A Comparison of Product Network and Social Network Based Recommendation Engines for Twitter Users

Shawndra Hill
OPIM
The Wharton School
University of Pennsylvania
shawndra@wharton.upenn.edu

Adrian Benton
OPIM
The Wharton School
University of Pennsylvania
adrianb@wharton.upenn.edu

Christophe Van den Bulte
Marketing
The Wharton School
University of Pennsylvania
vdbulte@wharton.upenn.edu

Introduction

The recent explosion of data available from social media has enabled both product and social networks to be linked to user attributes, such as demographics, and to business dependent variables, such as purchasing and fraud. Beyond the growing dataset social media makes available on user attributes like demographics, deriving value from social networks has become easier because social network data are also available through the Application Programming Interfaces (APIs) of sites such as Twitter, Facebook, and YouTube. Such data enable researchers and practitioners to determine the features of users' social networks, which can predict other characteristics of given users. In this study, we explore the extent to which we can use attributes, such as how many of a user's friends follow a particular brand, to predict customer engagement, which is measured by the brands users follow on Twitter. This research makes predictions for hundreds of thousands of users and for hundreds of brands.

Online social networks have been studied to explore a wide range of research questions, in a wide variety of fields, that relate to our work. In the Information Systems literature, the most notable of these concern the spread of information and influence on social networks (Aral et al. 2009b). It is widely known that it is difficult to identify influence in observational data because it is difficult to separate from confounding factors of product and brand adoption such as homophily, the notion that similar people cluster together from influence, or contagion, the idea that people influence others to take certain actions (Shalizi and Thomas 2012). While recent work on influence relates to our study because it links social network features to product adoption outcomes, researchers assume that their task is to identify influence, in particular distinguishing it from homophily.

In this paper, we instead focus on social network-based prediction, which does not rely on knowing how and why people are connected but the network structure and demographics of the brand audience. In our work, we capitalize on the fact that we know how brand preferences are correlated among friends because these preferences are made visible to us online because we can observe the brands consumers follow on social media. Our goal is to highlight differences in our ability to predict for different brands based on both brand and user characteristics. In addition, we compare two types of network that are in play for most online firms, the social network (a network of friends) and the product network (a network of products connected through consumers who are not necessarily friends). Recent work in Information Systems and Marketing has studied the fact that both product and consumer networks are important to searching for information (Goldenberg et al. 2012) and content on the web. In this work, we compare the two networks' value for predicting which brands people like.

Social network-based prediction on a large scale — predicting individuals' attributes based on those of their friends (Domingos and Richardson 2001) — is a relatively new practice. Social network-based prediction has proven to be successful in domains such as targeted marketing telecommunications services (Hill et al. 2006b), online targeted advertising (Provost et al. 2009), the adoption of online services (Aral et al. 2009a), online searches (Goldenberg et al. 2012), and fraud detection (Hill et al. 2006a). However, earlier studies investigated only one product in one context at a time, and when researchers did have access to information about different products or business outcomes, data on the social networks of users was extremely limited. Therefore, there is very little knowledge about which product types and services are conducive to accurate social network-based predictions.

We compare the predictive performance of a social network-based recommendation engine to the performance of product network based recommendation algorithm based on the networks formed by Twitter users and the brands they follow, across product and user categories. The task is to predict which of more than 600 brands in 15 industry categories are followed Twitter users. We compare the approaches using categories of both users and brands. First, we compare the results of the two recommendation algorithms by the size of the users' social network and the product network. Second, we compare the results by the brands' audience — in particular, whether the brands have a target "niche" audience.

Our main contribution is an exploration of the performance of social network-based predictions over a large number of brands and product categories. We find that performance varies by a number of factors, including the specificity (how skewed the audience is to a particular demographic) and size of the brand audience and the size of the user's social network. Previous work in social networks explored homophily, the idea that similar people are connected. Our results confirm that identifying similar people is important for predicting a brand's followers, because brand audiences have particular demographics, and the social network-based approach performs extremely well as long as the brand has a significant number of users or, in the case of Twitter, followers.

Our work has implications for both research and practice. To our knowledge, this is the first work to compare the two types of approaches often used by firms to make recommendations to users. We find in our study that there are some brands for which social network-based prediction works extremely well and others for which product network-based prediction works extremely well. Beyond this difference in brand performance, we show that performance also varies by user type – in particular by the size of the network of the user.

Data

We collected a large database of more than 700 brands in 15 different industries that have a following on Twitter. We then tracked all the Twitter followers of each brand in our database, and created a sample of these followers by using their first-degree social networks. We then collected brand-related content from Facebook in order to assess and explain the performance of our recommendation predictions for users based on different aspects of the brands, including audience demographics. The steps for data collection and the descriptive statistics and plots are described below.

