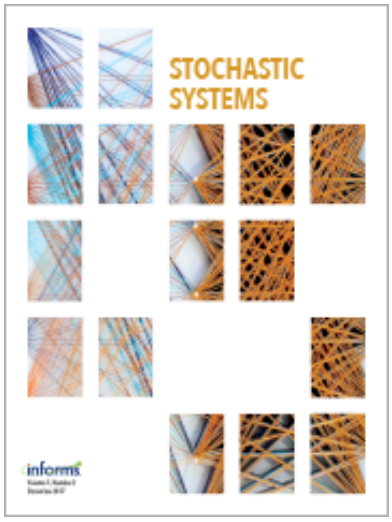
This article was downloaded by: [216.73.217.172] On: 15 November 2026, At: 05:56
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Distributed user Profiling via Spectral Methods

Dan-Cristian Tomozei, Laurent Massoulié

To cite this article:
Dan-Cristian Tomozei, Laurent Massoulié (2014) Distributed user Profiling via Spectral Methods. Stochastic Systems 4(1):1-43. https://doi.org/10.1287/11-SSY036

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*Stochastic Systems*. Copyright 2014 The Author(s). https://doi.org/10.1287/11-SSY036, used under a Creative Commons Attribution License: http://creativecommons.org/licenses/by/4.0/."

Copyright © 2014, The author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Stochastic Systems
2014, Vol. 4, No. 1, 1–43
DOI: 10.1214/11-SSY036

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

DISTRIBUTED USER PROFILING VIA SPECTRAL METHODS

By Dan-Cristian Tomozei* and Laurent Massoulié†

EPFL and INRIA†*

User profiling is a useful primitive for constructing personalised services, such as content recommendation. In the present paper we investigate the feasibility of user profiling in a distributed setting, with no central authority and only local information exchanges between users. We compute a *profile vector* for each user (i.e., a low-dimensional vector that characterises her taste) via spectral transformation of observed user-produced ratings for items. Our two main contributions follow:

(i) We consider a low-rank probabilistic model of user taste. More specifically, we consider that users and items are partitioned in a constant number of classes, such that users and items within the same class are statistically identical. We prove that without prior knowledge of the compositions of the classes, based solely on few random observed ratings (namely $O(N \log N)$ such ratings for N users), we can predict user preference with high probability for unrated items by running a local vote among users with similar profile vectors. In addition, we provide empirical evaluations characterising the way in which spectral profiling performance depends on the dimension of the profile space. Such evaluations are performed on a data set of real user ratings provided by Netflix.

(ii) We develop distributed algorithms which provably achieve an embedding of users into a low-dimensional space, based on spectral transformation. These involve simple message passing among users, and provably converge to the desired embedding. Our method essentially relies on a novel combination of gossiping and the algorithm proposed by Oja and Karhunen.

1. Introduction. Recommendation systems have attracted much interest lately, mostly because of their relevance to core businesses of several major companies (e.g. Amazon, Netflix, Yahoo) who offer large catalogues of products to a vast user base. While the advertisement of highly popular items is straightforward, a significant portion of business stems from sales of only mildly popular items. The latter cannot be advertised indiscriminately, and must be recommended to the "right" users, through targeted recom-

Received September 2011.

Keywords and phrases: Spectral decomposition, random matrix, message passing, distributed spectral embedding, distributed recommendation system.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

mendations. Such companies dispose of large storage and computational resources which enable a centralised computation of recommendations.

In this paper we take a different perspective on the problem of recommendation. Namely, we aim to develop strategies suited to *distributed* operation, where the burden of recommendation is not offloaded to the server, but is rather shared among the users. More specifically, we propose the following two-stage approach for generating recommendations:

- In the first stage, distributed algorithms assign coordinates (or *profiles*) to the users within a certain *profile space*, such that proximity in this space translates to proximity of user taste for content. We say that such algorithms perform *user profiling*.
- In the second stage, recommendations are obtained via simple and distributed algorithms which rely on the primitive of user profiling. We thereby avoid the need for complex machine learning techniques.

The performance of such an approach will depend heavily on the properties of the considered embedding of users in the profile space. For this reason, the focus in this paper is on the first stage of the process, i.e., *user profiling*. Namely, we argue that *spectral* profiling techniques retrieve hidden structure.

The techniques employed in a centralised setting for generating content recommendation are widely known under the generic name of "collaborative filtering". They are typically implemented by a provider who wishes to offer a recommendation service to a large customer base. In such a setting, the information requested from the customers (or users) is typically related both to their identity (via the registration procedure) and to their taste (via the opinions they express regarding the items).

It is not clear to which extent identity information characterises user taste. Moreover, the nature of such information gives rise to privacy concerns. On the contrary, the opinions that users express about items constitute the truly relevant data for solving the problem. For this reason, we advocate a purely agnostic approach to recommending content, which does not use information about the real identity of users, or the nature of content.

Opinions are expressed in the form of ratings assigned by a user to the items she has already purchased. Ratings characterise the satisfaction of a user with respect to a specific item. They are discrete and range from a lowest to a highest value (e.g., number of stars). In particular, the mere fact that a user has consumed or not a specific item can be regarded as a *binary rating*. In this paper we consider the latter form of rating.

Since the number of items on offer from the provider is overwhelmingly large, the vast majority of users only consume a small fraction of items.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Hence, for a typical user, only a small number of ratings are known. In the case of binary ratings, if an item has not been consumed it does not necessarily follow that the user dislikes it. It is possible that the user is simply unaware of the item's existence. Hence, in this case we cannot distinguish between missing ratings and disliked items.

A recent illustration of the possible machine learning techniques and of the corresponding performance comes from the Netflix prize competition [17]. The goal of the competition was to design an algorithm that, when trained on a data set made publicly available by Netflix, would manage to improve prediction accuracy by 10% (measured via Root Mean Squared Error) compared to the proprietary Cinematch algorithm. The designers of such an algorithm [13] were awarded a prize of $1M three years after the start of the competition. It is important to note that the last two years of the three were spent trying to improve the gain in prediction from 8.42% to 10.04%.

The strenuous advancement of the Netflix prize suggests the existence of an important obstacle in the way of achieving high prediction accuracy. A possible explanation is given in the study conducted by Amatriain et al. [1], which showed that when presented with a movie title several times, users provide inconsistent ratings. Hence, there exists an implicit noise in any collection of ratings due to the fact that human taste is variable and not easily quantified. The authors [1] used the RMSE to characterise the distance between two sets of ratings assigned by the same users to the same movies. They found RMSE values between 0.55 and 0.63. It is arguable whether these specific values can be compared to the target RMSE value 0.8563 of Netflix (the considered movies in the conducted study [1] are indeed a subset of the ones in the Netflix data set, but their number is significantly smaller, the user base is different and also much smaller, etc.). However what they suggest is that there exists a RMSE threshold (of the order of the Netflix target RMSE) that cannot be overcome by any recommendation algorithm, seeing as users themselves do not necessarily provide consistent ratings. Since user preference has a significant random component, parameters of a probabilistic model are the best suited to characterize user taste. Throughout this paper we make the following

ASSUMPTION 1. *The taste of each user is characterised by a certain probability distribution defined on the set of all possible ratings for the set of items (which includes the possibility that an item is "not rated"). A user's observed ratings are obtained by sampling her corresponding multi-dimensional distribution.*

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We now restrict ourselves to the binary rating model, where we do not distingush between an item that was not rated and a "disliked" item (i.e., 0-rated items). We describe a natural way of representing the observed purchases. Denote the set of users by $\mathcal{U}$ and the set of items by $\mathcal{F}$. Consider a rectangular matrix $S \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{F}|}$ which we call *the rating matrix* (in the literature it is also referred to as the "term-document" matrix). Each row corresponds to a user and each column corresponds to an item. An entry S_{ui} corresponds to the user-item pair $(u,i) \in \mathcal{U} \times \mathcal{F}$. The entry holds 1 if user u has purchased item i, and 0 otherwise. As previously stated, most of the zero entries in the rating matrix correspond to cases in which a user has not considered purchasing a specific item.

By Assumption 1, each row of the rating matrix is a realisation of a $\{0,1\}$ random $|\mathcal{F}|$-dimensional vector. Relying only on matrix S, we need to assign profiles to the users, such that users with similar taste have similar profiles.

The instances of the problem we consider are extremely large, it is not uncommon to have a user base $\mathcal{U}$ of the order of millions and a catalogue of items $\mathcal{F}$ of the order of tens of thousands. In the case of binary ratings, simply representing the probability distribution in Assumption 1 for a single user requires an exponential amount of memory $2^{|\mathcal{F}|}$. We are thus constrained to consider very simple approximations of such probability laws, for the sake of computational tractability.

Like most proposed models of user taste found in the literature (e.g., [11]), we consider a low-rank model of user taste. Our model is *probabilistic*. More specifically, we make the following

ASSUMPTION 2. *Each entry S_{ui} of the rating matrix is given by an independent Bernoulli random variable of parameter $\bar{S}_{ui}$. The matrix $\bar{S} = (\bar{S}_{ui})$ has rank K, where $K \ll |\mathcal{U}|$ and $K \ll |\mathcal{F}|$.*

In Section 3 we propose a user profiling technique based on the Singular Value Decomposition of matrix S. For a probabilistic model of user taste satisfying Assumption 2, and under further weak statistical assumptions, we prove that a simple voting scheme among users with similar profiles manages to produce accurate recommendations for most of the items with high probability. Furthermore, we use actual movie ratings to compute the profiles of anonymous users of the Netflix system. We find that users with similar profiles have indeed similar taste in movies.

Motivated by the ability to recover hidden structure of the spectral techniques, in the second part of this paper (Section 4), we design a distributed algorithm that computes individual spectral profiles based on local exchanges

among users. We prove almost sure convergence of the algorithm and provide evaluations on a synthetic trace. We conclude in Section 5.

2. Related work. Keshavan et al. [11] consider the problem of low-rank matrix reconstruction from sparsely observed noisy entries. Their modeling approach is more general and overlaps to some extent with the one we adopt in this paper. They propose a reconstruction method using gradient descent on a matrix manifold starting from an initial point given by the SVD of a "trimmed" version of the observed matrix. For a given rank r "well-behaved" $\alpha n \times n$ matrix, given $|E| = \Omega(n \log n)$ uniformly revealed noisy entries, they give a probabilistic upper bound [11, Theorems 1.2 and 1.3] on the Frobenius distance between the reconstructed matrix and the original one as a function of the spectral radius of the noise matrix. We focus on a binary low-rank probabilistic model of user taste, which we analyse using simpler methods. We are interested in retrieving the structure of the hidden model, not its exact parameters; *e.g.*, for a given user we care to deduce the correct ranking of items according to her preference from the noisy observations, without quantifying in absolute this preference.

Instead of using the low-rank fitting of a target partially observed matrix, Srebro et al. [21] take a different path and investigate low-norm factorisation. They advocate this method because it allows unbounded dimensinality in the feature space. Moreover, it can be formulated as a semi-definite optimisation problem. The authors give error bounds for $n \times m$ binary matrix reconstruction from uniformly distributed sample entries of size at least $n \log n$.

In Section 3 we consider a low-rank probabilistic model of user taste. Users are partitioned into classes, such that users within the same class are statistically identical. We show (Theorems 1 and 2) that the profile vectors corresponding to each user computed via spectral methods are clustered around distinct points corresponding to the classes. In this respect, our results are related to spectral clustering. There is a vast literature on the topic of spectral clustering, of which we now give a brief overview.

Among the most relevant work, Ng et al. [18] propose a clustering algorithm based on spectral decomposition. Our results provide a comprehensive analysis by giving conditions under which the underlying partition into classes is retrieved exactly.

In [6], Dasgupta et al. propose an algorithm based on iterative splitting of groups into two subgroups. In contrast, we obtain the desired groups in one go. In [15], McSherry proposes a different clustering method based on projections onto the column space of the original matrix. In both [6] and [15], the probabilistic model of the observed similarity matrix is akin to

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

ours (which extends the classical "planted partition" model). However, our model is less expressive in the sense that we do not consider "mixtures" of the distributions that characterise classes of users, and moreover in our case each user class accounts for a constant fraction of the total number of users. That said, the required separability conditions are similar, and we establish our results under far less stringent conditions on the average degree of the observed matrix. Namely, we require an average degree of order $\Omega(\log(N))$ while they require an order of $\Omega(\log(N)^6)$.

The recent paper [20] by Shi et al. discusses rationales for choosing which eigenvectors to use when performing spectral clustering. This issue is to a large extent complementary to the ones we address in this paper. We could rely on [20] to specify which eigenvectors to keep in our profiling context.

In Section 3.2 we extend the previous results in the more general setting of content recommendation, where a sparse so-called rating matrix (or "term-document" matrix) is available. The literature in this field is extensive, for brevity we mention only a few significant works: [4, 7, 14].

In Section 4.1 we propose a method for computing the eigenvectors of the adjacency matrix of a graph in a distributed manner. A variant of the method was briefly described in [23]. Eigenvector extraction is the object of Oja's algorithm [19]. This basic algorithm was refined by Borkar and Meyn [2]. None of these approaches is distributed however. Our contribution in Section 4.1 consists precisely in augmenting these methods to make them distributed.

A significant contribution towards computing the top k eigenvectors of a symmetric weighted adjacency matrix in a distributed fashion was brought by Kempe and McSherry [10]. The setting is similar to the one we consider in Section 4.1. The authors give bounds on the required running time of their algorithm. Due to the fact that we explicitly introduce noise in our iterations, obtaining such bounds is more difficult in our case. Both our algorithm and the distributed gossiping algorithm in [10] perform iterations to converge to the desired eigenvectors. In the latter, at each iteration all participating nodes first perform a coordinate-update step followed by an orthonormalisation step (that lasts for a determined number of rounds ensuring a bounded error). By explicitly separating the two steps in each iteration, synchronised time becomes a necessary assumption. In contrast, our approach facilitates an asynchronous implementation, since the coordinate-update and the orthonormalisation steps are performed simultaneously with different gains (and thus on different time scales). Furthermore, the algorithm we propose in Section 4.1 uses few elementary computations at each node, whereas in [10] at every iteration each node needs to perform a Cholesky factorisation.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

A similar approach to the one presented in this paper has been taken in a recent publication [12]. The authors aim to determine the eigenvectors of a deterministic matrix based on random sparse observations. They derive useful bounds on convergence time. However, the gossiping stage in their proposed algorithm is treated as a "black box". We explicitly construct an algorithm that incorporates two stages: gossiping, performed on a faster time scale, and Oja's method, performed on a slower time scale. Moreover, we explicitly determine multiple eigenvectors, whereas the authors of [12] focus on determining a single eigenvector, and argue that the extension can be achieved. Finally, we propose an asynchronous algorithm. We show on synthetic data that the latter determines the desired eigenvectors.

