Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

To Recommend or Not? A Model-Based Comparison of Item-Matching Processes

Serina Chang, Johan Ugander
Stanford University
serinac@cs.stanford.edu, jugander@stanford.edu

Abstract

Recommender systems are central to modern online platforms, but a popular concern is that they may be pulling society in dangerous directions (e.g., towards filter bubbles). However, a challenge with measuring the effects of recommender systems is how to compare user outcomes under these systems to outcomes under a credible counterfactual world without such systems. We take a model-based approach to this challenge, introducing a dichotomy of process models that we *can* compare: (1) a "recommender" model describing a generic item-matching process under a personalized recommender system and (2) an "organic" model describing a baseline counterfactual where users search for items without the mediation of any system. Our key finding is that the recommender and organic models result in dramatically different outcomes at both the individual and societal level, as supported by theorems and simulation experiments with real data. The two process models also induce different trade-offs: regularization improves the mean squared error of matches in both settings, but at the cost of less diverse (less radical) items chosen in the recommender model but more diverse (more radical) items chosen in the organic model. These findings provide a formal framework for how recommender systems may be fundamentally altering how we interact with content, in a world increasingly mediated by such systems.

Introduction

Personalized recommender systems guide the modern online experience. These systems suggest friendships and groups on social networking sites (Gupta et al. 2013; Kloumann and Kleinberg 2014), recommend movies and music on content platforms (Nguyen et al. 2014; Anderson et al. 2020), and filter news on the web to consumers (Das et al. 2007). As recommender systems increasingly shape the content we consume, we have become more critical of their potential for unintended societal consequences. For example, one concern that has received attention by academics (Nguyen et al. 2014; Flaxman, Goel, and Rao 2016; Rastegarpanah, Gummadi, and Crovella 2019) and the public (Singer 2011) alike is that these systems may be pulling society towards "filter bubble" dynamics, where individuals become isolated from viewpoints besides their own (Pariser 2011).

However, a key challenge with measuring the true effects of recommender systems is that we only observe user outcomes under recommender systems, but we are missing a credible assessment of user outcomes in the absence of these systems. To understand definitively whether recommender systems have a significant impact on filter bubbles, or any other societal phenomena, we need a systematic way to compare these two scenarios and their outcomes against each other. To address this need, in this work we introduce a dichotomy of process models that describes these two worlds: one with and one without recommender systems. In our *recommender model*, a personalized recommender system serves items to users; in our *organic model*, the user searches for items "organically" without the mediation of any system. Our objective is to systematically compare these two models, with the organic model acting as a baseline counterfactual against which the item-matching behavior of the recommender system can be evaluated.

Consider the example of a user who is new to a movie distribution platform and seeking to watch one movie for the night. Under our organic model, the user searches for movies herself. She wants to watch the movie that best fits her interests, but does not know the contents of any of the candidate movies. During this search process she has access only to noisy signals of each movie (e.g., trailers), and based on these noisy signals, she must estimate each movie's content. To complete the matching process, she chooses to watch the movie that she estimates best matches her own interests.

Compare this first model against the case where the platform has a personalized recommender system. Under our recommender model, the situation is now flipped: whereas the user knew her own interests well but previously had to estimate the contents of the movies, now the system knows the movies on the platform well, but must estimate the user's interests. In a similar manner, the system gathers a noisy sample from the user (e.g., by asking her to name one movie she recently enjoyed). Based on this noisy sample, the system then estimates her interests and recommends her the movie it believes is the best match.

What connects these two settings is the act of trying to match users to items with only noisy information about one side; what differs, however, is which information is known and who is doing the matching (the user vs. the recommender system). We find that this switch in perspective between the two models results in significant differences in outcomes. We characterize these differences through the lenses of both (1) individual metrics (what is the expected loss for a given user? Does this differ across users?) and (2) population metrics (what is the average user loss? Which items tend to be selected overall?). We show that the two models diverge in important ways for all three metrics; for example, the organic model favors mainstream users, while the recommender model serves all users equally well. Furthermore, we document the notably varying effects of regularization on each model. While regularization can reduce average user loss in both models, it also causes the models to diverge further, encouraging the recommender model to choose increasingly similar items while leading the organic model to diversify its selection.

Thus, even in these distilled settings, the intersection of our models, metrics, and algorithmic decisions creates a rich environment in which we can investigate the effects of recommender systems. We summarize our contributions as:

1. A framework of two contrasting models that capture organic search and recommended item-matching as comparable processes;
2. Theorems proving key differences between the models from both the individual and population perspectives;
3. Simulations demonstrating that our findings translate from the theorem settings to realistic data (MovieLens).

By introducing a formal framework and deriving analytical results, our work makes significant progress towards characterizing the true impacts of recommender systems. Such formal analysis is highly valuable because we can show that our findings about the differences between organic search and recommender systems generalize to a broad class of models representing each process. Our simulations furthermore demonstrate that our results translate to real-world data, using over a million movie ratings from real users and movies. Collectively, our work lays out a new, principled approach for how the effects of recommender systems can be meaningfully analyzed relative a counterfactual world without such systems.

Related Work

The widespread adoption of personalized recommender systems has led to diverse investigations of the potential societal consequences of these systems. One body of literature tackles bias in recommender systems (Chen et al. 2020; Bozdag 2013), examining how they may systematically underrepresent minority views (Stoica and Chaintreau 2019), serve predictions of uneven quality across user groups (Yao and Huang 2017), or fail to recommend valuable items to certain users, e.g., showing job opportunities in STEM fields to fewer women than men (Lambrecht and Tucker 2019). Empirical observations, however, have often been mixed in nature, e.g., documenting how systems sometimes favor long-tail items (Fleder and Hosanagar 2009; Brynjolfsson, Hu, and Simester 2011) or (over-)favor popular items (Abdollahpouri, Burke, and Mobasher 2017).

Another topic of societal concern is the possibility that personalized recommendations are pushing individuals into "filter bubbles" (Pariser 2011). Social media users are known to selectively share content and connect to friends who agree with their existing opinions (An et al. 2014; Garimella et al. 2018), and there is some evidence that recommender systems exacerbate this dynamic of ideological segregation (Bakshy, Messing, and Adamic 2015; Flaxman, Goel, and Rao 2016). Recommender systems seem to have a narrowing effect in other contexts as well: e.g., the sets of movies (Nguyen et al. 2014) and songs (Anderson et al. 2020) recommended to users tend to be less diverse than the content that users find on their own. However, some studies have pointed out that the role of recommender systems is modest compared to the impact of user choice (e.g., whether to click on a recommended story) on narrowing consumption diversity (Bakshy, Messing, and Adamic 2015). Other works have made the case that recommender systems actually *increase* diversity in exposure (Flaxman, Goel, and Rao 2016) and widen users' interests (Hosanagar et al. 2013).

Part of the reason why it is so challenging to measure the true effects of recommender systems—and perhaps why empirical studies have not been able to reach a consensus—is that we typically only observe user outcomes under the recommender system, but we cannot assess user outcomes in the absence of these systems. Claims, e.g., that users are entering filter bubbles, must be understood relative to some counterfactual baseline. Thus, it is often helpful to design models that enable us to "observe" and compare these hypothetical realities. Models have been used to analyze the impacts of recommender systems on polarization and user opinions (Dandekar, Goel, and Lee 2013); to investigate their role in accelerating hegemonic dynamics (Stoica and Chaintreau 2019) and inequalities in group visibility on social media (Fabbri et al. 2020); to simulate the effects of recommender systems on item, user, and rating diversity in movies (Szlávik, Kowalczyk, and Schut 2011); and finally, to understand the consequences of feedback loops, when the recommender system is trained on data that was influenced by the system (Chaney, Stewart, and Engelhardt 2018).

Using models to study the impact of recommender systems can be powerful when it is difficult to observe "what-ifs" in real life. In the present work, we take a model-based approach to ask arguably the most fundamental "what-if" question about recommender systems: what if there were no recommender system at all? We analyze our models using first-principles measures such as the expected item match for each user, their expected loss, and the variance over the population of matched items. These measures form the building blocks for many of the downstream phenomena of interest related to recommender systems, including polarization, filter bubbles, user satisfaction, user retention, and bias and fairness. Implicitly embedded in these more complex processes are the metrics we study, and thus, our work forms a foundation upon which future researchers can build.

