

# The effect of global and local influence models on the quality of recommendations

Nikolaos Mantas  
*Informatics and Telecommunications Engineering*  
*School of Engineering*  
*University of Western Macedonia*  
Kozani, Greece  
nmantas@uowm.gr

Magdalini Eirinaki  
*Computer Engineering Department*  
*San Jose State University*  
San Jose, CA, USA  
magdalini.eirinaki@sjsu.edu

George Karetzos  
*Department of Computer Engineering*  
*Technology Education Institute of Thessaly*  
Greece  
karetzos@teilar.gr

Malamati Louta  
*Informatics and Telecommunications Engineering*  
*School of Engineering,*  
*University of Western Macedonia*  
Kozani, Greece  
louta@uowm.gr

Iraklis Varlamis  
*Department of Informatics and Telematics*  
*Harokopio University of Athens*  
Athens, Greece  
varlamis@hua.gr

**Abstract**— *The quality of recommendations on social networks is a combination of the richness of the available information and the ability of algorithms and architectures to take advantage of this information in favor of the users. Recommendation algorithms have to address several problems, such as information sparsity, scalability of algorithms, concept drift etc. In this dynamic and complex environment, it is important to provide solutions that enrich information when it is necessary to fill the gaps and at the same time to scale solutions so that they can handle the ever increasing data sizes and flows. In this work, we extend our previous work on recommender systems for social networks that studied global influence and trust metrics and their applications. More specifically, we introduce a local influence model, which is relied on the formation of local user networks based on common interests and study the performance of the new model, both stand-alone and in combination with the global one. Results show a promising improvement on the similarity between a target user and the users recommended based on the users selected to influence the recommendations for that target user.*

**Keywords**—*Social networks, trust, recommendation, clustering.*

## I. INTRODUCTION

Social network analysis has been an active research domain for many decades, but has gained even bigger attention with online social networks, where the number of interactions between users has dramatically increased and the semantics of these interactions have flourished. For example, in social networking sites, people form friendship, trust, or follower bonds with other users, exchange content and messages with other users or groups of users, approve or disapprove the actions or contents of their fellows or any other user in the network, etc. In this complex and semantics-dense environment

there exist: a) works that attempt to represent online social interactions using networks, with nodes (or supernodes) and edges (or hyperedges), that carry weight, direction or type information [19], b) works that attempt to analyze networks [25] and extract useful information [8] and c) works that focus on the issues that may arise from the analysis of such networks [3].

The identification of user communities and the development of recommendation systems for users or content are common applications in online social networks. The modeling of influence propagation among the actors (nodes) in a social network is another problem of great interest to the research community. This is of crucial importance in customer networks, where the attraction of new customers and the retention of existing customers are of high business interest [35], [5]. The link of network analysis with the concept of “word of mouth” and viral marketing made influential users crucial for the promotion and endorsement of new products or ideas. A simple notion of user influence relates to the number of users directly or indirectly connected to a user, but the variety of edge types, the existence of direction or (positive or negative) weights in edges increases the complexity of the problem in a real scenario [15].

In the typical scenario of social recommendations, two types of information are combined: local information, which represents the relationships among the user and his/her friends/connections (within a certain circle), and global information, which corresponds to the reputation or influence of a user in the entire social network [26]. The recommendation strategy followed in this work is based on locating users that are most probable to influence a user’s

decision. Thus, it looks first for users that are directly connected to the user through trust connections (i.e. belong to the user's "circle of trust") [30] and then for users that share similar interests with the target user ("user cluster"). This combination of local influences which are based a) on trust and b) on the similarity of interests, is expected to improve recommendations' quality, since the users that are selected to influence the target user have more features in common with the target user.

In this work, we extend our previous work that employed the notion of trust [9] among users and the notion of global "influence" of users based on various influence metrics, such as the degree, closeness, betweenness and centrality or the hub, authority or PageRank scores [4], [27] and [28]. We add a local "influence" factor, which is based on users who, based on their similar interests, are grouped together in the same cluster with the target user. Through experimental evaluation, we study the effect of local and global influence factors, and show that local influence can lead to better recommendations.

The remaining of this paper is organized as follows. In the Section II, we present an overview of related research works on influence, trust and recommendations in social networks and provide a summary of the global influence model that we first introduced in [27]. In Section III, we present the extended influential model and in Section IV, we discuss our experimental evaluation results. Finally, Section V summarizes our findings and presents our next steps.

## II. RELATED WORK

The study and analysis of Web 2.0 media, such as social networks, blogs, wikis, forums etc. has gained a big momentum, resulting in an increase of research in the related fields. Among the several facets of these social media, trust, influence, and ranking are receiving a lot of attention.

Several researchers have addressed the topic of trust prediction and propagation. Most of them suggest classification models, such as SVM-based methods ([18], [22]) in order to assign trust class labels using features such as user profile, user interactions, product reviews, and trust relations. A different approach is that of Lim et. al. [24], that employs the "Trust Antecedent" framework proposed in management science and introduces quantitative - instead of qualitative - features, such as ability, benevolence and integrity in the prediction process. Some other works focus on how trust is propagated in a network of people ([9], [21]). In our work we assume that the trust between a pair of users in a social network is already known, either explicitly or implicitly. In addition, trust propagation is thought to be covered by the more general notion of "influence" within such a network.