3. Spectral recovery of probabilistic taste. We begin by analysing a simple setting in which the observations consist of measures of similarity between users. We prove that a profiling technique based on the spectral decomposition of the square matrix regrouping the observed similarities between pairs of users successfully recovers hidden structure.

We apply these findings to the case in which ratings of items by users are observed. We show that a simple distributed voting algorithm provides asymptotically accurate predictions for most items.

Finally, we observe the benefits of spectral profiling on a real trace.

We make use of notations described in Table 1. Unless otherwise indicated, all vectors are column vectors.

3.1. *Similarity-based profiling.* Denote by $N = |\mathcal{U}|$ the number of users. Let us consider in this first stage that we are given partial observations of user taste similarity in the form of a symmetric matrix

$$A \in \{0,1\}^{N\times N}.$$

Namely, for any two users u, v the elements $A_{uv} = A_{vu}$ take value 1 if users u and v have been evaluated as similar, and value 0 if the users are deemed

TABLE 1
General notations

$A^\top$	The transposition of matrix A
$x^\top$	The transposition of column vector x
e	The all-ones column vector
$\mathrm{diag}(\alpha)$	The $K \times K$ diagonal matrix having the elements on the main diagonal given by the K-dimensional vector α
$\|y\|_\alpha$	The α-norm of vector y, where $(0 < \alpha_k < 1)_{k=1}^K$ and y are column vectors with the same dimension, $\|y\|_\alpha := \sqrt{\sum_k \alpha_k y_k^2}$
$y(k)$, y_k	The k-th element of column vector y

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

dissimilar, or if the similarity between the two users has not been evaluated. By convention $A_{uu} = 0$.

We propose the following *spectral* representation of users based on these partial similarity observations. For some fixed dimension L, extract the *normalised* eigenvectors $x_1, \ldots, x_L$ corresponding to the L largest magnitude eigenvalues of matrix A. We define the *profile space* as $\mathbb{R}^L$. In the profile space, to each user u there corresponds a scaled row vector $\sqrt{N} z_u^\top$, where

$$z_u^\top = (x_1(u), \ldots, x_L(u)).$$

We refer to this vector as the *profile of user* u. The scaling factor $\sqrt{N}$ is introduced to compensate for the fact that the eigenvectors of the $N \times N$ matrix A are taken of norm 1. In what follows, we propose a simple probabilistic user taste model for which this spectral representation of users enables us to retrieve hidden structure.

3.1.1. *Statistical model.* We assume the following probabilistic model of user taste: The N users are partitioned into K classes $C_1, \ldots, C_K$, such that users within the same class are statistically identical. We denote the size of class C_k, $1 \le k \le K$, by $|C_k| = \alpha_k N$, for fixed $\alpha_k > 0$, which are such that $\sum_k \alpha_k = 1$. For any user $u \in \mathcal{U}$ we denote by $k(u)$ her unique corresponding class.

Each pair of classes $1 \le k, \ell \le K$ is characterised by probabilities $b_{k\ell} = b_{\ell k}$ as follows. User $u \in C_k$ and user $v \in C_\ell$ are similar with probability $b_{k\ell}$ and dissimilar with probability $1 - b_{k\ell}$. Moreover, for all pairs of users, their similarity is observed with probability $p = \frac{\omega}{N}$, where ω is a parameter of the model. The similarity of unobserved pairs of users is set to 0 by default.

Equivalently, for all ordered pairs of users $u < v$, the observed similarities $A_{uv} = A_{vu}$ are the outcome of independent Bernoulli random variables of parameters $pb_{k(u)k(v)}$. As previously stated, the diagonal elements are all null, $A_{uu} = 0$.

Given the observed similarity matrix A, without knowledge of the $K \times K$ profile similarity matrix $B = (b_{k\ell})_{k,\ell}$, we wish to recover the partition of users in the unknown classes C_k using their spectral representation. Below we provide sufficient scaling assumptions for which such recovery is possible.

In Table 2 we summarise the notations we have introduced so far:

3.1.2. *Scaling assumptions.* Let us describe the dependence of the various parameters of the model on the number of users.

We have made the Assumption 2, namely that our model is low-rank. Thus, we consider that the number of classes K, as well as the fraction α_k

TABLE 2
Notations and conventions

u, v	Indices referring to users
k, ℓ	User profile indices
$k(u)$	Index of the profile containing user u: $u \in C_{k(u)}$
$\{C_k\}_{k=1}^K$	Partition of users into K unknown disjoint profiles
α	Vector grouping fractions of users per profile: $\lvert C_k\rvert = \alpha_k N$
B	Constant unknown $K \times K$ matrix of probabilities
$p = \frac{\omega}{N}$	Probing probability
L	Dimension of profile space

of users in each class, and the similarity probability matrix $B = (b_{k\ell})_{k,\ell}$ are constant as the number of users in the system grows.

However, we assume that the probing probability p vanishes as the number of users grows. Namely, ω goes to infinity slower than N,

$$\omega \to_N \infty,\ \omega = o(N). \tag{3.1}$$

Hence, each user probes on average a fraction $p = \frac{\omega}{N} \to 0$ of users with whom she evaluates taste similarity. Equivalently, matrix A can be regarded as the adjacency matrix of a random graph on the set of users having average degree of $\Theta(\omega)$.

3.1.3. *Hidden structure recovery.* Theorem 1 below states that under mild conditions, for large N, the spectral representations $\sqrt{N} z_u^\top$ of users are clustered according to their respective classes C_k. Consider the vector

$$\alpha := (\alpha_k)_{k=1}^K.$$

The α-norm of a K-dimensional vector t is $\|t\|_\alpha^2 = \sum_k \alpha_k t_k^2$ (see Table 1). Define the following constant matrix

$$M := (b_{k\ell}\alpha_\ell)_{1 \le k,\ell \le K}.$$

Before stating the theorem, we introduce the following conditions:

(3.2a) The profile space dimension is upper bounded $L \le \operatorname{rank}(M)$.

(3.2b) The L largest magnitude eigenvalues of M have distinct absolute values.

(3.2c) The corresponding eigenvectors $y_1, \dots, y_L$, normalised for the α-norm, satisfy $t_k^\top \ne t_\ell^\top, 1 \le k \ne \ell \le K$, where $t_k^\top := (y_1(k), \dots, y_L(k))$.

(3.2d) $\omega \ge C \log(N)$, for some absolute constant C.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

THEOREM 1. *Under assumptions* (3.2) *and for some positive constants* $c_1 > 0$ *and* $0 < c_2 < 1/4$ *the following claim holds: with high probability* $1-O(N^{-c_1})$*, an increasingly large fraction* $1-O(\omega^{-\frac{1}{2}+2c_2})$ *of users* u *is such that their scaled profile vectors are in a ball of vanishing radius centered in a constant vector* $t_{k(u)}$ *that corresponds to their class* $k(u)$*:*

$$\|\sqrt{N} z_u^\top - t_{k(u)}^\top\| = O(\omega^{-c_2}).$$

Each vector $t_k^\top$ corresponds to a class C_k. If two classes k, ℓ were such that $t_k = t_\ell = t$, the theorem ensures that the profiles of users of both classes would be grouped around the constant vector t in the profile space. Hence, it would be impossible to distinguish between the users of the two classes based solely on their profile vectors. Condition (3.2c) ensures that

$$\|t_k^\top - t_\ell^\top\| = \Omega(1),$$

thus guaranteeing the ability to distinguish between distinct classes for a large enough number of users.

We prove that eigenvectors corresponding to non-zero eigenvalues of matrix A can be used to recover the hidden classes. Condition (3.2a) ensures selection of a suitable dimension L. We impose the technical condition (3.2b) for presentation ease.

Condition (3.2d) gives a lower bound of $\log N$ on order of the required average user neighbourhood size ω. The theorem states that for ω growing at least as fast as $\log N$, the initial partition in profiles C_k can be recovered with high probability for almost all users.

REMARK 1. Concerning condition (3.2d), it is plausible that an even lower requirement for ω (i.e. constant) suffices. Such a case has been explored in the context of bounding the second eigenvalue of the adjacency matrix of a sparse random graph to $O(\sqrt{\omega})$, by removing high degree outlier nodes from the graph (see [9, 5]).

REMARK 2. Other flavours of matrices than the adjacency matrix A could be considered for spectral analysis (e.g. Laplacian matrix, normalised Laplacian matrix). We do not address either of these scenarios in the present work.

We now give the main steps in the proof of Theorem 1. The auxiliary lemmas are proved in the appendix. Consider the matrix

$$\bar{A} := (pb_{k(u)k(v)})_{u,v},$$

which, according to our model, is equal to the expectation $\mathbb{E}A$ of the partially observed similarity matrix A. We can write $A = \bar{A} + Q$, where $Q := A - \bar{A}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

The theorem relies on the fact that the block matrix $\bar{A}$ imposes the eigenvalues and eigenvectors of A, while the perturbation matrix Q has little influence therein, as follows from the lemmas below:

LEMMA 1. *The top L largest magnitude eigenvalues of $\bar{A}$ have distinct absolute values and are order of $\Theta(\omega)$. The normalised eigenvectors $(\bar{x}_\ell)_{\ell=1}^L$ corresponding to these eigenvalues are constant on indices corresponding to each user class. Specifically, using the y_ℓ defined in (3.2c), we can write*

$$\bar{x}_\ell(u) = \frac{y_\ell(k(u))}{\sqrt{N}}, \ \forall u \in \mathcal{U}, \ 1 \le \ell \le L. \tag{3.3}$$

Use the following ordering of the eigenvalues of $\bar{A}$ and A:

$$\begin{aligned} |\bar{\lambda}_1| > |\bar{\lambda}_2| > \cdots > |\bar{\lambda}_L| \\ |\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_L|. \end{aligned} \tag{3.4}$$

We also denote by $\bar{x}_k$ and x_k the corresponding *normalised* eigenvectors.

To control the influence of the perturbation matrix Q we use the following

LEMMA 2. *Consider a square $N \times N$ symmetric 0-diagonal random matrix A such that its elements $A_{ij} = A_{ji}$ are independent Bernoulli random variables with parameters $\mathbb{E}A_{ij} = p_{ij} = a_{ij}\omega N^{-1}$, where the a_{ij} are constant and $\omega = \Omega(\log N)$. Then with high probability the spectral radius of the matrix $A - \mathbb{E}A$ satisfies the upper bound $\rho(A - \mathbb{E}A) \le O(\sqrt{\omega})$.*

We provide a proof of this lemma in Appendix B which relies on the work of Feige and Ofek [9].

Denote by $D = \operatorname{diag}(b_{k(u)k(u)}\frac{\omega}{N}, 1 \le u \le N)$. By application of Lemma 2 to matrix A, we get that the spectral radius of $Q_0 := A - \mathbb{E}A$ is upper bounded by $O(\sqrt{\omega})$ with high probability. Since $\rho(D) \le O(\frac{\omega}{N})$, we have that the spectral radius of $Q = Q_0 - D$ is also upper bounded by $O(\sqrt{\omega})$ with high probability.

The previous two lemmas are instrumental in proving that $\bar{A}$ and A have the same spectral structure:

LEMMA 3. *Using the ordering (3.4) for the eigenvalues of $\bar{A}$ and A, it holds that for all $1 \le k \le K$*

$$\big|\, |\lambda_k| - |\bar{\lambda}_k| \,\big| \le O(\sqrt{\omega}) \qquad \textit{whp}, \tag{3.5}$$

$$\sin(\widehat{x_k, \bar{x}_k}) \le O(\omega^{-1/4}) \qquad \textit{whp}. \tag{3.6}$$

Note that (3.5) also follows from Weyl's inequality.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We conclude by an averaging argument. Lemma 3 shows that with high probability have $\|x_\ell - \bar{x}_\ell\|^2 \leq O(\omega^{-1/2})$, for all $1 \leq \ell \leq L$. Condition (3.2c) guarantees that $\|t_k^\top - t_l^\top\| = \Omega(1)$.

The fraction of users that have their profile vectors at a distance larger than some constant $a > 0$ from the vector t_k corresponding to their class is

$$\frac{1}{N}\left|\left\{u : \|z_u^\top - t_{k(u)}^\top\| \geq a\right\}\right| \leq \frac{1}{N}\sum_{u=1}^{N} \frac{\|\sqrt{N}z_u^\top - t_{k(u)}^\top\|^2}{a^2}$$
$$\overset{(3.3)}{=} \sum_{\ell=1}^{L} \frac{\|x_\ell - \bar{x}_\ell\|^2}{a^2} = O(a^{-2}\omega^{-1/2}).$$

Thus we will be able to conclude the result of the theorem if we can find an a such that $a = o(1)$ and $a^{-2}\omega^{-1/2} = o(1)$. It is easy to see that for instance $a = \omega^{-1/6}$ satisfies these conditions. □

3.2. *Application: Extension to content recommendation.* Let us now consider the scenario exposed in Section 1. Denote again by $N = |\mathcal{U}|$ the number of users and by $F = |\mathcal{F}|$ the number of items. Assume without loss of generality that $N > F$. We can write $F = \gamma N$, with $0 < \gamma < 1$. We consider the rectangular observed rating matrix

$$S \in \{0,1\}^{N \times F},$$

where $S_{ui} = 1$ if user u has rated and liked item i and 0 otherwise. If an entry S_{ui} is null, it is not necessarily true that user u dislikes item i (the item might have simply not been rated).

We propose the following representation of users based on this collected information. For some dimension L, extract the L *normalised left singular vectors* $x_1, \ldots, x_L$ of S, corresponding to its L largest singular values. Like in the previous subsection, consider the L-dimensional profile space $\mathbb{R}^L$, in which we associate a scaled row vector $\sqrt{N}z_u^\top$ to user u, where

$$z_u^\top = (x_1(u), \ldots, x_L(u)).$$

The scaling by $\sqrt{N}$ is again due to the fact that the singular vectors are normalised.

Recall that the Singular Value Decomposition (SVD) of a rectangular real matrix $S = X\Sigma Y^\top$ is always well defined, where matrices $X \in \mathbb{R}^{N \times F}$ and $Y \in \mathbb{R}^{F \times F}$ are unitary and matrix $\Sigma \in \mathbb{R}^{F \times F}$ is positive diagonal.