Models and Metrics

Our models describe two contrasting processes through which users could be matched with items of choice, e.g., movies, news articles, or consumer products. In this section, we introduce the notation and formal dynamics of our models. Then, we define the individual and population-level metrics we will use to compare the models' outcomes.

Model Definition

In both models, we will have m users and n items. Each user and each item has a latent position; for example, this could represent movie attributes in the movie context or ideology in the context of news articles. We denote user positions as $x_i \in \mathbb{R}^d$ and item positions as $y_j \in \mathbb{R}^d$, where d represents the dimensionality of the latent space. Overall, the set $\mathcal{X} = \{x_i\}_{i=1}^m$ represents all user positions and the set $\mathcal{Y} = \{y_j\}_{j=1}^n$ represents all item positions. Our analysis will make very weak distance-based assumptions about the interest/utilities of users for items: we simply assume that user i's enjoyment of item j is monotonically decreasing in the distance between x_i and y_j. Thus, under both models, the decision-making agent (either the user or the recommender system) wants to match the user to the item whose position is closest to hers.

Organic model. In this model, the user will "organically" search through the collection of items and choose one for herself (Figure 1a). When surveying each item j, she only has access to a noisy sample of its true position, y_j. She draws her sample $z_j^{(i)}$ from $N(y_j, \Sigma_i)$, where Σ_i represents the covariance in her noise. Then, the user makes an estimate $\hat{y}_j^{(i)}$ of item j's true position. In the simplest case, the user takes the maximum likelihood estimate (MLE) based on her single sample of y_j:

$$\hat{y}_j^{(i,\text{MLE})} = z_j^{(i)}. \tag{1}$$

We compare this MLE to an alternate case where the user applies some form of shrinkage (James and Stein 1961) to her estimates. For example, if the user assumes a Gaussian prior on the item positions, she could take a maximum a posteriori (MAP) estimate of each, shrinking the estimate towards the item mean:

$$\hat{y}_j^{(i,\text{MAP})} = (\Sigma_{\text{item}}^{-1} + \Sigma_i^{-1})^{-1}(\Sigma_{\text{item}}^{-1}\mu_{\text{item}} + \Sigma_i^{-1} z_j^{(i)}), \tag{2}$$

where μ_{item} and Σ_{item} are the mean and covariance of the item distribution, respectively. This style of shrinkage can also be thought of as regularization towards the item mean. In an empirical setting without a clear prior, Empirical Bayes shrinkage (Efron and Morris 1976) would be preferred.

After surveying all n items, the user will have constructed estimates for every item's position. Since we assume utility is monotonically decreasing in distance, the user will then choose the item whose estimate is closest to her own position, x_i. Let $k_i^{(\text{org})}$ represent the item chosen by user i in the organic model. Then,

$$k_i^{(\text{org})} = \underset{j \in [n]}{\arg\min}\, ||x_i - \hat{y}_j^{(i)}||. \tag{3}$$

We let $y_{k_i}^{(\text{org})}$ denote the position of the chosen item $k_i^{(\text{org})}$, and will later study properties of $y_{k_i}^{(\text{org})}$ as a random variable that inherits its randomness from the user's estimates, $\hat{y}_j^{(i)}$.

Recommender model. In this model, a recommender system takes on the burden of search instead of the user (Figure 1b). When user i comes onto the platform, the system gathers a noisy sample of i's true position, x_i. This sample $z_i^{(r)}$ is drawn from $N(x_i, \Sigma_r)$, where Σ_r represents the covariance in the recommender's noise. Then, just as the user estimated item positions based on her noisy samples, the system makes an estimate $\hat{x}_i^{(r)}$ of user i's true position based on its sample of the user. Again, we first consider the case where the recommender system takes the MLE:

$$\hat{x}_i^{(r,\text{MLE})} = z_i^{(r)}. \tag{4}$$

We will again compare the MLE to the case where the recommender system applies shrinkage to its estimates of user positions. With a Gaussian prior, the MAP estimate of user i's position is

$$\hat{x}_i^{(r,\text{MAP})} = (\Sigma_{\text{user}}^{-1} + \Sigma_r^{-1})^{-1}(\Sigma_{\text{user}}^{-1}\mu_{\text{user}} + \Sigma_r^{-1} z_i^{(r)}), \tag{5}$$

where μ_{user} and Σ_{user} are the mean and covariance of the user distribution, respectively. In the absence of a prior, Empirical Bayes shrinkage would again be preferred.

As a reversal of the organic model, we assume that the recommender system has perfect knowledge of $\mathcal{Y}$, the set of true item positions, but can only estimate the user's true position. Let $k_i^{(\text{rec})}$ represent the chosen item for user i under the recommender model. Similar to the organic model, the system will choose the item whose position is closest to $\hat{x}_i^{(r)}$:

$$k_i^{(\text{rec})} = \underset{j \in [n]}{\arg\min}\, ||\hat{x}_i^{(r)} - y_j||. \tag{6}$$

Here we again let $y_{k_i}^{(\text{rec})}$ denote the position of the matched item $k_i^{(\text{rec})}$, where $y_{k_i}^{(\text{rec})}$ is a random variable that inherits its randomness from the system's estimates, $\hat{x}_i^{(r)}$.

We deliberately set up the recommendation process to be as similar as possible to the organic process—the structure of the noise and logic in the choice function are identical—so that we can compare the models on the basis of who/what controls the matching and what information they have access to. We do not, for example, model the process of users reacting to recommendations. Instead, we assume that $k_i^{(\text{rec})}$ will be the user's match, comparing it directly to the match $k_i^{(\text{org})}$ under the organic model. Furthermore, we note that both models are made up of distinct modules (sampling, estimation, choice), each of which could be swapped out for more complex processes. For example, in the organic model, one might be interested in other decision-making rules for the user, such as those that make explore/exploit trade-offs or capture risk aversion (in the case of non-uniform sample noise), or in the recommender model, the process of estimating user positions could be derived from a specific algorithm. Thus, our models describe flexible item-matching processes that can be extended to encompass many real-world systems.

Metrics

We are interested in comparing the outcomes of these models through two lenses, at the individual and at the population levels. At the individual level, we will treat the squared

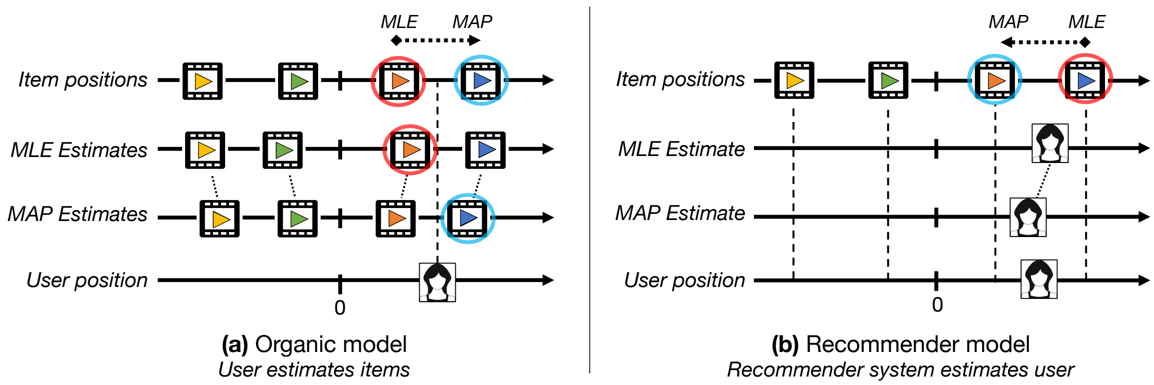

Figure 1: Model schematics. (a) In the organic model, the user samples and estimates item positions, and chooses the estimate closest to her own position. When shrinkage is applied, the user's estimates of the items shift inward, and the user's choice switches from the inner to outer item. (b) In the recommender model, the system samples and estimates the user's position, and chooses the item closest to its estimate. When shrinkage is applied, the system's estimate of the user shifts inward, and the system's choice switches from the outer to inner item.

distance between the user and their matched item as a user-level loss. Let $\mathcal{Y}_k = \{y_{k_1}, y_{k_2}, \cdots, y_{k_m}\}$ represent the multiset of matched item positions over all users; note that $|\mathcal{Y}_k| = |\mathcal{X}| = m$, and that each item position y_j might appear 0 times, once, or multiple times in $\mathcal{Y}_k$. Then, the loss for user i is defined as

$$l_i(\mathcal{X}, \mathcal{Y}_k) = ||x_i - y_{k_i}||^2. \quad (7)$$

As we analyze $l_i(\mathcal{X}, \mathcal{Y}_k)$ in the following section, we will also derive $\mathbb{E}[y_{k_i}]$ and $\mathrm{Var}[y_{k_i}]$ along the way; that is, the expected position and variance of the matched item for user i. We choose these metrics to study because they form the building blocks of many downstream phenomena of interest. For example, if the expected match for a user is not the user's own position, the matching process might eventually shift user preferences or opinions, and if different users systematically experience different losses (as we will see happens under certain settings of our models), this implies inequities in the model as it provides matches of differential quality to users on the basis of their preferences.

