

# Methods for Selecting Nodes for Maximal Spread of Influence in Recommendation Services

Bogdan Gliwa and Anna Zygmunt<sup>(✉)</sup>

AGH University of Science and Technology, Al. Mickiewicza 30,  
30-059 Kraków, Poland  
{bgliwa,azygmunt}@agh.edu.pl

**Abstract.** Social network analysis is a tool to assess social interactions between people e.g. in the Internet. One of the most active areas in this field are modeling influence of users and finding influential users. These areas have many applications, e.g., in marketing, business or politics. Several models of influence have been described in literature, but there is no single model that best describes the process of spreading entities (e.g. information, behaviour) through the network. Interesting and practical problem is how to choose a small number of users that will guarantee maximal spread of entities over the whole network (*influence maximization problem*). In this paper we studied this problem using various centrality metrics with different models of influence propagation. Experiments were conducted on three, real-world datasets regarding the domain of recommendation services.

**Keywords:** Influence · Influence maximization problem · Influence models · Centrality measures · Social networks · Social media

## 1 Introduction

Social network analysis (SNA) is an active and attractive area of research in computer science. It describes social relationships of users in the form of a network. Such a network, however, is not homogeneous and we can distinguish some parts (called groups) that are more dense than others. One of the directions of research is analysis of groups, their structure and dynamics. Another important aspect of analysis is the position of users in a network which may be described using centrality metrics. To the most interesting and promising directions of research we can also include finding influential users and analysis of influence that social network's users have on each other. Such studies are widely used mainly in business, marketing, but also in epidemiology and politics. In today's viral marketing, potential customers will pass each other information about the company, services or products. In epidemiology, it is important to detect sources of an epidemic and identifying trends.

Recommendation services are more and more popular among the Internet's users. People use them to express their opinions about products or services.

Such portals help them to make a decision about buying a product or using a service. Consumers often take into consideration the recommendations and opinions of trusted people so the use of algorithms that determine influentials can help, among others, in conducting a successful advertising campaign, reducing the evolution of the disease or predicting trends of popularity of a political party, a piece of music or a film. There are many interesting questions: why one user is more influential than other, what conditions must arise in the neighborhood that the user will become influential, if one is influential, it means that it has good network position or maybe some individual characteristics are important? Several models of influence have been proposed (e.g., IC, LT, LTC). In the context of these works, an interesting direction seems to be how to choose a small number of users (within predefined budget for given number of people) that will guarantee maximal spread over the whole network, which is called *influence maximization problem*.

The majority of work regarding the topic of influence in social networks has been dedicated to studying three different aspects of influence analysis: (1) influence modeling, (2) finding influential users, (3) influence maximization problem. In [16] we compared various influence models and proposed our new algorithm which takes into account users behaviours and their expertise.

In this paper we study the problem of influence maximization. We present an approach based on integrating various influence models (IC, ICN, LT, LTC, LTB) with different centrality measures (degrees, closeness, PageRank, eigenvector, betweenness). We study this in three different recommendation services: *Last.fm* (music recommendations), *Flixster* (movies recommendations) and *Yelp* (focused on reviews on local businesses, e.g., restaurants).

The remainder of this paper is organized as follows: Sect. 2 presents an overview of research in the area of the influence maximization problem. Section 3 gives a description of concept and methods. Section 4 presents experimental evaluation, and Sect. 5 provides some final conclusions and presents future directions.

## 2 Related Work

Presenting interactions between users of the given social media site in the form of social network allows discovering important users as well as observing how much influenced they are. Understanding and modeling the spread of influence is an important topic in social network analysis and therefore attracts many researchers to this area.

Spread of an entity (e.g., information or products) in the social network can be seen as cascading actions taken by adopting users (e.g., through commenting, evaluation or purchase the entity). Typically, a group of the initiators adopt the entity, then *infect* their neighbors (e.g., by sending messages, recommendations), which in turn may also decide on the entity adoption continuing its spread. Analysis of this process relates to a study of the structure of social networks, as well as the ability and the susceptibility of the individual users during the adoption of the entity [6].

Influence modeling is often seen as a variation of an epidemic spread problem. Many models have been proposed, namely Susceptible-Infected-Recovered (SIR [11]) model and its variations (e.g., SIS, SIRS, SEIS). These approaches assume an implicit network and do not include the chronological sequence of infections. Therefore, they are only able to predict general trends such as the percentage of population being infected or cured.

