


Article

Model Description of Similarity-Based Recommendation Systems

Takafumi Kanamori ^{1,2,*}  and Naoya Osugi ³¹ Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan² RIKEN AIP, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan³ Recruit Technologies Co., Ltd., GranTokyo South Tower, 1-9-2 Marunouchi, Chiyoda-ku, Tokyo 100-6640, Japan

* Correspondence: kanamori@c.titech.ac.jp

Received: 4 June 2019; Accepted: 11 July 2019; Published: 17 July 2019



Abstract: The quality of online services highly depends on the accuracy of the recommendations they can provide to users. Researchers have proposed various similarity measures based on the assumption that similar people like or dislike similar items or people, in order to improve the accuracy of their services. Additionally, statistical models, such as the stochastic block models, have been used to understand network structures. In this paper, we discuss the relationship between similarity-based methods and statistical models using the Bernoulli mixture models and the expectation-maximization (EM) algorithm. The Bernoulli mixture model naturally leads to a completely positive matrix as the similarity matrix. We prove that most of the commonly used similarity measures yield completely positive matrices as the similarity matrix. Based on this relationship, we propose an algorithm to transform the similarity matrix to the Bernoulli mixture model. Such a correspondence provides a statistical interpretation to similarity-based methods. Using this algorithm, we conduct numerical experiments using synthetic data and real-world data provided from an online dating site, and report the efficiency of the recommendation system based on the Bernoulli mixture models.

Keywords: recommendation; similarity measures; bernoulli mixture models; completely positive matrix

1. Introduction

In this paper, we study recommendation problems, in particular, the *reciprocal recommendation*. The reciprocal recommendation is regarded as an edge prediction problem of random graphs. For example, a job recruiting service provides preferable matches between companies and job seekers. The corresponding graph is a bipartite graph, where nodes are categorized into two groups: job seekers and companies. Directed edges from one group to the other are the expression of the user's interests. Using this, the job recruiting service recommends unobserved potential matches between users and companies. Another common example is online dating services. Similarly, the corresponding graph is expressed as a bipartite graph with two groups, i.e., males and females. The directed edges are the preference expressions among users. The recommendation system provides potentially preferable partners to each user. The quality of such services depends entirely on the prediction accuracy of the unobserved or newly added edges. The edge prediction has been widely studied as a class of important problems in social networks [1–5].

In recommendation problems, it is often assumed that similar people like or dislike similar items, people, etc. Based on this assumption, researchers have proposed various similarity measures. The similarity is basically defined through the topological structure of the graph that represents the

relationship among users or items. Neighbor-based metrics, path-based metrics, and random walk based metrics are commonly used in this type of analysis. Then, a similarity matrix defined from the similarity measure is used for the recommendation. Another approach is employing the statistical models, such as stochastic block models [6], that are used to estimate network structures, such as clusters or edge distributions. The learning methods using statistical models often achieve high prediction accuracy in comparison to similarity-based methods. Details on this topic are reported in [7] and the references therein.

The main purpose of this paper is to investigate the relationship between similarity-based methods and statistical models. We show that a class of widely applied similarity-based methods can be derived from the Bernoulli mixture models. More precisely, the Bernoulli mixture model with the expectation-maximization (EM) algorithm [8] naturally derives a completely positive matrix [9] as the similarity matrix. The class of completely positive matrices is a subset of doubly nonnegative matrices, i.e., positive semidefinite and element-wise nonnegative matrices [10]. Additionally, we provide an interpretation of completely positive matrices as a statistical model satisfying exchangeability [11–14]. Based on the above argument, we connect the similarity measures using completely positive matrices to the statistical models. First, we prove that most of the commonly used similarity measures yield completely positive matrices as the similarity matrix. Then, we propose an algorithm that transforms the similarity matrix to the Bernoulli mixture model. As a result, we obtain a statistical interpretation of the similarity-based methods through the Bernoulli mixture models. We conduct numerical experiments using synthetic data and real-world data provided from an online dating site, and report the efficiency of the recommendation method based on the Bernoulli mixture models.

Throughout the paper, the following notation is used. Let $[n]$ be $\{1, \dots, n\}$ for a positive integer n . For the matrices A and B , $A \leq B$ denotes the element-wise inequality and $0 \leq A$ denotes that A is entry-wise non-negative. The same notation is used for the comparison of vectors. The Euclidean norm (resp. 1-norm) is denoted as $\|a\|$ (resp. $\|a\|_1$). For the symmetric matrix A , $O \preceq A$ means that A is positive semidefinite.

In this paper, we mainly focus on directed bipartite graphs. The directed bipartite graph $G = (X, Y, E)$ consists of the disjoint sets of nodes, $X = \{x_1, \dots, x_n\}$, $Y = \{y_1, \dots, y_m\}$, and the set of directed edges $E \subset (X \times Y) \cup (Y \times X)$. The sizes of X and Y are n and m , respectively. Using the matrices $A = (a_{ij}) \in \{0, 1\}^{n \times m}$ and $B = (b_{ji}) \in \{0, 1\}^{m \times n}$, the adjacency matrix of G is given by

$$\tilde{A} = \begin{pmatrix} O & A \\ B & O \end{pmatrix} \in \{0, 1\}^{(n+m) \times (n+m)},$$

where $a_{ij} = 1$ (respectively $b_{ji} = 1$) if and only if $(x_i, y_j) \in E$ (respectively $(y_j, x_i) \in E$). For the directed graph, the adjacency matrix \tilde{A} is not necessarily symmetric. In many social networks, each node of the graph corresponds to each user with attributes such as the age, gender, preferences, etc. In this paper, an observed attribute associated with the node $x_i \in X$ (resp. $y_j \in Y$) is expressed by a multi-dimensional vector x_i (resp. y_j). In real-world networks, the attributes are expected to be closely related to the graph structure.

2. Recommendation with Similarity Measures

We introduce similarity measures commonly used in the recommendation. Let us consider the situation that each element in X sends messages to some elements in Y , and vice versa. The messages are expressed as directed edges between X and Y . The observation is, thus, given as a directed bipartite graph $G = (X, Y, E)$. The directed edge between nodes is called the expression of interest (EOI) in the context of the recommendation problems [15]. The purpose is to predict an unseen pair $(x, y) \in X \times Y$ such that these two nodes will send messages to each other. This problem is called the reciprocal recommendation [15–25]. In general, the graph is sparse, i.e., the number of observed edges is much fewer than the number of all possible edges.

In social networks, similar people tend to like and dislike similar people, and are liked and disliked by similar people as studied in [15,26]. Such observations motivated us to define similarity measures. Let $\text{sim}(i, i')$ be a similarity measure between the nodes $x_i, x_{i'} \in X$. In a slight abuse of notation, we write $\text{sim}(j, j')$ to indicate a similarity measure between the nodes $y_j, y_{j'} \in Y$. Based on the observed EOIs, the score of x_i 's interest to $y_j \in Y$ for $i \in [n]$ is defined as

$$\text{score}(i \rightarrow j) = \frac{1}{n} \sum_{i' \in [n]} \text{sim}(i, i') a_{i'j}. \quad (1)$$

If $x_{i'} \in X$ is similar to x_i and the edge $(x_{i'}, y_j)$ exists, the user x_i gets a high score even if $(x_i, y_j) \notin E$. In the reciprocal recommendation, $\text{score}(j \rightarrow i)$ defined by

$$\text{score}(j \rightarrow i) = \frac{1}{m} \sum_{j' \in [m]} \text{sim}(j, j') b_{ji} \quad (2)$$

is also important. The reciprocal score between x_i and y_j , $\text{score}(i \sim j)$, is defined as the harmonic mean of $\text{score}(i \rightarrow j)$ and $\text{score}(j \rightarrow i)$ [15]. This is employed to measure the affinity between x_i and y_j .

Table 1 shows popular similarity measures including graph-based measures and a content-based measure [1]. For the node $x_i \in X$ in the graph $G = (X, Y, E)$, let s_i (resp. \bar{s}_i) be the index set of outer-edges, $\{j | (x_i, y_j) \in E\}$ (resp. in-edges, $\{j | (y_j, x_i) \in E\}$) and $|s|$ be the cardinality of the finite set s . In the following, the similarity measures based on outer-edges are introduced on directed bipartite graphs. The set of outer-edges s_i can be replaced with \bar{s}_i to define the similarity measure based on in-edges.

In graph-based measures, the similarity between the nodes x_i and $x_{i'}$ is defined based on s_i and $s_{i'}$. Some similarity measures depend only on s_i and $s_{i'}$, and others may depend on the whole topological structure of the graph. In Table 1, the first group includes the Common Neighbors, Parameter-Dependent, Jaccard Coefficient, Sørensen Index, Hub Depressed, and Hub Promoted. The similarity measures in this group are locally defined, i.e., $\text{sim}(i, i')$ depends only on s_i and $s_{i'}$. The second group includes SimRank, Adamic-Adar coefficient, and Resource Allocation. They are also defined from the graph structure. However, the similarity between the nodes x_i and $x_{i'}$ depends on the topological structure more than s_i and $s_{i'}$. The third group consists of the content-based similarity, which is defined by the attributes associated with each node.

Below, we supplement the definition of the SimRank and the content-based similarity.