Again, we claim that users with similar taste will be mapped to close-by locations in the profile space. More specifically, we show that local voting in

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

the profile space provides users with accurate recommendations for most of the items. In what follows, we give a probabilistic model of user taste and scaling assumptions for which we prove these claims.

3.2.1. *Statistical model.* We consider a probabilistic model of user taste similar to the one in Section 3.1.1. The N users are partitioned into K disjoint classes $C_1, \dots, C_K$ and the F items are partitioned into K' disjoint classes $D_1, \dots, D_{K'}$. Users and items in the same class are statistically identical. The size of user class C_k is denoted by

$$|C_k| = \alpha_k N, \quad 1 \le k \le K,$$

while the size of item class $D_{k'}$ is denoted by

$$|D_{k'}| = \beta_{k'} F, \quad 1 \le k' \le K'.$$

The $(\alpha_k)_k$ and $(\beta_{k'})_{k'}$ are strictly positive and sum to 1:

$$\sum_k \alpha_k = \sum_{k'} \beta_{k'} = 1, \quad \alpha_k > 0,\ \beta_{k'} > 0.$$

For any user $u \in \mathcal{U}$ we denote by $k(u)$ her unique corresponding user class, and for any item $i \in \mathcal{F}$ we denote by $k'(i)$ its unique corresponding item class.

Pairs of user and item classes $(C_k, D_{k'})$, $1 \le k \le K, 1 \le k' \le K'$, are characterised by probabilities $r_{kk'}$ in the following way: Any user $u \in C_k$ likes any item $i \in D_{k'}$ with probability $r_{kk'}$ and dislikes it with probability $1 - r_{kk'}$. For all user-item pairs $(u \in \mathcal{U}, i \in \mathcal{F})$, u decides to rate i with probability $p = \frac{\omega}{N}$. Equivalently, any element S_{ui} of the observed rating matrix S is obtained by drawing a Bernoulli random variable of parameter $\bar{S}_{ui} := p \cdot r_{k(u)k'(i)}$.

Given the observed rating matrix S, without knowledge of the $K \times K'$ affinity matrix $R := (r_{kk'})_{k,k'}$, we wish to recover the partition of users in the K classes C_k by making use of their spectral representation. We provide sufficient scaling conditions in what follows.

We summarise the notations in Table 3.

3.2.2. *Scaling assumptions.* We make again Assumption 2 and take the number of classes of users K and of items K' to be constant with N. We assume that the fraction of users in each user class α_k, the fraction of items in each item class $\beta_{k'}$, the ratio between the number of items and the number of users $\gamma = \frac{F}{N}$, as well as the class affinity matrix $R = (r_{kk'})_{k,k'}$ are constant with respect to N. Thus, the class sizes, as well as the total number of items grow linearly with N.

TABLE 3
Notations

N	Number of users
$F = \gamma N$	Number of items
$k'(i)$	Index of the class containing item i: $i \in D_{k'(i)}$
$\{C_k\}_{k=1}^{K}$	Partition of users into K disjoint classes
$\{D_{k'}\}_{k'=1}^{K'}$	Partition of items into K' disjoint classes
α	Vector grouping fractions of users per class: $\lvert C_k\rvert = \alpha_k N$
β	Vector grouping fractions of items per class: $\lvert D_{k'}\rvert = \beta_{k'} F$
R	Constant unknown $K \times K'$ affinity matrix

We assume that the rating probability p vanishes as the number of users N grows to infinity. Specifically, we again impose condition (3.1) on parameter ω. Note that the expected number of items rated by a user is order of $\Theta(\omega)$.

We now formulate an extension of Theorem 1 that we can apply in this setting.

3.2.3. *Content recommendation.* Consider vectors

$$\alpha = (\alpha_k)_{k=1}^{K}, \ \beta = (\beta_{k'})_{k'=1}^{K'},$$

and define the following constant square matrix

$$G := R \operatorname{diag}(\beta) R^{\top} \operatorname{diag}(\alpha) \in \mathbb{R}^{K \times K}.$$

We impose the following conditions, similar to (3.2):

(3.7a) The dimension of the profile space is upper bounded by $L \leq \operatorname{rank}(G)$.

(3.7b) The L largest eigenvalues of matrix G are distinct.

(3.7c) The corresponding eigenvectors $\{g_\ell\}_{\ell=1}^{L}$ normalised for the α-norm satisfy: $\chi_k \neq \chi_\ell$, $1 \leq k < \ell \leq K$, where $\chi_k = (g_1(k), \ldots, g_L(k))$.

(3.7d) $\omega \geq C \log N$ for some constant C.

We can now state the following

THEOREM 2. *Under conditions (3.7), with high probability* $1 - O(N^{-c_1})$ *(where* $c_1 > 0$*), an increasing fraction* $1 - O(\omega^{-\frac{1}{2}+2c_2})$ *of users* u *is such that* $\|\sqrt{N} z_u^{\top} - \chi_{k(u)}^{\top}\| = O(\omega^{-c_2})$*, where* $0 < c_2 < 1/4$.

We briefly explain how this seemingly distinct setup can be mapped to the previous one. Define the following transformation which produces a square

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

matrix:

$$\tau : S \mapsto A = \begin{bmatrix} 0 & S \\ S^\top & 0 \end{bmatrix} \in \mathbb{R}^{(N+F)\times(N+F)},$$

where $S^\top$ denotes the transposition of matrix S. The spectrum of such a matrix is symmetrical (i.e. if σ is an eigenvalue of A, then so is $-\sigma$). Furthermore, the absolute values of the eigenvalues of A are the singular values of S and its singular components are determined by the eigenvectors of A.[1]

We do not give the proof of Theorem 2, since follows closely that of Theorem 1. Essentially, we use the transformation τ to obtain a square matrix, and subsequently we apply Lemma 2 and technical lemmas that we reproduce in Appendix D that play the role of Lemmas 1 and 3 in the proof.

3.2.4. *Characterising performance of a simple voting algorithm.* Let us now analyse a simple recommendation algorithm that relies on local voting in the profile space. We have defined for each user u a scaled L-dimensional profile vector which we denoted by $\sqrt{N}z_u^\top$. In Section 4 we propose a method for computing such profile vectors in a distributed fashion based solely on local information exchange.

Say we wish to characterise the taste of user u for item f. Consider a fixed constant $d > 0$. We define the *d-vicinity* of u as the set of users that have profile vectors at a Euclidean distance of at most d from $\sqrt{N}z_u^\top$ in the profile space. More formally,

$$B(u,d) := \left\{ v \in \mathcal{U} : \|z_u^\top - z_v^\top\|_2 \leq \frac{d}{\sqrt{N}} \right\}.$$

Votes for item f are collected among users in $B(u,d)$ reflecting their appreciation for item f. We denote the number of collected votes by

$$V_u(f) := \sum_{v \in B(u,d)} S_{vf}. \tag{3.8}$$

The following proposition guarantees that, for well chosen d and as the number of users grows to infinity, from the perspective of most users u, the quantities $V_u(f)$ provide an accurate ranking for items with high probability.

[1] To see this, consider an eigenvalue σ of A and its corresponding eigenvector $\zeta = [\begin{smallmatrix} x \\ y \end{smallmatrix}]$, with $x \in \mathbb{R}^{N\times 1}$ and $y \in \mathbb{R}^{F\times 1}$. Since $A\zeta = \sigma\zeta$, we can write:

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Sy \\ S^\top x \end{bmatrix} = \sigma \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } A \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} -Sy \\ S^\top x \end{bmatrix} = -\sigma \begin{bmatrix} x \\ -y \end{bmatrix},$$

and thus $x^\top Sy = y^\top S^\top x = \sigma\|x\|^2 = \sigma\|y\|^2$. Then, we have that x is a left singular vector, and that y is a right singular vector for matrix S. For more details see for instance [22].

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Recall that by (3.7d), $\omega \geq C \log N$.

PROPOSITION 1. *Consider a "polling" radius $d = 2\rho_0$ with $0 < \rho_0 < \frac{1}{6}\inf_{k,k'} \|\chi_k - \chi_{k'}\|$ and a large enough constant C. Then for all but a vanishing fraction of users u, the following property is true: Let $\mathcal{I} \subset \mathcal{F}$ be a sample of items of constant size picked uniformly at random. Then with high probability the item ranking given by the $(V_u(f))_{f\in\mathcal{I}}$ coincides with the item ranking given by the $(r_{k(u)k'(f)})_{f\in\mathcal{I}}$.*

PROOF. We consider a fixed user class C_k. For a specific item $f \in D_\ell$, define $V_k(f)$ to be the total number of votes it received from users in class C_k. It is distributed as a binomial: $\mathrm{Bin}(\alpha_k N, r_{k\ell}\omega/N)$. We want $V_k(f)$ to be contained in the interval

$$V_k(f) \in [\alpha_k r_{k\ell}\omega(1 - \varepsilon_k/r_{k\ell}), \alpha_k r_{k\ell}\omega(1 + \varepsilon_k/r_{k\ell})], \tag{3.9}$$

for a given ε_k. A standard application of a Chernoff bound gives:

$$\mathbb{P}\left[|\mathrm{Bin}(\alpha_k N, r_{k\ell}\omega/N) - \alpha_k r_{k\ell}\omega| \geq \varepsilon_k \alpha_k \omega\right] \leq 2e^{-\alpha_k r_{k\ell}\omega h(\varepsilon_k/r_{k\ell})},$$

where $h(\cdot)$ is a convex increasing function such that $h(0) = h'(0) = 0$.

Let us generalise and define the following events:

$$\mathcal{E}_{k,f} = \{|V_k(f) - \alpha_k r_{k\ell}\omega| < \varepsilon_k \alpha_k \omega\},$$

where ε_k are chosen such that $0 < \varepsilon_k = \inf_{\ell,\ell':|r_{k\ell}-r_{k\ell'}|\neq 0} \frac{|r_{k\ell}-r_{k\ell'}|}{3}$. Then by definition, it must be that on $\mathcal{E}_k := \bigcap_f \mathcal{E}_{k,f}$ the following desirable "separability" property holds: for two items $f \in D_\ell$ and $f' \in D_{\ell'}$, if $r_{k\ell} > r_{k\ell'}$, then $V_k(f) > V_k(f')$. In other words, the votes of users of class C_k preserve the inherent ranking of items.

Let us bound the probability of the event $\mathcal{E}_k$:

$$\mathbb{P}(\mathcal{E}_k) \geq 1 - \sum_\ell |D_\ell| 2e^{-\alpha_k r_{k\ell}\omega h(\varepsilon_k/r_{k\ell})}.$$

We impose the following condition on the constant C of (3.7d):

$$C \geq F_1(\{\alpha_k\}, \{r_{k\ell}\}),$$

where

$$F_1(\{\alpha_k\}, \{r_{k\ell}\}) = \sup_{k,\ell} \frac{1}{\alpha_k r_{k\ell} h(\varepsilon_k/r_{k\ell})}.$$

Then, since $|D_\ell| = \Theta(N)$, it follows that all the events $\mathcal{E}_k$ occur with high probability (i.e., with probability converging to 1 as the number of users N goes to infinity). We have proved the following

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

LEMMA 4. *If $\omega \geq C \log N$ and $C \geq F_1(\{\alpha_k\}, \{r_{k\ell}\})$, then with high probability for any class C_k and any two items $f \in D_\ell$ and $f' \in D_{\ell'}$ such that $r_{k\ell} > r_{k\ell'}$, we have that $V_k(f) > V_k(f') + C''\omega$ for some positive constant C''.*

Denote the total number of votes of user $u \in C_k$ by V_u. It is distributed as a sum of binomials: $\sum_\ell \mathrm{Bin}(|D_\ell|, \frac{\omega}{N} r_{k,\ell})$. The expected number of votes is thus $\mathbb{E}(V_u) = \sum_\ell |D_\ell| \frac{\omega}{N} r_{k\ell} = \Theta(\omega) = m_k \omega$, where $m_k = \sum_\ell \beta_\ell r_{k\ell}$.

By a similar Chernoff bound argument for a given $\delta > 0$ we find

$$\mathbb{P}(V_u \geq m_k \omega + \delta\omega) \leq e^{-m_k \omega h(\frac{\delta}{m_k+\delta})}.$$

Hence,

$$\mathbb{P}(\exists u : V_u \geq m_{k(u)} \omega + \delta\omega) \leq \sum_k |C_k| e^{-m_k \omega h(\frac{\delta}{m_k+\delta})}.$$

Define

$$F_2(\{m_k\}) = \sup_k \frac{1}{m_k h(\frac{\delta}{m_k+\delta})}.$$

The above probability goes to 0 as N goes to infinity under the additional conditions on the constant C stated in the following

LEMMA 5. *Denote $\bar{m} = \max_k m_k$. If $C \geq F_2(\{m_k\})$, all users u verify with high probability*

$$V_u \leq (m_{k(u)} + \delta)\omega \leq (\bar{m} + \delta)\omega.$$

Let us now characterise the performance of the distributed voting scheme. By Assumption (3.7c), the constant vectors corresponding to the user classes $(\chi_k)_k$ are such that $\|\chi_k - \chi_{k'}\| = \Omega(1)$, for any $1 \leq k \neq k' \leq K$.

By hypothesis we have that $\rho_0 < \frac{1}{6} \inf_{k,k'} \|\chi_k - \chi_{k'}\|$. A user u of class C_k queries other users having profile vectors within the ball $B(\sqrt{N} z_u, 2\rho_0)$ about some item f. Thus, for any u such that $\|\sqrt{N} z_u - \chi_k\| \leq \rho_0$, all users in $B(\chi_k, \rho_0)$ are necessarily queried. *Henceforth we only take interest in such users*, as they constitute a fraction of the total number of users that goes to 1 as N goes to infinity.