Kempe et al. in [10] investigated the *Linear Threshold Model (LT)* and *Independent Cascade (IC)* to simulate exact propagation of information in social networks. The *IC* model may be considered as a push model, because active nodes independently try to propagate information to their neighborhoods. In contrast, the *LT* model may be described as a pull model, where inactive nodes become activated when the majority of their neighbors (with respect to a certain threshold) have already accepted the information. Both models have remained a strong foundation for successive studies. The *Linear Threshold Colors (LT-C)* model was introduced to capture product adoption rather than influence [1]. Each node after activation can end up in one of three states/colors: *Adopt*, *Promote* or *Inhibit*. The two latter states correspond to the users who report positively (*Promote*) or discourage the product without actually adopting it (*Inhibit*). In [21] *Linear Threshold Value (LT-V)* assumes a distinction between users who cannot adopt certain product due to, e.g., a high price, and those who can. In [2], Chen et al. defined the *Independent Cascade Model with Negative Opinions (IC-N)* that allows for modeling influence of both positive and negative opinions. New model, taking into account (apart from user's connections) both users behaviours and their expertise was proposed in [16].

In our previous experiment described in [16] we showed that the accuracy of particular model highly depends on the kind of the entity that is being spread through the network and parameters that are associated with it (e.g., price, knowledge, accessibility).

The field of research dedicated to influence maximization problem was formalized by Kempe et al. in [10]. The idea is to find minimal set of nodes (within given budget), that will eventually result in maximal spread of information over the whole network. The problem is NP-hard under IC (Independent Cascade) and LT (Linear Threshold) model.

There are several approaches to solve influence maximization problem. In the simplest one, the centrality measure is chosen (e.g., degree centrality [20], closeness centrality [13], betweenness centrality [5], eigenvector centrality [18], PageRank [17]) together with selected model of propagation. The measure is then calculated for each node and those nodes with the highest value are selected. Centrality measures are based only on the structure of graph and nodes position. However, several efficient algorithms for effective implementation of the calculations [12] have been proposed.

Kempe in [10] proposed to use greedy algorithm and Monte Carlo simulation. Initial nodes are chosen in such way that in subsequent time steps they activate the greatest number of their neighbors. Process of neighbors activation is random, so selecting the appropriate node in each step is nondeterministic.

Therefore, the node activation process is repeated in several hundred iterations of the Monte Carlo simulation, checking thus the number of neighbors that a node is able to activate in each iteration. The node with the highest average number of active neighbors in all iterations is chosen [18].

Goyal in [7] presented an approach that completely omitted models of propagation. To determine the initial set of nodes, the method used only traces of the previous propagating information in the network. Method did not model neither the network nor actions taken by users. Narayanam proposed in [15] approach based on game theory. Presented algorithm *SPIN* (Sharply Value-Based Influential Nodes) turned out to be more time efficient than greedy algorithms.

Chen et al. in [3] presented new algorithms for influence maximization that employed community structure to improve nodes' selection. There are also some approaches to influence maximization in dynamic social networks (when the network changes over time) [22].

### 3 Concept and Methods

In this section we describe basic definitions used in the task of influence maximization problem and later, we provide descriptions of influence propagation models used in the experiments.

#### 3.1 Basic Definitions

To describe different models of influence propagation, we introduced the following definitions:

- a *social network* is a directed weighted graph denoted  $G = (V, E, W)$  where  $V$  is a set of nodes representing users of the network,  $E$  is a set of directed edges representing social relationship between the users, and  $W$  is a set of weights for the edges. Notation  $V_{in}(v)$  stands for all in-neighbors of the node  $v$  and  $V_{out}(v)$  denotes all out-neighbors of the node  $v$ . The weight  $w_e \in W$  of the directed edge  $(u, v)$  between the nodes  $u$  and  $v$  is an indicator of the influence strength and its weight implies probability of  $u$  influencing  $v$ . The edge weight is calculated as stated in Eq. 1

$$w_{u,v} = \frac{r_{v,u}}{\sum_{n \in V_{in}(v)} r_{v,n}}, w_{u,v} \in [0; 1], \quad (1)$$

where  $r_{a,b}$  means the number of entities that user  $a$  rated after user  $b$ . This formulation implies the following property:

$$\sum_{n \in V_{in}(u)} w_{n,u} = 1. \quad (2)$$

- an *entity* represents a thing that is being spread through the network, e.g., a movie, an artist, or a product. Each user may evaluate the entity with appropriate numeric rating/evaluation.

- a *node activation* is an action of accepting the entity by a given node. Such an action depends on a particular social service and may correspond to: liking, tagging, sharing, evaluating or commenting the entity. Time variable must be associated with each activation to let us order the sequence of nodes' activations.  $A_e$  denotes a set of active nodes during propagation of the entity  $e$ .
- a *seed set* is an initial set of active nodes that begins the propagation in the network. In the task of influence maximization a seed set is chosen by the order of nodes' ranks that can be determined using different metrics.
- a *coverage* is a final set of activated nodes at the end of simulation (when the propagation of entities is finished). Intuitively, it is a spread or a range of the entity in the network.

Influence maximization task involves the selection of a method of ranking nodes to achieve the maximum coverage using given size of seed set. In experiments we repeated each simulation for every seed size in a given model 100 times and all figures present mean values of a coverage.

