DECISION SCIENCES
A JOURNAL OF THE DECISION SCIENCES INSTITUTE

Decision Sciences
Volume 52 Number 3
June 2021

Taste Transitivity for Collaborative Filtering: A Stochastic Network Dynamics Approach

Minghong Xu†
The Johns Hopkins Carey Business School, Baltimore, MD, e-mail: xu.minghong@jhu.edu

Siddhartha Bhattacharyya
Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL

ABSTRACT

We develop a stochastic actor-based network model for online movie reviews and analyze social network dynamics to study the user-movie taste networks constructed from real movie review datasets. Examining such taste networks provides useful insights into factors underlying the performance of collaborative filtering-based recommenders. Our results show that similar taste transitivity effect does exist that support collaborative filtering methods to recommend movies based on user taste similarity. We also investigate the role of movie popularity, genre, user gender, age, and geographic location in the taste network evolution. The findings provide insights to improve current movie recommendation systems. The two- mode network analyses approach taken in this article will also be broadly useful in obtaining better understanding of factors that drive user appreciations for varied products. [Submitted: February 28, 2017. Revised: June 10, 2018. Accepted: October 23, 2018.]

Subject Areas: Collaborative Filtering, Decision Support System, Movie Recommender System, Network Dynamics, Statistical Inference Making, and Stochastic Simulation.

INTRODUCTION

User-generated reviews today are widely available on the Internet. Online stores make user reviews as an important component of their consumer decision support systems. Consumers submit reviews on e-merchant web sites or professional review forums to provide their opinions on different products. Such "opinion" is typically rated on a scale (rating) or can be categorized as like, dislike, or neutral. Ratings, as an additional piece of information beyond product descriptions, are searched by consumers in online business, to compare and find the "best fit" product or service from available options (Wang, Guo, Zhang, Wei, & Chen,

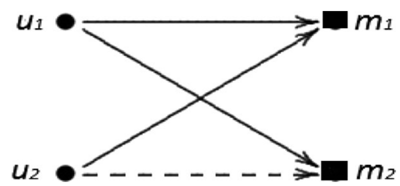
†Corresponding author.

2016). Understanding the process of consumer information search is very important in the study of decision making and in the design of decision support systems (Cook, 1993). Sellers and researchers use ratings as a rich data source to analyze customer behaviors and shopping preferences. E-business sites also utilize such analyses to support consumer decision-making. Recommendation engines are one of the most widely implemented and studied applications of this type. These systems aim to address the product selection problem, by using data on purchases, product ratings, and user profiles to predict which products are best suited to a particular user (Fleder & Hosanagar, 2009). Surveys of recommendation system research, methods, developments for decision support, and industrial applications are provided in Adomavicius and Tuzhilin (2005), Bobadilla, Ortega, Hernando, and Gutierrez (2013), Liang (2008), Linden, Smith, and York (2003), Sharma and Gera (2013), and Lu, Wu, Mao, Wang, and Zhang (2015). There are two basic approaches adopted by recommendation engines, content-based filtering, and collaborative filtering. Content-based systems utilize a set of discrete characteristics of an item in order to recommend additional items with similar properties. Collaborative filters, a more common approach, collect human judgments or ratings and match together people who share the same information needs or the same tastes (Herlocker et al., 1999). The collaborative filtering-based recommendations build models from user taste information (items purchased or liked, which are referred to as *tastes* throughout this article to reflect preferences or positive opinions) and recommend products bought or liked by other consumers who are similar in tastes. These two approaches can be combined in hybrid systems.

Collaborative filters are well studied by researchers, and implemented by various industries as successful prediction models. These type of recommendation engines rely on similar tastes between users. They are built on the assumption that people with similar tastes in the past will share similar tastes in the future, that is, similar tastes are transitive from past purchase history to future purchase decisions. Figure 1 depicts a single transitivity phenomenon considering two users, who like one product in common, and will like another product in common in the future. The success of collaborative filtering-based recommendations relies on the existence of such transitivity effect, which captures the persistence of similar tastes over time. However, this essential premise underlying the effectiveness of collaborative filtering-based recommendations has not been examined; there has been no study to establish whether taste similarity is actually transitive, and to quantify the extent of such transitivity effect in any dataset on user reviews or ratings of products.

This article uses network analysis methods to statistically examine whether similar taste transitivity is significant in movie review datasets, which are the basis for movie recommendations. Our study shows how stochastic actor network models (Snijders, 1996, 2001) can be used to empirically examine the factors, which can drive successful recommendations. Consumers' appreciation of products may also be driven by the "rich get richer" effect, whereby popular items tend to attract more attention. Both these effects can contribute to the success of product recommendations, with one resulting from users' similar tastes which is referred to as *similar taste transitivity effect* (or briefly, *taste transitivity effect*), and the other arising from items' popularity, which is denoted as *popularity effect*. This article examines the key premise underlying collaborative filtering-based

Figure 1: Similar taste transitivity depiction.
Nodes u_1 and u_2 denote customers. Nodes m_1 and m_2 denote products. Edges from u_i to m_i denote user u_i likes product m_i. The dotted line shows action in the future.



recommendations–taste transitivity. The focus is on validating the significance of taste transitivity and in estimating the size of this effect, together with other factors, which may drive successful recommendations. User activity is another potentially important factor, which may influence the effectiveness of recommendations–are active users more likely to appreciate additional movies in the future. We thus consider user activity and its interaction with movie popularity as additional factors for effective recommendations. Our analysis also considers user and movie characteristics like user's age and gender, and movie genre, and examines whether effective recommendation may vary by such characteristics. This research brings a much needed deeper understanding of which factors are significant, and their relative strengths, in driving successful recommendations. Such understanding can provide valuable insights to enhance the design of collaborative filtering-based recommendation engines. The novel method for network-based analyses of review data presented in this article provides a general approach for determining which factors drive successful recommendations, and will be broadly useful in the design of collaborative filtering-based recommendation engines across different types of products and e-business platforms.

Figure 1 also illustrates an alternative way of depicting the data used in recommendations. The user-product taste information can be viewed as a network composed of two *modes* (sets of different types) of nodes: nodes representing users and nodes representing products. Edges in this taste network only occur between nodes of different modes, specifically from a user node to a product node, denoting that the user likes (gives a high rating on) the product. Because the edges arise from users' actions to indicate their appreciations of products, this network is also a network describing consumer choices and behavior, or an *actor-based social network* where social actors play crucial roles in changing their ties and their behavior (Snijders, 1996, 2001). Transitivity in two-mode networks (as depicted in Figure 1) is also known as the two-mode version of closure. It was introduced by Robins and Alexander (2004) and has been discussed by social network studies in various types of two-mode networks.

In this study, the online rating data is viewed from this social network standpoint, specifically as a two-mode network. Because recommendation systems are interested in predicting which products a user might like, this study focuses on "likes" or positive ratings on products by users. Consideration of neutral and negative ratings is a topic for future work. Prior research on consumer decision support systems has mostly focused on developing algorithms to generate recommendations or product ranking lists (Wang et al., 2016), and these algorithms are driven by similarities between users, or common features/characteristics between products. We consider dimensions of both users and products, and infer a two-mode social network representation of recommendation engines. We investigate the dynamic aspects of the network representing consumer behavior, that is, how the network evolves and how edges are formed or dissolved. Edge changes in the network indicate changes in user tastes, and modeling this evolution process helps determine factors that are significant in driving user appreciation of products. The goal is to analyze the evolution of the taste network over time, make statistical inferences on the underlying process and factors that drive network formation, and thereby to address the question of what drives the success of collaborative filtering-based recommendations. We argue that by taking a social network view of data and understanding the extent and significance of different factors that drive this network formation, we can improve the design and performance of current recommendation engines. Because the factors influencing successful recommendations may vary by product, such analyses can inform the design of recommenders for different types of products across different e-business sites.

We utilize the *stochastic actor-oriented network models* of Snijders (1996, 2001), which are a class of statistical inference making models on longitudinal network data. These models provide statistical estimates of factors that drive changes in social networks using Markov Chain-based network simulation. Given that movie recommendation engines are one of the most widely discussed types of recommender applications, we focus on movie recommendations in this article. We use movie review datasets from Amazon, an electronic commerce site, and MovieLens, an electronic community site, to develop the taste networks. We quantitatively measure similar taste transitivity effect, movie popularity effect, and two other effects, to disentangle different factors that drive the user-movie taste network evolution, and to answer the following research questions:

- Will people who like one movie in common like more other movies in common?
- Are there other factors besides the taste transitivity effect, which drive the network formation?

This is the first study to systematically examine the premise underlying collaborative filtering-based recommendation engines–that user tastes are transitive. Our results provide evidence that the taste transitivity effect does exist across different movie review datasets. Besides taste transitivity, our findings reveal that other factors also play important roles in driving successful recommendations; such as movie popularity and genre, user activity level, and age. For example, we find that movie popularity is strongly significant and should be considered in making recommendations. User activity level can also play a distinct role, and in interaction with movie popularity. We find that while popularity is a useful factor to consider, its impact on active users may be limited; and that active users are more likely to be willing to explore and appreciate niche movies. Our analyses also reveal that movie genre and user age have a significant effect and can be usefully considered in designing collaborative filtering recommenders. The importance of taste transitivity, movie popularity, and user activity is seen to vary by movie genre. A surprising finding on user age is that taste transitivity is not significant for younger aged (under 24) users. Such results bring novel insights to improve current collaborative filtering recommendation systems for consumer decision support.