We denote by $V_u(f)$ the number of votes collected by u for f. Then the difference $W_u(f) := V_u(f) - V_k(f)$ is the error that u makes when estimating $V_k(f)$. It can be seen as the difference between the unwanted votes of users in various classes $C_{k'}$ whose profile vectors fall within $B(\sqrt{N} z_u, 2\rho_0)$ (and hence fall outside the ball of radius ρ_0 around the constant vector corresponding to their own class $B(\chi_{k'}, \rho_0)$) and the votes of users in C_k who fall outside $B(\sqrt{N} z_u, 2\rho_0)$ (and implicitly outside of $B(\chi_k, \rho_0)$).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We can bound $\sum_f |W_u(f)|$ as follows:

$$\sum_f |W_u(f)| \leq \sum_{\substack{u' \in C_k: \\ \sqrt{N} z_{u'} \notin B(\chi_k, \rho_0)}} V_{u'} + \sum_{\substack{u' \notin C_k: \\ \sqrt{N} z_{u'} \notin B(\chi_{k(u')}, \rho_0)}} V_{u'} \leq N_{out} \cdot (\bar{m} + \delta)\omega,$$

where $N_{out} = |\{u' : \sqrt{N} z_{u'} \notin B(\chi_{k(u')}, \rho_0)\}| = o(N)$, as shown in Theorem 2.

We call an item f "misclassified" by u if

$$V_u(f) = V_k(f) + W_u(f) \notin [\alpha_k \omega(r_{k\ell} - \varepsilon_k - \varepsilon_k/2), \alpha_k \omega(r_{k\ell} + \varepsilon_k + \varepsilon_k/2)].$$

If this is the case in $\mathcal{E}_k = \bigcap_f \mathcal{E}_{k,f}$, then necessarily $|W_u| \geq \alpha_k \omega \frac{\varepsilon_k}{2}$, and the number of misclassified items is upper bounded as

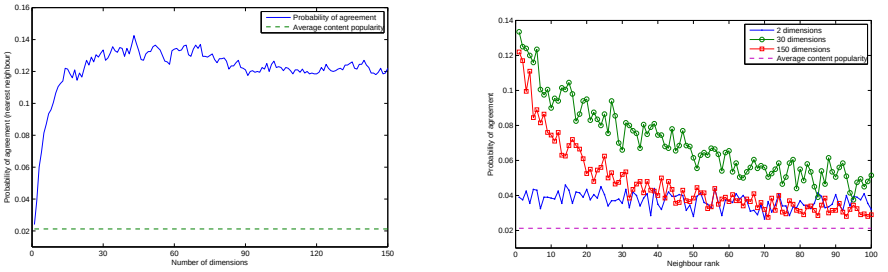
$$|\{f \text{ misclassified by } u\}| \leq \frac{\sum_f |W_u(f)|}{\alpha_k \omega \frac{\varepsilon_k}{2}} \leq \frac{\bar{m} + \delta}{\alpha_k \frac{\varepsilon_k}{2}} N_{out} = o(N).$$

Thus, a constant sample of items chosen at random are well ranked (i.e., the order of the $V_u(f)$ is consistent with the order of the $r_{k\ell}$) with high probability for all users u (within the same class of items the ordering is irrelevant). □

Since ρ_0 might be difficult to estimate, in practice the polling radius can be increased up to a distance that ensures collecting on average $\Theta(\omega) = \mathbb{E}[V_u]$ votes (i.e., an average degree of the order of ω). Another option is to use a weighted average of the collected votes with weights given by a function of the distance to the polled user in the profile space. This point is left for future investigation.

3.3. *Spectral profiling in practice.* In this section we evaluate the benefits of spectral techniques on a real trace provided by Netflix. The Netflix data set contains about 10^8 user ratings for $17,770$ movies by $480,000$ users. The ratings are given in the form of an integer number of stars, ranging from 1 to 5.

We evaluate taste similarity between users that are assigned close-by profiles in the spectral embedding. We do this as follows: We select a set of 2000 users and a set of 2000 movies from the Netflix data set, such that the selected users have given roughly the same number of ratings within the selected movie set. The presence of a rating is viewed in this setting as a sign that the user has viewed that particular movie, and is therefore considered as a binary form of appreciation (the lack of a rating denoting a potential lack of interest for that content). Subsequently, for each user we hide the



(a) Average frequency of agreement with nearest neighbour

(b) Average frequency of agreement with neighbours

FIG 1. *Profiling Proximity.*

rating of one content at random. Using the remaining observed ratings, we build a sparse observed $0-1$ rating matrix S, and we compute the spectral profiles $\sqrt{N}z_u^\top$ of the users. For each user, an ordered list of neighbours is implicitly defined, from the closest to the farthest one in the profile space. We compute over the set of users the average frequency of the occurence of the following event: "a user at distance k has rated the hidden content". The average is taken over the set of users. This "frequency of agreement" reflects the taste proximity of users.

In Figure 1 we plot this "frequency of agreement" for different values of the dimensionality L of the embedding. Content popularity (i.e. the fraction of users having rated it) ranges from 0.45% to 5.3%, and the average popularity of content is 2.13%. We show this value on our plots for comparison. In Figure 1(a) we vary the dimension from 2 to 150 and plot the average frequency of agreement with the nearest neighbour. We notice that there is a peak in this average frequency around roughly 30 dimensions. Subsequently the plotted frequency decays slowly. In Figure 1(b) we plot the average frequency of agreement for 2, 30 and 150 dimensions for the 100 closest neighbours in the profile space. We conclude that an embedding of rank 30 is appropriate to characterise user taste for the selected users in the trace.

It is important to observe that the frequency decays with distance in the profile space. This indicates that spectral profiling manages to capture user taste proximity.

4. Oja's algorithm and beyond. We now propose a method for extracting the eigenvectors of the adjacency matrix of a graph in a distributed manner. Eigenvector extraction is the object of Oja's algorithm [19], the basic version of which is refined by Borkar and Meyn in [2].

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Consider a sequence of symmetric square random matrices $(A_k \in \mathbb{R}^{N\times N})_k$ of common finite mean $A \in \mathbb{R}^{N\times N}$. Oja and Karhunen [19] proposed the following stochastic approximation algorithm for determining the s top eigenvectors of A:

$$\tilde{X}_k = X_{k-1} + A_k X_{k-1}\Gamma_k, \tag{4.1}$$

$$X_k = \tilde{X}_k R_k^{-1}, \tag{4.2}$$

where X_k is an $N \times L$ matrix, Γ_k is a diagonal matrix of gains and R_k^{-1} is a matrix achieving the orthonormalisation of the columns of $\tilde{X}_k$. They prove almost sure convergence under typical assumptions on the sequence of gains, assuming unit multiplicity of the top s eigenvalues, probability density uniformly bounded away from 0 for each of the A_k, and almost sure boundedness and symmetry of the A_k, as well as statistical independence.

A simpler single-step form of the algorithm was also proposed by Oja and Karhunen; the latter produces only an orthonormal basis of the subspace spanned by the top s eigenvectors. The convergence of a slightly modified version thereof when A is positive definite was showed by Borkar and Meyn [2], who introduced the additional factor $\frac{1}{1+\mathrm{Tr}(X_k X_k^\top)}$ and added the additional i.i.d. $\mathcal{N}(0, I)$ noise sequence ξ_k:

$$X_k - X_{k-1} = \frac{a_k}{1 + \mathrm{Tr}(X_k X_k^\top)}[(I - X_{k-1}X_{k-1}^\top)A_{k-1}X_{k-1} + \xi_k].$$

Here (a_k) is an almost typical gain sequence:

$$\sum_k a_k = \infty, \quad \sum_k a_k^2 < \infty, \quad \sup_k \frac{\sum_{n\geq k} a_n^2}{a_k} < \infty.$$

None of these approaches are distributed however. Our contribution consists precisely in augmenting these methods to make them distributed.

4.1. *A method for distributed spectral profiling.* In Section 3 we have demonstrated the benefits of spectral profiling. We have seen that, given a set of N users, it is sufficient that each of them contact on average only $\Omega(\log(N))$ other users at random and determine similarity with them to essentially manage to characterise the profiles of everyone, by application of the spectral transformation.

In this section we consider a (sparse) graph obtained via such a probing process. We develop message-passing algorithms that enable all users to individually compute their spectral profile, while allowing communication only between pairs of neighbours. The algorithms we propose in this section only

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

require the connectivity property of the graph (and hence no specific bound on neighbourhood sizes). However, the scenario of a sparse graph is most appealing, as it fits very well with the model we introduced in the previous section. We consider that the adjacency matrix of the graph does not change over time and prove our claims in this setting. However we conjecture that our results hold even when this matrix varies independently from one time step to the next, so that all the matrices in this sequence have the same finite mean (like in the assumptions of the Oja-Karhunen algorithm).

Let us thus consider a network described by an undirected graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$, where $\mathcal{U}$ is the set of N nodes (also designated interchangeably by "users"), and $\mathcal{E}$ is the set of edges connecting these nodes. We denote by $\mathcal{N}_u$ the set of neighbours of u, and also write $u \sim v$ to indicate that two nodes u, v are neighbours. Nodes that share a link compute their "similarity" value, i.e., a real number A_{uv} which reflects pairwise taste similarity (in Section 3 we were only considering binary values). The computed values define a weighted (undirected) graph with adjacency matrix A.

In this section we show that nodes are able to compute individually via message passing their coordinates in an L-dimensional profile space. These coordinates form a collection of L linearly independent vectors which span the vector space generated by the L eigenvectors corresponding to the top L largest magnitude eigenvalues of A. For binary similarity values, if we interpret the aforementioned computed coordinates as user profiles, Theorem 1 and its corollaries still apply under the same assumptions (3.2); thus, for the considered class-based model of user taste, clusters corresponding to the different classes still emerge. We show this in Appendix E.

Let us now describe the proposed method for distributed user profile computation.

By definition $A_{uu} = 0$ (and thus $\mathrm{Tr}(A) = 0$), thus the matrix is not positive semidefinite (it has necessarily negative eigenvalues). We need to alter the matrix A to guarantee positive semidefiniteness, without changing its eigenvectors. To achieve this, pick for example either one of the two solutions below:

- Compute the value $\Delta := \max_u \sum_v |A_{uv}|$ (e.g., via a distributed voting scheme), and subsequently set the diagonal values of A to a value larger than Δ, e.g., $A_{uu} := \Delta + \epsilon$ for some $\epsilon > 0$. This procedure simply adds $\Delta + \epsilon$ to the eigenvalues of A, thus rendering them positive (Δ corresponds to the maximum degree of the graph in the binary case, and it is known that $|\lambda_1| < |\Delta|$).
- Use matrix A^2, which has the same eigenvectors as A and eigenvalues equal to the square of those of A. The advantage of this second method

is that the ranking of the magnitudes of the eigenvalues is preserved. However, to avoid direct communication with distance-2 neighbours, local gossiping should be used (as described in Section 4.2).

In the rest of this section we consider that one of the solutions above is used and thus that matrix A is rendered positive semidefinite.

Given some fixed number L of target eigenvectors to be extracted, each user u maintains at all time t three sets of variables:

1. An L-dimensional row vector $X_u(t)$, the sought-for eigenvector coordinates;
2. An $L \times L$ matrix $\Phi_u(t)$ and a scalar $\Psi_u(t)$, both playing an auxiliary role in the calculation.

For ease of presentation, we assume slotted time $t = 0, 1, 2, \ldots$, and synchronous updates at all peers. Asynchronous versions will be described and tested in the next section.

Our algorithm then takes the following form:

(4.3)
$$X_u(t+1) - X_u(t) = \frac{a(t)}{Y_u(t)} \left[\sum_{v \sim u} A_{vu} X_v(t) - N X_u(t) \Phi_u(t) + \xi_u(t+1) \right].$$

In the above, $a(t)$ is a gain parameter to be specified, $\xi_u(t+1)$ is a noise term deliberately introduced by user u, and the denominator $Y_u(t)$ is taken equal to

$$Y_u(t) = \max \left(1, |\Psi_u(t)|, \frac{1}{NL^2} \sum_{k,\ell=1}^{L} |(\Phi_u(t))_{k\ell}| \right). \tag{4.4}$$

It is readily seen that the update (4.3) can be computed locally at u, solely relying on variables local to node u and inputs $X_v(t)$ from u's neighbours $v \in \mathcal{N}_u$. The same is also true for the updates of variables Φ_u and Ψ_u, which take the forms:

$$\Phi_u(t+1) = \Phi_u(t) + b(t) \sum_{v \sim u} (\Phi_v(t) - \Phi_u(t)) + f_u(t+1) - f_u(t), \tag{4.5}$$

where $b(t)$ is a gain parameter, $f_u(t)$ is a $L \times L$ matrix, specified by

$$f_u(t) = X_u^\top(t) \sum_{v \sim u} A_{uv} X_v(t), \tag{4.6}$$

and

$$\Psi_u(t+1) = \Psi_u(t) + b(t) \sum_{v \sim u} (\Psi_v(t) - \Psi_u(t)) + g_u(t+1) - g_u(t), \tag{4.7}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

where $g_u(t)$ is a scalar, specified by

$$g_u(t) = X_u(t)X_u^\top(t). \tag{4.8}$$

Before stating the main result of this section, we introduce the technical conditions that will be required from the gain sequences $a(t)$, $b(t)$:

$$a(t),\ b(t) \in [0,1],\ t > 0, \tag{4.9a}$$

$$\sum_{t>0} a(t) = \sum_{t>0} b(t) = +\infty, \tag{4.9b}$$

$$\sum_{t>0} a(t)^2 < +\infty,\ \sum_{t>0} b(t)^2 < +\infty, \tag{4.9c}$$

$$\lim_{t\to\infty} \frac{a(t)}{b(t)} e^{K\sum_{s=1}^{t} a(s)} = 0,\ K > 0 \tag{4.9d}$$

Note that these conditions are satisfied for instance upon taking $a(t) = 1/(t\log(t))$, and $b(t) = t^{-2/3}$. Indeed, with this choice for $a(t)$, it is readily seen that

$$\sum_{s=1}^{t} a(s) \sim \log(\log(t)),$$

and (4.9b) follows. In addition, the quantity in (4.9d) then reads

$$\frac{1}{t^{1/3}\log(t)} e^{K(1+o(1))\log(\log(t))} \le \frac{(\log(t))^{2K-1}}{t^{1/3}},$$

where we have used the upper bound of 1 on the term $o(1)$, and property (4.9d) readily follows.