The population-level analysis, meanwhile, is concerned with average user loss and the overall collection of matched items. Observe that the average loss over users simply becomes the mean squared error (MSE) of the matches:

$$\mathrm{MSE}(\mathcal{X}, \mathcal{Y}_k) = \frac{1}{m}\sum_{i=1}^{m} l_i(\mathcal{X}, \mathcal{Y}_k) = \frac{1}{m}\sum_{i=1}^{m} ||x_i - y_{k_i}||^2. \quad (8)$$

These matches are based on estimates, where shrinkage is known to reduce the MSE of an estimator. As such, there are opportunities in both models to reduce average user loss by having the user / system apply shrinkage during estimation. However, while shrinkage might reduce MSE, that is not its only effect. In Figure 1, we illustrate a basic intuition for a key result: due to the reversal in who is doing the estimating and what is being estimated, shrinkage acts as a "diversifying" force in the organic model, but it serves as a "homogenizing" force in the recommender model. We see that shrinkage in the organic model shifts the user's estimates of the items inward (i.e., towards 0), so items that are further out now have a better chance of being chosen. Meanwhile, shrinkage in the recommender model shifts the system's estimate of the *user* inward, so now items closer to the center are likelier to be chosen.

To quantify this intuition, we also study $\mathrm{Var}[\mathcal{Y}_k]$ as another population-level metric, the variance of the matched item positions $\mathcal{Y}_k$. This variance has numerous interpretations in the real world: for example, seeing what kinds of items are being selected may guide content creators as they decide what to generate next. Furthermore, there is evidence that users adjust their preferences to better align with content that they are matched with, whether because they were persuaded by the content (Diehl, Weeks, and Zúñiga 2015), or by the very fact that it was recommended to them (Summers, Smith, and Reczek 2016). Such dynamics suggest that users may become more heterogeneous if they are matched with a more diverse set of items, and more homogeneous if they are all matched to similar content. With repeated rounds of matching and opinion formation, a great level of heterogeneity in users could lead to polarization or radicalization; conversely, increasing homogeneity could result in filter bubbles and a lack of diverse perspectives for those on the platform.

Model Analysis

To develop a theoretical understanding of how these two models behave, we begin with simplified instances, where we assume that the user and item positions come from Gaussian distributions in a one-dimensional space. For now, we will assume that user and item positions are drawn independently from $N(0, \sigma^2_{\text{user}})$ and $N(0, \sigma^2_{\text{item}})$, respectively (since $d = 1$, we replace covariance matrices Σ with scalar variances, σ^2). We also focus here on the asymptotic setting where the number of items n approaches ∞ (later we examine simulations with different finite values of n). Interestingly, we will see that even this simplified setting is sufficient to induce the phenomena we seek to understand, and that the results remain qualitatively similar when we explore non-Gaussian multidimensional user and item positions learned from data.

In this section, we organize our results into a series of theorems. In Theorems 1.1–2.2, we analyze how the models differ in terms of individual metrics, by deriving 3 quantities for each model: (1) $\mathbb{E}[y_{k_i}]$, the expected item match; (2) $\mathrm{Var}[y_{k_i}]$, the variance in the match; (3) $\mathbb{E}[l_i]$, the expected loss, for a single user i at any position x_i. In Theorems 3–4, we study how the models differ at the population level, showing that even though shrinkage can reduce MSE for both processes, a key difference is that shrinkage increases the variance of matched items under the organic model but decreases this variance under the recommender model.

Theorem 1.1. In the organic MLE model as $n \to \infty$,

1. $\mathbb{E}[y_{k_i}^{(\text{org,MLE})}] \to \frac{\sigma_{\text{item}}^2}{\sigma_{\text{item}}^2+\sigma_i^2} x_i$
2. $\mathrm{Var}[y_{k_i}^{(\text{org,MLE})}] \to \left(\frac{1}{\sigma_{\text{item}}^2} + \frac{1}{\sigma_i^2}\right)^{-1}$
3. $\mathbb{E}[l_i^{(\text{org,MLE})}] \to \left(\frac{\sigma_{\text{item}}^2}{\sigma_{\text{item}}^2+\sigma_i^2} - 1\right)^2 x_i^2 + \left(\frac{1}{\sigma_{\text{item}}^2} + \frac{1}{\sigma_i^2}\right)^{-1}$.

Proof. In the organic MLE model, the user makes an estimate of each item position, $\hat{y}_j^{(i,\text{MLE})} = z_j^{(i)}$, where $z_j^{(i)}$ is a sample drawn from $N(y_j, \sigma_i^2)$. In the limit as $n \to \infty$, there will be a sample drawn arbitrarily close to any point in $\mathbb{R}$, including x_i. Given that the user's chosen item $k_i^{(\text{org,MLE})}$ produced a sample arbitrarily close[1] to x_i, and that the chosen item's position $y_{k_i}^{(\text{org,MLE})}$ was drawn from $N(0, \sigma_{\text{item}}^2)$, marginalizing over the conjugate Gaussian prior furnishes the limiting expectation and variance expressions above.

Recall that in estimation, mean squared error can be decomposed into the sum of the squared bias and the variance of the estimator. If we view the position of the matched item, $y_{k_i}^{(\text{org,MLE})}$, as an estimate for user i's true position, x_i, then their expected squared distance is $(\mathbb{E}[y_{k_i}^{(\text{org,MLE})}] - x_i)^2 + \mathrm{Var}[y_{k_i}^{(\text{org,MLE})}]$, thus yielding $\mathbb{E}[l_i^{(\text{org,MLE})}]$. ☐

Theorem 1.2. In the organic MAP model as $n \to \infty$,

1. $\mathbb{E}[y_{k_i}^{(\text{org,MAP})}] \to x_i$
2. $\mathrm{Var}[y_{k_i}^{(\text{org,MAP})}] \to \left(\frac{1}{\sigma_{\text{item}}^2} + \frac{1}{\sigma_i^2}\right)^{-1}$
3. $\mathbb{E}[l_i^{(\text{org,MAP})}] \to \left(\frac{1}{\sigma_{\text{item}}^2} + \frac{1}{\sigma_i^2}\right)^{-1}$.

Proof. When the user uses MAP, her estimate of item j's position becomes $\hat{y}_j^{(i,\text{MAP})} = \frac{\sigma_{\text{item}}^2}{\sigma_{\text{item}}^2+\sigma_i^2} z_j^{(i)}$. In the limit, there will be a sample $z_j^{(i)}$ such that $\frac{\sigma_{\text{item}}^2}{\sigma_{\text{item}}^2+\sigma_i^2} z_j^{(i)}$ is arbitrarily close to x_i. Given that the matched item $k_i^{(\text{org,MAP})}$ produced a sample that limits to $\frac{\sigma_{\text{item}}^2+\sigma_i^2}{\sigma_{\text{item}}^2} x_i$, we can again derive the posterior distribution of this item's position $y_{k_i}^{(\text{org,MAP})}$, knowing that it was also drawn from $N(0, \sigma_{\text{item}}^2)$. The correction terms cancel out, yielding the limiting expectation x_i. The variance does not change compared to the MLE version of the model. Since $\mathbb{E}[y_{k_i}^{(\text{org,MAP})}]$ is now an unbiased estimate for x_i, the user's limiting expected loss is only the variance. ☐

Theorem 2.1. In the recommender MLE model as $n \to \infty$,

1. $\mathbb{E}[y_{k_i}^{(\text{rec,MLE})}] \to x_i$
2. $\mathrm{Var}[y_{k_i}^{(\text{rec,MLE})}] \to \sigma_r^2$
3. $\mathbb{E}[l_i^{(\text{rec,MLE})}] \to \sigma_r^2$.