This study is the first to apply stochastic network generation modeling to online review data. Results show that the two-mode network analyses approach helps analyze review data to uncover different factors, which drive successful recommendations. This network-based analysis approach will be broadly useful for informing the design of recommenders for various types of products to improve consumer decision support; future work can consider similar analyses on other product types.

The rest of the article is organized as follows. Section "Literature Review" reviews the related literature. Section "Stochastic Actor-Oriented Networking Modeling Approach" presents the network analyses methodology used for analyses. Section "Models and Theories" discusses models and related theories, and Section "Data Description" describes the datasets used in this study. Results and insights from the study are discussed in Section "Results and Discussions," and Section "Conclusion and Future Works" concludes the article and offers directions for future work.

LITERATURE REVIEW

The work in this article relates to three streams of research. The first is on the mechanism of online recommendation generation engines. Two recent survey articles reviewed related research, different techniques used, and current trends, especially in the context of extensive online data available today (Bhosale & Pande, 2015; Lu et al. 2015). Some other related studies include Li and Hitt (2008) which developed a model that examines how idiosyncratic preferences of early buyers can affect long-term consumer purchase behavior as well as the social welfare created by review systems; Fleder and Hosanagar (2009) which analyzed the effect of recommendation engines on the diversity of sales; and Lee, Hosanagar, and Tan (2015) which studied the social influence of prior ratings and, in particular, investigated any differential impact of prior ratings by strangers ("crowd") versus friends. This study is different in that we investigate the premise and rationale behind the success of recommendations by developing a stochastic two-mode network model and examining different factors, which are significant in users' appreciation of movies. These are the factors, which drive the success of collaborative filtering-based recommendations, and our study helps to provide insights into designing more effective consumer decision support systems.

The second stream of literature is on social networks and theories. Sociology researchers have identified social theories based on generative mechanisms that are directly relevant to the emergence and coevolution of social networks (Monge & Contractor, 2003). The theories that are applied to understand consumer behavior in movie appreciation include *Homophily* (Coleman, 1957; Laumann, 1966; McPherson & Smith-Lovin, 1987; Marsden, 1988; Carley, 1991; Ibarra, 1993, 1995), *Preferential Attachment* and *Power Law* (Barabasi, 2002), *User Choices* and *Theories of Self-interests* (Coleman, 1986), and *Theories of Contagion* (Burt, 1980, 1987; Contractor & Eisenberg, 1990). In the social network area, two-mode networks are often called *affiliation* networks (Wasserman & Faust, 1994) because they are used to represent the association (duality) between social actors and activities, groups, or events with which the actors may be affiliated. Examples of social entities linked by two-mode networks include individuals and groups (Breiger, 1974), team members and knowledge (Carley, 1991), individuals and issues (Breiger, 2000), organizations and items in their political agendas (Mische & Pattison, 2000), directors and corporate boards (Robins & Alexander, 2004), and organizational members and issues (Doreian, Batagelj, & Ferligoj, 2004). Our study contributes to this literature by inferring a two-mode social network representation of recommender systems.

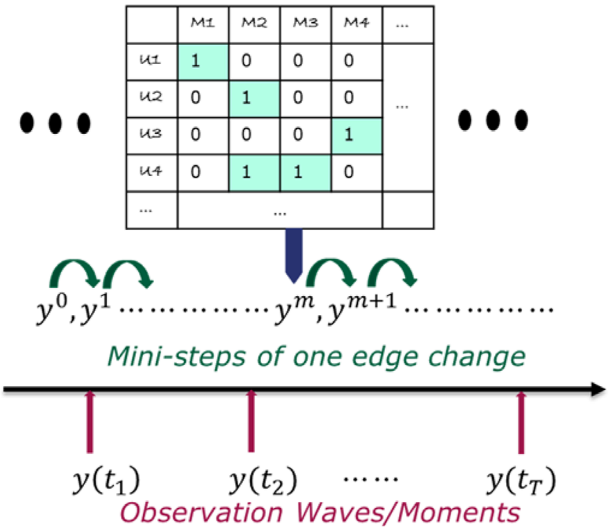
The last stream of literature related to our work is on statistical models developed to estimate factors that drive network formation. Among these models, exponential random graph models, also known as p* models (Wasserman & Pattison, 1996; Anderson, Wasserman, & Crouch, 1999; Pattison & Wasserman, 1999; Robins, Pattison, & Wasserman, 1999), and dynamic actor-oriented models (Snijders, 2001, 2005; Snijders, Steglich, & Schweinberger, 2007; Snijders, Koskinen, & Schweinberger, 2010) are the most widely used. The p* models are applicable for a single observation on a stochastic directed graph, while the stochastic actor-oriented network models are used on longitudinal data with more than two observations. Previous studies on two-mode network dynamics include Conaldi, Lomi, and Tonellato (2012) which studied the affiliation network of problem solving in an open source software project; and Snijders, Lomi, and Torlo (2013) which proposed a stochastic actor-oriented model for the co-evolution of two-mode and one-mode networks. This article extends the thread of two-mode network dynamics analysis to study the taste association between users and products, which is an important type of relation in e-business.

STOCHASTIC ACTOR-ORIENTED NETWORKING MODELING APPROACH

Stochastic actor-oriented network models (Snijders, 2001) analyze and draw statistical inference on the underlying process and factors, which drive social network formation. In these models the evolution of the network is the result of actions by individuals in the network. Network status is then observed at discrete time points (*observation waves*). The network evolution is not bound in a special way to the observations, but is regarded as a continuous process with observations as snapshots of the network over time. The dynamics of the network can thus be modeled by a continuous time Markov chain model defined over the space of all possible networks with a given node set.

Specifically, we consider longitudinal data on the user-movie taste network of N users and M movies. The two-mode network is defined through a binary array $Y(t)$ of size $N \times M$. An edge in the network records the existence of an appreciation/like from an element of the user set to an element of the movie set. While Y is stochastic, there are repeated observations on $Y(t)$, that is, $y(t_1),\ y(t_2)\ldots,\ y(t_T)$,

Figure 2: Illustration on mini-steps and observation waves.



at time points $t_1,\ t_2 \cdots,\ t_T$ (with $T \geq 2$). Typically a user reviews a movie once after watching it, and cases where an individual likes a movie and dislikes the same movie later are rare. An edge in the user-movie taste network is thus considered to persist after it is established; in other words, the changes in the network are 'upward' only.

The dynamics of the taste network is analyzed by modeling the edge changes. There are two aspects of these changes: (1) the moment when the changes occur, modeled by the rate function, and (2) the particular changes that users make, or the edges chosen to form, which is modeled by the objective function. Although we have time stamps on the reviews, the time of submission may be different from the time that the user watches and likes the movie. In addition, movie recommendations are interested in predicting which movies a user might like, rather than when a user might like the movies. Thus, we adopt a constant rate of change, which allows the changes to happen at any time between observation waves. The network evolution is then regarded as the result of many unobserved small changes that occur between the consecutive observation waves. Such small changes are called *mini-steps*, in which the network differs only by one edge. Figure 2 shows the process by which the network changes through multiple mini-steps from one observation to another.

The choices of changes are modeled in the objective functions. The function f for user i is a weighted sum of different factors which drive the changes, denoted by

$$f_i\ (y) = \sum_{k=1}^{L} \beta_k\, s_{ik}\,(y)\,,$$

indicating how likely it is for the user i to change her network in a particular way. It thus models network changes as the stochastic result of factors included in the function. The factors or terms s_{ik} represent network features that are hypothesized to play a role in the network evolution. The s_{ik} may be: (i) purely structural features, which are also called "endogenous" network effects, such as reciprocity, transitivity, etc., (ii) dependent on node-specific covariates, which are also called "exogenous" characteristics, or (iii) defined by interactions between structural network effects and node-specific covariates. The statistics $s_{ik}(y)$ can be calculated explicitly based on the observed network structure of y. The purpose of modeling is then to estimate factors, which realize the observed network y.

The weights β_k are statistical parameters indicating the strength of the corresponding factor s_{ik}, controlling for all other terms in the model. Greater values of β_k are expected to lead to higher values of the statistics $s_{ik}(Y(t_{j+1}))$ for all users. The principle of estimation is to determine the parameters β_k such that, summed over all users in all observation waves, the expected values of these statistics are equal to the observed values (Snijders, 2005):

$$s_k^{obs} = \sum_{j=1}^{T-1} \sum_{i=1}^{N} s_{ik}\left(y^{obs}\left(t_{j+1}\right)\right) \quad (k = 1, \cdots, L).$$

This approach is called *method of moments* by fitting "observed" to "expected" values. Because the expected values cannot be calculated explicitly, they are estimated from simulations. The method of moments is implemented using a stochastic approximation algorithm, which is a descendant of the Robbins-Monro (1951) algorithm. It is a recursive procedure to obtain approximate solutions to the moment equation. The estimation procedure is an instance of MCMC (Markov Chain Monte Carlo simulation), and the detailed simulation and approximation procedure can be found in Snijders (2005). The procedure is implemented in the R package RSIENA,[i] which carries out the statistical method of moments estimation for the evolution of social networks on the basis of computer simulations. In this study, the simulation run is iterated in RSIENA until all the estimates are adequately converged.[ii] The estimate of β_k can be interpreted as the log odds ratio of the probability that an edge is chosen over another edge, where the two connecting objects have exactly the same network position and the same values on all variables included in the model, but the second object is one unit less than the first object in the value of statistics $s_{ik}(y)$. The significance of the parameter can be tested by referring the t-ratio (estimate divided by standard error) to a standard normal distribution.

