Paradigms for Decentralized Social Filtering Exploiting Trust Network Structure

Cai-Nicolas Ziegler and Georg Lausen

Institut für Informatik, Group DBIS
Universität Freiburg, Germany

{cziegler,lausen}@informatik.uni-freiburg.de

Abstract. Recommender systems, notably collaborative and hybrid information filtering approaches, vitally depend on neighborhood formation, i.e., selecting small subsets of most relevant peers from which to receive personal product recommendations. However, common similarity-based neighborhood forming techniques imply various drawbacks, rendering the conception of decentralized recommender systems virtually impossible. We advocate trust metrics and trust-driven neighborhood formation as an appropriate surrogate, and outline various additional benefits of harnessing trust networks for recommendation generation purposes. Moreover, we present an implementation of one suchlike trust-based recommender and perform empirical analysis to underpin its fitness when coupled with an intelligent, content-based filter.

1 Introduction

Automated recommender systems [15] intend to provide people with recommendations of products they might appreciate, taking into account their past product ratings profile and history of purchase or interest. Most successful systems apply so-called social filtering techniques [14], particularly collaborative filtering [8]: for each active[1] user, these systems compute neighborhoods of like-minded peers, making use of some interest similarity measure in n-dimensional vector space. Hereafter, products are recommended based upon proposals of neighbors, e.g., products that *many* neighbors assigned *positive* ratings.

However, similarity-based neighborhood formation implies several computation-intensive processes, owing to the $O(|A|^2)$ complexity bottleneck when making recommendations for all $|A|$ members. Clearly, sensible operation thus becomes largely confined to *centralized* scenarios only, e.g., communities bearing manageable numbers of users and blessed with massive server cluster leverage. For decentralized scenarios, among those peer-to-peer systems, Semantic Web and the Grid, the above-mentioned similarity-based neighborhood formation scheme fails. Recall that these systems may comprise *millions* of users. Moreover, single entities, e.g., agents, machine-readable homepages, etc., commonly possess *partial* views of the entire system only.

[1] The term "active" identifies the user demanding recommendation services.

R. Meersman, Z. Tari (Eds.): CoopIS/DOA/ODBASE 2004, LNCS 3291, pp. 840–858, 2004.

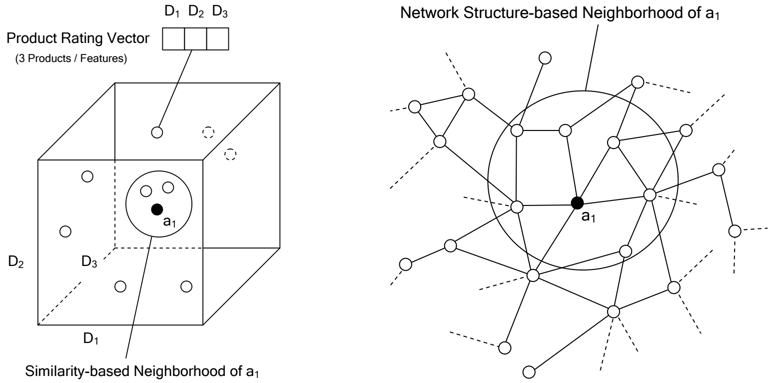

Fig. 1. Similarity-based versus network-based neighborhood formation

On the other hand, those entities part of decentralized systems are generally embedded into so-called social networks, emanating from diverse kinds of relationships holding between entities. We intend to exploit these existing network structures for efficient and scalable neighborhood formation, hence superseding similarity-based neighborhood formation schemes (see Figure 1). Moreover, we focus on one particular type of interpersonal relationships, namely trust networks. For instance, the advent of FOAF[2] networks, weaving "webs of acquaintances" [7], has been paving the way for an infrastructure of trust on the Semantic Web, and numerous communities commence incorporating concepts of trust into their very information models [9].

Our paper aims at conceiving one such trust-based recommender system aware of characteristic features and specific problems pertaining to *decentralized* application scenarios. Hereby, the underlying research combines results from our prior work on trust propagation models [35], taxonomy-driven filtering techniques for sparse data [37,37], and positive correlation between attitudinal similarity and interpersonal trust [34]. These components are seamlessly integrated into one coherent framework. Ample empirical evaluation based upon "real world" data outlines our approach's fitness for decentralized settings and investigates positive impacts that exploitation of trust network structure may have on computational efficiency and overall recommendation quality in particular.

2 Related Work

Recent studies [32] have shown that people tend to prefer receiving recommendations from people they *know* and *trust*, i.e., friends and family-members, rather than from online recommender systems. Some researchers have therefore commenced to focus on computational trust models as appropriate means to supplement or replace current collaborative filtering approaches: Kautz et al. [14] mine social network structures in order to render fruitful information exchange and collaboration feasible. Olsson [23] proposes

[2] an abbreviation for "Friend of a Friend".

an architecture combining trust, collaborative filtering and content-based filtering in one single framework, giving only vague information and insight, though. Another agent-based approach has been presented by Montaner et al. [21], who introduce so-called opinion-based filtering. Hereby, Montaner states that trust should be *derived* from user similarity, implying that friends are exactly those people that resemble our very attitudinal behavior. However, Montaner's model only extends to the agent world and does not reflect evidence acquired from *real-world* social studies concerning trust formation.

Assuming non-benevolent environments, Mui proposes an approach called collaborative sanctioning [22] for recommendation generation. His approach builds upon situation-dependent reputation and trust models. Massa [19] reasons about trust networks as suitable means for neighborhood formation when extreme product-user matrix sparseness prevails and common collaborative filtering schemes fail to infer similarity. However, Massa does not show the effectiveness of trust with respect to recommendation quality.

