

IMP: A Message-Passing Algorithm for Matrix Completion

Byung-Hak Kim, Arvind Yedla, and Henry D. Pfister
Department of Electrical and Computer Engineering, Texas A&M University
Email: {bhkim,yarvind,hpfister}@tamu.edu

Abstract—A new message-passing (MP) method is considered for the matrix completion problem associated with recommender systems. We attack the problem using a (generative) factor graph model that is related to a probabilistic low-rank matrix factorization. Based on the model, we propose a new algorithm, termed IMP, for the recovery of a data matrix from incomplete observations. The algorithm is based on a clustering followed by inference via MP (IMP). The algorithm is compared with a number of other matrix completion algorithms on real collaborative filtering (e.g., Netflix) data matrices. Our results show that, while many methods perform similarly with a large number of revealed entries, the IMP algorithm outperforms all others when the fraction of observed entries is small. This is helpful because it reduces the well-known cold-start problem associated with collaborative filtering (CF) systems in practice.

I. INTRODUCTION

An important new inference problem, called the *matrix completion* problem, has recently come to light; it combines many elements of compressed sensing and collaborative filtering. This problem involves the recovery of a data matrix from incomplete (or corrupted) information and is of great practical interest over a wide range of fields [1]. The basic idea is summarized well in the following quote:

“In its simplest form, the problem is to recover a matrix from a small sample of its entries, and comes up in many areas of science and engineering including collaborative filtering, machine learning, control, remote sensing, and computer vision... Imagine now that we only observe a few entries of a data matrix. Then is it possible to accurately—or even exactly—guess the entries that we have not seen?”
- Candes and Plan [2]

In the Netflix challenge, for example, one is given a subset of large data matrix in which rows are users and columns are movies (e.g., see the Netflix Prize [3]). An overwhelming portion of the user-movie matrix (e.g., 99%) is unknown and the observation matrix is very sparse because most users rate only a few movies. Randomness in the ratings process implies that one can also interpret the ratings as noisy observations of some true matrix.

This work was supported in part by the National Science Foundation under Grant No. 0747470 and the Texas Norman Hackerman Advanced Research Program under Grant No. 000512-0168-2007. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

The goal is to predict the rating that a user would give, to a movie he/she has not watched, based on the observed ratings. In other words, the problem is to recover missing ratings of a data matrix using the subset of observed movie ratings. In general, it would seem that this problem is difficult, if not impossible. However, if the unknown matrix has some structure, then approximate recovery is possible. Recent progress on the matrix completion problem can be largely divided into two areas:

- 1) The first area considers efficient models and practical algorithms. For the matrix completion problem, many researchers use models based on the assumption that the matrix has low rank. This assumption allows one to reformulate the problem into rank (or nuclear norm) minimization problem under certain incoherence assumptions [1]. For exact and approximate matrix completion, these models lead to convex relaxations, and semi-definite programming (SDP) [4][5][6][7], and Bayesian-based approaches [8].
- 2) The second area involves exploration of the fundamental limits of these methods. Prior work has developed some precise relationships between sparse observation models and the recovery of missing entries under the restriction of low-rank matrices or clustering models [2][9][10][11][12]. This area is closely related with the practical issues known as the cold-start problem of the recommender system [13]. That is, giving recommendations to new users who have submitted only a few ratings, or recommending new items that received only a few ratings from users. In other words, how many ratings are needed to generate good recommendations?

Unlike this prior work, this paper considers an important subclass of the matrix completion problem where the entries (drawn from a finite alphabet) are modeled by a (generative) factor graph. Based on this factor graph model, we propose a MP based algorithm, termed IMP, to estimate missing entries. This algorithm seems to share some of the desirable properties demonstrated by MP in its successful application to modern coding theory [14]. The IMP algorithm tries to combine the benefits of soft clustering of users/movies into groups and message-passing based on the unknown groups to make predictions. In addition, simulation results for cold-start settings (i.e., less than 0.5% randomly sampled entries) show that the cold start problem is reduced greatly by IMP

