear Threshold Rank (FwLTR), which is the natural interpretation of the Linear Threshold Rank on directed networks.

We experimentally analyze the ranking properties for several networks in which the influence resistance threshold follows different schemes. Here we consider three different schemes: (1) uniform, in which all players have the same value; (2) random, where each player is assigned a threshold u.a.r. in a prescribed interval; and (3) determined by the value of another centrality measure on the actor. Our results show that the selection has a clear impact on the ranking, even quite significant and abrupt in some cases. We conclude that the social networks ranks that provide the best assignments for the individual thresholds are FwLTR and the well-known PageRank.

Keywords: Social Networks · Centrality measures · (Forward) Linear Threshold Rank · Social Media Analytics.

* Supported by MCIN/AEI/10.13039/501100011033 under grant PID2020-112581GB-C21 (MOTION).

1 Introduction

Nowadays the importance of social networks as a marketing tool is growing rapidly and spanning through diverse social networks and objectives. Think for example about viral marketing, that seeks to spread information about a product or service from person to person by word of mouth or sharing via the internet or email. The goal of viral marketing is to inspire individuals to share a marketing message (or opinion) to friends, family, and other individuals to create exponential growth in the number of its recipients. The so-called *spread of influence* process is used here with a very specific goal and needs to be maximized until it stabilizes and no additional convictions are possible. How to select the initial group of individuals that start the spreading process is a key issue.

Traditional social networks are labelled graphs in which the nodes represent individuals and the edges represent weighted relations between them. There have been several proposal for modeling the spread of influence in a network. Among them, the *Independent Cascade* (IC) model is a stochastic model proposed in [7]. It is based on the assumption that whenever a node is activated, it will do (stochastically) attempt to activate a neighbor. The Linear Threshold (LT) model is a deterministic model for influence spread based on some ideas of collective behavior [9]. In the LT model the strength of the tie between every pair of actors quantifies the capacity of one to influence the other and, additionally, each actor opposes a resistance to be influenced. A node gets influenced when their active predecessors can exert enough influence to overpass its resistance. The node resistance is quantified by a threshold that quantifies the minimum required percentage of the total income weight needed to convince the actor. Several computational problems have been proposed that attempt to solve the problem of selecting the initial set of participants to spread the product in the most effective. The two most relevant being the *influence maximization* [9, 17] and the *target set selection problems* [4]. Another line of research uses the spread of influence process to define centrality measures [2, 14].

Trying to experimentally evaluate centrality measures based on the linear threshold model, we came with a lack of suitable data sets. Most of the social networks you can find in repositories do not include any kind on information about the resistance of a participant to be influenced. Therefore, as no information about suitable threshold values to run the linear threshold model was collected, a single majority rule was assumed. However this rule might not be a realistic one in many scenarios. In this paper, we attempt to shed some light on the effect that the selection of threshold has on one of the centrality measures based on the LT model. We focus our analysis in the FwLTR rank which measures the influence that a node together with its immediate successors can exert [2]. FwLTR is the natural interpretation of the Linear Threshold Rank (LTR) on directed networks.

We performed a series of experiments in order to assess how the threshold values affect the FwLTR. We considered a subset of the social networks used in [14]. To evaluate the effect of the threshold selection, we fix the method to assign the influence thresholds and compute the FwLTR ranking. Over the ranking, we compute the traditional statistics and then we select some relevant sets of participants, the ten in the top of the ranking, the 10% in the top, and the participants whose rank is in the 10% of the highest values, and compute the number of nodes influenced by those sets.

We run three kind of experiments. In the first one we assign to each node a common threshold θ. This type of uniform assignment was the usual scheme used in the literature until now, with $\theta = 0.5$ to implement a simple majority rule. However, we try different values for θ others than 0.5. In the second experiment we assign random thresholds in different ranges. In the third experiment we follow a completely different schema and use the value of another centrality measure on the node to determine its threshold.

Besides more specific comments given later, we comment here on the main discovered trends. Our first set of experiments clearly show that, in most of the treated networks, the range of threshold values more relevant to the maximization of the influence is within $[0.2, 0.5]$, where it is possible to accumulate greater proportion of the total influence diffused in groups of actors of small size. Interestingly enough, we found that a random distribution of the thresholds behaves practically the same with respect to our measures for the diffusion of influence whether they are in $(0, 0.5]$ or $(0, 1]$. This suggests that the fact that a minority fraction of the actors have high thresholds affects little in the actor's ability to influence in general. Finally, experiments also suggest that the FwLTR centrality and the well-known PageRank provide the best thresholds, in the sense that they generate favorable conditions for a great influence expansion even when starting with reduced sets. In nearly all cases, it is enough to consider the first 10 top-ranked actors to reach the entire network, or the vast majority of it.

2 Centrality Measures

We outline the centrality measures used in different parts of this paper. We have focused only on some of the most significant ones, and refer the reader to existing surveys (e.g., [16]) for a wider overview on other measures. As we will see, some of them are based on the relevance and the topological properties of the nodes, e.g., PageRank, while others focus on their influence over the network, e.g., the Linear Threshold Rank and the Independent Cascade Rank. We start describing the topology-based centrality measures, which were the ones used traditionally as centrality measures, and continue describing the influence-based ones that have been proposed more recently. As usual, we assume that the social network is represented by a directed graph $G = (V, E)$.