3 Advocacy for Trust-Based Neighborhood Formation

We investigate social network structures in order to easily assemble personalized neighborhoods for active users a. To give an example of network-based neighborhood formation, a's neighborhood may comprise exactly those peers being closest in terms of *link distance*, necessitating simple breath-first search instead of $O(|A|)$ complexity, which is required for computing similarity measures between one single a and all other individuals in the system. More specifically, we exclusively focus on *trust* relationships, motivated by reasons given below:

- **Security and attack-resistance.** Closed communities generally possess efficient means to control the user's identity and penalize malevolent behavior. Contrarily, decentralized systems cannot prevent deception and insincerity. Spoofing and identity forging thus become facile to achieve and allow for luring people into purchasing products which may provide some benefit for attackers a_o [16,34,24]. For instance, to accomplish suchlike attacks, agents a_o simply have to copy victim a_v's rating profile and add excellent ratings for products b they want to trick a_v into buying. Owing to high similarities between rating profiles of a_o and a_v, b's probability of being proposed to a_v quickly soars beyond competing products' recommendation likelihood. On the other hand, only proposing products from people the active user deems most trustworthy inherently solves this issue, hence excluding perturbations from unknown and malevolent agents from the outset.
- **Recommendation transparency.** One of the major disadvantages of recommender systems refers to their lacking transparency, i.e., users would like to understand *why* they were recommended particular goods [10]. Algorithmic clockworks of recommenders actually resemble black boxes. Hence, when proposing products from users based upon complex similarity measures, most of these "neighbors" probably being unknown to the active user, recommendations become difficult to follow. On the other hand, recommendations from trustworthy people clearly exhibit higher acceptance probability. Recall that trust metrics operate on naturally grown social

network structures while neighborhoods based upon interest similarity represent pure artefacts, computed according to some invisible scheme.
- **Correlation of trust and similarity.** Sinha and Swearingen [32] found that people tend to prefer receiving recommendations from people they *know* and *trust*, i.e., friends and family-members, rather than from online recommender systems. Moreover, positive mutual impact of attitudinal similarity on interpersonal attraction counts among one of the most reliable findings of modern social psychology [3], backing the proverbial saying that "birds of a feather flock together". Analyzing data obtained from an online community, we provided first empirical evidence of correlation between trust and interest similarity [34].
- **Mitigating the new-user cold-start problem.** One major weakness that collaborative filtering systems are faced with is the so-called new-user cold-start problem [20]: newbie members generally have issued few product ratings only. Consequently, owing to common product-user matrix sparseness and low profile overlap, appropriate similarity-based neighbors are difficult to find, entailing poor recommendations. The whole process is self-destructive, for users discontinue using the recommender system before the latter reaches acceptable performance. Trust networks alleviate cold-start issues by virtue of comparatively high network connectivity. Neighborhood formation hence becomes practicable even for users that explicitly trust one person only, taking into account an abundant transitive trust closure (see Section 4.2 for details).

Note that when computing neighborhoods based upon types of social relationships other than trust, e.g., geographical proximity, acquaintanceship, etc., the above benefits may become partially exploited only.

4 Proposed Approach

Subsequent paragraphs briefly outline our decentralized, trust-based recommender system's core constituents. Both of its essential ingredients, namely our taxonomy-driven similarity measure and our Appleseed group trust metric, have been documented before [35,34,37,36]. The major contribution of the underlying work hence consists in gluing these components together in one unifying framework and exposing synergetic effects by means of empirical analysis.

4.1 Information Model

The infrastructure we suppose allows facile mapping into diverse scenarios. For instance, Semantic Web compliance can be accomplished via FOAF networks, weaving webs of personal, machine-readable homepages.

- **Set of agents** $A = \{a_1, a_2, \dots, a_n\}$. Set A contains all agents part of the community. Globally unique identifiers are assigned through URIs.
- **Set of products** $B = \{b_1, b_2, \dots, b_m\}$. All products considered are comprised in set B. Hereby, unique identifiers may refer to product descriptions from an online shop agreed upon, such as Amazon.com (*http://www.amazon.com*), or globally accepted codes, like ISBNs in case of books.

– **Set of partial trust functions** $T = \{t_1, t_2, \ldots, t_n\}$. Every agent $a_i \in A$ has one partial trust function $t_i : A \rightarrow [-1, +1]^{\perp}$ that assigns direct, i.e., explicit, trust to its peers. Functions $t_i \in A$ are partial since agents generally only rate small subsets of the overall community, hence rendering t_i sparse:

$$t_i(a_j) = \begin{cases} p, & \text{if } \mathrm{trust}(a_i, a_j) = p \\ \perp, & \text{if no trust statement for } a_j \text{ from } a_i \end{cases} \tag{1}$$

We define high values for $t_i(a_j)$ to denote high trust from a_i in a_j, and negative values to express distrust, respectively. Values around zero indicate low trust, not to be confused with explicit distrust [18].

– **Set of partial rating functions** $R = \{r_1, r_2, \ldots, r_n\}$. In addition to functions $t_i \in T$, every $a_i \in A$ has one partial function $r_i : B \rightarrow [-1, +1]^{\perp}$ that expresses his liking or dislike of product $b_k \in B$. No person can rate every available product, so functions $r_i \in B$ are necessarily partial.

$$r_i(b_k) = \begin{cases} p, & \text{if } \mathrm{rates}(a_i, b_k) = p \\ \perp, & \text{if no rating for } b_k \text{ from } a_i \end{cases} \tag{2}$$

Intuitively, high positive values for $r_i(b_k)$ denote that a_i highly appreciates b_k, while negative values express dislike, respectively.

– **Taxonomy** C **over set** $D = \{d_1, d_2, \ldots, d_l\}$. Set D contains categories for product classification. Each category $d_e \in D$ represents one specific topic that products $b_k \in B$ may fall into. Topics express broad or narrow categories. The partial taxonomic order $C : D \rightarrow 2^D$ retrieves all immediate sub-categories $C(d_e) \subseteq D$ for topics $d_e \in D$. Hereby, we require that $C(d_e) \cap C(d_h) = \emptyset$ holds for all $d_e, d_h \in D, e \neq h$, hence imposing tree-like structuring, similar to single-inheritance class hierarchies known from object-oriented languages. Leaf topics d_e are topics with zero outdegree, formally $C(d_e) = \perp$, i.e., most specific categories. Furthermore, taxonomy C has exactly one top element $\top$, which represents the most general topic and has zero indegree.

– **Descriptor assignment function** $f : B \rightarrow 2^D$. Function f assigns a set $D_k \subseteq D$ of product topics to every product $b_k \in B$. Note that products may possess *several* descriptors, for classification into one single category generally entails loss of precision.

We suppose all information about agents a_i, their trust relationships t_i and ratings r_i stored in machine-readable homepages distributed throughout the Web. Contrarily, taxonomy C, set B of products and descriptor assignment function f must hold *globally* and therefore offer public accessibility. Central maintenance of this information hence becomes inevitable. Later on, we will demonstrate that such sources of information for product categorization already exist for certain application domains.

4.2 Trust-Based Neighborhood Formation

The computation of trust-based neighborhoods constitutes one pivotal pillar of our approach. Clearly, neighborhoods are subjective, reflecting every agent $a_i's$ very beliefs about the accorded trustworthiness of immediate peers.

