Evaluating Similarity Measures: A Large-Scale Study in the Orkut Social Network

Ellen Spertus
Mills College
5000 MacArthur Blvd.
Oakland, CA 94613
spertus@mills.edu

Mehran Sahami, Orkut Buyukkokten
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
{sahami, orkut}@google.com

ABSTRACT

Online information services have grown too large for users to navigate without the help of automated tools such as collaborative filtering, which makes recommendations to users based on their collective past behavior. While many similarity measures have been proposed and individually evaluated, they have not been evaluated relative to each other in a large real-world environment. We present an extensive empirical comparison of six distinct measures of similarity for recommending online communities to members of the Orkut social network. We determine the usefulness of the different recommendations by actually measuring users' propensity to visit and join recommended communities. We also examine how the ordering of recommendations influenced user selection, as well as interesting social issues that arise in recommending communities within a real social network.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*data mining*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services; I.5 [**Computing Methodologies**]: Pattern Recognition

General Terms

Algorithms, measurement, human factors

Keywords

Data mining, collaborative filtering, recommender system, similarity measure, online communities, social networks

1. INTRODUCTION

The amount of information available online grows far faster than an individual's ability to assimilate it. For example, consider "communities" (user-created discussion groups) within Orkut, a social-networking website (http://www.orkut.com) affiliated with Google. The original mechanisms for users to find communities were labor-intensive, including searching for keywords in community titles and descriptions or browsing other users' memberships. Four months after its January 2004 debut, Orkut had over 50,000 communities, providing the necessity and opportunity for data-mining for automated recommendations. There are now (May 2005) over 1,500,000 communities.

While there are many forms of recommender systems [3], we chose a collaborative filtering approach [13] based on overlapping membership of pairs of communities. We did not make use of semantic information, such as the description of or messages in a community (although this may be an area of future work). Our recommendations were on a per-community, rather than a per-user basis; that is, all members of a given community would see the same recommendations when visiting that community's page. We chose this approach out of the belief, which was confirmed, that community memberships were rich enough to make very useful recommendations without having to perform more computationally intensive operations, such as clustering of users or communities or computing nearest neighbor relations among users. Indeed, Sarwar et al. have found such item-based algorithms to be both more efficient and successful than user-based algorithms [13]. By measuring user acceptance of recommendations, we were able to evaluate the absolute and relative utility of six different similarity measures on a large volume of data.

2. MEASURES OF SIMILARITY

The input data came from the membership relation $\mathcal{M} = \{(u,c) \mid u \in \mathcal{U}, c \in \mathcal{C}\}$, where $\mathcal{C}$ is the set of communities with at least 20 members and U the set of users belonging to at least one such community. When we began our experiment in May 2004, $|\mathcal{C}| = 19,792$, $|\mathcal{U}| = 181,160$, and $|\mathcal{M}| = 2,144,435$. Table 1 summarizes the distribution.

All of our measures of community similarity involve the *overlap* between two communities, i.e., the number of com-

Table 1: Distribution of community memberships

	min	max	median	σ
Users per community	20	9077	50	230.5
Communities per user	1	4173	6	28.0

The definitive version was published in KDD '05, August 21-24, 2005 http://doi.acm.org/10.1145/1081870.1081956

mon users. If a base community b and a (potentially) related community r are considered as sets of users, the overlap is $|B \cap R|$, where we use capital letters to represent the set containing a community's members. Note that overlap cannot be the sole factor in relatedness, as the size of communities varies greatly. If we only considered overlap, practically every community would be considered related to the "Linux" community, which was the most popular, with 9,077 members. The similarity measures in the next section normalize the overlap in different ways.

2.1 Similarity Measure Functions

Each similarity measure we consider is presented as a (possibly asymmetric) function of b and r indicating how appropriate the related community r is as a recommendation for the base community b. We do not use the result of the function as an absolute measure of similarity, only to rank recommendations for a given base community.