Proof. In the recommender MLE model, the system makes an estimate of the user's true position, $\hat{x}_i^{(r,\text{MLE})} = z_i^{(r)}$, where $z_i^{(r)}$ is a sample drawn from $N(x_i, \sigma_r^2)$. Then, the system finds the item whose position is closest to $\hat{x}_i^{(r,\text{MLE})}$. Echoing the previous proofs, in the limit there will be an item arbitrarily close to $\hat{x}_i^{(r,\text{MLE})}$. Thus, the distribution of chosen item position, $y_{k_i}^{(\text{rec,MLE})}$, limits to the distribution of $\hat{x}_i^{(r,\text{MLE})}$, which has an expected value x_i and variance σ_r^2. Since $y_{k_i}^{(\text{rec,MLE})}$ is an unbiased estimate for x_i, the user's limiting expected loss is again just the variance. ☐

Theorem 2.2. In the recommender MAP model as $n \to \infty$,

1. $\mathbb{E}[y_{k_i}^{(\text{rec,MAP})}] \to \frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2} x_i$
2. $\mathrm{Var}[y_{k_i}^{(\text{rec,MAP})}] \to \left(\frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2}\right)^2 \sigma_r^2$
3. $\mathbb{E}[l_i^{(\text{rec,MAP})}] \to \left(\frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2} - 1\right)^2 x_i^2 + \left(\frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2}\right)^2 \sigma_r^2$.

Proof. In the MAP setting, the recommender model's estimate of the user's position becomes $\hat{x}_i^{(r,\text{MAP})} = \frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2} z_i^{(r)}$. In the limit the chosen item's position, $y_{k_i}^{(\text{rec,MAP})}$, will be arbitrarily close to $\hat{x}_i^{(r,\text{MAP})}$. Thus the distribution of $y_{k_i}^{(\text{rec,MAP})}$ limits to the distribution of $\hat{x}_i^{(r,\text{MAP})}$, which is simply a sample drawn from $N(x_i, \sigma_r^2)$ then scaled by $\frac{\sigma_{\text{user}}^2}{\sigma_{\text{user}}^2+\sigma_r^2}$, furnishing the limiting expectation and variance above. The limiting expected squared loss is obtained from the squared bias $(\mathbb{E}[y_{k_i}^{(\text{rec,MAP})}] - x_i)^2$ plus the limiting variance. ☐

Theorems 1.1–2.2 imply that if we treat the matched item position y_{k_i} as an estimate of the user position x_i, then the organic MLE model serves biased matches to users, while the recommender MLE model does not. In particular, the organic MLE model is biased in that, relative to x_i, the expected match for user i is contracted towards the center of

[1]Formally, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\min_{z\in\{z_1^{(i)},\ldots,z_n^{(i)}\}} |x_i - z| < \epsilon) = 1,$$

where the probability is over the process generating each item position y_j from $N(0, \sigma_{\text{item}}^2)$ and each sample $z_j^{(i)}$ from $N(y_j, \sigma_i^2)$.

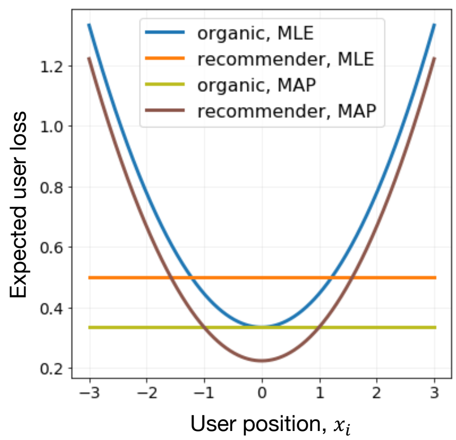

Figure 2: Expected user loss as a function of user position (Theorems 1.1–2.2); $\sigma^2_{\text{item}} = 1$ and $\sigma^2_i = \sigma^2_r = 0.5$.

the item distribution (Theorem 1.1). Furthermore, since the bias term grows with x_i^2, users closer to the center will be less impacted by the bias than users who are farther out. This is a fundamental difference between the MLE versions of the processes: the organic model favors users who are closer to the center of the item distribution, while the recommender model is agnostic to user position. We visualize these relationships in Figure 2.

Interestingly, the trend reverses when we switch from MLE to MAP: the organic model becomes unbiased, and the recommender model becomes favorable to central users. Some of this effect is particular to our form of shrinkage here (namely, that the user perfectly erases the bias in the organic model by taking the MAP based on the true prior distribution of items), but in general, shrinkage will always increase bias for the recommender model, and, if the item distribution is heaviest around its mean, shrinkage will also reduce bias for the organic model.

From these four theorems, we can also establish that (in the limit of many items) switching from MLE to MAP will *always* reduce average user loss (i.e., MSE) in the organic model and sometimes reduce MSE in the recommender model. First, Theorems 1.1–1.2 show that in the organic model, expected loss strictly decreases for every user (with non-zero x_i) from MLE to MAP; thus, the average loss must fall as well. From Theorem 2.2, we can imagine the conditions under which MAP will achieve a lower MSE in the recommender model: if the user variance σ^2_{user} is small, then the bias terms will be smaller because all of the x_i^2's will be closer to 0, and the relative reduction that MAP achieves on the variance, scaling by a factor of $(\frac{\sigma^2_{\text{user}}}{\sigma^2_{\text{user}}+\sigma^2_r})^2$, will be larger. More generally, we can conclude that there are certainly reasons to switch from MLE to MAP for both models, which makes the consequences of the switch all the more interesting: the individual-level consequences which we have analyzed, in terms of biased matches and favored users, and population-level consequences, which we analyze next in Theorems 3–4.

Theorem 3. In the organic model as $n \to \infty$,

1. $\text{Var}[\mathcal{Y}_k^{(\text{org,MLE})}] \to \frac{m-1}{m}\left[(\frac{\sigma^2_{\text{item}}}{\sigma^2_{\text{item}}+\sigma^2_i})^2\sigma^2_{\text{user}} + (\frac{1}{\sigma^2_{\text{item}}} + \frac{1}{\sigma^2_i})^{-1}\right]$
2. $\text{Var}[\mathcal{Y}_k^{(\text{org,MAP})}] \to \frac{m-1}{m}\left[\sigma^2_{\text{user}} + (\frac{1}{\sigma^2_{\text{item}}} + \frac{1}{\sigma^2_i})^{-1}\right]$.

Proof. In the proofs of Theorems 1.1–2.2, we studied a single individual, deriving results for a user at a specific position x_i. We now seek to analyze the variance of $\mathcal{Y}_k = \{y_{k_i}\}_{i=1}^m$, integrating over user positions in $\mathcal{X} = \{x_i\}_{i=1}^m$.

The process of generating y_{k_i}, the position of user i's matched item, can be seen as first drawing a random user position x_i from $N(0, \sigma^2_{\text{user}})$, and then considering the randomness from matching. Unconditional on x_i, then, y_{k_i} is the sum of two random variables. Using the results from Theorem 1.1 for the organic MLE model, the first variable limits to $\frac{\sigma^2_{\text{item}}}{\sigma^2_{\text{item}}+\sigma^2_i}x_i$ (from $\mathbb{E}[y_{k_i}^{(\text{org,MLE})}]$), with variance $(\frac{\sigma^2_{\text{item}}}{\sigma^2_{\text{item}}+\sigma^2_i})^2\sigma^2_{\text{user}}$. The second variable, capturing randomness in matching, has limiting variance $\text{Var}[y_{k_i}^{(\text{org,MLE})}] = (\frac{1}{\sigma^2_{\text{item}}} + \frac{1}{\sigma^2_i})^{-1}$. The limiting expected variance of $\mathcal{Y}_k^{(\text{org,MLE})}$ is the sum of these two variances multiplied by $\frac{m-1}{m}$, since there are m elements in $\mathcal{Y}_k^{(\text{org,MLE})}$.