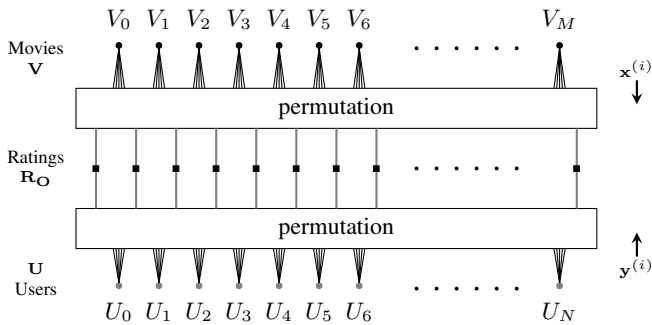


Figure 1. The factor graph model for the matrix completion problem. The graph is sparse when there are few ratings. Edges represent random variables and nodes represent local probabilities. The node probability associated with the ratings implies that each rating depends only on the movie group (top edge) and the user group (bottom edge). Synthetic data can be generated by picking i.i.d. random user/movie groups and then using random permutations to associate groups with ratings. Note $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$ are the messages from movie to user and user to movie during iteration i for the Algorithm 1.

in comparison to other methods on real collaborative filtering (or Netflix) data matrices.

The paper is structured as follows. After defining the factor graph model in Section II, we introduce the IMP algorithm in Section III. In Section IV, we discuss the algorithm performance via experimental results, and give conclusions in Section V.

II. FACTOR GRAPH MODEL

Consider a collection of N users and M movies when the set O of user-movie pairs have been observed. The main theoretical question is, “How large should the size of O be to estimate the unknown ratings within some distortion δ ?”. Answers to this question certainly require some assumptions about the movie rating process. So we begin differently by introducing a probabilistic model for the movie ratings. The basic idea is that *hidden* variables are introduced for users and movies, and that the movie ratings are conditionally independent given these hidden variables. It is convenient to think of the hidden variable for any user (or movie) as the *user group* (or *movie group*) of that user (or movie) and this can be viewed as a simplistic assumption about the psychological nature of movie preferences [15][16]. In this context, the rating associated with a user-movie pair depends only on the user group and the movie group.

Since the number of movie groups are very small compared to the number of movies, this idea is similar to mapping movies to a low-dimensional movie group. Each movie group may correspond to a genre (e.g., comedy, drama, action, ...). Each user group tries to capture sets of users that have similar taste in movies. For example, a movie may be classified as a comedy, and a user may be classified as a comedy lover. The model may use 20 to 40 such groups to locate each movie and user in a multidimensional space. It then predicts a user’s rating of a movie according to the movie’s rating on the dimensions that person cares about most since similar user/movie map to similar groups in the low-dimensional (group) space.

The goal is to design a probabilistic mapping such that reflects group associations in the low-dimensional (group) space. Let there be g_u user groups, g_v movie groups, and define $[k] \triangleq \{1, 2, \dots, k\}$. The user group of the n -th user, $U_n \in [g_u]$, is a discrete random variable drawn from $\Pr(U_n = u) \triangleq p_U(u)$ and $\mathbf{U} = U_1, U_2, \dots, U_N$ is the user group vector. Likewise, the movie group of the m -th movie, $V_m \in [g_v]$, is a discrete random variable drawn from $\Pr(V_m = v) \triangleq p_V(v)$ and $\mathbf{V} = V_1, V_2, \dots, V_M$ is the movie group vector. Then, the rating of the m -th movie by the n -th user is a discrete random variable $R_{nm} \in \mathcal{R}$ (e.g., Netflix uses $\mathcal{R} = [5]$) drawn from $\Pr(R_{nm} = r | U_n = u, V_m = v) \triangleq w(r|u, v)$ and the rating R_{nm} is *conditionally independent* given the user group U_n and the movie group V_m . Let \mathbf{R} denote the rating matrix and the observed submatrix be \mathbf{R}_O with $O \subseteq [N] \times [M]$. In this setup, some of the entries in the rating matrix are observed while others must be predicted. The conditional independence assumption in the model implies that

$$\Pr(\mathbf{R}_O | \mathbf{U}, \mathbf{V}) \triangleq \prod_{(n,m) \in O} w(R_{nm} | U_n, V_m).$$

Specifically, we consider the factor graph (composed of 3 layers, see Figure 1) as a randomly chosen instance of this problem based on this probabilistic model. The key assumptions are that these layers separate the influence of user groups, movie groups, and observed ratings. A random permutation is used to map the edges attached to user nodes to the edges attached to movie nodes.

