Employing Personality Feature to Rank the Influential Users in Signed Networks

Amir Mohammadinejad, Reza Farahbakhsh, Noël Crespi
Institut Mines-Télécom, Télécom SudParis, CNRS UMR 5157 SAMOVAR, France
amir.mohammadinejad, reza.farahbakhsh, noel.crespi}@it-sudparis.eu

***Abstract*—Social networks are an important part of the everyday activities of a big part of people. Different type of social-based activities (e.g. online product shopping, question answering forums and etc.) create a vast connection between users. One of the most important features of these networks is knowledge sharing. This knowledge usually provides better insight for the users and consequently has a direct impact on the decision made by them. For example, online shopping members usually take their decision based on this shared information. But the main issue is there are a huge amount of shared knowledge without an accurate mechanism to determine their validity. One approach is to count more on the influential users opinions in the system and toward this end, several ranking algorithms have been proposed. But the existing algorithms for users ranking don't consider the personality features of users in their methodology. In this paper, we use this new feature of personality in the ranking algorithm for influential user detection in signed networks. We used Optimism and Pessimism scores as personality features of each user and employ it in the PageRank algorithm as a sample ranking algorithm and evaluated the new ranking results by using a new metric of credibility. To assess the performance of the proposed method, we applied it to a large dataset of Epinions signed networks. The results are compared with state-of-the-art expert finding algorithms which indicate that the personality feature can effectively improve the ranking and influential user detection accuracy.**

***Index Terms*—Influential Users, Link Analysis, Expert Detection, Ranking algorithms, Credibility, Social Networks.**

I. Introduction

Online social networks are increasing and impacting our lives and studies show most of the people have a membership at least in one of these social networks or somehow using them in their everyday interactions with others people. Apart from very popular social networks like Facebook and Tweeter, there are other types of social based networks such as Amazon, Ebay or small online shopping websites which have user or product profiling and even communication features which make it possible to apply the social networks concepts to them. In this era of the digital world, lots of people are registered in different social networks and take tremendous decisions based on their knowledge and information such as buying a product from online shopping websites or booking a hotel or restaurant. However, there are a tremendous amount of shared information and there is no mechanism yet to distinguish the validity of them in an accurate way. Thus, the knowledge shared by users on social networks could not be fully trusted.

Recently, with the popularity of knowledge sharing in a social network, this problem attracted lots of attention. One of the main direction in this domain is identifying influential users or experts and rely more on their opinion. By identifying the influential or expert users in social networks and determining their level of knowledge, the reliability of their provided information could be identified. Also in recommending systems, finding these users is important due to the fact that the preferable choices of them can be recommended to other users. The expert finding is one of the most important subjects for mining from (web-based) social networks. The task of expert finding is aimed at detecting the most influential and useful users in a network. These influential users defined as users who are more popular and more trusted among others. The problem of expert finding emerged many years ago to achieve reduced processing by selecting only influential users, achieve fast marketing query results, to address these users directly by 'targeted advertising' (so as to create public opinions or market awareness quickly and efficiently while spending much less on ineffective general advertising approach) and to improve the accuracy of the statistical results by avoiding the outliers and odd opinions contaminating the aggregated totals.

There are two approaches for finding the influential users in social networks. The first approach is analyzing the user's profile and the second one is user's link analysis. The links show the connection of users and the profiles shows their personal information such as age, city, gender, area of interests and etc. Most of the social networks include both link and profile information but with limitation to access them publicly.

Link analysis is one of the common methods to analyze the users connections and extract the needed information. Link analysis ranking (LAR) [1] is a method which ranks objects based on their links and the sign of links with each other. There are many studies which aim to rank the users considering their links [1]–[5]. However, to the best of our knowledge most of them focused on unsigned networks and there are few studies on signed networks with positive and negative links which are important to study the interactions in social media because the richness of a social network in most cases generally consists of a mixture of both positive and negative ones.

In signed networks, the link between users has positive or negative values. These signs present trust/distrust relation between users. The personality of each user has a direct impact on creating the signs of the links which affect the ranking calculations. In other words, the task of influential user detection by LAR method may greatly be affected by the personality of each user.

In this paper, considering the signs of the link, we first review most of the link ranking methods and then try to use the user's personality in ranking to find the most influential users. There are tremendous types of personalities in social science [6] which we use two main ones, namely optimism and pessimism (as user personality metrics) which can be calculated based on users propensity in relations (links) [7]. The optimism of a user shows how optimistic she thinks and in contrast, pessimism shows how pessimistic she thinks about the environment [8]. This personality feature is applicable to different ranking algorithms and we use these features in a sample ranking algorithm (PageRank) in order to verify the impact of them in ranking users and identifying the most influential ones. We call the new extended ranking algorithm as POPRank (Personality based on Optimist and Pessimist as a new feature for ranking algorithm). In order to evaluate the rankings, we used credibility criterion which relies upon the fact that better rankings should have more credibility values. The results showed that the added personality feature can effectively improve the ranking scores and has a meaningful impact on detecting the influential users.