In the organic MAP case, the process of generating terms in $\mathcal{Y}_k^{(\text{org,MAP})}$ can also be seen as first drawing a random user position x_i from $N(0, \sigma^2_{\text{user}})$, and then considering the randomness from matching. We do not scale x_i in this case, since $\mathbb{E}[y_{k_i}^{(\text{org,MAP})}]$ limits to x_i (Theorem 1.2). Each term can again be viewed as the sum of two random variables, with limiting variances σ^2_{user} and $(\frac{1}{\sigma^2_{\text{item}}} + \frac{1}{\sigma^2_i})^{-1}$ (from $\text{Var}[y_{k_i}^{(\text{org,MAP})}]$), respectively. So the limiting expected variance of $\mathcal{Y}_k^{(\text{org,MAP})}$ is their sum scaled by $\frac{m-1}{m}$. ☐

Theorem 4. In the recommender model as $n \to \infty$,

1. $\text{Var}[\mathcal{Y}_k^{(\text{rec,MLE})}] \to \frac{m-1}{m}(\sigma^2_{\text{user}} + \sigma^2_r)$
2. $\text{Var}[\mathcal{Y}_k^{(\text{rec,MAP})}] \to \frac{m-1}{m}\left[\frac{\sigma^2_{\text{user}}}{\sigma^2_{\text{user}}+\sigma^2_r}\sigma^2_{\text{user}}\right]$.

Proof. We take the same approach as in Theorem 3. In the recommender MLE model, the process of generating terms in $\mathcal{Y}_k^{(\text{rec,MLE})}$ can be seen as drawing a random user position x_i from $N(0, \sigma^2_{\text{user}})$, then adding a noise term for matching that has limiting variance $\text{Var}[y_{k_i}^{(\text{rec,MLE})}] = \sigma^2_r$ (Theorem 2.1). Again, we do not scale x_i, since $\mathbb{E}[y_{k_i}^{(\text{rec,MLE})}]$ limits to x_i (Theorem 2.1). The sum of these limiting variances is $\sigma^2_{\text{user}} + \sigma^2_r$, and again we scale by $\frac{m-1}{m}$.

In the recommender MAP model, each term in $\mathcal{Y}_k^{(\text{rec,MAP})}$ can be seen as the sum of two random variables, capturing randomness in x_i and in matching. The first variable limits to $\frac{\sigma^2_{\text{user}}}{\sigma^2_{\text{user}}+\sigma^2_r}x_i$ (from $\mathbb{E}[y_{k_i}^{(\text{rec,MAP})}]$, Theorem 2.2), with variance $(\frac{\sigma^2_{\text{user}}}{\sigma^2_{\text{user}}+\sigma^2_r})^2\sigma^2_{\text{user}}$. The second variable has limiting variance $\text{Var}[y_{k_i}^{(\text{rec,MAP})}] = \left(\frac{\sigma^2_{\text{user}}}{\sigma^2_{\text{user}}+\sigma^2_r}\right)^2\sigma^2_r$. Adding these vari-

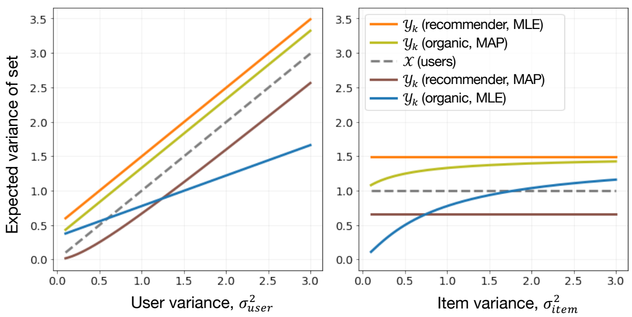

Figure 3: Variance of matched items Var$[\mathcal{Y}_k]$ (Theorems 3–4); $\sigma_i^2 = \sigma_r^2 = 0.5$ and $m = 300$. On the left, we fix $\sigma^2_{\text{item}} = 1$ and vary σ^2_{user}; on the right, we fix $\sigma^2_{\text{user}} = 1$ and vary σ^2_{item}.

ances together, simplifying, and multiplying by $\frac{m-1}{m}$ yields the limiting expression for Var$[\mathcal{Y}_k^{(\text{rec,MAP})}]$. $\square$

From Theorems 3 and 4, we can see that (in the limit of many items) the variance of matched items will *always* increase in the organic model when switching from MLE to MAP, but it will *always* decrease in the recommender model when making the same switch. We demonstrate these two effects in Figure 3: across different values for the user variance σ^2_{user} and item variance σ^2_{item}, the matched item variance for the recommender MLE model remains well above that of the recommender MAP, and the matched item variance for the organic MLE model is always below that of the organic MAP. This quantifies our earlier intuition from Figure 1 that shrinkage acts as a diversifying force in the organic model (increasing variance), but a homogenizing force in the recommender model (decreasing variance), and we have proven that this is always true in this setting.

Simulations

Our theorems demonstrated that even the simplest versions of the organic and recommender processes result in vastly different matching behavior. However, one may wonder how sensitive these results are to our assumptions (studying the asymptotic case, with single-dimensional, normally distributed user and item positions). In this section, we remove each of these assumptions, and show via simulation experiments that the key qualitative results we established hold for much more realistic settings as well.

First, we test how the individual-level and population-level metrics converge to their limits as we increase the number of available items. Then, we augment our process models to integrate multidimensional user and item embeddings learned from real ratings data.[2]

Exploring Finite Numbers of Items

First, in the single-dimensional Gaussian setting we analyze the behavior of our proposed metrics as the number of items n grows, complementing our analytical results where we focused on the asymptotic setting as $n \to \infty$. We consider both the MLE and MAP versions of the organic and recommender models and $n \in \{4, 6, \cdots, 200\}$. For the individual metrics, we simulate 5000 stochastic trials for each model and value of n, re-sampling per trial the positions of items and noisy samples, but fixing the location of the single hypothetical user to $x_i = 0.75$. For the population metric Var$[\mathcal{Y}_k]$, we simulate 500 stochastic trials, since there is far less variance in this metric, and re-sample positions of users, items, and noisy samples per trial.

As shown in Figure 4, these simulations are consistent with and extend our earlier analyses. In the organic model, when the user employs the MLE, a biased match is formed and the average position of the matched item, y_{k_i}, converges inward of $x_i = 0.75$; as expected, converging at 0.5, since $\sigma^2_{\text{item}} = 1$ and $\sigma_i^2 = 0.5$, and $\mathbb{E}[y_{k_i}^{(\text{org,MLE})}] = \frac{\sigma^2_{\text{item}}}{\sigma^2_{\text{item}} + \sigma_i^2} x_i$ (Theorem 1.1). Meanwhile, the recommender MLE model forms an unbiased match and the average y_{k_i} converges to $x_i = 0.75$. When MAP is used, the roles are reversed, as the organic model becomes unbiased and the recommender model becomes biased. Furthermore, every model converges to a different average Var$[\mathcal{Y}_k]$: the recommender MLE model has the highest variance, followed by organic MAP, then organic MLE, then finally recommender MAP (following the order we would expect from Figure 3, at $\sigma^2_{\text{user}} = \sigma^2_{\text{item}} = 1$), and this ordering is observed when there are as few as 10 items.

Empirical Analysis with MovieLens Data

In this section, we further extend our simulations by incorporating multidimensional user and item positions fitted on real ratings data. We use the MovieLens 1M dataset[3], which contains 1,000,209 ratings from approximately 6,000 MovieLens users on 3,900 movies (Harper and Konstan 2015). First, we filter the ratings matrix to only keep users that have rated at least 50 movies, then filter to keep movies with at least 50 ratings. After filtering, we are left with a rating matrix $R \in \mathbb{R}^{m' \times n}$, where we have $m' = 4,297$ users and $n = 2,514$ movies remaining.