This model attempts to exploit correlation in the ratings based on similarity between users (and movies). It also tries to include the noisy rating process in the model and reduce the impact of corrupted ratings on prediction by dimension reduction. These advantages allows one to approximate real Netflix data generation process more closely than other simpler factor models. In fact this model can be seen as a generalization of [8] and [11]. It is also important to note that this is a *probabilistic generative model* which generalizes the clustering model in and also allows one to evaluate different learning algorithms on synthetic data and compare the results with theoretical bounds (see [17] for details).

III. THE IMP ALGORITHM

A. Initializing $w(r|u, v)$ for Group Ratings

The IMP algorithm requires reasonable initial estimates, of the observation model $w(r|u, v)$, to get started. To get these estimates, we cluster users (and movies) first. The basic method uses a variable-dimension vector quantization (VDVQ) clustering algorithm and the standard codebook splitting approach known as the generalized Lloyd algorithm (GLA) to generate codebooks whose size is any power of 2 [18]. Though our approach was motivated by the VDVQ clustering algorithm, it turns out to be equivalent to soft K -means clustering with an appropriate distance measure. So we will refer VDVQ clustering as soft K -means clustering.

The soft K -means clustering algorithm is based on the alternating minimization of the average distance between users

Algorithm 1 IMP Algorithm

Step I: Initialization of $w(r|u, v)$ via Algorithm 2 and randomized initialization of user/movie group probabilities $\mathbf{x}_{m \rightarrow n}^{(0)}$ and $\mathbf{y}_{n \rightarrow m}^{(0)}$.

Step II: Recursive update for user/movie group probabilities

$$\mathbf{y}_{n \rightarrow m}^{(i+1)}(u) \propto \mathbf{y}_n^{(0)}(u) \prod_{k \in \mathcal{V}_n \setminus m} \sum_v w(r|u, v) \mathbf{x}_{k \rightarrow n}^{(i)}(v)$$
$$\mathbf{x}_{m \rightarrow n}^{(i+1)}(v) \propto \mathbf{x}_m^{(0)}(v) \prod_{k \in \mathcal{U}_m \setminus n} \sum_u w(r|u, v) \mathbf{y}_{k \rightarrow m}^{(i)}(u)$$

Step III: Update $w(r|u, v)$ and output probability of rating R_{nm} given observed ratings

$$\hat{p}_{R_{nm}|\mathbf{R}_O}^{(i+1)}(r) \propto \sum_{u,v} \mathbf{y}_{n \rightarrow m}^{(i+1)}(u) w(r|u, v) \mathbf{x}_{m \rightarrow n}^{(i+1)}(v)$$

(or movies) and codebooks (that contain no missing data). This leads to alternating application of *nearest neighbor* and *centroid* rules. The distance is computed only on the elements both vectors share. In the case of users, one can think of this Algorithm 2 as a “ K -critics” algorithms which tries to design K critics (i.e., people who have seen every movie) that cover the space of all user tastes and each user is given a soft “degree of assignment (or soft group membership)” to each of the critics which can take on values between 0 and 1. After soft-clustering users/movies each into user/movie groups, we estimate $w(r|u, v)$ by computing the soft frequency of each rating for each user-movie group pair.