### 3.2 IC Model

The *Independent Cascade Model* (denoted *IC*) is a one of the basic models used in the influence modeling which was studied by Kempe in [10]. Each edge has assigned activation probability corresponding to a weight  $w$  described in Sect. 3.1. Each node can be either in one of two states – *Active* or *Inactive*. The process of nodes activation starts with a seed  $A_e^0$  and during the process continues to activate further nodes as follows: when a node  $u$  becomes activated in step  $t$ , it gets only single chance to activate each of its inactive neighbors, denoted  $n$  with a probability  $p = w_{u,n}$ . Once the node is activated, it cannot be deactivated. The process runs until no more activations are possible.

### 3.3 ICN Model

The *Independent Cascade Model with Negative Opinions* (denoted *ICN*) [2] is an extension of *IC* model that allows for modeling of both positive and negative opinions. Each node has one of three states – *neutral*, *positive* or *negative*. The node is treated as activated in a timestep  $t$  if it is positive or negative in a timestep  $t$  and *neutral* in the previous timestep. *ICN* model has a parameter  $q$  (*quality factor*) describing the probability of staying *positive* when was activated by *positive* neighbour. Initially, all nodes from seed set are either *positive* (with probability  $q$ ) or *negative* (with probability  $1 - q$ ). In each timestep, they try to activate their inactive neighbours. If *negative* node infects one of its neighbour, then such neighbour becomes also *negative*, but in the case when infecting node is *positive* then the neighbour may become either *positive* (with probability  $q$ ) or *negative* (with probability  $1 - q$ ). The process of activation stops when no more activations are possible.

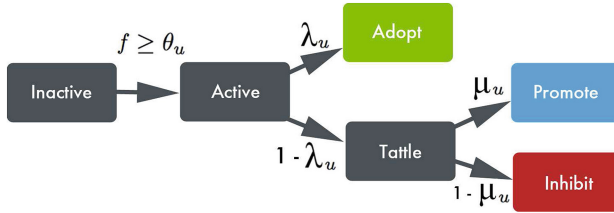
In experiments we set value of probability  $q$  to value 0.7.

### 3.4 LT Model

The *Linear Threshold Model* (denoted *LT*) [10] assumes directed, weighted graph  $G$ . Nodes can also be either *Active* or *Inactive*. Each node  $v \in V$  selects random, chosen with uniform distribution, threshold number  $\theta_v$ . At the beginning, only a seed set is active. With each timestep, active nodes ( $A_e$ ) try to activate their inactive neighbors. If the sum of weights of all active in-neighbor’s edges is greater than or equal to the node threshold, the node becomes activated. The active nodes cannot become inactive and, again, the process ends when no more activations are possible.

### 3.5 LTC Model

The *Linear Threshold with Colors Model* (denoted *LTC*) [1] introduced three additional states to differentiate the level of entities adoption (as described in Fig. 1).



**Fig. 1.** Node states in LTC model (Color figure online).

Except new states, the main difference between *LT* model and *LTC* model is the condition of activation. Nodes activate its neighbours with similar rules as specified in the *LT* model, but an activation condition differs significantly. Activation function  $f$  for a node  $u$  during propagation of an entity  $e$  is defined by Eq. 3, where  $s_{min}$  and  $s_{max}$  are the lowest and the highest ranking scores respectively, and  $s_{n,e}$  is the rating score that the neighbor user  $n$  assigned to the entity  $e$ .

$$f(u, e) = \frac{\sum_{n \in V_{in}(u), n \in A_e} w_{n,u} (s_{n,e} - s_{min})}{s_{max} - s_{min}}, f \in [0; 1]. \tag{3}$$

Once the node changes its state to *Active*, it again picks a random value with uniform distribution from  $[0; 1]$ . If the value is grater than  $\lambda$ , the node fully adopts the entity (green). If not, it goes to *Tattle* state which indicates users that cannot afford the entity or do not want to buy the entity, but are still expressing opinions about it. In the next random process, the node has a  $\mu$  probability to eventually get into *Promote* (blue) state and  $1 - \mu$  probability to completely inhibit the entity and express negative opinion (red).

### 3.6 LTB Model

*Linear Threshold Behavioral Model* (denoted *LTB*) [16] is based on *LT* model, but enhances the activation condition from *LTC* model. It incorporates into the influence propagation model additional parts connected with personality (*Activity* parameter) and expertise (*Expertise* parameter) to fulfill the criteria of influence defined by two sociologists, Katz and Lazarsfeld in [9] (also confirmed in the context of online social networks by [4]).

The *Expertise* property is a node score that tries to describe the competence of the node in a given field of interest. General expertise  $\gamma_u$  of the node  $u$  in a social network is measured as a fraction of the number of entities rated by the user over the number of all entities (as described in Eq. 4):

$$\gamma_u = \frac{r_u}{\sum_{v \in V} r_v}, \gamma_u \in [0; 1], \quad (4)$$

where  $r_u$  denotes the number of all ratings issued by a node  $u$ . We assume that the more entities the user rates, the more knowledgeable they are.