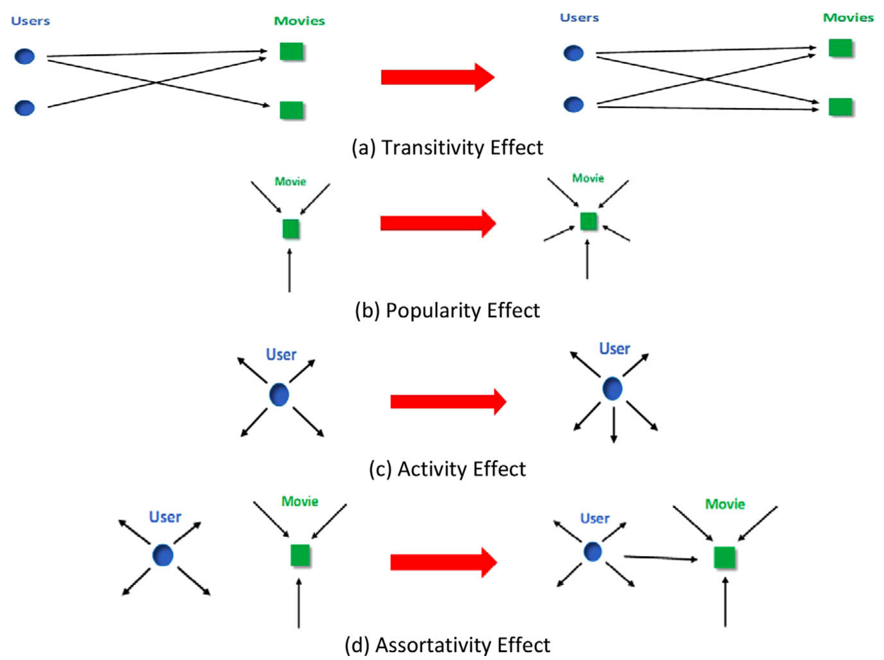
MODELS AND THEORIES

There are four endogenous network effects estimated in the study, each motivated by different social theories. The four effects are illustrated in Figure 3

The taste transitivity effect measures the extent to which individuals who like one common movie will like more movies in common. It reveals how strongly

[i] Simulation investigation for empirical network analysis.
[ii] Overall maximum convergence ratio is less than 0.25 as suggested in the manual for RSiena.

Figure 3: Illustration of four endogenous structural network effects.



similar tastes are transitive from the past to the future, which is the assumption for successful collaborative filtering-based movie recommendations. The taste transitivity effect reflects the social mechanism of homophily, whereby similarities in age, gender, education, prestige, social class, and occupation have been identified to affect various social networks' formation. In the decision sciences literature, similarities have been demonstrated to impact how purchase decisions are made, and the establishment of preference relations. Al-Natour, Benbasat, and Cenfetelli (2008) showed that perceived process similarity exerts positive and significant effects on users' perceptions of the decision aids' usefulness and trustworthiness. Janakiraman and Niraj (2011) found a significant effect of geographic proximity on what to buy, how to buy, and where to buy in the personal computer market. In this study, we analyze whether similarity in movie tastes plays a role in the evolution of the taste network. The extent of the transitivity effect thus reflects the tendency of users with similar tastes in movies (who have liked common movies in the past) to like the same movies in the future.

In social networks, in-degree often reflects a node's "popularity" in the relation defined by the edges, like friendship or advisorship. The in-degree popularity effect measures the extent to which current in-degrees contribute to the probability of receiving more in-links. A positive popularity effect leads to the power law distribution for node degrees, which is widely reported in many empirical networks (Sundararajan, Provost, Oestreicher-Singer, & Aral, 2013). The process can be explained by preferential attachment where quantities distribute among individuals or objects according to how much they already have. The popularity effect also illustrates the social mechanism of contagion, whereby exposure increases the likelihood that network members will adopt others' beliefs, assumptions, and attitudes. Thus, in the user-movie taste network, it models how exposure to positive ratings increases the user's probability of liking the movie.

The other two endogenous terms that we analyze are out-degree activity effect and out-in degree assortativity effect. The activity effect measures the extent to which individuals who currently like many movies continue liking more movies. Similar to the optimal stopping rules for product information search (Branco, Sun, & Villas-Boas, 2012), the examining of activity effect in this study helps to understand consumers' trade-off between the costs of time and energy spent in searching and watching more movies, and the potential benefits obtained from the activities. The activity effect thus helps evaluate how such trade-offs can impact the success of recommendations. The assortativity effect measures the tendency of active users (with high out-degrees) to preferably affiliate with popular movies (with high in-degrees). This can be regarded with an interaction term between the popularity effect and the activity effect.

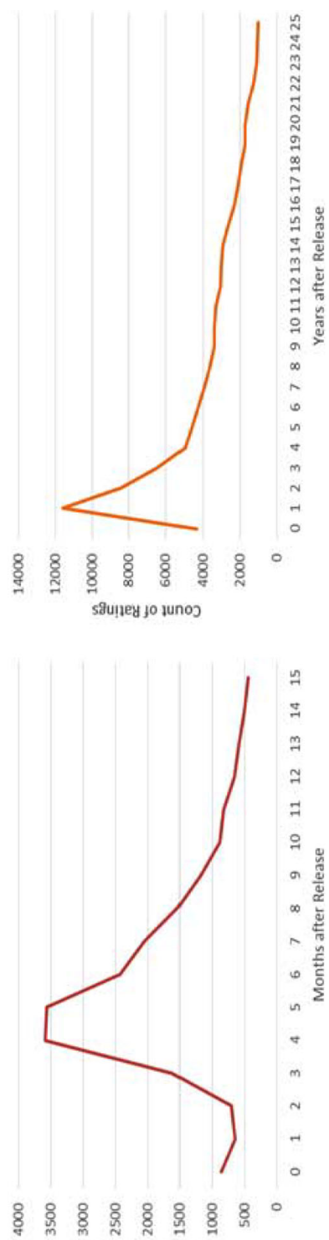
Besides the four structural effects, exogenous covariates or characteristics of users and movies are also used. These include movie genre and user demographic information such as gender, age range, and geographic region.

DATA DESCRIPTION

The proposed models are analyzed on three datasets from Amazon (McAuley et al. 2015a, 2015b) and MovieLens (https://grouplens.org/datasets/movielens/). Amazon, the largest e-commerce retailer in the United States, lists products from books, DVDs, software, electronics to apparel, furniture, food, and other products. It offers the function for consumers to comment on listed products, including various movie products. Most users of the Amazon movie review platform are Amazon shoppers who purchase movie products from Amazon and then submit reviews on Amazon. Thus, the Amazon users can be considered as amateur movie reviewers, or ordinary consumers who not only consume movie products but also other products on Amazon. MovieLens, which has a different business model than Amazon, is one of the most popular movie recommender systems and electronic community web sites. MovieLens is designed exclusively for recommending movies instead of selling movie products directly. MovieLens platform users may watch movies from other sources and then submit reviews on it. These users go to the MovieLens platform because they are interested in watching and reviewing movies, and sharing their opinions in the electronic community. Results from these two platforms are compared to draw insights for improving movie recommendation engines.

For each dataset we construct three observation waves. The time windows used to construct the observation waves are chosen considering the review peak time periods. A plot of the number of movie reviews on Amazon against the number of months because movie release reveals a peak of reviews at four to five months after release. A similar plot on the MovieLens reviews shows the peak at two years after movie release. The two plots are presented in Figure 4. As discussed, the data

Figure 4: Plot of number of reviews on Amazon and Movielens.



(a) Number of Movie Reviews on Amazon by Months after Movie Release

(b) Number of Movie Reviews on MovieLens by Years after Movie Release

Figure 5: Illustrations on moving time window.

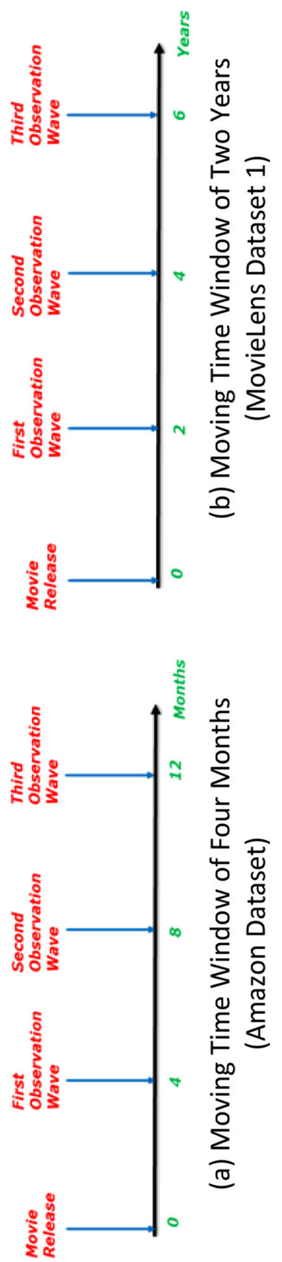


Table 1: Descriptive statistics for Amazon Dataset.