We are now in a position to state this section's main result:

THEOREM 3. *Assume that the gains $a(t)$, $b(t)$ verify the conditions (4.9). Assume further that the noise terms $\xi_u(t)$ are i.i.d, white Gaussian noise. Assume finally that the overlay graph over which peers communicate is connected, and that matrix A is positive semidefinite (see discussion above). Then the distributed updating algorithm (4.3–4.8) verifies the following property: With probability 1, the columns of $X(t) := (X_u(t))_{u\in\mathcal{U}}$ converge to a collection of L orthonormal vectors spanning the vector space associated with the L largest eigenvalues of A.*

The proof of the theorem is given in Section F. In what follows, we provide some background and intuition for it.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Consider first the main equation, (4.3). If we ignore the denominator $Y_u(t)$, the noise term $\xi_u(t+1)$, and replace the term $NX_u(t)\Phi_u(t)$ by $\sum_v X_u(t)f_v(t)$, where $f(t)$ is as given in (4.6), this equation reads, written in matrix form:

$$X(t+1) - X(t) = a(t)\left[AX(t) - X(t)X^\top(t)AX(t)\right].$$

This is in fact the celebrated Oja algorithm, proposed by Oja and Karhunen [19] to extract precisely the eigenvectors of the largest eigenvalues of A. Oja's algorithm is subject to some stability issues, that Borkar and Meyn [2] proposed to fix by scaling down the right-hand side of the previous equation by some factor $Z(t) = 1 + \sum_{u,k} X_{u,k}^2(t)$, and by adding an extra noise term $\xi(t+1)$ in the bracket in the right-hand side. Thus, the update rule they considered reads:

$$X(t+1) - X(t) = \frac{a(t)}{Z(t)}\left[AX(t) - X(t)X^\top(t)AX(t) + \xi(t+1)\right], \tag{4.10}$$

and is proved in ([2]) to converge with probability 1 to the desired eigenvectors, under assumptions (4.9b),(4.9c) on the gains $a(t)$, and similar conditions on the noise $\xi(t)$ as in our theorem.

However, algorithm (4.10) does not lend itself to a distributed implementation, since neither of the two terms $X^\top(t)AX(t)$ or $Z(t)$ can be computed locally by the users.

To solve this issue, we introduce the auxiliary local variables Φ_u, Ψ_u. The dynamics (4.5–4.7) according to which they evolve is best understood by setting to zero the input terms $f_u(t+1) - f_u(t)$ and $g_u(t+1) - g_u(t)$ in the right-hand side. It then becomes apparent that these dynamics perform local averaging (also known as gossiping in [3]). Thus these eventually converge to a state where all variables $\Phi_u(t)$ coincide with the average $(1/N)\sum_v \Phi_v(0)$ of the original entries.

We can now provide a heuristic argument for the theorem. On a fast time scale, characterised by the gain parameters $b(t)$, the gossiping dynamics converge to almost constant vectors, with

$$\begin{aligned} \Phi_u(t) &\equiv \tfrac{1}{N}\textstyle\sum_v f_v(t), \quad u \in \mathcal{U}, \\ \Psi_u(t) &\equiv \tfrac{1}{N}\textstyle\sum_v g_v(t) \quad u \in \mathcal{U}. \end{aligned}$$

Then on a slower time scale dictated by the gain parameters $a(t)$, the variables of interest $X_u(t)$ follow dynamics very close to (4.10). Indeed, the auxiliary parameters Φ_u, Ψ_u track accurately the desired terms $X^\top(t)AX(t)$ and $Z(t)$ respectively.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

A couple of remarks are in order. The stabilisation by the scaling factor $Z(t)$ in (4.10) seems insufficient in the presence of the additional dynamics (4.5, 4.7). This leads us to introduce our alternative stabilisation via $Y_u(t)$ in (4.4). Also, in problems with dynamics at two time scales a common assumption on the gain parameters is that $a(t)/b(t) \to 0$. In the present case, a stronger form of time scale separation (namely, condition (4.9d)) is needed, to prevent reinforcing instabilities between the two dynamics.

4.2. *Evaluation of an asynchronous version.* In this section we present numerical evaluations on synthetic data. We exhibit convergence of an asynchronous version of the distributed coordinate assignment scheme presented in the previous section on synthetic data generated according to the model presented in Section 3.1.

In Section 4.1 we showed that the distributed algorithm (4.3–4.5) converges almost surely towards L linearly independent vectors spanning the vector space generated by the eigenvectors corresponding to the top L magnitude eigenvalues of the adjacency matrix A.

In the following, we evaluate the asynchronous version of the algorithm. In this setting, each node u keeps track of its own coordinates X_u as well as the gossiped variables Φ_u and Ψ_u. However, instead of explicitly imposing a timescale separation via gains $a(t)$ and $b(t)$ while enforcing a synchronised evolution of all the quantities, we impose distinct rates at which the updates are performed. Namely, the coordinate updates (4.3) are performed independently according to Poisson processes of rate λ, while gossiping (4.4–4.5) is performed independently "pairwise" according to Poisson processes of rate $\mu \gg \lambda$. By pairwise we mean that a pair of nodes $(u, v) \in \mathcal{E}$ will exchange and update their values for Φ and Ψ at rate μ similarly to the randomised gossiping technique from [3].

Furthermore, we replace the adjacency matrix by its square A^2. We choose to do so, since the latter is positive semidefinite, has the same eigenvectors as A, and a spectrum composed of the squared eigenvalues of A. Implicitly, the eigenvectors are ordered according to the magnitude of the eigenvalues of A, instead of their actual values. In turn, this modification alters the function f_u from (4.6), which becomes:

$$f_u^{(2)} = \left(\sum_{v\sim u} A_{uv} X_v\right)^\top \left(\sum_{v\sim u} A_{uv} X_v\right). \tag{4.11}$$

The algorithm executed at each node is summarised in Algorithm 1.

Since sparsity is not preserved by taking the square of A, we cannot simply use A^2 as a new adjacency matrix. Thus, we need to compute products

Algorithm 1 Distributed profiling algorithm

NODE(u)::UPDATE-LOCAL() at rate λ
Local Variables: $X_u, \Pi_u, w_u, X_u^0, \Pi_u^0, \Psi_u, \Phi_u$

1: **if** $w_u = 1$ **then**
2: Retrieve partial product vectors Π_v from $v \in \mathcal{N}_u$
3: $X_u := X_u + \gamma \frac{\sum_{v\sim u} A_{uv}\Pi_v - NX_u\Phi_u}{Y_u(\Psi_u,\Phi_u)}$
4: **else**
5: Retrieve vectors X_v from $v \in \mathcal{N}_u$
6: $\Pi_u := \sum_{v\sim u} A_{uv}X_v$
7: **end if**
8: $w_u := 1 - w_u$

LINK(u, v)::GOSSIP() at rate μ

1: Retrieve local variables at u and v
2: **for** (H, h) in $\{(\Phi, f^{(2)}), (\Psi, g)\}$ **do**
3: $\alpha := \frac{H_u+H_v}{2}$
4: **for** i in $\{u, v\}$ **do**
5: $\delta_i := h_i(X_i, \Pi_i) - h_i(X_i^0, \Pi_i^0)$
6: $H_i := \alpha + \delta_i$
7: **end for**
8: **end for**
9: $X_u^0 := X_u$, $\Pi_u^0 := \Pi_u$, $X_v^0 := X_v$, $\Pi_u^0 := \Pi_u$

A^2X specifically: For some node u, every other call to the UPDATE-LOCAL() procedure computes the partial products $\Pi_u = \sum_{v\sim u} A_{uv}X_v$ (State 6 in the Algorithm). Subsequently, the neighbours' partial product vectors Π_v are used in the coordinate update procedure (4.3) at State 3 and for the gossiping of $f_u^{(2)} = \Pi_u^\top \Pi_u$. Each node additionally stores its previous vectors X_u^0 and Π_u^0 for use in the GOSSIP() procedure.

The only piece of global information required is the number of users in the system (or an approximation thereof). In the UPDATE-LOCAL() procedure, we used a fixed gain γ. The noise component ξ is omitted in the algorithm. Even so, noise is intrinsic to the algorithm as it is introduced by both the gossip averaging and by the fact that exchanges are asynchronous.

For evaluation of convergence, we used a synthetic data trace generated according to our model from Section 3.1. This trace considers 2200 users clustered in 4 classes of sizes $200, 500, 600$ and 900. The probability matrix pB characterising the classes is

$$pB = \frac{1}{100}\begin{pmatrix} 0.5 & 1.5 & 2.0 & 1.0 \\ 1.5 & 0.55 & 1.0 & 2.0 \\ 2.0 & 1.0 & 0.45 & 4.0 \\ 1.0 & 2.0 & 4.0 & 0.5 \end{pmatrix}.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

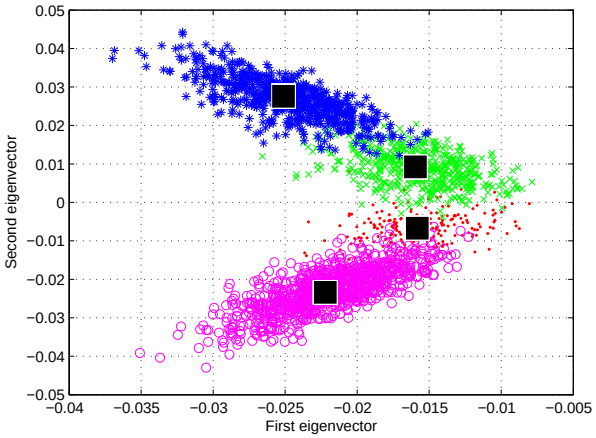


FIG 2. *Pure and noisy user profiles.*

We consider an adjacency matrix A generated using the aforementioned parameters. This matrix is sparse (the average node degree is 40). For visualisation ease, we consider the case $L = 2$ (i.e. the profile space is a 2-dimensional plane). In Figure 2 we plot the first two eigenvectors of the expected adjacency matrix $\bar{A}$. They are constant on each of the 4 classes, hence the plot is constituted of 4 points, which are depicted as the four black squares. They represent the pure user profiles which characterise each of the four classes. Additionally, we plot the first two eigenvectors of the "noisy" adjacency matrix A, with elements belonging to a specific class marked with the same symbol. We have the visual confirmation that despite the sparsity of matrix A, the noisy profiles are grouped around the pure profiles.

We choose a gain $\gamma = 0.001$, an update rate $\lambda = 0.2$ and a gossip rate $\mu = 10$. We initialise the algorithm at a random state. In Figure 3 we plot the time evolution of the proportion of the mass of the two coordinate vectors $X_{\cdot 1}$ and $X_{\cdot 2}$ (aggregated across users) that falls on the space orthogonal to the 2-dimensional eigenspace generated by the first two eigenvectors of matrix A^2. Additionally, we plot the scalar product of the two coordinate vectors. After roughly 400 time units, we observe convergence towards orthogonal vectors spanning the desired eigenspace.

5. Conclusions. In this paper we addressed the problem of distributed user profiling and recommendation.

We first showed that spectral techniques constitute an appealing approach, and obtained novel results on their efficiency, thereby improving

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

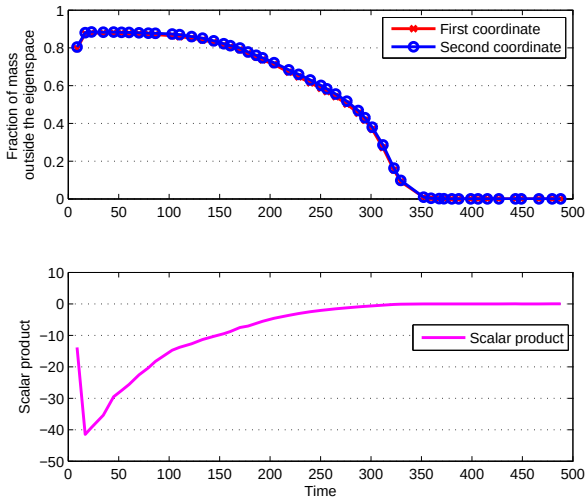


FIG 3. *Convergence of the asynchronous distributed algorithm.*

upon previous literature on the subject of spectral clustering. We showed that for a low-rank probabilistic model of user taste, a simple distributed algorithm based on local votes in the profile space asymptotically achieves accurate prediction of user preference.

We developed techniques for computing eigenvectors in a distributed manner. Our solution combines ideas from Oja's algorithm with gossiping algorithms. From a theoretical standpoint, it essentially relies on a special form of multiple time scale stochastic approximation. The resulting technique may have other applications besides user profiling.

Finally we evaluated our proposed methods on synthetic and actual data traces. We thereby validated our analysis in observing convergence to the desired eigenvectors. We could further show, based on the Netflix prize dataset, that accurate recommendations could be made at limited communication cost based on our spectral embedding.

Several research directions can be envisioned to take this work further. One intriguing problem concerns privacy. While our methods do not rely on direct exchange of sensitive private information, they may nevertheless lead to private information leakage. A distributed solution avoiding the issue is yet unavailable.

Other directions concern the fine tuning of the methods. The issue of analytical selection of the number of eigenvectors has not been addressed here. The recent work of Shi et al. [20] could be an appealing solution.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

APPENDIX A: PROOF OF LEMMA 1

Consider an eigenvalue $\bar{\lambda} \neq 0$ of $\bar{A}$ and a corresponding normalised eigenvector $\bar{x}$, $\|\bar{x}\|_2 = 1$. For $1 \leq u \leq N$,

$$(\bar{A}\bar{x})(u) = \frac{\omega}{N}\sum_{v=1}^{N} b_{k(u)k(v)}\bar{x}(v) = \frac{\omega}{N}\sum_{\ell=1}^{K} b_{k(u)\ell} \sum_{v:k(v)=\ell} \bar{x}(v) = \bar{\lambda}\bar{x}(u).$$

For a large enough N, we have that $|C_k| > 1$, for all k. This is true, since the size of each class grows linearly with N. Then for all k and for all $u, u' \in C_k$, with $u' \neq u$ it follows that $\bar{x}(u') = \bar{x}(u)$.

Denote the value of $\bar{x}(u)$ by $\hat{y}(k(u))$. Then, for all $1 \leq \ell \leq K$,

$$\sum_{\ell'=1}^{K} \alpha_{\ell'} b_{\ell\ell'} \hat{y}(\ell') = \frac{\bar{\lambda}}{\omega}\hat{y}(\ell).$$

Thus $\hat{y}$ is an eigenvector of the $K \times K$ matrix $M = B\mathrm{diag}(\alpha)$ corresponding to its eigenvalue $\frac{\bar{\lambda}}{\omega}$. Since M is a constant matrix, its eigenvalues are also constants. Hence it must be that there exists a constant c such that $\bar{\lambda} = c\omega = \Theta(\omega)$. By Condition (3.2b) we conclude that the top L magnitude eigenvalues of $\bar{A}$ have distinct absolute values.