The *Activity* property ( $\alpha$  parameter) is a node score expressing the user engagement in social media and adopting abilities based on their personality (categories of adopters proposed by Ryan and Gross in [19] linked with their eagerness to accept information, namely: innovators, early adopters, early majority, late majority, and laggards). It has the following formula – Eq. 5:

$$\alpha_u = \frac{1}{|R_u|} \sum_{e \in R_u} \frac{i_{e,u}}{|I_e|}, \alpha_u \in [0; 1], \quad (5)$$

where:  $R_u$  is a set of entities rated by  $u$ ;  $I_e$  is a sequence of users who rated entity  $e$ , ordered ascendingly by rating date;  $i_{e,u}$  is an index of  $u$  in the sequence  $I_e$ . Top 2.5% of the user with a lower  $\alpha$  parameter may be called *Innovators*, the next 13.5% - *Early Adopters*, etc. In other words, the lower  $\alpha$  parameter is, the more eager the node is to accept the entity.

Activation condition (denoted as  $Influence(u, e)$ ) to compute total influence exerted on the node  $u \in N$  by its active neighbors while evaluating the entity  $e$  has the following form:

$$Influence(u, e) = \frac{\sum_{n \in V_{in}(u), n \in A_e} w_{n,u} (s_{n,e} - s_{min}) \gamma_n}{s_{max} - s_{min}}. \quad (6)$$

Similar to the *LTC* model,  $s_{n,e}$  is the rating score that user  $n$  assigned to the entity  $e$ ; and  $s_{min}$ ,  $s_{max}$  are the lowest and the highest ranking scores respectively. Such function guarantees, that the more experienced the user is, the more power they have to influence others. In addition, higher neighbors' ratings of a particular entity increase overall influence.

The nodes' states are the same as in the *LT* model, that is: *Inactive* and *Active*. Each node  $u \in V$  picks random threshold number  $\theta_u$  between 0 and  $\alpha_u$  - the more innovative the user is, the more likely they are to adopt the new entity. In each timestep, each of the inactive nodes computes total value of *Influence*

exerted on it by its neighbors. Once the sum exceed threshold  $\theta_u$ , the node  $u$  becomes *Active*. The propagation continues until no more inactive nodes are left.

## 4 Results

In this section we provide the description of datasets, then we present experiments on full dataset. Due to the expensive computational complexity of two centrality metrics – closeness and betweenness centrality, they were not calculated for *Flixster* and *Yelp* dataset. We decided to reduce graphs generated from these datasets and we demonstrate results for one chosen model of influence propagation (IC).

### 4.1 Description of Datasets

Experiments were conducted on three real world data sets from the domain of social services:

**Last.fm**<sup>1</sup> - an online music service with songs ratings. We utilized a dataset that was released for the Workshop on Information Heterogeneity and Fusion in Recommender System (HetRec 2011)<sup>2</sup>. The dataset consists of 1892 users with 25.4K relations among them. Listening history is represented in the form of the number of times a user listened to a song performed by an artist. The dataset also includes tagging history with date timestamps, but does not provide explicit user ratings, so it is calculated based on the frequency of playing songs of artists by users. Exact way of assigning scores for songs was described in [16]. In *Last.fm*, an entity symbolize an artist, and node activation indicates an action of tagging or listening to an artist for the first time.

**Flixster**<sup>3</sup> - a social network for movies aimed at sharing opinions, making reviews and rating movies. *Flixster* dataset was published by Jamali et al. [8] and contains 1 M users, 14 M undirected friendships relations between users, and 8.2M ratings. In this case, an entity is a movie, and node activation represents an action of rating a movie.

**Yelp**<sup>4</sup> - a social network focused on enabling users to publish opinions on companies such as restaurants, cafes etc. The dataset was published as a part of Yelp Challenge<sup>5</sup>. It consists of 1.6 M reviews published by 366 K users with 2.9M social connections for 61 K business companies. In this dataset, the following definitions are assumed: an entity describes a business company and node activation means an action reviewing a company by a user.

In all datasets, nodes that did not rate any entity or do not have any neighbours, were excluded.

<sup>1</sup> <http://www.last.fm>.

<sup>2</sup> <http://ir.ii.uam.es/hetrec2011>.

<sup>3</sup> <http://www.flixster.com>.

