Modeling Collaborative Similarity with the Signed Resistance Distance Kernel

Jérôme Kunegis, Stephan Schmidt, Şahin Albayrak [1],
Christian Bauckhage [2] and **Martin Mehlitz**
{kunegis, stephan.schmidt, sahin}@dai-lab.de,
christian.bauckhage@telekom.de, martin.mehlitz@googlemail.com

Abstract. We extend the resistance distance kernel to the domain of signed dissimilarity values, and show how it can be applied to collaborative rating prediction. The resistance distance is a graph kernel inspired by electrical network models where edges of a graph are interpreted as electrical resistances. We model the similarity between users of a large collaborative rating database using this signed resistance distance, generalizing the previously known regular resistance distance kernel which is limited to nonnegative values. We show that the signed resistance distance kernel can be computed effectively using the Moore-Penrose pseudoinverse of the Laplacian matrix of the bipartite rating graph, leading to fast computation based on the eigenvalue decomposition of the Laplacian matrix. We apply this technique to collaborative rating prediction on the Netflix Prize corpus, and show how our new kernel can replace the traditional Pearson correlation for rating prediction.

1 Introduction

In the field of information retrieval, the filtering and recommendation of items to users is usually done in a content-based manner, meaning that the content of items is analyzed in order to provide recommendations. Collaborative filtering, by contrast, bases its item rankings on ratings collected from users of the recommendation system.

A collaborative filtering system usually consists of a database of users, items such as text documents or movies, and a collection of ratings users give to items. The collected database of ratings is usually sparse, as each single user has generally only rated a small part of all available items.

To make recommendations, a collaborative filtering system has to rank items. To rank items, a score has to be calculated for each item, based on the profile of the user receiving the recommendation. These scores can be interpreted as rating predictions, meaning that the recommendation system will recommend items the user will probably like.

Different algorithms exist for predicting ratings, most based on the calculation of similarities between users, and sometimes between items. In this paper, we describe a rating prediction algorithm using a new graph kernel based on the signed resistance distance.

[1] DAI-Labor, Technische Universität Berlin, Germany
[2] Telekom Laboratories, Berlin, Germany

The signed resistance distance we define differs from the regular resistance distance in the literature in that in can be applied to similarity measures taking on negative values as well, such as the Pearson correlation. The regular (unsigned) resistance distance can only be used with nonnegative values, and thus cannot be applied to rating prediction, as ratings take on negative values.

Based on a known result about the regular resistance distance, we show how the signed resistance distance can be computed effectively using the Moore-Penrose pseudoinverse of the Laplacian of the correlation matrix.

We evaluate our approach by comparing it to the Pearson correlation based prediction algorithm. The evaluation is performed on the Netflix Prize corpus.

The reminder of this paper is organized as follows. Section 2 introduces the basic collaborative filtering techniques In Section 3, we define the notation used in the paper. Section 4 describes the basic collaborative filtering algorithm in detail. In Section 5, we define the regular (unsigned) resistance distance. Section 6 motivates and defines the signed resistance distance, and presents a closed-form expression for it. Section 7 describes how the signed resistance distance can be applied to rating prediction, giving a signed resistance distance-based algorithm. Section 8 presents the evaluation results, and Section 9 concludes the paper and gives future research directions.

2 Related Work

In this section, we present the basic collaborative filtering methods. Methods based on the resistance distance are introduced later.

Collaborative filtering systems appeared in the 1990s, with systems such as GroupLens [14], Ringo [15], MovieLens, etc. These systems all used a simple neighborhood finding and similarity-based rating prediction approach that consists in calculating a weighted average of known item ratings, taking the user-user correlation values as weights. Variants of these methods can be found in [1,6]. An overview of state-of-the-art collaborative prediction algorithms can be found in [7].

Collaborative filtering algorithm are classified by various methods:

- User-based approaches calculate similarities between users, and use these to weight known item ratings. Item-based

approaches work the other way around, taking user ratings, and weighting them by item-item similarities.
- Implementations accessing the rating database directly are called memory-based, and implementations first building a model and then using this model are called model-based.

The approach developed in this paper is model-based, as a signed resistance distance matrix is precomputed. It can be used in user-based and item-based variants.

Other approaches that have been used in collaborative filtering include other graph-theoretic approaches [10], linear algebraic approaches [8,9], and probabilistic methods [16,17].