2.1.1 L1-Norm

If we consider the base and related communities to be vectors $\overrightarrow{b}$ and $\overrightarrow{r}$, where the i^{th} element of a vector is 1 if user i is a member and 0 if not, we can measure the overlap as the product of their L1-norms:

$$L1(\overrightarrow{b}, \overrightarrow{r}) = \frac{\overrightarrow{b} \cdot \overrightarrow{r}}{\parallel \overrightarrow{b} \parallel_1 \cdot \parallel \overrightarrow{r} \parallel_1}$$

This quantity can also be expressed in set notation, where we use a capital letter to represent the set containing a community's members:

$$L1(B, R) = \frac{|B \cap R|}{|B| \cdot |R|}$$

Note that this evaluates to the overlap between the two groups divided by the product of their sizes. When the base community is held constant (as when we determine the base community's recommendations), this evaluates to the overlap divided by the size of the related community, favoring small communities. Kitts et al. [9] reported this to be a successful measure of similarity in their recommender system.

2.1.2 L2-Norm

Similarly, we can measure the overlap with the product of the L2-norms ("cosine distance" [3, 6, 12]) of $\overrightarrow{b}$ and $\overrightarrow{r}$:

$$L2(\overrightarrow{b}, \overrightarrow{r}) = \frac{\overrightarrow{b} \cdot \overrightarrow{r}}{\parallel \overrightarrow{b} \parallel_2 \cdot \parallel \overrightarrow{r} \parallel_2}$$

In set notation:

$$L2(B, R) = \frac{|B \cap R|}{\sqrt{|B| \cdot |R|}}$$

Note that the square-root in the denominator causes L2 to penalize large communities less severely than L1.

Observe that the L2-norm presented here is equivalent to the widely used cosine coefficient applied to binary data. Moreover, while Pearson correlation has been used previously in recommender systems where ranking data is available, we did not use this measure here since it is generally considered inappropriate for binary data.

2.1.3 Pointwise Mutual-Information: positive correlations (MI1)

Information theory motivates other measures of correlation, such as "mutual information" [2]. We chose pointwise mutual information where we only count "positive" correlations (membership in both B and R). Such a formulation essentially focuses on how membership in one group is predictive of membership in another (without considering how base non-membership in a group effects membership in another group), yielding:

$$MI1(b, r) = \mathrm{P}(r, b) \cdot \lg \frac{\mathrm{P}(r, b)}{\mathrm{P}(r) \cdot \mathrm{P}(b)}$$

2.1.4 Pointwise Mutual-Information: positive and negative correlations (MI2)

Similarly, we can compute the pointwise mutual information with both positive and negative correlations (e.g., membership in both B and R, or non-membership in both groups). Again, we don't compute the full expected mutual information, since we believe cross-correlations (e.g., how membership in B affects non-membership in R) tend to be distortive with the recommendation task since such cross-correlations are plentiful but not very informative. This yields:

$$MI2(b, r) = \mathrm{P}(r, b) \cdot \lg \frac{\mathrm{P}(r, b)}{\mathrm{P}(r) \cdot \mathrm{P}(b)} + \mathrm{P}(\bar{r}, \bar{b}) \cdot \lg \frac{\mathrm{P}(\bar{r}, \bar{b})}{\mathrm{P}(\bar{r}) \cdot \mathrm{P}(\bar{b})}$$

2.1.5 Salton (IDF)

Salton proposed a measure of similarity based on inverse document frequency scaling (tf-idf) [12]:

$$IDF(b, r) = P(r|b) \cdot (-\lg \mathrm{P}(r))$$

$$IDF(B, R) = \frac{|B \cap R|}{|B|} \cdot (-\lg \frac{|R|}{|\mathcal{U}|})$$

2.1.6 Log-Odds

We first considered the standard log-odds function, which measures the relative likelihood that presence or absence in a base community predicts membership in a related community:

$$LogOdds0(b, r) = \lg \frac{\mathrm{P}(r|b)}{\mathrm{P}(r|\bar{b})}$$

Empirically, we found this generated the exact same rankings as using the L1-Norm, which makes sense because:

1. Logarithm is monotonic and, while affecting scores, does not affect rankings.
2. Constant factors, such as $|B|$, do not affect rankings.
3. For $|B| \ll |U|$, $\mathrm{P}(r|\bar{b}) \approx \mathrm{P}(r)$

We formulated a different log-odds metric, which measures whether membership in the base community is likelier to predict membership or absence in the related community:

$$LogOdds(b, r) = \lg \frac{\mathrm{P}(r|b)}{\mathrm{P}(\bar{r}|b)}$$

Table 2: Average size of top-ranked community for each measure

measure	Average size		
	rank 1	rank 2	rank 3
L1	332	482	571
L2	460	618	694
MI1	903	931	998
MI2	966	1003	1077
IDF	923	985	1074
LogOdds	357	513	598

Table 3: Agreement in top-ranked results between measures. For example, MI1 and IDF rank the same related community first for 98% of base communities. Correlations greater than 85% are in bold.

	L1	**L2**	**MI1**	**MI2**	**IDF**
L2	.70				
MI1	.41	.60			
MI2	.39	.57	**.96**		
IDF	.41	.59	**.98**	**.97**	
LogOdds	**.88**	.79	.46	.44	.46

2.2 Discussion

For a given measure, we refer to the related community yielding the highest value to be the *top-ranked* related community relative to a base community. The average size of top-ranked communities for each measure, which varies greatly, is shown in Table 2. Table 3 shows how often two functions yield the same top-ranking result. Table 4 shows the top recommendations for the "I love wine" community. Note that MI1, MI2, and IDF favor very large communities, while L1 and LogOdds favor small communities.

Note that in addition to the obvious correlations between the two mutual-information functions (96%), there is a very strong correlation between IDF and the mutual-information functions (97-98%). Manipulation of the formulas for MI1 and IDF shows:

$$\begin{aligned} MI1(b,r) &= \mathrm{P}(r,b)\cdot \lg \frac{\mathrm{P}(r,b)}{\mathrm{P}(r)\cdot \mathrm{P}(b)} \\ &= \mathrm{P}(r|b){\cdot}\mathrm{P}(b)\cdot \lg \mathrm{P}(r|b) - \mathrm{P}(r|b){\cdot}\mathrm{P}(b)\cdot \lg \mathrm{P}(r) \\ &= \mathrm{P}(r|b){\cdot}\mathrm{P}(b)\cdot \lg \mathrm{P}(r|b) \\ &\quad -\mathrm{P}(r|b)\cdot [1-\mathrm{P}(\bar{b})]\cdot \lg \mathrm{P}(r) \\ &= \mathrm{P}(r|b){\cdot}[\mathrm{P}(b)\cdot \lg \mathrm{P}(r|b) + \mathrm{P}(r|b){\cdot}\mathrm{P}(\bar{b})\cdot \lg \mathrm{P}(r)] \\ &\quad -\mathrm{P}(r|b)\cdot \lg \mathrm{P}(r) \end{aligned}$$

Substituting $IDF(b,r) = -P(r|b)\cdot \lg \mathrm{P}(r)$, we get:

$$\begin{aligned} MI1(b,r) &= \mathrm{P}(r|b)\cdot \left[\mathrm{P}(b)\cdot \lg \mathrm{P}(r|b) + \mathrm{P}(\bar{b})\cdot \lg \mathrm{P}(r)\right] \\ &\quad +IDF(b,r) \end{aligned}$$

Since for virtually all communities b, $P(b) \ll P(\bar{b})$, we can approximate:

$$MI1(b,r) \approx IDF(b,r) + \mathrm{P}(r|b)\cdot \mathrm{P}(\bar{b})\cdot \lg \mathrm{P}(r)$$

Thus, MI1 yields a ranking that can be thought of as starting with the ranking of IDF and perturbing the score of each element in the ranking by $\mathrm{P}(r|b)\cdot \mathrm{P}(\bar{b})\cdot \lg \mathrm{P}(r)$, which generally is not great enough to change the relative ranking of the top scores, leading to MI1 and IDF often giving the same ranking to top-scoring communities. (Note that this perturbation quantity is given only to explain the high correlation between MI1 and IDF. Statistically, it is meaningless, since b and $\bar{b}$ cannot simultaneously hold.)