The dataset was collected through the following process: **(1)** A set of 734 widely recognized brands across 15 product types was identified from various online sources. **(2)** Next, 631 Twitter handles were found for brands that had a social media presence on Twitter. If a brand had multiple Twitter handles, we chose the handle with the most followers. **(3)** Using the Twitter API, the follower network of all 631 brands was collected, resulting in a network of approximately 18 million brand followers. **(4)**

Random samples were created from each brand's follower network. Only users who had between 1 and 2,000 followers were considered in this network in order to avoid capturing celebrities', brands', and companies' handles, which tend to have massive numbers of followers. In addition, all users were required to follow at least two of the 631 brands. For each brand in our dataset, a user meeting these criteria was selected a large number of users at random from that brand's network. This process was repeated for all brands in our data, resulting in a sample of 223,517 users. **(5)** For all sampled users, the Twitter API was queried to collect the user identifications of their Twitter friends and followers. Thus, in addition to the network of brand followers, we also constructed the users' first-degree networks. **(6)** For each of these 631 brands, the Facebook Ads API was queried in order to retrieve aggregate-level demographic characteristics of the Facebook users who liked these brands. The demographic features were collected for 624 brands. During this process, we advertised our lab homepage on Facebook in order to get access to Facebook level demographics on brands and TV shows. In doing so we were able to get the proportions of users that follow brands of a certain demographic – for example, the proportion of women that follow Tide. The aggregate-level demographic features and their mean proportions are listed in Table 1. These features of the brands are used to explain our performance.

The data we collected are complex and present two highly interesting dimensions: the brand nicheness and the user type. Table 1 displayed features of the brand audience that we will use to calculate nicheness below and Figure 1 provides some detail about the distributions of the characteristics of the users. We plot the distribution of the number of users' followers and the number of brands followed by users.

The data we collected uniquely enables testing the recommendation systems using both types of networks (product and social). We can compare different approaches across many brands. In this paper, we focus on two different aspects of the brands described above: 1) popularity on Twitter and 3) audience demographics (in particular, whether a brand has a specific “niche” audience). In addition, we focus on user characteristics: the number of followers the user has and the number of brands the user follows.

Table 1. Demographic features collected from the Facebook Ad API and their mean proportion in the collection of brands

Demographic dimension	Demographic feature	Mean proportion
Gender	Men	37%
	Women	63%
Age	13–17	8%
	18–20	15%
	21–24	18%
	25–34	23%
	35–49	22%
	50–54	6%
	55–64	6%
	65+	3%
Education level	In high school	9%
	In college	13%
	Graduated college	78%
Family status	Is a parent	41%
Ethnicity	Hispanic	8%
	Non-Hispanic	92%

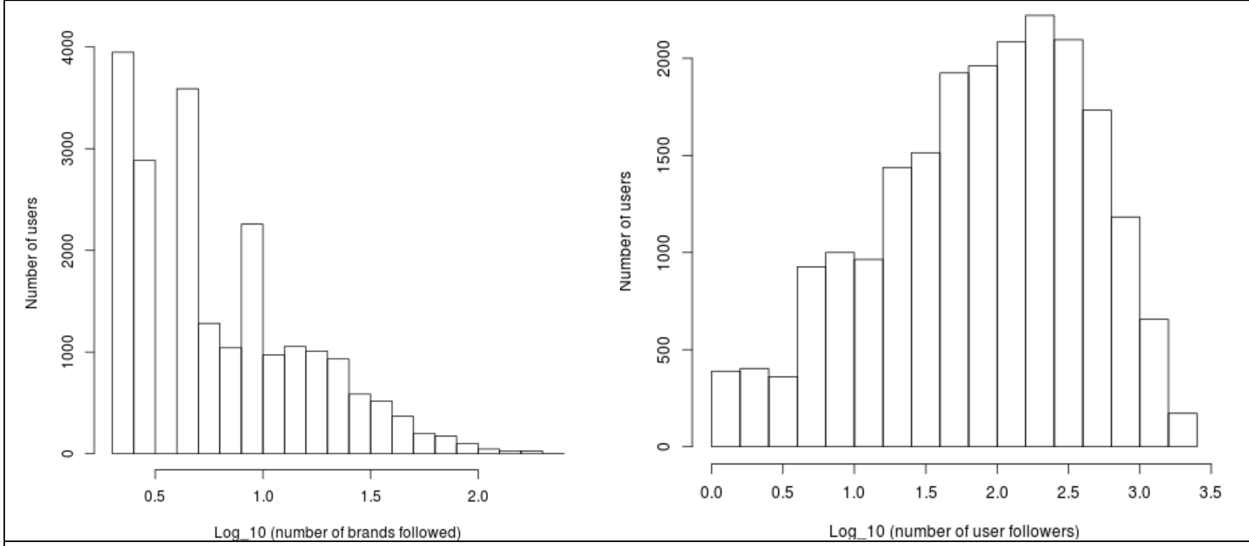

Figure 1. User feature. Distribution of the number of brands followed per user (left), and (right) the number of followers per user, log (base 10) transformed.