	First Observation		Second Observation		Third Observation
Density	0.029		0.047		0.051
Average degree	3.145		5.125		5.566
Number of ties	478		779		846
Distance		301		67	
Jaccard Index		0.614		0.921	

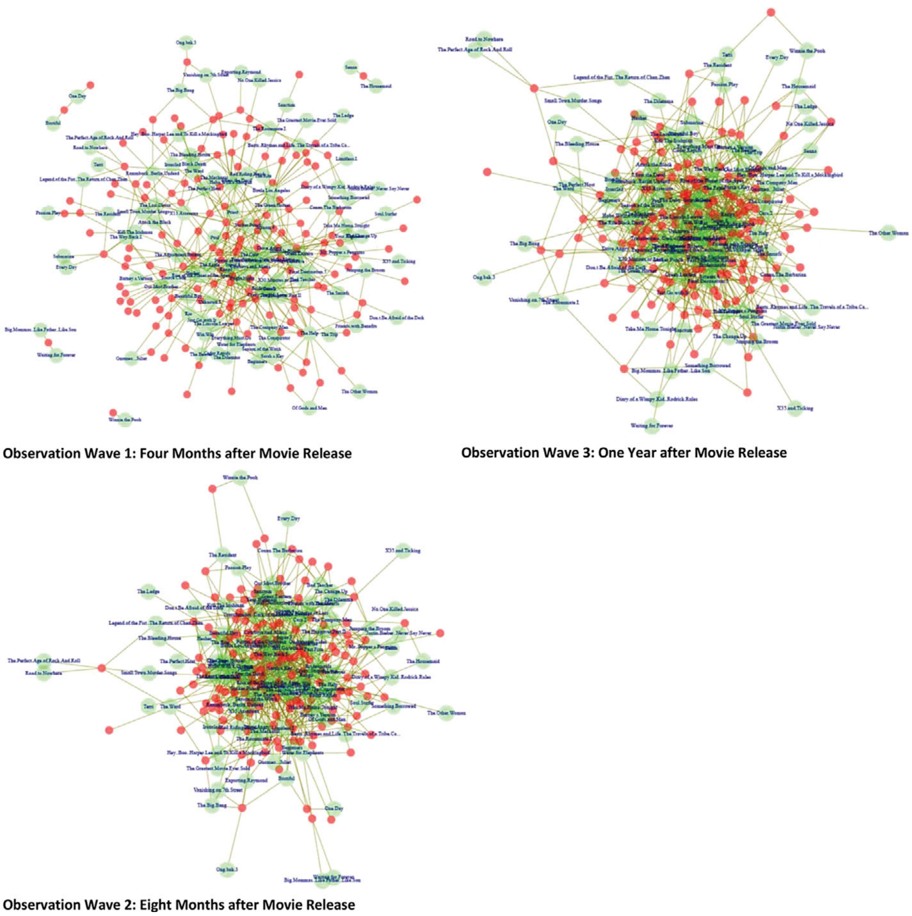
is filtered and only positive ratings (above four out of five on both Amazon and MovieLens) are considered.

The first dataset, which we refer to as Amazon Dataset, uses positive ratings from Amazon on movies released in an eight-month period from January to August, 2011. Based on the peak review period at four to five months after release, a moving time window of four months is adopted. An illustration of the moving time window on the Amazon Dataset is given in Figure 5(a). In three observation waves, the dataset tracks users' tastes towards movies in a one-year period. We consider this as the newly released movie dataset.

The SIENA models are actor-based models which explain user behavior and describe how the observed networks result from the actors' behavior. Because recommendation engines typically model behavior of users who watch multiple movies, this study focuses on the "returning" customers, and on testing the transitivity effect. For the Amazon Dataset, we remove outlying users who have either very low or very high activity levels (or the out-degrees, which represent how many movies a user likes across the three observation waves). The users who like fewer than three movies in three observation waves are removed because absence in one or more observation waves may indicate that the users are inactive on Amazon that their reviews on Amazon may not fully reveal their movie tastes. After inspecting the profiles of active users, we also remove those who like more than 22 movies in three observation waves. This excludes professional reviewers, whose behavior is distinct from most Amazon users who are amateur movie reviewers; such typical users are the main target of movie recommenders. This filtering of users also helps avoid spam users or bots (e.g., one user posted a total of 2,084 reviews from January 2011 to October 2012 and gave 5 stars in 1,928 of these reviews).

This dataset includes 110 movies and 152 users through the three observation waves, with a total of 262 nodes and 846 ties. The descriptive statistics of the networks in the three observation waves are provided in Table 1, which gives *density*, *average degree*, and *number of ties*. The *distance* measures the number of edge changes between the two observation waves. The *Jaccard index*, also known as the Jaccard similarity coefficient, is a statistic used for comparing the similarity of networks in two waves. Figure 6 provides a visualization of the Amazon Dataset network evolution over the three observation waves (using a Fruchterman-Reingold layout). Because it is an "upward only" network, we observe that the density, degree, and edge count all go up.

Figure 6: Amazon user-movie taste network evolution.
Observation Wave 1: Four months after movie release.
Observation Wave 2: Eight months after movie release.
Observation Wave 3: One year after movie release.



For these 110 movies, we also download genre information from IMDB and three big genres are identified: Action, Comedy and Drama. To construct genre indicators as movie covariates (exogenous variables in the model), the genre indicator is set to 1 if the movie has the named genre label, and 0 if it does not. It is not uncommon for a movie to have several genre labels, so certain movies have more than one genre indicator to be 1.

The second dataset is constructed using positive ratings and genre information on MovieLens. It includes movies released in fifteen years from 1996 to 2010 and is referred to as MovieLens Dataset 1. Based on observed peak of user reviews at two years after release (Figure 4b), a moving time window of two years is

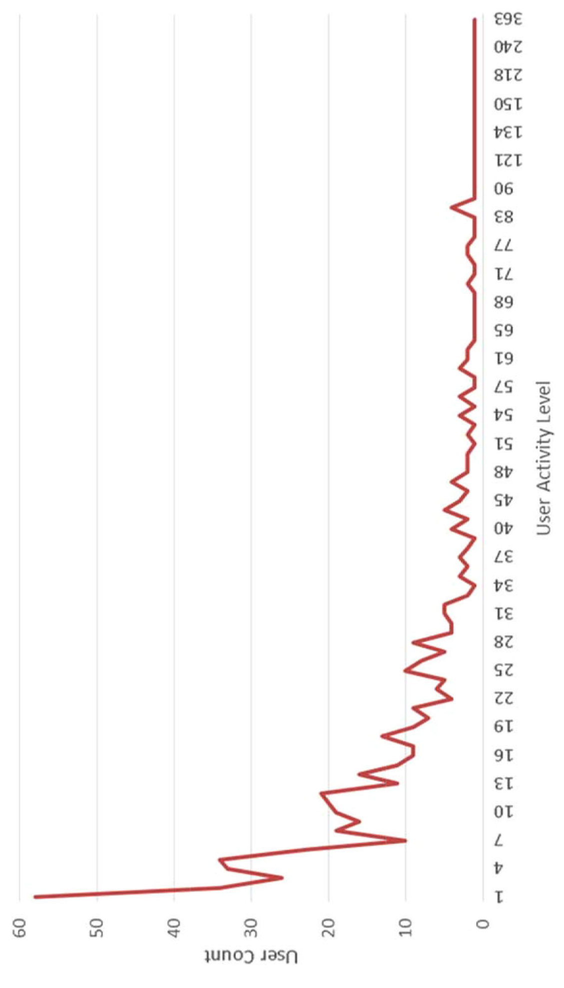


Figure 7: User activity level plot on MovieLens Dataset 1.

Table 2: Descriptive statistics for MovieLens Dataset 1.

	First Observation	Second Observation	Third Observation
Density	0.022	0.036	0.047
Average degree	16.877	27.556	35.864
Number of ties	1,367	2,232	2,905
Distance		865	673
Jaccard index		0.612	0.768

adopted as shown in Figure 5(b). This dataset tracks users' tastes towards movies in three observation waves over a period of six years after movie release. Compared to the Amazon Dataset, we consider this as the long-term released movie dataset.

To inspect the heterogeneity among actors in their out-degrees, Figure 7 plots the user activity levels. The x-values are the user activity levels, or out-degrees, and the y-values denote the numbers of users with that particular activity levels. It can be observed that the user counts drop sharply at low activity levels of 1 through around 30. The numbers of users tend to be stable at activity levels from roughly 30 to 65, followed by a flat tail, which goes totally flat after activity level of 90. For the study in this article, we focus on the medium-active users, who have activity levels from 30 to 70 in the 20-year period, which is roughly the middle one-third of users according to the plot in Figure 7. The low-active and high-active users are removed from the final dataset for the same reasons as for the Amazon Dataset. We also examine the sensitivity of results on data from different out-degree ranges, as discussed later.

The MovieLens Dataset 1 includes 759 movies and 65 users through the three observation waves, with a total of 824 nodes and 2,891 links. The descriptive statistics of the networks are provided in Table 2. This is a much larger network and we observe higher average degrees than those in Amazon Dataset. The genre information is available on MovieLens and we are able to identify five genres, given the larger number of movies in this dataset. These five genres are: Action, Comedy, Drama, Adventure, and Crime, and five corresponding genre indicators are constructed as movie covariates in the same way as for the Amazon Dataset.