Table 1. Definition of similarity measures $\text{sim}(i, i')$ between the nodes x_i and $x_{i'}$. The right column shows whether the similarity measure is a completely positive similarity kernel (CPSK); see Section 4.

Similarity	Definition/Condition of $S = (\text{sim}(i, i'))$	CPSK
Common neighbors [27]	$ s_i \cap s_{i'} $	✓
Parameter-dependent [28]	$ s_i \cap s_{i'} / (s_i s_{i'})^\lambda, \lambda \geq 0$	✓
Jaccard coefficient [29]	$ s_i \cap s_{i'} / s_i \cup s_{i'} $	✓
Sørensen index [30]	$ s_i \cap s_{i'} / (s_i + s_{i'})$	✓
Hub depressed [31]	$ s_i \cap s_{i'} / \max\{ s_i , s_{i'} \}$	✓
Hub promoted [32]	$ s_i \cap s_{i'} / \min\{ s_i , s_{i'} \}$	×
SimRank [33]	$S = cP^T S P + D, c \in (0, 1)$	✓
Adamic-adar coefficient [34]	$\sum_{k \in s_i \cap s_{i'}} 1 / \log \bar{s}_k $	✓
Resource allocation [31]	$\sum_{k \in s_i \cap s_{i'}} 1 / \bar{s}_k $	✓
Content-based similarity [17]	$\frac{1}{A} \sum_{a \in [A]} \sum_{b \in V_a} \mathbf{1}[(x_i)_a = b, (x_{i'})_a = b]$	✓

SimRank:

SimRank [33] and its reduced variant [35] are determined from the random walk on the graph. Hence, the similarity between the two nodes depends on the whole structure of the graph. For $c \in (0, 1)$, the similarity matrix $\tilde{S} = (\tilde{S}_{ij})_{i,j \in [n+m]}$ on $X \cup Y$ is given as the solution of

$$\tilde{S}_{ii'} = c \sum_{k \in s_i, k' \in s_{i'}} \frac{\tilde{S}_{kk'}}{|s_i| |s_{i'}|}$$

for $i \neq i'$, while the diagonal element \tilde{S}_{ii} is fixed to 1. Let $P \in [0, 1]^{(n+m) \times n+m}$ be the column-normalized adjacency matrix defined from the adjacency matrix of $G = (X, Y, E)$. Then, \tilde{S} satisfies $\tilde{S} = cP^T \tilde{S}P + D$, where D is a diagonal matrix satisfying $1 - c \leq D_{ii} \leq 1$. In the reduced SimRank, D is defined as $(1 - c)I$. For the bipartite graph, the similarity matrix based on the SimRank is given as a block diagonal matrix.

Content-Based Similarity:

In RECON [17,21], the content-based similarity measure is employed. Suppose that $x_i = (x_{ia})_{a \in [Q]} \in \prod_{a \in [Q]} V_a$ is the attributes of the node x_i , where $V_a, a \in [Q]$ are finite sets and $x_{ia} \in V_a$. The continuous variables in the features are appropriately discretized. The similarity measure is defined using the number of shared attributes, i.e.,

$$\text{sim}(i, i') = \frac{1}{Q} \sum_{a \in [Q]} \mathbf{1}[x_{ia} = x_{i'a}] = \frac{1}{Q} \sum_{a \in [Q]} \sum_{b \in V_a} \mathbf{1}[x_{ia} = b] \cdot \mathbf{1}[x_{i'a} = b].$$

In RECON, the score is defined from the normalized similarity, i.e.,

$$\text{score}(i \rightarrow j) = \sum_{j'} \frac{a_{ij'}}{\sum_k a_{ik}} \text{sim}(j', j).$$

The similarity-based recommendation is simple but the theoretical properties have not been sufficiently studied. In the next section, we introduce statistical models and consider the relationship to similarity-based methods.

3. Bernoulli Mixture Models and Similarity-Based Prediction

In this section, we show that the similarity-based methods are derived from Bernoulli mixture models (BMMs). BMMs have been employed in some studies [36–38] for block clustering problems. Here, we show that the BMMs are also useful for recommendation problems.

Suppose that each node belongs to a class $c \in [C]$. Let π_c (respectively π'_c) be the probability that each node in X (respectively Y) belongs to the class $c \in [C]$. We assume that the class at each node is independently drawn from these probability distributions. Though the number of classes, C , can be different in each group, here we suppose that they are the same for simplicity. When the node $x_i \in X$ in the graph belongs to the class c , the occurrence probability of the directed edge from x_i to $y_j \in Y$ is defined by the Bernoulli distribution with the parameter $\alpha_{cj} \in (0, 1)$. As previously mentioned, the adjacency matrix of the graph consists of $A = (a_{ij})$ and $B = (b_{ji})$. We assume that all elements of A and B are independently distributed. For each $x_i \in X$, the probability of $(a_{ij})_{j \in [m]}$ is given by the BMM,

$$\sum_{c \in [C]} \pi_c \prod_{j \in [m]} \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1 - a_{ij}},$$

and the probability of the adjacency submatrix $A = (a_{ij})$ is given by

$$P(A) = \prod_{i \in [n]} \sum_{c \in [C]} \pi_c \prod_{j \in [m]} \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}. \quad (3)$$

In the same way, the probability of the adjacency submatrix B is given by

$$P(B) = \prod_{j \in [m]} \sum_{c \in [C]} \pi'_c \prod_{i \in [n]} \beta_{ci}^{b_{ji}} (1 - \beta_{ci})^{1-b_{ji}}, \quad (4)$$

where $\beta_{ci} \in (0, 1)$ is the parameter of the Bernoulli distribution. Hence, the probability of the whole adjacency matrix \tilde{A} is given by

$$P(\tilde{A}; \Psi) = P(A)P(B), \quad (5)$$

where Ψ is the set of all parameters in the BMM, i.e., $\pi_c, \pi'_c, \alpha_{cj}$ and β_{ci} for $i \in [n], j \in [m]$ and $c \in [C]$. One can introduce the prior distribution on the parameter α_{cj} and β_{ci} . The beta distribution is commonly used as the conjugate prior to the Bernoulli distribution.

The parameter is estimated by maximizing the likelihood for the observed adjacency matrix \tilde{A} . The probability $P(\tilde{A}; \Psi)$ is decomposed into two probabilities, $P(A)$ and $P(B)$, which do not share the parameters. In fact, $P(A)$ depends only on π_c and α_{cj} and $P(B)$ depends only on π'_c and β_{ci} . In the following, we consider the parameter estimation of $P(A)$. The same procedure works for the estimation of the parameters in $P(B)$.

The expectation-maximization (EM) algorithm [8] can be used to calculate the maximum likelihood estimator. The auxiliary variables used in the EM algorithm have an important role in connecting the BMM with the similarity-based recommendation methods. Using the Jensen's inequality, we find that the log-likelihood $\log P(A)$ is bounded below as

$$\begin{aligned} \log P(A) &= \sum_i \log \sum_c r(c|i) \frac{\pi_c \prod_{j \in Y} \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}}{r(c|i)} \\ &\geq J(r, \Psi; A) := \sum_{i,c} r(c|i) \log \frac{\pi_c}{r(c|i)} + \sum_{i,j,c} r(c|i) \log \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}, \end{aligned} \quad (6)$$

where the parameter $r = (r(c|i))_{c,i}$ is positive auxiliary variables satisfying $\sum_{c \in [C]} r(c|i) = 1$. In the above inequality, the equality holds when $r(c|i)$ is proportional to $\pi_c \prod_j \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}$. The auxiliary variable $r(c|i)$ is regarded as the class probability of $x_i \in X$ when the adjacency matrix is observed.

In the EM algorithm, the lower bound of the log-likelihood, i.e., the function $J(r, \Psi; A)$ in (6) is maximized. For this purpose, the alternating optimization method is used. Firstly the parameter Ψ is optimized for the fixed r , and secondly, the parameter r is optimized for the fixed Ψ . This process is repeatedly conducted until the function value $J(r, \Psi; A)$ converges. Importantly, in each iteration the optimal solution is explicitly obtained. The following is the learning algorithm of the parameters:

$$\text{Step 1: } \begin{cases} \pi_c &\leftarrow \frac{1}{n} \sum_i r(c|i), \\ \alpha_{cj} &\leftarrow \frac{1}{n\pi_c} \sum_i r(c|i) a_{ij}, \end{cases} \quad (7)$$

$$\text{Step 2: } r(c|i) \leftarrow \frac{\pi_c \prod_j \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}}{\sum_c \pi_c \prod_j \alpha_{cj}^{a_{ij}} (1 - \alpha_{cj})^{1-a_{ij}}}. \quad (8)$$

The estimator of the parameter Ψ is obtained by repeating (7) and (8).