Finally, since $\bar{x}(u) = \hat{y}(k(u))$ we have that

$$1 = \|\bar{x}\|_2 = \sqrt{N}\|\hat{y}\|_\alpha \iff \|\hat{y}\|_\alpha = \frac{1}{\sqrt{N}}.$$

Since $y = \frac{\hat{y}}{\|\hat{y}\|_\alpha}$, we must have that $x(u) = \hat{y}(k(u)) = \frac{y(k(u))}{\sqrt{N}}$, which proves (3.3). □

Note that a consequence of this result is that $\bar{A}$ can have a number of non-zero eigenvalues that is at least L (by Condition (3.2a)) and at most K (i.e. the maximum number of eigenvalues of matrix M).

APPENDIX B: PROOF OF LEMMA 2

We start by establishing a simple result of stochastic dominance.

By definition a random variable X is dominated for the convex ordering by another random variable Y (written as $X \leq_{cx} Y$) if for any convex function $f : [c, d] \to \mathbb{R}$ such that $\mathbb{E}f(X)$ and $\mathbb{E}f(Y)$ exist, we have $\mathbb{E}f(X) \leq \mathbb{E}f(Y)$.

It is known that (see for instance [16], Theorem 1.5.20): If $X \leq_{cx} Y$, then there exist $\hat{X}$ and $\hat{Y}$ with the same distributions as the original variables, but which are such that $\mathbb{E}(\hat{Y}|\hat{X}) = \hat{X}$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Another result found in [16] states that for some closed interval $[c,d]$, if $X:\Omega\to[c,d]$ and $Y:\Omega\to\{c,d\}$ are two random variables such that $\mathbb{E}X=\mathbb{E}Y$, then $X\leq_{cx}Y$.

In this latter setting, we wish to establish a variant of the former result. Namely, for some closed interval $[c,d]$, if $X:\Omega\to[c,d]$, we wish to construct a random variable $\hat{Y}:\Omega\to\{c,d\}$ supported on the extremities $\{c,d\}$ of the interval, such that $\mathbb{E}X=\mathbb{E}\hat{Y}$ and $\mathbb{E}(\hat{Y}|X)=X$. To achieve this, pick a uniformly distributed random variable $U\sim\mathcal{U}[0,1]$ independent of X. We define the random variable $\hat{Y}:\Omega\to\{c,d\}$ as $\hat{Y}=F(X,U)$, where

$$F(x,u)=d-\mathbb{1}_{\{x\leq u(d-c)+c\}}(d-c).$$

Then

$$\mathbb{E}(\hat{Y}|X)=\mathbb{E}(F(X,U)|X)=d-\frac{d-X}{d-c}(d-c)=X.$$

Since $\hat{Y}$ can possibly take only two values, we can compute the corresponding probabilities:

(B.1)
$$\mathbb{P}(\hat{Y}=d)=\int_c^d P_X(dx)\int_0^{\frac{x-c}{d-c}}du=\frac{\mathbb{E}X-c}{d-c}\text{ and }\mathbb{P}(\hat{Y}=c)=\frac{d-\mathbb{E}X}{d-c},$$

and hence $\mathbb{E}X=\mathbb{E}\hat{Y}$.

Let us now proceed with the proof of the Lemma.

Denote by $Q:=A-\mathbb{E}A$. Denote further by $p_{max}=\max_{ij}p_{ij}$ and by $p_{min}=\min_{ij}p_{ij}$. Then the elements of Q all belong to the interval $[-p_{max},1-p_{min}]$. For a large enough N such that $p_{max}<1$, there exist $\alpha=\frac{1-p_{min}}{1-p_{max}}>1$ and $p=p_{max}$, such that $[-p_{max},1-p_{min}]\subset\alpha[-p,1-p]$.

Consider a symmetric matrix U of independent uniformly distributed random variables $\{U_{ij}=U_{ji}\sim\mathcal{U}[0,1]\}_{i<j}$. Then for each entry Q_{ij} of Q, such that $i<j$, which we regard as a random variable with values in the interval $\alpha[-p,1-p]$, we construct the random variable $Z_{ij}=F(Q_{ij},U_{ij})$ which has the desired property written in matrix form

$$\mathbb{E}(Z|Q)=Q. \tag{B.2}$$

Note that the entries of Z are mutually independent by construction. Furthermore, the random variables defined as $\{Y_{ij}=\alpha^{-1}Z_{ij}+p,\ Y_{ji}=Y_{ij}\}_{i<j}$ are mutually independent Bernoulli random variables of parameter p and hence form the adjacency matrix Y of an Erdos-Renyi graph of parameters (N,p). Denote $\bar{Y}:=\mathbb{E}Y=p(ee^\top-I)$, where e is the all-ones column vector and $e^\top$ denotes transposition.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Let us now prove that the spectral radius of $Z = \alpha(Y - \bar{Y})$ is upper bounded by $O(\sqrt{\omega})$ with high probability.

Since $\omega = \Omega(\log N)$, and the $\{Y_{ij}\}_{i<j}$ are mutually independent, we can apply the results from [9]. Let y be any vector of norm 1 and denote $u := \frac{1}{\sqrt{N}}e$. We can decompose y as follows: $y = ax + bu$, where $a^2 + b^2 = 1$, and x is a vector of norm 1 orthogonal to u, $x \perp u$.

$$|y^\top (Y - \bar{Y})y| \le 2\underbrace{|abx^\top (Y - \bar{Y})u|}_{T_1} + a^2 \underbrace{|x^\top (Y - \bar{Y})x|}_{T_2} + b^2 \underbrace{|u^\top (Y - \bar{Y})u|}_{T_3}.$$

Denote by δ_i the degree of node i and by $\bar{\delta} := \frac{\sum_i \delta_i}{N}$ the average degree. According to Lemma 2.2 from [9] and taking into account the fact that e is an eigenvector of $\bar{Y}$ we have

$$T_1 = |x^\top Y u| \le 2\sqrt{\bar{\delta}}, \text{ with probability } 1 - e^{-\Omega((N\omega)^{1/3})}.$$

By Theorem 2.5 and Claim 2.4 from [9], we have that for every constant $c_1 > 0$, there exists another constant $c_2 > 0$ such that:

$$|x^\top Y x| \le c_2\sqrt{\omega}, \text{ with probability } 1 - N^{-c_1}. \tag{B.3}$$

We will restrict ourselves to constants $c_1 > 1$ for reasons that will become apparent later in the proof. Thus, we can bound the second term with probability $1 - N^{-c_1}$:

$$\begin{aligned} T_2 &\le |x^\top Y x| + |x^\top \bar{Y} x| \overset{(B.3)}{\le} O(\sqrt{\omega}) + \left|\sum_i x_i \sum_{j:i\neq j} p x_j\right| \\ &= O(\sqrt{\omega}) + \left|-p\sum_i x_i^2\right| = O(\sqrt{\omega}) + \Theta\left(\frac{\omega}{N}\right). \end{aligned}$$

Finally, using a Chernoff bound we find

$$T_3 = |\bar{\delta} - \omega| = O(\sqrt{\omega}), \text{ with prob. } 1 - e^{-\Omega(N)}.$$

Thus, $\rho(Z) = O(\sqrt{\omega})$, with probability $1 - N^{-c_1}$. Furthermore, there exists a constant $a > 0$ such that $\rho(Z) < aN$.

Let us now finally characterise the spectral radius of Q.

Using the fact that the spectral radius is a convex function and by Jensen's inequality, we get

$$\rho(Q) = \rho(\mathbb{E}[Z|Q]) \le \mathbb{E}[\rho(Z)|Q]$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We have a random variable $R := \rho(Z)$ supported on $[0, aN]$, such that $\mathbb{P}(R > t) \leq O(N^{-c_1})$ for $t = O(\sqrt{\omega})$ and we wish to deduce that the conditional expectation $S := \mathbb{E}(R|Q)$ is also upper bounded by $O(\sqrt{\omega})$ with high probability.

Since R and S have countable state spaces, it makes sense to consider

$$\beta(s) := \mathbb{P}(R > t | S = s).$$

Since $R : \Omega \to [0, aN]$, and since $\mathbb{E}(R|S) = S$, we have that $\beta(s)(aN) + (1 - \beta(s))t \geq s$, and thus, $\beta(s) \geq (s - t)/(aN - t)$. Denoting $\gamma := \mathbb{P}(S > t + 1)$, we have

$$\mathbb{P}(R > t) = E(\beta(S)) \geq E(\beta(S)\mathbb{1}_{\{S>t+1\}}) \geq \frac{t + 1 - t}{aN - t}\gamma = \frac{\gamma}{aN - t}.$$

Hence

$$\gamma = \mathbb{P}(S > t + 1) \leq (aN - t)\mathbb{P}(R > t) = (aN - t)O(N^{-c_1}) = o(1),$$

since we considered $c_1 > 1$. □

APPENDIX C: PROOF OF LEMMA 3

We will show the two claims (3.5) and (3.6) by induction. Denote (3.5) by $\mathcal{P}_k$ and (3.6) by $\mathcal{Q}_k$.

We will begin by proving $\mathcal{P}_1$ and $\mathcal{Q}_1$. Since we make extensive use of Lemma 2, it is implied that all inequalities in this proof hold with high probability in the sense of Lemma 2 (that is with probability $1 - N^{-c}$ for $c > 0$). All vectors are column vectors and we denote by $x^\top$ the transpose of vector x. Furthermore, for two vectors x and y by $x \perp y$ we mean that their scalar product $x^\top y$ equals zero.

Using the variational characterisation of eigenvalues and Lemma 2, we get:

$$\big||\lambda_1| - |\bar{\lambda}_1|\big| \overset{\text{wlog}}{=} |\lambda_1| - |\bar{\lambda}_1| \leq |\lambda_1| - |x_1^\top \bar{A} x_1| \leq |x_1^\top (A - \bar{A}) x_1| \leq O(\sqrt{\omega}),$$

which proves $\mathcal{P}_1$.

We denote the first eigenvector of A by $x_1 = a_1\bar{x}_1 + b_1\bar{y}_1$, where $a_1^2 + b_1^2 = 1$, $\bar{x}_1$ is the first eigenvector of $\bar{A}$ and $\bar{x}_1 \perp \bar{y}_1$. Then, by making use of $\mathcal{P}_1$, there exist positive constants θ_1 and θ_2 such that

$$|\bar{\lambda}_1| - \theta_1\sqrt{\omega} \leq |\lambda_1| \leq a_1^2 |\bar{x}_1^\top \bar{A} \bar{x}_1| + b_1^2 |\bar{y}_1^\top \bar{A} \bar{y}_1| + \theta_2\sqrt{\omega}.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We took into account the symmetry of $\bar{A}$ and the fact that $\bar{A}\bar{x}_1 = \bar{\lambda}_1\bar{x}_1$. By the Courant-Fischer theorem, we get the following inequality:

$$|\bar{\lambda}_1| - \theta\sqrt{\omega} \leq |\bar{\lambda}_1| - b_1^2(|\bar{\lambda}_1| - |\bar{\lambda}_2|),\ \theta > 0 \tag{C.1}$$

and since the top L eigenvalues of $\bar{A}$ are distinct (by Condition (3.2b)), it holds that $0 < |\bar{\lambda}_1| - |\bar{\lambda}_2| = \Theta(\omega)$. We get that $b_1^2 \leq O(\frac{1}{\sqrt{\omega}})$, thus proving $\mathcal{Q}_1$.

In order to generalise this result, we make use of the following simple lemma:

LEMMA 6. *Let x_1 and $\bar{x}_1$ be two non-orthogonal vectors of norm 1 such that*

$$1 - (x_1^\top \bar{x}_1)^2 \leq O\left(\frac{1}{\sqrt{\omega}}\right). \tag{C.2}$$

Then for all vectors x_2 of norm 1 such that $x_2 \perp x_1$, $(x_2^\top \bar{x}_1)^2 \leq O(\frac{1}{\sqrt{\omega}})$.

PROOF. We have $x_1 = a\bar{x}_1 + b\bar{y}_1$, where $\bar{x}_1 \perp \bar{y}_1$, $\|\bar{y}_1\| = 1$ and $a^2 + b^2 = 1$. By hypothesis (C.2), $b^2 \leq O(\frac{1}{\sqrt{\omega}})$. Thus,

$$x_2^\top x_1 = a x_2^\top \bar{x}_1 + b x_2^\top \bar{y}_1 = 0,$$

and thus, since $a \neq 0$,

$$|x_2^\top \bar{x}_1|^2 = \frac{b^2}{1-b^2}|x_2^\top \bar{y}_1|^2 \leq \theta \frac{1}{\sqrt{\omega}} \frac{\sqrt{\omega}}{\sqrt{\omega} - \theta},$$

where $\theta > 0$ is a constant. □

We proceed by complete induction. We showed $\mathcal{P}_1$ and $\mathcal{Q}_1$. Now assume $\mathcal{P}_\ell$ and $\mathcal{Q}_\ell$ are true for all $1 \leq \ell < k$. We wish to show $\mathcal{P}_k$ and $\mathcal{Q}_k$.

Let us write $x_k = \alpha\bar{x}_k + \beta\bar{y} + \gamma\bar{z}$, where $\bar{y} \in \text{Span}\{\bar{x}_{k+1}, \bar{x}_{k+2}, \dots\}$ and $\bar{z} \in \text{Span}\{\bar{x}_1, \dots, \bar{x}_{k-1}\}$ and $\alpha^2 + \beta^2 + \gamma^2 = 1$. Lemma 6 and the induction hypothesis show that $\gamma^2 \leq O(\frac{1}{\sqrt{\omega}})$. Since

$$|x_k^\top \bar{A} x_k| \leq (1-\gamma^2)|\bar{\lambda}_k| + \gamma^2|\bar{\lambda}_1| \leq |\bar{\lambda}_k| + O(\sqrt{\omega}),$$

we can conclude that

$$\big||\lambda_k| - |\bar{\lambda}_k|\big| \overset{\text{wlog}}{=} |\lambda_k| - |\bar{\lambda}_k| \leq |\lambda_k| - |x_k^\top \bar{A} x_k| + O(\sqrt{\omega}) \leq O(\sqrt{\omega}),$$

thus proving $\mathcal{P}_k$.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We have that,

$$\begin{aligned} |\lambda_k| &\leq |x_k^\top \bar{A} x_k| + |x_k^\top (A-\bar{A}) x_k| \leq \alpha^2 |\bar{\lambda}_k| + \beta^2 |\bar{\lambda}_{k+1}| + \gamma^2 |\bar{\lambda}_1| + O(\sqrt{\omega}) \\ &\leq |\bar{\lambda}_k| + \frac{a}{\sqrt{\omega}} (|\bar{\lambda}_1| - |\bar{\lambda}_k|) - \beta^2 (|\bar{\lambda}_k| - |\bar{\lambda}_{k+1}|) + \theta\sqrt{\omega}, \end{aligned}$$

where a and θ are positive constants. Without loss of generality take $||\lambda_k| - |\bar{\lambda}_k|| = |\bar{\lambda}_k| - |\lambda_k|$ and using $\mathcal{P}_k$ we get

$$\beta^2 \leq \frac{\varphi\sqrt{\omega} + \frac{a}{\sqrt{\omega}}(|\bar{\lambda}_1| - |\bar{\lambda}_k|)}{|\bar{\lambda}_k| - |\bar{\lambda}_{k+1}|} \leq O\left(\frac{1}{\sqrt{\omega}}\right), \ \varphi > 0$$

thus proving $\mathcal{Q}_k$. □

APPENDIX D: TECHNICAL LEMMAS FOR PROVING THEOREM 2

We make the following notation:

$$\bar{S} := (pr_{k(u)k'(i)})_{u,i} = \mathbb{E}S.$$

The following Lemmas characterise the structure of the singular decomposition of $\bar{S}$ and S. They show that the two matrices have the same spectral structure.