Learning user and item distributions from data. We apply a collaborative filtering algorithm to R in order to infer latent embeddings $\mathcal{U} = [u_i]_{i=1}^{m'}$, $u_i \in \mathbb{R}^d$, for each user i, and $\mathcal{V} = [v_j]_{j=1}^{n}$, $v_j \in \mathbb{R}^d$, for each movie j. Collaborative filtering uses the history of interactions between all users and items to infer latent representations, where users and items are encoded into low-dimensional spaces such that if a user has given an item a high rating, their representations should be "similar", and if a user has given an item a low rating, their representations should be further apart. Similarity can be measured in different ways (e.g., inner products, Euclidean distance); since our models take a distance-based perspective to utility, we implement a collaborative filtering procedure (Khoshneshin and Street 2010) that embeds users and items into a unified Euclidean space where items that are closer to users are more attractive to them.

[2]All code to run our models and simulations is available at https://github.com/serinachang5/org_rec_simulations.

[3]https://grouplens.org/datasets/movielens/1m.

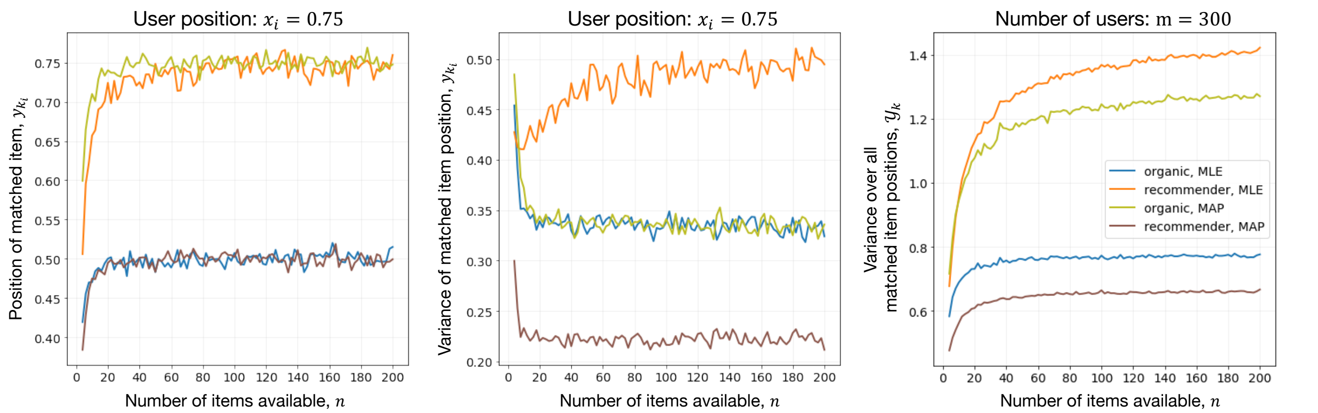

Figure 4: Simulation results over different values of n. As before, we set $\sigma^2_{\text{user}} = \sigma^2_{\text{item}} = 1$, $\sigma^2_i = \sigma^2_r = 0.5$, and $m = 300$. For the individual metrics, we run 5000 trials for each model and n, and evaluate the mean (left) and variance (middle) of the matched item position y_{k_i} for a user positioned at $x_i = 0.75$. For the population metric Var$[\mathcal{Y}_k]$, we run 500 trials for each setting and evaluate the mean variance over the matched item positions $\mathcal{Y}_k$ (right).

In this framework, given a user embedding u_i and movie embedding v_j, their predicted rating $\hat{r}_{ij}$ is

$$\hat{r}_{ij} = \mu + b_i - ||u_i - v_j||^2, \tag{9}$$

where μ is the global average rating in R, b_i is the user bias term capturing users that tend to rate higher or lower, and $||u_i - v_j||^2$ is the squared distance between the two embeddings. This framework naturally integrates our definition of user loss—the squared distance between a user and her matched item—since a user's predicted rating is exactly the negative squared distance, translated by $\mu + b_i$ (which remains constant per user). Furthermore, this formulation fits with the choice function in our models: just as we assume that the recommender system will choose the item that minimizes distance to its estimate of the user, that very same item here would be the one that maximizes predicted rating.

To learn these embeddings, we define the following objective function, where we aim to minimize the regularized loss over all observed ratings in R:

$$\min_{\mathcal{U},\mathcal{V},\mathcal{B}} \sum_{i,j} w_{ij}[(r_{ij} - \hat{r}_{ij})^2] + \lambda(||u_i - v_j||^2 + b_i^2). \tag{10}$$

Here $\mathcal{B}$ represents the set of all user biases b_i, and $w_{ij} \in \{0, 1\}$ indicates whether the rating r_{ij} is observed in R. We then use gradient descent to update the latent parameters $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{B}$ with respect to the regularized loss (see details in Appendix). After learning $\mathcal{U}$, i.e., the embeddings of the users in the MovieLens dataset, we sample from this empirical distribution to generate m "test" users. This creates a new matrix of user positions, $\mathcal{X} \in \mathbb{R}^{m\times d}$, which we use in our simulations to represent unseen users. Our simulations do not rely on unseen movies, however, so we can directly set $\mathcal{Y}$, the item positions in our simulations, to $\mathcal{V}$, the movie positions directly inferred from the MovieLens dataset.

Simulations with MovieLens embeddings. Using this approach, we learn user and movie embeddings of dimension $d = 5$ from the MovieLens dataset. In Figure 5, we display the learned user and movie distributions. We see that the

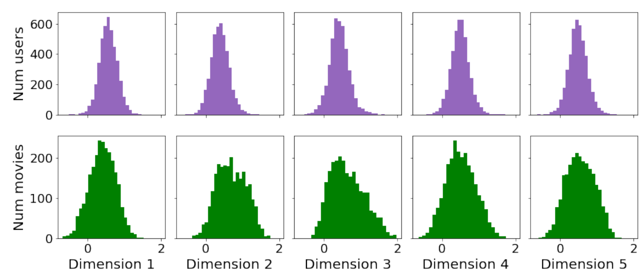

Figure 5: Distribution of learned user (top) and movie (bottom) embeddings from the MovieLens 1M dataset.

user distribution is less dispersed than the movie distribution in each dimension, and the movie distribution is sometimes asymmetric (clearly non-Gaussian). For both embeddings, the covariance between dimensions is low.

We then simulate the organic and recommender models with $m = 300$ users and $n = 2,514$ movies (the number of movies in the filtered MovieLens dataset). For the organic model, we fix each user's noise covariance Σ_i to $0.5 \cdot \Sigma_{\text{item}}$, where Σ_{item} was the empirical covariance fitted on $\mathcal{Y}$, the movie embeddings. Similarly, for the recommender model, we fixed the system's noise covariance Σ_r to $0.5 \cdot \Sigma_{\text{user}}$, where Σ_{user} was the empirical covariance fitted on $\mathcal{X}$, the user embeddings. We scaled the user / movie covariance in this way, as opposed to using spherical noise, so that the size of the noise per dimension would scale with the variance of the estimated population in that dimension, which we believed was more realistic. When shrinkage is applied within each model, we no longer have the user or system construct MAP estimates with Gaussian priors: the prior distribution is neither Gaussian nor known. Instead, we implemented shrinkage parameterized by a scalar $\alpha \in [0, 1]$, which interpolates between the MLE estimate and the mean over all MLE estimates. That is, user i's shrunken estimate $\hat{y}_j^{(i,S)}$ for

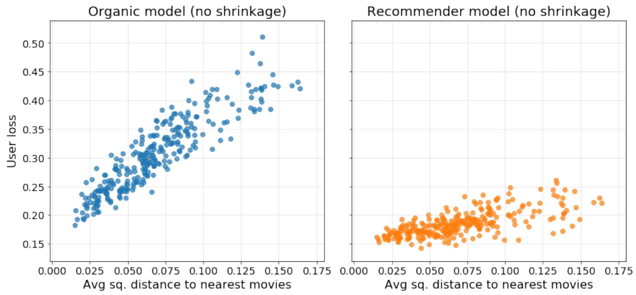

Figure 6: Individual-level results from MovieLens experiments. We plot each user's average squared distance to their 10 nearest movies (as a measure of centrality) vs. the user's loss, averaged over 500 trials. In order to see data points more clearly, we truncated outliers beyond the 95th percentile of the x-axis (showing 285 out of 300 users).