Using the auxiliary variables $r(c|i)$, one can naturally define the “occurrence probability” of the edge from x_i to y_j . Here, the occurrence probability is denoted by $\text{score}(i \rightarrow j)$. Note that the auxiliary variable $r(c|i)$ is regarded as the conditional probability that x_i belongs to the class c . If x_i belongs to the class c , the occurrence probability of the edge (x_i, y_j) is α_{cj} . Hence, the occurrence probability of the edge (x_i, y_j) is naturally given by

$$\begin{aligned}\text{score}(i \rightarrow j) &:= \sum_c r(c|i) \alpha_{cj} = \sum_c r(c|i) \frac{1}{n\pi_c} \sum_{i'} r(c|i') a_{i'j} \\ &= \frac{1}{n} \sum_{i'} \text{sim}(i, i') a_{i'j},\end{aligned}\quad (9)$$

where the updated parameter α_{cj} in (7) is substituted. The similarity measure $\text{sim}(i, i')$ on X in the above is defined by

$$\text{sim}(i, i') = \sum_c \frac{r(c|i)r(c|i')}{\pi_c} = \frac{\sum_c \pi_c r(i|c)r(i'|c)}{r(i)r(i')} = \frac{r(i, i')}{r(i)r(i')}, \quad (10)$$

where $r(i) := 1/n$, $r(i|c) := r(c|i)r(i)/\pi_c$ and

$$r(i, i') := \sum_{c \in [C]} \pi_c r(i|c)r(i'|c). \quad (11)$$

The equality $\sum_i r(i, i') = r(i') = 1/n$ holds for r satisfying the update rule (7). The above joint probability $r(i, i')$ clearly satisfies the symmetry, $r(i, i') = r(i', i)$. This property is the special case of the finite exchangeability [11,13]. The exchangeability is related to the de Finetti's theorem [39], and the statistical models with the exchangeability have been used in several problems such as Bayes modeling and classification [12,40,41]. Here, we use the finite exchangeable model for the recommendation systems.

Equation (9) gives an interpretation of the heuristic recommendation methods (1) using similarity measures. Suppose that a similarity measure $\text{sim}(i, i')$ is used for the recommendation. Let us assume that the corresponding similarity matrix $S = (\text{sim}(i, i'))_{i, i' \in [n]}$ is approximately decomposed into the form of the mixture model r in (10), i.e.,

$$S_{ii'} \approx \frac{\sum_{c \in [C]} \pi_c r(i|c)r(i'|c)}{r(i)r(i')}. \quad (12)$$

Then, $\text{score}(i \rightarrow j)$ defined from S is approximately the same as that computed from the Bernoulli mixture model with the parameter Ψ that maximizes $J(r, \Psi; A)$ for the fixed $r(c|i)$ associated with S . On the other hand, the score for the recommendation computed from the Bernoulli mixture uses the maximum likelihood estimator Ψ that attains the maximum value of $J(r, \Psi; A)$ under the optimal auxiliary parameter $r(c|i)$. Hence, we expect that the learning method using the Bernoulli mixture model will achieve higher prediction accuracy as compared to the similarity-based methods, if the Bernoulli mixture model approximates the underlying probability distribution of the observed data.

For $i, i' \in [n]$, the probability function $r(i, i')$ satisfying (11) leads to the $n \times n$ positive semidefinite matrix $(r(i, i'))_{i, i' \in [n]}$ with nonnegative elements. As a result, the ratio $r(i, i')/r(i)r(i')$ is also positive semidefinite with nonnegative elements. Let us consider whether the similarity measures in Table 1 yield the similarity matrix with expression (10). Next, we demonstrate that the commonly used similarity measures meet the assumption (12) under a minor modification.

4. Completely Positive Similarity Kernels

For the set of all n by n symmetric matrices S_n , let us introduce two subsets of S_n ; one is the completely positive matrices and the other is doubly nonnegative matrices. The set of completely positive matrices is defined as $\mathcal{C}_n = \{S \in S_n | \exists N \geq 0 \text{ s.t. } S = NN^T\}$, and the set of doubly nonnegative matrices is defined as $\mathcal{D}_n = \{S \in S_n | S \geq 0, S \succeq O\}$. Though the number of columns of the matrix N in the completely positive matrix is not specified, it can be bounded above by $n(n+1)/2 + 1$. This is because \mathcal{C}_n is expressed as the convex hull of the set of rank one matrices $\{qq^T | q \in \mathbb{R}^n, q \geq 0\}$ as shown in [11]. The Carathéodory's theorem can be applied to prove the assertion. More detailed analysis of the matrix rank for the completely positive matrices is provided by [42]. Clearly, the completely

positive matrix is doubly nonnegative matrix. However, [10] proved that there is a gap between the doubly nonnegative matrix and completely positive matrix when $n \geq 5$.

The similarity measure that yields the doubly nonnegative matrix satisfies the definition of the kernel function [43]. The kernel function is widely applied in machine learning and statistics [43]. Here, we define the completely positive similarity kernel (CPSK) as the similarity measure that leads to the completely positive matrix as the Gram matrix or similarity matrix. We consider whether the similarity measures in Table 1 yield the completely positive matrix. For such similarity measures, the relationship to the BMMs is established via (10).

Lemma 1. (i) Let $B = (b_{ij})$ and $C = (c_{ij})$ be $n \times n$ completely positive matrices. Then, their Hadamard product $B \circ C = (b_{ij}c_{ij})_{i,j \in [n]}$ is completely positive. (ii) Let $\{B_k\} \subset \mathcal{C}_n$ be a sequence of $n \times n$ completely positive matrices and define $B = \lim_{k \rightarrow \infty} B_k$. Then, B is the completely positive matrix.

Proof of Lemma 1. (i) Suppose that $B = FF^T$ and $C = GG^T$ such that $F = (f_{ik}) \in \mathbb{R}^{n \times p}$ and $G = (g_{i\ell}) \in \mathbb{R}^{n \times q}$. Then, $(B \circ C)_{ij} = \sum_k f_{ik}f_{jk} \sum_\ell g_{i\ell}g_{j\ell} = \sum_{k,\ell} (f_{ik}g_{i\ell})(f_{jk}g_{j\ell})$. Hence, the matrix $H = (h_{ij}) \in \mathbb{R}^{n \times pq}$ such that $h_{i,(k-1)q+\ell} = f_{ik}g_{i\ell} \geq 0$ satisfies $B \circ C = HH^T$. (ii) It is clear that \mathcal{C}_n is a closed set. \square

Clearly, the linear sum of the completely positive matrices with non-negative coefficients yields completely positive matrices. Using this fact with the above lemma, we show that all measures in Table 1 except the HP measure are the CPSK. In the following, let $\mathbf{a}_i = (a_{i1}, \dots, a_{im})^T \in \{0, 1\}^m$ for $i \in [n]$ be non-zero binary column vectors, and let A be the matrix $A = (a_{ij}) \in \{0, 1\}^{n \times m}$. The index set s_i is defined as $s_i = \{k \mid a_{ik} = 1\} \subset [m]$.

Common Neighbors

The elements of the similarity matrix are given by

$$S_{ii'} = |s_i \cap s_{i'}| = \mathbf{a}_i^T \mathbf{a}_{i'} \geq 0.$$

Hence, $S = AA^T$ holds. The common neighbors similarity measure yields the CPSK.

Parameter-Dependent:

The elements of the similarity matrix are given by

$$S_{ii'} = \frac{\mathbf{a}_i^T \mathbf{a}_{i'}}{\|\mathbf{a}_i\|_1^\lambda \|\mathbf{a}_{i'}\|_1^\lambda}.$$

Hence, we have $S = DAA^TD^T$, where D is the diagonal matrix whose diagonal elements are $1/\|\mathbf{a}_1\|_1^\lambda, \dots, 1/\|\mathbf{a}_n\|_1^\lambda$. The Parameter-Dependent similarity measure yields the CPSK.

Jaccard Similarity:

We have $|s_i \cap s_{i'}| = \mathbf{a}_i^T \mathbf{a}_{i'}$ and $|s_i \cup s_{i'}| = m - \bar{\mathbf{a}}_i^T \bar{\mathbf{a}}_{i'}$, where $\bar{\mathbf{a}}_i = \mathbf{1} - \mathbf{a}_i$. Hence, the Jaccard similarity matrix $S = (S_{ii'})_{i,i' \in [n]}$ is given by

$$S_{ii'} = \frac{\mathbf{a}_i^T \mathbf{a}_{i'}}{m - \bar{\mathbf{a}}_i^T \bar{\mathbf{a}}_{i'}}.$$

Let us define the matrices S_0 and $T_{(k)}$ respectively by $(S_0)_{ii'} = \mathbf{a}_i^T \mathbf{a}_{i'}/m$ and $(T_{(k)})_{ii'} = (\bar{\mathbf{a}}_i^T \bar{\mathbf{a}}_{i'}/m)^k$. The matrix S is then expressed as $S = S_0 \circ \sum_{k=0}^{\infty} T_{(k)}$. Lemma 1 (i) guarantees that $T_{(k)}$ is the CPSK since $T_{(1)}$ is the CPSK. Hence, the Jaccard similarity measure is the CPSK.

Sørensen Index:

The similarity matrix $S = (S_{ii'})$ based on the Sørensen Index is given as

$$S_{ii'} = \frac{2\mathbf{a}_i^T \mathbf{a}_{i'}}{\|\mathbf{a}_i\|^2 + \|\mathbf{a}_{i'}\|^2} = 2 \sum_{k=1}^m \int_0^\infty a_{ik} e^{-t\|\mathbf{a}_i\|^2} a_{i'k} e^{-t\|\mathbf{a}_{i'}\|^2} dt.$$

The integral part is expressed as the limit of the sum of the rank one matrix, $\sum_\ell \mathbf{b}_k(t_\ell) \mathbf{b}_k^T(t'_\ell) (t_\ell - t_{\ell-1})$, where $t'_\ell \in [t_{\ell-1}, t_\ell]$ and $\mathbf{b}_k(t)$ is the n -dimensional vector defined by $(\mathbf{b}_k(t))_i = a_{ik} e^{-t\|\mathbf{a}_i\|^2} \geq 0$. Hence, the Sørensen index is the CPSK.