The main contribution of this study can be summarized as follows:

- A comprehensive overview on different ranking algorithms based on link analysis.
- Adding personality as a new feature to rank users which can be appended to any ranking algorithm in order to provide better rankings for networks with positive and negative (signed) links.
- Introduce a new property of users namely credibility as a new measure to analyze the performance of ranking.

The remaining of the paper is organized as follows. In section II we overview the existing methods used for expert detection. The proposed methodology is presented in section III. Section IV presents the evaluation of our work. Finally, section V concludes the paper and presents some ideas as future lines of this study.

II. Related Works

this section provides a comprehensive literature review of related studies on expert detection and link analysis. In general, the researches for the problem of expert finding can be categorized into two main groups, (i) Authority ranking approaches [9]–[16] and (ii) non-authority ranking approaches [17]–[24].

The authority ranking approaches are based on link analysis for finding the influential users. These techniques which are based on web page rankings, evaluate the connections and relationships between users of a network and it is used in a situation which there is no access to the profile of users. For example, Kardan et al. in [25] find the experts to solve the problem regarding whose knowledge in social networks should be shared, which is based on PageRank method. Later in [26] they extend the experts detection in online communities. S. Chen et al. proposed an integrated PageRank method in [9] for the maximization problem to select the seeds in signed networks. Jurczyk et al. in [15] discover the users authorities in question answer communities by adjusting the HITS method [3]. X. Kong et al. in [10] tried to calculate the authors impacts in the author-paper network by a new algorithm based on PageRank scores. H. Zhu et al. in [11] proposed an expert finding framework using Topical Random Surfer (TRS) which is originally used for web page ranking. Bouguessa et al. [14] identify the experts in question-answering forums by ranking the users regarding the validity of their answers.

On the other hand, the non-authority ranking approaches are based on information retrieval from the activity and profile of the users. This class of ranking methods aims to find the experts using the information included in the profile of the users as well as analyzing their activities and posts in a given social network. For example, H. Deng et al. in [20] tried to develop weighted language, topical and hybrid model based on them for expert finding in DBLP bibliography and Google scholar dataset. Chen et al. [21] proposed a model for expert detection using the user activity analysis in rating the comments in question answering systems. D. mimno et al. [23] created a user model in order to determine the expertise level of reviewers based on papers. Also, J. Went et al. in [24] tried to find the influential users in Tweeter using the topical similarity among users.

In addition, there are methods which utilizing both link and profile information to increase the accuracy of detection task. For instance, J. Zhang et al. in [19] proposed a propagation-based approach that takes into consideration of both person local information and network information (e.g. relationships between persons). Z. Zhao et al in [17] declared that some parts of the available information in question answering systems are missing and then they find the experts using matrix completion technique and users similarity to fill the gap. Balog et al. [27] introduced a probability model for expert identification based on users topical profile in multilingual systems. Guo et al. [22] presented a method to find the best related user regarding a specific question by constructing the users profile by discovering latent topics and interests of users. Lu et al. in [28] used the question sessions and user profiles to build the network graph. Then using this graph, they proposed two expert detection method based on semantic propagation and semantic language model. Shahriari et al. in [18] proposed a new method to identify the experts using overlapping community detection.

The presented personality feature that is used in ranking algorithms falls in the first category (link analysis). The proposed LAR methods use the links between users to rank them. However, each user has her own characteristic of making links that affects link analysis. Hence, this feature has direct impact on the ranking which is not considered. In this manuscript, we will take to the account the impact of users personality (based on their opinions) and try to rank them in order to identify the most influential ones among them.

III. Signed network ranking algorithms

In social networks, there are two main approaches regarding the influential users detection problem as mentioned above. We introduce two measures as the personality of each user that can be added to any ranking algorithm in order to improve the performance of the ranking. First, we review most of the existing algorithms of ranking users including their shortcoming and then we will try to apply and utilize the personality measures on them. To this end, we take into account the sign of the links in signed networks and used Optimist and Pessimist scores of each user as their personality.

A. Existing Ranking Algorithms

According to the link analysis in social network, we first describe the baselines approaches and algorithms for the link analysis which rank the users in order to find the influential ones and then we will describe the proposed method that can be applied to ranking algorithm and effectively rank the users in order to identify the most influential ones.

1) In degree: The most common and simple way to find the influential users is verifying the number of coming links in a particular domain network and label the users with the most in-degree as expert [3]. More positive links in a trust\distrust relations mean more expertise a user has on that network. This method is used when there is only the information about the connection between users. However, this method is not very accurate because it only considers the positive in-links without considering the users who made the links.

2) Popularity or Prestige: This method is based on positive and negative links received by a user [29]. The main idea of Prestige is that the users who have received plenty of positive links should be ranked high and the ones who have received many negative links should be ranked low.

$$popularity_i = \frac{|IN_i^{(+)}| - |IN_i^{(-)}|}{|IN_i^{(+)}| + |IN_i^{(-)}|} \quad (1)$$

where $IN_i^{(+)}$ and $IN_i^{(-)}$ are positive and negative links received by user i respectively. Considering the signs of the links is a positive point of this method, yet it lacks utilizing the personality of the users who making the links in order to define a weight for links which indicate the importance of user's votes toward others.