The signed resistance has been used before for collaborative filtering [2–5, 13]. In these cases, a multipartite graph was considered, where the different vertex types correspond to the different entities in the collaborative filtering system, such as users and documents. This method will be described in Section 5.

3 Definitions

This section gives the notation used in this paper. Symbols used in one section exclusively are defined there.

$\mathcal{U}$	the set of users
$u \in \mathcal{U}$	a user
$\mathcal{I}$	the set of items
$i \in \mathcal{I}$	an item
$R \in (\mathbb{R} \cup \{\varnothing\})^{\mathcal{U}\times\mathcal{I}}$	the sparse rating matrix
$\varnothing$	denotes a missing rating
$r(u,i)$	a rating in R
$\tilde{r}(u,i)$	a rating prediction for the pair (u,i)
$\mathcal{U}(i) \subseteq \mathcal{U}$	the set of users having rated item i

4 Basic Algorithm

In this section, we describe the basic collaborative rating prediction algorithm based on the Pearson correlation. This algorithm will be adapted to the signed resistance distance in the following sections.

Let $u \in \mathcal{U}$ be a user and $i \in \mathcal{I}$ an item. We want to predict a rating of i for u, so we assume that $r(u,i) = \varnothing$.

A rating prediction can be obtained by taking the average of known ratings for i, weighted by the user-user correlations:

$$\tilde{r}(u,i) \quad = \quad \frac{\sum_{u'\in\mathcal{U}(i)} \operatorname{sim}(u,u')r(u',i)}{\sum_{u'\in\mathcal{U}(i)} \operatorname{sim}(u,u')} \tag{1}$$

Here, $\operatorname{sim}(u,u')$ is a measure of the similarity between users u and u', which we define as the Pearson correlation between these two users ratings:

$$\operatorname{sim}(u,u') \quad = \quad \frac{\sum_i (r(u,i)-\bar{r}(u))(r(u',i)-\bar{r}(u'))}{\sigma_u \sigma_{u'}} \tag{2}$$

The sums are taken over all items both users have rated. σ_u is the standard deviation of u's ratings.

This algorithm can be declined in numerous ways:

- Mirroring the role of users and items results in the item-based equivalent.
- Normalizing each user's ratings usually improves the results. For each user, adjust the ratings linearly to zero mean and unit variance. This approach is motivated by the fact that each user may interpret the rating scale differently, for instance covering situations in which a user only give ratings above the middle value available.
- The Pearson correlation can be multiplied with a confidence value, usually based on the number of ratings the users have in common [7].
- The correlation between users can be based on all items at least one user has rated. In this case, a default value has to be chosen as a default value for items rated by only one user [12].

The next sections will show how $\operatorname{sim}(u,u')$ can be replaced by graph kernels.

5 Resistance Distance

This section motivates the use of the resistance distance and presents a method for computing it.

The resistance distance is a distance function defined on vertices of a graph inspired from electrical resistance networks. When an electrical current is applied to an electrical network of resistors, the whole network acts as a single resistor whose resistance is a function of the individual resistances. In such an electrical network, any two nodes can be taken as the endpoint of the total resistance, giving a function defined between every pair of nodes. As shown in [11], this function is a metric, usually called the *resistance distance*.

Intuitively, two nodes further apart are separated by a greater equivalent resistance, while nodes closer to each other lead to a small resistance distance. This distance function has been used before to perform collaborative filtering [2–4]. As the resistance distance by default only applies to nonnegative values, these previous works use it on nonnegative data, such as document view counts.

The paper [13] uses the same signed resistance distance as this paper, but fails to give a closed-form expression, which is the basis for an efficient calculation of the corresponding similarity matrix.

We now establish a closed-form expression giving the resistance distance between all node pairs based on [11].

Definitions. The following graph-theoretic notation is used.

- $G = (V, E)$ is the complete undirected graph. We will assume it is connected.
- V is the set of graph vertices, corresponding to nodes in the electrical network. $|V| = n$.
- $x, y \in V$ are nodes of G
- E is the set set of graph edges, corresponding to connections in the electrical network. $|E| = m$.
- (x, y) is an individual edge
- $a, b \in V$ are the two endpoints for calculation of the equivalent resistance
- i_{xy} is the current flowing through the edge (a, b). The order of a and b is significant: $i_{ab} = -i_{ba}$.
- $\mathbf{v}_x$ is the electric potential at node x. Potentials are defined up to an additive constant.
- r_{xy} is the resistance value of edge (x, y). Note that $r_{xy} = r_{yx}$.