Hub Promoted:

The hub promoted similarity measure does not yield the positive semidefinite kernel. Indeed, for the adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

the similarity matrix based on the Hub Promoted similarity measure is given as

$$S = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 1 & 1/2 & 1 \end{pmatrix}.$$

The eigenvalues of S are 1 and $(2 \pm \sqrt{5})/2$. Hence, S is not positive semidefinite.

Hub Depressed:

The similarity matrix is defined as

$$S_{ii'} = \frac{\mathbf{a}_i^T \mathbf{a}_{i'}}{\max\{\|\mathbf{a}_i\|^2, \|\mathbf{a}_{i'}\|^2\}} = \mathbf{a}_i^T \mathbf{a}_{i'} \min\{1/\|\mathbf{a}_i\|^2, 1/\|\mathbf{a}_{i'}\|^2\}.$$

Since the min operation is expressed as the integral $\min\{x, y\} = \int_0^1 \mathbf{1}[t \leq x] \cdot \mathbf{1}[t \leq y] dt$ for $x, y \geq 0$, we have

$$S_{ii'} = \sum_{k=1}^m \int_0^1 a_{ik} \mathbf{1}[t \leq 1/\|\mathbf{a}_i\|^2] \cdot a_{i'k} \mathbf{1}[t \leq 1/\|\mathbf{a}_{i'}\|^2] dt.$$

In the same way as the Sørensen Index, we can prove that the Hub Depressed similarity measure is the CPSK.

SimRank:

The SimRank matrix S satisfies $S = cP^T S P + (1 - c)D$ for $c \in (0, 1)$, where $P \geq 0$ is properly defined from \tilde{A} and D is a diagonal matrix such that the diagonal element d_{ii} satisfies $1 - c \leq d_{ii} \leq 1$. The recursive calculation yields the equality $S = \sum_{k=0}^\infty c^k (P^k)^T D P^k$, meaning that S is the CPSK.

Adamic-Adar Coefficient:

Given the adjacency matrix $A = (a_{ij})$, the similarity matrix $S = (S_{ii'})$ is expressed as

$$S_{ii'} = \sum_{k \in s_i \cap s_{i'}} \frac{1}{\log |s'_k|} = \sum_k \frac{a_{ik} a_{i'k}}{\log \sum_{\ell} a_{\ell k}},$$

where $\frac{a_{ik} a_{i'k}}{\log \sum_{\ell} a_{\ell k}}$ is set to zero if $\sum_{\ell} a_{\ell k} \leq 1$. Hence, we have $S = ADA^T$, where D is the diagonal matrix with the elements $D_{kk} = 1 / \log \sum_{\ell} a_{\ell k}$ for $\sum_{\ell} a_{\ell k} \geq 2$ and $D_{kk} = 0$ otherwise. Since $S = NN^T$ with $N = AD^{1/2}$ holds, the similarity measure based on the Adamic-adar coefficient is the CPSK.

Resource Allocation:

In the same way as the Adamic-adar coefficient, the similarity matrix is given as

$$S_{ii'} = \sum_{k \in s_i \cap s_{i'}} \frac{1}{|s'_k|} = \sum_k \frac{a_{ik} a_{i'k}}{\sum_{\ell} a_{\ell k}},$$

where the term $\frac{a_{ik} a_{i'k}}{\sum_{\ell} a_{\ell k}}$ is set to zero if $\sum_{\ell} a_{\ell k} \leq 1$. We have $S = ADA^T$, where D is the diagonal matrix with the elements $D_{kk} = 1 / \sum_{\ell} a_{\ell k}$ for $\sum_{\ell} a_{\ell k} \geq 2$ and $D_{kk} = 0$ otherwise. Since $S = NN^T$ with $N = AD^{1/2}$ holds, the similarity measure based on resource allocation is the CPSK.

Content-Based Similarity:

The similarity matrix is determined from the feature vector of each node as follows,

$$S_{ii'} = \frac{1}{Q} \sum_{a \in [Q]} \sum_{b \in V_a} \mathbf{1}[x_{ia} = b] \cdot \mathbf{1}[x_{i'a} = b].$$

Clearly, S is expressed as the sum of rank-one matrix $c_{a,b} c_{a,b}^T$, where $(c_{a,b})_i = \mathbf{1}[x_{ia} = b] / \sqrt{Q} \geq 0$. Hence, Content-based similarity is the CPSK.

5. Transformation from Similarity Matrix to Bernoulli Mixture Model

Let us consider whether the similarity matrix allows the decomposition in (10) for sufficiently large C . Then, we construct an algorithm providing the probability decomposition (10) that approximates the similarity matrix.

5.1. Decomposition of Similarity Matrix

We show that a modified similarity matrix defined from the CPSK is decomposed into the form of (10). Suppose the n by n similarity matrix S is expressed as (10). Then, we have

$$S_{ii'} r(i') = \frac{r(i, i')}{r(i)},$$

where $r(i) = r(i') = 1/n$. Taking the sum over $i' \in X$, we find that the equality

$$S\mathbf{1}/n = \mathbf{1} \tag{13}$$

should hold. If the equality (13) is not required, the completely positive matrix S will be always decomposed into the form of (10) up to a constant factor. The equality (13) does not necessarily hold even when the CPSK is used. Let us define the diagonal matrix D as

$$D_{ii} = \max_{i'} (S\mathbf{1})_{i'} - (S\mathbf{1})_i \geq 0, \quad i \in [n],$$

and let \tilde{S} be

$$\tilde{S} = \frac{n}{\max_i (\tilde{S}\mathbf{1})_i} (S + D). \quad (14)$$

Then, $\tilde{S}\mathbf{1}/n = \mathbf{1}$ holds. Since S is the completely positive matrix, also \tilde{S} is the completely positive matrix. Suppose that \tilde{S}/n^2 is decomposed into FF^T with the non-negative matrix $F = (f_1, \dots, f_C) \in \mathbb{R}^{n \times C}$. Then,

$$\frac{1}{n^2} \tilde{S} = \sum_{c \in [C]} f_c f_c^T = \sum_{c \in [C]} \|f_c\|_1^2 \frac{f_c}{\|f_c\|_1} \frac{f_c^T}{\|f_c\|_1}.$$

Let us define $\pi_c = \|f_c\|_1^2$ and $r(i|c) = (f_c)_i / \|f_c\|_1 \geq 0$. Since $\mathbf{1}^T \tilde{S}\mathbf{1}/n^2 = 1$, we have $\sum_c \pi_c = 1$. Moreover, the equality $\tilde{S}\mathbf{1}/n = \mathbf{1}$ guarantees

$$\frac{1}{n} (\tilde{S}\mathbf{1})_i = n \sum_c \|f_c\|_1^2 \frac{f_c}{\|f_c\|_1} = n \sum_c \pi_c r(i|c) = 1,$$

meaning that $\sum_c \pi_c r(i|c) = r(i) = 1/n$ for $i \in X$. Hence, we have

$$\tilde{S}_{ii'} = \frac{\sum_c \pi_c r(i|c) r(i'|c)}{r(i) r(i')}.$$

The modification (14) corresponds to the change of the balance between the self-similarity and the similarities with others.

5.2. Decomposition Algorithm

Let us propose a computation algorithm to obtain the approximate decomposition of the similarity matrix S . Once the decomposition of S is obtained, the recommendation using the similarity measure is connected to the BMMS. Such a correspondence provides a statistical interpretation of the similarity-based methods. For example, the conditional probability $r(c|i)$ is available to categorize nodes into some classes according to the tendency of their preferences once a similarity matrix is obtained. The statistical interpretation provides a supplementary tool for similarity-based methods.

The problem is to find π_c and $r(i|c)$ such that the equation $\frac{\sum_c \pi_c r(i|c) r(i'|c)}{r(i) r(i')}$ approximates the similarity matrix $S = S_{ii'}$, where $r(i) = 1/n = \sum_c \pi_c r(i|c)$ for $n = |X|$. Here, we focus on the similarity matrix on X . The same argument is clearly valid for the similarity matrix on Y .