B. Message-Passing Updates of Group Vectors

Using the model from Section II, we describe how message-passing can be used for the prediction of hidden variables based on observed ratings. Ideally, we could perform exact inference of our factor graph model. But exact learning and inference for this model is intractable, so we turn to approximate message-passing algorithms (e.g., the sum-product algorithm) [19]. The basic idea is that the local neighborhood of any node in the factor graph is tree-like (see [17] for details). For iteration i , we simplify notation by denoting the message from movie m to user n by $\mathbf{x}_{m \rightarrow n}^{(i)}$ and the message from user n to movie m by $\mathbf{y}_{n \rightarrow m}^{(i)}$. The iteration is initialized with

$$\mathbf{x}_{m \rightarrow n}(v) = \mathbf{x}_m(v) = p_V(v), \mathbf{y}_{n \rightarrow m}(u) = \mathbf{y}_n(u) = p_U(u).$$

The set of all users who rated movie m is denoted \mathcal{U}_m and the set of all movies whose rating by user n was observed is denoted \mathcal{V}_n . The exact update equations are given in Algorithm 1. The group probabilities are randomly initialized by assuming that the initial group (of the user and movie) probabilities are uniform across all groups.

C. Approximate Matrix Completion

Since the primary goal is the prediction of hidden variables based on observed ratings, the IMP algorithm focuses on estimating the distribution of each hidden variable given the observed ratings. In particular, the outputs of the algorithm (after i iterations) are estimates of the distributions for R_{nm} ,

Algorithm 2 Initializing Group Ratings (shown only for users)

Step I: Initialization

Let $i = j = 0$ and $c_m^{(0,0)}(0)$ be the average rating vector of users for movie m .

Step II: Splitting of critics

Set

$$c_m^{(i+1,j)}(u) = \begin{cases} c_m^{(i,j)}(u) & u = 0, \dots, 2^i - 1 \\ c_m^{(i,j)}(u - 2^i) + z_m^{(i+1,j)}(u) & u = 2^i, \dots, 2^{i+1} - 1 \end{cases}$$

where the $z_m^{(i+1,j)}(u)$ are i.i.d. random variables with small variance.

Step III: Recursive soft K -means clustering for $c_m^{(i,j)}(u)$ for $j = 1, \dots, J$.

1. Each user is assigned a soft group membership $\pi_n(u)$ to each of the critics using

$$\pi_n^{(i,j)}(u) \propto \exp\left(-\beta \sqrt{\frac{1}{|\mathcal{V}_n|} \sum_{m \in \mathcal{V}_n} \left(c_m^{(i,j)}(u) - R_{nm}\right)^2}\right)$$

where $\mathcal{V}_n = \{m \in [M] \mid (n, m) \in O\}$ and $g_u = 2^{i+1}$.

2. Update all critics as

$$c_m^{(i,j+1)}(u) \propto \sum_n \pi_n^{(i,j)}(u) c_m^{(i,j)}(u).$$

Step IV: Repeat Steps II and III until the desired number of critics g_u is obtained.

Step V: Estimate of $w(r|u, v)$

After clustering users/movies each into user/movie groups with the soft group membership $\pi_n(u)$ and $\tilde{\pi}_m(v)$, compute the soft frequencies of ratings for each user/movie group pair as

$$w(r|u, v) \propto \sum_{(n,m) \in O: R_{nm}=r} \pi_n(u) \tilde{\pi}_m(v).$$

U_n , and V_m . They are denoted, respectively, as

$$\hat{p}_{R_{nm}|\mathbf{R}_O}^{(i+1)}(r) \propto \sum_{u,v} \mathbf{y}_{n \rightarrow m}^{(i+1)}(u) w(r|u, v) \mathbf{x}_{m \rightarrow n}^{(i+1)}(v)$$
$$\hat{p}_{U_n|\mathbf{R}_O}^{(i+1)}(u) \propto \mathbf{y}_n^{(0)}(u) \prod_{k \in \mathcal{V}_n} \sum_v w(r|u, v) \mathbf{x}_{k \rightarrow n}^{(i)}(v)$$
$$\hat{p}_{V_m|\mathbf{R}_O}^{(i+1)}(v) \propto \mathbf{x}_m^{(0)}(v) \prod_{k \in \mathcal{U}_m} \sum_u w(r|u, v) \mathbf{y}_{k \rightarrow m}^{(i)}(u).$$