LEMMA 7. *For $L \leq K'$, the top L largest singular values of $\bar{S}$ are distinct and of order $\Theta(\omega)$. The normalised left-singular vectors $(\bar{x}_\ell)_{\ell=1}^L$ corresponding to these singular values are constant on indices corresponding to each user class. Specifically, using the g_ℓ defined in (3.7c), we can write*

$$\bar{x}_\ell(u) = \frac{g_\ell(k(u))}{\sqrt{N}}, \ \forall u \in \mathcal{U}, \ 1 \leq \ell \leq L. \tag{D.1}$$

PROOF. Consider a non-zero singular value of $\bar{S}$, $\sigma > 0$ and corresponding left and normalised right singular vectors $\bar{x}$ and $\bar{y}$. Then we can write:

$$\begin{aligned} (\bar{S}^\top \bar{x})(i) &= \sum_{u=1}^N \bar{S}_{ui} \bar{x}(u) = \sigma \bar{y}(i) \overset{\text{not.}}{=} \sigma h(k'(i)), \\ (\bar{S}\bar{y})(i) &= \sum_{i=1}^F \bar{S}_{ui} \bar{y}(i) = \sigma \bar{x}(u) \overset{\text{not.}}{=} \sigma \hat{g}(k(u)), \end{aligned}$$

since $\bar{S}_{ui} = pr_{k(u)k'(i)}$ depends only on the class of u, $k(u)$, and the class of i, $k'(i)$. Since $\bar{y}$ is a right-singular vector of $\bar{S}$, it is also an eigenvector of $\bar{S}^\top \bar{S}$ corresponding to eigenvalue σ^2, and thus after some simplification we

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

have that

$$\frac{\sigma^2}{\gamma\omega^2}\hat{g} = R\mathrm{diag}(\beta)R^\top \mathrm{diag}(\alpha)\hat{g},$$

that is the K-dimensional vector $\hat{g}$ is an eigenvector of the constant matrix $R\mathrm{diag}(\beta)R^\top\mathrm{diag}(\alpha)$ corresponding to the eigenvalue $\frac{\sigma^2}{\gamma\omega^2}$. Since the eigenvalues of a constant matrix are constant, it follows that there exists a constant λ such that $\sigma = \lambda\omega = \Theta(\omega)$. By Condition (3.7b) it follows that the largest L singular values of $\bar{S}$ are distinct. Moreover, matrix $R\mathrm{diag}(\beta)R^\top\mathrm{diag}(\alpha)$ can have at most K distinct non-zero eigenvalues and hence the same holds for the singular values of $\bar{S}$.

Finally, we have that

$$1 = \|\bar{x}\|_2 = \sqrt{N}\|\hat{g}\|_\alpha.$$

Thus, if we consider a vector $g = \frac{\hat{g}}{\|\hat{g}\|_\alpha}$ normalised under the α-norm, Equation (3.3) holds. □

We denote the singular values of $\bar{S}$ and S by

$$\begin{aligned} &\bar{\sigma}_1 > \bar{\sigma}_2 > \cdots > \bar{\sigma}_L > 0 \\ &\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L > 0. \end{aligned} \tag{D.2}$$

and the corresponding left and right normalised singular vectors by $\bar{x}_k$, $\bar{y}_{k'}$, x_k, and $y_{k'}$.

LEMMA 8. *For all* $1 \leq k \leq K$

$$|\sigma_k - \bar{\sigma}_k| \leq O(\sqrt{\omega}) \ \textit{whp}, \tag{D.3}$$

$$\sin(\widehat{x_k, \bar{x}_k}) \leq O(\omega^{-1/4}) \ \textit{whp}, \tag{D.4}$$

$$\sin(\widehat{y_k, \bar{y}_k}) \leq O(\omega^{-1/4}) \ \textit{whp}. \tag{D.5}$$

PROOF. For $\bar{A} = \tau\bar{S}$ we denote by $\bar{\zeta}_k^+$ the normalised eigenvector corresponding to the eigenvalue $\bar{\sigma}_k$ and by $\bar{\zeta}_k^-$ the normalised eigenvector corresponding to the eigenvalue $-\bar{\sigma}_k$. We introduce similar notation for the eigenvectors of $A = \tau S$, namely ζ_k^+ and ζ_k^-. Then it holds that

$$\bar{\zeta}_k^+ = \frac{1}{\sqrt{2}}\begin{bmatrix} \bar{x}_k \\ \bar{y}_k \end{bmatrix}, \bar{\zeta}_k^- = \frac{1}{\sqrt{2}}\begin{bmatrix} \bar{x}_k \\ -\bar{y}_k \end{bmatrix}, \zeta_k^+ = \frac{1}{\sqrt{2}}\begin{bmatrix} x_k \\ y_k \end{bmatrix}, \zeta_k^- = \frac{1}{\sqrt{2}}\begin{bmatrix} x_k \\ -y_k \end{bmatrix}.$$

By Condition (3.7b), we can apply a slightly modified Lemma 3 to matrix A. The only modification we need to make is to change the considered ordering of the eigenvalues – instead ordering them by largest magnitude, we order them decreasingly by value. Since we can apply Lemma 2 to $A - \bar{A}$, it is straightforward to see that the proof also holds in this setting.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

We have thus that (D.3) holds and that with high probability

$$\sin(\widehat{\zeta_k^+, \bar{\zeta}_k^+}) \leq O(\omega^{-1/4}).$$

We now show that (D.4) holds as well. For (D.5) the analysis is similar. We write

$$\begin{aligned} 1 - \langle \zeta_k^+, \bar{\zeta}_k^+ \rangle^2 = 1 - \frac{1}{4}(\langle x_k, \bar{x}_k \rangle + \langle y_k, \bar{y}_k \rangle)^2 &\leq O(\omega^{-1/2}), \\ - \langle \zeta_k^+, \bar{\zeta}_k^- \rangle^2 = -\frac{1}{4}(\langle x_k, \bar{x}_k \rangle - \langle y_k, \bar{y}_k \rangle)^2 &\leq 0. \end{aligned}$$

By summing the two expressions we get that: $\frac{1}{2}(1 - \langle x_k, \bar{x}_k \rangle^2) + \frac{1}{2}(1 - \langle y_k, \bar{y}_k \rangle^2) \leq O(\omega^{-1/2})$. □

APPENDIX E: APPLICABILITY OF THEOREM 1 IN THE SETTING OF SECTION 4

In Section 3.1 we had defined the profile of user u as $\sqrt{N} z_u^\top$, a scaled row vector containing the u-th coordinate of each of the top L eigenvectors of matrix A (i.e., the user profiles are scaled rows of the $N \times L$ matrix $X = (x_1, \ldots, x_L)$ of the top L normalised eigenvectors of A). The algorithms we investigate in Section 4 produce slightly different coordinates for the users: these coordinates form a collection of L linearly independent vectors which span the vector space generated by the L eigenvectors corresponding to the top L largest magnitude eigenvalues of A.

Specifically, for some unknown full rank $L \times L$ matrix $W = (w_1, \ldots, w_L)$ of linear coefficients, we redefine the profile of user u as a scaled row vector $\sqrt{N} \hat{z}_u^\top$ containing the u-th coordinates of an orthonormal basis of the space spanned by the top L eigenvectors (i.e., the rows of the matrix XW): $\hat{z}_u = ((Xw_1)(u), \ldots, (Xw_L)(u))$. Even in this setting, Theorem 1 and its corrolaries still apply under the same assumptions (3.2). Hence, for a large number of users, clusters will emerge, and there is no need to know the matrix of linear coefficients W.

To give an intuition as to why this result still holds, recall Lemma 1 which shows that the top L largest eigenvectors $\bar{X}$ of the block matrix $\bar{A}$ are constant on indices within the same user class. Hence, a linear combination thereof has the same property. Thus, the redefined user profiles concentrate around constant vectors $(\hat{t}_k^\top)_k$ corresponding to the user classes: $\hat{t}_k^\top := ((Yw_1)(k), \ldots, (Yw_L)(k))$, where $Y = (y_1, \ldots, y_L)$ is the matrix of eigenvectors of $M = B\mathrm{diag}(\alpha)$ normalised under the α-norm. As previously stated, the following condition needs to hold in order to distinguish users of

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

separate classes:

The normalized eigenvectors Y under the α-norm are such that $\hat{t}_k^\top \neq \hat{t}_\ell^\top,\ k \neq \ell$. (3.2c′).

Due to the fact that matrix W is full rank, one can show that condition (3.2c′) is equivalent to (3.2c). Moreover, the averaging argument in the proof of the theorem yields:

$$\begin{aligned}\frac{1}{N}\left|\left\{u : \|\hat{z}_u^\top - \hat{t}_{k(u)}^\top\| \geq a\right\}\right| &\leq \frac{1}{N}\sum_{u=1}^{N}\frac{\|\sqrt{N}\hat{z}_u^\top - \hat{t}_{k(u)}^\top\|^2}{a^2} \\ &= a^{-2}\sum_{u=1}^{N}\sum_{\ell=1}^{L}\left[\sum_{i=1}^{L} w_\ell(i)[x_i(u) - \bar{x}_i(u)]\right]^2 \\ &\overset{\text{Cauchy-Schwartz}}{\leq} \|W\|_F^2 \sum_{\ell=1}^{L}\frac{\|x_\ell - \bar{x}_\ell\|^2}{a^2} = O(a^{-2}\omega^{-1/2}),\end{aligned}$$

since the Frobenius norm $\|W\|_F$ of matrix W is constant (all of its elements are subunitary).

APPENDIX F: PROOF OF THEOREM 3

The main ingredient in the proof is the following

LEMMA 9. *The update equation (4.3) is such that for all $t > 0$,*

$$\|X(t+1)\| \leq e^{K\sum_{s=1}^{t} a(s)}\left(\|X(1)\| + M\right) \tag{F.1}$$

for some positive constant K (i.e., that does not depend on t) and some almost surely finite random variable M.

PROOF. Rewrite Equation (4.3) in matrix form as

$$X(t+1) - X(t) = a(t)\left[F((X,\Phi,\Psi)(t)) + D^{-1}(t)\xi(t+1)\right]$$

for some suitable function F, and where $D(t)$ denotes the $N \times N$ diagonal matrix with diagonal entries $Y_u(t)$. By the specific form of the terms $Y_u(t)$, and their role in the function F, it is readily seen that the latter verifies

$$\|F(X,\Phi,\Psi)\| \leq K_1\|X\|$$

for some suitable constant K_1. This readily implies that

$$\|X(t+1)\| \leq (1 + K_1 a(t))\|X(t)\| + a(t)\eta(t),$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

where the $\eta(t) = \|\xi(t+1)\|$ are iid. By induction, one then establishes that

$$\|X(t+1)\| \leq \prod_{s=1}^{t}(1+K_1 a(s)) \left[\|X(1)\| + \sum_{s=1}^{t} a(s)\eta(s)\right].$$

Denote now $\bar{\eta}$ the expectation of $\eta(s)$, and let $M(t) := \sum_{s=1}^{t} a(s)[\eta(s) - \bar{\eta}]$. It is readily seen that $M(t)$ is a uniformly integrable martingale, and hence the supremum $\sup_{s>0} M(s)$ is almost surely finite; denote it by $\hat{M}$. It then follows from the above equation that

$$\|X(t+1)\| \leq \prod_{s=1}^{t}(1+K_1 a(s)) \left[\|X(1)\| + \sum_{s=1}^{t} \bar{\eta} a(s) + \hat{M}\right].$$

Using the elementary inequality $1 + x \leq e^x$, one deduces:

$$\|X(t+1)\| \leq e^{\sum_{s=1}^{t} K_1 a(s)} \left[\|X(1)\| + e^{\sum_{s=1}^{t} \bar{\eta} a(s)} + \hat{M}\right].$$

The result (F.1) then follows by setting $K = K_1 + \bar{\eta}$, and $M = 1 + \max(0, \hat{M})$. □

An additional result that is used is the following

LEMMA 10. *For a given sequence $\epsilon(t)$ with $\lim_{t\to\infty} \epsilon(t) = 0$, the sequence $z(t)$ defined as*

$$z(t+1) = (1 - \lambda b(t))z(t) + b(t)\epsilon(t+1)$$

converges to 0 as t goes to ∞.

PROOF. By induction, one can deduce from the previous expression the identity

$$z(t+1) = z(1)\prod_{s=1}^{t}(1-\lambda b(s)) + \sum_{s=1}^{t} \epsilon(s)b(s) \prod_{\sigma=s+1}^{t} (1-\lambda b(\sigma)).$$

Elementary analysis can then be used to deduce from this last display, assumptions (4.9b),(4.9c) on gains $b(t)$, and convergence of $\epsilon(t)$ to 0 that $z(t)$ also converges to 0.