Influence in social networks, a topic extensively studied in the pre-WWW era [32], have attracted a lot of attention. Domingos et al. [2] have represented the identification of influencers as a combinatorial optimization problem, according to which given a fixed number of nodes that can be initially activated or infected, the set of nodes with maximum influence over the entire network is found. Kempe et al. [13] have proposed two basic information diffusion models i.e., Linear Threshold Model and the Independent Cascade Model. The

social network is represented as a Directed Graph and each node in the graph is considered to be either active or inactive. Probability of an inactive node to become active depends on the number of the considered inactive node's active neighbors. This way other nodes which are neighbors to the recently activated node might also tend to become active. Several researchers have focused on Information Cascade (IC) notion proposing various machine learning algorithms ([25], [16], [13], [36], [7]). Although these approaches have improved over traditional social network analysis metrics, they rely solely on the link structure of social network and do not take into consideration other significant parameters, such as activity, rate of updates, and trust among users. Similarly, researchers have investigated the identification of likely influential users through link analysis techniques [25], as well as user activity-related parameters in order to identify influential users in social networks [14]. Xie et al. [34] identify influencers in a social network by analyzing features like influencing capabilities of the user and their friends, influence capability of particular action and developing a model which would predict the cascading triggered by a user or an action. Javari et al. [11] also pointed out that the nature of relationship depicts an imperative feature for considering influence propagation.

Another important issue to consider is scalability. Chen et al. [1] have proposed an algorithm to improve influence maximization by making it scalable to millions of nodes and edges. A new MIA (Maximum influence Arborescence) algorithm is proposed, where the local structures of each node are used to approximate the influence propagation. As a first step, maximum influence paths (MIP) between every pair of nodes is calculated using Dijkstra's algorithm. MIPs with probability less than a threshold are ignored, restricting influence to a local region. Subsequently, MIPs with same starting or ending nodes are merged into arborescence (MAP) structures. Only influence propagated through these structures is considered.

Ranking on the web is mainly based on the analysis of the web graph, expressed by hyperlinks. In the case of blogs, various classification algorithms that exploit explicit (EigenRumor [23]) and/or silent (BlogRank [17]) links have been proposed. All these algorithms represent blogs as a graph based on hyperlinks and then apply on it PageRank in order to provide an overall ranking of blogs. However, all these algorithms provide a static measure of blog importance that does not reflect the temporal aspects accompanying the evolution of the blogosphere. Voudigari et al. [31] proposed the "Rank degree" algorithm, which works with undirected graphs and explores them based on the degree values of nodes and their rankings. The first  $s$  nodes are selected randomly and form the seed set. For each node in the seed set, all connected nodes and their corresponding degrees are found. Based on the degrees of connected nodes, the nodes are ranked in a descending order. They consider the top- $k$  friends and these are added as new seeds to the seed set. This process is repeated for all the nodes in the initial set and the duplicate values are removed. Lately, some effort has been done to also integrate the content in the ranking process, when ranking twitterers (TwitterRank [33]).

In [29] we introduced a collaborative rating mechanism, which employs direct and indirect information from a user's neighborhood. Specifically, it exploits the explicit connections between users and other implicit connections in order to provide personalized rankings. In [20] we presented a global rating model for the blogosphere. The model distinguishes between explicit links between blogs and implicit links between individual posts. The model also captures the time dimension of links, to punish blogs that artificially receive a large number of links in a small period of time and are ignored.

To the best of our knowledge, this is the first extensive study of the effect of both local "influence" that relies on the formation of local user networks based on common interests, as well as personalized aspects of "influence" such as trust, in ranking and recommending other users or content.

#### A. Global Influence (GI) model

The Global Influence (GI) model that we first introduced in [27] assumes that users are connected in a network of trust, with positive trust edges. The result is a directed graph representation of the network, in which a target user is influenced directly by his/her circle of trust and indirectly by one or more overall influential users of the social network. As a result, the collaborative recommender system that suggests items based on what the user's trustees recommend is extended with items that users in the immediate network of trust of the user recommend and with items that users with an important position in the graph (e.g. centrality, prestige etc.) recommend.

The global influence model results in a global ranking of all users in a social network, based on their position in the social graph and their connections to all other users. In essence, the global influence  $GI_{(i)}$  of user  $i$  is an indication of the importance of this user in the whole social graph and is a linear combination of six global metrics of node importance: degree, closeness and betweenness centrality, hub and authority score and PageRank score as follows:

$$GI(i) = w_d G_d(i) + w_c G_c(i) + w_b G_b(i) + w_h G_h(i) + w_a G_a(i) + w_p G_p(i) \quad (1)$$

Consequently, the influence score for any user  $j$  is a function of the ratings/trust provided for the user  $j$  by a) user  $i$  ( $LAS_c(i, j)$ ), b) the network of trust of user  $i$ , and c) the globally influential users:

$$INF_c(i, j) = f \left( \begin{array}{l} LAS_c(i, j), \\ \sum_{\substack{(k, j) \in E \\ (i, k) \in E}} w_k \cdot LAS_c(k, j), \\ \sum_{(m, j) \in E} GI(m) \cdot LAS_c(m, j) \end{array} \right) \quad (2)$$

### III. EXTENDED INFLUENTIAL MODEL

In this section, we present the basic characteristics and assumptions of our proposed model, which is an extension of the Global Influence model presented in the previous section. Many of the key facts of this model are the same as in [27] hence we focus on the differences between the two models.