3) Exponential ranking: In this probabilistic algorithm, the negative links are taken into consideration [30]. The idea behind this ranking algorithm is to decrease the rank of the users if they receive negative links. Also, it relies on that the user links should not be distrusted if she has a negative reputation and in fact, they just need to be trusted less. Particularly, the users with negative reputation should not be assumed completely trust-less (as if she point negative to another user, we assume it as positive) instead, her judgment should be considered less. The expected reputation is calculated as $a = A^T P$ where a is a pillar vector, A is adjacency matrix, P is a positive definite pillar probability vector with $|P|_1 = 1$ which is calculated recursively as follows:

$$P(t+1) = \frac{exp(\frac{1}{\mu} A^T P(t))}{|exp(\frac{1}{\mu} A^T P(t))|} \quad (2)$$

where μ specifies the amount of noise in selecting the highest reputable judge. This algorithm emphasizes the importance of negative links and the fact that enemy of a user enemy should not be considered as a friend. Indeed, their assumption which is based on social balance theory does not consider the importance of both positive and negative labeled users who make the links.

4) HITS: The HITS algorithm mainly relies on the fact that the way the links go has more information than just shared content [3]. This algorithm has two update rules namely authority and hub to rank the web pages. It assumes that each user has its own hub and authority value. Hubs are users which links to other users and authorities receive incoming links. First, an initial weight is assigned to hub and authority. Then in a specified repetitive iteration, the authority and hub will be updated until they converge as follows:

$$hub(i) = \sum_{j \in E_{ji}} authority(j) \quad (3)$$

$$authority(i) = \sum_{j \in E_{ij}} hub(j) \quad (4)$$

At the end of each iteration, weights are normalized under a norm such as In-degree, Salsa, Max-norm and etc. However, if a page makes several links to many good authorities, the hub score of it will be enhanced (so it will be ranked high).

5) Bias and Deserve: In this algorithm which is similar to HITS, bias of a user is its tendency to trust/distrust other users and deserve of a user reflects the true trust a user deserves [31]. A user is biased if her tendency of making trust/distrust connection to other users is high. The algorithm can work for both signed and unsigned networks. The update rules of the Deserve and Bias are as follows respectively:

$$Deserve_i(t+1) = \frac{1}{|d^{in}(i)|} \sum_{k \in d^{in}(i)} [w_{ki}(1 - X_{ki}(t))] \quad (5)$$

$$Bias_i(t+1) = \frac{1}{2|d^{out}(i)|} \sum_{k \in d^{out}(i)} [w_{ki} - Deserve_k(k)] \quad (6)$$

where $d^{in}(i)$ is the set of all receiving links by user i and $d^{out}(i)$ is the set of all outgoing links from user i, w_{ki} is the trust score from user k to user i (the weight of the links between users which is 1 for positive links and -1 for negative links). $X_{ki}(t)$ represents the effect of bias of user k on its outgoing link to user i at time t and is computed as: $X_{ki}(t) = max\{0, Bias_k \times w_{ki}\}$. This method suffers the same problem as HITS that is a user can show herself trustful if she rate users that deserve high positive values negatively and users that deserve high negative values positively which make her bias almost zero (trusted user).

6) PageRank: The PageRank algorithm performs a random walk in a given network to rank the nodes based on their connections [2]. The PageRank algorithm was proposed in order to rank the web pages regarding their hyperlinks to each other. Consider we have $P_i, P_2, ..., P_N$ pages that should be ranked. The update rule of the algorithm is as follows:

$$PR(P_i) = \alpha \sum_{P_j \in M(P_i)} \frac{PR(P_j)}{L(P_j)} + (1-\alpha)\frac{1}{N} \tag{7}$$

Where $M(P_i)$ is the set of pages that link to P_i, $L(P_j)$ is the number of out going links from page P_j, N is the total number of pages and α is a damping factor. α is added as a coefficient to the formula to guarantee that the algorithm does not accidentally end up with an infinite series of PageRanks. For implementation, an initial ranking will perform to the nodes and then they will be updated until convergence. The original PageRank algorithm does not consider the negative links and in fact, it is created for unsigned networks ranking. Also, nature (personality) of the nodes are not considered which can effectively change the ranking scores.

7) PageTrust: PageTrust is an extension of PageRank algorithm which considers both positive and negative links. The idea behind this algorithm is to decrease the random walk encounters to the pages which have negative incoming links [32].

$$\begin{aligned} PageTrust_i(t+1) &= (1 - Z_{ii}(t)) \\ &.[\alpha \sum_{j,(j.i)\in G^+} \frac{PageTrust_j(t)}{|d_j^{(+)}|} + (1-\alpha)\frac{1}{N}] \end{aligned} \tag{8}$$

where α is damping factor as PageRank, G^+ is sub-graph of positive links, $d_j^{(+)}$ is outgoing links in positive sub-graph from node j and Z is a matrix which is calculated as $Z(t+1) = T(t)P(t)$, where T is the transition matrix at time t which is calculated as the row-normalized version of the sub-graph with positive links. P is the distrust matrix that considering the negative links is calculated iteratively as follows:

$$P_{ij}(t+1) = \begin{cases} 1 & if(i \neq j; (i,j) \in G^- \\ 0 & if(i = j; (i,j) \in G^- \\ Z_{ij}(t+1) & otherwise \end{cases} \tag{9}$$

where G^- is sub-graph with negative links. The algorithm is promised to improve the PageRank accuracy by enabling it for both signed and unsigned networks yet it sustains the problem of PageRank to involve the personality of users.