The last dataset, which we refer to as MovieLens Dataset 2, includes user demographic information (gender, age and geographic region), and is used to examine the roles of user demographics in driving the taste network evolution. To maintain stable demographic information for analysis, one year's data (2002) is used to help avoid changes in ages or geographic locations. Given the one-year duration of this data, three observation waves are constructed using a fixed time window of four months. User activity levels are plotted in Figure 8. The numbers of users drop dramatically as activity levels vary from 1 to 15. At activity levels of 16 to 43, the user counts are stable at around 5 to 10. At higher activity levels, the numbers of users drop below 5. As with the earlier datasets, here again we focus on the medium-active users who are the typical target of recommendation engines. We select data on theses medium-active users who have activity levels

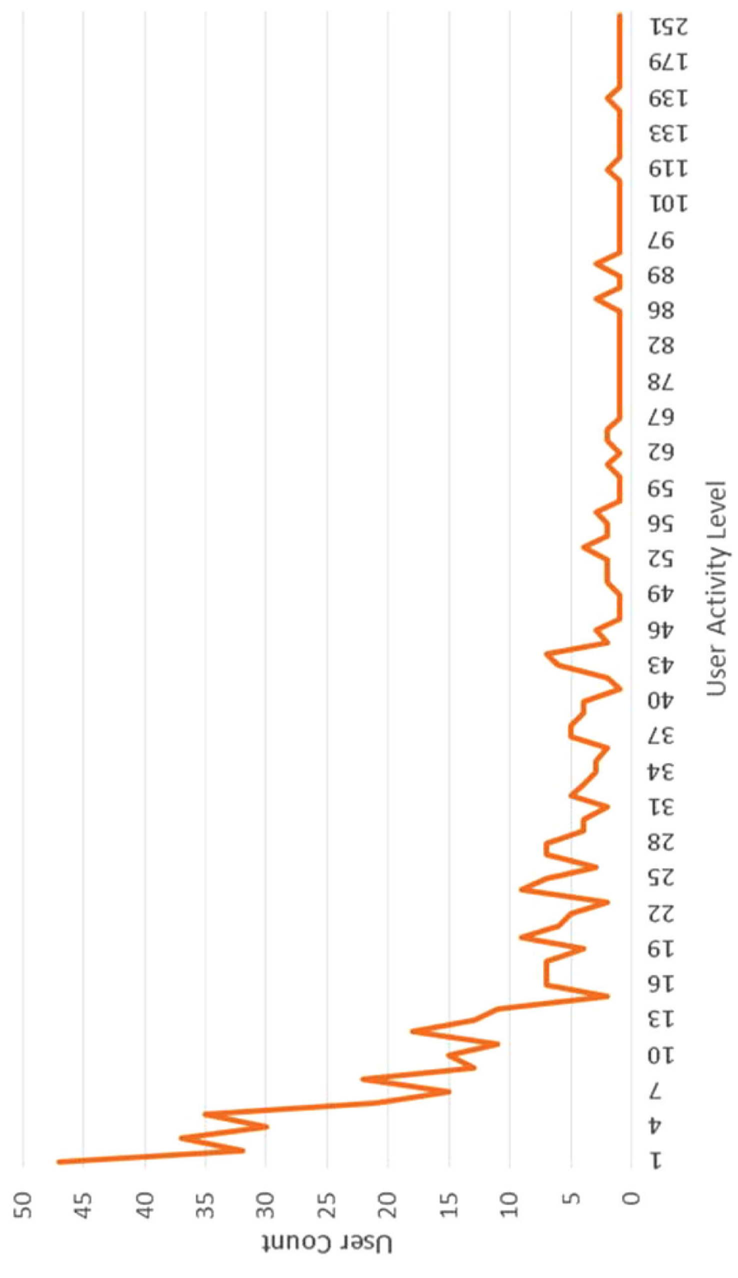


Figure 8: User activity level plot on MovieLens Dataset 2.

Table 3: Descriptive statistics for MovieLens Dataset 2.

	First Observation		Second Observation		Third Observation
Density	0.022		0.026		0.030
Average degree	16.736		20.069		23.069
Number of ties	1,205		1,445		1,661
Distance		240		216	
Jaccard index		0.834		0.870	

between 20 and 40. Compared with the activity plot for the MovieLens Dataset 1, the plot for MovieLens Dataset 2 has a longer tail, and we take this longer tail into consideration in the sensitivity analysis.

A set of 778 movies and 72 users are analyzed through the three observation waves, with a total of 850 nodes and 1661 links. Table 3 provides the descriptive statistics of the networks. The demographic information is processed for all users, and specifies gender (Male and Female), seven possible age ranges (under 18, age 18–24, age 25–34, age 35–44, age 45–49, age 50–55, and age 56+), and four possible regions mapped from zip-codes to Northeastern, Midwestern, Southern, and Western as per the US Census Bureau. To avoid multicollinearity, only one gender indicator (1 for Male and 0 for Female), six age range indicators (age 56+ is dropped), and three region indicators (Western is dropped) are used in the models.

RESULTS AND DISCUSSIONS

This section presents results of the stochastic actor network models to estimate the significances and sizes of different effects on the three datasets described above.

Estimates on Amazon Dataset

On the Amazon Dataset, four models are analyzed: a basic model including only the taste transitivity effect; a partial model, which adds the movie popularity effect; a full model having taste transitivity effect, movie popularity effect, user activity effect, and user-movie assortativity effect; and finally the genre model, which also includes three genre indicators. Results of the four models are compared in Table 4.

The basic model shows that the taste transitivity effect is significant, which indicates a strong tendency of two Amazon users who like one movie in common to like more movies in common. It thus demonstrates support for the key underlying premise for collaborative filtering-based recommendations; given strong taste transitivity points to success of recommendations based on collaborative filtering. After controlling for the movie popularity in the partial model, the level of taste transitivity effect is found to drop, but remains significant. The strongly significant popularity effect shows that the success of collaborative filtering systems also arises partially from the movie popularity. Popularity is thus a factor that can be usefully incorporated in making recommendations.

Table 4: Estimated stochastic actor-based models on Amazon Dataset.

Effects	Basic	Partial	Full	Genre
Taste transitivity	0.1247*** (0.0175)	0.0583** (0.0235)	0.1030*** (0.0236)	0.1032*** (0.0228)
Movie popularity		0.4174*** (0.0761)	0.9899** (0.3554)	0.9382** (0.3378)
User activity			−0.3144 (0.3674)	−0.4002 (0.3691)
User-movie assortativity			−0.3100 (0.1783)	−0.2797 (0.1692)
Action				−0.2148 (0.1337)
Comedy				−0.1950 (0.1237)
Drama				−0.0271 (0.1223)
Wald χ^2 statistics (df)	50.645*** (1)	142.226*** (2)	160.788*** (4)	170.907*** (7)

$** \ p < .05, \ *** \ p < .01.$

Activity and assortativity effects are included in the full model but neither of them is found to be significant at the 5% level (the assortativity effect comes close to the 5% level). The taste transitivity effect becomes stronger while the significance of popularity effect reduces; this indicates that activity and assortativity effects help to further disentangle the transitivity effect from the popularity effect, and to attribute more driving power to the taste transitivity effect. It is interesting to observe that estimates of both activity and assortativity effects are negative. A negative activity effect means that active users will tend to be less active in the future. The marginally significant negative assortativity effect implies that popular movies are less likely to attract active users, or that active users are less likely to provide positive ratings for popular movies. The implication for movie recommendation engine design is that while popularity is a useful factor to consider, its impact on active users may be limited. The fourth model shows that controlling for genre has no substantial influence on the four effects. None of the genre indicators is significant at the 5% level, meaning that users do not exhibit an inclination towards specific genres.

Estimates on MovieLens Dataset 1, Sensitivity, and Genre Analyses

A full model including taste transitivity effect, movie popularity effect, user activity effect and user-movie assortativity effect is estimated on MovieLens Dataset 1, and then five genre indicators are added. Results are presented in Table 5. All the four estimated effects are strongly significant in the models, both with or without genre information. The taste transitivity effect is strongly significant in driving the taste network evolution, which is similar to what is observed for Amazon Dataset. The popularity effect is more powerful (as indicated by the larger estimates) compared

Table 5: Estimated stochastic actor-based models on Movielens Dataset 1.

Effects	Without Genre	With Genre
Taste transitivity	0.0389***	0.0397***
	(0.0029)	(0.0028)
Movie popularity	1.3157***	1.2925***
	(0.2198)	(0.2410)
User activity	−0.5591***	−0.5612***
	(0.0583)	(0.0634)
User-movie	−0.1431***	−0.1407***
assortativity	(0.0431)	(0.0479)
Action		0.0281
		(0.0688)
Comedy		0.1002
		(0.0647)
Drama		0.0044
		(0.0631)
Adventure		0.2460***
		(0.0719)
Crime		0.2714***
		(0.0718)
Wald χ^2 statistics (df)	1262.219***	1265.497***
	(4)	(9)

$** \ p < .05$, $*** \ p < .01$.

to that in Amazon Dataset. Both activity and assortativity effects are negative, and they are significant. These imply that active users will like fewer movies, and that they have a tendency towards less popular movies. Controlling for genre in the MovieLens Dataset 1 does not have an influence on the four major effects. The Adventure and Crime genre indicators are found to be significant, which suggest that movie genre has an impact on the formation of appreciation links towards movies. Additional models are developed to examine the role of genre, as described later.

