AI Communications 00 (2008) 1–17
DOI 10.3233/AIC-2008-0431
IOS Press

Key figure impact in trust-enhanced recommender systems

Patricia Victor [a,*], Chris Cornelis [a], Martine De Cock [a] and Ankur M. Teredesai [b]
[a] *Applied Math & CS, UGent, 9000 Gent, Belgium*
E-mails: {Patricia.Victor, Chris.Cornelis, Martine.DeCock}@UGent.be
[b] *Institute of Technology, UW Tacoma, Tacoma, WA, USA*
E-mail: ankurt@u.washington.edu

Abstract. Collaborative filtering recommender systems are typically unable to generate adequate recommendations for newcomers. Empirical evidence suggests that the incorporation of a trust network among the users of a recommender system can significantly help to alleviate this problem. Hence, users are highly encouraged to connect to other users to expand the trust network, but choosing whom to connect to is often a difficult task. Given the impact this choice has on the delivered recommendations, it is critical to guide newcomers through this early stage connection process. In this paper, we identify several classes of key figures in the trust network, namely mavens, frequent raters and connectors. Furthermore, we introduce measures to assess the influence of these users on the amount and the quality of the recommendations delivered by a trust-enhanced collaborative filtering recommender system. Experiments on a dataset from Epinions.com support the claim that generated recommendations for new users are more beneficial if they connect to an identified key figure compared to a random user.

Keywords: Trust network, recommender system, cold start problem, social network analysis

1. Introduction

Systems that guide users through the vast amounts of online information are gaining tremendous importance. Among such applications are recommender systems (RSs) [1,33], which, given some information about their users' profiles and relationships, suggest items that might be of interest to them. One of the most widely used recommendation techniques is collaborative filtering (CF) [32], which typically works by identifying users whose tastes are similar to those of the particular user and by recommending items that they have liked. However, CF recommender systems still face important challenges, one of their main weaknesses being the *cold start problem*: new users have not rated a significant number of items, and cannot properly be linked with similar users;[1] hence, accurate and adequately personalized recommendations are difficult to generate.

The cold start (CS) problem receives a lot of attention from the RS community, see e.g. [2,23,30] for some recent work. One of the promising directions suggests that the incorporation of a *trust network* (in which the agents are connected by trust scores indicating how much they trust and/or distrust each other) can significantly help alleviate the CS problem, primarily because the information included in trust statements about a RS's user can be propagated and aggregated, and hence more people and products can be matched [23,42,45].

Since the trust information in such a *trust-enhanced RS* has a significant direct influence on the delivered recommendations (both amount and quality), it is beneficial for users to connect to other users in the trust network as soon as possible (see e.g. [11,23]). This is however not straightforward for CS users because they are new to the system and they often do not know which users will have the best impact on the generated recommendations. As research has shown that interactivity and transparency are two key factors to a better understanding and acceptance of RSs (see e.g. [15, 37]), it is worthwhile to guide newcomers through this

*Corresponding author: Patricia Victor, Applied Math & CS, UGent, Krijgslaan 281 (S9), 9000 Gent, Belgium. E-mail: Patricia.Victor@UGent.be.

[1] Note that the phrase *cold start* has also been used to describe the situation where recommendations are required for items that have never been rated (see e.g. [36]). In this paper however, *cold start* refers to the situation where recommendations are required for users that have rated only very few items, the so-called *cold start users*.

connection process by providing suggestions and by explaining the effect of making trust connections.

In this paper, we identify different user classes in the RS network as mavens (knowledgeable users who write a lot of reviews), connectors (with a lot of connections in the trust network), and frequent raters (who rate a lot of reviews). We claim that it is more beneficial for new users to connect to one of these *key figures* as opposed to connecting to a random user. Verifying this claim involves investigating both the quality (accuracy) as well as the amount (coverage) of the delivered recommendations. This accuracy-coverage trade-off is comparable to the precision-recall trade-off in information retrieval. We deal with the problem on a local level within the trust network. The main questions to be answered are:

1. If a cold start user a has a user b in his web of trust, how does this affect the quality and the amount of the recommendations generated for a?
2. Based on this, how can we quantify the accuracy and the coverage impact of user b for cold start user a?
3. What can we conclude about the impact of a particular key figure b for the cold start users in a trust-enhanced recommender system in general?

As shorter propagation chains lead to more accurate predictions, we propose to measure b's impact on the accuracy for a based on how often b is on a shortest path from a to an item. To this end, we use a modification of the well-known betweenness measure from social network analysis (SNA) [41]. Furthermore, user b is vital for a when b rates a lot of items and these items are only rated by b. Omitting such a high impact user from the web of trust results in a fragmented network with many items appearing in isolated fragments that are not accessible anymore from a; hence we propose a modification of an existing fragmentation measure to assess the impact of b on the coverage for a.

In Section 2, we describe classical CF RSs and explain how trust-enhanced RSs can help alleviate the CS problem. To benefit from these trust algorithms, a new user needs to know which users are best to connect to. In Section 3, we identify different classes of key figures: mavens, connectors, and frequent raters. To investigate the influence of these key figures on the generated recommendations, in Section 4 we introduce new measures that are based on the concepts of betweenness and fragmentation. In Section 5, we show by a number of experiments that it is more beneficial for new users to connect to key figures rather than making random connections. To evaluate the techniques we propose in this paper, we use a large dataset from Epinions,[2] a prominent e-commerce site that gives users the opportunity to include other users (based on their quality as reviewers of all kinds of consumer goods) in their own web of trust (WOT). The results can be generalised to other trust-based RSs. We conclude the paper with a discussion of future research directions.

In [40], we reflected on a first effort of measuring the impact of key figures in the Epinions trust network. Although the dataset and the aim is the same, the results in this paper are substantially different from those in [40] because we use different quality and coverage measures that have a clear foundation in social network analysis.

2. Related work

Recommender systems [1,33] are often used to accurately estimate the degree to which a particular user will like a particular item. Such algorithms come in many flavours, such as content-based, collaborative filtering and trust-based methods; the latter two being the ones most relevant to our current efforts.

2.1. Classical CF RSs

Content-based systems suggest items similar to the ones that the user previously liked. They tend to have their recommendation scope limited to the immediate neighbourhood of the users' past purchase or rating record. For instance, if a customer of a video rental store has only ordered romantic movies, the system will continue to recommend related items only, and not explore other interests of the user. In this sense, RSs can be improved significantly by (additionally) using collaborative filtering [32], which typically identifies users whose tastes are similar to those of the given user and recommends items that they have liked.

CF algorithms produce a rating for an item i that is new to a user a. This new rating is based on a combination of the ratings of the nearest neighbours (similar users) already familiar with item i. The classical CF-formula is given by (1). The unknown rating $p_{a,i}$ for an item i for a user a is predicted based on the mean $\overline{r}_a$ of ratings by a for other items, as well as on the ratings $r_{u,i}$ by other users u for i. The formula also takes into account the similarity $w_{a,u}$ between users a and u,

[2] www.epinions.com.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 3

usually calculated as Pearson's Correlation Coefficient (PCC) [19]:

$$p_{a,i} = \overline{r}_a + \frac{\sum_{u=1}^{k} w_{a,u}(r_{u,i} - \overline{r}_u)}{\sum_{u=1}^{k} w_{a,u}}. \qquad (1)$$

Throughout this paper, the *rating coverage* for a user a, or coverage for short, refers to the ratio of the amount of items for which $p_{a,i}$ as in (1) can be calculated versus the total amount of items available in the RS. For any RS algorithm, an increase in coverage is only beneficial when the accuracy does not drop significantly, while an accuracy increase is not useful when there are too few ratings that can be predicted. Hence, coverage and accuracy results should be evaluated together.

The effectiveness (accuracy and coverage) of CF based RSs is significantly affected by the number of ratings available for each user: the more ratings are available, the better the quality of the recommendations (see e.g. [35]). Moreover, generating recommendations is only possible for users who have rated at least two items because the PCC requires at least two ratings per user. An important problem arises with cold start users: being new users, they have rarely rated a significant number of items, and since they usually constitute a sizeable portion of the RS's user community (see e.g. [24]), it is very important to address this problem. Consequently, it has received considerable attention from the RS community in the last years.

Most of the approaches combine rating data with content data to alleviate the CS problem, such as Middleton et al. [25] who work with information delivered by ontologies, and Park et al. [30] who focus on simple filterbots. Ahn [2] and Huang et al. [21] only use rating data: the former introduces a similarity measure which takes into account the proximity of the ratings, the rating impact and item popularity, while in the latter approach the set of CF neighbours is extended by exploring transitive associations between the items and users.

2.2. *Trust-enhanced RSs*

Trust-enhanced RSs can alleviate the CS problem by using additional information coming from a trust network in which the users are connected by trust scores indicating how much they trust and/or distrust each other. Such trust networks can be generated automatically, e.g. inferred through the similarity of rating behaviour [29,31,42] or based on a user's history of making reliable recommendations [27]. Another approach is to ask the RS's users explicitely to issue trust statements about other users [5,12,16,22,45]. A nice example is Golbeck's FilmTrust [12], an online social network combined with a movie rating and review system in which users are asked to evaluate their acquaintances on a scale from 1 to 10. The movie recommender system uses the weighted mean of the item ratings from a selected set of users; the weights represent the trust that the target user has in the selected users. FilmTrust is a non commercial venture, but trust-based systems are also being used in e-commerce applications.