#### A. Cluster Influence (CI) Model

The Global Influence (GI) model as described in our previous work [27], analyzes the entire social graph and identifies the most influential users across the network. In this study we introduce a local model of influence, namely Cluster Influence (CI) model, which focuses on the identification of the most influential users within user groups (clusters) that share common interests. The main assumption behind the CI model is that apart from counting on user influence, it is of equal importance to search for influencers that have similar behavior or common interests with the target user.

The locality of CI model is based on the identification and rewarding of the most influential users within each user cluster and not at a global scale. Thus, it is not necessary to analyze the entire network in order to calculate a user's influence, or in order to aggregate information (e.g. ratings) for the influencers, but we can focus on measuring the influence of each user to the users belonging to the same cluster. Consequently, the first step is to group users based on their interests and then to calculate the influence of each user based on the relationships with other users in the same cluster.

In the light of the aforementioned, the  $CI_{(i)}$  of a user  $i$  is an indication of his/her importance within the cluster he/she belongs to, and can be defined as a linear combination of popular social network metrics of importance and/or influence. In this work, we build on the same social network metrics that we employed in [27]. So  $CI_{(i)}$  is defined as follows in a similar manner to Eq. 1, albeit within the cluster the user  $i$  belongs to:

$$CI(i) = w_d C_d(i) + w_c C_c(i) + w_b C_b(i) + w_h C_h(i) + w_a C_a(i) + w_p C_p(i) \quad (3)$$

where  $C_d$  stands for degree centrality,  $C_c$  for closeness centrality,  $C_b$  for betweenness centrality,  $C_h$  and  $C_a$  for hub and authority metrics respectively and  $C_p$  for the PageRank metric. Using Eq. 3, we can define the importance of each metric through the respective weights.

According to the above, our proposed model for social networks quantifies the influence  $INF_c(i, j)$  of a user  $i$  to another user  $j$  using a function that combines: a) the direct influence of user  $i$  to user  $j$  ( $LAS_c(i, j)$ ), b) the total influence of the network of trust of user  $i$  ( $LAS_c(k, j) \forall k \in NT_i$ , where  $NT_i$  comprises the users that user  $i$  trusts –  $i$ 's network of trust), and c) the total influence from the influential users of the cluster that user  $i$  belongs to ( $LAS_c(n, j) \forall n \in CL_i$ , where  $CL_i$  comprises the users that have similar interests to user  $i$ ):

$$INF_c(i, j) = f \left( \begin{array}{l} LAS_c(i, j), \\ \sum_{k \in NT_i} w_k \cdot LAS_c(k, j), \\ \sum_{n \in CL_i} CI(n) \cdot LAS_c(n, j) \end{array} \right) \quad (4)$$

This function could be implemented as the weighted sum of the three factors of Eq. 4:

$$INF_c(i, j) = w_{local} \cdot LAS_c(i, j) + w_{collab} \sum_{k \in NT_i} w_k \cdot LAS_c(k, j) + w_{cluster} \cdot \sum_{n \in CL_i} CI(n) \cdot LAS_c(n, j) \quad (5)$$

The definition given in Eq. 5 has a dual meaning. On one hand, a member of a social network decides based on his/her own preferences but is also influenced by the preferences (and recommendations) of the people he/she trusts. On the other hand, he/she is influenced by the most influential members of the whole network, given that they share common interests.

The weights used in the proposed model are normalized to sum to 1 and represent the significance of the three types of influence:  $w_{local}$  for the user's own beliefs,  $w_{collab}$  for the user's extended network beliefs and  $w_{cluster}$  for influential user's beliefs. In addition, each component weighs differently each participant, and assigns a different weight  $w_k$  to each user  $k$ , who is in the trust network of user  $i$  and a different weight  $CI_{(n)}$  to each influential user  $n$ . The latter is proportional to the importance of user  $n$  in the cluster he/she belongs to.

The proposed model can be appropriately adapted to include users who are new in a social network and haven't yet formed a network of trust. In this case, the influence score can be based on the cluster influence and adjusted as the user starts to make more network relationships.

The final outcome of our proposed model is to provide for each user a personalized list of influence scores for other users who have common interests. This list of influence scores can be used to rank the users, and this ranking can be subsequently used to generate personalized recommendations to the current user  $i$ .

### B. User interests representation

Our objective is to create clusters of users who have common interests. Depending on the social network, user interests may vary. For example, in the blogosphere we can have different types or thematic categories of blogs, in a product review site we can have various product categories and users that review products in each category etc.