Betweenness - In the betweenness centrality, a node is more important if it belongs to the shortest path between any pair of nodes in the graph [5]. For every node $i \in V$, define

$$\mathsf{Btwn}(i) = \sum_{s,t \in V - \{i\}} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths between s y t, and $\sigma_{st}(i)$ the number of such shortest paths that pass by i.

PageRank - One of the most popular centrality measures is the PageRank [13], that Google uses to assign relevance to web pages. A web page is more relevant if other important web pages point to it. It uses a parameter $\alpha \in (0, 1]$, that represents the probability that a user keeps jumping from a web page to another through the links that are between them (and thus, $1-\alpha$ represents the probability that the user goes to a random web page). Let A be the adjacency matrix of G (i.e., $a_{ij} = 1$ if $(i, j) \in E$, and 0 otherwise), the PageRank (PR) of i is given by

$$\mathsf{PgR}(i) = (1 - \alpha) + \alpha \sum_{j \in V} \frac{a_{ji}\ \mathsf{PgR}(j)}{\delta^+(j)}$$

where $\delta^+(i)$ is the out degree of $i \in V$.

Perhaps the two most prevalent diffusion models in computer science are the *Independent Cascade* model [7] and the *Linear Threshold* model [9] (see also [15]). We define their corresponding influence-based centrality measures:

Independent Cascade Rank - It is an influence-based centrality measure [10] based on the Independent Cascade (IC) Model [7], which is a stochastic model. It is based on the assumption that whenever a node is activated, it will (stochastically) do attempt to activate each actor he targets. Given an activated node $i \in V$, any neighbor j such that $(i, j) \in E$ will be activated with a probability p_{ij}. When a new actor is activated, 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.

Given an initial core $X \subseteq V$ and a probability $p \in [0, 1]$ (where $\forall (i, j) \in E : p_{ij} = p$), the influence spread of X is denoted by $F'(X, p)$. The Independent Cascade Rank of a node $u \in V$ is then defined as:

$$\mathsf{ICR}(u, p) = \frac{|F'(u, p)|}{\max_{v \in V}\{|F'(v, p)|\}}.$$

Linear Threshold Rank - It is based on the Linear Threshold (LT) Model [9]. Every node has an influence threshold, which represents the resistance of this node to be influenced by others. In the model, every edge (u, v) also has a weight representing the influence that node u has over node v. In practice, that weight is set to one since it is very different to quantify that concept.

The influence algorithm starts with an initial predefined set of activated nodes. At every iteration, the active nodes will influence their neighbors. When the total influence that a node receives exceeds its influence threshold, then this node will become active and join the set of active nodes. The algorithm stops when the set of active nodes converges, i.e., when no new nodes are influenced.

Given an initial set of active nodes $X \subseteq V$ and a threshold assignment $\theta : V \to \mathbb{N}$, let $F_t(X) \subseteq V$ denote the set of activated nodes at the t-th iteration of the spreading process. At the first step ($t = 0$) only the nodes in X are active, which means that $F_0(X) = X$. At the $t + 1$ iteration, a node i will be activated if, and only if, the sum of all the weights of the edges $\{i, j\}$, where j is already active, is higher than the resistance (or influence threshold) of i, i.e., $\sum_{j \subseteq F_t(X)} w_{ij} \geq \theta(i)$, In practice, it is usual to consider that all the edge weights are equal to one and thus a node i will be activated if, and only if,

$$\frac{\mid F_t(X) \cap \mathcal{N}(i) \mid}{|\mathcal{N}(i)|} \geq \theta(i),$$

where, $\mathcal{N}(u) = \{v \mid (u, v) \in E \vee (v, u) \in E\}$. Observe, that the process is monotonic, therefore it stops after at most $n = |V|$ steps. Thus can define the spread of X as $F(X) = F_n(X)$. The *Linear Threshold Rank* [14] of a node $i \in V$ is given by

$$\mathsf{LTR}(i) = \frac{|F(\{i\} \cup \mathcal{N}(i))|}{n}.$$

Forward Linear Threshold Rank - It is a centrality measure very similar to the LTR, but with a different initial activation set [2]. Formally, let $\mathcal{N}^+(i) = \{j \in V \mid (i, j) \in E\}$, then the Forward Linear Threshold Rank of a node $i \in V$ is given by

$$\mathsf{FwLTR}(i) = \frac{|F(\{i\} \cup \mathcal{N}^+(i))|}{n}.$$

3 Statistical and Top Measures

For analysing the results of a centrality measure on its own, three statistical metrics will be used later in the experimental part: the number of different ranks assigned (#), their standard deviations (σ) from the mean and, Gini coefficient of the ranks.

The Gini coefficient [3, 6] comes originally from the field of sociology as a measure of the inequality of populations with respect to different criteria (e.g., wealth spread). It is often represented graphically through the Lorenz curve [11], which shows wealth distribution by plotting the population percentile by income on the horizontal axis and cumulative income on the vertical axis. The Gini coefficient is equal to the area below the line of perfect equality (i.e., 0.5) minus the area below the Lorenz curve, divided by the area below the line of perfect equality.