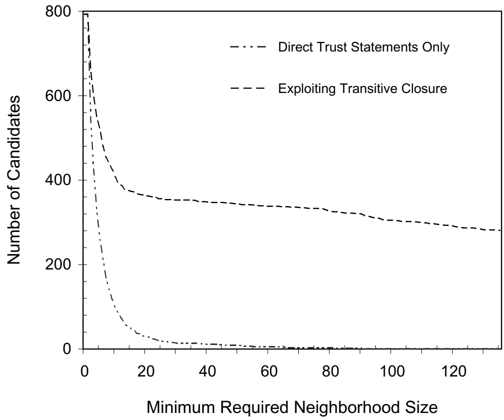

Fig. 2. Reach of direct trust versus transitive closure

Network Connectivity. However, as has been indicated before, trust functions t_i assigning *explicit* trust ratings are generally sparse. Likewise considering *indirect* trust relationships, hence exploiting the "conditional transitivity" property of trust [1], the assembly of neighborhoods that contain M most trustworthy peers becomes possible even for larger M, e.g., $M \geq 50$. Figure 2 backs our hypothesis, analyzing the connectivity of 793 users from the All Consuming (*http://www.allconsuming.com*) community. The figure shows how many agents, indicated on the y-axis, satisfy the minimum neighborhood size criterion given along the x-axis. For instance, while 49 people have issued 15 or more *direct* trust statements, 374 users are able to reach 15 or more peers when also considering the *transitive closure* of trust relationships. While the trust outdegree curve decays rapidly, the transitive closure curve's fallout decelerates drastically as the number of candidate persons drops below 400, thus revealing the presence of one highly connected trust cluster.

The above result relates to the classical theorem on random graphs [6].[3] Therein, Erdős and Rényi proved that in large graphs $G = (V, E)$, assuming E randomly assigned, the probability of getting a single gigantic component jumps from zero to one as E/V increases beyond the critical value 0.5. However, Erdős and Rényi supposed undirected graphs, in contrast to our assumption of *directed* trust relationships.

Massa [19] conducted experiments on top of the well-known Epinions rating community (*http://www.epinions.com*), revealing that "trust-aware techniques can produce trust scores for very high numbers of peers". Neighborhood formation thus becomes facile to achieve when considering reachability of nodes via trust paths.

[3] Watts and Strogatz [33] have shown that social networks exhibit diverse "small-world" properties making them different from random graphs, such as high clustering coefficients $C(p)$. Barabási and Albert [2] have investigated further distinctive features, such as the scale-free nature of social networks, not present in random graphs. Even though, the above-mentioned theorem holds for random graphs and social networks alike.

Trust Propagation Models. Trust-based neighborhood detection for a_i, using those "trust-aware techniques" mentioned by Massa, implies *deriving* trust values for peers a_j not directly trusted by a_i, but one of the persons the latter agent trusts directly or indirectly. The trust network's high connectivity allows assembling top-M trusted neighborhoods with potentially large M.

Numerous scalar metrics [4,17] have been proposed for computing trust between two given individuals a_i and a_j. We hereby denote computed trust weights by $t_i^c(a_j)$ as opposed to explicit trust $t_i(a_j)$. However, our approach requires metrics that compute top-M *nearest trust neighbors*, and not evaluate trust values for any two given agents. We hence opt for local group trust metrics [35], which have only been attracting marginal interest until now. The most important and most well-known local group trust metric is Levien's Advogato metric [17]. However, the metric can only make *boolean* decisions with respect to trustworthiness, simply classifying agents into trusted and untrusted ones.

Appleseed [35], our own proposal for local group trust computation, allows more fine-grained analysis, assigning continuous trust weights for peers within trust computation range. Rankings thus become feasible. Appleseed's principal concepts derive from spreading activation models [27], which have been conceived for modelling human semantic memory, and random graph walk theory, similar to Brin's famous PageRank approach [25]. Appleseed operates on partial trust graph information, exploring the social network within predefined ranges only and allowing the neighborhood detection process to retain scalability. Hereby, high ranks are accorded to trustworthy peers, i.e., those agents which are largely trusted by others with high trustworthiness. These ranks are used later on for selecting agents deemed suitable for making recommendations.

4.3 Measuring User Similarity and Product-User Relevance

Trust allows selecting peers with overall above-average interest similarity [34]. However, for each active user a_i, some highly trusted peers a_j having completely opposed interests generally exist. The proposition that interpersonal attraction, and hence trust, implies attitudinal similarity does not always hold true. Supplementary filtering, preferably content-based, e.g., considering a_i's major fields of interest, thus becomes indispensable.

For this purpose, we apply taxonomy-driven methods to likewise compute user similarity $c(a_i, a_j)$ and product-user relevance $c_b(a_i, b_k)$ [34,37]. We designed these metrics with decentralized scenarios in mind: in suchlike settings, common filtering metrics based upon rating vector similarity [31,5] tend to fail [19], owing to information sparseness implied by virtually unconstrained product sets and sparse, largely implicit, rating information. Subsequent sections briefly summarize the principle ideas of taxonomy-driven filtering.

Profile Generation. In contrast to generic feature-based filtering, product categories still play an important role, but we have them arranged in a taxonomy and not separate from each other. Products b_k bear topic descriptors $d_{k_e} \in f(b_k)$ that relate these b_k to taxonomic nodes. Several classifications per product are possible, hence $|f(b_k)| \geq 1$. Each product liked by the user infers some interest score for those $d_{k_e} \in f(b_k)$. Since these categories d_{k_e} are arranged in a taxonomy, C, we can also infer a fractional interest for all *super-topics* of d_{k_e}. Hereby, remote super-topics are accorded less interest score than super-topics close to d_{k_e}.

Assume that $(p_0, p_1, \ldots, p_q)$ gives the taxonomic path from top element $p_0 = \top$ to node $p_q = d_{k_e}$. Function $\mathrm{sib}(p)$ returns the number of p's siblings, while $\mathrm{sco}(p)$ returns p's score:

$$\forall m \in \{0, 1, \ldots, q-1\} : \mathrm{sco}(p_m) = \kappa \cdot \frac{\mathrm{sco}(p_{m+1})}{\mathrm{sib}(p_{m+1}) + 1} \tag{3}$$

Similar to Sarwar's framework for common collaborative filtering techniques [28], scores are normalized, i.e., all topic score that a_i's profile assigns to nodes from taxonomy C amounts to some fixed value s. Hence, high product ratings from agents with short product rating histories have higher impact on profile generation than product ratings from persons issuing rife ratings. Score s is divided evenly among all products that contribute to a_i's profile makeup. Factor κ permits fine-tuning the extent of super-topic score inference, depending on the underlying taxonomy's depth and granularity. Figure 3 demonstrates the assignment of score for three topic descriptors, accorded score $s_i = 10, i \in \{1, 2, 3\}$ each.