For defining interest-based user clusters, it is necessary to quantify user's interests, represent them and then group users based on this representation to the space of interests. For this purpose, we represent each user  $U_i$  as a weighted vector in the space of interests as follows:

$$\vec{U}_i = (w_1, w_2, \dots, w_s) \quad (6)$$

where  $s$  is the number of different interests or thematic categories, while the weight  $w_s$  expresses the value for user  $U_i$  associated with the  $S_{th}$  interest or category.

The vector-based representation of user's interests can also be mapped to a  $M \times S$  matrix of users - categories, consisting of  $M$  rows representing the number of users and  $S$  columns representing the different categories or interests. The matrix values for row  $r_i$  represent the weighted vector of interests of user  $u_i$ .

### C. User Clustering

Given a vector space model representation of user interests as the one described above, it is possible to use any vector-based measure of user similarity, such as cosine similarity

$$\cos(i, j) = \frac{u_i \cdot u_j}{\|u_i\| \|u_j\|} = \frac{\sum_{k=1}^n (w_{i,k} \cdot w_{j,k})}{\sqrt{\sum_{k=1}^n w_{i,k}^2} \sqrt{\sum_{k=1}^n w_{j,k}^2}} \quad (7)$$

or a simple similarity measure that counts the number of co-rated items (i.e. all non-zero weights in Eq. 7 are equal to 1), or any user distance measure, such as Euclidian or Manhattan distance. The combination of a similarity or distance measure and any clustering algorithm (density-based or centroid-based, flat or hierarchical) will result to a partitioning of the set of users to clusters of similar interests. Depending on the clustering algorithm, the resulting clusters may differ, they may overlap or not, they can have convex or arbitrary shape. By applying a centroid based clustering algorithm (e.g. k-Means [10]) we will end up with clusters where all users resemble to all other users and all users will belong in a cluster, whereas using a density based algorithm (such as DB-SCAN [6]) we will end up with clusters where users have similar interests to many other users in the cluster but not to all of them and we will have users that do not cluster together with other users (noise). Finally, using graph-partitioning algorithms, such as METIS [12] results in non-overlapping and equally sized user clusters, which can combine both user interest similarity and user proximity in the graph if the original user's trust network is expanded with more similarity-based edges.

## IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed mechanism for providing recommendations to social network users. Specifically, we study several scenarios regarding the contribution of the local models (L, CL) and the cluster influence model (CI). In addition, we examine the performance of our model for different parameter values.

The Epinions dataset was used for the evaluation. The dataset contains information (user ratings) for product reviews written by the members of the Epinions community and trust scores between the community members. Moreover, it contains information about the author and subject of each review, from which we extract the interests of each author.

In our experiments we found the 10 most popular product categories (i.e. the ones with the most reviews) and we subsequently selected a subset of the original dataset that comprises users that review products in these categories. To study the results thoroughly, we divided the dataset into two equal size subsets. Dataset *A* includes users with a narrow community of trust (users that trust a limited number of users, i.e. between 5 and 10), while dataset *B* includes users with an extended circle of trust (having more than 30 trust bonds).

In order to evaluate the proposed model, we examine whether all users  $U$ , suggested to the user  $U_i$ , have common interests. The similarity between two users  $U_i$  and  $U_j$  is defined as the number of items (reviews in the case of the Epinions dataset) they have co-rated. The y-axis in all the following figures is labeled "average similarity", because it averages the similarity values of all users in the examined dataset.

We first apply the k-means clustering algorithm and examine the effect of the parameter  $k$ , which determines the number of clusters in the cluster influence model, and

parameter  $m$ , which refers to the number of recommendations for each user. Then, we compare our proposed cluster influence (CI) model with global influence (GI) that was introduced into our previous study [27]. Finally, we evaluate each model individually and combine collaborative local with cluster influence.

#### A. Defining the parameter $k$

The first experiment examines the effect of parameter  $k$  in the cluster influence model. For this reason, various  $k$  values, in the range [2, 10] are employed for k-means clustering of both subsets  $A$  and  $B$ . We used Euclidian distance as user distance measure.

For each user  $U$ , a registered list of suggested users was generated using the cluster influence model (Eq. 5), using the appropriate weights ( $w_{local} = 0, w_{collab} = 0, w_{cluster} = 1$ ).

We then calculate the cluster influence rankings in all user clusters using all centrality scores ( $C_d$  using degree centrality,  $C_c$  using closeness centrality,  $C_b$  using betweenness centrality,  $C_h$  using hub score,  $C_a$  using authority score and  $C_p$  using PageRank score) and select the top- $m$  users from each list. These are the recommended users. To calculate the average similarity of users, the top 20 users ( $m = 20$ ) were used.

Fig. 1 and Fig. 2 illustrate the performance of the cluster influence model in datasets  $A$  and  $B$ , respectively, for different values of  $k$  between 2 and 10. As we can see, PageRank, betweenness and degree centrality outperformed all other social network analysis metrics.