Definition 1. *Given a list of values X of size n, the Gini coefficient of X is calculated as follows:*

$$\mathit{Gini} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n \sum_{i=1}^{n} x_i}$$

The Gini coefficient is a value in $[0, 1]$, where 0 indicates a completely equitable distribution of values, and 1 represents the most radical inequality. This coefficient is lately being very much used in data science as a measure for quantifying the fairness of data distributions coming from other scientific areas.

In [8] the case of selecting sets of actors with high centrality values was studied as initial trigger sets. Specifically, they used the Degree, Betweenness and Eigenvector centralities, but a mixed influence diffusion model was used instead of the usual IC model or LT model, and only two networks of a relatively small size (of the order of a thousand nodes), both of academic collaboration (similar to the ArXiv network that we use here). However, we were inspired by the idea in that article for the introduction of three metrics related to the maximization of influence, and we study some variants that use measures of centrality other than the FwLTR.

The following variables are conceived as natural measures that can allow us to see if the individual actors with the highest FwLTR in the network are able to exert an amount of influence meaningfully together. We define:

Top10 - This parameter is the proportion of actors that are influenced if we execute the influence expansion algorithm taking as initial activating set the 10 actors with the highest FwLTR, i.e.,

$$\mathsf{Top10}(G) = \frac{|F(X_{10})|}{n}$$

where X_{10} is the 10 actors of G with the highest FwLTR.

Top10%Actors - From the FwLTR ranking of actors, let Y_{10} be the 10% with the greatest capacity to disseminate direct influence. Then,

$$\mathsf{Top10\%Actors}(G) = \frac{|F(Y_{10})|}{n}$$

Top10%Values - Analogously, let Z_{10} be the set of actors whose influence make up the first 10% of different ranking values (ties included). Then,

$$\mathsf{Top10\%Values}(G) = \frac{|F(Z_{10})|}{n}$$

4 Networks

Table 1 summarizes the characteristics of the networks considered for our experiments. The structural characteristics of the networks are described by seven common attributes: the number of vertices, the number of edges, whether the graph is weighted, whether the graph is directed, the average clustering coefficient, and the size of the main core.

The *average clustering coefficient* (ACC) is the average of the local clustering coefficients in the graph. The local clustering coefficient C_i of a node i is the number of triangles T_i in which the node participates normalized by the maximum number of triangles that the node could participate in.

$$\mathrm{ACC} = \frac{1}{n}\sum_{i=1}^{n} C_i, \quad \text{where } C_i = \frac{T_i}{\delta_i(\delta_i - 1)}$$

where δ_i is the degree of the node i, and $n = |V|$. Given a graph G and $k \in \mathbb{Z}^+$, a k-*core* is the maximal induced subgraph of G where every node has at least degree k. The *main core* is a k-core of G with the highest k.

Table 1. Characteristics of the real networks under consideration (in alphabetical order). ACC = Average Clustering Coefficient, MC = size of the main core. When the diameter is ∞, the diameter of the biggest connected component is provided.

Data set	n	m	Directed?	Edge-weighted?	ACC	Diameter	MC
Amazon	334863	925872	✗	✗	0.3967	44	497
ArXiv	5242	14496	✗	✗	0.5296	∞ (17)	44
Caida	26475	106762	✓	✓	0.2082	17	50
ENRON	36692	183831	✗	✗	0.4970	11	275
Epinions	75879	508,837	✓	✗	0.1378	14	422
Gnutella	62586	147892	✓	✗	0.0055	11	1004
Higgs	256491	328132	✓	✓	0.0156	19	10
Wikipedia	7115	103689	✓	✗	0.1409	7	336

5 Experiments

Although the LT model contemplates an individual threshold for each actor in the network, the existing studies so far have always considered a threshold of 0.5 equal in all actors. In practice, this means that a majority activation criterion is assumed: a node is activated (or influenced) when at least half of its neighbors have been activated. Our main objective in this work is to start the study on how the dissemination of information on networks is affected when we consider other options for setting those thresholds, how many network actors end up being influenced by that dissemination process, and to which extend the top-ranked actors guarantee that dissemination.

Among the experiments performed, we would like to point out here the most interesting ones. The first experiment still considers the same threshold for all the actors but, apart from the standard 0.5, we consider other values in $[0, 1]$. The second experiment sets random thresholds to the actors following a random uniform distribution. Finally, the third experiment sets to each node a threshold that depends on its centrality value according to several different measures. All the algorithms were implemented in ANSI C++, using GCC 7.5.0 for compiling the code. The experimental evaluation was performed on a cluster of computers with Intel® Xeon® CPU 5670 CPUs of 12 nuclei of 2933 MHz and (in total) 32 Gigabytes of RAM.[1]

[1] Infrastructure available at the Research & Development Lab (**RD**lab) of the Universitat Politècnica de Catalunya - BarcelonaTech.

5.1 Uniform Thresholds

The first experiment carried out consists in calculating the FwLTR of the network actors for the cases in which the threshold is the same for each of them and is defined as $1/4$, $1/2$, $3/4$ and 1. The results of this experiments are summarized in Table 2. For this experiment we obtain six metrics of the distribution of the FwLTR on the actors.

Table 2. Experimental results when considering uniform thresholds for $\theta \in \{1/4, 1/2, 3/4, 1.00\}$.