Using these, one can minimize various types of prediction error. For example, minimizing the mean-squared prediction error results in the conditional mean estimate (see Figure 2)

$$\hat{r}_{n,m}^{(i)} = \sum_{r \in \mathcal{R}} r \hat{p}_{R_{nm}|\mathbf{R}_O}^{(i)}(r).$$

D. Density Evolution (DE) Analysis

DE is a well-known technique for analyzing probabilistic message-passing inference algorithms that was originally developed to analyze belief-propagation decoding of error-correcting codes and was later extended to more general inference problems [20]. It works by tracking the distribution of the messages passed on the graph under the assumption that the local neighborhood of each node is a tree. While this assumption is not rigorous, we consider that, in Figure 1, the outgoing edges from each user node are attached to movie

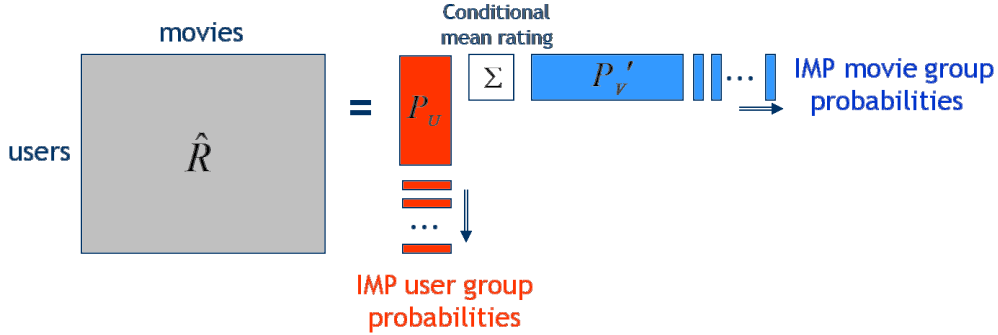


Figure 2. Minimum mean square estimator (MMSE) estimates \hat{R} can be written as a matrix factorization. Each element of Σ represents the conditional mean rating of $w(r|u, v)$ given u, v and each row of P_U/P_V represents a user/movie group probabilities. In contrast to the basic low-rank matrix model, we add non-negativity (to Σ, P_U and P_V) and normalization constraints (to both P_U and P_V).

nodes via random permutations. This is identical to the model used for irregular LDPC codes [21]. For this problem, the messages passed during inference consist of belief functions for user groups (e.g., passed from movie nodes to user nodes) and movie groups (e.g., passed from user nodes to movie nodes). We have derived the DE equations for this problem and currently in process of doing analysis based on them (see [17] for details). Like LDPC codes, we expect to see that the performance of Algorithm 1 depends heavily on the degree structure of the factor graph.

IV. SIMULATION RESULTS WITH REAL DATA MATRICES

A. Details of Training

The key challenge of matrix completion problem is predicting the missing ratings of a user for a given item based only on very few known ratings in a way that minimizes some per-letter metric $d(r, r')$ for ratings. To provide further insights into the proposed factor graph model and the IMP algorithm, we compared our results against three other algorithms: OptSpace [6], SET [7] and SVT [4]. Due to time and space constraints, we have chosen three algorithms among all the available algorithms. OptSpace and the more recent SET appear to be the best (this is also apparent from experimental results), and can handle reasonably large matrix sizes. In some cases, the programs are publicly available (e.g., [6][4]) and others (e.g., [7]) have been obtained from their respective authors. Our program is also publicly available at [22].

To make a fair comparison between different algorithms/models whose complexity varies widely, we have created two smaller submatrices from the real Netflix dataset:

- **Netflix Data Matrix 1** is a matrix given by the first 5,000 movies and users. This matrix contains 280,714 user/movie pairs. Over 15% of the users and 43% of the movies have less than 3 ratings.
- **Netflix Data Matrix 2** is a matrix of 5,035 movies and 5,017 users by selecting some 5,300 movies and 7,250 users and avoiding movies and users with less than 3 ratings. This matrix contains 454,218 user/movie pairs. Over 16% of the users and 41% of the movies have less than 10 ratings.