A well-known trust-enhanced example is Epinions.com, an e-commerce site where users can rate products and include users in their personal web of trust. In [23,24], Massa et al. investigate how trust can be incorporated into the CF process by conducting experiments on a dataset from Epinions. They propose a special case of (1) in which the weights $w_{a,u}$ are replaced by trust information $t_{a,u}$ [23]. The formula is given by (2). In this approach, trust is interpreted as a numerical value which ranges between 0 and 1, denoting absence and full presence of trust, respectively.

$$p_{a,i} = \overline{r}_a + \frac{\sum_{u=1}^{k} t_{a,u}(r_{u,i} - \overline{r}_u)}{\sum_{u=1}^{k} t_{a,u}}. \qquad (2)$$

The main strength of trust-enhanced recommender systems is their use of *trust propagation operators*; mechanisms to estimate the trust transitively by computing how much trust an agent a has in another agent c, given the value of trust for a trusted third party (TTP) b by a and c by b. Although there is much debate about the most suitable propagation operator(s), see e.g. [11,14,18,23,34,44], all of them agree on the case of atomic direct propagation, namely that if a trusts b and b trusts c, then it is inferred that a trusts c.

Golbeck's TidalTrust [11] and Massa's MoleTrust [22] are specifically designed for propagation of trust only. They both choose multiplication as propagation operator and take into account a maximum propagation depth and a minimum trust value below which users are not allowed to interfere in the recommendation process, but the ways these two thresholds are determined differ significantly.

Another, very recent, research path is the propagation of trust and distrust, which obviously requires new propagation operators. For a short discussion we refer to Section 6.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 4

By combining (propagated) trust information with the available ratings, more users and (consequently) more items get covered by the RS, even if only few trust statements per user are available [23]. In particular, a prediction $p_{a,i}$ can be calculated when a trusts at least one user u to a degree $t_{a,u} \neq 0$ and u already rated i. It is demonstrated in [23] that the coverage for CS users increases significantly when they connect to the trust network.

As shown in e.g. [13,23], the trade-off between accuracy and coverage turns out to be advantageous for trust-enhanced RSs, and especially for CS users. Golbeck [13] and Massa et al. [23] report that using only the information coming from trusted acquaintances, and from users who are trusted by trusted people in turn, makes the recommendations significantly more accurate and also more personalized. Hence, it is beneficial for a new user to connect to the trust network as soon as possible. But, as demonstrated in the following section, it is often the case that CS users in the classical sense (people who provided only a few product ratings) are also CS users in the trust sense, meaning that they issued only a few, or no trust statements at all. Therefore, we propose to guide the new users during the connection process by suggesting to connect to key figures who have a positive impact on the coverage while maintaining sufficient accuracy.

To our knowledge, no studies have been conducted on the influence of key figures in a trust-based RS. However, there exists some work on the impact of CF users, in particular studies about identifying users who influence the buying behaviour of other users and hence boost the sales of particular items [3,9]. These approaches differ from ours, as they do not specifically measure the impact on the coverage and accuracy, and do not focus on cold start users. Furthermore, we use characteristics of trust-based CF networks to define the key figures.

3. Users in the Epinions dataset

Epinions.com is a popular e-commerce site where users can write reviews about products and assign a rating to them. Guha et al. [14] compiled a dataset containing 1,560,144 reviews (written by 326,983 users) that received 25,346,057 ratings by 163,634 different users. Reviews are evaluated by assigning a helpfulness rating which ranges from 'not helpful' ($1/5$) to 'most helpful' ($5/5$). Note that we do not have information about consumer products and product ratings, but work with reviews and review ratings instead; in other words, we evaluate a 'review recommender system'. Hence, in this context, an item denotes a review of consumer goods.

3.1. Cold start users

We focus on users who have evaluated at least one review. In this group, 59,767 users rated only one review, 20,159 only two, 11,216 exactly three and 7322 exactly four. These cold start users constitute about 60% of all review raters in the Epinions community. The relative numbers of users are given in Table 1 where the cold start users are denoted by CS1 (exactly one review), CS2 (two reviews), CS3 and CS4.

Besides evaluating reviews, users can also evaluate other users based on their quality as a reviewer. This can be done by including them in their WOT (i.e., a list of reviewers whose reviews and ratings were consistently found to be valuable[3]) or by putting them in their block list (i.e. a list of authors whose reviews were consistently found to be offensive, inaccurate or low quality,[3] thus indicating distrust). These evaluations make up the Epinions WOT graph consisting of 131,829 users and 840,799 non self-referring trust or distrust relations (see also [14]). About 85% of the statements are labelled as trust, which is reflected in the average number of users in a WOT (5.44) and in a block list (0.94). Due to the large portion of trust statements, we focus on trust information only in the remainder of the paper.

The trust graph consists of 5866 connected components (i.e., maximal undirected connected subgraphs). The largest component (LC) contains 100,751 users, while the size of the second largest component is only 31. Hence, in order to receive more trust-enhanced recommendations, users should connect to the largest component. But as shown in Table 1, this cluster does not even contain half of the cold start users. This, combined with the fact that cold start users evaluate only a few users (as shown in the third and fourth row

Table 1
CS users in the dataset

	CS1	CS2	CS3	CS4
% of review raters	36.52	12.32	6.85	4.47
% in LC	18.43	30.85	38.34	44.88
Mean # trust rel	0.27	0.51	0.72	0.99
Mean # distrust rel	0.03	0.05	0.06	0.09

[3]www.epinions.com/help/faq/, accessed on February 12, 2008.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 5

of Table 1), illustrates that cold start users in the classical sense are very often cold start users in the trust sense as well.

Better results can be expected when newcomers connect to a large component of the trust graph, but they may encounter difficulties in finding the most suitable people to connect to. Therefore, we define three user classes and locate them in the network.

3.2. *Key figures*

The first class of key figures are *mavens*, people who write a lot of reviews. This term is borrowed from Gladwell's book [10] in which mavens are defined as knowledgeable people who want to share their wisdom with others. Out of the three user classes mavens are the most visible, and hence the ones which are the easiest to evaluate: the more reviews someone writes, the better a new user can form an opinion on him and decide to put him in his personal WOT or not.

Unlike mavens, *frequent raters* are not always so visible. They do not necessarily write a lot of reviews but evaluate a lot of them, and hence are an important supplier for the recommender system: it is not possible to generate predictions without ratings. By including a frequent rater in a trust network, more items can be reached, which has a direct influence on the coverage of the system.

While mavens and frequent raters are not necessarily bound to the trust network, *connectors* are: they connect a lot of users and occupy central positions in the trust network. Such users issue a lot of trust statements (many outlinks) and are often at the receiving end as well (many inlinks). The strength of connectors lies not in their rating capacity or visibility, but in their ability to reach a large group of users through trust propagation. When a trust-enhanced algorithm has to find a path from one user to another, a connector will be part of the propagation chain more often than a random user, and propagation chains containing connectors will on average be shorter than other chains. Shorter chains have a positive influence on the accuracy of the trust estimations and recommendations, as discussed in [11].

Figure 1 shows a diagram with examples of each type: the darker the node, the more reviews the user wrote (maven). The larger the node, the more reviews the user evaluated (frequent rater). The trust network is denoted by the arrows representing trust relations; connectors are characterized by many incoming and outgoing arrows.

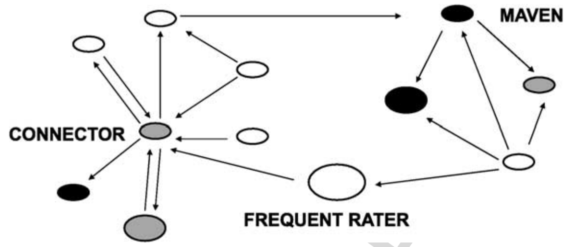

Fig. 1. Key figures example.

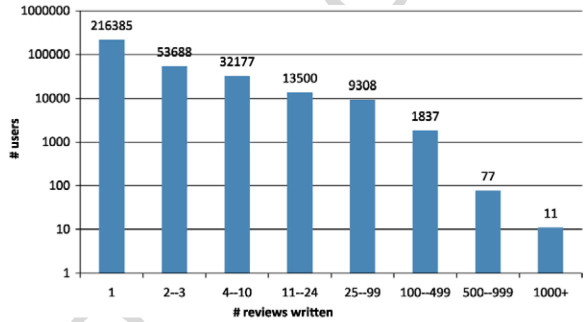

Fig. 2. Number of reviews written vs. number of authors.