8) Distance Algorithm: This simple algorithm ranks the web pages based on their shortest logarithmic distance from each other [33]. The distance algorithm between two pages i and j is $Distance_{ij} = -log \prod_{S \in path(i,j)} \frac{1}{O(S)}$ where $O(S)$ refers to out degree of user S. Then the ranking score of page j is equal to $Rank_j = \frac{\sum_{i=1} Distance_{ij}}{N}$. This algorithm is as simple as in-degree which ranks the web pages based on their distance (number of edges between them). It can be used to rank the users based on their distance as well, yet it suffers the same problems as in-degree method.

9) Ontology Ranking Algorithms: This is the other branches of ranking algorithm which is usually used in semantic web and tries to decrease the amount of overloaded data [34]. The main idea behind this algorithm is providing relevant information regarding a user query and rank the related information as high as possible so the searcher can easily access it. The problem of these algorithms is satisfactory of the users which are not guaranteed.

Considering all of the mentioned ranking algorithms, we noticed that the PageRank is the most common algorithm for ranking the users and observed that many existing ranking methods used this algorithm as their baseline for comparison. Hence, in this study we consider PageRank as the base ranking algorithm and add the personality feature to it to verify its effect. As long as the Optimist and Pessimist score of each user is defined based on their in and out links, we consider the Prestige algorithm as another evaluation ranking algorithm which uses the in-links for the ranking calculation.

B. Personality as a ranking feature

We add personality as a new feature to PageRank algorithm and propose a new ranking namely POPRank, in order to see how much this feature can improve the ranking. The PageRank has been originally proposed for networks with only positive links which is unable to be used directly for signed networks. We modified it to perform better and also can be used for signed networks. The added personality is consist of two social science features, namely Optimism and Pessimism which are added to the algorithm in order to improve the ranking accuracy.

1) Optimism and Pessimism: Optimist users are those who think positive about everything around them and make more positive (trust) links to other users. This personality makes the other users establish positive links to her as well. Therefore, an optimist user usually has both trust links to others and trusted links from others. In contrast, pessimist users are those who think negative about their environment and make more negative (distrust) links. We say that a pessimist user usually has both distrust links to others and distrusted links from others. We try to calculate the optimism and pessimism scores of the users from their rates (votes) toward external items (e.g. Epinion data set). The optimist and pessimist scores of the users are defined and calculated in [35] which are used to rank the users in the sign prediction problem. We will use this definition and add them as a feature in the ranking algorithms (in the case of this paper to the PageRank) to verify its impact on link ranking methods and influential user detection. The optimist and pessimist scores are defined as follows:

Consider there are N items $I_1, I_2, ..., I_N$, the set of items with low average rating scores rated by user u_i are:

$$OptLow_i = \left\{ I_k | r_{ik} \neq 0 \wedge \bar{r}_k \leq \frac{(1+z)}{2} \right\} \tag{10}$$

where r_{ik} indicates the rating score from user u_i to item I_k and $\bar{r}_k$ denotes the users average rating score toward I_k. If the rates are in the range of $[1, z]$, we consider scores in $[1, (1+z)/2]$ as low and $[((1+z)/2)+1, z]$ as high scores. The set of items which have low average scores and are scored high by user u_i are as follows:

$$OptHigh_i = \left\{ I_k | I_k \in OptLow_i \wedge r_{ik} > \frac{(1+z)}{2} \right\} \quad (11)$$

Likewise, the set of items with high average rating scores rated by user u_i are:

$$PessHigh_i = \left\{ I_k | r_{ik} \neq 0 \wedge \bar{r}_k > \frac{(1+z)}{2} \right\} \quad (12)$$

And the set of items which have high average scores and are scored low by user u_i are as follows:

$$PessLow_i = \left\{ I_k | I_k \in PessHigh_i \wedge r_{ik} \leq \frac{(1+z)}{2} \right\} \quad (13)$$

If the user u_i has rated above the average then she is more optimistic. Hence, the optimism score of user u_i is $Optimism_i = \frac{|OptHigh_i|}{|OptLow_i|}$. Accordingly, the pessimism score of user u_i is $Pessimism_i = \frac{|PessLow_i|}{|PessHigh_i|}$. These two quantities will be used as a coefficient in ranking algorithms, therefore they will be normalized to the range of [0,1] in order to adjust the values and prevent diverge.