By virtue of inference of fractional interest for super-topics, one may establish high user similarity for users which have not even rated *one single* product in common. According to our scheme, the more score two profiles have accumulated in same branches, the higher their computed similarity.

Similarity Computation. Taxonomy-driven interest profiles form the grounding for our novel filtering paradigm. Similarity computation between agents a_i, a_j, and between agents a_i and products b_k[4], respectively, requires some distance metric. For our approach, we apply common nearest-neighbor techniques, namely Pearson correlation [8,31] and cosine distance known from information retrieval. Hereby, profile vectors map *category* score vectors from C instead of plain product-rating vectors. For users a_i and a_j with profiles $\boldsymbol{v_i}$ and $\boldsymbol{v_j} \in [0, s]^{|D|}$, respectively, Pearson correlation is defined as below:

$$c(a_i, a_j) = \frac{\sum_{k=0}^{|D|} (v_{i_k} - \overline{v_i}) \cdot (v_{j_k} - \overline{v_j})}{\sqrt{\sum_{k=0}^{|D|} (v_{i_k} - \overline{v_i})^2 \cdot \sum_{k=0}^{|D|} (v_{j_k} - \overline{v_j})^2}} \tag{4}$$

Hereby, $\overline{v_i}$ and $\overline{v_j}$ give mean values for vectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. In our case, because of profile score normalization, both are identical, i.e., $\overline{v_i} = \overline{v_j} = s \,/\, |D|$. Values for $c(a_i, a_j)$ range from -1 to $+1$, where negative values indicate negative correlation, and positive values positive correlation, respectively.

[4] Supposing implicit product ratings, the generation of taxonomy-driven profiles for products b_k equates profile generation for pseudo-user a_θ having implicitly rated b_k only.

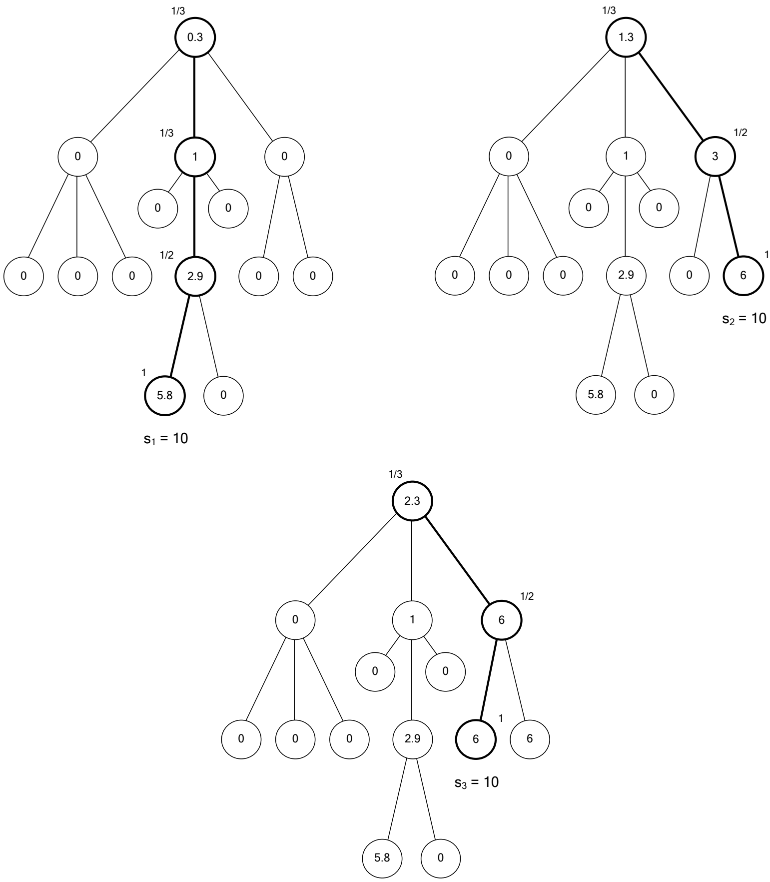

Fig. 3. Assigning three topic descriptors with overall profile score $s = 30$

4.4 Recommendation Generation

Candidate recommendation products b_k for the active user a_i are taken from the set of products that a_i's top-M neighbors have implicitly rated, discounting those products that a_i already knows. We hence obtain set B_i of candidate products. Next, all $b_k \in B_i$ need to be weighted according to their *relevance* for a_i. Relevance $w_i(b_k)$ hereby depends on two factors:

- **Accorded trust $t_i^c(a_j)$ of peers a_j mentioning b_k.** Trust-based neighborhood formation supersedes finding nearest neighbors based upon interest similarity. Likewise, similarity ranks $c(a_i, a_j)$ become substituted by trust weights $t_i^c(a_j)$ for computing the predicted relevance of a_j for a_i.

– **Content-based relevance $c_b(a_i, b_k)$ of product b_k for user a_i.** Besides mere trustworthiness of peers a_j rating product b_k, the content-based relevance of b_k for the active user a_i is likewise important. For example, one may consider the situation where even close friends recommend products not fitting our interest profile at all.

We then define relevance $w_i(b_k)$ for the active user a_i as follows:

$$w_i(b_k) = \frac{q \cdot c_b(a_i, b_k) \cdot \sum_{a_j \in A_i(b_k)} \rho(a_i, a_j)}{|A_i(b_k)| + \Upsilon_R}, \tag{5}$$

where

$$A_i(b_k) = \{a_j \in \text{clique}(a_i) \mid r_j(b_k) \neq \bot\}$$

and

$$q = (1.0 + |f(b_k)| \cdot \Gamma_T)$$

Hereby, $\text{clique}(a_i)$ denotes $a_i's$ neighborhood, Γ_T and Υ_R represent fine-tuning parameters: large Υ_R makes popular items acquire particularly high relevance weight. Factor Υ_R rewards topics bearing extensive content descriptions, i.e., large $|f(b_k)|$. Function $\rho(a_i, a_j)$ gives a_j's significance for a_i. Refer to [37] for extensive discussions of parameterizations and their effects. Therein, the above framework has been presented in context of similarity-based hybrid filtering, i.e., $\rho(a_i, a_j) := c(a_i, a_j)$.

Since we now suppose trust-based neighborhoods, $\rho(a_i, a_j) := t_i^c(a_j)$ holds.

5 Empirical Analysis

The following sections present empirical results obtained from evaluating our trust-based approach for decentralized social filtering. Hereby, we gathered information from an online community featuring both trust network information and product rating data. Our analysis mainly focused on pinpointing the impact that latent information kept within the trust network, namely positive correlation between interpersonal trust and attitudinal similarity [34], may have on recommendation quality. We performed empirical offline evaluations applying metrics well-known from information retrieval, e.g., precision, recall and Breese score [5].