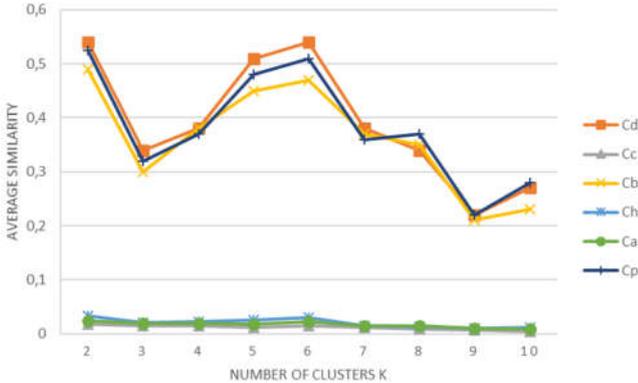


Fig. 1. The effect of parameter  $k$  of the cluster influence model in dataset A

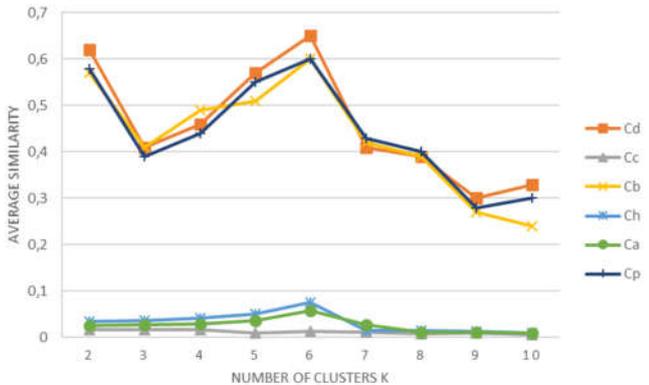


Fig. 2. The effect of parameter  $k$  of the cluster influence model in dataset B

In addition, the figures show that the number of clusters influences significantly the performance of the proposed model. Specifically, when clusters are between 3 and 6, we see an increase in model efficiency as  $k$  increases. For values of  $k$  greater than 6, the efficiency of the proposed model declines. The maximum performance of the model is shown for  $k = 6$ . The results are similar in both subsets  $A$  and  $B$ , showing that the effect of the size of network of trust is similar in all cases.

#### B. The number of recommended users

In this scenario, we examine how the number of recommendations affects the performance of our cluster influence model. We recall that for each cluster of users, we calculate user's influence using Eq. 3. The importance of a user is related with weight  $CI_{(n)}$  of Eq. 5, which is used to calculate his/her influence to other users of the same group. Finally, top  $m$  users with the highest ranking are proposed.

Fig. 3 and Fig. 4 illustrate the performance of the cluster influence model in datasets  $A$  and  $B$ , respectively, for different values of  $m$  ( $m = 3, 5, 10, 15, 20, 25, 30$ ) setting  $k$  value (i.e. number of introduced clusters) equal to 6 ( $k = 6$ ).

As already mentioned, the similarity between two users is defined as the number of items they have co-rated. When the list of the top- $m$  users includes users with similar interests, the similarity is expected to increase as  $m$  increases. This is due to the fact that the number of items that have been rated by both the user and a user in the list is expected to increase. When the top- $m$  list contains users with completely different interests, the similarity will remain stable. The aforementioned are also confirmed by figures 3 and 4, as an increase in the value of the parameter  $m$  causes an increase in the average similarity.

The maximum values of the average similarity are shown in dataset  $B$ , which consists of users with many trusted nodes in their circle. The  $C_p$  (PageRank) metric appears to outperform other metrics across all datasets when the list of suggested users includes up to 15 users. On the other hand, when the list of recommendations has more than 15 users, the  $C_d$  (Degree Centrality) metric outperforms all other metrics across all datasets. The curve of  $C_b$  (Betweenness Centrality) metric has a similar behavior but is less efficient than  $C_p$  and  $C_d$ . For all other metrics, there are no similarities between the user's interests and those of the top- $m$  ranked users in the clusters of the network.

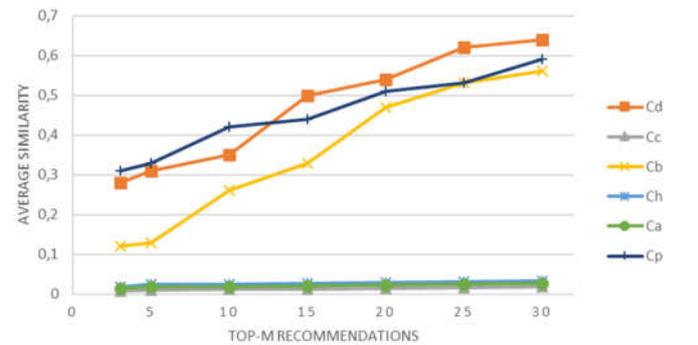


Fig. 3. The effect of number of recommendations in cluster influence model in dataset A

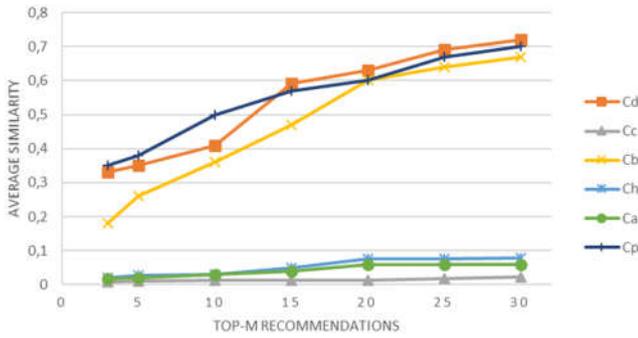


Fig. 4. The effect of number of recommendations in cluster influence model in dataset B

Using more recommended users (i.e. selecting higher  $m$  values), results in including more users that are more likely to have common interests with the target user. It is observed that for small values of  $m$  (less than 15) the average similarity is low, while we found that large values of  $m$  (over 30) do not contribute much to our model, yielding almost the same average similarity (thus, they are not depicted in the figures); In the latter case, computational complexity and respective resource requirements are increased.