2) POPRank: The original PageRank algorithm is a vote by all the other pages to show how important a page is (a link to a page counts as a vote). In fact, it does not consider the users who make the connections. Using this algorithm, we consider each page as a user and take into account the validity of users who make a connection with a specific user. In other words, to calculate the rank score of a user, we consider the coming links (same as PageRank) and the personality of users making them in POPRank algorithm. As mentioned above, optimism and pessimism are two quantities that provide us the possibility to measure the personality. The idea of using personality is that when an optimist user makes a positive link, her vote should be considered less (we will decrease her vote impact) and in contrast when she makes a negative link, her vote should be considered more (we will increase her vote impact). A similar theory is used for pessimist user, meaning that, her negative votes will be decreased and her positive ones will be increased. In this ranking, we will apply PageRank separately on sub-graph with positive links G^+ and sub-graph with negative links G^-. The update rules of POPRank are as follows:

$$POPRank^+(P_i) = (1-\alpha)\frac{1}{N} + \alpha \sum_{P_j \in M(P_i)} \frac{PR^+(P_j)}{L^+(P_j)} \times Per_j \quad (14)$$

$$POPRank^-(P_i) = (1-\alpha)\frac{1}{N} + \alpha \sum_{P_j \in M(P_i)} \frac{PR^-(P_j)}{L^-(P_j)} \times Per_j \quad (15)$$

where $L^+(P_j)$ and $L^-(P_j)$ are the number of positive and negative outgoing links from node j, respectively. Similar to PageRank algorithm, it starts with some initial condition for both positive and negative PageRanks vectors and after enough iterations, it converges to the final rank vectors. In social science, a person can be optimist or pessimist. Taken this into account, we consider personality as follows:

$$Per_j = max\left\{Optimism_j, Pessimism_j\right\} \quad (16)$$

where the $Optimism_j$ and $Pessimism_j$ are the optimism and pessimism scores of the user u_j and are calculated as mentioned above. The final rank vector POPRank is calculated by:

$$POPRank(P_i) = POPRank^+(P_i) - POPRank^-(P_i) \quad (17)$$

The convergence of this algorithm is assured since it is the same as standard PageRank algorithm with the same computational complexity.

C. Credibility as a measure to analyze ranking

The previous studies [36] indicate that the trust can emerge among users with two main factors: the first one is familiarity and the second one is similarity. That is when the users know (familiarity) or resemble (similarity) each other, they trust each other more. Hence, these two measures can calculate the trust score toward the users in social networks which shows the credibility of them. According to this, a user's credibility is defined by familiarity and similarity. In this paper, we use similarity to calculate the credibility. W. hu et al. in [37], [38] used the similarity of neighbors to calculate the credibility and concluded that popular ranked users have more credibility. They also showed that users credibility of a network has a direct relation with its ranking so it can be used to compare the rankings. We will use the credibility of users as the evaluation criteria which can confirm and verify the ranking outcomes. The credibility indicates the votes of a user's neighbor towards her. In other words, the credibility of a user reflects her expected trust value in the network. The value of the credibility does not consider only the number of coming links instead, it depends on their quality. The credibility of user u_i is calculated as follows:

$$\begin{aligned} Credibility(u_i) = \frac{1}{|M^i(u_i)|} \sum_{u_p \in M^i(u_i)} W_{u_p u_i} \\ .Sim(u_p, u_i).Credibility(u_p) \end{aligned} \quad (18)$$

where $M^i(u_i)$ denotes the set of all incoming links to node u_i and $W_{u_p u_i}$ presents the link weight from user u_p to user u_i. There are several methods such as the correlation coefficient, the cosine similarity measure, and the euclidean distance that can be used to calculate the distance of two end points and return a quantitative value to represent the similarity between users. In trust network, a user's similarity depends on its neighbors [39], while user tends to trust similar users like her. According to this, in this context, we use the *Jaccard Distance* to model the similarity between u_p and u_i, which is $Sim(u_p, u_i) = \frac{|F_p \cap F_i|}{|F_p \cup F_i|}$. $F_p \cap F_i$ is the set of two users common neighbors and $F_p \cup F_i$ is the set of two users total neighbors. Note that in signed networks, the weights $W_{u_p u_i}$

TABLE I
THE MAIN CHARACTERISTIC OF THE EPINION DATA SET

Total number of users	131,828
Total trust ratings	841,372
Number of filtered users	49,289
Trust ratings	507,592
Positive trust ratings	434,694
Negative trust ratings	72,898
Number of Items	139,738
Number of Items' ratings	664,824

are -1 or +1 and the credibility value lies in the range of [-1,1] for such networks.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed algorithm by using a real-world signed network in the context of influential user detection using a ranking algorithm involving each nodes personality. We also used personality (optimist and pessimist) of each node which is application dependent meaning that the definition and calculation of it can be vary in different data sets. The evaluations have two parts. The aim of the first one is to show that algorithm works correctly. In other words, the aim of first evaluation is not showing that our algorithm always, or in most of the cases, produces better rankings when compared to the baselines. Instead, we demonstrate that our algorithm produces such rankings that are useful in the sense that they produce rankings that are distinct and competitive with the ones produced by baseline and high quality of link analysis. The second evaluation compares the performance of the proposed algorithm with baseline ones using credibility. As we discussed, credibility is a criterion which can verify and show the validity of rankings.