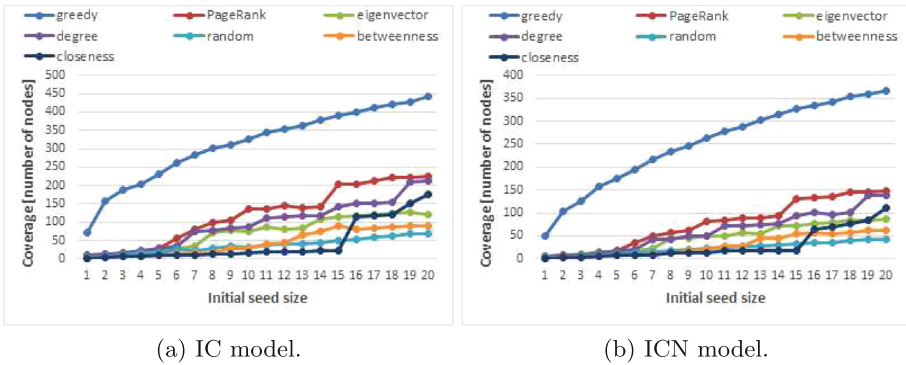
<sup>4</sup> <http://www.yelp.com>.

<sup>5</sup> [http://www.yelp.com/dataset\\_challenge](http://www.yelp.com/dataset_challenge).



## 4.2 Influence Maximization

**LastFm Dataset.** LastFm is the smallest dataset among tested ones. Figure 2a and b presents results (which are very similar to each other) for models IC and ICN, respectively. Greedy algorithm provides the best coverage. Good results are also obtained using PageRank and out-degree metrics. Eigenvector and betweenness gave worse coverage. Interesting behaviour can be observed for closeness centrality i.e. for some seed sizes the coverage is very small (even worse than random selection of nodes), but later, it has sudden increase. It may be explained by the fact that closeness metrics find nodes that have central position in network (nodes with the smallest total distances from all other nodes) and therefore, are close to each other (often may be neighbours), so adding more nodes with the highest values of this metrics does not improve the coverage until we add a node that has higher values of other metrics.



**Fig. 2.** LastFm - models based on IC.

Figure 3a shows results for LT model. The tendency is similar to the previous cases (but eigenvector centrality gives comparable results to out-degree). In LTC model (Fig. 3b) the difference between various centrality metrics is small (the tendency is also similar). In LTB model (Fig. 3c) the centrality metrics do not improve at all the total coverage, they may be even worse than random selection of nodes. It is also worth noticing that models LTC have 10 times less coverage than LT model, and LTB - 20 times.

Table 1a shows calculation times for different centrality measures. We provided such comparison only for LastFm, because only in this dataset we could calculate all tested centrality metrics. We can spot that out-degree, eigenvector and PageRank are very fast, but betweenness and closeness very slow. Greedy algorithm gives best results in every case, but is also very expensive computationally – in Table 1b we can see mean times for choosing a single node. The difference of times between different models is very small (the largest value is for LTC model).