The common findings from results on the Amazon and MovieLens platforms are: (i) transitivity of similar tastes does exist and significantly drives the network formation; (ii) movie popularity plays a strong role in driving user appreciation of movies; and (ii) the more active the users are, the more likely it is for their movie appreciation activities to be driven by their tastes rather than by movie popularity; active users can thus be seen as more likely to be willing to explore and appreciate niche movies. These results provide strong support for success of movie recommendations based on users' similar tastes. The findings also suggest that movie recommendation engines should consider movie popularity when making recommendations to less active users, while recommendations for active users should be based more on shared tastes of users. The results from the two platforms demonstrate different patterns when users are considering newly released movies (in Amazon Dataset), and "older" or previously released movies (in MovieLens

Dataset 1). A user's probability of appreciating an older movie is significantly higher if it is a popular movie.

We perform a sensitivity test on the MovieLens Dataset 1 to check the robustness of our results. Two different sets of medium-active users are selected using different activity level cutoffs. To assess robustness of findings on the original dataset with users having activity levels from 30 to 70, two comparison datasets are taken: the first comparison set includes users with out-degrees from 20 to 50, and the second set includes users who have activity levels from 40 to 90. For each set, we fit both the full model as well as the genre model. The results are provided in Table 6, which show that our conclusions based on medium-active users are consistent regardless of the activity levels considered.

Sensitivity analysis on MovieLens Dataset 1 confirms that movie genre plays a role in the appreciation link formation; we thus conduct additional analysis to further examine the role of genre. We test the role of genre using a data set with users having activity levels from 40 to 90, which provides us the largest number of nodes and links to analyze. Five genre subsets are constructed: Action, Comedy, Drama, Adventure, and Crime, according to the respective genre indicators. Key statistics of the five subsets are shown in Table 7, and Table 8 presents comparative results of the full models.

The taste transitivity effect is found to be significant in all genres, though it is smallest among Drama movies. The significances of popularity, activity, and assortativity effects vary between different genres. The popularity effect is significant for Comedy, Drama and Adventure movies, but not for Action or Crime movies. The activity effect is significantly negative for Comedy and Drama movies, but is positive for Action and Adventure movies. The assortativity effect is negative in all the five subsets but is only significant for Comedy, Drama, or Adventure movies.

This genre analysis provides valuable inputs on using genre to improve movie recommendations. The results suggest that users' similar tastes in Action, Adventure, and Crime movies are strongly transitive to the same type of movies, as indicated by the larger estimates of the transitivity effect than those in Tables 4, 5, and 6. User appreciation for Action and Crime movies is driven primarily by the taste transitivity effect, and the popularity effect is not significant. A negative activity effect is observed in Tables 4, 5 and 6 for the overall data, indicating a general tendency of active users to be less active in the future. However, the activity effect is found to be positive for Action and Adventure movies, implying that active users in these two genres will be more active in the future. The insight for movie recommendation engines is to recommend more Action or Adventure movies to users who are active in these two genres. A difference between these two genres is that popularity and assortativity effects are significant for Adventure movies. Thus, for Adventure movies, recommendation engines should consider movie popularity for less active users, while recommendations for active users should be based more on similar tastes.

Table 6: Sensitivity analysis on MovieLens Dataset 1.

Effects	**20–50 Full**	**20–50 Genre**	**30–70 Full**	**30–70 Genre**	**40–90 Full**	**40–90 Genre**
Taste transitivity	0.0658***	0.0674***	0.0389***	0.0397***	0.0326***	0.0330***
	(0.0041)	(0.0044)	(0.0029)	(0.0028)	(0.0020)	(0.0021)
Movie popularity	1.7153***	1.7100***	1.3157***	1.2925***	1.7616***	1.7196***
	(0.2296)	(0.2313)	(0.2198)	(0.2410)	(0.2570)	(0.2695)
User activity	−0.6935***	−0.6985***	−0.5591***	−0.5612***	−0.3525***	−0.3536***
	(0.0712)	(0.0700)	(0.0583)	(0.0634)	(0.0721)	(0.0908)
User-movie assortativity	−0.2605***	−0.2602***	−0.1431***	−0.1407***	−0.2276***	−0.2237***
	(0.0579)	(0.0592)	(0.0431)	(0.0479)	(0.0472)	(0.0486)
Action		0.0154		0.0281		0.2291***
		(0.0725)		(0.0688)		(0.0668)
Comedy		0.0782		0.1002		0.1313**
		(0.0660)		(0.0647)		(0.0597)
Drama		0.0092		0.0044		0.0183
		(0.0649)		(0.0631)		(0.0611)
Adventure		0.4126***		0.2460***		0.2396***
		(0.0799)		(0.0719)		(0.0733)
Crime		0.3349***		0.2714***		0.2923***
		(0.0772)		(0.0718)		(0.0718)
Wald χ^2 statistics (df)	1510.667***	1877.492***	1262.219***	1265.497***	1338.048***	1312.371***
	(4)	(9)	(4)	(9)	(4)	(9)

$**\ p < .05$, $***\ p < .01$.

Table 7: Descriptive statistics for Movielens Dataset 1 genre subsets.

	Action	Comedy	Drama	Adventure	Crime
Number of movies	*154*	*265*	*358*	*118*	*114*
Number of users	*36*	*55*	*57*	*42*	*46*
Average degree	*11.389*	*13.297*	*17.135*	*8.556*	*7.203*

Table 8: Movielens Dataset 1 genre test results.

Effects	Action	Comedy	Drama	Adventure	Crime
Taste	0.1634***	0.0905***	0.0645***	0.1973***	0.1967***
transitivity	(0.0191)	(0.0102)	(0.0058)	(0.0196)	(0.0229)
Movie	3.0071	1.974***	2.2206***	3.1555***	1.3840
popularity	(1.8061)	(0.4965)	(0.3502)	(1.1365)	(0.9423)
User activity	0.2426	−0.5378**	−0.4224***	0.1270	−0.5329
	(0.7933)	(0.1867)	(0.1142)	(0.4760)	(0.4086)
User-movie	−0.8301	−0.3556**	−0.4135***	−0.9923**	−0.3777
assortativity	(0.5496)	(0.1444)	(0.0905)	(0.3889)	(0.3528)
Wald χ^2	159.324***	380.549***	609.214***	191.005***	170.795***
statistics (df)	(4)	(4)	(4)	(4)	(4)

$** \ p < .05$, $*** \ p < .01$.

Estimates on MovieLens Dataset 2, Sensitivity, and Age Group Analyses

The role of user demographic information is examined using the MovieLens Dataset 2. The full model with four effects is estimated and then the demographic information is added, first the gender, secondly the age ranges, and lastly the geographic regions of user location. Results are presented in Table 9. The taste transitivity effect is seen to be strongly significant in all the four models. The activity effect is also consistently negative and significant in all models. Neither the popularity effect nor the assortativity effect is found to be significant at the 5% level. The lack of significance on popularity and assortativity effects here may be due to the short time period (one year) analyzed in this dataset. Recall that the movie review peak on MovieLens occurs two years after movie release in Figure 4; thus, the popularity related effects might not have full impact within one year. Our findings on nonpopularity related effects are consistent with those on the previous two datasets, which confirm that the findings on this MovieLens Dataset 2 are valid. Neither the gender indicator, nor any of the region indicators are significant. There is thus no evidence that difference in gender or difference in geographic region affects the appreciation link formation. Interestingly, two younger age indicators (under 18 and age 18–24) show a positive and significant effect, with larger coefficients than those for the other four age groups. These indicate that user age has an impact on the taste network formation, with the under-18 and the 18–24 age group users being more likely to form movie appreciation links. This suggests that younger age users show greater enthusiasm for appreciating movies compared

Table 9: Estimated stochastic actor-based models on Movielens Dataset 2.

Effects	W.O. demo.	With Gender	With Age	With Region
Taste transitivity	0.1368*** (0.0402)	0.1375*** (0.0396)	0.1312*** (0.0403)	0.1473*** (0.0405)
Movie popularity	0.8073 (0.6028)	0.8911 (0.5869)	1.3411 (0.6728)	0.8451 (0.6519)
User activity	−1.3032*** (0.1491)	−1.2868*** (0.1438)	−1.2051*** (0.1697)	−1.2986*** (0.1516)
User-movie assortativity	0.0276 (0.1658)	0.0076 (0.1598)	−0.1064 (0.1877)	0.0129 (0.1731)
Gender		−0.1182 (0.3401)		
Under 18			4.5671** (1.5747)	
Age 18–24			2.2191** (0.8433)	
Age 25–34			1.0385 (0.7436)	
Age 35–44			0.3276 (0.7468)	
Age 45–49			0.9528 (0.9171)	
Age 50–55			1.0934 (0.8621)	
Northeast				−0.5653 (0.4234)
Midwest				−0.0637 (0.3734)
South				−0.7564 (0.5257)
Wald χ^2 statistics (df)	1642.279*** (4)	1600.217*** (5)	1568.779*** (10)	1913.903*** (7)

$** \; p < .05, \; *** \; p < .01.$

to the other age groups; recommendation engines, may thus find it useful to use customer age information while making movie recommendations.