A. Dataset

To evaluate our work we used Epinions dataset gathered from Stanford Large Network dataset Collection (SNAP)[1]. The Epinions website is a general consumer review site and its dataset consists of two types of ratings, trust relation among users (members of the site can decide whether to trust each other or not) and users rating on items (the rate of users regarding the items of the website). This dataset includes 131,828 users and 841,372 trust ratings. The main characteristic of the dataset is presented in table I. In our evaluations, we did a filtering step and omitted who has no links and only considered 49,289 users with links with 507,592 trust ratings as links for the input of each algorithm (table I).

B. Evaluation

For the first part of the evaluation, we implemented the PageRank and Prestige algorithms to obtain our performance benchmark. We used 664,824 item rating by users in order to calculate the optimist and pessimist score of each user. In order to compare POPRank with PageRank and Prestige we

[1]https://snap.stanford.edu/data/

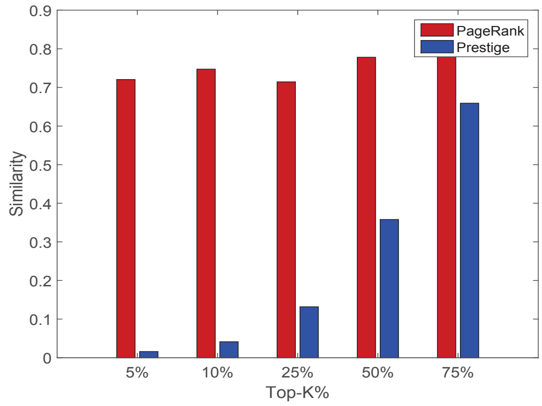

Fig. 1. Similarity of POPRank with each approach in found influential users. X axis represents different percentages of top found influential users and Y axis shows the similarity with POPRank

use Spearman's rank correlation which measures the similarity of two rankings:

$$Similarity = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (19)$$

Here, x and y are rankings by two algorithms and $\bar{x}$ and $\bar{y}$ are average ranks. We compare the effectiveness of our proposed rank algorithm with the benchmark algorithms. To compute the rank coefficient, a portion of the highest ranked nodes in the merged graph according to x are considered. As a default, we considered 10% highest ranked nodes but we also varied the target percentage (5%, 10%, 25%, 50% and 75%) to observe how the accuracy varies with result size. For damping factor, we used $\alpha = 0.85$ as a parameter of the ranking algorithm. Also, it is worth mentioning that the run time of the PageRank algorithm is O(m+n) in which n is the number of nodes and m is the number of edges. POPRank has similar complexity to calculate the users ranks. In particular, the proposed algorithm has more complexity to calculate the personality which is O(m+n+k) in which k is the number of item ratings.

Figure 1 compares POPRank with PageRank and Prestige in different percentages of top found influential users. To compare the algorithms we ignored the users which have no links to others and the ones whom item rating are not available because they have no impact on ranking algorithms. This Figure presents the percentage of common found influential users in different percentages of data between POPRank and other two algorithms. The result is shown in Table II. This, confirms that POPRank performance is near to PageRank but far from Prestige. The similarity of POPRank and PageRank is maintained with different percentages of data.

We expect that the common found users should be increased if we consider more and larger percentage of data. In the Figure 1 the similarity of POPRank and Prestige increased when we added more data. However, in top-25% the similarity of POPRank and PageRank decreased. This can happen if we are comparing the similarity in the beginning or middle of the x-axis of this figure because for each next step of comparison

TABLE II
COMMON FOUND INFLUENTIAL USERS WITH POPRANK

Top-K%	5%	10%	25%	50%	75%
PageRank	72.05	74.72	71.44	77.81	80.89
Prestige	1.60	04.14	13.18	35.79	65.91

(top-N%) the new found users could be different, but it can not happen at the end of the x-axis because the users who are added are same. This decrement, indicate that the users found by POPRank and PageRank are different in top-25% of the found users.

The other perception of this experiment was the difference of POPRank and Prestige. Prestige is based on coming positive and negative links and the personality is based on the user votes (links) to the items. Nevertheless, the similarity of these two concepts did not affect the POPRank ranking. Particularly, the personality of users involved in POPRank will not force it to be dependent only on the received links.

For the second experiment, we verify the ranking of nodes by all the three algorithms using the credibility values. The credibility of nodes is used as the criterion to evaluate and analyze the performance of algorithms. We say that nodes with more credibility should be ranked higher than those with less one. Taking it into account, we compare the top found nodes in different algorithms with nodes with more credibility. The evaluation was conducted with different percentages of top found nodes. We partitioned the result of each ranking algorithm in different percentages. For each percentage of found nodes, we sum up their credibility and compare it with different algorithms. Figure 2 shows the normalized credibility values of each algorithm for different percentages of top found influential users.