In the Epinions dataset, we define a maven as someone who has written at least 100 reviews (M-100+), a frequent rater as someone who has evaluated at least 2500 reviews (F-2500+), and a connector as someone who has an in+out degree of at least 175 (C-175+). With these definitions,[4] the community contains 1925 mavens, 1891 frequent raters and 1813 connectors. These thresholds are chosen for a number of reasons. Firstly, the characteristics of the key figures must be distinctive. For example, among all authors (i.e., users who wrote at least one review), the average number of reviews written is 4.77 while the maximum is 1496. Obviously, a user who has written merely 5 reviews cannot be regarded as a maven. Figure 2 shows the distribution of the number of reviews per author; there are over 300,000 authors. The users who wrote more than 100 reviews constitute about 0.6% of all review writers, which we consider a good representation of the 'true' mavens: they certainly exhibit the desired behaviour and the size of the group is still large enough to diversify (we refer to Section 5.3 for a further discussion on this topic). The thresholds for frequent raters and connectors are obtained analogously, each of them representing about 1% of the corresponding user sets: the F-2500+ and C-175+ sets

[4]Note that we cannot refine the definitions by taking into account additional information such as the length or the class of the reviews, because the dataset does not contain any other information.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 6

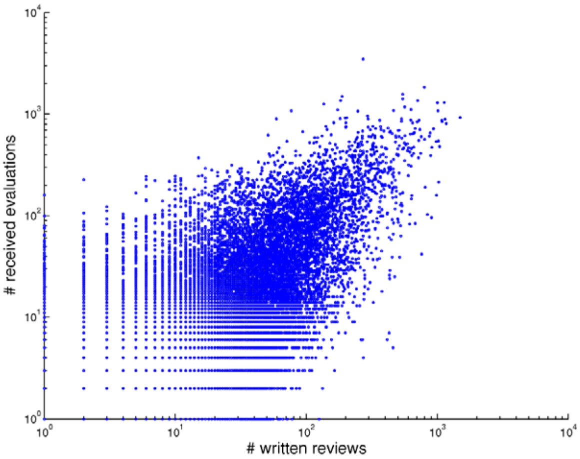

Fig. 3. Reviews written vs. evaluations received.

constitute about 1.2% of the raters and 1.4% of the trust graph members respectively. Secondly, the thresholds are also chosen such that the different key figure sets have similar sizes; this enables us to perform the analysis in the following paragraph in a fairer way. In Section 5, we experiment with other thresholds as well.

The sets of connectors and mavens share a large number of users, which is not surprising because mavens are visible through the reviews they write, making it more likely for others to connect to them by trust statements. This is illustrated by Fig. 3: the horizontal axis corresponds to the number of reviews a user has written; the more to the right a user is, the more of a maven he is. The vertical axis corresponds to the number of evaluations a user has received. The higher someone is on that axis, the more inlinks he receives (and the more of a connector he will be). In particular, the conditional probability $P(\text{M-100+}|\text{C-175+}) \approx 0.52$. More surprising is the relation between connectors and frequent raters, namely $P(\text{F-2500+}|\text{C-175+}) \approx 0.64$. The intersection of the maven set and the frequent rater set also contains many users (933), so there clearly is a strong overlap between the different groups of key figures. This indicates that users who are active on one front are often active on other fronts as well.

Note that these findings may be influenced by Epinions' 'Income Share program' and the benefits of being selected as a category lead, top reviewer or advisor.[5] Some of these classes are related to the key figures we defined, though our approach for identifying key figures only relies on objective data, while the selection in the Income Share program is partially subjective. Note that Epinions' interface also has an impact on the visibility and relatedness of the user classes.

Although their characteristics may be influenced by the specific situation, the three user classes can be detected in many kinds of trust-based RSs, and hence the results in the remainder of the paper can easily be generalised. In the following sections, we investigate the impact of the identified key figure types in the trust network by means of new social network analysis measures.

4. Measuring the impact of trusted users

In this section we tackle the first two questions raised in the introduction: we zoom in on a user a and we inspect a user b in the web of trust of a. More in particular, we propose a way to quantify the impact of b on the coverage and the accuracy of the recommendations generated for a through the trust network. In the remainder, we use $\mathit{WOT}(a)$ to denote the web of trust of a. A straightforward approach is to remove b from $\mathit{WOT}(a)$ and to compare the *accuracy* and the *coverage* in the resulting network with the initial situation.

A classical way to measure the accuracy of recommendations is by using the leave one out method, which consists of hiding a rating first and then predicting its value and determining the deviation. In particular, the *mean absolute error* (MAE) metric [19] is computed as in Eq. (3): N denotes the number of available ratings for a, $r_{a,i}$ the actual rating and $p_{a,i}$ the predicted rating:

$$\mathit{MAE}(a) = \frac{\sum_{i=1}^{N} |p_{a,i} - r_{a,i}|}{N}. \tag{3}$$

Better prediction algorithms have lower MAE's. The accuracy change $\mathit{AC}(b, a)$ is obtained by subtracting the MAE after excluding the ratings and trust links provided by b, from the MAE when taking into account all available ratings and links.

Definition 1 (Accuracy change). The change in accuracy caused by user b for user a is defined as:

$$\mathit{AC}(b, a) = \mathit{MAE}(a) - \mathit{MAE}(a, -b),$$

in which $\mathit{MAE}(a, -b)$ denotes the MAE when b is omitted from $\mathit{WOT}(a)$.

[5] www.epinions.com/help/faq/show_~faq_recognition, accessed on February 12, 2008.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 7

Consequently, a positive *AC* denotes higher prediction errors when taking into account the ratings and links provided by user b. Formula (3) only takes into account items i for which a rating $r_{a,i}$ is available. Since the problem with cold start users in the first place is that they have rated only very few items, the value of N in (3) is typically low. Even worse: for a cold start user a who rated only one item so far, the leave one out method can not even be used as it hides the sole rating available for a, leaving the recommender system clueless. In Section 4.1, we therefore propose the use of a betweenness measure as a more informative way to assess the impact of user b on the accuracy for a.

The coverage for a relates to the number of items that are accessible from a, either directly or through trust propagation. In the remainder, let $Acc_0(a)$ denote the set of items that are rated by a, i.e. the set of items that are accessible from a in zero propagation steps. Through propagation, more items can become accessible from a. We use $Acc_n(a)$ to denote the set of items that are accessible in n steps from a but not less, i.e., items i that have n intermediary nodes on the shortest path from a to i.

Definition 2 (Accessible items). The set of items accessible from a in n propagation steps, but not less, is defined as

$$\begin{aligned} Acc_0(a) &= \{i \mid \text{item } i \text{ is rated by } a\},\\ Acc_n(a) &= \bigcup\{Acc_{n-1}(u) \mid u \in \mathit{WOT}(a)\}\\ &\quad \setminus \bigcup\{Acc_k(a) \mid k = 0,\ldots,n-1\}. \end{aligned}$$

Note that $|Acc_n(a)|$ is the number of new items for which a rating can be predicted with (2) using n propagation steps. In a similar way we define $Acc_n(a,-b)$ as the set of items still accessible from a after omitting b from a's web of trust.

Definition 3 (Accessible items after omission).

$$\begin{aligned} &Acc_0(a,-b) = Acc_0(a)\\ &Acc_n(a,-b)\\ &\quad = \bigcup\{Acc_{n-1}(u) \mid u \in \mathit{WOT}_{-b}(a)\}\\ &\qquad \setminus \bigcup\{Acc_k(a,-b) \mid k = 0,\ldots,n-1\}, \end{aligned}$$

in which $\mathit{WOT}_{-b}(a) = \mathit{WOT}(a) \setminus \{b\}$.

Note that normalizing the difference $|Acc_1(a)| - |Acc_1(a,-b)|$ by dividing it by the total amount of items available in the RS results in very small values as a RS typically contains thousands of items. Instead of looking at the number of items still accessible from a after the removal of b and relating this to the total amount of items in the RS, we therefore focus on the number of items that is lost when b is omitted from a's web of trust, and relate this to the total number of items accessible from a. To this end we propose in Section 4.2 an adaptation of an existing fragmentation measure.

4.1. Betweenness

As shorter propagation chains yield more accurate predictions, one way of measuring the impact of users is by counting how often they are on shortest paths leading to items. To quantify this, we use the following measure which is inspired by the well known betweenness measure, commonly used to locate users who have a large influence on the flow in a network (see e.g. [7, 8,41]).

Definition 4 (Betweenness). Let a be a user and b a member of $\mathit{WOT}(a)$. The betweenness of b for a on level n is defined as:

$$B_n(b,a) = \frac{1}{|Acc_n(a)|} \sum_{i \in Acc_n(a)} \left(\frac{\tau_{ai}(b)}{\tau_{ai}}\right),$$

in which τ_{ai} is the number of different shortest paths from user a to item i and $\tau_{ai}(b)$ is the number of those shortest paths that contain b.

Note that $B_n(b,a) \in [0,1]$. Also remark that a shortest path from a to i containing b always contains the edge from a to b as its first link.