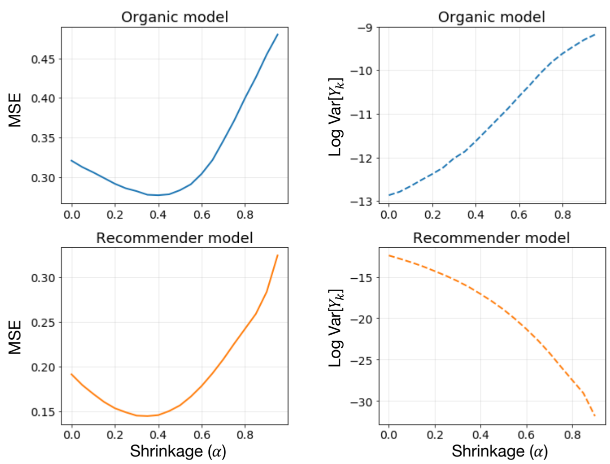

Figure 7: Population-level results from MovieLens experiments. We assess the effect of varying the shrinkage parameter α on the MSE and variance of matched item positions $\text{Var}[\mathcal{Y}_k]$, running 500 trials for each model and value of α.

movie j becomes

$$\hat{y}_j^{(i,S)} = (1-\alpha)\hat{y}_j^{(i,\text{MLE})} + \alpha\Big(\frac{1}{n}\sum_{k=1}^{n}\hat{y}_k^{(i,\text{MLE})}\Big), \qquad (11)$$

and the recommender system's shrunken estimate of user i is the equivalent interpolation between $\hat{x}_i^{(r,\text{MLE})}$ and the mean MLE estimate over all users. This family of estimators generalizes James–Stein shrinkage and Empirical Bayes estimators, which correspond to specific recipes for choosing α (or α_i, different for each user).

Despite the lopsided movie distribution and expanded number of dimensions, we find that the key qualitative results from our theoretical analyses hold. In Figure 6, we see that the organic MLE model continues to strongly favor more central users, compared to the recommender MLE model, which is much more even-handed across user positions. In our earlier setting with items drawn from $N(0, \sigma^2_{\text{item}})$, the centrality of a user could be summarized by the absolute value of x_i, but here we need a more empirical measure; for example, we use the user's average squared distance to their 10 nearest movies (where higher average distance corresponds to lower centrality).

Secondly, we simulate different amounts of shrinkage ($\alpha \in \{0.05, 0.1, \cdots, 0.95\}$) and evaluate the impacts on our population-level metrics, the average user loss (MSE) and the variance of matched item positions $\text{Var}[\mathcal{Y}_k]$. In multiple dimensions, we compute the generalized variance as the product of the eigenvalues of the covariance matrix of $\mathcal{Y}_k$; we log-transform this quantity to make it more interpretable. As we increase the amount of shrinkage, at first this improves MSE for both models, but eventually MSE begins to increase; the optimal level of shrinkage seems to fall around $\alpha \approx 0.4$ for both models. However, even though the MSE curves look similar, the $\text{Var}[\mathcal{Y}_k]$ curves completely diverge: the more we increase α, the more $\text{Var}[\mathcal{Y}_k]$ grows in the organic model, and the more it shrinks in the recommender model (Figure 7). This matches our earlier findings that shrinkage has a "diversifying" effect on the organic model but a "homogenizing" effect on the recommenders.

Conclusions and Future Work

We have introduced two contrasting models of item-matching processes: one describing a generic personalized recommender system, and the other describing a setting where users search for items organically without the mediation of any system. In both cases, we have a decision-making agent trying to match users to items with limited information, but the difference lies in who/what is doing the matching, and what information they have access to vs. what is being estimated. Comparing the two, we have seen that this simple switch in perspective results in dramatic differences at both the individual and population levels. For example, in the MLE versions of the process models, the recommender model serves unbiased item matches for the users, while the organic model does not, and the organic model favors central users, while the recommender model does not. Applying shrinkage has notably diverging effects: while it can reduce MSE in both models, shrinkage leads the recommender model to choose increasingly similar sets of items, while leading the organic model to diversify its selections.

Most interestingly, these results are general. Our models of organic search and personalized recommendation encompass many specific instantiations of each process (e.g., flexible to different user behavior or recommendation algorithms). Furthermore, we are able to prove our findings under asymptotic settings for a large class of these models; within this class, the differences between the models are guaranteed under any choice of parameters. Finally, through simulations, we show that our findings are robust to changes in user and item distribution, single and multiple dimensions, and asymptotic and finite settings.

We have worked to study a core counterfactual question about recommender systems: if there were no recommender system, how would matching be different? Our analyses provide evidence that the use of recommender systems fundamentally alters how humans interact with content, as we reveal pervasive differences between recommended and

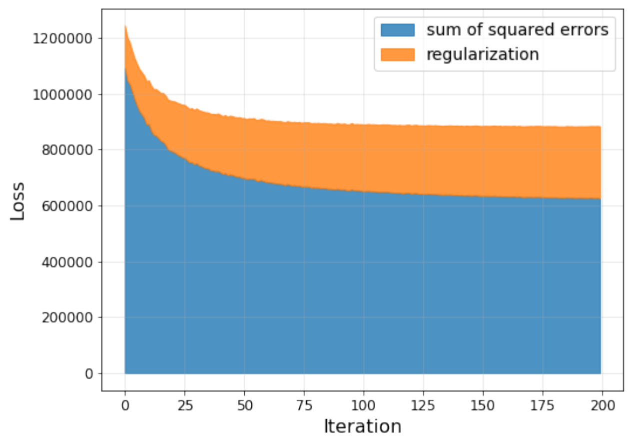

Figure A1: Regularized loss over iterations, when learning user and movie embeddings from MovieLens ratings data.

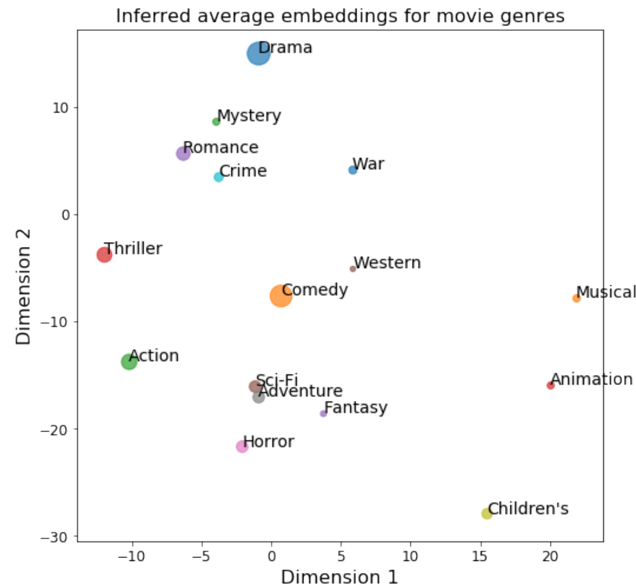

Figure A2: Visualization of the learned movie embeddings. We performed t-SNE to convert the 5-dimensional embeddings into 2 dimensions. Here, we visualize the average embedding for every movie genre with at least 50 movies in the filtered MovieLens dataset (all movies with at least 50 ratings in MovieLens 1M). The size of the datapoint corresponds to the number of movies in the genre; for example, the largest genre Drama has 938 movies.

organic processes of item-matching. Our model-based approach also has its limitations: for example, seeking maximum comparability between the organic search and recommender model, we made simplifying assumptions about how users might behave in both search and recommended settings. Future work could incorporate more realistic user models, such as modeling the process of choosing among a list of recommendations.

Simulations, even with real-world data, also have their limitations, so we hope that future work can tie the theory of our models to real-world user studies and experiments; for example, through randomized controlled trials where participants are assigned to unmediated organic search versus recommended item-matching processes. We also hope that future studies can build upon the formal framework established in this paper, using our models as building blocks to analyze diverse potential long-term effects of recommender systems, such as polarization, filter bubbles, user retention rates, or fairness across users or products. By modeling repeated choices and incorporating learning dynamics where users learn from the items they consume (Diehl, Weeks, and Zúñiga 2015), it should be possible to analyze how the phenomena we document here, pertaining to single choices, compound over time and affect the evolution of users' opinions and behaviors.