In electrical networks, the current entering a node must be equal to the current leaving that node. This relation is known

as *Kirchhoff's law*, and can be expressed as $\sum_{(y,x)} i_{xy} = 0$ for all $x \in V - \{a, b\}$. We assume that a current i will be flowing through the network from a to b, and therefore we have

$$\sum_{(y,a)} i_{ay} = i, \quad \sum_{(y,b)} i_{by} = -i$$

Using the identity matrix $I \in [0,1]^{V \times V}$, we express these relations as:

$$\sum_{(y,x)} i_{xy} \quad = \quad i(I_{xa} - I_{xb}) \tag{3}$$

The relation between currents and potentials is given by Ohm's law: $\mathbf{v}_x - \mathbf{v}_y = r_{xy} i_{xy}$ for all edges (x, y).

We define the adjacency matrix $A \in \mathbb{R}^{V \times V}$ to be based on the inverse resistance values, called conductances. Let

$$A_{xy} \quad = \quad \begin{cases} 1/r_{xy} & (x,y) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Let $D \in \mathbb{R}^{V \times V}$ be the diagonal degree matrix, giving the sum of conductances of adjacent edges to each node:

$$D_{xx} \quad = \quad \sum_{(y,x)} 1/r_{xy} \tag{5}$$

Finally the Laplacian matrix is given by

$$L \quad = \quad D - A \tag{6}$$

We will now show that the equivalent resistance $\bar{r}_{ab}$ between a and b is given by:

$$\begin{aligned} \bar{r}_{ab} &= (I_a - I_b) L^+ (I_a - I_b)^T \\ &= L^+_{aa} + L^+_{bb} - L^+_{ab} - L^+_{ba} \end{aligned} \tag{7}$$

where L^+ is the Moore-Penrose pseudoinverse of L [11].

The proof follows from recasting Equation (3) as:

$$\sum_{(y,x)} \frac{1}{r_{xy}} (v_x - v_y) \quad = \quad i(I_{xa} - I_{xb})$$

Combining over all $x \in V$:

$$\begin{aligned} D\mathbf{v} - A\mathbf{v} &= i(I_a - I_b) \\ L\mathbf{v} &= i(I_a - I_b) \end{aligned}$$

Let L^+ be the Moore-Penrose pseudoinverse of L, then because $\mathbf{v}$ is contained in the row space of L [11], we have $L^+ L\mathbf{v} = \mathbf{v}$, and we get

$$\mathbf{v} \quad = \quad L^+ i(I_a - I_b)$$

Which finally gives the equivalent resistance between a and b as

$$\begin{aligned} \tilde{r}_{ab} &= (v_a - v_b)/i \\ &= (I_a - I_b)^T \mathbf{v}/i \\ &= (I_a - I_b)^T L^+ (I_a - I_b) \end{aligned}$$

A symmetry argument shows that $\tilde{r}(a,b) = \tilde{r}(b,a)$ as expected. As proved in [11], $\tilde{r}$ is a metric.

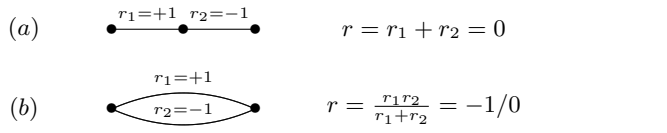

Figure 1. Applying the sum rules to negative resistance values leads to contradictions.

6 Signed Resistance Distance

In this section, we define the requirements for a signed version of the resistance distance, derive a modified node-based expression, and give a proof for a closed-form expression giving the signed resistance distance for all node pairs. Figure 1. shows two examples in which we allow negative resistance values in Equation (7): two parallel edges, and two serial edges. In these examples, we will use the sum rules that hold for electrical resistances: resistances in series add up and conductances in parallel also add up.