### C. Comparison of CI and GI Influence Models

In this comparative study, we focus on the contribution of the proposed cluster influence model (CI) compared to the global influence model (GI) model, as presented in [27], aiming at improving the quality of the provided personalized recommendations. As already mentioned, the *GI* model analyzes the entire social graph and identifies the most influential users across the network, while the newly introduced *CI* model focuses on the identification of most influential users within the user clusters of common interests.

Both models conclude that for the particular type of social network and considered datasets, PageRank and Degree Centrality are those that mostly affect the average similarity and are thus better in identifying influence. Thus, in the experiments performed subsequently, we used these metrics, applying the corresponding weights ( $w_d = 1, w_c = w_b = w_h = w_a = w_p = 0$  for degree centrality and  $w_p = 1, w_d = w_c = w_b = w_h = w_a = 0$  for PageRank) to equation 3.

Fig. 5 and Fig. 6 show the performance of the two models for datasets *A* and *B*, respectively, for different values of  $m$ . The curves  $GI_d$  and  $GI_p$  represent the average similarity of the *GI* model using only the Degree Centrality and PageRank metrics, respectively, while the curves  $CI_d$  and  $CI_p$  represent the average similarity of the *CI* model using the same metrics.

As we can see, the *CI* model outperforms the *GI* model in both datasets. Specifically, for the Degree Centrality metric, the average improvement of the model is about 17%, with a highest value of 25% when  $m = 15$ . For PageRank metric, the average improvement of the model depends on the dataset. In the dataset *A*, which includes users with a small circle of trust, the average improvement for all values of  $m$  is also about 17%, while in dataset *B*, with users with a large circle of trust, the average improvement is about 8%.

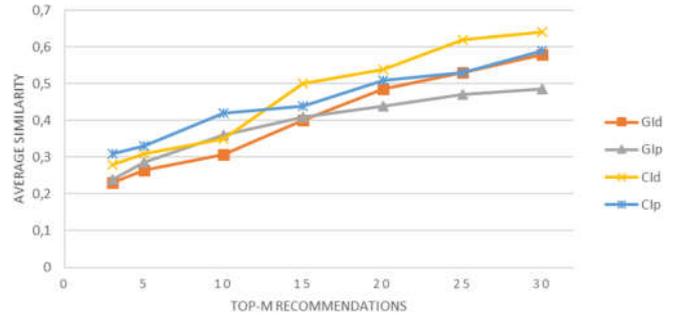


Fig. 5. Comparison of CI and GI influence model in dataset A

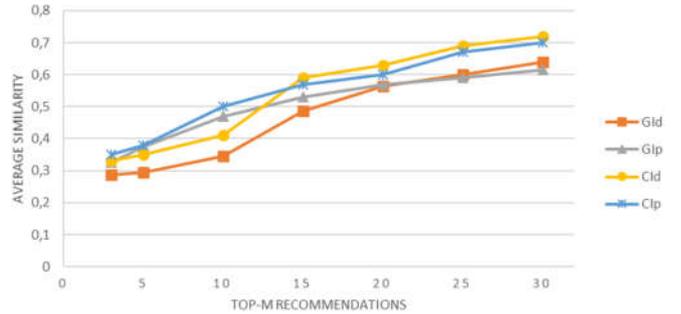


Fig. 6. Comparison of CI and GI influence model in dataset B

Based on the above, for this type of social network and dataset, we recommend using PageRank metric for small values of  $m$  (under 15) and users with a small circle of trust. In all other cases, the use of the Degree Centrality metric yields improved results.

### D. Evaluating the proposed model contribution

In this experiment, we study the contribution of our proposed model. We evaluate individually each model, namely, the Local (*L*), the Collaborative Local (*CL*) and Cluster Influence model considering the Degree Centrality metric ( $CI_d$ ) (setting the respective weight in Eq. 3 to 1 and the remaining weights to 0). In addition, based on the results of this experiment, which show that the *CL* model outperforms the *L* one, we combine the *CL* model with  $CI_d$  model, called  $CL/CI_d$ .

For each user  $U_x$  in datasets *A* and *B*, we generate a ranked list of recommended users, using the individual models *L*, *CL* and  $CI_d$ . Then, we select and recommend the top- $m$  users from each list.

In order to comparatively evaluate the performance of the models, we use a trust-based baseline method *T*, which is the average similarity between  $U_i$  and the users  $U_t$  to whom  $U_i$  is connected via an explicit trust link ( $(U_i, U_t) \in E$ ).