5.1 Dataset Acquisition

Currently, few online communities suit requirements articulated in Section 4.1, i.e., are able to provide both trust *and* product rating information. To our best knowledge, Epinions (*http://www.epinions.com*) and All Consuming count among the only prospective candidates. Epinion's major drawbacks are twofold: first, owing to an immense product range diversity, most ratable products lack content meta-information. Taxonomy-based filtering thus becomes unfeasible. Second, rating information sparseness is beyond measure. For instance, Massa pointed out that only 8.34% of all ratable products have 10 or more reviews.

We therefore opted for the All Consuming community, which has its product range thoroughly confined to the domain of books. Required taxonomic background knowledge C, along with descriptors $f(b_k)$ for virtually all English books b_k, were mined from Amazon.com's Web pages and Web services.

The All Consuming dataset crawl, launched on May 10, 2004, offers information about $3,441$ users, mentioning $10,031$ distinct book titles in $15,862$ implicit book ratings. The accompanying trust network consists of $4,282$ links. Both book and trust ratings are boolean, i.e., non-quantifiable with respect to the extent of appreciation and confidence, respectively. Consequently, book ratings express *full appreciation* and trust statements express *full trust* only. Amazon.com's book taxonomy contains $15,525$ distinct topics, each connected to the root node via one single unique topic chain. For $9,374$ of all $10,031$ books, $31,157$ descriptors pointing to Amazon.com's book taxonomy were found. Book ratings referring to one of those $6,55\%$ books not having valid taxonomic content descriptors were discarded.

One can see that using the All Consuming dataset only partially exploits functionalities our trust-based recommender system is able to unfold. For instance, our Appleseed trust metric [35] has been conceived with continuous trust and distrust statements in mind, whereas All Consuming only offers statements of *full trust*.

5.2 Evaluation Framework

The principal objective of our evaluation was to match the trust-based neighborhood formation scheme against other, more common approaches. Hereby, all benchmark systems were devised according to the same algorithmic clockwork, based upon the recommendation generation framework defined in Equation 4.3. Their only difference pertains to the kind of neighborhood formation, depending on function $\rho(a_i, a_j)$, which identifies the *relevance* of peers a_j for the active user a_i. The following list gives all recommender setups used for experimentation:

- **Trust-based recommender.** Filtering based on trust defines $\rho(a_i, a_j) := t_i^c(a_j)$, as indicated before in Section 4.4. Trust ranks are computed by applying the Appleseed group trust metric [35], thus assigning more weight to products recommended by highly trusted peers.
- **Advanced hybrid approach.** Hybrid filtering likewise exploits content-driven and collaborative filtering facilities. Designed to eliminate intrinsic drawbacks of both mentioned types, this approach currently represents the most promising paradigm for crafting superior recommender systems. The hybrid recommender we propose features *similarity-based* neighborhood formation, requiring $\rho(a_i, a_j) := c(a_i, a_j)$. Since metric $c(a_i, a_j)$ computes the proximity between users a_i, a_j according to purely *content-based information* about products that a_i and a_j have rated, our scheme well complies with Pazzani's "collaboration via content" approach [26]. In studies conducted prior to this work, we tested our hybrid technique and substantiated its superior performance over common benchmark recommender systems [37]. However, note that its applicability is largely restricted to *centralized* scenarios only, necessitating similarity computations $c(a_i, a_j)$ for all pairs $(a_i, a_j) \in A \times A$.

– **Purely content-based filter.** Purely content-driven recommender systems ignore aspects of collaboration among peers and focus on content-based information only. We simulate one suchlike recommender by supposing $\rho(a_i, a_j) := \mathrm{rnd}_{[0,1]}(a_j)$, where function $\mathrm{rnd}_{[0,1]} : A \times A \rightarrow [0, 1]$ randomly assigns relevance weights to pairs of agents. Neighborhood formation thus amounts to an arbitrary sampling of users, devoid of meaningful similarity criteria. Discarding collaboration, recommendations generated are not subject to mere random, though. They rather depend on product features, i.e., measure $c_b(a_i, b_k)$. Hence this recommender's purely content-based nature.

Past efforts have shown that intelligent hybrid approaches tend to outperform purely content-based ones [13,26]. We are particularly interested in beneficial ramifications resulting from trust-based neighborhood formation as opposed to random neighborhoods. Supposing that latent semantic information about interpersonal trust and its positive association with attitudinal similarity, endogenous to the very network, has forged sufficiently strong bonds, we conjecture that the overall recommendation quality of our trust-based approach surpasses filtering based upon content only.

Setup. The evaluation framework we established intends to compare the "utility" of recommendation lists generated by all three recommenders. Measurement is achieved by applying metrics well-known from information retrieval, i.e., precision and recall, implemented according to Sarwar [29], and Breese's half-life utility metric [11], known as Breese score [5] or weighted recall. Hereby, we borrowed various ideas from machine learning cross-validation methods. First, we selected all users a_i with more than five ratings and discarded those having less, owing to the fact that reasonable recommendations are beyond feasibility for these cases. Moreover, users having low trust connectivity were likewise discounted.

Next, we applied K-folding, dividing every user $a_i's$ implicit ratings $R_i := \{b \in B \mid r_i(b) \neq \perp\}$ into $K = 5$ disjoint "slices" of preferably equal size. Hereby, four randomly chosen slices constitute agent a_i's *training set* R_i^x, thus containing approximately 80% of implicit ratings $b \in R_i$. These ratings then define a_i's profile from which final recommendations are computed. For recommendation generation, a_i's residual slice $(R_i \setminus R_i^x)$ is retained and not used for prediction. This slice, denoted T_i^x, contains about 20% of a_i's ratings and constitutes the *test set*, i.e., those products the recommendation algorithms intend to "guess". For our experiments, we considered *all five* combinations $(R_i^x, T_i^x), 1 \leq x \leq 5$ of user a_i's slices, hence computing five complete recommendation lists for every a_i that suffices the before-mentioned criteria.

Parameterization. For our first experiment, neighborhood formation size was set to $M = 20$, and we provided top-20 recommendations for each active user's training set R_i^x. Proximity between profiles, based upon R_i^x and the original ratings R_j of all other agents a_j, was hereby computed anew for each training set R_i^x of a_i.

In order to promote the impact that collaboration may have on eventual recommendations, we adopted $\Upsilon_R = 2.25$, thus rewarding books occurring frequently in ratings R_j

of the active user a_i's immediate neighborhood. For content-based filtering, this parameter exerts marginal influence only. Moreover, we assumed propagation factor $\kappa = 0.75$, and topic reward $\Gamma_T = 0.1$.