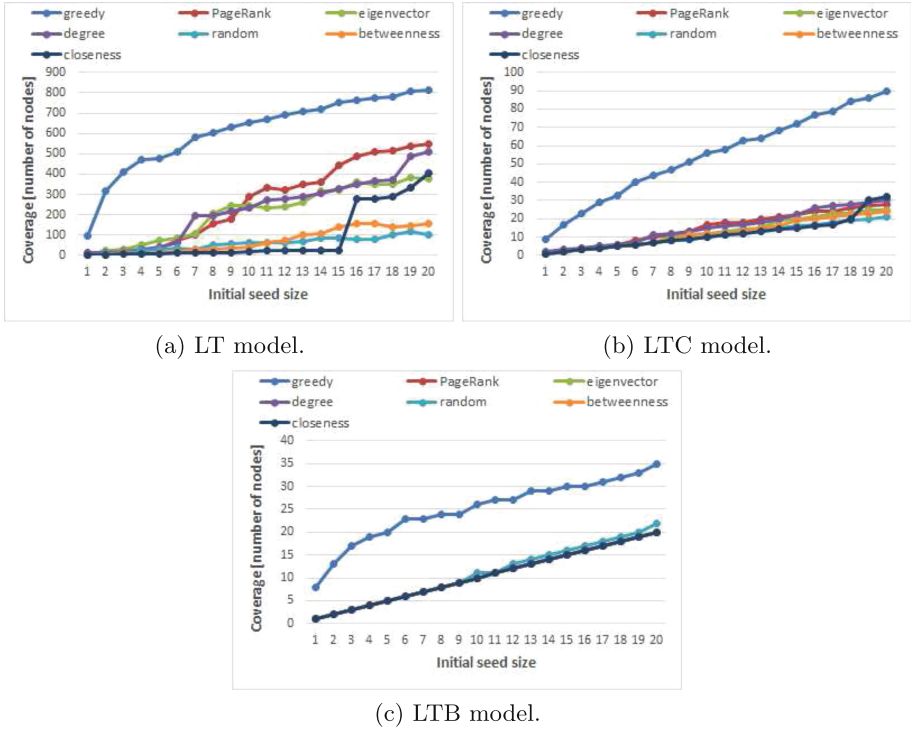


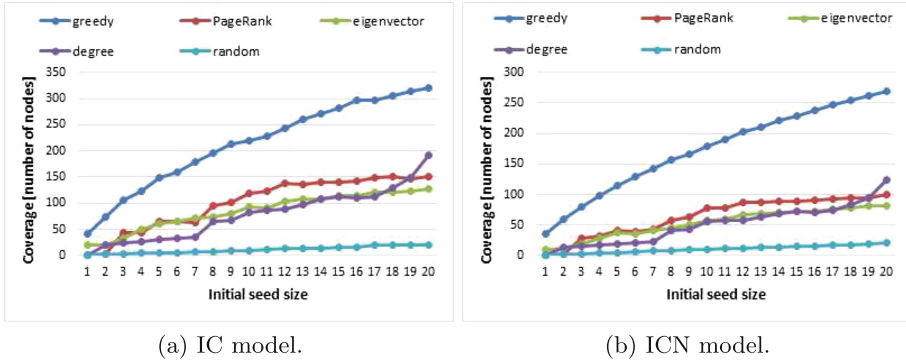
Fig. 3. LastFm - models based on LT.

Table 1. Calculation times for LastFm dataset.

(a) Centrality metrics (times for all nodes)		(b) Greedy algorithm in different models (times for choosing a single node)	
Algorithm of metrics	Time [s]	Propagation model	Mean time [s]
Out-degree	0.008	IC	0.014
Eigenvector	0.198	ICN	0.013
PageRank	0.136	LT	0.037
Betweenness	73.504	LTC	0.095
Closeness	2 262.865	LTB	0.014

**Flixter Dataset.** In IC and IC models (Fig. 4a, b) we can observe analogous behaviour as it was presented in LastFm dataset - among centrality metrics, the best results are for PageRank, little worse results for out-degree and eigenvector centrality metrics.

For LT model (Fig. 5a) the situation is little different – eigenvector centrality provides better results than PageRank. Good coverage is also obtained using out-degree. In LTC model (Fig. 5b) PageRank has the best results among centrality



**Fig. 4.** Flixter - models based on IC.

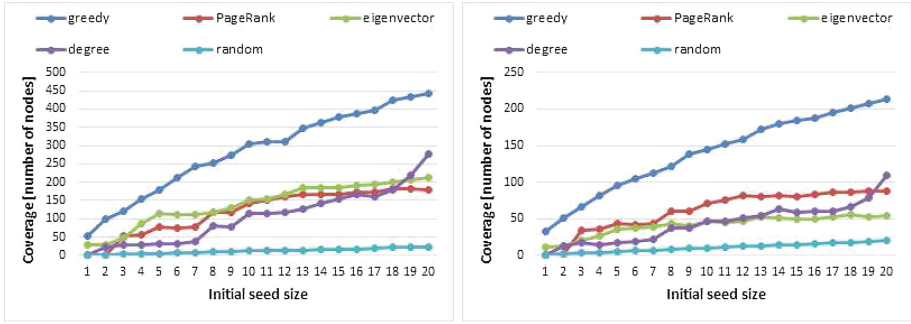
metrics (but for bigger seed sizes out-degree reaches higher coverage). In LTB model (Fig. 5c) the selection of centrality metrics has almost no influence on the coverage. One can also notice that random nodes' selection gives worse results in this dataset in comparison with *LastFm* dataset.

**Yelp Dataset.** Figure 6a and b present results for IC and ICN models. We can see that they behave very similar to each other. In comparison with previous datasets, PageRank and out-degree also provide a good coverage, but eigenvector centrality gave much worse results. One can observe that the coverage for this dataset is 10 times bigger than it was in previous datasets.

For LT and LTC models (Fig. 7a and b, respectively) the results are very similar to presented ones for IC and ICN models. In LTB model (Fig. 7c) centrality metrics have some effect on the final value of coverage (bigger than in previous datasets). The results obtained using centrality metrics are closer (than in previous datasets) to the coverage of greedy algorithm, and even in some cases greedy algorithm produced worse results than results from centrality measures (e.g. in LT model for selection only one node). One can notice that random selection of nodes gives bad results.

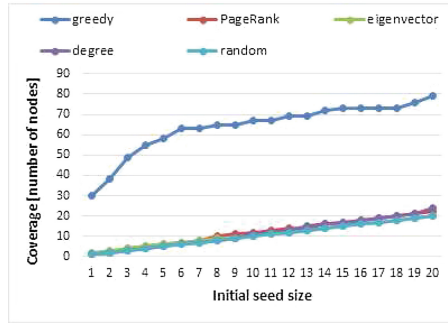
### 4.3 Influence Maximization on Reduced Graphs