Fig. 7 and Fig. 8 represent the average similarity of a user with  $m$  recommended users for datasets *A* and *B*. As illustrated in the figures, the baseline reference curve *T* increases when the size of user's trust circle increases, with improved values acquired in *B* dataset. This is due to the fact that users have a large circle of trust and therefore many recommendations from which they can choose. However, both the *L* model and *CL* model improve the performance of the baseline *T*.

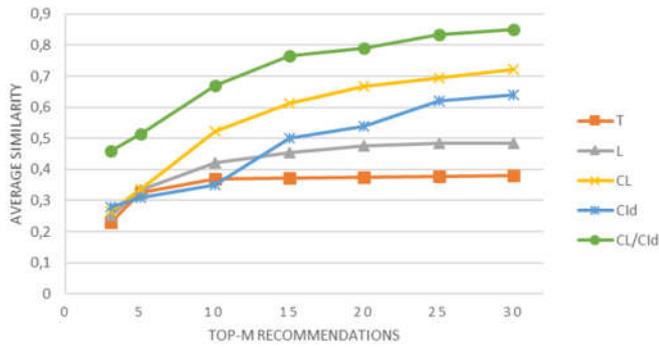


Fig. 7. Comparison of CI and GI influence model in dataset A

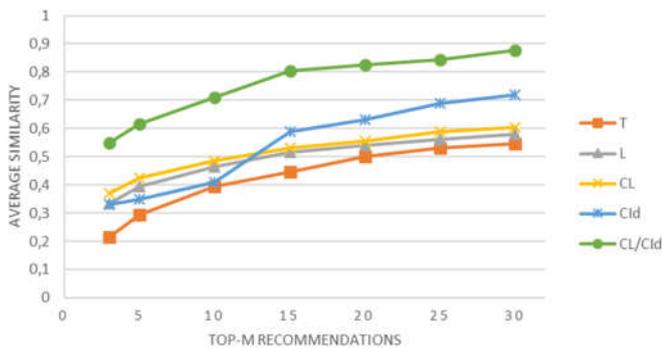


Fig. 8. Comparison of CI and GI influence model in dataset B

The  $CL$  model significantly improves the baseline method  $T$  and the improvement is greater in dataset  $A$  that includes users with a small circle of trust. The contribution of  $CL$  is greater than the local  $L$  model, which is based only on the direct of the user's circle of trust. This implies that it is useful for a recommendation model to check for suggestions beyond the direct neighbors of a node, in the extended neighborhood of users.

Studying the performance of  $CI_d$ , it is concluded that outperforms  $L$  model. In some cases, where users have a large circle of trust (dataset  $B$ ) and the model makes a lot of recommendations (more than 15), the  $CI_d$  outperforms the  $CL$  model. An explanation of this is that users in dataset  $B$  have many direct or indirect neighbors, so these users are probably connected to some users of the graph, who are also influential. This is an indication that recommendation mechanisms can be improved by using models based on influential or central users. These models can be valuable, mainly in the absence of local sources of recommendation when a user is new and hasn't yet formed a network of trust.

Recommendation models based entirely on user influence cannot affect the entire network, especially if it consists of thousands of users. However, when combining user's environment and the opinion of influential users, the results are significantly improved.

The  $CL/CI_d$  curve represents the performance of the model, combining  $CL$  and  $CI_d$  models and applying the relative weights accordingly to the Eq. 5 ( $w_{local} = 0$ ,  $w_{collab} = 0.5$ ,  $w_{cluster} = 0.5$ ). The improvement introduced to the provided recommendations applying the combined model with respect to

the corresponding  $CL$  curve for all values of  $m$  is about 35% in dataset  $A$  and about 42% in dataset  $B$ .

## V. CONCLUSIONS

This work extended previous work on global influence models that focused on the influence of users in social networks. In contrast to the previous Global Influence (GI) model that analyzed the entire social graph and identified the most influential users across the network, this work introduces a local model of Cluster-based Influence (CI). The  $CI$  model focuses on the identification of most influential users within clusters that comprise users with common interests, considering that it is more efficient to take into account users who are important in the community and have similar behavior and common interests. The results of the current study demonstrate the applicability of the  $CI$  model and are promising towards using local, interest-based models to measure user influence.

## REFERENCES

- [1] Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1029-1038). ACM.
- [2] Domingos, P., & Richardson, M. (2001, August). Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 57-66). ACM.
- [3] Eirinaki, M., Gao, J., Varlamis, I., & Tserpes, K. (2018). Recommender systems for large-scale social networks: A review of challenges and solutions.
- [4] Eirinaki, M., Louta, M. D., & Varlamis, I. (2014). A trust-aware system for personalized user recommendations in social networks. *IEEE transactions on systems, man, and cybernetics: systems*, 44(4), 409-421.
- [5] Eirinaki, M., Moniz, N., & Potika, K. (2016, October). Threshold-bounded influence dominating sets for recommendations in social networks. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)* (pp. 408-415). IEEE.
- [6] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).
- [7] Estevez, P. A., Vera, P., & Saito, K. (2007, August). Selecting the most influential nodes in social networks. In *2007 International Joint Conference on Neural Networks* (pp. 2397-2402). IEEE.
- [8] Freeman, L. C. (2017). *Research methods in social network analysis*. Routledge.
- [9] Golbeck, J. (2008). Trust on the world wide web: a survey. *Foundations and Trends® in Web Science*, 1(2), 131-197.
- [10] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666.
- [11] Javari, A., & Jalili, M. (2014). Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(2), 24.
- [12] Karypis, G., & Kumar, V. (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1), 359-392.
- [13] Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). 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.