3. EXPERIMENT DESIGN

We designed an experiment to determine the relative value of the recommendations produced by each similarity measure. This involved interleaving different pairs of recommendations and tracking user clicks. Specifically, we measured the efficacy of different similarity measures using pair-wise binomial sign tests on click-through data rather than using traditional supervised learning measures such as precision/recall or accuracy since there is no "true" labeled data for this task (i.e., we do not know what are the correct communities that should be recommended to a user). Rather, we focused on the task of determining which of the similarity measures performs best on a relative performance scale with regard to acceptance by users.

3.1 Combination

When a user viewed a community page, we hashed the combined user and community identifiers to one of 30 values, specifying an ordered pair of similarity measures to compare. Let S and T be the ordered lists of recommendations for the two measures, where $S = (s_1, s_2, \ldots, s_{|S|})$ and $T = (t_1, t_2, \ldots, t_{|T|})$ and $|S| = |T|$. The recommendations of each measure are combined by Joachims' "Combined Ranking" algorithm [7], restated in Figure 1. The resulting list is guaranteed to contain the top k_S and k_T recommendations for each measure, where $k_T \leq k_S \leq k_T + 1$ [7, Theorem 1].

3.2 Measurements

Whenever a user visited a community, two measures were chosen and their recommendations interleaved, as discussed above. This was done in a deterministic manner so that a given user always saw the same recommendations for a given community. To minimize feedback effects, we did not regenerate recommendations after the experiment began.

A user who views a base community (e.g., "I love wine") is either a member (denoted by "M") or non-member (denoted by "n"). (We capitalize "M" but not "n" to make them easier to visually distinguish.) In either case, recommendations are shown. When a user clicks on a recommendation, there are three possibilities: (1) the user is already a member of the recommended community ("M"), (2) the user joins the recommended community ("j"), or (3) the user visits but does not join the recommended community ("n"). The combination of base and related community memberships can be combined in six different ways. For example "M→j" denotes a click where a member of the base community clicks on a recommendation to another community to which she does not belong and joins that community. Traditionally, analyses of recommender systems focus on "M→j", also known informally as "if you like this, you'll like that" or formally as "similarity" or "conversion". "M→n" recommendations are considered distracters, having negative utility, since they waste a user's time with an item not of interest. Before running the experiment, we decided that the measures should be judged on their "M→j" performance.

Other interpretations are possible: "M→n" links could be considered to have positive utility for any of the following

Table 4: Top recommendations for each measure for the "I love wine" community, with each recommended community's overlap with the base community and size. The size of "I love wine" is 2400.

	L1	**L2**	**MI1**	**MI2**	**IDF**	**LogOdds**
1	Ice Wine (Eiswein) (33/51)	Red Wine (208/690)	Japanese Food/Sushi Lovers (370/3206)	Japanese Food/Sushi Lovers (370/3206)	Japanese Food/Sushi Lovers (370/3206)	Japanese Food/Sushi Lovers (370/3206)
2	California Pinot Noir (26/41)	Cheeses of the World (200/675)	Red Wine (208/690)	Red Wine (208/690)	Photography (319/4679)	Photography (319/4679)
3	Winery Visitor - Worldwide (44/74)	I love red wine! (170/510)	Cheeses of the World (200/675)	Cheeses of the World (200/675)	Red Wine (208/690)	Linux (299/9077)

Figure 1: Joachims' "Combine Rankings" algorithm [7]