Therefore, the constructions of Figure 1. would result in a total resistance of zero for case (a), and an undefined total resistance in case (b). However, the graph of Figure 1 (a) could result from two users a and b having a positive and a negative correlation with a third user c. Intuitively, the resulting distance between a and b should take on a negative value. In the graph of Figure 1 (b), the intuitive result would be $r = -1/2$. What we would like is for the sign and magnitude of the equivalent resistance to be handled separately: The sum rules should hold for the *absolute values* of the resistance similarity values, while the sign should obey a product rule. These requirements are summarized in Figure 2.

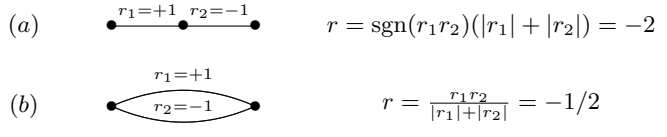

Figure 2. Applying modified sum rules resolves the contradictions.

To achieve the serial sum equation proposed in Figure 2., we propose the following interpretation of a negative resistance:

- An edge carrying a negative resistance value acts like the corresponding positive resistance in series with a component that negates potentials.

A component that negates electric potential cannot exist in physical electrical networks, because it violates an invariant of electrical circuit: The invariant stating that potentials are only defined up to an additive constant. However, as we will see below, the potential inversion gets canceled out in the calculations, yielding results independent of any additive constant. For this reason, we will talk of negative resistances, but avoid the term resistor in this context.

Before giving a closed-form expression for the signed resistance distance, we provide three intuitive examples validating our definition in Figure 3.

- Example (a) shows that, as a path of resistances in series gets longer, the resulting resistance increases. This conditions applies to the regular resistance distance as well as to

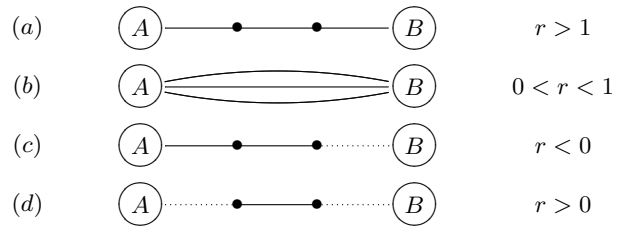

Figure 3. Example configurations of signed resistance values. The total resistance is to be calculated between the nodes A and B. All edges have unit absolute resistance. Edges with negative resistance values are shown as dotted lines. For each case, we formulate a condition that should hold for any signed resistance distance.

the signed resistance distance. In this case, the total resistance should be higher than one.

- Example (b) shows that a higher number of parallel resistances decreases the resulting resistance value. Again, this is true for both types of resistances. In this example, the total resistance should be less than one.
- Examples (c) and (d) show that in a path of signed resistances, the total resistance has the sign of the product of individual resistances. This condition is particular to the signed resistance distance.

We will now show how Kirchhoff's law has to be adapted to support our definition of negative resistances. We adapt Equation (3) by applying the absolute value to the resistance weight.

$$\sum_{(y,x)} \frac{1}{|r_{xy}|}(v_x - \operatorname{sgn}(r_{xy})v_y) \quad = \quad 0 \tag{8}$$

where $\operatorname{sgn}(x)$ denotes the sign function. Defining a modified degree matrix and Laplacian, we get

$$\begin{aligned} \bar{D}_{xx} &= \sum_{(y,x)} |1/r_{xy}| && (9)\\ \bar{L} &= \bar{D} - A && (10)\\ \tilde{r}_{ab} &= (I_a - I_b)\bar{L}^+(I_a - I_b)^T && (11)\\ &= \bar{L}^+_{aa} + \bar{L}^+_{bb} - \bar{L}^+_{ab} - \bar{L}^+_{ba} \end{aligned}$$

The proof follows analogously to the proof for the regular resistance distance by noting that $\mathbf{v}$ is again contained in the row space of $\bar{L}$.

$$\bar{L}^+\bar{L}\mathbf{v} \quad = \quad \mathbf{v}$$

From which the result follows.

As with the regular resistance distance, the signed resistance distance is symmetric: $\tilde{r}(a,b) = \tilde{r}(b,a)$.

7 Implementation

This section describes how the correlation can be replaced by a similarity measure based on the signed resistance distance.