Evaluation Metrics. We adopted evaluation measures similar to precision and recall known from information retrieval. Remember that for some given number of returned items, recall indicates the percentage of relevant items that were returned, and precision gives the percentage of returned items that are relevant.

Sarwar [29] presents some adapted variant of recall, recording the percentage of test set products $b \in T_i^x$ occurring in recommendation list $P_i^x : \{1, 2, \ldots, 20\} \rightarrow B$ with respect to the overall number of test set products $|T_i^x|$:

$$\text{Recall} = 100 \cdot \frac{|T_i^x \cap \Im P_i^x|}{|T_i^x|} \tag{6}$$

Note that $\Im P_i^x$ denotes the *image* of map P_i^x, i.e., all books part of the recommendation list.

Accordingly, precision represents the percentage of test set products $b \in T_i^x$ occurring in P_i^x with respect to the size of the recommendation list:

$$\text{Precision} = 100 \cdot \frac{|T_i^x \cap \Im P_i^x|}{|\Im P_i^x|} \tag{7}$$

Breese [5] refines Sarwar's adaptation of recall by introducing *weighted* recall, or Breese score. Breese hereby proposes that the expected utility of a recommendation list is simply the *probability* of viewing a recommended product that is actually relevant, i.e., taken from the test set, times its utility, which is either 0 or 1 for implicit ratings. Moreover, he posits that each successive product in a list is less likely to be viewed by the active user with *exponential decay*. Parameter α denotes the viewing half-life, which is the number of the product on the list such that there is a 50% chance that the active agent, represented by training set R_i^x, will review that product. Interestingly, when assuming $\alpha = \infty$, Breese score is identical to Sarwar's definition of recall.

In order to obtain "global" metrics, i.e., precision, recall, and Breese score for the entire system and not only one single agent, we averaged the respective metric values for all evaluated users.

5.3 Experiments

We conducted three diverse experiments. The first compares the effects of neighborhood formation on recommendation quality when assuming raters with varying numbers of ratings. The second investigates neighborhood size sensitivity for all three candidate schemes, while the third measures overlap of neighborhoods.

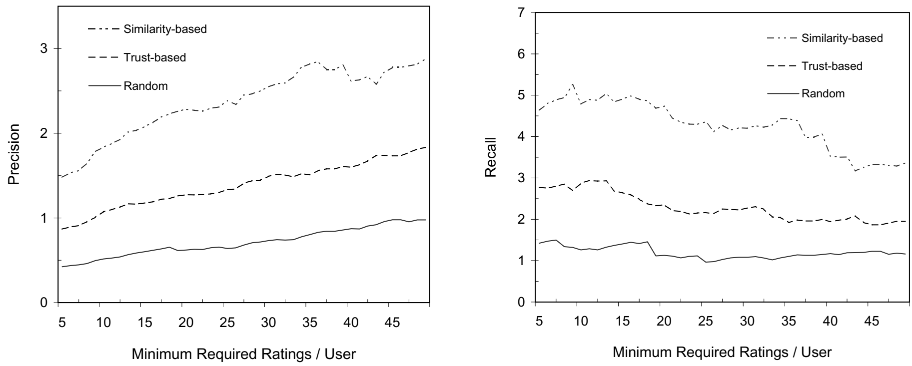

Fig. 4. Unweighted precision and recall, investigating neighborhood formation

Neighborhood Formation Impact. For the first experiment, performance was analyzed by computing unweighted precision and recall (see Figure 4), and Breese score with half-life $\alpha = 5$ and $\alpha = 10$ (see Figure 5). For each indicated chart, the *minimum number* of ratings that users were required to have issued in order to be considered for recommendation generation and evaluation are expressed by the horizontal axis. Since all users with less than five ratings were ignored from the outset, performance evaluations start with all users having at least five ratings. Clearly, larger x-coordinates imply less agents considered for measurement.

Remarkably, all four charts confirm our principal hypothesis that hybrid approaches outperform purely content-based ones. Hence, promoting products that like-minded agents have voted for increases recommendation quality considerably. Next, we observe that our trust-based recommender significantly exceeds its purely content-based counterpart, but cannot reach the hybrid approach's superior score. These results again corroborate our assumption that trust networks contain latent knowledge that reflects attitudinal similarity between trusted agents. Clearly, trust-based neighborhood formation can only *approximate* neighborhoods assembled by means of similarity. However, recall that similarity-based neighborhood formation exhibits poor scalability, owing to its $O(|A|^2)$ complexity that arises from computing proximity measures $c(a_i, a_j)$ for all pairs $(a_i, a_j) \in A \times A$. Trust-based clique formation, on the other hand, does scale and lends itself well for decentralized settings.

– **Precision.** Interestingly, precision (see Figure 4) steadily increases even for content-based filtering. The reason for this phenomenon lies in the very nature of precision: for users a_i with test sets T_i^x smaller than the number $|P_i^x|$ of recommendations received, there is not even a chance of achieving 100% precision.
– **Recall.** Degradation takes place for all curves when increasing x, particularly pronounced for our hybrid recommender. Sample inspections of the All Consuming dataset suggest that infrequent raters favor bestsellers and popular books. Consequently, recommending popular books, promoted by large factor $\Upsilon_R = 2.25$, represents an appropriate guess for that particular type of users. However, when

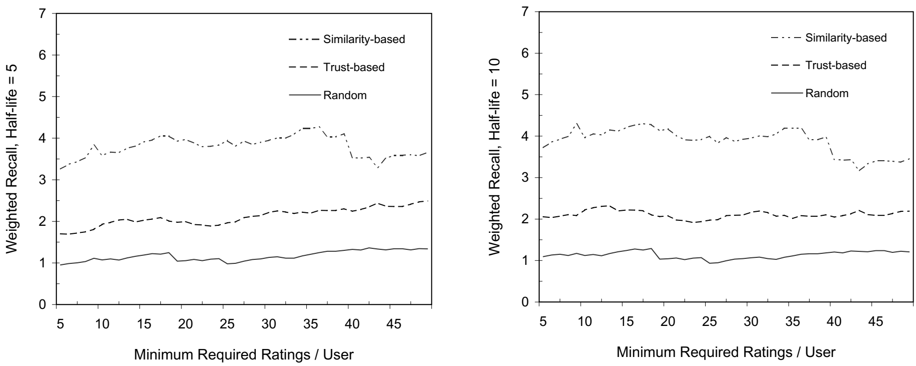

Fig. 5. Weighted recall, using half-life $\alpha \in \{5, 10\}$, for analyzing neighborhood formation

considering users possessing more refined profiles, simple "cherry picking" [11] does not apply anymore.