Input: ordered recommendation lists $S = (s_1, s_2, \ldots, s_{|S|})$ and $T = (t_1, t_2, \ldots, t_{|T|})$ where $|S| = |T|$
Call: **combine** $(S, T, 0, 0, \emptyset)$
Output: combined ordered recommendation list D

combine$(S, T, k_s, k_t, D)\{$
  if $(k_s < |S| \wedge k_t < |T|)$
    if $(k_s = k_t)$ $\{$
      if $(S[k_s + 1] \notin D)\{D := D + S[k_s + 1];\}$
      combine$(S, T, k_s + 1, k_t, D)$;
    $\}$ else $\{$
      if $(T[k_t + 1] \notin D)\{D := D + T[k_t + 1];\}$
      combine$(S, T, k_s, k_t + 1, D)$;
    $\}$
  $\}$
$\}$

Table 5: Clicks on recommendations, by membership status in the base and recommended communities, as counts and as percentages of total clicks. The last column shows the conversion rate, defined as the percentage of non-members clicking on a related community who then joined it ($\frac{j}{n+j}$).

membership in base community	**membership in recommended community**				
	M (member)	n (non-member)	j (join)	**total**	**conversion rate**
M (member): number of clicks	36353	184214	212982	433549	54%
percent of total clicks	4%	20%	24%	48%	
n (non-member): number of clicks	8771	381241	77905	467917	17%
percent of total clicks	1%	42%	9%	52%	
total: number of clicks	45124	565455	290887	901466	34%
percent of total clicks	5%	63%	32%	100%	

reasons:

1. As the user found the link sufficiently interesting to click on, it was of more utility than a link not eliciting a click.

2. The user is genuinely interested in the related community but does not want to proclaim her interest, as membership information is public and some communities focus on taboo or embarrassing topics. For example, a recommendation given for the popular "Chocolate" community is "PMS". Note that this effect is specific to social networks and not, for example, Usenet groups, where the user's list of communities is not revealed to other users.

Similarly, it is unclear how to value clicks from a base community that the user does not belong to. Does an "n→j" click indicate failure, since the base community was not joined by the user, but the recommended community was, indicating a degree of dissimilarity? Or is it of positive utility, since it helped a user find a community of interest? For these reasons, we tracked all clicks, recording the user's membership status in the base and recommended communities for later analysis. (We did not track whether users returned to communities in the future because of the logging overhead that would be required.)

3.3 User Interface

On community pages, our recommendations were provided in a table, each cell of which contained a recommended community's name, optional picture, and link (Figure 2). Recommendations were shown by decreasing rank from left to right, top to bottom, in up to 4 rows of 3. For aesthetic reasons, we only showed entire rows; thus, no recommendations were displayed if there were fewer than 3. We also provided a control that allowed users to send us comments on the recommendations.

4. RESULTS

We analyzed all accesses between July 1, 2004, to July 18, 2004, of users who joined Orkut during that period. The system served 4,106,050 community pages with recommendations, which provides a lower bound on the number of views. (Unfortunately, we could not determine the total number of views due to browser caching.) There were 901,466 clicks on recommendations, 48% by members of the base community, 52% by non-members (Table 5). Clicks to related communities to which the user already belonged were rare, accounting for only 5% of clicks. The most common case was for a non-member of a base community to click through and not join a related community (42%).

We defined *conversion rate* (also called *precision*) as the percentage of non-members who clicked through to a community who then joined it. The conversion rate was three times as high (54%) when the member belonged to the base community (from which the recommendation came) than not (17%).

4.1 Relative performance of different measures

We compared each measure pairwise against every other measure by analyzing clicks of their merged recommendations. If the click was on a recommendation ranked higher by measure L2 than measure L1, for example, we considered it a "win" for L2 and a loss for L1. If both measures ranked it equally, the result was considered to be a tie. Table 6 shows the outcomes of all clicks, with conversions by members ("M→j") and non-members of the base community ("n→j") broken out separately.