This problem is similar to the non-negative matrix factorization (NMF) [44]. However, the standard algorithm for the NMF does not work since we have the additional constraint, $\sum_c \pi_c r(i|c) = 1/n$. Here, we use the extended Kullback–Leibler (ext-KL) divergence to measure the discrepancy [45]. The ext-KL divergence between the two matrices $C = (c_{ij})$ and $D = (d_{ij})$ with nonnegative elements is defined as

$$\text{KL}(C, D) = \sum_{ij} c_{ij} \log \frac{c_{ij}}{d_{ij}} - \sum_{ij} c_{ij} + \sum_{ij} d_{ij} \geq 0. \quad (15)$$

The minimization of the ext-KL divergence between $S_{ii'}$ and the model $r(i, i')/r(i)r(i')$ is formalized by

$$\begin{aligned} \min_{r(i|c) > 0, \pi_c > 0} & - \sum_{ij} S_{ij} \log \frac{\sum_c \pi_c r(i|c) r(j|c)}{r(i) r(j)} + \sum_{ij} \frac{\sum_c \pi_c r(i|c) r(j|c)}{r(i) r(j)}, \\ \text{s. t. } & \sum_c \pi_c r(i|c) = r(i) = 1/n, \sum_c \pi_c = 1, \sum_i r(i|c) = 1. \end{aligned}$$

This is equivalent with

$$\begin{aligned}
& \min_{r(i|c)>0, \pi_c>0} - \sum_{ij} S_{ij} \log \sum_c \pi_c r(i|c) r(j|c) \\
& \text{s. t. } \sum_c \pi_c r(i|c) = r(i) = 1/n, \sum_c \pi_c = 1, \sum_i r(i|c) = 1.
\end{aligned} \tag{16}$$

There are many optimization algorithms that can be used to solve nonlinear optimization problems with equality constraints. A simple method is the alternating update of π_c and $r(i|c)$ such as the coordinate descent method [46]. Once $r(i|c)$ is fixed, however, the parameter π_c will be uniquely determined from the first equality constraint in (16) under a mild assumption. This means that the parameter π_c cannot be updated while keeping the equality constraints. Hence, the direct application of the coordinate descent method does not work. On the other hand, the gradient descent method with projection onto the constraint surface is a promising method [47,48]. In order to guarantee the convergence property, however, the step-length should be carefully controlled. Moreover, the projection in every iteration is computationally demanding. In the following, we propose a simple method to obtain an approximate solution of (16) with an easy implementation.

The constraint $\sum_c \pi_c r(i|c) = r(i) = 1/n$ is replaced with the condition that the KL-divergence between the uniform distribution and $\sum_c \pi_c r(i|c)$ vanishes, i.e.,

$$\frac{1}{n} \sum_i \log \frac{1/n}{\sum_c \pi_c r(i|c)} = 0.$$

We incorporated this constraint into the objective function to obtain tractable algorithm. Eventually, the optimization problem we considered is

$$\begin{aligned}
& \max_{r(i|c)>0, \pi_c>0} \sum_{ij} S_{ij} \log \sum_c \pi_c r(i|c) r(j|c) + \frac{\lambda}{n} \sum_i \log \sum_c \pi_c r(i|c) \\
& \text{s. t. } \sum_c \pi_c = 1, \sum_i r(i|c) = 1,
\end{aligned}$$

where the minimization problem is replaced with the maximization problem and λ is the regularization parameter. For a large λ , the optimal solution approximately satisfies the equality constraint $\sum_c \pi_c r(i|c) = 1/n$.

For the above problem we use the majorization-minimization (MM) algorithm [49]. Let $a_{cij}, c \in [C], i, j \in [n]$ and b_{ci} be the auxiliary positive variables satisfying $a_{cij} = a_{cji}$, $\sum_c a_{cij} = 1$ and $\sum_c b_{ci} = 1$. Then, the objective function is bounded below by

$$\begin{aligned}
& \sum_{ij} S_{ij} \log \sum_c \pi_c r(i|c) r(j|c) + \frac{\lambda}{n} \sum_i \log \sum_c \pi_c r(i|c) \\
& \geq \sum_{c,i,j} S_{ij} a_{cij} \log \frac{\pi_c r(i|c) r(j|c)}{a_{cij}} + \frac{\lambda}{n} \sum_{c,i} b_{ci} \log \frac{\pi_c r(i|c)}{b_{ci}}.
\end{aligned}$$

For fixed π_c and $r(i|c)$, the optimal a_{cij} and b_{ci} are explicitly obtained. The optimal solutions of π_c and $r(i|c)$ for a given a_{cij} and b_{ci} are also explicitly obtained. As a result, we obtain the following algorithm to compute the parameters in the Bernoulli mixture model from the similarity matrix S . Algorithm 1 is referred to as the SM-to-BM algorithm.

The convergence of the SM-to-BM algorithm is guaranteed from the general argument of the MM algorithm [49].

Note that the SM-to-BM algorithm yields an approximate BMM model, even if the similarity matrix S is not completely positive such as the Hub Promoted. However, the approximation accuracy is thought to be not necessarily high, since not-CPSK such as the Hub Promoted does not directly correspond to the exchangeable mixture model (10).

Algorithm 1: SM-to-BM algorithm.

Input: Similarity matrix $S = (S_{ii'})_{i,i' \in [n]}$ and the number of classes C .

Step 1. Initial values of auxiliary variables $a_{cii'}$ and b_{ci} are defined.

Step 2. Repeat (i) and (ii) until the solution converges to a point:

(i) For given $a_{cii'}$ and b_{ci} :

$$\pi_c \leftarrow \frac{\sum_{i,i'} S_{ii'} a_{cii'} + \frac{\lambda}{n} \sum_i b_{ci}}{\sum_{c,i,i'} S_{ii'} a_{cii'} + \frac{\lambda}{n} \sum_{c,i} b_{ci}}, \quad r(i|c) \leftarrow \frac{2 \sum_{i'} S_{ii'} a_{cii'} + \frac{\lambda}{n} b_{ci}}{2 \sum_{i,i'} S_{ii'} a_{cii'} + \frac{\lambda}{n} \sum_i b_{ci}}$$

(ii) For given r_{ic} and π_c :

$$a_{cii'} \leftarrow \frac{r_{ic} r_{i'c} \pi_c}{\sum_c r_{ic} r_{i'c} \pi_c}, \quad b_{ci} \leftarrow \frac{r_{ic} \pi_c}{\sum_c r_{ic} \pi_c}.$$

Step 3. Terminate the algorithm with the output: “The similarity matrix S is approximately obtained from the Bernoulli mixture model with π_c and the auxiliary variable $r(c|i) = n \cdot r(i|c) \pi(c)$.”

6. Numerical Experiments of Reciprocal Recommendation

We conducted numerical experiments to ensure the effectiveness of the BMMs for the reciprocal recommendation. We also investigated how well the SM-to-BM algorithm works for the recommendation. In numerical experiments, we compare the prediction accuracy for the recommendation problems.

Suppose that there exist two groups, $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$. Expressions of interest between these two groups are observed and they are expressed as directed edges. Hence, the observation is summarized as the bipartite graph with directed edges between X and Y . If there exists two directed edges (x, y) and (y, x) between $x \in X$ and $y \in Y$, the pair is a preferable match in the graph. The task is to recommend a subset of Y to each element in X and vice versa based on the observation. The purpose is to provide potentially preferable matches as much as possible.

There are several criteria used to measure the prediction accuracy. Here, we use the mean average precision (MAP), because the MAP is a typical metric for evaluating the performance of recommender systems; see [5,50–56] and references therein for more details.

Let us explain the MAP according to [50]. The recommendation to the element x is provided as the ordered set of Y , i.e., $y_{(1)}, y_{(2)}, \dots, y_{(m)}$, meaning that the preferable match between x and $y_{(1)}$ is regarded to be most likely to occur compared to $y_{(2)}, \dots, y_{(m)}$. Suppose that for each $x \in X$, the preferable matches with elements in the subset $\hat{Y}_x \subset Y$ are observed in the test dataset. Let us define z_i as $z_i = 1$ if $y_{(i)}$ is included in \hat{Y}_x and otherwise $z_i = 0$. The precision at the position k is defined as $P@k = \frac{1}{k} \sum_{i=1}^k z_i$. The average precision v_x is then given as the average of $P@k$ with the weight z_k , i.e.,

$$v_x = \frac{\sum_{k=1}^m z_k P@k}{\sum_{k=1}^m z_k}.$$

Note that v_x is well defined unless $\sum_{i=1}^m z_i$ is zero. For example, we have $v_x = 1$ for $\hat{Y}_x = \{y_{(1)}, \dots, y_{(m')}\}$ with $m' \leq m$, and $v_x = 1 - \frac{m'}{m-m'} \sum_{k=m'+1}^m \frac{1}{k}$ for $\hat{Y}_x = \{y_{(m'+1)}, \dots, y_{(m)}\}$ with $0 \leq m' < m$. In the latter case, $v_x = 1/m$ for $m' = m-1$, and $v_x = \frac{1}{m} + \frac{1}{2(m-1)}$ for $m' = m-2$. The MAP is defined as the mean value of v_x over $x \in X$. The high MAP value implies that the ordered set over Y generated by the recommender system is accurate on average. We use the normalized MAP that is the ratio of the above MAP and the expected MAP for the random recommendation. The normalized

MAP is greater than one when the prediction accuracy of the recommendation is higher than that of the random recommendation.

The normalized discounted cumulative gain (NDCG) [5,50,57] is another popular measure in the literature of information retrieval. However, the computation of the NDCG requires the true ranking over the node. Hence, the NDCG is not available for the real-world data in our problem setup.