To establish robustness of these findings, we perform a sensitivity test by comparing models estimated from two sets of users corresponding to different activity-level ranges. The user activity level plot in Figure 8 shows that user counts drop dramatically as activity levels increase from 1 to 15. At activity levels of 16 to 43, the user counts are stable at around 5 to 10. The main analysis in Table 9 above is thus based on users having activity levels of 20 to 40. At higher activity levels, the numbers of users are seen to drop, to under 5 at activity levels of 45 to 70, and then stay mostly at 1 for activity levels beyond 70. For the comparison dataset, we consider users with activity levels ranging from 30 to 70; these correspond to around the middle one-third of users, according to the plot in Figure 8.

Table 10: Sensitivity analysis on MovieLens Dataset 2.

Effects	20–40 Full	20–40 Age	30–70 Full	30–70 Age
Taste transitivity	0.1368*** (0.0402)	0.1385*** (0.0412)	0.1055*** (0.0189)	0.1093*** (0.0205)
Movie popularity	0.8073 (0.6028)	1.5275** (0.6685)	0.0210 (0.4489)	0.4834 (0.6279)
User activity	−1.3032*** (0.1491)	−1.1567*** (0.1774)	−1.0256*** (0.1014)	−0.9401*** (0.1281)
User-movie assortativity	0.0276 (0.1658)	−0.1607 (0.1902)	0.0852 (0.1014)	−0.0099 (0.1374)
Under 18		4.0338** (1.7923)		1.8197*** (0.5799)
Age 18–24		1.6184** (0.7314)		2.2374*** (0.6519)
Age 25–34		0.4139 (0.5398)		1.4295** (0.4867)
Age 35–44		−0.3298 (0.5635)		1.2506** (0.4952)
Age 45–49		0.3271 (0.8127)		0.8440 (0.4955)
Wald χ^2 statistics (df)	1642.279*** (4)	1710.144*** (9)	2616.077*** (4)	2992.89*** (9)

$** \ p < .05$, $*** \ p < .01$.

Results on the sensitivity analysis, comparing full models with and without age indicators from the two datasets are provided in Table 10. Because there are no users aged above 56 in the comparison set, five age indicators are used in Table 10. The results show that our findings based on medium-active users are consistent; transitivity effect and activity effect are significant in the network generation process, and user age is seen as another factor that impacts link formation. Both datasets show significant and larger coefficients for the two younger age indicators (under 18 and age 18–24), compared with the other age groups; this provides clear evidence of younger users having greater enthusiasm for appreciating movies, a finding which can be useful in customizing recommendations for younger users.

To better discern age related differences, age group test is performed. In order to obtain sufficient samples, we combine the two age groups of under 18 and 18–24 into one group of users with age under 24. We then compare this to the group of users with age in the 25–34 range. Key statistics on the two groups are listed in Table 11. The numbers of nodes and average degrees in the two subsets are seen to be comparable, which is an important consideration for exploring age-related differences in the two data subsets. Table 12 provides age-related results from the full models. Surprisingly, the significance of taste transitivity effect disappears in the under 24 group. A possible and intuitive explanation is that young people have varying tastes, or their tastes in movies are still developing. This can be compared with the result on the 25–34 age group, where the taste transitivity effect

Table 11: Descriptive statistics for Movielens Dataset 2 age group subsets.

	Under 24 Group	Age 25–34 Group
Number of movies	*238*	*392*
Number of users	*21*	*31*
Average Degree	*15.635*	*17.398*

Table 12: Movieslens Dataset 2 age group test results.

Effects	Under 24 Group	Age 25–34 Group
Taste transitivity	0.1201	0.2601**
	(0.1364)	(0.1027)
Movie popularity	8.0231	2.8671
	(5.5332)	(2.7450)
User activity	0.6007	−0.9867
	(1.4334)	(0.7125)
User-movie assortativity	−2.2403	−0.5111
	(1.8526)	(0.9008)
Wald χ^2 statistics (df)	50.778***	344.129***
	(4)	(4)

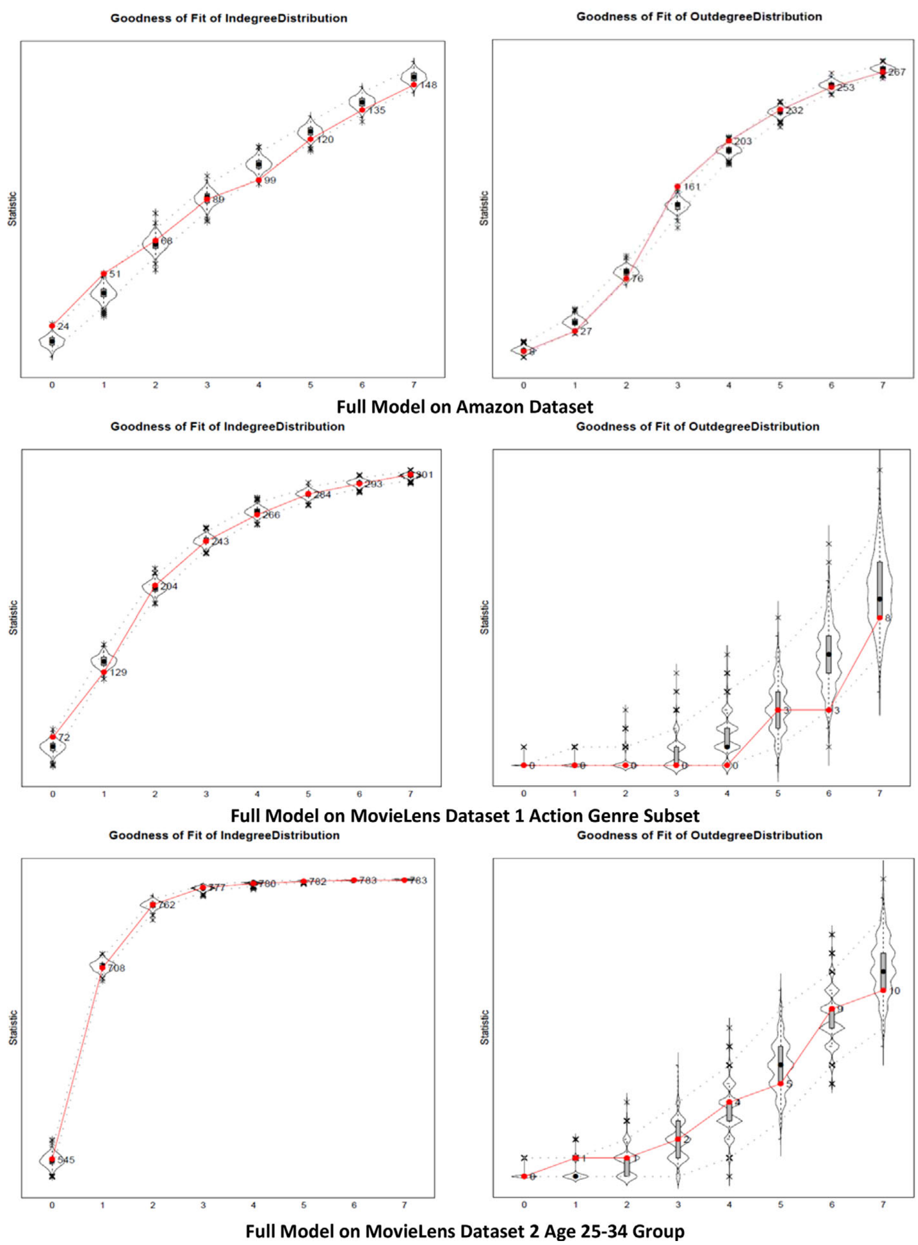
$**\ p < .05$, $***\ p < .01$.

is significant. The insight for movie recommendation engines is thus to recommend a broader range of movies to younger audiences whose movie tastes may still be developing. In addition, the activity effect is positive in the under 24 age group, indicating that younger users with higher activity may like more movies. This finding coincides with the evidence in the main analysis, where younger users are seen to exhibit greater enthusiasm for appreciating movies. A plausible reason for this phenomenon can be that young users are eager to watch movies after they are clear of the age 18 restriction of MPAA film ratings, and like to explore different types of movies.

Goodness of Fit Diagnostics

We perform goodness-of-fit diagnostics to evaluate the ability of our models to reproduce two key characteristics of the user-movie taste network, the movie in-degree distribution and the user out-degree distribution. We analyze the models by comparing the observed degree distributions with the corresponding values in all the networks simulated during the model estimation processes (recall that we iterate the simulation runs until sufficient convergence is obtained). Figure 9 provides three of the full model diagnostics: on the Amazon Dataset, on the MovieLens Dataset 1 Action genre subset and on the MovieLens Dataset 2 age 25–34 group. The solid lines represent the cumulative distributions of in- and out-degrees for the observed networks. The violin-plots present the distributions of the statistics in simulated networks as combinations of box plots and smooth approximations to the density, with the observed values superimposed. Model

Figure 9: Goodness-of-fit diagnostics for in-degree and out-degree distributions.



fit is considered good when all the observed values fall within the range of the simulated network statistics. Plots in Figure 9 demonstrate the ability of our models to reproduce the observed networks satisfactorily. A Wald test for joint significance of factors is carried out for every model as well. Chi-square statistics with degrees of freedom and significances are reported in tables for each model. The strong significances of chi-square statistics in all the Wald tests, combined with the goodness-of-fit diagnostics on degree distributions, show that the user-movie taste network evolution depends on the endogenous network effects and characteristics of users and movies that we analyze in this study.