We say that a measure *dominates* another if, in their pairwise comparison, the former has more "wins". For example, L2 dominates L1. This definition, combined with the data in Table 6, yielded a total order (to our surprise) among the measures: L2, MI1, MI2, IDF, L1, LogOdds. The same total order occurred if only "n→j" clicks were considered. The order was different if all clicks were considered: L2, L1, MI1, MI2, IDF, LogOdds.

4.2 Conversion rates

There was great variance in conversion rate by recommended community. We examined the 93 recommended communities that were clicked through to more than 1000 times. Unsurprisingly, the ten with the lowest conversion rate all were about sex (e.g., Amateur Porn). Note that members of the base community were far more willing than non-members to join, perhaps because they had already shown their willingness to join a sex-related community. At the other extreme, none of the ten with the highest conversion rate were sexual (e.g., Flamenco). Table 7 provides selected data by each membership combination. Unsurprisingly, for all 93 base communities, members were more likely than non-members to join the recommended community.

4.3 User comments

Users were also able to submit feedback on related communities. Most of the feedback was from users who wanted recommendations added or removed. Some complained about inappropriate recommendations of sexual or political communities, especially if they found the displayed image offensive. A few objected to our generating related community recommendations at all, instead of allowing community creators to specify them. In one case, poor recommendations destroyed a community: The creator of a feminist sexuality community disbanded it both because of the prurient recommendations that appeared on her page and the disruptive new members who joined as a result of recommendations from such communities. We agreed with her that the recommendations were problematic and offered to remove them. While anecdotal, this example illustrates how a recommendation can have unanticipated consequences that cannot be captured in simple statistical measures. (An informal discussion of users' behavior when we allowed them to choose related communities can be found elsewhere [14].)

5. POSITIONAL EFFECTS

During the above experiment, we became curious how the relative placement of recommendations affected users' selections and performed a second experiment.

5.1 Design

After determining that L2 was the best measure of similarity, we recomputed the recommendations and studied the effect of position on click-through. While in our original experiment we displayed up to 12 recommendations in decreasing rank, for this experiment we displayed up to 9 recommendations in *random* order, again ensuring that each

Table 6: The relative performance of each measure in pairwise combination on clicks leading to joins, divided by base community membership status, and on all clicks. Except where numbers appear in italics, the superioriority of one measure over another was statistically significant ($p < .01$) using a binomial sign test [10].

measures		M → j			n → j			all clicks		
		win	equal	loss	win	equal	loss	win	equal	loss
L2	MI1	6899	2977	4993	2600	1073	1853	30664	12277	20332
L2	MI2	6940	2743	5008	2636	1078	1872	31134	11260	19832
L2	IDF	6929	2697	5064	2610	1064	1865	30710	11271	20107
L2	L1	7039	2539	4834	2547	941	1983	28506	13081	23998
L2	LogOdds	8186	1638	4442	2852	564	1655	34954	6664	18631
MI1	MI2	3339	9372	1855	1223	3401	683	14812	37632	7529
MI1	IDF	3431	8854	1891	1139	3288	629	14671	37049	7758
MI1	LogOdds	7099	3546	3341	2514	1213	1193	29837	13869	13921
MI1	L1	6915	1005	6059	2547	407	2338	27786	4308	29418
MI2	IDF	1564	11575	1031	533	4266	359	6003	47885	4490
MI2	LogOdds	6920	3959	3177	2484	1418	598	2881	15308	13188
MI2	L1	6830	950	6419	*2383*	*362*	*2333*	26865	3872	29864
IDF	L1	6799	1006	6304	*2467*	*392*	*2352*	27042	4069	29755
IDF	LogOdds	6691	3804	3096	2452	1378	1085	28224	15013	13330
L1	LogOdds	6730	518	5975	2521	108	2059	31903	2097	24431