	θ	σ	#	**Gini**	**Top10**	**Top10%Actors**		**Top10%Values**	
Amazon	0.25	0.001358	1146	0.697121	0.00658478	0.754688	(33486)	0.628117	(518)
	0.50	0.000037	224	0.411186	0.000979505	0.301711	=	0.00142745	(25)
	0.75	0.000031	203	0.372209	0.00078241	0.210316	=	0.000958601	(24)
	1.00	0.000017	153	0.326807	2.9863e-05	0.0999991	=	5.0767e-05	(17)
ArXiv	0.25	0.149279	435	0.854073	0.648417	0.670355	(524)	0.656429	(104)
	0.50	0.005950	169	0.540466	0.00267074	0.263068	=	0.00839374	(20)
	0.75	0.004339	137	0.519565	0.00209844	0.167684	=	0.0034338	(14)
	1.00	0.003578	118	0.513171	0.00190767	0.0999618	=	0.0022892	(12)
Caida	0.25	0.234984	2033	0.827517	0.890954	0.979339	(2647)	0.979339	(2426)
	0.50	0.014930	1234	0.704623	0.114145	0.555694	=	0.362606	(137)
	0.75	0.012896	1174	0.693649	0.0957129	0.47169	=	0.266931	(130)
	1.00	0.001261	158	0.503259	0.000377715	0.0999811	=	0.000642115	(17)
ENRON	0.25	0.268885	2786	0.833348	0.110406	0.845307	(3669)	0.841655	(426)
	0.50	0.034267	2322	0.752884	0.0292707	0.588139	=	0.442222	(236)
	0.75	0.013211	1852	0.715277	0.0168429	0.39202	=	0.148643	(190)
	1.00	0.009538	1448	0.726360	0.000272539	0.0999945	=	0.00411534	(151)
Epinions	0.25	0.034836	1395	0.961041	0.0462051	0.612976	(7587)	0.609418	(375)
	0.50	0.001259	828	0.830114	0.014879	0.484587	=	0.0422778	(85)
	0.75	0.001185	814	0.828013	0.0144045	0.370326	=	0.0377838	(84)
	1.00	0.000343	327	0.707198	0.000131789	0.0999881	=	0.000487618	(37)
Gnutella	0.25	0.299963	593	0.892894	0.974148	0.979692	(6258)	0.979788	(6716)
	0.50	0.000173	111	0.689648	0.00615154	0.97402	=	0.00672674	(13)
	0.75	0.000172	111	0.689222	0.00594382	0.956076	=	0.00640718	(12)
	1.00	0.000070	52	0.545161	0.00015978	0.0999904	=	0.000111846	(7)
Higgs	0.25	0.000010	77	0.279112	0.000557524	0.147467	(25649)	0.000557524	(10)
	0.50	0.000006	52	0.218244	0.00035089	0.13094	=	0.000323598	(8)
	0.75	0.000006	53	0.214459	0.000362586	0.125685	=	0.000304104	(6)
	1.00	0.000005	46	0.192700	3.89877e-05	0.0999996	=	2.33926e-05	(6)
Wikipedia	0.25	0.026782	309	0.875415	0.32298	0.366268	(711)	0.324666	(37)
	0.50	0.006443	234	0.766315	0.00196767	0.359663	=	0.00477864	(25)
	0.75	0.006125	238	0.763096	0.00196767	0.288686	=	0.00463809	(26)
	1.00	0.005943	238	0.761196	0.00140548	0.0999297	=	0.00337316	(24)

Observing the sharp differences obtained in the results of this first experiment, we decided to do more experiments for the Gnutella and ArXiv networks with fine-grained threshold values between 0.2 and 0.5, and for Amazon, Caida, ENRON, Epinions and Wikipedia between 0.2 and 0.4. On the contrast, the remaining networks showed a considerable degree of monotony in metrics despite threshold variation. Table 3 shows the refined results for the Caida and Gnutella networks.[2] Here we only expose and comment on the most relevant results.

The extreme variation of Top10%Values due to widespread ties in the allocation of the FwLTR questions its usefulness as a metric for maximizing influence, since its behavior is not justified by the variation in the other measures.

Note that there is an obvious correlation between Top10%Actors and the Gini, since the more unequal a distribution is, the greater the proportion of the value accumulated by a small group of actors will be. However, this does not necessarily occur in Top10, where the set of actors has a fixed size of 10, regardless of the size of the network considered. The value of the Gini decreases when the influence threshold is increased, but the behavior of Top10 seems independent to this trend (see, e.g., Wikipedia or Caida; see Table 3).

We see a curious rise and fall of the cardinality of the set of Top10%Values in the Gnutella network between the thresholds 0.26 and 0.34 (see Table 3), which is not reflected in the variation of the other metrics for the same range of thresholds. There is also a sudden drop after 0.48, reflected in all metrics except Top10%Actors, which suffers only a small decrease. This suggests that, despite been the distribution of the actors' FwLTR radically affected and its influence severely limited, 10% of the most influential actors is a robust and large enough set for expanding its influence to most of the Gnutella network at those threshold magnitudes. This behavior is not observed in ArXiv, where Top10%Actors follow an online behavior with the other metrics. Despite Gnutella's dimensions being larger than ArXiv's, the most significant differences between these two networks, in terms of the parameters that we have obtained from them, is that Gnutella is directed and ArXiv is not, and ArXiv has greater diameter and ACC (around 100 times greater), and a much smaller main core. These data suggest that ArXiv is a much more interconnected and distributed network than Gnutella, whose parameters indicate that most nodes have few interconnections, but there are a subset of nodes with very high degrees, which would be very capable of diffusing the influence effectively, and therefore would be in the Top10%Actors. Gnutella's Gini is especially high compared to these networks, a fact that could validate this explanation.