Example 1. In the first scenario in Fig. 4, 3 items are accessible from a. b_1 is on the only shortest path from a to i_1 as well as on one of the two shortest paths from a

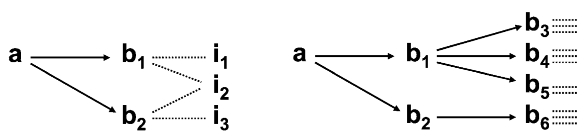

Fig. 4. Example: scenarios 1 and 2.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 8

to i_2, hence we obtain:

$$B_1(b_1, a) = \frac{1}{3} \cdot \left(1 + \frac{1}{2}\right) = \frac{1}{2}.$$

Similarly, $B_1(b_2, a) = 1/2$. However, when focusing on items reached in an additional propagation step (scenario 2), the betweenness of b_1 and b_2 is no longer equal. Because b_1 connects to more users, a can reach more items through b_1 than through b_2. In other words, b_1 is more of a connector than b_2: $B_2(b_1, a) = 8/11$, while $B_2(b_2, a) = 3/11$. In the above we presuppose that all items on level 2 are different from i_1, i_2 and i_3. Note that if, e.g., i_3 were one of the two items rated by b_5, the betweenness of b_1 would decrease (7/10) because he is not on the shortest path to i_3.

This example illustrates that betweenness rewards connectors. If user b is the only one in a's web of trust to have rated a particular item i, then for that i the maximal value of $\tau_{ai}(b)/\tau_{ai}$ is added, namely 1. In this sense, betweenness also rewards frequent raters who contribute to the coverage.

$B_n(b, a)$ gives an indication of the absolute impact of b on the coverage of the recommendations for a, but it does not provide information on how b compares to other members of a's WOT. However, this is a determining factor for the real impact of b on the recommendations generated for a. A strong WOT contains strong users who rate many items and link to other strong users. Adding b to such a WOT is less beneficial than adding b to a weak WOT: in the latter case, a will often reach more previously unreachable items through b, whereas less items are unreachable in a strong WOT (thanks to the strong members). In other words, b will have a more significant influence when a has a weak WOT. We can represent the *WOT strength* by the betweenness of the best user of the WOT besides the key figure, and compare this value to the betweenness of the key figure.

Definition 5 (Betweenness utility). The betweenness utility of user b for user a on level n is defined as:

$$BU_n(b, a) = B_n(b, a) - \max_{u \in \mathit{WOT}_{-b}(a)} B_n(u, a).$$

4.2. Fragmentation

Instead of focusing on shortest paths, user b's influence can also be measured by the reduction in cohesion of the network which occurs if b is deleted from a's WOT. User b is vital for a when he rates a lot of items and when a lot of these items are only rated by b. Deleting such a high impact user from a WOT results in a fragmented network with many items appearing in isolated fragments. For a user a we study the fragmentation in the undirected graph corresponding to the network like the ones depicted in Fig. 4, i.e., the graph that contains as its nodes all users and items accessible from a in zero or more propagation steps, and the links that lead to them as its edges.

Example 2. In the first as well as in the second scenario of Fig. 4 all items are initially in one fragment. If we remove b_1 from *WOT*(a) in the first scenario, two fragments arise, namely $\{i_1\}$ and $\{i_2, i_3\}$. Similarly, in the second scenario, 9 fragments (of which 8 are islands, i.e. containing only 1 item) are obtained after deleting the edge from a to b_1.

To quantify the fragmentation impact, we count the number of pairs of items that become disconnected from each other, i.e., items that are in separate fragments after removal of b. Note that a fragment containing s items contains exactly $s \cdot (s - 1)$ connected item pairs, since all items in the same fragment are connected to each other. The following measure, which is a modification of the traditional fragmentation measure (see e.g. [4,6]), is based on this.

Definition 6 (Fragmentation). Let a be a user and b a member of *WOT*(a). The fragmentation of b for a on level n is defined as

$$F_n(b, a) = 1 - \frac{\sum_{j=1}^{k} s_j(s_j - 1)}{|Acc_n(a)| \cdot (|Acc_n(a)| - 1)},$$

in which k is the number of fragments after removing b from *WOT*(a), and s_j is the number of items in the jth fragment.

The numerator describes the situation after the removal of b: there are k fragments and each jth fragment contains $s_j \cdot (s_j - 1)$ pairs of connected items, hence the numerator is the total number of connected item pairs after removal of b. The denominator on the other hand describes the original state of the network, i.e. before omitting b from *WOT*(a): all $|Acc_n(a)|$ items are in the same fragment (i.e. minimal fragmentation) and this fragment contains $|Acc_n(a)| \cdot (|Acc_n(a)| - 1)$ connected item pairs.

A user b who has only rated items that are also reachable through other users will yield $F_n(b, a) = 0$, because the situation after deletion does not differ from

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 9

the minimal fragmentation situation. In other words, the fragmentation measure rewards b's original contribution to the coverage for a: when b is removed from $WOT(a)$, items that have only been rated by b become separate fragments. The more islands, the more $F_n(b, a)$ approaches 1, the ideal situation. Note that $F_n(b, a) \in [0, 1]$.

Example 3. In the first scenario of Fig. 4 it holds that

$$F_1(b_1, a) = F_1(b_2, a) = \frac{2}{3}.$$

In the second scenario of Fig. 4, we obtain $F_2(b_1, a) = 104/110 \approx 0.95$ while $F_2(b_2, a) = 54/110 \approx 0.49$, which reflects that b_1 plays a more vital role than b_2 in the web of trust of a.

Much work has been done on the vulnerability of networks to disconnection. A large part of it focuses on cutpoint problems, such as the min-k-cut or the min-k-vertex sharing problem (e.g. [26]). The latter tries to minimize the number of deleted users to achieve a k-way partition. This problem is complementary to ours, as we know the number of users to be deleted: in our experiments we typically remove one user from the WOT and study the effect.

When assessing the influence of a particular user, it is best to take into account fragmentation and betweenness together: users that have an equal fragmentation score might still be distinguished based on betweenness, and vice versa.

Example 4. For scenario 3 in Fig. 5 we obtain:

$$\begin{aligned} &F_1(b_1, a) = 6/12, \qquad && B_1(b_1, a) = 3/8,\\ &F_1(b_2, a) = 0, && B_1(b_2, a) = 2/8,\\ &F_1(b_3, a) = 6/12, && B_1(b_3, a) = 3/8 \end{aligned}$$

while in scenario 4 it holds that:

$$\begin{aligned} &F_1(b_1, a) = 0, \qquad && B_1(b_1, a) = 3/8,\\ &F_1(b_2, a) = 0, && B_1(b_2, a) = 1/8,\\ &F_1(b_3, a) = 6/12, && B_1(b_3, a) = 4/8. \end{aligned}$$

If we focus on fragmentation only, then the influence of b_3 is the same in both scenarios. However, it is clear that b_3 in scenario 4 is more beneficial, because he has rated more items, and more item ratings help to obtain more accurate predictions. This is reflected in the betweenness value for b_3, which is higher in scenario 4.

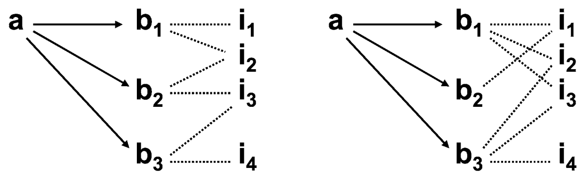

Fig. 5. Example: scenarios 3 and 4.

Analogously, although b_1 has the same betweenness in both scenarios, it is clear that he is more beneficial in scenario 3, since in scenario 4 all items rated by b_1 can also be reached through other users. This is reflected by a higher fragmentation value for b_1 in scenario 3.

Although in theory the fragmentation impact of b for a can range from 0 to 1, in practice its upper bound is determined by the behaviour of all users in a's web of trust, more in particular by the number of items that they rated in common. While for the betweenness measure different users can score well simultaneously by occurring frequently on (different) shortest paths, for the fragmentation score they are in competition with each other. Fragmentation rewards original contributions, so the more items are rated by more than one user, the harder it is for individual users to achieve a high fragmentation score. We call the practical upper bound on $F_n(b, a)$ the *room for originality*. It is defined as:

$$F_n^{\max}(a) = 1 - \frac{|Com_n(a)| \cdot (|Com_n(a)| - 1)}{|Acc_n(a)| \cdot (|Acc_n(a)| - 1)},$$

in which $Com_n(a)$ represents the set of items in $Acc_n(a)$ that are accessible through more than one user of a's WOT:

$$Com_n(a) = \bigcap \{Acc_n(a, -x) \mid x \in WOT(a)\}.$$

Note that $F_n^{\max}$ is the same for all users in a's web of trust. $F_n^{\max}$ is reached when a single user of $WOT(a)$ reaches all non common items. This corresponds to the maximal fragmentation situation possible in practice.

Example 5. In scenario 4 of Fig. 5, there is only one non common item, which is reached by b_3. $Com_1(a) = \{i_1, i_2, i_3\}$, hence

$$F_1^{\max}(a) = 1 - \frac{3 \cdot 2}{4 \cdot 3} = \frac{1}{2}.$$

This value is indeed reached at $F_1(b_3, a) = 1/2$. In scenario 1 of Fig. 4 on the other hand, $Com_1(a) = \{i_2\}$.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 10

In this case $F_1^{\max}(a)$ is 1 which indicates that there is more room for original contribution than in scenario 4. Even though in absolute terms $F_1(b_1, a) = 2/3$ from scenario 1 is higher than $F_1(b_3, a) = 1/2$ from scenario 4, user b_3 from scenario 4 exhibits a stronger behaviour as he filled the room for original contribution maximally while user b_1 from scenario 1 only managed to fill two thirds.