– **Breese score.** Scores for half-life $\alpha = 5$ and $\alpha = 10$ (see Figure 5) exhibit marginal variance with respect to unweighted recall. However, degradation for increasing x becomes less pronounced when supposing lower α[5], i.e., $\alpha = 10$ and eventually $\alpha = 5$.

As a matter of fact, the above experiment corroborates our hypothesis that trust networks, in contrast to arbitrary connections between agents, bear inherent information about similarity that improves recommendation quality.

Neighborhood Size Sensitivity. The second experiment analyzes the impact of the neighborhood's size on evaluation metrics. Note that we omitted charts for weighted recall, owing to minor deviations from unweighted recall only. Figure 6 indicates scores for precision and recall for increasing neighborhood size $|M|$ along the horizontal axis.

Both charts exhibit similar tendencies for each neighborhood formation scheme. As it comes to similarity-based neighborhood formation, performance of the hybrid approach steadily increases at first. Upon reaching its peak at $|M| = 25$, further increasing neighborhood size $|M|$ does not entail any gains in precision and recall, respectively. This result well aligns with Sarwar's investigations for baseline collaborative filtering techniques [30]. Undergoing slight downward movements between $|M| = 10$ and $|M| = 15$, the content-based scheme's performance curve catches up softly. Basically, increasing the neighborhood size for our content-based filter equates to offering more candidate products[6] and easing "cherry-picking" [11] by virtue of large $\Upsilon_R = 2.25$.

In contrast to both other techniques, the trust-based approach shows comparatively insensitive to increasing neighborhood size $|M|$. As a matter of fact, its performance only

[5] Recall that unweighted recall equates Breese score with $\alpha = \infty$.

[6] Note that only products rated by neighbors are considered for recommendation.

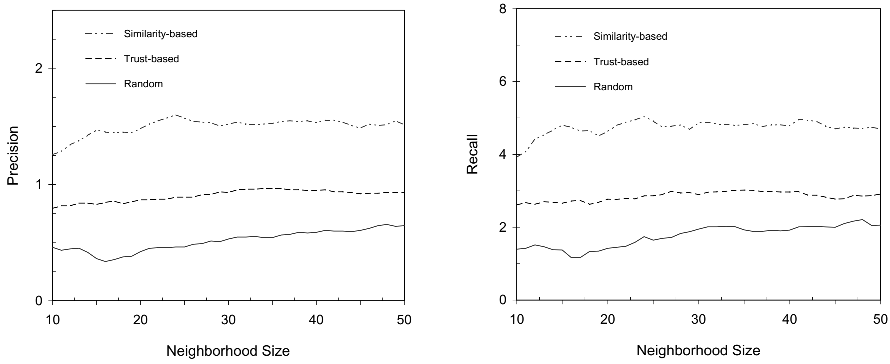

Fig. 6. Unweighted precision and recall for varying neighborhood sizes

marginally improves. We attribute this observation to trust's "conditional transitivity" [1] property and Huang et al.'s investigations of transitive associations for collaborative filtering [12]: exploitation of *transitive* trust relationships, i.e., opinions of friends of friends, only works to a certain extent. However, with increasing network distance from the trust source, these peers do not satisfactorily reflect interest similarity anymore and thus represent weak predictors only. Besides empirical evidence of positive correlation between interpersonal trust and attitudinal similarity, as well as its positive impact on recommendation quality, we regard this aspect as one of the most important findings of our studies.

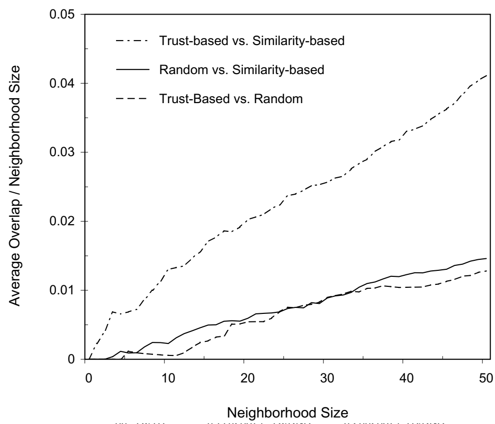

Fig. 7. Unweighted precision and recall for varying neighborhood sizes

Neighborhood Overlap Analysis. Eventually, we compared neighborhoods formed by those three techniques. For any unordered pair $\{p, q\}$ of our three neighborhood formation techniques, we measured the number of agents a_j occurring in *both* x-sized cliques of every active user $a_i \in A$, and normalized the figure by clique size x and the number of agents $|A|$:

$$s^x(\{p, q\}) = \frac{\sum_{a_i \in A} |\,\text{clique}_p^x(a_i) \cap \text{clique}_q^x(a_i)|}{|A| \cdot x} \tag{8}$$

Figure 7 shows all three plots of $s^x(\{p, q\}), x \in [0, 50]$. All curves exhibit tendencies of approximatively linear rise for increasing neighborhood size x, for the probability of overlap rises when neighborhoods become larger. Consequently, supposing clique size $x = |A|$, 100% overlap holds.

As expected, both curves displaying overlap with randomly formed neighborhoods only marginally differ from each other. On the other hand, overlap between trust-based and similarity-based cliques significantly exceeds these two baseline plots, showing that trust-based and similarity-based neighborhoods are considerably more similar to each other than pure random would allow. The above experiment again strongly corroborates our hypothesis that interpersonal trust and attitudinal similarity correlate.

6 Conclusion

In this paper we introduced an approach to exploit trust networks for product recommendation making. Superseding common collaborative approaches with trust-based filtering becomes vital when envisaging *decentralized* recommender system infrastructures, lacking central authorities. With suchlike settings in mind, we issued an advocacy for trust, pointing out several beneficial aspects of this type of relationships with respect to product recommendations. We also devised a new hybrid recommender framework that makes use of trust-based neighborhood formation and taxonomy-driven selection of suitable products.

Moreover, we provided ample empirical evidence to show that network structures emanating from relationships of interpersonal trust, in contrast to random associations between users, exhibit traits of attitudinal similarity which significantly improve recommendation quality. However, we also found that trust's tight coupling with similarity becomes lost when overly exploiting *transitive* relationships.

For our experiments, we used real-world data mined from the All Consuming book reading community which offers both rating *and* trust information about its users. Note that most reputation and rating systems based upon trust models only use synthesized rather than real trust data, therefore allowing largely limited analysis of trust semantics only. However, we would like to base our investigations upon richer datasets in order to make our results more reliable. Unfortunately, few communities currently exist that offer accessible bulk information about both trust relationships and product rating data of its users. We expect this situation to change within the next years to come, owing to an increasing public interest in trust networks, which is particularly promoted by the advent of weblogs and the Semantic Web.