All centrality metrics were only calculated on *LastFm* dataset, but for *Flixter* and *Yelp* it was not possible due to high computation complexity. Therefore, we decided to reduce graphs from these datasets. The reduction was conducted using *RandomWalk* algorithm [14]. In this subsection we provide results only for IC model – Fig. 8a and b. After the reduction, the number of nodes in the graph generated from *Flixter* dataset shrunked 6 times and in analogous graph from *Yelp* – 8 times, so it changes graphs significantly and conclusions drawn from these reduced graphs may be different than ones from full graphs. In both datasets the results obtained using PageRank, out-degree and eigenvector centrality are very similar. We can observe that betweenness produced very small



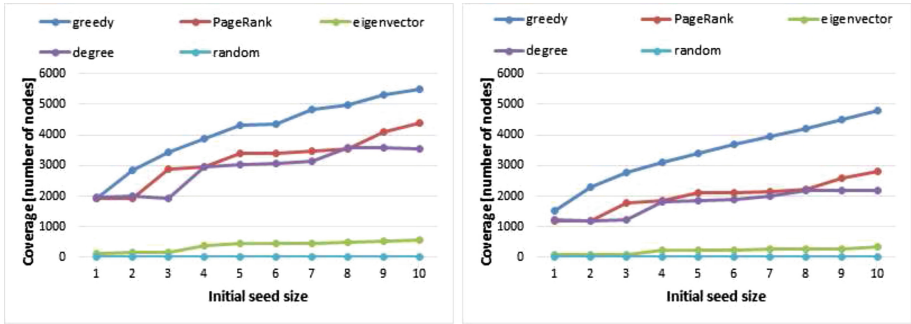
(a) LT model.

(b) LTC model.



(c) LTB model.

Fig. 5. Flixter - models based on LT.

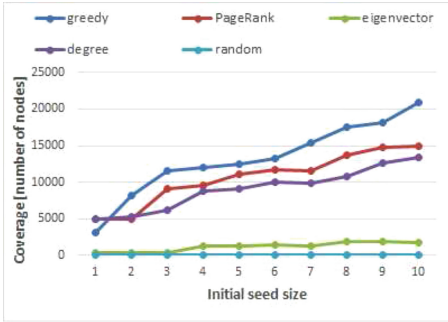


(a) IC model.

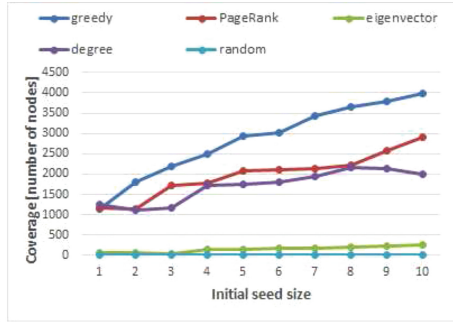
(b) ICN model.

Fig. 6. Yelp - models based on IC.

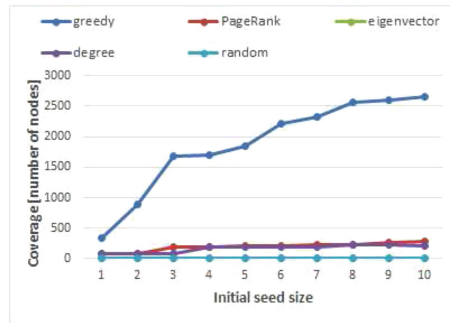
coverage. Regarding the centrality metrics, the behaviour is similar to observed one in the case of *LastFm* i.e. the selection using this metric for some high ranked nodes gives poor results, but at some point there is a sudden increase, and in the case of reduced *Yelp* dataset its coverage was higher than the coverage of other centrality metrics.



(a) LT model.

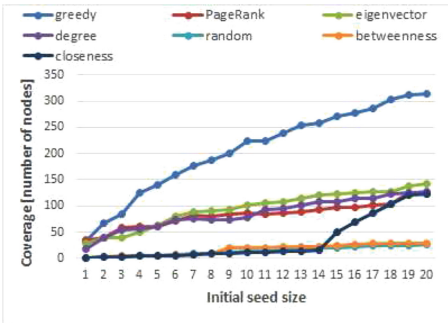


(b) LTC model.

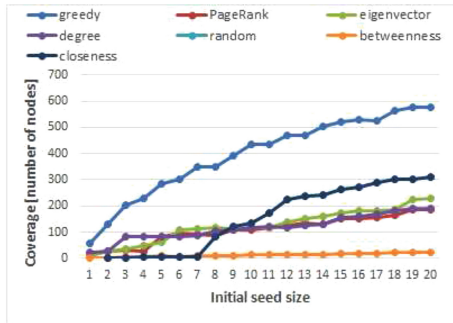


(c) LTB model.

Fig. 7. Yelp - models based on LT.



(a) Flixtter



(b) Yelp

Fig. 8. Reduced graphs - IC models.

## 5 Conclusion and Future Work

In this paper we studied using various centrality metrics in the influence maximization problem. We considered different models of influence propagation.