We take these considerations into account by normalizing the fragmentation utility w.r.t. the room for originality. Note that FU_n as well as BU_n range from -1 to 1.

Definition 7 (Fragmentation utility). The fragmentation utility of user b for user a on level n is defined as:

$$FU_n(b,a) = \frac{F_n(b,a) - \max_{u \in WOT_{-b}(a)} F_n(u,a)}{F_n^{\max}(a)}.$$

5. Results and discussion

To answer the third question raised in the introduction, we performed two kinds of experiments to investigate the influence of key figures on the coverage and accuracy of CS recommendations. Table 2 gives an overview of the measures we evaluated.

5.1. Contribution of key figures

In the first experiment, we analyse the role of key figures in a cold start user's WOT and compare them with random WOT members. To this aim, we only consider CS users who have exactly one key figure of a specific type in their WOT. For instance, the set of CS2 users who are connected with exactly one maven of type M-1000+. We denote such a set as U and represent a user of U by a. The corresponding key figure is denoted by k_a, and a randomly chosen member of a's WOT by r_a, i.e., $r_a \in WOT(a) \setminus \{k_a\}$. The results for the SNA measures in this experiment can be found in Tables 3–5. A column (row) corresponds to a specific

Table 2

Notations used in Sections 4 and 5

k	Superscript for key figure	r	Superscript for random WOT user
a	Superscript for best alternative WOT user		
$AC(b,a)$	Accuracy change	AAC	Average accuracy change
$B_n(b,a)$	Betweenness	$F_n(b,a)$	Fragmentation
AB_n	Average betweenness	AF_n	Average fragmentation
$BU_n(b,a)$	Betweenness utility	$FU_n(b,a)$	Fragmentation utility
$DBU(a)$	Betweenness utility difference between $BU(random\ key, a)$ and $BU(random\ active, a)$	$DFU(a)$	Fragmentation utility difference between $FU(random\ key, a)$ and $FU(random\ active, a)$
$ADBU$	Average betweenness utility difference	$ADFU$	Average fragmentation utility difference

Table 3

Evaluation for frequent raters (F), mavens (M) and connectors (C) on L1. Experiment 1, average betweenness and fragmentation for the key figure

Type (#)	AB_1^k (σ^B)				AF_1^k (σ^F)			
	CS1	CS2	CS3	CS4	CS1	CS2	CS3	CS4
F-100000 (2)	0.90 (0.20)	0.86 (0.25)	0.85 (0.27)	0.85 (0.24)	0.94 (0.20)	0.88 (0.26)	0.88 (0.28)	0.90 (0.21)
F-50000 (36)	0.85 (0.26)	0.83 (0.25)	0.80 (0.26)	0.80 (0.29)	0.89 (0.25)	0.88 (0.24)	0.87 (0.25)	0.85 (0.28)
F-10000 (459)	0.89 (0.26)	0.85 (0.28)	0.84 (0.28)	0.83 (0.28)	0.92 (0.24)	0.89 (0.26)	0.89 (0.26)	0.89 (0.26)
F-2500 (1394)	0.85 (0.31)	0.80 (0.34)	0.73 (0.39)	0.71 (0.38)	0.88 (0.29)	0.84 (0.32)	0.77 (0.37)	0.76 (0.36)
M-1000 (11)	0.75 (0.34)	0.75 (0.31)	0.69 (0.36)	0.72 (0.35)	0.80 (0.33)	0.82 (0.28)	0.75 (0.35)	0.77 (0.35)
M-500 (77)	0.80 (0.33)	0.73 (0.37)	0.68 (0.38)	0.70 (0.36)	0.84 (0.31)	0.78 (0.35)	0.74 (0.36)	0.75 (0.35)
M-100 (1837)	0.91 (0.24)	0.85 (0.30)	0.83 (0.32)	0.81 (0.33)	0.93 (0.22)	0.89 (0.27)	0.87 (0.30)	0.85 (0.30)
C-1000 (47)	0.88 (0.24)	0.82 (0.28)	0.79 (0.30)	0.79 (0.31)	0.92 (0.22)	0.88 (0.23)	0.85 (0.28)	0.85 (0.28)
C-500 (253)	0.81 (0.33)	0.78 (0.34)	0.72 (0.36)	0.74 (0.34)	0.84 (0.31)	0.83 (0.32)	0.78 (0.34)	0.81 (0.31)
C-175 (1513)	0.86 (0.30)	0.80 (0.35)	0.77 (0.36)	0.72 (0.38)	0.89 (0.27)	0.83 (0.33)	0.82 (0.33)	0.77 (0.36)

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 11

Table 4

Evaluation for frequent raters, mavens and connectors on L1. Experiment 1, average betweenness and fragmentation for a random WOT member

Type (#)	AB_1^r (σ^B)				AF_1^r (σ^F)			
	CS1	CS2	CS3	CS4	CS1	CS2	CS3	CS4
F-100000 (2)	0.07 (0.19)	0.01 (0.02)	0.05 (0.15)	0.04 (0.15)	0.09 (0.20)	0.06 (0.16)	0.09 (0.22)	0.03 (0.08)
F-50000 (36)	0.14 (0.26)	0.08 (0.15)	0.11 (0.20)	0.13 (0.25)	0.19 (0.27)	0.14 (0.22)	0.14 (0.24)	0.16 (0.26)
F-10000 (459)	0.17 (0.30)	0.16 (0.29)	0.13 (0.24)	0.12 (0.23)	0.21 (0.33)	0.18 (0.28)	0.16 (0.27)	0.17 (0.28)
F-2500 (1394)	0.21 (0.33)	0.23 (0.34)	0.23 (0.34)	0.21 (0.32)	0.27 (0.37)	0.27 (0.37)	0.30 (0.39)	0.24 (0.35)
M-1000 (11)	0.21 (0.33)	0.15 (0.22)	0.17 (0.28)	0.16 (0.28)	0.24 (0.33)	0.21 (0.27)	0.22 (0.32)	0.18 (0.29)
M-500 (77)	0.19 (0.30)	0.20 (0.30)	0.18 (0.29)	0.15 (0.26)	0.26 (0.35)	0.26 (0.33)	0.22 (0.33)	0.20 (0.30)
M-100 (1837)	0.21 (0.33)	0.20 (0.32)	0.23 (0.34)	0.19 (0.32)	0.26 (0.36)	0.26 (0.36)	0.28 (0.38)	0.25 (0.36)
C-1000 (47)	0.13 (0.24)	0.12 (0.21)	0.11 (0.21)	0.14 (0.26)	0.14 (0.25)	0.15 (0.25)	0.14 (0.23)	0.14 (0.25)
C-500 (253)	0.22 (0.33)	0.20 (0.31)	0.17 (0.28)	0.15 (0.24)	0.28 (0.36)	0.26 (0.35)	0.23 (0.32)	0.21 (0.31)
C-175 (1513)	0.25 (0.34)	0.24 (0.35)	0.25 (0.35)	0.25 (0.35)	0.30 (0.39)	0.31 (0.38)	0.29 (0.39)	0.31 (0.39)

Table 5

Evaluation for frequent raters, mavens and connectors on L1. Experiment 1, average betweenness and fragmentation for the best alternative WOT member

Type (#)	AB_1^a (σ^B)				AF_1^a (σ^F)			
	CS1	CS2	CS3	CS4	CS1	CS2	CS3	CS4
F-100000 (2)	0.10 (0.20)	0.08 (0.16)	0.12 (0.22)	0.11 (0.18)	0.11 (0.20)	0.09 (0.17)	0.13 (0.24)	0.10 (0.15)
F-50000 (36)	0.13 (0.24)	0.14 (0.20)	0.16 (0.23)	0.18 (0.27)	0.15 (0.26)	0.18 (0.24)	0.21 (0.25)	0.22 (0.29)
F-10000 (459)	0.10 (0.25)	0.14 (0.27)	0.14 (0.26)	0.15 (0.26)	0.13 (0.28)	0.18 (0.30)	0.18 (0.30)	0.20 (0.30)
F-2500 (1394)	0.13 (0.28)	0.18 (0.32)	0.24 (0.35)	0.26 (0.35)	0.16 (0.32)	0.22 (0.36)	0.29 (0.40)	0.32 (0.39)
M-1000 (11)	0.20 (0.29)	0.20 (0.25)	0.24 (0.30)	0.22 (0.29)	0.24 (0.33)	0.28 (0.32)	0.30 (0.34)	0.26 (0.31)
M-500 (77)	0.17 (0.30)	0.23 (0.32)	0.27 (0.33)	0.24 (0.29)	0.21 (0.34)	0.28 (0.36)	0.33 (0.37)	0.31 (0.35)
M-100 (1837)	0.09 (0.23)	0.14 (0.29)	0.16 (0.30)	0.18 (0.31)	0.11 (0.27)	0.17 (0.32)	0.19 (0.34)	0.22 (0.36)
C-1000 (47)	0.09 (0.20)	0.15 (0.24)	0.17 (0.24)	0.18 (0.27)	0.12 (0.24)	0.19 (0.28)	0.22 (0.29)	0.22 (0.31)
C-500 (253)	0.17 (0.30)	0.19 (0.30)	0.24 (0.32)	0.22 (0.30)	0.21 (0.34)	0.24 (0.35)	0.30 (0.36)	0.29 (0.36)
C-175 (1513)	0.13 (0.28)	0.19 (0.33)	0.21 (0.33)	0.26 (0.36)	0.16 (0.32)	0.22 (0.37)	0.25 (0.38)	0.31 (0.40)

user group (key figure), e.g., a M-100 is a maven who wrote at least 100 and at most 499 reviews.