Appendix

Details from MovieLens Simulations

Recall that we defined the following objective function, which minimizes the regularized loss over all observed ratings in the rating matrix R (Equation 10):

$$\min_{\mathcal{U},\mathcal{V},\mathcal{B}} \sum_{i,j} w_{ij}[(r_{ij} - \hat{r}_{ij})^2] + \lambda(||u_i - v_j||^2 + b_i^2).$$

We can minimize this loss using gradient descent, with updates in each step defined as

$$b_i \leftarrow b_i + \gamma(e_{ij} - \lambda b_i), \tag{12}$$
$$u_i \leftarrow u_i - \gamma(u_i - v_j)(e_{ij} + \lambda), \tag{13}$$
$$v_j \leftarrow v_j + \gamma(u_i - v_j)(e_{ij} + \lambda), \tag{14}$$

where b_i is the bias of user i, u_i is user i's embedding, v_j is movie j's embedding, γ is the step size, and $e_{ij} = r_{ij} - \hat{r}_{ij}$ indicates the error on the current rating.

In our experiments, we ran 200 iterations of gradient descent, where each iteration looped through every observed rating in R and updated the latent parameters accordingly. In Figure A1, we demonstrate that the loss approximately converged within this number of iterations. We used $\lambda = 0.01$, which allowed regularization to account for around 30% of the overall loss, while the remaining percentage came from the sum of squared errors between the predicted and observed ratings. We decreased the step size γ over iterations, where, for a given iteration t, we set $\gamma^{(t)}$ to

$$\gamma^{(t)} = \max\left(0.01, \gamma_0 \cdot \frac{1}{1+\beta t}\right), \tag{15}$$

where $\gamma_0 = 0.2$ indicates the initial step size and $\beta = 0.05$ controlled the rate at which the step size decayed.

We also demonstrate that even in 5 dimensions, our learned embeddings capture reasonable representations of movie genres. Each movie in the MovieLens 1M dataset is tagged with one or more genres, such as Drama or Comedy. In Figure A2, we plot the average movie embedding (after performing t-SNE) of every genre that mapped to at least 50 movies in our filtered dataset. The resulting lay-out of genres follows what we would expect: for example, the genres Children's, Animation, and Musical are close to each other; Sci-Fi, Adventure, and Fantasy are tightly grouped; and, on the other side of the plot, Drama is close to Mystery, Romance, Crime and War. These checks confirm that our inferred embeddings capture the real data, and thus bolster the validity of our experimental results, where we used these embeddings to represent user and item positions in our models.

Broader Impact

Our work is primarily motivated by broader impacts: our goal is to develop a principled framework through which we can meaningfully assess the impacts of recommender systems on society. Recommender systems are ubiquitous on the web and social media platforms, and have the potential for large-scale negative consequences such as pulling individuals into filter bubbles, increasing population-level polarization, or exacerbating social inequalities. However, in order to distill the role of recommender systems in contributing to these social phenomena, we need to compare user outcomes under recommender systems to a credible assessment of user outcomes without recommender systems. Thus, we develop two contrasting models that capture each of these worlds and systematically compare them, so that we can analyze the consequences of recommender systems relative to a counterfactual world without then.

As a model-based approach, we establish key general insights about recommender systems, but we do not make any claims about specific platforms nor do we offer instruction on how real-world recommender systems should be designed. We use real movie ratings data from MovieLens to demonstrate that our theoretical results translate to real-world settings, not to demonstrate superior performance on the movie recommendation task. The test users in our simulation experiments are synthetic, generated from a distribution learned from the real users; the real users are also anonymized in the data. Practitioners should be aware of the limitations and theoretical nature of our work. They should not directly apply our findings to their domains, but we hope that they appreciate our main takeaways: that recommender systems fundamentally alter how humans interact with content, and that seemingly minor algorithmic decisions (such as regularization) can have major effects on user outcomes.

Acknowledgements

This work was supported in part by funding from the Stanford Program on Democracy and the Internet and ARO MURI award #W911NF-20-1-0252. S. C. was supported by an NSF Graduate Research Fellowship. We thank Jan Overgoor, Amel Awadelkarim, Arjun Seshadri, Kaitlyn Zhou, and anonymous reviewers for helpful comments.

References

Abdollahpouri, H.; Burke, R.; and Mobasher, B. 2017. Controlling popularity bias in learning-to-rank recommendation. In *RecSys*, 224–232.

An, J.; Quercia, D.; Cha, M.; Gummadi, K.; and Crowcroft, J. 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, 3(12).

Anderson, A.; Maystre, L.; Anderson, I.; Mehrotra, R.; and Lalmas, M. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *WWW*, 2155–2165.

Bakshy, E.; Messing, S.; and Adamic, L. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6238): 1130–1132.

Bozdag, E. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15: 209–227.

Brynjolfsson, E.; Hu, Y.; and Simester, D. 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8): 1373–1386.

Chaney, A.; Stewart, B.; and Engelhardt, B. 2018. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *RecSys*, 224–232.

Chen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; and He, X. 2020. Bias and Debias in Recommender System: A Survey and Future Directions. *arXiv:2204.00241*.

Dandekar, P.; Goel, A.; and Lee, D. T. 2013. Biased assimilation, homophily, and the dynamics of polarization. *PNAS*, 110(15): 5791–5796.

Das, A. S.; Datar, M.; Garg, A.; and Rajarm, S. 2007. Google news personalization: scalable online collaborative filtering. In *WWW*, 271–280.

Diehl, T.; Weeks, B.; and Zúñiga, H. 2015. Political persuasion on social media: Tracing direct and indirect effects of news use and social interaction. *New Media & Society*, 18(9): 1875–1895.

Efron, B.; and Morris, C. 1976. Multivariate empirical Bayes and estimation of covariance matrices. *The Annals of Statistics*, 4(1): 22–32.

Fabbri, F.; Bonchi, F.; Boratto, L.; and Castillo, C. 2020. The Effect of Homophily on Disparate Visibility of Minorities in People Recommender Systems. In *ICWSM*, 165–175.

Flaxman, S.; Goel, S.; and Rao, J. M. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80: 298–320.

Fleder, D.; and Hosanagar, K. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5): 697–712.

Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2018. Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *WWW*, 913–922.

Gupta, P.; Goel, A.; Lin, J.; Sharma, A.; Wang, D.; and Zadeh, R. 2013. WTF: The who to follow service at twitter. In *WWW*, 505–514.

Harper, F. M.; and Konstan, J. A. 2015. The MovieLens Datasets: History and Context. In *ACM Transactions on Interactive Intelligent Systems*.

Hosanagar, K.; Fleder, D.; Lee, D.; and Buja, A. 2013. Will the Global Village Fracture into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science*, 60(4): 805–823.

James, W.; and Stein, C. 1961. Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1961*, volume 1, 361–380.

Khoshneshin, M.; and Street, W. N. 2010. Collaborative filtering via euclidean embedding. In *RecSys*, 87–94.

Kloumann, I. M.; and Kleinberg, J. M. 2014. Community membership identification from small seed sets. In *KDD*, 1366–1375.

Lambrecht, A.; and Tucker, C. 2019. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, 65(7).

Nguyen, T. T.; Hui, P.; Harper, F. M.; Terveen, L.; and Konstan, J. A. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *WWW*, 677–686.

Pariser, E. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin.

Rastegarpanah, B.; Gummadi, K. P.; and Crovella, M. 2019. Fighting Fire with Fire: Using Antidote Data to Improve Polarization and Fairness of Recommender Systems. In *WSDM*, 231–239.

Singer, N. 2011. The Trouble With the Echo Chamber Online. *New York Times*.

Stoica, A.; and Chaintreau, A. 2019. Hegemony in Social Media and the effect of recommendations. In *1st FATES Workshop, WWW*, 575–580.

Summers, C.; Smith, R.; and Reczek, R. 2016. An Audience of One: Behaviorally Targeted Ads as Implied Social Labels. *J. Consumer Research*, 43(1): 156–178.

Szlávik, Z.; Kowalczyk, W.; and Schut, M. 2011. Diversity Measurement of Recommender Systems under Different User Choice Models. In *ICWSM*, 369–376.

Yao, S.; and Huang, B. 2017. Beyond Parity: Fairness Objectives for Collaborative Filtering. In *NeurIPS*, 2925–2934.