6.1. Gaussian Mixture Models

The graph is randomly generated based on the attributes defined on each node. The size of X and Y is 1000. Suppose that $x_i \in X$ has the profile vector $\mathbf{x}'_i \in \mathbb{R}^{100}$ and the preference vector $\mathbf{x}''_i \in \mathbb{R}^{100}$. Thus, the attribute vector of x_i is given by $(\mathbf{x}'_i, \mathbf{x}''_i) \in \mathbb{R}^{200}$. Likewise, the attribute vector $(\mathbf{y}'_j, \mathbf{y}''_j) \in \mathbb{R}^{200}$ of $y_j \in Y$ consists of the profile vector $\mathbf{y}'_j \in \mathbb{R}^{100}$ and the preference vector $\mathbf{y}''_j \in \mathbb{R}^{100}$. For each $x_i \in X$, 100 elements in Y , for example, $y_{k_1}, \dots, y_{k_{100}}$ are randomly sampled. Then, the Euclidean distance between the preference vector \mathbf{x}''_i of x_i and the profile vector \mathbf{y}'_{k_j} of y_{k_j} , i.e., $\|\mathbf{x}''_i - \mathbf{y}'_{k_j}\|$ is calculated for each y_{k_j} . Then, the 10 closest y_{k_j} from x_i in terms of the above distance are chosen and directed edges from x_i to the 10 chosen nodes in Y are added. In the same way, the edges from Y to X are generated and added to the graph. The training data is obtained as a random bipartite graph. Repeating the same procedure again with a different random seed, we obtain another random graph as a test data.

The above setup imitates practical recommendation problems. Usually, a profile vector is observed for each user. However, the preference vector is not directly observed, while the preference of each user can be inferred via the observed edges.

In our experiments, the profile vectors and preference vectors are identically and independently distributed from the Gaussian mixture distribution with two components, i.e.,

$$\mathbf{x}'_i, \mathbf{x}''_i, \mathbf{y}'_j, \mathbf{y}''_j \sim_{i.i.d.} \frac{1}{2}N_{100}(\mathbf{0}, I_{100}) + \frac{1}{2}N_{100}(\mathbf{1}, I_{100}),$$

meaning that each profile or preference vector is generated from $N_{100}(\mathbf{0}, I_{100})$ or $N_{100}(\mathbf{1}, I_{100})$ with probability 1/2. Hence, each node in X is roughly categorized into one of two classes, i.e., $\mathbf{0}$ or $\mathbf{1}$, that is the mean vector of the preference, \mathbf{x}''_i . When the class of x_i is $\mathbf{0}$ (resp. $\mathbf{1}$), the edge from x_i is highly likely to be connected to y_j having the profile vector generated from $N_{100}(\mathbf{0}, I_{100})$ (resp. $N_{100}(\mathbf{1}, I_{100})$). Therefore, the distribution of edges from X to Y will be well approximated by the Bernoulli mixture model with $C = 2$. Figure 1 depicts the relationship between the distribution of attributes and edges from X to Y . The same argument holds for the distribution of edges from Y to X .

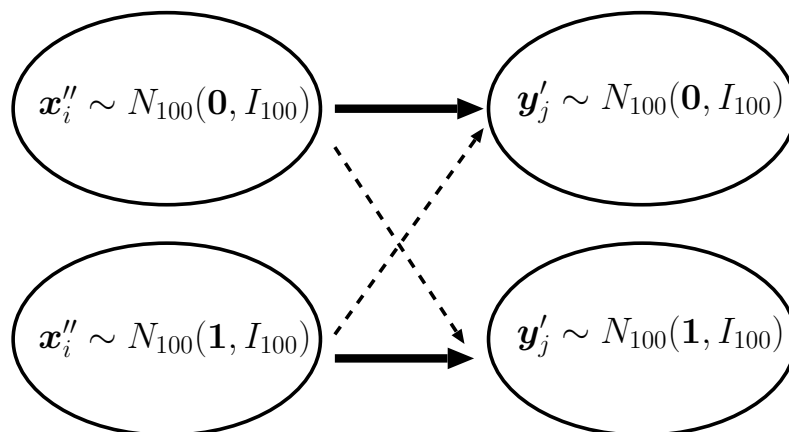


Figure 1. Edges from X to Y . The bold edges mean that there are many edges between the connected groups. The broken edges mean that there are few edges between the connected groups.

In this simulation, we have focused on the recommendation using similarity measures based on the graph structure. The recommendation to each node of the graph was determined by (1),

where the similarity measures in Table 1 or the one determined from the Bernoulli mixture model (10) were employed. Table 2 shows the averaged MAP scores with the median absolute deviation (MAD) over 10 repetitions with different random seeds. In our experiments, the recommendation based on the BMMs with the appropriate number of components outperformed the other methods. However, the BMMs with a high number of components showed low prediction accuracy.

Table 2. Mean average precision (MAP) values of similarity-based methods under synthetic data. The bold face indicates the top two MAP scores.

Similarity	MAP (MAD)
Common Neighbors	1.889 (± 0.413)
Cosine	1.907 (± 0.431)
Jaccard Coefficient	2.115 (± 0.421)
Sørensen Index	2.021 (± 0.369)
Hub Depressed	2.231 (± 0.376)
Hub Promoted	2.053 (± 0.301)
SimRank	2.853 (± 0.610)
Adamic-Adar coefficient	2.188 (± 0.587)
Resource Allocation	1.950 (± 0.516)
Bernoulli Mixture ($C = 2$)	5.599 (± 1.811)
Bernoulli Mixture ($C = 5$)	4.552 (± 1.766)
Bernoulli Mixture ($C = 10$)	2.821 (± 1.164)
Bernoulli Mixture ($C = 50$)	1.382 (± 0.449)
Bernoulli Mixture ($C = 100$)	1.535 (± 0.555)

Below, we show the edge prediction based on the SM-to-BM algorithm in Section 5. The results are shown in Table 3. The number of components in the Bernoulli mixture model was set to $C = 2$ or $C = 5$. Given the similarity matrix S , the SM-to-BM algorithm yielded the parameter π_c and $r(i|c)$. Next, edges were predicted through the formula (10) using π_c , $r(i|c)$ and $r(i) = \sum_c \pi_c r(i|c)$. The averaged MAP scores of this procedure are reported in the column of “itr:0”. We also examined the edge prediction by the BMMs with the parameter updated from the one obtained by the SM-to-BM algorithm, where the update formula is given by (7) and (8). The “itr:10” (resp. “itr:100”) column shows the MAP scores of the edge prediction using 10 times (resp. 100 times) updated parameter. In addition, the “BerMix” shows the MAP score of the BMMs with the updated parameter from the randomly initialized parameter.

In our experiments, we found that the SM-to-BM algorithm applied to commonly used similarity measures improved the accuracy of the recommendation. The MAP score for the “itr:0” method achieved a higher accuracy than the original similarity-based methods. The updated parameter from “itr:0”, however, did not improve the MAP score significantly. The results of “itr:10” and “itr:100” for similarity measures were almost the same when the model was the Bernoulli mixture model with $C = 2$. This is because the EM algorithm with 10 iterations achieved the stationary point of this model in our experiments. We confirmed that there was yet a gap between the likelihood of the parameter computed by the SM-to-BM algorithm and the maximum likelihood. However, the numerical results indicate that the SM-to-BM algorithm provides a good parameter for the recommendation in the sense of the MAP score.

Table 3. MAP values of updated Bernoulli mixture models with the SM-to-BM algorithm under synthetic data. The results of Bernoulli mixture models with $C = 2$ and $C = 5$ are reported. The bold face indicates the top two MAP scores in each column.

SM-to-BM: $C = 2$			
Similarity	itr:0	itr:10	itr:100
Common Neighbors	5.646 (± 1.041)	5.572 (± 1.041)	5.572 (± 1.041)
Cosine	4.725 (± 1.119)	4.549 (± 1.119)	4.549 (± 1.119)
Jaccard Coefficient	5.421 (± 0.643)	5.373 (± 0.643)	5.373 (± 0.643)
Sørensen Index	5.013 (± 2.223)	4.964 (± 2.223)	4.964 (± 2.223)
Hub Depressed	5.417 (± 1.756)	5.262 (± 1.756)	5.262 (± 1.756)
Hub Promoted	5.120 (± 0.563)	5.165 (± 0.563)	5.165 (± 0.563)
SimRank	3.848 (± 1.630)	4.377 (± 1.279)	4.379 (± 1.264)
Adamic-Adar coefficient	4.348 (± 1.170)	4.404 (± 1.170)	4.404 (± 1.170)
Resource Allocation	4.435 (± 0.552)	4.385 (± 0.552)	4.385 (± 0.552)
BerMix. (Random ini.)	1.297 (± 0.446)	5.718 (± 2.013)	5.911 (± 2.087)
SM-to-BM: $C = 5$			
Similarity	itr:0	itr:10	itr:100
Common Neighbors:	5.059 (± 0.939)	5.557 (± 0.939)	5.137 (± 0.939)
Cosine	4.442 (± 0.901)	4.070 (± 0.901)	3.948 (± 0.901)
Jaccard Coefficient	5.167 (± 1.745)	4.765 (± 1.745)	4.792 (± 1.745)
Sørensen Index	5.675 (± 0.773)	5.294 (± 0.773)	5.189 (± 0.773)
Hub Depressed	5.408 (± 1.807)	4.668 (± 1.807)	4.391 (± 1.807)
Hub Promoted	5.078 (± 0.702)	4.815 (± 0.702)	5.008 (± 0.702)
SimRank	4.121 (± 1.274)	3.592 (± 1.150)	3.615 (± 1.447)
Adamic-Adar coefficient	5.284 (± 1.166)	4.909 (± 1.166)	5.084 (± 1.166)
Resource Allocation	4.884 (± 0.751)	4.499 (± 0.751)	4.263 (± 0.751)
BerMix. (Random ini.)	1.080 (± 0.446)	3.705 (± 1.925)	4.810 (± 1.268)

6.2. Real-World Data

We show the results for real-world data. The data was provided from an online dating site. The set X (resp. Y) consists of $n = 15,925$ males and $m = 16,659$ females. The data were gathered from 3 January 2016 to 5 June 2017. We used 130,8126 messages from 3 January 2016 to 31 October 2016 as the training data. Test data consists of 177,450 messages from 1 November 2016 to 5 June 2017 [55]. The proportion of edges in the test set to all data set is approximately 0.12.