Table 3 contains the average betweenness and fragmentation values of a key figure (AB_1^k and AF_1^k resp.), while Table 4 contains the average betweenness and fragmentation of a random other WOT member (AB_1^r and AF_1^r). Finally, Table 5 contains the results for the best alternative user (AB_1^a and AF_1^a). Note that AB_1^a and AF_1^a represent the average WOT strength.

The formula for the average betweenness value of the key figures for cold start users who are connected with exactly one key figure of a certain type is given by (4); the other formulas are analogous:

$$AB_1^k = \frac{\sum_{a \in U} B_1(k_a, a)}{|U|}. \qquad (4)$$

For each table we also included the standard deviations, which are denoted by σ^B and σ^F for the betweenness and fragmentation averages respectively.

A key figure is clearly very influential for a CS user, with an average AB_1^k of 0.80 and an average AF_1^k of 0.84. As expected, the betweenness and fragmentation values for a random WOT user are significantly lower. Frequent raters score somewhat higher than connectors and mavens, with an average AB_1^k of 0.83 and an average AF_1^k of 0.87. This is not surprising because frequent raters are the real suppliers for the RS. Hence, it is more difficult for members of such a WOT (containing a frequent rater) to obtain a high betweenness and fragmentation value, than for members of another WOT. This explains why AB_1^a and AF_1^a are generally lower for CS users connected to a frequent rater. For instance, AB_1^a is on average 0.15 for frequent raters, as opposed to 0.18 and 0.20 for connectors and mavens, respectively.

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 12

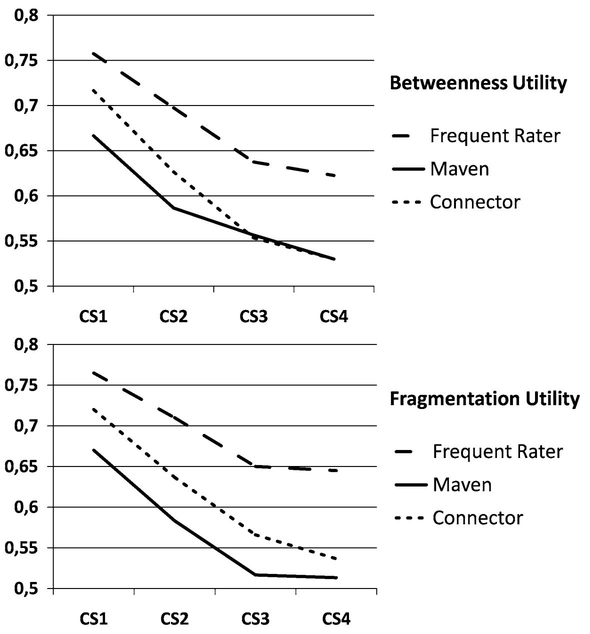

Fig. 6. Experiment 1, average betweenness and fragmentation utility.

Figure 6 depicts the course of the average betweenness and fragmentation utility of the different key figure types. Recall that the utility compares the impact of the key figure (B_1^k and F_1^k) to that of the best alternative user in the WOT (B_1^a and F_1^a). For the fragmentation utility values, also another contextual factor is taken into account, namely the room for originality. We did not include the originality results, as all of them approach to 1.

The figure clearly shows us that the use of having a key figure in a WOT decreases as the new user becomes more active. Indeed, as is illustrated in Table 1, more active CS users rate more items and issue more trust statements; consequently, the WOT sizes become larger. This means that there is a higher chance that one of the WOT members is a stronger user, yielding higher values for B_1^a and F_1^a, and lower values for the key figures.

We claimed that including connectors in a WOT yields shorter propagation chains because they connect more users and reach more reviews. Therefore, besides the above experiment (*level 1*, L1), we also measured the influence of coverage by propagating trust information one step (*level 2*, L2). Specifically, this means that if a trusts b and b trusts u, $t_{a,u}$ in (2) equals 1.

For example, the results for CS3 and fragmentation are shown in Table 6. As can be seen, the average F_2 values are actually lower than their L1-counterparts. However, it is important to realize that the amount of

Table 6

Evaluation of connectors for CS3 users on L2, experiment 1

	C-1000	C-500	C-175
AF_1^k	0.85	0.78	0.82
AF_2^k	0.58	0.66	0.77
AFU_1^k	0.63	0.48	0.56
AFU_2^k	0.70	0.67	0.71
AF_1^a	0.22	0.30	0.25
AF_2^a	0.16	0.18	0.18
Avg. $F_1^{\max}$	1.00	1.00	1.00
Avg. $F_2^{\max}$	0.60	0.67	0.77

new items that is provided e.g. by a C-500 via one step propagation is almost 20 times the amount delivered by a C-500 on the first level; for instance, for CS3 users connected to a C-500, Acc_1 contains 33,102 reachable items, while Acc_2 contains 641,758 items. Hence, the lower values can be explained by the fact that more items reached through the connector are also reached through other WOT members, which is illustrated by the lower $F^{\max}$ values on level 2.

On the other hand, the average fragmentation utility AFU^k increases compared to level 1. As indicated by the last four rows of the table, this is due both to weaker AF_1^a values, and to the fact that there is less room left for originality on level 2. To conclude, it is clear that trust propagation and connectors have a strong positive impact on the coverage of the RS.

As mentioned earlier, an increase in coverage is beneficial only to the extent that the accuracy does not drop significantly. Therefore, the average *AC* values (*AAC*, see formula (5)), were also computed and are shown in Table 7:

$$AAC = \frac{\sum_{a \in U} AC(k_a, a)}{|U|}. \qquad (5)$$

Note that no results are generated for the CS1 group: formula (2) uses the mean of a user's ratings, but the leave one out method already hides the sole rating of a CS1 user.

Since items are rated on a scale from 1 to 5, the extreme values of *AC* and *AAC* are -4 and 4. Because we use the leave one out method, we can only take into account items that are rated by the cold start user, i.e., items of Acc_0. Hence, on level 1, *AAC* measures the average accuracy change for items that are immediately accessible through users of a WOT-list, i.e., items that are in Acc_0 and in Acc_1. On level 2, we consider items that become accessible through trust propagation

AIC ios2a v.2008/05/27 Prn:1/10/2008; 14:07 F:aic431.tex; VTEX/R p. 13

Table 7

Experiment 1, average accuracy change

Type	*AC*		
	CS2	CS3	CS4
F-100000	−0.23	0.04	0.08
F-50000	−0.04	−0.09	0.05
F-10000	0.16	−0.02	0.00
F-2500	−0.06	0.03	−0.03
M-1000	0.05	−0.14	−0.12
M-500	0.02	−0.01	0.04
M-100	0.16	0.08	0.04
C-1000 (L1)	0.01	0.04	0.02
C-500 (L1)	0.06	−0.05	0.05
C-175 (L1)	0.01	0.03	−0.04
C-1000 (L2)	0.07	0.03	0.05
C-500 (L2)	−0.01	0.00	0.02
C-175 (L2)	−0.01	0.00	−0.04

(items in Acc_0 and in Acc_2, but not in Acc_1); the values are obtained by subtracting the MAE of the predictions generated by information reached through TTPs (trusted third parties) other than the connector, from the MAE of the predictions based on all TTPs (including the connector). Hence, positive accuracy changes denote higher prediction errors when taking into account the key figure.

The results on level 1 demonstrate that the absence or presence of a key figure in a WOT does not significantly affect the accuracy. In other words, the key figures have a positive effect on the coverage (as shown above), while maintaining sufficient accuracy. The results for L2 lead to the same conclusion.

5.2. *Benefit over random users*

The number of users in experiment 1 is fairly small compared to the total number of CS users; for example, 84.36% of the CS4 users have no F-2500 in their WOT, as opposed to 7.34% whose WOT contains exactly one. To take into account a larger group of users, we also conducted an experiment with groups of cold start users who have no key figure of a particular type in their WOT. We denote such a group by U. The goal of the experiment is then to investigate the effect of adding a key figure to such a CS user's WOT. For instance, we connect a M-100, M-500 or M-1000 to each CS2 user whose WOT does not contain a maven.

In particular, for one experiment, we calculate for each user a in a given group U the difference $DFU(a)$ between the fragmentation utilities $FU_1(b_1, a)$ and $FU_1(b_2, a)$, in which b_1 represents a randomly chosen key figure of a given type and b_2 a randomly chosen member of the set of all active users; active users are those who rated at least one user or one item, hence this set contains key figures as well. Analogously, *DBU* is defined for betweenness. In other words, *DFU* and *DBU* measure the extra gain when connecting to a key figure instead of to a random user.