Methods

We compare two different approaches to making product recommendations. First, in a product network approach, we associate users by the number of brands they share. Users who share many brands are counted as similar. Secondly, we take a social network-based marketing approach, which bases recommendations on the brands to which immediate followers connect.

Product network-based system

Given a particular test user, u, who is known to follow the set of brands A, we calculate the similarity between this user and all training-set users as follows: For each training-set user, v, who follows the set of brands B, we calculate the similarity of u to v as sim(u, v) = |intersection(A, B)|/|A|. We then select the most similar users from the training set, K, and rank recommendations based on the popularity of brands among similar users. We empirically set the parameter K to 20, because we found that the performance of the system plateaued at that point. Brands in the set of input brands are omitted from the recommendations list.

Social network-based system

Given a particular test user, this approach finds all the followers of that particular test user, excluding any user in the test set. We rank recommendations based on the popularity of brands in the test user's local follower network. In other words, the brand that would be recommended first to a user is the brand that is followed by the largest number of the user's friends. As in the product network based system, brands which belong to the test user's input brands are not included in the recommendation rankings.

Evaluation

Overall performance

First, the set of users was randomly split into training and test sets for 10-fold cross-validation, with 21,027 users in each test set. We used recommendation systems to make predictions for test-set users based on the training-set users. These systems were posed the following problem: given a particular test user who follows N brands, give the system N-1 input brands that the user follows. Given the user's local network and the N-1 brands it follows, attempt to predict the N^{th} output brand the user follows. The held-out brands that the systems attempted to predict were selected in round-robin fashion from the brands that test users followed.

We identified two recommendation systems to assess: a collaborative filtering K-nearest neighbor approach and a social network-based approach, as described above. These systems were evaluated on their recall after a set number of recommendations for a set of hold out brands for each user. In other words, given K recommendations, for what proportion of users in our test set could we recommend the N^{th} held-out brand they follow within the K recommendations? The two recommendation engine methods were then evaluated in two ways as described below: first, across categories, considering all possible test users and held-out brand pairs, and secondly, within categories, considering only recommendations within a product category.

Comparing across brand and user types

In addition to evaluating these methods' performance as a function of the number of overall recommendations made over test users, we also assessed their performance along various dimensions, including brands' popularity and audience skewness. We also assessed performance based on the users' popularity.

We divided the samples by two user feature, the number of followers) and the number of input brands, *N-1* input brands. In addition, we compared the performance of the recommenders by the held-out brand features, the number of brand followers/popularity, gender skew, age skew, education level skew. The skew features correspond to the symmetric *KL*-divergence from the observed distribution over the demographic groups for the held-out brand compared to the mean distribution for those demographic groups over all brands in our data. For brand demographic distribution, *D*, and mean distribution over all brands, *M*, over *n* groups, the symmetric *KL*-divergence between D and M is defined as $\sum_{i=1}^{n} \ln(\frac{D(i)}{M(i)})D(i) + \ln(\frac{M(i)}{D(i)})M(i)$. A higher *KL*-divergence thus corresponds to a less typical demographic make-up of the brand's followers. Suppose that, on average, 75% of brand followers are female and 25% of followers are male. In this case, a brand that is followed by an equal number of males and females would have a higher KL-divergence, or be considered a less typical demographic distribution than a brand with 77% female followers. When evaluating the two systems over these dimensions, the recall for each system is reported, with the number of recommendations fixed to 20, which is a reasonable number of recommendations for a firm to make. We present the results in the next section, in two parts: first comparing the results using the dimensions of brand and user popularity and, secondly, using product category and audience skew.

Results

Overall performance

Figure 2 displays the overall performance of the social network-based and collaborative filtering-based methods across all brands. When considering a low number of recommendations, the social network-based method performs better than the collaborative filtering system. However, the higher the number of recommendations made, the more the social network-based system's performance dramatically degrades . This trend occurs because the social network-based approach runs out of items to recommend. If a users' friends only follow 5 brands, then 10 brands cannot be recommended, instead the social network-based approach would only be able to recommend at most 5.