The findings thus far indicate that within the majority of the analyzed networks, resistance thresholds falling within the range of approximately $[0.2, 0.5]$ emerge as particularly significant. Within this range, small actor groups are able to induce a significant proportion of the total network influence.

5.2 Random Thresholds

In this experiment, we set random thresholds to the actors according to a random uniform distribution in $[0, 1]$, and later on two different intervals.

[2] Results for the remaining ones can be found in the Appendix.

Table 3. Experimental results when considering uniform thresholds θ: refined zoom in [0.2, 0.4] and [0.2, 0.5].

	θ	σ	#	Gini	Top10	Top10%Actors		Top10%Values	
Caida	0.20	0.365273	1455	0.747064	0.907347	0.995392	(2647)	0.995392	(6562)
	0.22	0.338180	1615	0.775445	0.906327	0.995392	=	0.995392	(5223)
	0.24	0.322551	1740	0.790862	0.904363	0.98916	=	0.994334	(4594)
	0.26	0.218601	2077	0.825349	0.871124	0.978357	=	0.978357	(2225)
	0.28	0.197560	2173	0.822601	0.860057	0.933975	=	0.930727	(1818)
	0.30	0.142933	2291	0.792907	0.787762	0.92729	=	0.918829	(1399)
	0.32	0.130483	2355	0.783452	0.765703	0.913957	=	0.899452	(1302)
	0.34	0.062324	2366	0.720697	0.187875	0.87101	=	0.795958	(408)
	0.36	0.057867	2395	0.714390	0.18085	0.859754	=	0.791917	(386)
	0.38	0.049777	2424	0.704256	0.184816	0.849745	=	0.769481	(334)
	0.40	0.044526	2402	0.697489	0.171709	0.838414	=	0.745647	(277)
Gnutella	0.20, 0.22	0.373341	280	0.819895	0.974547	0.988176	(6258)	0.987873	(212)
	0.24	0.373303	277	0.819943	0.974547	0.988176	=	0.987361	(186)
	0.26	0.299963	593	0.892894	0.974148	0.979692	=	0.979788	(6716)
	0.28	0.299944	594	0.892910	0.974148	0.979692	=	0.979788	(6715)
	0.30, 0.32	0.297771	615	0.894629	0.974148	0.979596	=	0.979596	(6606)
	0.34, 0.38	0.090475	581	0.986210	0.971863	0.981753	=	0.971863	(610)
	0.40, 0.42	0.080219	666	0.986708	0.971783	0.981657	=	0.971783	(495)
	0.44	0.080126	663	0.986709	0.971783	0.981657	=	0.971783	(497)
	0.46, 0.48	0.080126	662	0.986709	0.971783	0.981657	=	0.971783	(495)
	0.50	0.000173	111	0.689648	0.00615154	0.97402	=	0.00672674	(13)

Table 4. Experimental results (for 100 executions) for random thresholds under a random uniform distribution.

Network	σ	#	Gini	Top10	Top10%Actors		Top10%Values	
Amazon	0.007479	2317	0.002218	1	1	(33486)	1	(1492)
ArXiv	0.316806	1478	0.208256	0.793209	0.793209	(524)	0.793209	(234)
Caida	0.007981	3248	0.002557	1	1	(2647)	1	(396)
ENRON	0.248287	13117	0.083801	0.918347	0.918347	(3669)	0.918347	(1390)
Epinions	0.269755	5816	0.259223	0.628316	0.637106	(7587)	0.628316	(636)
Gnutella	0.405467	2288	0.768062	0.972805	0.992394	(6258)	0.978238	(264)
Higgs	0.003500	9404	0.709183	0.00924399	0.139252	(25649)	0.0245662	(1472)
Wikipedia	0.142819	1048	0.277731	0.327056	0.383696	(711)	0.333942	(109)

Random Uniform Distribution - For this experiment we assigned thresholds in $[0, 1]$ following a uniform probability distribution. In order to determine centrality of the actors, we simulate the influence expansion several times, each with new randomly assigned values, and finally we get the average FwLTR of the actors for those executions. Table 4 shows the results for this experiment when performing 100 executions.

We can see that there are no large differences for any of the variables between the two experiments and that, for the vast majority of networks, a uniform random distribution is very conducive to maximizing influence, achieving very satisfactory results with only the top 10 actors of the ranking. However, we also see that we rarely achieve more expansion if we increase the set to the first 10% of actors. We also see, for the first time, that the Top10 sets have managed to expand the influence over all the actors in the Amazon and Caida networks.