In the numerical experiments, half of the users were randomly sampled from each group, and the corresponding subgraph with the training edges were defined as the training graph. On the other hand, the graph with the same nodes in the training graph and the edges in the test edges were used as the test graph. Based on the training graph, the recommendation was provided and was evaluated on the test graph. The same procedure was repeated over 20 times and the averaged MAP scores for each similarity measure were reported in Table 4. In the table, the MAP score of the recommendation for X and Y were separately reported. So far, we have defined the similarity measure based on out-edges from each node of the directed bipartite graph, referred to as “Interest”. On the other hand, the similarity measure defined by in-edges is referred to as “Attract”. For the BMMs, “Attract” means that the model of each component is computed under the assumption that each in-edge is independently generated, i.e., the probability of $(a_{ij})_{i \in [n]}$ is given by $\prod_i \alpha_{ic}^{a_{ij}} (1 - \alpha_{ic})^{1-a_{ij}}$ when the class of $y_j \in Y$ is c . In the real-world datasets, the SM-to-BM algorithm was not used, because the dataset was too large to compute the corresponding BMMs from similarity matrices.

As shown in the numerical results, the recommendation based on the BMMs outperformed the other methods. Some similarity measures such as the Common Neighbors or Adamic-Adar coefficient showed relatively good results. On the other hand, the Hub Promoted measure, that is not the CPSK,

showed the lowest prediction accuracy. As well as the result for synthetic data, the BMMs with two to five components produced high prediction accuracy. Even for medium to large datasets, we found that the Bernoulli mixture model with about five components worked well. We expect that the validation technique is available to determine the appropriate size of components. Also, the similarity with “Interest” or “Attract” can be determined from the validation dataset.

Table 4. MAP scores for real-world data. The bold face indicates the top two MAP scores in each column.

Similarity	Recomm. of Y to X MAP (MAD):	Recomm. of X to Y MAP (MAD)
Common Neighbors:Interest	6.267 (± 0.806)	2.893 (± 0.343)
Common Neighbors:Attract	2.053 (± 0.276)	8.813 (± 0.757)
Cosine:Interest	3.496 (± 0.324)	3.699 (± 0.309)
Cosine:Attract	2.746 (± 0.276)	6.108 (± 0.546)
Jaccard Coefficient:Interest	4.098 (± 0.373)	4.066 (± 0.294)
Jaccard Coefficient:Attract	3.288 (± 0.362)	7.777 (± 0.724)
Sørensen Index:Interest	4.205 (± 0.363)	3.996 (± 0.293)
Sørensen Index:Attract	3.205 (± 0.319)	7.910 (± 0.620)
Hub Depressed:Interest	4.370 (± 0.369)	4.106 (± 0.291)
Hub Depressed:Attract	3.366 (± 0.379)	8.364 (± 0.613)
Hub Promoted:Interest	1.959 (± 0.300)	2.691 (± 0.334)
Hub Promoted:Attract	1.662 (± 0.262)	2.641 (± 0.263)
SimRank:Interest	2.079 (± 0.164)	6.336 (± 0.423)
SimRank:Attract	5.100 (± 0.775)	3.158 (± 0.193)
Adamic-Adar coefficient:Interest	6.216 (± 0.701)	2.970 (± 0.308)
Adamic-Adar coefficient:Attract	2.209 (± 0.267)	8.300 (± 0.632)
Resource Allocation:Interest	5.521 (± 0.679)	3.557 (± 0.262)
Resource Allocation:Attract	2.713 (± 0.298)	6.875 (± 0.660)
Bernoulli Mixture (C = 2):Interest	4.578 (± 0.734)	15.061 (± 4.106)
Bernoulli Mixture (C = 2):Attract	10.625 (± 2.054)	12.825 (± 3.323)
Bernoulli Mixture (C = 5):Interest	5.055 (± 0.813)	17.394 (± 3.271)
Bernoulli Mixture (C = 5):Attract	10.362 (± 1.981)	10.348 (± 2.514)
Bernoulli Mixture (C = 10):Interest	4.263 (± 0.772)	15.013 (± 4.042)
Bernoulli Mixture (C = 10):Attract	10.451 (± 1.133)	8.786 (± 1.195)
Bernoulli Mixture (C = 50):Interest	5.664 (± 1.873)	14.288 (± 6.409)
Bernoulli Mixture (C = 50):Attract	10.029 (± 3.933)	8.436 (± 3.929)
Bernoulli Mixture (C = 100):Interest	2.910 (± 0.436)	8.525 (± 1.199)
Bernoulli Mixture (C = 100):Attract	5.980 (± 1.612)	5.119 (± 0.464)

7. Discussions and Concluding Remarks

In this paper, we considered the relationship between the similarity-based recommendation methods and statistical models. We showed that the BMMs are closely related to the recommendation using completely positive similarity measures. More concretely, both the BMM-based method and completely positive similarity measures share exchangeable mixture models as the statistical model of the edge distribution. Once this was established, we proposed the recommendation methods using the EM algorithm to BMMs to improve similarity-based methods.

Moreover, we proposed the SM-to-BM algorithm that transforms a similarity matrix to parameters of the Bernoulli mixture model. The main purpose of the SM-to-BM algorithm is to find a statistical model corresponding to a given similarity matrix. This transformation provides a statistical interpretation for similarity-based methods. For example, the conditional probability $r(c|i)$ is obtained from the SM-to-BM algorithm. This probability is useful to categorize nodes, i.e., users, into some classes according to the tendency of their preferences once a similarity matrix is obtained. The SM-to-BM algorithm is available as a supplementary tool for similarity-based methods.

We conducted numerical experiments using synthetic and real-world data. We numerically verified the efficiency of the BMM-based method in comparison to similarity-based methods. For the

synthetic data, the BMM-based method was compared with the recommendation using the statistical model obtained by the SM-to-BM algorithm. We found that the BMM-based method and the SM-to-BM method provide a comparable accuracy for the reciprocal recommendation. Since the synthetic data is well approximated by the BMM with $C = 2$, the SM-to-BM algorithm was thought to reduce the noise in the similarity matrices. In the real-world data, the SM-to-BM algorithm was not examined, since our algorithm using the MM method was computationally demanding for a large dataset. On the other hand, we observed that the BMM-based EM algorithm was scalable for a large dataset. A future work includes the development of computationally efficient SM-to-BM algorithms.

It is straightforward to show that the stochastic block models (SBMs) [6] are also closely related to the recommendation with completely positive similarity measures. In our preliminary experiments, however, we found that the recommendation system based on the SBMs did not show a high prediction accuracy in comparison to other methods. We expect that detailed theoretical analysis of the relation between the similarity measure and statistical models is an interesting research topic that can be used to better understand the meaning of the commonly used similarity measures.

Author Contributions: Conceptualization, T.K. and N.O.; Methodology, T.K.; Software, T.K.; Validation, T.K., N.O.; Formal Analysis, T.K.; Investigation, T.K.; Resources, T.K.; Data Curation, T.K.; Writing—Original Draft Preparation, T.K.; Writing—Review & Editing, T.K.; Visualization, T.K.; Supervision, T.K.; Project Administration, T.K.; Funding Acquisition, N.O.

Funding: T.K. was partially supported by JSPS KAKENHI Grant Number 15H01678, 15H03636, 16K00044, and 19H04071.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, P.; Xu, B.; Wu, Y.; Zhou, X. Link prediction in social networks: The state-of-the-art. *Sci. China Inf. Sci.* **2015**, *58*, 1–38. [\[CrossRef\]](#)
2. Liben-nowell, D.; Kleinberg, J. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1019–1031. [\[CrossRef\]](#)
3. Hasan, M.A.; Zaki, M.J. A Survey of Link Prediction in Social Networks. In *Social Network Data Analytics*; Springer Science+Business Media: Berlin/Heidelberg, Germany, 2011; pp. 243–275.
4. Lü, L.; Zhou, T. Link prediction in complex networks: A survey. *Phys. A* **2011**, *390*, 1150–1170. [\[CrossRef\]](#)
5. Agarwal, D.K.; Chen, B.C. *Statistical Methods for Recommender Systems*; Cambridge University Press: Cambridge, UK, 2016.
6. Stanley, N.; Bonacci, T.; Kwitt, R.; Niethammer, M.; Mucha, P.J. Stochastic Block Models with Multiple Continuous Attributes. *arXiv* **2018**, arXiv:1803.02726.
7. Mengdi, W. Vanishing Price of Decentralization in Large Coordinative Nonconvex Optimization. *SIAM J. Optim.* **2017**, *27*, 1977–2009.
8. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser.* **1977**, *39*, 1–38. [\[CrossRef\]](#)
9. Berman, A.; Shaked-Monderer, N. *Completely Positive Matrices*; World Scientific Publishing Company Pte Limited: Singapore, 2003.
10. Burer, S.; Anstreicher, K.M.; Dür, M. The difference between 5×5 doubly nonnegative and completely positive matrices. *Linear Algebra Its Appl.* **2009**, *431*, 1539–1552. [\[CrossRef\]](#)
11. Diaconis, P. Finite forms of de Finetti's theorem on exchangeability. *Synth. Int. J. Epistemol. Methodol. Philos. Sci.* **1977**, *36*, 271–281. [\[CrossRef\]](#)
12. Wood, G.R. Binomial Mixtures and Finite Exchangeability. *Ann. Probab.* **1992**, *20*, 1167–1173. [\[CrossRef\]](#)
13. Diaconis, P.; Freedman, D. Finite Exchangeable Sequences. *Ann. Probab.* **1980**, *8*, 745–764. [\[CrossRef\]](#)
14. De Finetti, B. *Theory of Probability*; Wiley: Hoboken, NJ, USA, 1970.
15. Xia, P.; Liu, B.; Sun, Y.; Chen, C. Reciprocal Recommendation System for Online Dating. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France, 25–28 August 2015; pp. 234–241.