CONCLUSION AND FUTURE WORKS

Collaborative filtering-based recommendations rely on similarity in users' tastes towards products or services, and assume that similar tastes are transitive from earlier observed behavior to future purchase decisions. This underlying essential premise for successful recommendations, has, however, not been established in earlier research. This article presents a first study on whether such taste transitivity is significant in user-review datasets, which form the basis for collaborative filtering-based recommendations. It examines the significance of taste transitivity, and estimates the strength of this effect, together with other factors, which play roles in successful recommendations.

Understanding the factors and underlying process driving successful recommendations is key to the design of more effective recommenders and consumer decision support systems. Using movie recommendations as the example, this article is the first to apply a stochastic network analyses approach on online review data to statistically analyze different factors, which contribute towards making effective recommendations. It constructs actor-based networks from real review data sets, and analyzes the dynamics of these networks representing consumer behavior. Examining the evolution of such taste networks over time provides valuable inputs into recommendation engine design for decision support. Based on results from three movie review datasets from Amazon and MovieLens, we find that the similar taste transitivity effect is significant across different electronic platforms. It thus provides support for collaborative filtering methods to recommend movies based on user taste similarity. Our findings also suggest that movie popularity should not be neglected. The popularity effect can be very large in certain contexts, such as for movies released more than one year ago; and its impact on active users may be limited. Besides taste transitivity and popularity effects, results from this study show that other factors can also play important roles in improving the effectiveness of recommendation engines; these include movie genre, user activity level and user age.

The two-mode network analyses approach used in this study will be broadly useful in obtaining a better understanding of factors that drive user appreciation for varied products in e-business. An extension to other types of products is a topic of ongoing research. A limitation of this approach is the computing limit arising from the need to process the large state space in the Markov Chain modeling. An important area of future work is in the improvement of current tools so that larger scale datasets and networks can be efficiently analyzed.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6)734–749.

Al-Natour, S., Benbasat, I., & Cenfetelli, R. T. (2008). The effects of process and outcome similarity on users' evaluations of decision aids. *Decision Sciences*, *39*(2), 175–211.

Anderson, C., Wasserman, S., & Crouch, B. (1999). A p* Primer: Logit models for social networks. *Social Networks*, *21*, 37–66.

Barabasi, A. (2002). *Linked: The new science of networks*. Cambridge: Perseus.

Bhosale, N. S., & Pande, S. S. (2015). A survey on recommendation system for big data applications. *Data Mining and Knowledge Engineering*, *7*(1), 42–44.

Bobadilla, J., Ortega, F., Hernando, A., & Gutierrez, A. (2013). "Recommender systems survey" *Knowledge-Based Systems*, *46*, 109–132.

Branco, F., Sun, M., & Villas-Boas, J. M. (2012). "Optimal search for product information" *Management Science*, *58*(11), 2037–2056.

Breiger, R. (1974). "The duality of persons and groups" *Social Forces*, *53*(2), 181–190.

Breiger, R. (2000). "A toolkit for practice theory" *Poetics*, *27*(2–3), 91–115

Burt, R. S. (1980). "Models of network structure" *Annual Review of Sociology*, *6*, 79–141

Burt, R. S. (1987). "Social contagion and innovation: cohesion versus structural equivalence" *American Journal of Sociology*, *92*, 1287–1335

Carley, K. M. (1991). "A theory of group stability" *American Sociological Review*, *56*(3), 331–354

Coleman, J. S. (1957). *Community conflict*. New York: Free Press.

Coleman, J. S. (1986). *Individual interests and collective action: Selected essays*. New York: Cambridge University Press.

Conaldi, G., Lomi, A., & Tonellato, M. (2012). "Dynamic models of affiliation and the network structure of problem solving in an open source software project" *Organizational Research Methods*, *15*(3), 385–412.

Contractor, N. S., & Eisenberg, E. M. (1990). Communication networks and new media in organizations. In J. Fulk, & C. W. Steinfield (Eds.), *Organizations and communication technology*. Newbury Park: Sage,143–172.

Cook, G. J. (1993). "An empirical investigation of information search strategies with implications for decision support system design" *Decision Sciences*, *24*(3), 683–698

Doreian, P., Batagelj, V., & Ferligoj, A. (2004). "Generalized blockmodeling of two-mode network data" *Social Networks*, *26*(1), 29–53

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

Golbeck, J. (2006). Generating predictive movie recommendations from trust in social networks. *Proceedings 4th International Conference on Trust Management*, Berlin, Heidelberg: Springer, 93–104

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, Berkeley, CA, USA, 230–237

Ibarra, H. (1993). "Personal networks of women and minorities in management: A conceptual framework" *Academy of Management Review*, *18*, 56–87

Ibarra, H. (1995). "Race, opportunity, and diversity of social circles in managerial networks" *Academy of Management Journal*, *38*, 673–703

Janakiraman, R., & Niraj, R. (2011). "The impact of geographic proximity on what to buy, how to buy, and where to buy: Evidence from high-tech durable goods market" *Decision Sciences*, *42*(4), 889–919

Laumann, E. O. (1966). *Prestige and association in an urban community*. Indianapolis: Bobbs-Merrill.

Lee, Y., Hosanagar, K., & Tan, Y. (2015). "Do I follow my friends or the crowd? Information cascades in online movie ratings" *Management Science*, *61*(9), 2241–2258

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456–474

Liang, T. P. (2008). Recommendation systems for decision support: An editorial introduction. *Decision Support Systems*, *45*(3), 385–386

Linden, G., Smith, B., & York, J. (2003). "Amazon.com recommendations: Item-to-item collaborative filtering" *IEEE Internet Computing*, *7*(1), 76–80

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). "Recommender system application developments: A survey" *Decision Support Systems*, *74*, 12–32

Marsden, P. V. (1988). "Homogeneity in confiding relations" *Social Networks*, *10*, 57–76

McAuley, J., Targett, C., Shi, Q., & van den Hengel, A. (2015a). Image-based recommendations on styles and substitutes. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, Chile, 43–52

McAuley, J., Pandey, R., & Leskovec, J. (2015b). Inferring networks of substitutable and complementary products. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia, 785–794

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face to face groups. *American Sociological Review*, *52*, 370–379

Mische, A., & Pattison, P. (2000). Composing a civic arena: Publics, projects and social settings. *Poetics*, *27*(2–3), 163–194

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.

Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks, II: Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, *52*, 169–193

Robbins, H., & Monro, S. (1951). “A stochastic approximation method” *Annals of Mathematical Statistics*, *22*(3), 400–407

Robins, G., & Alexander, M. (2004). “Small worlds among interlocking directors: Network structure and distance in bipartite graphs” *Computational & Mathematical Organizational Theory*, *10*(1), 69–94

Robins, G., Pattison, P., & Wasserman, S. (1999). “Logit models and logistic regressions for social networks, III: Valued relations” *Psychometrika*, *64*, 371–394

Sharma, L., & Gera, A. (2013). “A survey of recommendation system: Research challenges” *International Journal of Engineering Trends and Technology*, *4*(5), 1989–1992

Snijders, T. A. B. (1996). “Stochastic actor-oriented models for network change” *Journal of Mathematical Sociology*, *21*, 149–172

Snijders, T. A. B. (2001). “The statistical evaluation of social network dynamics” *Sociological Methodology*, *31*, 361–395

Snijders, T. A. B. (2005). Models for longitudinal network data. In *Models and methods in social network analysis (Ch. 11)*. New York: Cambridge University Press,215–247.

Snijders, T. A. B., Steglich, C. E. G., & Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences*. Mahwah: Lawrence Erlbaum,41–71.

Snijders, T. A. B., Koskinen, J. H., & Schweinberger, M. (2010). “Maximum likelihood estimation for social network dynamics” *Annals of Applied Statistics*, *4*, 567–588

Snijders, T. A. B., Lomi, A., & Torlo, V. J. (2013). “A Model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice” *Social Networks*, *35*, 265–276

Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). “Research commentary—Information in digital, economic, and social networks” *Information Systems Research*, *24*(4), 883–905

Wang, H., Guo, X., Zhang, M., Wei, Q., & Chen, G. (2016). “Predicting the incremental benefits of online information search for heterogeneous consumers” *Decision Sciences*, *47*(5), 957–988

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York and Cambridge: Cambridge University Press.

Wasserman, S., & Pattison, P. (1996). "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*" *Psychometrika*, *61*, 401–425

Minghong Xu is an assistant professor in information systems at the Johns Hopkins Carey Business School, Baltimore, United States of America. She received her BS degree in computer science from University College Dublin, Dublin, Ireland, and her MS degree in Operations Research from Columbia University, New York, United States of America. She completed her PhD in management information systems at the University of Illinois, Chicago, United States of America. Her research interests include analytics, big data, network analysis, dynamic pricing, and supply chain management. Her work has been presented at national and international conferences including INFORMS, AMA, and AAAI.

Siddhartha Bhattacharyya is Professor of Information and Decision Sciences in the College of Business Administration at the University of Illinois, Chicago. He received his PhD from the University of Florida in 1993. His research addresses data intensive applications, machine learning, information management, agent-based models, and evolutionary computation. His work appears in journals like Complex Systems, Decision Support Systems, Decision Sciences, Evolutionary Computation, MIS Quarterly, IEEE Transactions on Evolutionary Computation, INFORMS Journal of Computing, Journal of Economic Dynamics and Control, Social Networks, and in various conferences like ICDM, KDD, HT. He teaches courses in business data mining and technology project management and has received multiple teaching awards.