Also, we hide 1,000 randomly selected user/movie entries as a validation set S . The performance is evaluated using the root mean squared error (RMSE) of prediction on this set defined by

$$\sqrt{\sum_{(n,m) \in S} (\hat{r}_{n,m} - r_{n,m})^2 / |S|}.$$

We primarily focused on the RMSE as a function of the average number of observation ratings per user (i.e., how many ratings, $|O|$, are needed to get each algorithm in shape). Simulations were performed in the very small sample regime (e.g., much less than 0.5% of ratings) by varying the randomly selected average number of observed ratings per user between 1 and 30 and the average results are shown in Figure 3. Note that the choice of parameters for each algorithm (e.g., g_u and g_v for IMP and rank for others) was optimized over the validation set S by running each algorithm multiple times. For IMP, we used hard K-means clustering (i.e., soft K-means clustering with large β) for Algorithm 2 Step III to improve the speed of $w(r|u, v)$ initialization. Also, to make a fair comparison with algorithms that provide unbounded predictions, we clip the out-of-range predictions (i.e., ratings greater than 5 or less than 1), if there are any.

B. Discussion

Our results do shed some light on the performance of recommender systems based on the MP framework. First, we have verified that IMP really does improve the cold-start problem. From simulation results on Netflix submatrices in Figure 3, we clearly see while other matrix completion algorithms perform similarly with large amounts of revealed entries, the IMP algorithm can estimate the matrix very well only after a few observed entries. The performance of other algorithms for users with fewer than 5 ratings is generally poorer than that of the simple movie average algorithm that uses the average rating for each movie as the prediction. The IMP algorithm, however, performs considerably better on users with a very few ratings. This better threshold performance (see the steep RMSE decay) of the IMP algorithm in comparison to other algorithms helps to reduce the cold start problem. It is worth noting that the simple K-means clustering (used for $w(r|u, v)$ initialization) performs worse than movie average

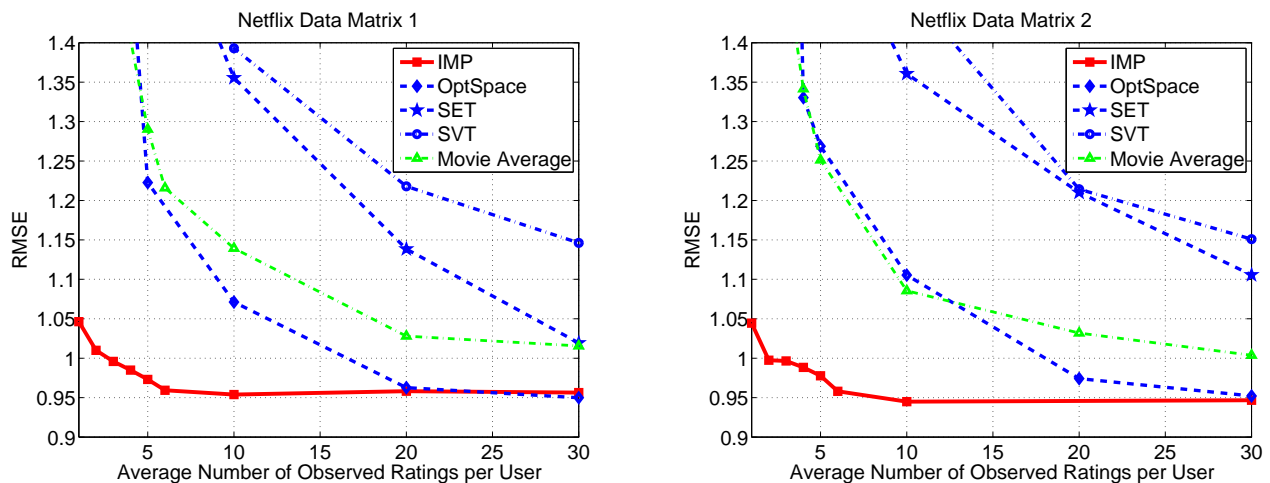


Figure 3. Remedy for the Cold-Start Problem: RMSE performance is compared with other different competing algorithms [6][7][4] on the validation set versus the average number of observations per user for Netflix submatrices.