Table 7: Conversion rates by status of membership in base community, for communities to which more than 1000 clicks on recommendations occurred.

	member of base community				non-member of base community			
Related community	M→M	M→j	M→j	conversion rate	n→M	n→n	n→M	conversion rate
10 communities with highest conversion rates	583	2273	6984	75%	198	3454	2017	37%
10 communities with lowest conversion rates	326	1984	826	29%	68	26287	472	1.8%
all 93 communities	13524	54415	52614	46%	3488	127819	19007	17%

user always saw the same ordering of recommendations for a given community. By randomizing the position of recommendations, we sought to measure ordering primacy effects in the recommendations as opposed to their ranked quality.

5.2 Results

We measured all 1,279,226 clicks on related community recommendations from September 22, 2004, through October 21, 2004. Table 8 shows the relative likelihood of clicks on each position. When there was only a single row, the middle recommendation was clicked most, followed by the leftmost, then rightmost recommendations, although the differences were not statistically significant. When there were two or three rows, the differences were very significant ($p < .001$), with preferences for higher rows. P-values were computed using a Chi-Squared test comparing the observed click-through rates with a uniform distribution over all positions [10].

6. CONCLUSION AND FUTURE PLANS

Orkut's large number of community memberships and users allowed us to evaluate the relative performance of six different measures of similarity in a large-scale real-world study. We are not aware of any comparable published large-scale experiments. We were surprised that a total order emerged among the similarity measures and that L2 vector normalization showed the best empirical results despite other measures, such as log-odds and pointwise mutual information, which we found more intuitive. For future work, we would like to see how recommendations handpicked by community owners compare.

Just as we can estimate *communities'* similarity through common users, we can estimate *users'* similarity through common community memberships: i.e., user A might be similar to user B because they belong to n of the same communities. It will be interesting to see whether L2 also proves superior in such a domain. We could also take advantage of Orkut's being a social network [8], i.e., containing information on social connections between pairs of users. In addition to considering common community memberships, we could consider distance between users in the "friendship graph". Users close to each other (e.g., friends or friends-of-friends) might be judged more likely to be similar than distant strangers, although some users might prefer the latter type of link, since it would introduce them to someone they would be unlikely to meet otherwise, perhaps from a different country or culture.

Similarly, friendship graph information can be taken into account when making community recommendations, which would require that recommendations be computed on a per-user (or per-clique), rather than per-community, basis. In such a setting, we could make community recommendations based on weighted community overlap vectors where weights are determined based on the graph distances of other community members to a given user. This is a fertile area for future work and yet another example of how the interaction

Figure 2: Displays of recommendations for three different communities

(a) Namorado para o Bulldog (b) IUT Civil Engineering (c) Collaborative Filtering

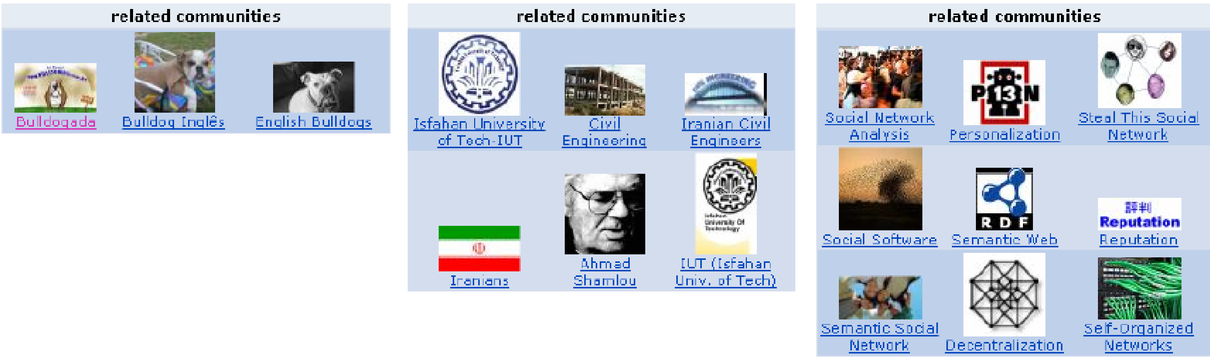

Table 8: The relative likelihood of clicks on link by position when there are (a) one, (b) two, or (c) three rows of three recommendations.