Figures 7 and 8 depict the average utility differences *ADFU* and *ADBU* for each user group when a specific key figure is added to the WOT. The formula for *ADFU* is given by (6), the formula for the average betweenness utility difference is analogous:

$$\begin{aligned} ADFU &= \frac{\sum_{a \in U} DFU(a)}{|U|} \\ &= \frac{\sum_{a \in U} (FU_1(b_1, a) - FU_1(b_2, a))}{|U|}. \end{aligned} \quad (6)$$

Connecting to a key figure is clearly more beneficial than connecting to a random user. For instance, the fragmentation utility of an added key figure increases on average with 0.41 compared to the utility of the randomly added user.

The figures also show that, in general, the more active the key figure is, the more advantageous it is to have such a user in a WOT. For instance, users who are connected with a F-10000+ have a larger *DFU* and *DBU* than users connected with a F-2500. This phenomenon also occurs with mavens and connectors, which confirms once again that users who are active on one front (being a maven or connector) are often active on other fronts as well (being a frequent rater, i.e., boosting the number of accessible items).

Note that the differences become larger for more active cold start users. As Table 8 proves, this is because the utility of randomly added users decreases more rapidly than the utility of key figures when the cold start user rates more items.

Table 8 also provides an explanation why the differences for betweenness in Fig. 8 are much smaller than those for fragmentation in Fig. 7. Indeed, recall that WOT users only receive a strictly positive fragmentation value when they deliver new items, while the corresponding betweenness value still increases when delivering items for which a prediction can already be generated (in other words, common items). This explains why random users will yield higher *BU* values than *FU* values.

The accuracy change for the second experiment is calculated as the mean of all $AC(b, a)$ values over one

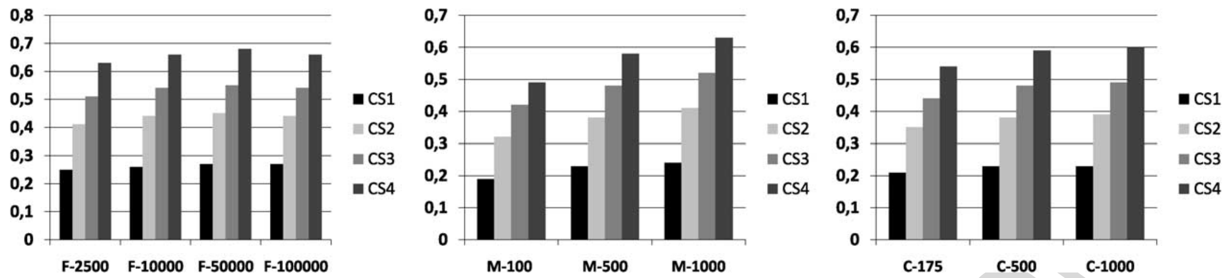

Fig. 7. Experiment 2, average fragmentation difference between key figures and random users.

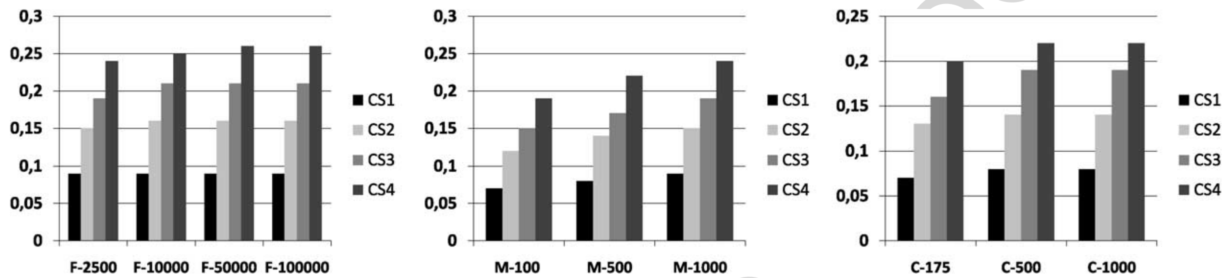

Fig. 8. Experiment 2, average betweenness difference between key figures and random users.

Table 8

Experiment 2, average fragmentation and betweenness utility

Type	AFU_1				ABU_1			
	CS1	CS2	CS3	CS4	CS1	CS2	CS3	CS4
F-100000	0.9997	0.9993	0.9992	0.9990	0.9998	0.9997	0.9995	0.9994
F-50000	0.9991	0.9982	0.9978	0.9973	0.9995	0.9990	0.9988	0.9985
F-10000	0.9964	0.9928	0.9909	0.9894	0.9981	0.9962	0.9951	0.9940
F-2500	0.9883	0.9772	0.9716	0.9654	0.9940	0.9878	0.9846	0.9817
Random	0.7298	0.5516	0.4432	0.3209	0.9076	0.8350	0.7892	0.7408
M-1000	0.9917	0.9849	0.9798	0.9794	0.9960	0.9917	0.9908	0.9891
M-500	0.9794	0.9595	0.9500	0.9398	0.9898	0.9801	0.9738	0.9700
M-100	0.9574	0.9210	0.9059	0.8739	0.9783	0.9591	0.9501	0.9363
Random	0.7402	0.5640	0.4459	0.3451	0.9106	0.8403	0.7976	0.7489
C-1000	0.9876	0.9975	0.9703	0.9136	0.9934	0.9880	0.9842	0.9812
C-500	0.9832	0.9677	0.9615	0.9524	0.9912	0.9837	0.9808	0.9723
C-175	0.9705	0.9490	09329	0.9598	0.9855	0.9732	0.9653	0.9573
Random	0.7463	0.5772	0.4650	0.3453	0.9132	0.8470	0.8035	0.7558

experiment. Note that the WOT of the CS users now contains an extra user, viz. the added key figure. The results are shown in Table 9. Because we compute the MAE by predicting existing ratings and CS users rate very few items, there is only a small chance that an added key figure will provide a rating for an item which is rated by the CS user but not by other members of his WOT. The small accuracy changes may therefore indicate that the extra ratings provided by the key figure do not significantly affect the predictions (that can already be generated by the ratings of actual WOT members).

5.3. Discussion

For new users, choosing the right WOT members can come as an overwhelming task. Therefore, the recommender system can guide and interact with such CS users by proposing a (random) list of members which

Table 9

Experiment 2, accuracy change

Type	AC_2		
	CS2	CS3	CS4
F-100000	0.014	−0.008	−0.002
F-50000	0.009	0.010	−0.001
F-10000	0.006	0.002	0.001
F-2500	0.003	0.005	0.001
M-1000	0.005	0.010	0.007
M-500	0.004	0.006	0.006
M-100	0.001	0.005	0.01
C-1000 (L1)	−0.009	0.000	−0.001
C-500 (L1)	0.005	0.011	0.000
C-175 (L1)	0.000	0.004	0.001

are worth exploring because they have an immediate and positive impact on the generated recommendations. Such 'suggestion lists' are a common technique in social networking sites. For example, in FilmTrust,[6] Golbeck encourages users to expand their network by showing two lists of users which people can connect to: a set of random users and a set of random people with no friends in the network. LinkedIn[7] and Live QnA[8] provide similar services with their 'Just joined LinkedIn' and 'Meet a QnA superstar' lists respectively.

Such systems can be further refined. Because not every user has the same likes and dislikes, the system can propose several types of (random) users, think for instance of a 'mainstream' key figure who rates a lot of popular items, or one with more distinct preferences. Furthermore, the system could narrow down the selection and present more 'tailor-made' key figures if the user has indicated that he is only interested in some specific item categories. Of course, the key figures only appear as suggestions; a new user can always check whether the candidates are worth to be included in his web of trust.

A possible consequence of our technique is that mavens and frequent raters eventually become connectors too, since the more people connect to key figures, the higher the number of inlinks they will have and hence the more of a connector they will be. Note that we showed in Section 3.2 that this phenomenon already occurs in a moderate form in the original dataset.

A related side effect is the appearance of clusters around established users in the trust network. If this clustering is undesirable, it can be restricted by choosing appropriate thresholds for the key figure selection. If one chooses high thresholds, a small number of 'true' key figures are obtained, which might lead to a small number of star-like clusters. This can be avoided by low thresholds, yielding many key figures. By generating random suggestion lists of these key figures, the network can remain more equally connected. In other words, the occurrence of strong clusters diminishes, but along with it also the power of the selected key figures, because we have shown that less active key figures yield lower betweenness and fragmentation values. Hence, it is clear that the thresholds must be chosen carefully in agreement with the characteristics of the RS's network, and that a trade-off should be made between the desired performance and network topology.

The results clearly illustrate that generated recommendations for new users are more beneficial if they connect to mavens, frequent raters or connectors compared to random users. Hence, aside from interaction and personalization, another benefit of our technique is the ability to better explain the effect of WOT users on coverage and accuracy of the system, which is a new step in the development of more transparant recommender systems. For this reason, we think that the incorporation of our technique might be a good asset for existing and future trust-enhanced RSs.