In all cases, the best results were achieved using greedy algorithm, which may be time-consuming when we select many nodes. Therefore, selection of nodes based on their ranking from centrality metrics may be a good option. Generally, very good results are provided when we select nodes on the basis of ranking from PageRank. Using out-degree is also a good alternative. Betweenness and closeness produced significantly worse results. Moreover, they are also very expensive computationally and rather should not be considered for nodes selection.

Analyzing the results of choosing different methods of nodes selection in various propagation models, we can notice similar behavior and usage of centrality metrics gave much better results than random selection of nodes. In LTB model, such difference was the smallest (only in the case of *Yelp* dataset there is significant gain from choosing nodes based on the centrality metrics instead of choosing them randomly).

Future work will cover comparison with other methods of node selection. We are also planning to conduct comparisons of such methods on different datasets. Finally, we would like to repeat experiments with taking into account dynamics of networks (respecting the fact that edges can change in time).

**Acknowledgements.** The research reported in the paper was partially supported by the grant No. DOBR-BIO4/060/13423/2013 from the Polish National Centre for Research and Development.

The authors thank to Bartosz Niemczura, a student of Computer Science of AGH-UST, for his collaboration.

## References

1. Bhagat, S., Goyal, A., Lakshmanan, L.V.: Maximizing product adoption in social networks. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, pp. 603–612. ACM, New York (2012)
2. Chen, W., Collins, A., Cummings, R., Ke, T., Liu, Z., Rincón, D., Sun, X., Wang, Y., Wei, W., Yuan, Y.: Influence maximization in social networks when negative opinions may emerge and propagate. In: Proceedings of the Eleventh International Conference on Data Mining, SDM 2011, Mesa, Arizona, USA, 28–30 April 2011, pp. 379–390 (2011)
3. Chen, Y.C., Peng, W.C., Lee, S.Y.: Efficient algorithms for influence maximization in social networks. *Knowl. Inf. Syst.* **33**(3), 577–601 (2012). <http://dx.org/10.1007/s10115-012-0540-7>
4. Eccleston, D., Griseri, L.: How does web 2.0 stretch traditional influencing patterns. *Int. J. Mark. Res.* **50**(5), 591–616 (2008)
5. Freeman, L.C.: Centrality in social networks conceptual clarification. *Soc. Netw.* **1**(3), 215–239 (1979)
6. Godes, D., Mayzlin, D.: Firm-created word-of-mouth communication: evidence from a field test. *Mark. Sci.* **28**(4), 721–739 (2009)
7. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: A data-based approach to social influence maximization. *Proc. VLDB Endow.* **5**(1), 73–84 (2011)
8. Jamali, M., Ester, M.: A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 135–142. ACM, New York (2010)

9. Katz, E., Lazarsfeld, P.: *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Transaction Publishers, New Brunswick (2005)
10. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146. ACM (2003)
11. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics-II. The problem of endemicity. *Proc. Roy. Soc. London Math. Phys. Eng. Sci.* **138**(834), 55–83 (1932)
12. Kundu, S., Murthy, C.A., Pal, S.K.: A new centrality measure for influence maximization in social networks. In: Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K. (eds.) *PReMI 2011. LNCS*, vol. 6744, pp. 242–247. Springer, Heidelberg (2011)
13. Leavitt, H.J.: Some effects of certain communication patterns on group performance. *J. Abnorm. Soc. Psychol.* **46**(1), 38 (1951)
14. Lovasz, L.: *Random walks on graphs: A survey* (1993)
15. Narayanam, R., Narahari, Y.: A shapley value-based approach to discover influential nodes in social networks. *IEEE Trans. Autom. Sci. Eng.* **8**(1), 130–147 (2011)
16. Niemczura, B., Gliwa, B., Zygmunt, A.: Linear threshold behavioral model for the spread of influence in recommendation services. In: *The Second European Network Intelligence Conference, ENIC 2015*, pp. 98–105 (2015)
17. Page, L., Brin, S., Motwani, R., Winograd, T.: *The pagerank citation ranking: Bringing order to the web* (1999)
18. Zafarani, R., Abbasi, M.A., Liu, H.: Influence and homophily. In: Zafarani, R., Abbasi, M.A., Liu, H. (eds.) *Social Media Mining*. Cambridge University Press (2014)
19. Ryan, B., Gross, N.C.: Acceptance and diffusion of hybrid corn seed in two Iowa communities, vol. 372. *Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts* (1950)
20. Shaw, M.E.: Group structure and the behavior of individuals in small groups. *J. Psychol.* **38**(1), 139–149 (1954)
21. Wei, L., Laks, V.S.: Profit maximization over social networks. *CoRR* abs/1210.4211 (2012)
22. Zhuang, H., Sun, Y., Tang, J., Zhang, J., Sun, X.: Influence maximization in dynamic social networks. In: *2013 IEEE 13th International Conference on Data Mining (ICDM)*, pp. 1313–1318, December 2013