16. Li, L.; Li, T. MEET: A Generalized Framework for Reciprocal Recommender Systems. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, HI, USA, 29 October–2 November 2012; ACM: New York, NY, USA, 2012; pp. 35–44.
17. Pizzato, L.; Rej, T.; Chung, T.; Koprinska, I.; Kay, J. RECON: A Reciprocal Recommender for Online Dating. In Proceedings of the Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; ACM: New York, NY, USA, 2010; pp. 207–214.
18. Pizzato, L.; Rej, T.; Akehurst, J.; Koprinska, I.; Yacef, K.; Kay, J. Recommending People to People: The Nature of Reciprocal Recommenders with a Case Study in Online Dating. *User Model. User-Adapt. Interact.* **2013**, *23*, 447–488. [[CrossRef](#)]
19. Xia, P.; Jiang, H.; Wang, X.; Chen, C.; Liu, B. Predicting User Replying Behavior on a Large Online Dating Site. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014.
20. Yu, M.; Zhao, K.; Yen, J.; Kreager, D. Recommendation in Reciprocal and Bipartite Social Networks—A Case Study of Online Dating. In Proceedings of the Social Computing, Behavioral-Cultural Modeling and Prediction—6th International Conference (SBP 2013), Washington, DC, USA, 2–5 April 2013; pp. 231–239.
21. Tu, K.; Ribeiro, B.; Jensen, D.; Towsley, D.; Liu, B.; Jiang, H.; Wang, X. Online Dating Recommendations: Matching Markets and Learning Preferences. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; ACM: New York, NY, USA, 2014; pp. 787–792.
22. Hopcroft, J.; Lou, T.; Tang, J. Who Will Follow You Back?: Reciprocal Relationship Prediction. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; ACM: New York, NY, USA, 2011; pp. 1137–1146.
23. Hong, W.; Zheng, S.; Wang, H.; Shi, J. A Job Recommender System Based on User Clustering. *J. Comput.* **2013**, *8*, 1960–1967. [[CrossRef](#)]
24. Brun, A.; Castagnos, S.; Boyer, A. Social recommendations: Mentor and leader detection to alleviate the cold-start problem in collaborative filtering. In *Social Network Mining, Analysis and Research Trends: Techniques and Applications*; Ting, I., Hong, T.-P., Wang, L.S., Eds.; IGI Global: Hershey, PA, USA, 2011; pp. 270–290.
25. Gentile, C.; Parotsidis, N.; Vitale, F. Online Reciprocal Recommendation with Theoretical Performance Guarantees. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; pp. 8257–8267.
26. Akehurst, J.; Koprinska, I.; Yacef, K.; Pizzato, L.A.S.; Kay, J.; Rej, T. CCR—A Content-Collaborative Reciprocal Recommender for Online Dating. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, 16–22 July 2011; pp. 2199–2204.
27. Newman, M.E.J. Clustering and preferential attachment in growing networks. *Phys. Rev. Lett.* **2001**, *64*, 025102. [[CrossRef](#)] [[PubMed](#)]
28. Zhu, Y.X.; Lü, L.; Zhang, Q.M.; Zhou, T. Uncovering missing links with cold ends. *Phys. Stat. Mech. Its Appl.* **2012**, *391*, 5769–5778. [[CrossRef](#)]
29. Urbani, C.B. A Statistical Table for the Degree of Coexistence between Two Species. *Oecologia* **1980**, *44*, 287–289. [[CrossRef](#)] [[PubMed](#)]
30. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* **1948**, *5*, 1–34.
31. ZhouEmail, T.; Lü, L.; Zhang, Y.C. Predicting missing links via local information. *Eur. Phys. J.* **2009**, *71*, 623–630.
32. Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **2002**, *297*, 1551–1555. [[CrossRef](#)] [[PubMed](#)]
33. Jeh, G.; Widom, J. SimRank: A Measure of Structural-Context Similarity. In Proceedings of the Eighth ACM SIGKDD International Conference, Edmonton, AB, Canada, 23–25 July 2002; pp. 538–543.
34. Adamic, L.A.; Adar, E. Friends and neighbors on the Web. *Soc. Netw.* **2003**, *25*, 211–230. [[CrossRef](#)]
35. Zhu, R.; Zou, Z.; Li, J. SimRank on Uncertain Graphs. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2522–2536. [[CrossRef](#)]
36. Govaert, G.; Nadif, M. Block clustering with Bernoulli mixture models: Comparison of different approaches. *Comput. Stat. Data Anal.* **2008**, *52*, 3233–3245. [[CrossRef](#)]

37. Govaert, G.; Nadif, M. Fuzzy Clustering to Estimate the Parameters of Block Mixture Models. *Soft-Comput. Fusion Found. Methodol. Appl.* **2006**, *10*, 415–422. [[CrossRef](#)]
38. Amir, N.; Abolfazl, M.; Hamid, R.R. Reliable Clustering of Bernoulli Mixture Models. *arXiv* **2019**, arXiv:1710.02101.
39. Finetti, B.D. *Probability, Induction and Statistics: The Art of Guessing*; Wiley Series in Probability and Mathematical Statistics; Wiley: Hoboken, NJ, USA, 1972.
40. Niepert, M.; Van den Broeck, G. Tractability through exchangeability: A new perspective on efficient probabilistic inference. In Proceedings of the 28th AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.
41. Niepert, M.; Domingos, P. Exchangeable Variable Models. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; Xing, E.P., Jebara, T., Eds.; PMLR: Beijing, China, 2014; Volume 32, pp. 271–279.
42. Barioli, F.; Berman, A. The maximal cp-rank of rank k completely positive matrices. *Linear Algebra Its Appl.* **2003**, *363*, 17–33. [[CrossRef](#)]
43. Schölkopf, B.; Smola, A.J. *Learning with Kernels*; MIT Press: Cambridge, MA, USA, 2002.
44. Lee, D.D.; Seung, H.S. Algorithms for Non-negative Matrix Factorization. In Proceedings of the 13th International Conference on Neural Information Processing Systems, Denver, CO, USA, 28 November 2000; pp. 535–541.
45. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley Series in Telecommunications and Signal Processing; Wiley-Interscience: Hoboken, NJ, USA, 2006.
46. Wright, S.J. Coordinate Descent Algorithms. *Math. Program.* **2015**, *151*, 3–34. [[CrossRef](#)]
47. Bertsekas, D. *Nonlinear Programming*; Athena Scientific: Belmont, MA, USA, 1996.
48. Luenberger, D.; Ye, Y. *Linear and Nonlinear Programming*; Springer: Berlin/Heidelberg, Germany, 2008.
49. Lange, K. *MM Optimization Algorithms*; SIAM: Philadelphia, PA, USA, 2016.
50. Liu, T.Y. Learning to Rank for Information Retrieval. *Found. Trends Inf. Retr.* **2009**, *3*, 225–331. [[CrossRef](#)]
51. Kishida, K. *Property of Average Precision as Performance Measure for Retrieval Experiment*; Technical Report; NII-2005-014E; National Institute of Informatics: Tokyo, Japan, 2005.
52. Cormack, G.V.; Lynam, T.R. Statistical Precision of Information Retrieval Evaluation. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 6–10 August 2006; pp. 533–540.
53. McFee, B.; Lanckriet, G. Metric Learning to Rank. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 775–782.
54. Fukui, K.; Okuno, A.; Shimodaira, H. Image and tag retrieval by leveraging image-group links with multi-domain graph embedding. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 221–225.
55. Sudo, K.; Osugi, N.; Kanamori, T. Numerical study of reciprocal recommendation with domain matching. *Jpn. J. Stat. Data Sci.* **2019**, *2*, 221–240. [[CrossRef](#)]
56. Beitzel, S.M.; Jensen, E.C.; Frieder, O.; Chowdhury, A.; Pass, G. Surrogate Scoring for Improved Metasearch Precision. In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, 15–19 August 2005; pp. 583–584.
57. Wang, Y.; Wang, L.; Li, Y.; He, D.; Chen, W.; Liu, T.Y. A theoretical analysis of NDCG type ranking measures. In Proceedings of the 26th Annual Conference on Learning Theory (COLT 2013), Princeton, NJ, USA, 12–14 June 2013; pp. 583–584.