(a) n=28108, p=.12

1.00	1.01	.98

(b) n=24459, p<.001

1.04	1.05	1.08
.97	.94	.92

(c) n=1226659, p<.001

1.11	1.06	1.04
1.01	.97	.99
1.01	.94	.87

of data mining and social networks is becoming an exciting new research area [4] [11].

7. ACKNOWLEDGMENTS

This work was performed while Ellen Spertus was on sabbatical from Mills College and a visiting scientist at Google. We are grateful to Patrick Barry, Alex Drobychev, Jen Fitzpatrick, Julie Jalalpour, Dave Jeske, Katherine Lew, Marissa Mayer, Tom Nielsen, Seva Petrov, Greg Reshko, Catherine Rondeau, Adam Sawyer, Eric Sachs, and Lauren Simpson for their help on the Orkut project and to Corey Anderson, Alex Drobychev, Oren Etzioni, Alan Eustace, Keith Golden, Carrie Grimes, Pedram Keyani, John Lamping, Tom Nielsen, Peter Norvig, Kerry Rodden, Gavin Tachibana, and Yonatan Zunger for their help on this research or its exposition.

8. REFERENCES

[1] Breese, J.; Heckerman, D.; Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (Madison, Wisconsin, 1998). Morgan Kaufmann.

[2] Cover, T.M., and Thomas, J.A. *Elements of Information Theory*. Wiley, New York, 1991.

[3] Deshpande, M., and Karypis, G. Item-Based Top-N Recommendation Algorithms. *ACM Transactions on Information Systems* **22**(1) (January 2004), 143-177.

[4] Domingos, P. Prospects and Challenges for Multi-Relational Data Mining. *ACM SIGKDD Exploration Newsletter* **5**(1) (July 2003).

[5] Dumais, S.; Joachims, T.; Bharat, K.; Weigend, A. SIGIR 2003 Workshop Report: Implicit Measures of User Interests and Preferences. *SIGIR Forum* **37**(2) (Fall 2003).

[6] Harman, D. Ranking Algorithms. In W. B. Frakes and R. Baeza-Yates (ed.), *Information Retrieval: Data Structures & Algorithms* (chapter 14). Upper Saddle River, NJ, USA: Prentice Hall, 1992.

[7] Joachims, T. Evaluating Retrieval Performance Using Clickthrough Data. In *Proceedings of the SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval* (2002). ACM Press, New York, NY.

[8] Kautz, H.; Selman, Bart; Shah, M. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM* **45**(8) (March 1997).

[9] Kitts, B.; Freed, D.; Vrieze, M. Cross-Sell: A Fast Promotion-Tunable Customer-Item Recommendation Method based on Conditionally Independent Probabilities. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Boston, 2000). ACM Press, New York, NY, 437-446.

[10] Lehmann, E.L. Testing Statistical Hypotheses (second edition). Springer-Verlag, 1986.

[11] Raghavan, P. Social Networks and the Web (Invited Talk). In *Advances in Web Intelligence: Proceedings of the Second International Atlantic Web Intelligence Conference*, May 2004. Springer-Verlag, Heidelberg.

[12] Salton, G. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison Wesley, Reading, MA, 1989.

[13] Sarwar, B.; Karypis, G.; Konstan, J.; Reidl, J. Item-Based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the Tenth International Conference on the World Wide Web (WWW10)* (Hong Kong, 2001). ACM Press, New York, NY, 285-295.

[14] Spertus, Ellen. Too Much Information. Orkut Media Selections, January 19, 2005. Available online at "http://media.orkut.com/articles/0078.html".