The Prestige algorithm is based on positive and negative links received by a user and PageRank is based on a random walk to rank the nodes based on their connections while POPRank considers the personality of each user as a added value to rank them. As it shown, the nodes that identified and ranked high by POPRank have more credibility for top-5% and top-10% in compare to the others. In contrast for top-25% and rest, the PageRank has better credibility. This shows that for more influential users (top-5% and top-10%) POPRank has better performance. Also, as we observed in Figure 1, the similarity of POPRank and PageRank decreased in top-25% so we expect a meaningful difference in the credibility of them. The credibility increment of the PageRank in top-25% is beheld in Figure 2 (as we expected), showing that there is a meaningful difference between the found users by these algorithms here. The POPRank algorithm found the most influential users based on credibility in top-25% of its ranking whereas PageRank and Prestige found it in top-50% and top-75% respectively. Overall, the POPRank algorithm has better performance in identifying the influential users within top-5% and top-10% and can find the most influential users within top-25% which is better than other algorithms. In other words, POPRank outperformed the baseline ranking algorithms such as PageRank and Prestige.

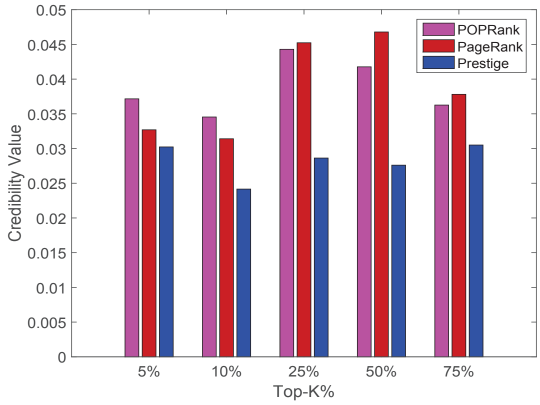

Fig. 2. Normalized Credibility of each ranking algorithm regarding different percentages of found influential users. X axis represents different percentages of top found users and Y axis shows the normalized credibility value of found users

TABLE III
COMPARISON OF PAGERANK AND POPRANK

Top-N	10	20	50	100	500
Common found	0	3	8	23	304
PageRank Credibility	0.017	0.024	0.028	0.031	0.035
POPRank Credibility	0.066	0.069	0.070	0.068	0.048

We noticed that the percentage of found influential users in POPRank have more credibility in comparison with PageRank and Prestige. This shows that leveraging the power of personality of each node can further improve the performance of expert finding. We also verified the Spearman correlation of more credible nodes with all algorithms. We found that the correlation for Prestige, Pagerank and POPRank are 5.73%, 16.09%, and 19.50% respectively. This again confirms that the users found by POPRank algorithm have more credibility in comparison with others. Furthermore, in Table III, we compared the performance of PageRank and POPRank in terms of the top ranked users.

In addition, although the similarity of them is high (72.05%) for top-5% of the ranked users, but it is different for the most top found ones. As we can see in this table, the number of commonly found users are quite low which indicates that POPRank makes a distinct ranking. To have a better understanding, Table III presents the normalized credibility of PageRank and POPRank as well. The comparison of the credibility values demonstrates that POPRank rankings always has higher credibility which indicates that it provides a better ranking. In the nutshell, the results show the positive impact of users personality in the rankings algorithm in order to find the influential users.

V. CONCLUSION AND FUTURE WORK

There are a huge amount of information and opinion on different topics and products shared in different social networks. However, one way to find the useful and trustful ones in relying on the influential users or experts whom can provide valuable shared knowledge. Toward this end, we first need to identify this set of influential users and to this end, the links between users and their profiles can be used. In this paper, we used two features from social science namely Optimism and Pessimism to add the personality of each user in the ranking algorithms. We applied the user's personality to PageRank algorithm and created a new ranking POPRank for signed networks. Next, we compared the influential users found by POPRank with two baseline approaches of ranking. The result demonstrates the efficiency of the proposed algorithm.

Two future directions can be taken into account: *(i) Trust propagation:* In the application of group decision making, when a problem occurs the group will discuss to find the solution. The fact is that the influential users are more trusted and has more effect on the final decision. We plan to investigate how a decision is made in a group by identifying the expert users related to the problem and propagating their information based on their trusted links. *(ii) Using profile information:* As a future guideline, we plan to use the profile of users in addition to their connections. In other words, the link analysis can identify the expert users but we plan to investigate the effect of each user profile on the accuracy of expert detection.

VI. Acknowledgment

This work is partially supported by ITEA3 CAP and the European Celtic-Plus project CONVINcE. We would like thank Dr. Rebecca Copeland for the constructive discussions.

References

[1] A. Borodin, G. O. Roberts, J. S. Rosenthal, and P. Tsaparas, "Link analysis ranking: algorithms, theory, and experiments," *ACM Transactions on Internet Technology (TOIT)*, vol. 5, no. 1, pp. 231–297, 2005.

[2] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." 1999.

[3] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[4] A. Y. Ng, A. X. Zheng, and M. I. Jordan, "Stable algorithms for link analysis," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2001, pp. 258–266.

[5] R. Lempel and S. Moran, "The stochastic approach for link-structure analysis (salsa) and the tkc effect," *Computer Networks*, vol. 33, no. 1, pp. 387–401, 2000.

[6] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, 1992.

[7] M. F. Scheier, C. S. Carver, and M. W. Bridges, "Optimism, pessimism, and psychological well-being." 2001.