The Pearson correlation $\operatorname{sim}(u,u')$ is a similarity measure in the sense that it is higher when u and u' are more similar. In an electrical network, the resistances intuitively represent dissimilarities: They separate nodes. Therefore, we define $r(a,b) = 1/\operatorname{sim}(a,b)$. A Pearson correlation of zero thus corresponds to an infinite resistance, which corresponds to the absence of a connection between the two nodes.

The adjacency matrix A used for calculating the resistance distance contains the inverted resistance, and is thus equal to the correlation matrix.

The resulting signed resistance distance matrix $\tilde{R}$ contains resistance values. To convert these to similarities, we invert each entry and arrive at a similarity matrix.

$$\begin{aligned} A_{ij} &= \operatorname{sim}(i,j) && (12)\\ \operatorname{sim}_{\text{res}}(i,j) &= 1/\tilde{r}(i,j) && (13) \end{aligned}$$

The kernel $\operatorname{sim}_{\text{res}}(i,j)$ can be used instead of the correlation $\operatorname{sim}(i,j)$ in any algorithm using the Pearson correlation.

For reasons of scalability, the Laplacian cannot be inverted exactly. Instead, we approximate matrix pseudoinversion with the eigenvalue decomposition of the Laplacian $L = QDQ^T$:

$$\begin{aligned} L^+ &= QD^+Q^T && (14)\\ &\approx Q_k D_k^+ Q_k^T && (15) \end{aligned}$$

Where Q_k and D_k denote Q and D truncated to the $k \ll m+n$ biggest eigenvalues and corresponding eigenvectors. D^+ can be computed by inverting non-null elements on the diagonal of D.

8 Evaluation

In this section, we describe the evaluation methodology and show the evaluation results.

We use the corpus of movie ratings available on the Netflix Prize website[3]. The corpus consists of a database of movie ratings given by the customers of the Netflix DVD rental service.

We evaluate the rating prediction algorithms by taking subsets of the corpus randomly, and letting each algorithm predict another disjoint subset of the corpus.

The accuracy of each prediction algorithm is estimated using two algorithms common for evaluating collaborative prediction algorithms [1]: the mean absolute error (MAE) and the root mean squared error (RMSE).

Given the n probe ratings $\bar{r}(u_j, i_j)$ for $1 \le j \le n$, the error measures are defined as:

$$\begin{aligned} \text{MAE} &= \frac{1}{n}\sum_{1\le j\le n} |\bar{r}(u_j,i_j) - \tilde{r}(u_j,i_j)| && (16)\\ \text{RMSE} &= \sqrt{\frac{1}{n}\sum_{1\le j\le n} (\bar{r}(u_j,i_j) - \tilde{r}(u_j,i_j))^2} && (17) \end{aligned}$$

Of these two, the root mean squared error (RMSE) is used on the Netflix Prize website for evaluation of rating prediction algorithms. We check the statistical significance of the results by estimate the error on the calculated measures.

We implemented the following algorithms:

- (CORR) As the baseline, the Pearson correlation-based algorithm.
- (CORR-SHIFT) A variant of the baseline algorithm using a shifted correlation that has proven empirically to provide better predictions on the corpus evaluated.
- (RES-DIST) The weighted mean algorithm using the (unsigned) resistance distance kernel.

[3] `http://www.netflixprize.com/`

- (RES-DIST-SIGNED) The weighted mean algorithm using the signed resistance distance kernel, as defined in this paper.

We show all algorithms in user-based and item-based variants. For the two resistance distance-based variants, we use a dimensionality reduction parameter $k = 100$, giving a good approximation and good runtime.

For all algorithms, we use the same normalization method, shifting each rating by the mean of the average user rating and average item rating. Table 1. shows the evaluation results.

Table 1. Performance of prediction algorithms

	User-based		Item-based	
	MAE	RMSE	MAE	RMSE
CORR	0.769	0.979	0.762	0.991
CORR-SHIFT	0.768	0.978	0.760	0.987
RES-DIST	0.999	1.305	0.990	1.296
RES-DIST-SIGNED	0.770	0.972	0.773	0.983

The evaluation shows the signed resistance distance kernel performing better than the baseline algorithm with regard to the root mean squared error. The (unsigned) resistance distance kernel performs very badly because of the presence of negative ratings. We note that the baseline algorithm achieves a lower mean absolute error.