Table 5. Experimental results (for 20 executions) for random thresholds under a random uniform distribution within $(0, 0.5]$.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.021248	519	0.008737	1	1 (33486)	1 (1192)
ArXiv	0.316359	646	0.213341	0.793209	0.793209 (524)	0.793209 (151)
Caida	0.031134	526	0.015298	1	1 (2647)	1 (117)
ENRON	0.249355	3265	0.089727	0.918347	0.918347 (3669)	0.918347 (380)
Epinions	0.266832	702	0.269977	0.629397	0.7452 (7587)	0.763597 (9994)
Gnutella	0.407975	424	0.768950	0.974579	0.991276 (6258)	0.991628 (12851)
Higgs	0.003177	4318	0.714451	0.00971964	0.139479 (25649)	0.0190182 (1027)
Wikipedia	0.139828	571	0.285285	0.327056	0.374982 (711)	0.334505 (77)

Table 6. Experimental results (for 100 executions) for random thresholds under a random uniform distribution within $[0.5, 1]$.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.163244	35500	0.712485	0.717867	0.719664 (33486)	0.717935 (9784)
ArXiv	0.228914	2007	0.240972	0.791683	0.791683 (524)	0.791683 (240)
Caida	0.246121	7448	0.609839	0.926308	0.926308 (2647)	0.926308 (818)
ENRON	0.208083	13455	0.152003	0.918075	0.918075 (3669)	0.918075 (1572)
Epinions	0.124726	11641	0.759954	0.537092	0.542825 (7587)	0.537092 (1365)
Gnutella	0.000176	1059	0.691910	0.00560828	0.973988 (6258)	0.856725 (145)
Higgs	0.000043	2344	0.528186	0.002234	0.141759 (25649)	0.00643687 (358)
Wikipedia	0.091907	1960	0.678521	0.319606	0.364301 (711)	0.331553 (214)

We observe that the number of different values produced in the ranking is much higher in all cases than those of the FwLTR with $\theta = 0.5$ (see Table 2) and they are more dispersed also, as evidenced by the standard deviation σ. The Gini also indicates that, in general, the ranking values are significantly more equally distributed when $\theta = 0.5$. We see a significant reduction in Gini in almost all networks (e.g., in the case of ENRON, it drops approximately from 0.75 to 0.08).

The exceptions are Gnutella and Higgs, where there is an ascent with respect to the majority criterion.

Random Uniform Distribution in Two Different Intervals - As a result of the previous experiment, we wondered what would happen if we assigned the thresholds uniformly at random but separately for domains (0, 0.5] and [0.5, 1]. By selecting these two intervals we can see what happens when the thresholds are in the more permissive or restrictive half for the transmission of influence. In Table 5 and Table 6 we can check the results.

The values of the variable Top10 in Table 6 suggest that for most of networks it is possible to achieve a good expansion of influence even when the thresholds of resistance are distributed in the most restrictive half of the interval. In some cases, we obtain practically the same expansion as when assigning the thresholds in $(0, 0.5]$, as is the case for the networks ArXiv, Wikipedia or ENRON. As expected, a smaller value is reached in general. The values of # are notably different between the two experiments, being the values of the second significantly larger than those of the first in almost all cases. This may indicate that, when the thresholds tend to be more restrictive, there is a more diverse ranking. Despite this, there does not appear to be any trend in the value of the Gini between the two experiments, and the standard deviation increases for some networks and decreases for others.

If we compare Table 5 with Table 4 in the previous section, we observe little difference in the variables that measure the expansion of influence. In fact, the only relevant improvement occurs on the Epinions network. We also see a general reduction in the number of rankings assigned to the actors (#), but very similar standard deviation and Gini values. These results indicate that a random distribution of the thresholds behaves practically the same with respect to the diffusion of influence whether they are in $(0, 0.5]$ or $(0, 1]$, suggesting that perhaps the fact that a minority fraction of the actors have high thresholds affects little to the ability to influence of the actors in general.

5.3 Thresholds Fixed by Other Centrality Measures

In this experiment we will use different centrality criteria to set the thresholds of the actors of a network. We will denote FwLTR_m to the FwLTR where the threshold θ is set to the value given by the centrality measure m; for example $\mathsf{FwLTR}_{\mathsf{ICR}}$ is the FwLTR ranking of the actors in a network when we set the resistance threshold of each actor to its ICR centrality value. We introduce these metrics to study the possibility that an actor's resistance to being influenced in a network is related to its centrality value, perhaps by a topological criterion or by own ability to spread influence, and in turn, we study the case that this related with the lack of these qualities.[3]

[3] How the thresholds are assigned according to the different centrality measures in each network can be seen in the Appendix.

Table 7. Experimental results for node thresholds assigned according to Btwn.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.144358	185	0.021906	0.986335	0.996363 (33486)	0.987983 (572)
ArXiv	0.321546	84	0.207490	0.793209	0.793209 (524)	0.793209 (11)
Caida	0.097087	10	0.009517	1	1 (2647)	1 (26228)
ENRON	0.269095	431	0.104976	0.915458	0.918347 (3669)	0.915458 (1392)
Epinions	0.277439	39	0.265431	0.628725	0.757126 (7587)	0.629173 (28)
Gnutella	0.410393	73	0.767687	0.974978	0.996421 (6258)	0.974579 (9)
Higgs	0.003131	198	0.817138	0.00899057	0.148863 (25649)	0.0103551 (109)
Wikipedia	0.145380	14	0.275292	0.32818	0.431061 (711)	0.339002 (55)