in the small sample regime (due to space limits, this curve is not shown). This implies that the improvement of IMP for the cold start problem comes from the MP update steps and not the clustering initialization. We believe this will be a major benefit of MP approaches to standard CF problems. Other than these important advantages, each output group has generative nature with explicit semantics. In other words, after learning the density, we can use them to generate synthetic data with clear meanings. These benefits do not extend to general low-rank matrix models easily.

V. CONCLUSIONS

This paper introduces a novel MP framework for the matrix completion problem associated with recommender systems. In contrast to prior work, we model the problem using a generative factor graph model. Based on the model, we introduce the IMP algorithm, which is a low complexity inference method that gives optimal performance when the graph is tree. We demonstrate the superiority of the IMP algorithm by the comparing results against three other algorithms. Simulations are performed with the focus on the cold-start setting (very sparse regime) using Netflix data submatrices. Results show that, while the methods perform similarly with large amounts of data, the IMP algorithm is superior for very small amounts of data and improves the cold-start problem for CF systems in practice. Another advantage of the IMP algorithm is that it can be analyzed using the technique of DE that was originally developed for MP decoding of error-correcting codes. We anticipate that, by including the effects of clustering, this analysis will help us understand the algorithm's impressive performance.

REFERENCES

- [1] E.J.Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, pp. 1–56, 2008.
- [2] E. Candes and Y. Plan, "Matrix completion with noise," to appear in *Proceedings of the IEEE*, 2009.
- [3] [Online]. Available: <http://www.netflixprize.com>
- [4] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," Oct. 2008, [Online]. Available: <http://arxiv.org/abs/0810.3286>.
- [5] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *Arxiv preprint cs.IT/0905.0044*, 2009.
- [6] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," June 2009, [Online]. Available: <http://arxiv.org/abs/0906.2027>.
- [7] W. Dai and O. Milenkovic, "Set: an algorithm for consistent matrix completion," Sept. 2009, [Online]. Available: <http://arxiv.org/abs/0909.2705>.
- [8] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in *Proceedings of the International Conference on Machine Learning*, vol. 25, 2008.
- [9] E. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," March 2009, [Online]. Available: <http://arxiv.org/abs/0903.1476>.
- [10] R. Keshavan, S. Oh, and A. Montanari, "Matrix completion from a few entries," Sept. 2009, [Online]. Available: <http://arxiv.org/abs/0901.3150v4>.
- [11] S. Aditya, O. Dabeer, and B. Dey, "A channel coding perspective of collaborative filtering," Aug. 2009, submitted to *IEEE Trans. on Inform. Theory* [Online]. Available: <http://arxiv.org/abs/0908.2494>.
- [12] S. Vishwanath, "Information theoretic bounds for low-rank matrix completion," Jan. 2010, [Online]. Available: <http://arxiv.org/abs/1001.2331>.
- [13] A. Schein, A. Popescul, L. Ungar, and D. Pennock, "Methods and metrics for cold-start recommendations," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM New York, NY, USA, 2002, pp. 253–260.
- [14] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, USA: The M.I.T. Press, 1963.
- [15] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 44–54.
- [16] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 160–168.
- [17] B.-H. Kim, A. Yedla, and H. D. Pfister, "Message-passing inference on a factor graph for collaborative filtering," April 2010, [Online]. Available: <http://arxiv.org/abs/1004.1003>.
- [18] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer Academic Publishers, 1992.
- [19] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [20] A. Montanari, "Estimating random variables from random sparse observations," Sept. 2007, [Online]. Available: <http://arxiv.org/abs/0709.0145>.
- [21] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [22] [Online]. Available: <http://www.ece.tamu.edu/~hpfister/software/imp>