9 Conclusion and Future Work

We have generalized the resistance distance graph kernel to the domain of negative values. We have used this kernel to implement collaborative rating prediction. By constrast, the related (unsigned) graph kernel can only be applied to non-negative data, such as link prediction and recommendation.

We have evaluated our approach by implementing it on a large rating database. The evaluation showed that using this signed resistance distance improves the accuracy of the prediction algorithm.

As future work, the following questions remain open:

- Other variants of rating prediction algorithms exist take make the results of the correlation algorithm better. They should be evaluated in conjunction with the signed resistance distance kernel.
- How does the signed resistance distance perform in other areas besides collaborative filtering? As it provides a similarity measure, it could be used anywhere a similarity measure such as the Pearson correlation is used.
- How does the signed resistance distance perform on multipartite datasets? Bipartite and tripartite datasets have only been analyzed for the regular resistance distance.

REFERENCES

[1] John S. Breese, David Heckerman, and Carl Kadie, 'Empirical analysis of predictive algorithms for collaborative filtering', in *Proc. Conf. Uncertainty in Artificial Intelligence*, pp. 43–52, (1998).

[2] François Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens, 'Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation', *IEEE Transactions on Knowledge and Data Engineering*, **19**(3), 355–369, (2007).

[3] François Fouss, Alain Pirotte, and Marco Saerens, 'The application of new concepts of dissimilarities between nodes of a graph to collaborative filtering', in *Proc. Workshop on Statistical Approaches for Web Mining*, pp. 26–37, (2004).

[4] François Fouss, Alain Pirotte, and Marco Saerens, 'A novel way of computing similarities between nodes of a graph, with application to collaborative recommendation', in *Proc. Int. Conf. on Web Intelligence*, pp. 550–556, (2005).

[5] François Fouss, Luh Yen, Alain Pirotte, and Marco Saerens, 'An experimental investigation of graph kernels on a collaborative recommendation task', in *Proc. Int. Conf. on Data Mining*, pp. 863–868, (2006).

[6] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl, 'An algorithmic framework for performing collaborative filtering', in *Proc. Int. Conf. on Research and Development in Information Retrieval*, pp. 230–237, (1999).

[7] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl, 'Evaluating collaborative filtering recommender systems', *ACM Trans. Inf. Syst.*, **22**(1), 5–53, (2004).

[8] Thomas Hofmann, 'Collaborative filtering via gaussian probabilistic latent semantic analysis', in *Proc. Int. Conf. on Research and Development in Information Retrieval*, pp. 259–266, (2003).

[9] Thomas Hofmann, 'Latent semantic models for collaborative filtering', *ACM Trans. Inf. Syst.*, **22**(1), 89–115, (2004).

[10] Zan Huang and Daniel D. Zeng, 'Why does collaborative filtering work? – recommendation model validation and selection by analyzing bipartite random graphs', in *Proc. Workshop of Information Technologies and Systems*, (2005).

[11] D. J. Klein and M. Randić, 'Resistance distance', *Journal of Mathematical Chemistry*, **12**(1), 81–95, (1993).

[12] Jérôme Kunegis, Andreas Lommatzsch, Martin Mehlitz, and Şahin Albayrak, 'Assessing the value of unrated items in collaborative filtering', in *Proc. Int. Conf. on Digital Information Management*, pp. 212–216, (2007).

[13] Jérôme Kunegis and Stephan Schmidt, 'Collaborative filtering using electrical resistance network models with negative edges', in *Proc. Industrial Conf. on Data Mining*, pp. 269–282. Springer-Verlag, (2007).

[14] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl, 'GroupLens: An Open Architecture for Collaborative Filtering of Netnews', in *Proc. Conf. on Computer Supported Cooperative Work*, pp. 175–186, (1994).

[15] Upendra Shardanand and Patti Maes, 'Social information filtering: Algorithms for automating "word of mouth"', in *Proc. Conf. on Human Factors in Computing Systems*, volume 1, pp. 210–217, (1995).

[16] L. Ungar and D. Foster, 'A formal statistical approach to collaborative filtering', in *Proc. Conf. on Automated Learning and Discovery*, (1998).

[17] Kai Yu, Anton Schwaighofer, Volker Tresp, Xiaowei Xu, and Hans-Peter Kriegel, 'Probabilistic memory-based collaborative filtering', *Trans. on Knowledge and Data Engineering*, **16**(1), 56–69, (2004).