Table 8. Experimental results for node thresholds assigned according to their ICR value.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.048237	159	0.002346	0.997814	1 (33486)	0.998719 (67)
ArXiv	0.016946	389	0.626108	0.0721099	0.603396 (524)	0.183518 (43)
Caida	0.058035	3025	0.664515	0.355354	0.95966 (2647)	0.913692 (364)
ENRON	0.016987	2322	0.691726	0.0317235	0.526927 (3669)	0.251199 (237)
Epinions	0.003255	1614	0.862370	0.0452035	0.599573 (7587)	0.152493 (169)
Gnutella	0.387287	267	0.801691	0.974643	0.990701 (6258)	0.983383 (106)
Higgs	0.000028	144	0.448542	0.00132948	0.15022 (25649)	0.0015907 (18)
Wikipedia	0.028498	781	0.744640	0.113563	0.355727 (711)	0.195643 (92)

Table 9. Experimental results for node thresholds assigned according to PgR.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.000000	1	0.000000	1	1 (33486)	1 (334863)
ArXiv	0.321327	84	0.207110	0.793209	0.793209 (524)	0.793209 (11)
Caida	0.066887	7	0.004494	1	1 (2647)	1 (26356)
ENRON	0.256195	360	0.085229	0.918347	0.918347 (3669)	0.918347 (37)
Epinions	0.275013	39	0.258302	0.628725	0.756837 (7587)	0.629173 (28)
Gnutella	0.410393	73	0.767687	0.974978	0.996421 (6258)	0.974579 (9)
Higgs	0.003319	189	0.785266	0.00894768	0.151744 (25649)	0.0101797 (83)
Wikipedia	0.145312	14	0.274871	0.32832	0.431061 (711)	0.339002 (55)

Table 10. Experimental results when the thresholds of the nodes are assigned according to their value of FwLTR.

Network	σ	#	Gini	Top10	Top10%Actors	Top10%Values
Amazon	0.004888	4	0.000024	1	1 (33486)	1 (334855)
ArXiv	0.322199	84	0.208631	0.793209	0.793209 (524)	0.793209 (11)
Caida	0.151072	20	0.023374	1	1 (2647)	1 (25895)
ENRON	0.296371	385	0.118261	0.918347	0.918347 (3669)	0.918347 (39)
Epinions	0.277610	39	0.265945	0.628725	0.757087 (7587)	0.629173 (28)
Gnutella	0.410393	73	0.767687	0.974978	0.996421 (6258)	0.974579 (9)
Higgs	0.003712	183	0.706886	0.00899837	0.157916 (25649)	0.0103473 (101)
Wikipedia	0.146984	14	0.285386	0.32832	0.430921 (711)	0.338721 (54)

$\mathsf{FwLTR}_{\mathsf{Btwn}}$ – When fixing the threshold according to Btwn (see Table 7), we can observe similar results to those of the experiment detailed in Table 4 for the thresholds distributed in a uniform random way. The standard deviation and the Gini are also very similar. Where the results of these two experiments really differ is in the number of distinct values in the ranking, which are significantly smaller now. These results suggest that we can expect achieve a very good expansion with very small initial sets, but note that we gain little or nothing by increasing the pool of the top 10 to the first 10%.

$\mathsf{FwLTR}_{\mathsf{ICR}}$ – When fixing the threshold according to ICR (see Table 8) the values of Top10, Top10%Actors and Top10%Values are lower than when fixed according to Btwn (compare with Table 7) for all networks except Amazon and Higgs, where we see a slight increase. The values of σ are also lower, indicating higher concentration, and gives more variety of values, which are also more unequally distributed, as indicated the Gini, except in the Higgs network, which undergoes the opposite behavior.

$\mathsf{FwLTR}_{\mathsf{PgR}}$ and $\mathsf{FwLTR}_{\mathsf{FwLTR}}$ – When fixing the threshold according to PgR and FwLTR (see Tables 9 and 10, respectively) we see extremely similar values in the variables related to the expansion of influence. We expected similarities, and we observe that they give very comparable threshold assignments. However, it is surprising to what extent the diffusion of the influence on the Top10 variables between both methods. The values obtained for the metrics σ, # and Gini are not disparate, which show similar results under both centralities. From these tables we deduce that assigning the thresholds according to the PgR and FwLTR centralities of the actors generates favorable conditions for a great influence expansion from reduced sets. In nearly all cases, it is enough to have the first 10 actors of the ranking to reach the entire network or its vast majority, except for two networks (Higgs and Wikipedia) whose individual characteristics seem to imply difficulties for the diffusion of influence, as by this and previous experiments suggest.

5.4 Complementaries

We have also performed an additional experiment that complements the previous one. We will denote by FwLTR_{1-m} the case in which the thresholds are set by the complement of the value obtained from the centrality m, that is, if an actor obtains a centrality value c in the measure m, we will assign a threshold equivalent to $1 - c$. In this way we can observe the case in which a greater centrality corresponds to a greater resistance to influence, and the opposite.

The $\mathsf{FwLTR}_{1-\mathsf{Btwn}}$ (see Appendix) behaves in the opposite way in the Top10 variables, where we barely managed to expand the influence. The values of the Gini they have also been reversed: the high values in Table 7 are here low, and vice versa. The standard deviation σ is very low in all cases, indicating that ranking values are produced very clustered around the mean.