[8] W. N. Dember, S. H. Martin, M. K. Hummer, S. R. Howe, and R. S. Melton, "The measurement of optimism and pessimism," *Current Psychology*, vol. 8, no. 2, pp. 102–119, 1989.

[9] S. Chen and K. He, "Influence maximization on signed social networks with integrated pagerank," in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. IEEE, 2015.

[10] X. Kong, J. Zhou, J. Zhang, W. Wang, and F. Xia, "Taprank: A time-aware author ranking method in heterogenous networks."

[11] H. Zhu, E. Chen, H. Xiong, H. Cao, and J. Tian, "Ranking user authority with relevant knowledge categories for expert finding," *World Wide Web*, vol. 17, no. 5, pp. 1081–1107, 2014.

[12] A. El-Korany, "Integrated expert recommendation model for online communities," *arXiv preprint arXiv:1311.3394*, 2013.

[13] J. Liu, Y.-I. Song, and C.-Y. Lin, "Competition-based user expertise score estimation," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 425–434.

[14] M. Bouguessa, B. Dumoulin, and S. Wang, "Identifying authoritative actors in question-answering forums: the case of yahoo! answers," in *Proceedings of the 14th ACM SIGKDD*. ACM, 2008.

[15] P. Jurczyk and E. Agichtein, "Discovering authorities in question answer communities by using link analysis," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 2007, pp. 919–922.

[16] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007.

[17] Z. Zhao, L. Zhang, X. He, and W. Ng, "Expert finding for question answering via graph regularized matrix completion," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 27, no. 4, 2015.

[18] M. Shahriari, S. Parekodi, and R. Klamma, "Community-aware ranking algorithms for expert identification in question-answer forums," in *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 2015, p. 8.

[19] J. Zhang, J. Tang, and J. Li, "Expert finding in a social network," in *Advances in Databases: Concepts, Systems and Applications*. Springer, 2007, pp. 1066–1069.

[20] H. Deng, I. King, and M. R. Lyu, "Formal models for expert finding on dblp bibliography data," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 163–172.

[21] B.-C. Chen, J. Guo, B. Tseng, and J. Yang, "User reputation in a comment rating environment," in *Proceedings of the 17th ACM SIGKDD*. ACM, 2011.

[22] J. Guo, S. Xu, S. Bao, and Y. Yu, "Tapping on the potential of q&a community by recommending answer providers," in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 921–930.

[23] D. Mimno and A. McCallum, "Expertise modeling for matching papers with reviewers," in *Proceedings of the 13th ACM SIGKDD*, 2007.

[24] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010.

[25] A. Kardan, A. Omidvar, and F. Farahmandnia, "Expert finding on social network with link analysis approach," in *Electrical Engineering (ICEE), 2011 19th Iranian Conference on*. IEEE, 2011.

[26] M. Rafiei and A. A. Kardan, "A novel method for expert finding in online communities based on concept map and pagerank," *Human-centric computing and information sciences*, vol. 5, no. 1, 2015.

[27] K. Balog, T. Bogers, L. Azzopardi, M. De Rijke, and A. Van Den Bosch, "Broad expertise retrieval in sparse data environments," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2007, pp. 551–558.

[28] Y. Lu, X. Quan, J. Lei, X. Ni, W. Liu, and Y. Xu, "Semantic link analysis for finding answer experts," *Journal of Information Science and Engineering*, vol. 28, no. 1, pp. 51–65, 2012.

[29] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.

[30] V. A. Traag, Y. E. Nesterov, and P. Van Dooren, *Exponential Ranking: taking into account negative links*. Springer, 2010.

[31] A. Mishra and A. Bhattacharya, "Finding the bias and prestige of nodes in networks based on trust scores," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011.

[32] C. De Kerchove and P. Van Dooren, "The pagetrust algorithm: How to rank web pages when negative links are allowed?" in *SDM*. SIAM, 2008, pp. 346–352.

[33] A. M. Z. Bidoki and N. Yazdani, "Distancerank: An intelligent ranking algorithm for web pages," *Information Processing & Management*, vol. 44, no. 2, pp. 877–892, 2008.

[34] A. S. Butt, A. Haller, and L. Xie, "Ontology search: An empirical evaluation," in *The Semantic Web–ISWC 2014*. Springer, 2014, pp. 130–147.

[35] G. Beigi, J. Tang, and H. Liu, "Signed link analysis in social media networks," *In the 10th International AAAI Conference on Web and Social Media (ICWSM-16)*, 2016.

[36] H. Mei, Y. Zhang, and X. Meng, "Csa: A credibility search algorithm based on different query in unstructured peer-to-peer networks," *Mathematical Problems in Engineering*, vol. 2014, 2014.

[37] W. Hu and Z. Gong, "Assessing the credibility of nodes on multiple-relational social networks," in *Web Information Systems Engineering–WISE 2014*. Springer, 2014, pp. 62–77.

[38] Hu, Weishu and Gong, Zhiguo, "Multi-relational reinforcement for computing credibility of nodes," *World Wide Web*, pp. 1–22, 2015.

[39] G. Jeh and J. Widom, "Simrank: a measure of structural-context similarity," in *Proceedings of the eighth ACM SIGKDD*. ACM, 2002.