Indeed, take any fixed $\varepsilon > 0$. Since $\epsilon(t) \to 0$, there exists some t_0 such that $\epsilon(t) < \frac{\varepsilon\lambda}{3}$, for all $t \geq t_0$. Then, we can write:

$$z(t+1) \leq z(1)\prod_{s=1}^{t}(1-\lambda b(s)) \tag{F.2}$$

$$+\sum_{s=1}^{t_0}\epsilon(s)b(s)\prod_{\sigma=s+1}^{t}(1-\lambda b(\sigma)) \tag{F.3}$$

$$+\frac{\varepsilon\lambda}{3}\sum_{s=1}^{t}b(s)\prod_{\sigma=s+1}^{t}(1-\lambda b(\sigma)). \tag{F.4}$$

Term (F.2) develops as

$$z(1)\prod_{s=1}^{t}(1-\lambda b(s)) = z(1)e^{\sum_{s=1}^{t}\log(1-\lambda b(s))} \leq z(1)e^{-\lambda\sum_{s=1}^{t}b(s)} \to 0,$$

by assumption (4.9c). Term (F.3) consists of t_0 terms (a finite number), each of which converges to 0 by the same argument as above. Hence, there exists t_1 such that for all $t \geq t_1$, we have

$$z(1)\prod_{s=1}^{t}(1-\lambda b(s)) + \sum_{s=1}^{t_0}\epsilon(s)b(s)\prod_{\sigma=s+1}^{t}(1-\lambda b(\sigma)) \leq \frac{2\varepsilon}{3}.$$

It can be shown that term (F.4) can be written as:

$$\frac{\varepsilon\lambda}{3}\sum_{s=1}^{t}b(s)\prod_{\sigma=s+1}^{t}(1-\lambda b(\sigma)) = \frac{\varepsilon\lambda}{3}\frac{1}{\lambda}\left(1-\prod_{s=1}^{t}(1-\lambda b(s))\right) \leq \frac{\varepsilon}{3}.$$

We have shown that for all $\varepsilon > 0$, there exists some $t_2 = t_0 \vee t_1$ such that for all $t > t_2$, we have $z(t+1) \leq \varepsilon$. □

The previous lemmas are now used to establish the following result:

LEMMA 11. *The auxiliary variables $\Phi_u(t)$, $\Psi_u(t)$ verify*

$$\begin{aligned}&\lim_{t\to\infty}|N\Phi_u(t) - \textstyle\sum_v f_v(t)| = 0,\\ &\lim_{t\to\infty}|N\Psi_u(t) - \textstyle\sum_v g_v(t)| = 0,\end{aligned} \tag{F.5}$$

i.e. asymptotically, these quantities do track accurately their intended targets.

PROOF. We shall only consider the case of $\Psi_u(t)$, the other one being entirely similar. Rewrite the update rule (4.5) in matrix form as

$$\Psi(t+1) = (I - b(t)\Lambda)\Psi(t) + g(t+1) - g(t), \tag{F.6}$$

where Λ is the so-called Laplacian matrix of the overlay graph: $\Lambda_{uu} = |\mathcal{N}_u|$, $\Lambda_{uv} = -\mathbb{1}_{u\sim v}$ for $u \neq v$. Recall that the Laplacian Λ is positive semi-definite,

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

with eigenvector $e = (1, \dots, 1)^\top$ associated to the eigenvalue 0. Also, when the overlay graph is connected, all other eigenvectors are associated with strictly positive eigenvalues $\lambda > 0$.

We have by definition of g:

$$\begin{aligned} g(t+1) - g(t) &= \{X_u(t+1)X_u(t+1)^\top - X_u(t)X_u(t)^\top\}_u \\ &= \{X_u(t+1)(X_u(t+1)^\top - X_u(t)^\top)\}_u \\ &\quad + \{(X_u(t+1) - X_u(t))X_u(t)^\top)\}_u \\ &= a(t)\{X_u(t+1)[F_u((X,\Phi,\Psi)(t))^\top]\}_u \\ &\quad + \{[F_u((X,\Phi,\Psi)(t))]X_u(t)^\top)\}_u \\ &\quad + a(t)\left\{X_u(t+1)\frac{\xi_u(t+1)^\top}{Y_u(t)} + \frac{\xi_u(t+1)}{Y_u(t)}X_u(t)^\top\right\}_u. \end{aligned}$$

We can rewrite this as $g(t+1) - g(t) = a(t)(r(t+1) + s(t+1) + w(t+1))$, where

$$\begin{aligned} r(t+1) &= \{(X_u(t+1) + X_u(t))[F_u((X,\Phi,\Psi)(t))^\top]\}_u, \\ s(t+1) &= \left\{[2X_u(t) + a(t)F_u((X,\Phi,\Psi)(t))]\frac{\xi_u(t+1)^\top}{Y_u(t)}\right\}_u, \\ w(t+1) &= a(t)\left\{\frac{\|\xi_u(t+1)\|^2}{Y_u^2(t)}\right\}_u \end{aligned}$$

By Lemma 9 and the bound on function F therein we have that

$$\|r(t+1)\| \le K' e^{2K\sum_{s=1}^t a(s)}.$$

Additionally,

$$\mathbb{E}[s(t+1)|\mathcal{F}_t] = 0.$$

Consider z an eigenvector of Λ corresponding to a non-zero eigenvalue $\lambda > 0$ and define

$$\begin{aligned} \epsilon(t+1) &:= \frac{a(t)}{b(t)} z^\top [r(t+1) + w(t+1)], \\ \epsilon'(t+1) &:= \frac{a(t)}{b(t)} z^\top s(t+1). \end{aligned}$$

Then $\lim_{t\to\infty} \epsilon(t) = 0$ almost surely. For the term $\frac{a(t)}{b(t)} z^\top r(t+1)$ the convergence follows from condition (4.9d), while for $\frac{a(t)}{b(t)} z^\top w(t+1)$ it follows from the fact that the $\|\xi_u\|$ have finite variance and from condition (4.9c).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

Denote $\hat{z}(t)$ the scalar product $z^\top \Psi(t)$. One then deduces from (F.6) the equation

$$\hat{z}(t+1) = (1 - \lambda b(t))\hat{z}(t) + b(t)(\epsilon(t+1) + \epsilon'(t+1)).$$

We wish to show that $\hat{z}(t)$ converges to 0. We know that the sequence $z(t)$ defined in Lemma 10 converges to 0. Consider $\Delta(t) := \hat{z}(t) - z(t)$. It verifies

$$\Delta(t+1) = (1 - \lambda b(t))\Delta(t) + b(t)\epsilon'(t+1),$$

with ϵ' such that $\mathbb{E}[\epsilon'(t+1)|\mathcal{F}_t] = 0$.

If we manage to show that $\Delta(t)$ converges to 0, then the same follows for $\hat{z}(t)$. The convergence of $\Delta(t)$ follows from Theorem 1.IV.26. (Robbins-Monro) of [8]. We need to check that the hypothesis of the latter are verified. We only give details for the following condition $\sum_t b(t)^2 \mathbb{E}[\|\epsilon'(t+1)\|^2] \leq +\infty$, as the others are immediately verified.

A sufficient condition is:

$$\sum_t b(t)^2 \mathbb{E}\left[\underbrace{\left(\frac{a(t)^2}{b(t)^2} e^{2K\sum_{s=1}^t a(s)}\right)}_{\to 0} (K''(1+\hat{M}))^2 \|\xi(t+1)\|^2\right] < \infty,$$

where $\hat{M}$ is the supremum of the martingale $M(t) = \sum_{s=0}^t a(s)(\eta(s) - \bar{\eta})$ from Lemma 9, with $\eta(s) = \|\xi(t+1)\|$, and $\bar{\eta} = \mathbb{E}\eta(t+1)$.

We need that $\mathbb{E}(\hat{M}^2\|\xi(t+1)\|^2) < \infty$ and that $\sum_t b(t)^2 < \infty$ to conclude.

$$\mathbb{E}(\hat{M}^2\|\xi(t+1)\|^2) \leq \frac{1}{2}[\mathbb{E}\hat{M}^4 + \underbrace{\mathbb{E}\|\xi(t+1)\|^4}_{<\infty}].$$

By Doob's inequality

$$\mathbb{P}(\sup_{s\leq t} M(s) \geq x) \leq \frac{\mathbb{E}M(t)}{x}.$$

Then

$$\mathbb{P}(\sup_{s\leq t} M(s) \geq x) = \mathbb{P}(\sup_{s\leq t} e^{\theta M(s)} \geq e^{\theta x}) \leq e^{-\theta x}\mathbb{E}[e^{\theta M(t)}] = e^{-\theta x} e^{\sum_{s=0}^t \varphi(\theta a(s))},$$

where $\varphi(y) := \log \mathbb{E}e^{y(\eta - \bar{\eta})}$ (by independence). Moreover, $\varphi(y) \approx Cy^2$, for a small enough y and some constant $C > 0$. Thus, there exists a large enough s^* such that $\varphi(\theta a(s)) \leq C(\theta a(s))^2$ for all $s \geq s^*$

$$\mathbb{P}(\sup_{s\geq 0} M(s) \geq x) \leq e^{-\theta x} e^{\sum_{s=0}^{s^*} \varphi(\theta a(s)) + \sum_{s>s^*} C(\theta a(s))^2} = C' e^{-\theta x},$$

since $\sum_s a(s)^2 < \infty$. Hence,

$$\mathbb{E}\hat{M}^4 = \mathbb{E}(\sup_{s\geq 0} M(s)^4) = \int_0^\infty \mathbb{P}(\hat{M}^4 > t)\, dt \leq A \int_0^\infty e^{-\theta t^{1/4}} = \frac{24A}{\theta^4} < \infty,$$

for some constant $A \geq 0$.

We thus obtain that when decomposing vector $\Psi(t)$ according to the eigenbasis of matrix Λ, one finds vanishing coordinates except along eigenvector e. Since the scalar product $e^\top \Psi(t)$ is always equal to $\sum_u g_u(t)$, by (F.6) the announced result follows. □

To conclude the proof of the theorem, note now that, by the previous lemma, and our specific choice of gain parameters $Y_u(t)$ in (4.4), for large enough t, Equation (4.3) reads in vector form

(F.7)
$$X(t+1) - X(t) = \frac{a(t)}{\max(1, \sum_{k,v} X_{vk}^2(t) + o(1))}[AX(t) - X(t)X^\top(t)AX(t) + o(\|X(t)\|) + \xi(t+1)].$$

As is readily seen, this coincides with the update rule (4.10), except for the $o(\cdot)$ terms. The analysis of [2] for establishing convergence of (4.10) also applies in fact to its perturbed version (F.7), and Theorem 3 follows. □

REFERENCES

[1] Amatriain, X., Pujol, J., and Oliver, N., I like it... i like it not: Evaluating user ratings noise in recommender systems. Volume 5535 of *Lecture Notes in Computer Science*, pages 247–258. Springer Berlin / Heidelberg, 2009. 10.1007/978-3-642-02247-0_24.

[2] Borkar, V. and Meyn, S. P., Oja's algorithm for graph clustering and markov spectral decomposition. In *ValueTools'08: Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools*, pages 1–7, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[3] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D., Randomized gossip algorithms. *IEEE/ACM Transactions on Networking*, 14(SI):2508–2530, 2006. MR2238556

[4] Chaudhuri, K. and Rao, S., Learning Mixtures of Product Distributions Using Correlations and Independence. In R. A. Servedio and T. Zhang, editors, *COLT*, pages 9–20. Omnipress, 2008.

[5] Coja-Oghlan, A., A spectral heuristic for bisecting random graphs. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'05, pages 850–859, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics. MR2298340

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.

[6] Dasgupta, A., Hopcroft, J., Kannan, R., and Mitra, P., Spectral clustering by recursive partitioning. In *ESA'06: Proceedings of the 14th Conference on Annual European Symposium*, pages 256–267, London, UK, 2006. Springer-Verlag. MR2347148

[7] Dasgupta, A., Kannan, R., Hopcroft, J., and Mitra, P., Spectral Clustering with Limited Independence, 2005.

[8] Duflo, M., *Méthodes récursives aléatoires*. Techniques stochastiques. Masson, Paris, Milan, Barcelone, 1990. MR1082344

[9] Feige, U. and Ofek, E., Spectral techniques applied to sparse random graphs. *Random Struct. Algorithms*, 27(2):251–275, 2005. MR2155709

[10] Kempe, D. and McSherry, F., A decentralized algorithm for spectral analysis. In *STOC*, pages 561–568, 2004. MR2121643

[11] Keshavan, R. H., Montanari, A., and Oh, S., Matrix completion from noisy entries. *Journal of Machine Learning Research*, 11:2057–2078, August 2010. MR2678022

[12] Korada, S. B., Montanari, A., and Oh, S., Gossip PCA. In *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS'11, pages 209–220, New York, NY, USA, 2011. ACM.

[13] Koren, Y., The BellKor Solution to the Netflix Grand Prize, 2009.

[14] Kumar, A. and Kannan, R., Clustering with Spectral Norm and the k-Means Algorithm. In *FOCS*, pages 299–308. IEEE Computer Society, 2010. MR3025203

[15] McSherry, F., Spectral partitioning of random graphs. *Proceedings FOCS*, pages 529–537, 2001. MR1948742

[16] Müller, A. and Stoyan, D., *Comparison Methods for Stochastic Models and Risks*. J. Wiley and Sons, 2002. MR1889865

[17] Netflix, Netflix prize. http://www.netflixprize.com.

[18] Ng, A. Y., Jordan, M. I., and Weiss, Y., On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2001.

[19] Oja, E. and Karhunen, J., On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Math. An. and App.*, 106(1), 1985. MR0780319

[20] Shi, T., Belkin, M., and Yu, B., Data spectroscopy: Learning mixture models using eigenspaces of convolution operators. In *ICML*, pages 936–943, 2008.

[21] Srebro, N., Rennie, J. D. M., and Jaakola, T. S., Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems 17*, pages 1329–1336. MIT Press, 2005.

[22] Stewart, G. W., On the early history of the singular value decomposition. *SIAM Review*, 35(4):551–566, 1993. MR1247916

[23] Tomozei, D.-C. and Massoulié, L., Distributed user profiling via spectral methods. *SIGMETRICS Perform. Eval. Rev.*, 38:383–384, June 2010.

EPFL IC-LCA2
BC256 (BC Building)
Station 14
1015 Lausanne
Switzerland
E-mail: dan-cristian.tomozei@epfl.ch

INRIA-Microsoft Research
Campus de l'École polytechnique
Bâtiment Alan Turing
1 rue Honoré d'Estienne d'Orves
91120 Palaiseau France
E-mail: laurent.massoulie@inria.fr

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 05:56 . For personal use only, all rights reserved.