In the $\mathsf{FwLTR}_{1-\mathsf{ICR}}$ (see Appendix) we see that we hardly managed to diffuse the influence in none of the Top10 variables. We also have low values of the standard deviation, but generally similar to those of the $\mathsf{FwLTR}_{\mathsf{ICR}}$. The Gini values are less high, indicating a more equitable distribution of ranks. In the same way, we see practically the same values for $\mathsf{FwLTR}_{1-\mathsf{PgR}}$ and $\mathsf{FwLTR}_{1-\mathsf{FwLTR}}$ (see Appendix) in all metrics. The expansion of influence is predictably restrictive unlike the cases described in the previous paragraph. The values of σ are very low and those of the Gini higher in most cases, compared to the previous ones obtained in the non-complementary version. The exceptions are in Higgs and Gnutella, where they previously had high values but are now diminished.

6 Conclusions

We have considered the spread of influence in the LT model and the absence of potential values for the threshold defining the individual resistance that is the basis for the model. We have selected to study the behaviour of the FwLTR rank in a popular selection of social networks, and studied different ways to fix the individual threshold of the actors. Our results show that the selection has a clear impact on the ranking. Surprisingly enough, we have found a abrupt variation of the ranking when setting the thresholds uniformly, being thresholds in the interval $[0.2, 0.5]$ the best selection. For the case of random thresholds, a similar phenomenon appears as there are almost no differences on the ranking when the individual thresholds are selected in $(0, 0.5]$ than in $(0, 1]$. Finally, we have been able to asses that the social networks ranks that provide the best assignments for the individual thresholds are PgR and FwLTR.

There are many ways to complement this study, for example by including additional social networks in the analysis which might provide insight on the best models to study influence spread in the LT model. Another extension will be to study how the threshold selection impacts on other influence spread problems like the influence maximization or the target set selection problems. We would also like to compare FwLTR to recently-proposed centrality measures, e.g. [1,12].

References

1. Arhachoui, N., Bautista, E., Danisch, M., Giovanidis, A.: A fast algorithm for ranking users by their influence in online social platforms. In: Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 526–533. Association for Computing Machinery (2022). https://doi.org/10.1109/ASONAM55673.2022.10068673
2. Blesa, M., García-Rodríguez, P., Serna, M.: Forward and backward linear threshold ranks. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 265–269. Association for Computing Machinery (2021). https://doi.org/10.1145/3487351.3488355
3. Ceriani, L., Verme, P.: The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. The Journal of Economic Inequality **10**(3), 421–443 (2012)
4. Chen, N.: On the approximability of influence in social networks. SIAM Journal on Discrete Mathematics **23**(3), 1400–1415 (2009). https://doi.org/10.1137/08073617X

5. Freeman, L.: A set of measures of centrality based on betweenness. Sociometry **40**(1), 35–41 (1977)
6. Gini, C.: Variabilitàe Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. C. Cuppini (1912)
7. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. Academy of Marketing Science Review **9**(1) (2001)
8. Hasson, S.T., Akeel, E.: Influence maximization problem approach to model social networks. In: 2019 International Conference on Advanced Science and Engineering (ICOASE). pp. 135–140 (2019). https://doi.org/10.1109/ICOASE.2019.8723703
9. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 137–146 (2003)
10. Kempe, D., Kleinberg, J., Tardos, E.: Influential nodes in a diffusion model for social networks. In: Intl. Colloquium on Automata, Languages and Programming (ICALP). vol. 3580, pp. 1127–1138. Lecture Notes in Computer Science (2005). https://doi.org/10.1007/11523468_91
11. Lorenz, M.O.: Methods of measuring the concentration of wealth. Publications of the American Statistical Association **9**(70), 209–219 (1905). https://doi.org/10.2307/2276207
12. Namtirtha, A., Dutta, A., Dutta, B., Sundararajan, A., Simmhan, Y.: Best influential spreaders identification using network global structural properties. Nature - Scientific Reports **11**(2254) (2021). https://doi.org/10.1038/s41598-021-81614-9
13. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. In: The Web Conference. Stanford Digital Library (1999)
14. Riquelme, F., Gonzalez-Cantergiani, P., Molinero, X., Serna, M.: Centrality measures in social networks based on linear threshold model. Knowledge-Based Systems **40**, 92–102 (2017). https://doi.org/10.1016/j.knosys.2017.10.029
15. Shakarian, P., Bhatnagar, A., Aleali, A., Shaabani, E., Guo, R.: The Independent Cascade and Linear Threshold Models, chap. 4, pp. 35–48. Briefs in Computer Science, Springer (2015). https://doi.org/10.1007/978-3-319-23105-1_4
16. Wan, Z., Mahajan, Y., Kang, B.W., Moore, T.J., Cho, J.H.: A survey on centrality metrics and their network resilience analysis. IEEE Access **9**, 104773–104819 (2021). https://doi.org/10.1109/ACCESS.2021.3094196
17. Zareie, A., Sakellariou, R.: Influence maximization in social networks: a survey of behaviour-aware methods. Social Network Analysis and Mining **13**(78) (2023). https://doi.org/10.1007/s13278-023-01078-9

Appendix

Additional data supporting this work can be found at
https://cs.upc.edu/~mjblesa/ASONAM.2024