6. Conclusions and future work

The key figures we have identified, and the measures we have proposed to evaluate their influence on CS recommendations, can provide useful clues to the RS for optimizing the process of guiding new users through the connection phase. Each key figure has its own characteristics; mavens are easy to evaluate, frequent raters provide a lot of ratings, and connectors help to reach more users and items. The new measures each reflect a different aspect of the influence on coverage: betweenness focuses on a key figure's ability to reach items via short propagation chains, while fragmentation focuses on its capacity for delivering new items. The utility measures take into account environmental factors such as the strength of the web of trust. The experimental results that we obtained clearly show that connecting to an identified key figure is more beneficial than including a randomly chosen user, with respect to coverage as well as accuracy.

[6] http://trust.mindswap.org/FilmTrust/.

[7] www.linkedin.com.

[8] http://qna.live.com/.

Our future work goes in several directions. First we want to investigate the potential of other key figures like hubs and authorities by using well-known evaluation measures such as HITS [20] and PageRank [28]. Another research path is the incorporation of distrust information into the recommendation process. Distrust could e.g. be used to debug a web of trust: suppose that a trusts b completely, b fully trusts c and a completely distrusts c. The latter information ensures that the propagated trust result (viz. a trusts c) is invalid and that a will not use information coming from c in the future. As such, trust and distrust-enhanced algorithms could be used to filter out false positives generated by other techniques such as CF. Distrust can also be exploited to alleviate the sparsity problem: through specific propagation operators that can handle trust as well as distrust, more users and items could be reached. But so far, only a few researchers have focused on trust (propagation) models that take into account distrust [14,18,38,44]. Guha et al. [14] and Ziegler et al. [44] use one propagated value that incorporates both trust and distrust, but, as explained in [39], potentially important information is lost when trust and distrust scales are merged into one. Jøsang et al. [18] and Victor et al. [38] keep trust and distrust values separated throughout the complete propagation process, the former by using a probabilistic subjective logic approach [17], and the latter by using a gradual approach with fuzzy logic concepts [43]. But despite these advances, much ground remains to be covered in this domain.

Acknowledgements

Patricia Victor would like to thank the Institute for the Promotion of Innovation through Science and Technology in Flanders for funding her research. Chris Cornelis would like to thank the Research Foundation-Flanders for funding his research. We thank Epinions.com for making the data available, in particular R. Guha, R. Kumar, P. Raghavan and A. Tomkins.

References

[1] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowl. Data Eng.* **17** (2005), 734–749.

[2] H.J. Ahn, A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, *Information Sciences* **178** (2008), 37–51.

[3] A.M. Rashid, G. Karypis and J. Riedl, Influence in ratings-based recommender systems: an algorithm-independent approach, in: *Proc. of SIAM International Conference on Data Mining*, 2005.

[4] A. Arulselvan, C.W. Commander, L. Elefteriadou and P.M. Pardalos, Detecting critical nodes in sparse graphs, *Computers and Operations Research* (2007), to appear.

[5] P. Bedi, H. Kaur and S. Marwaha, Trust based recommender system for the semantic web, in: *Proc. of IJCAI-07*, 2007, pp. 2677–2682.

[6] S.P. Borgatti, Identifying sets of key players in social networks, *Comput. Math. Org. Theor.* **12** (2006), 21–34.

[7] L.C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* **40**(1) (1977), 35–41.

[8] L.C. Freeman, Centrality in social networks I: Conceptual clarification, *Social Networks* **1** (1979), 215–239.

[9] P. Domingos and M. Richardson, Mining the network value of customers, in: *Proc. of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57–66.

[10] M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, Little Brown, 2000.

[11] J. Golbeck, Computing and applying trust in web-based social networks, PhD thesis, 2005.

[12] J. Golbeck and J. Hendler, FilmTrust: movie recommendations using trust in web-based social networks, in: *Proc. of CCNC2006*, 2006, pp. 282–286.

[13] J. Golbeck, Generating predictive movie recommendations from trust in social networks, *LNCS* **3986** (2006), 93–104.

[14] R. Guha, R. Kumar, P. Raghavan and A. Tomkins, Propagation of trust and distrust, in: *Proc. of WWW2004*, 2004, pp. 403–412.

[15] J. Herlocker, J. Konstan and J. Riedl, Explaining collaborative filtering recommendation, in: *Proc. of CSCW2000*, 2000, pp. 241–250.

[16] C. Hess, K. Stein and C. Schlieder, Trust-enhanced visibility for personalized document recommendations, in: *Proc. of SAC2006*, 2006, pp. 1865–1869.

[17] A. Jøsang, A logic for uncertain probabilities, *Int. J. Uncertain Fuzz* **9**(3) (2001), 279–311.

[18] A. Jøsang, S. Marsh and S. Pope, Exploring different types of trust propagation, *LNCS* **3986** (2006), 179–192.

[19] J. Herlocker, J. Konstan, L. Terveen and J. Riedl, Evaluating collaborative filtering recommender systems, *ACM Trans. Inform. Syst.* **22**(1) (2004), 5–53.

[20] J. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM* **46** (1999), 604–632.

[21] Z. Huang, H. Chen and D. Zeng, Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering, *ACM Trans. Inform. Syst.* **22**(1), (2004), 116–142.

[22] P. Massa, A. Avesani and R. Tiella, A trust-enhanced recommender system application: Moleskiing, in: *Proc. of SAC2005*, 2005, pp. 1589–1593.

[23] P. Massa and P. Avesani, Trust-aware collaborative filtering for recommender systems, *LNCS* **3290** (2004), 492–508.

[24] P. Massa and B. Bhattacharjee, Using trust in recommender systems: an experimental analysis, *LNCS* **2995** (2004), 221–235.

[25] S.E. Middleton, H. Alani, N.R. Shadbolt and D.C. De Roure, Exploiting synergy between ontologies and recommender systems, in: *Proc. of WWW2002 Semantic Web Workshop*, 2002.

[26] H. Narayan, S. Roy and S. Patkar, Approximation algorithms for min-k-overlap problems using the principal lattice of partitions approach, *Journal of Algorithms* **21** (1996), 306–330.

[27] J. O'Donovan and B. Smyth, Trust in recommender systems, in: *Proc. of IUI2005*, 2005, pp. 167–174.

[28] L. Page, S. Brin, R. Motwani and T. Winograd, The pagerank citation ranking: bringing order to the web, Technical report, Stanford Digital Library Technologies Project, 1998.

[29] M. Papagelis, D. Plexousakis and T. Kutsuras, Alleviating the sparsity problem of collaborative filtering using trust inferences, *LNCS* **3477** (2005), 224–239.

[30] S.-T. Park, D. Pennock, O. Madani, N. Good and D. De Coste, Naive filterbots for robust cold-start recommendations, in: *Proc. of SIGKDD2006*, 2006, pp. 699–705.

[31] G. Pitsilis and L. Marshall, A trust-enabled P2P recommender system, in: *Proc. of WETICE06*, 2006, pp. 59–64.

[32] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm and J. Riedl, Grouplens: An open architecture for collaborative filtering of netnews, in: *Proc. of CSCW1994*, 1994, pp. 175–186.

[33] P. Resnick and H. Varian, Recommender systems, *Commun. ACM* **40**(3) (1997), 56–58.

[34] M. Richardson, R. Agrawal and P. Domingos, Trust management for the semantic web, in: *Proc. of ISWC03*, 2003, pp. 351–368.

[35] U. Shardanand and P. Maes, Social information filtering: algorithms for automating "word of mouth", in: *Proc. of the SICGHI Conference on Human Factors in Computing Systems*, 1995, pp. 210–217.

[36] A.I. Schein, A. Popescul, L.H. Ungar and D.M. Pennock, Methods and metrics for cold-start recommendations, in: *Proc. of SIGIR2002*, 2002, pp. 253–260.

[37] K. Swearingen and R. Sinha, Beyond algorithms: an HCI perspective on recommender systems, in: *Proc. of ACM SIGIR Workshop on Recommender Systems*, 2001.

[38] P. Victor, C. Cornelis and M. De Cock, Enhanced recommendations through propagation of trust and distrust, in: *Proc. of WI-IAT2006 Workshops*, 2006, pp. 263–266.

[39] P. Victor, C. Cornelis, M. De Cock and P. Pinheiro da Silva, Towards a provenance-preserving trust model in agent networks, in: *Proc. of WWW2006 Models of Trust for the Web Workshop*, 2006.

[40] P. Victor, C. Cornelis, A.M. Teredesai and M. De Cock, Whom should I trust? The impact of key figures on cold start recommendations, in: *Proc. of SAC2008*, 2008, pp. 2014–2018.

[41] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.

[42] J. Weng, C. Miao and A. Gohl, Improving collaborative filtering with trust-based metrics, in: *Proc. of SAC2006*, 2006, pp. 1860–1864.

[43] L.A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965), 338–353.

[44] C.-N. Ziegler and G. Lausen, Propagation models for trust and distrust in social networks, *Information System Frontiers* **7** (2005), 337–358.

[45] C.-N. Ziegler, Semantic web recommender systems, in: *Proc. of the Joint ICDE/EDBT PhD Workshop*, 2004.