References

1. ABDUL-RAHMAN, A., AND HAILES, S. A distributed trust model. In *New Security Paradigms Workshop* (Cumbria, UK, September 1997), pp. 48–60.
2. BARABÁSI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science 286* (October 1999), 509–512.
3. BERSCHEID, E. Interpersonal attraction. In *The Handbook of Social Psychology*, D. Gilbert, S. Fiske, and G. Lindzey, Eds., 4th ed., vol. II. McGraw-Hill, New York, NY, USA, 1998.
4. BETH, T., BORCHERDING, M., AND KLEIN, B. Valuation of trust in open networks. In *Proceedings of the 1994 European Symposium on Research in Computer Security* (1994), pp. 3–18.
5. BREESE, J., HECKERMAN, D., AND KADIE, C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence* (Madison, WI, USA, July 1998), Morgan Kaufmann, pp. 43–52.
6. ERDŐS, P., AND RÉNYI, A. On random graphs. *Publicationes Mathematicae 5* (1959), 290–297.
7. GOLBECK, J., PARSIA, B., AND HENDLER, J. Trust networks on the Semantic Web. In *Proceedings of Cooperative Intelligent Agents* (Helsinki, Finland, August 2003).
8. GOLDBERG, D., NICHOLS, D., OKI, B., AND TERRY, D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM 35*, 12 (1992), 61–70.
9. GUHA, R. Open rating systems. Tech. rep., Stanford Knowledge Systems Laboratory, Stanford, CA, USA, 2003.
10. HERLOCKER, J., KONSTAN, J., AND RIEDL, J. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, PA, USA, 2000), pp. 241–250.
11. HERLOCKER, J., KONSTAN, J., TERVEEN, L., AND RIEDL, J. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems 22*, 1 (2004), 5–53.
12. HUANG, Z., CHEN, H., AND ZENG, D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems 22*, 1 (2004), 116–142.
13. HUANG, Z., CHUNG, W., ONG, T.-H., AND CHEN, H. A graph-based recommender system for digital library. In *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (Portland, OR, USA, 2002), ACM Press, pp. 65–73.
14. KAUTZ, H., SELMAN, B., AND SHAH, M. Referral Web: Combining social networks and collaborative filtering. *Communications of the ACM 40*, 3 (March 1997), 63–65.
15. KONSTAN, J. Introduction to recommender systems: Algorithms and evaluation. *ACM Transactions on Information Systems 22*, 1 (2004), 1–4.
16. LAM, S., AND RIEDL, J. Shilling recommender systems for fun and profit. In *Proceedings of the 13th Conference on World Wide Web* (New York, NY, USA, 2004), ACM Press, pp. 393–402.
17. LEVIEN, R., AND AIKEN, A. Attack-resistant trust metrics for public key certification. In *Proceedings of the 7th USENIX Security Symposium* (San Antonio, TX, USA, January 1998).
18. MARSH, S. *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling, Stirling, UK, 1994.
19. MASSA, P., AND BHATTACHARJEE, B. Using trust in recommender systems: an experimental analysis. In *Proceedings of the 2nd International Conference on Trust Management* (Oxford, UK, March 2004), C. Jensen, S. Poslad, and T. Dimitrakos, Eds., vol. 2995 of *LNCS*, Springer-Verlag.
20. MIDDLETON, S., ALANI, H., SHADBOLT, N., AND DE ROURE, D. Exploiting synergy between ontologies and recommender systems. In *Proceedings of the WWW2002 International Workshop on the Semantic Web* (Maui, HW, USA, May 2002), vol. 55 of *CEUR Workshop Proceedings*.

21. Montaner, M. *Collaborative Recommender Agents Based on Case-based Reasoning and Trust*. PhD thesis, Universitat de Girona, Girona, Spain, 2003.
22. Mui, L., Szolovits, P., and Ang, C. Collaborative sanctioning: Applications in restaurant recommendations based on reputation. In *Proceedings of the Fifth International Conference on Autonomous Agents* (Montreal, Canada, 2001), ACM Press, pp. 118–119.
23. Olsson, T. Decentralized social filtering based on trust. In *Working Notes of the AAAI-98 Recommender Systems Workshop* (Madison, WI, USA, 1998).
24. O'Mahony, M., Hurley, N., Kushmerick, N., and Silvestre, G. Collaborative recommendation: A robustness analysis. *ACM Transactions on Internet Technology 4*, 3 (August 2004).
25. Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: Bringing order to the Web. Tech. rep., Stanford Digital Library Technologies Project, 1998.
26. Pazzani, M. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review 13*, 5-6 (1999), 393–408.
27. Quillian, R. Semantic memory. In *Semantic Information Processing*, M. Minsky, Ed. MIT Press, Boston, MA, USA, 1968, pp. 227–270.
28. Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM Conference on Electronic Commerce* (Minneapolis, MN, USA, 2000), ACM Press, pp. 158–167.
29. Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Application of dimensionality reduction in recommender systems - a case study. In *ACM WebKDD Workshop* (Boston, MA, USA, August 2000).
30. Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International World Wide Web Conference* (Hong Kong, China, May 2001).
31. Shardanand, U., and Maes, P. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems* (1995), vol. 1, pp. 210–217.
32. Sinha, R., and Swearingen, K. Comparing recommendations made by online systems and friends. In *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries* (Dublin, Ireland, June 2001).
33. Watts, D., and Strogatz, S. Collective dynamics of "small-world" networks. *Nature 393* (June 1998), 440–442.
34. Ziegler, C.-N., and Lausen, G. Analyzing correlation between trust and user similarity in online communities. In *Proceedings of the 2nd International Conference on Trust Management* (Oxford, UK, March 2004), C. Jensen, S. Poslad, and T. Dimitrakos, Eds., vol. 2995 of *LNCS*, Springer-Verlag, pp. 251–265.
35. Ziegler, C.-N., and Lausen, G. Spreading activation models for trust propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service* (Taipei, Taiwan, March 2004), IEEE Computer Society Press.
36. Ziegler, C.-N., Lausen, G., and Schmidt-Thieme, L. Taxonomy-driven computation of product recommendations. In *Proceedings of the 2004 ACM CIKM Conference on Information and Knowledge Management* (Washington D.C., USA, November 2004), ACM Press. To appear.
37. Ziegler, C.-N., Schmidt-Thieme, L., and Lausen, G. Exploiting semantic product descriptions for recommender systems. In *Proceedings of the 2nd ACM SIGIR Semantic Web and Information Retrieval Workshop 2004* (Sheffield, UK, July 2004).