- [14] Kim, E. S., & Han, S. S. (2009, July). An analytical way to find influencers on social networks and validate their effects in disseminating social games. In *2009 International Conference on Advances in Social Network Analysis and Mining* (pp. 41-46). IEEE.
- [15] Kim, E. D. J., Keng, B. J. L., & Padmanabhan, K. (2016). *U.S. Patent No. 9,262,537*. Washington, DC: U.S. Patent and Trademark Office.
- [16] Kimura, M., Saito, K., & Nakano, R. (2007, July). Extracting influential nodes for information diffusion on a social network. In *AAAI* (Vol. 7, pp. 1371-1376).
- [17] Kritikopoulos, A., Sideri, M., & Varlamis, I. (2006, October). BlogRank: ranking weblogs based on connectivity and similarity features. In *Proceedings of the 2nd international workshop on Advanced architectures and algorithms for internet delivery and applications* (p. 8). ACM.
- [18] Liu, H., Lim, E. P., Lauw, H. W., Le, M. T., Sun, A., Srivastava, J., & Kim, Y. (2008, July). Predicting trusts among users of online communities: an opinions case study. In *Proceedings of the 9th ACM conference on Electronic commerce* (pp. 310-319). ACM.
- [19] Liu, Y., Safavi, T., Dighe, A., & Koutra, D. (2018). Graph summarization methods and applications: A survey. *ACM Computing Surveys (CSUR)*, 51(3), 62.
- [20] Louta, M., Varlamis, I.: Blog rating as an iterative collaborative process. In: Wallace, M., et al. (eds.) *Semantics in Adaptive and Personalized Services. Springer Series on Studies in Computational Intelligence SCI* 279, pp. 187-203. Springer, Berlin/Heidelberg (2010).
- [21] Massa, P., & Avesani, P. (2005, July). Controversial users demand local trust metrics: An experimental study on epinions. com community. In *AAAI* (pp. 121-126).
- [22] Matsuo, Y., & Yamamoto, H. (2009, April). Community gravity: measuring bidirectional effects by trust and rating on online social networks. In *Proceedings of the 18th international conference on World wide web* (pp. 751-760). ACM.
- [23] Nakajima, S., Tatemura, J., Hino, Y., Hara, Y., & Tanaka, K. (2005, May). Discovering important bloggers based on analyzing blog threads. In *Annual Workshop on the Weblogging Ecosystem*.
- [24] Nguyen, V. A., Lim, E. P., Jiang, J., & Sun, A. (2009, December). To trust or not to trust? predicting online trusts using trust antecedent framework. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 896-901). IEEE.
- [25] Richardson, M., & Domingos, P. (2002, July). Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 61-70). ACM.
- [26] Tang, J., Aggarwal, C., & Liu, H. (2016, April). Recommendations in signed social networks. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 31-40). International World Wide Web Conferences Steering Committee.
- [27] Varlamis, I., Eirinaki, M., & Louta, M. (2010, August). A study on social network metrics and their application in trust networks. In *2010 International Conference on Advances in Social Networks Analysis and Mining* (pp. 168-175). IEEE.
- [28] Varlamis, I., Eirinaki, M., & Louta, M. (2013). Application of social network metrics to a trust-aware collaborative model for generating personalized user recommendations. In *The Influence of Technology on Social Network Analysis and Mining* (pp. 49-74). Springer, Vienna.
- [29] Varlamis, I., Louta, M.: Towards a personalized blog site recommendation system: a collaborative rating approach. *The International IEEE Workshop on Semantic Media Adaptation and Personalization, SMAP*, San Sebastián, Spain, December 14-15, (2009).
- [30] Victor, P., De Cock, M., & Cornelis, C. (2011). Trust and recommendations. In *Recommender systems handbook* (pp. 645-675). Springer, Boston, MA.
- [31] Vouidigari, E., Salamanos, N., Papageorgiou, T., & Yannakoudakis, E. J. (2016, August). Rank degree: An efficient algorithm for graph sampling. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 120-129). IEEE Press.
- [32] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.
- [33] Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). Twitterank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 261-270). ACM.
- [34] Xie, J., Kelley, S., & Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)*, 45(4), 43.
- [35] Zhang, M., Guo, L., Hu, M., & Liu, W. (2017). Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation. *International Journal of Information Management*, 37(3), 229-240.
- [36] M. Kimura, K. Yamakawa, K. Saito, and H. Motoda (2008). Community analysis of influential nodes for information diffusion on a social network. *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (pp. 1358-1363), IJCNN.