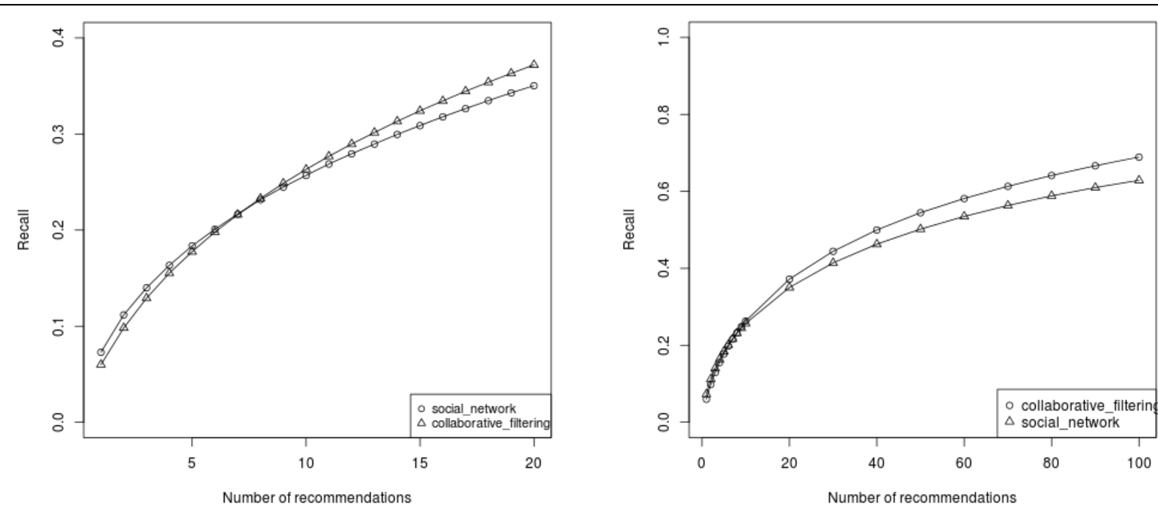

Figure 2. Across Categories. Overall recall of the *K*-nearest neighbor collaborative filtering and social network-based systems. The social network-based approach performs well at first, but the collaborative filtering approach overtakes it. The left plot focuses on the top 20 recommendations.

While the social network-based approach outperforms the product network based approach when making predictions across categories, it does not do so when making predictions only within categories. This is demonstrated in Figure 3, which displays the two strategies' performances when making only within-category recommendations. Though we highlight only three categories, a consistent pattern emerges: when making within-category recommendations, the collaborative filtering-based approach performs as well or better than the social network-based approach.

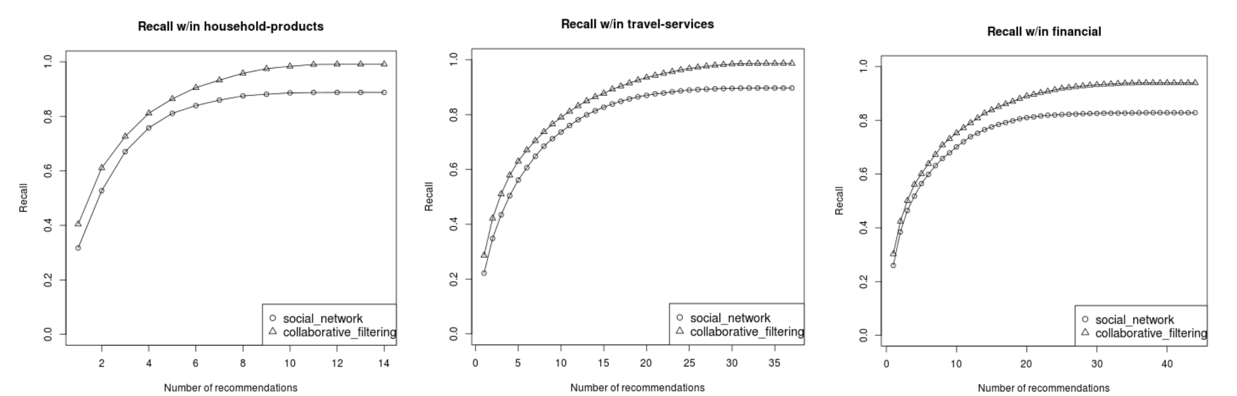

Figure 3. Within-Categories Analysis. Overall recall of the *K*-nearest neighbor collaborative filtering and social network-based systems. The collaborative filtering approach typically outperforms the social network system when making recommendations within categories. The household products, travel services, and financial categories are presented here.

Comparing across brand and user types

We first compare the systems' performance by the popularity of both brand and user. Figure 4 (left) displays the performance of both systems as a function of the number of followers for the held-out brand and the test user. For each approach, we present five lines representing the predicted brands divided by number of followers. The red lines present results from the social network-based approach, and the blue lines those of the collaborative filtering approach. Bin 5 contains the brands with the most followers, and

Bin 1 the brands with the fewest followers. On the horizontal axis, we divide users by their number of followers. The results suggest that users with large numbers of followers are crucial for the social network-based method to perform well and that both systems more easily predict more popular brands, although the social network-based method appears to be more sensitive to popularity of the brand. The social network-based approach does not perform well with unpopular brands. To make this data easier to interpret, these results are also plotted in three dimensions (3D) on the right side of Figure 4.

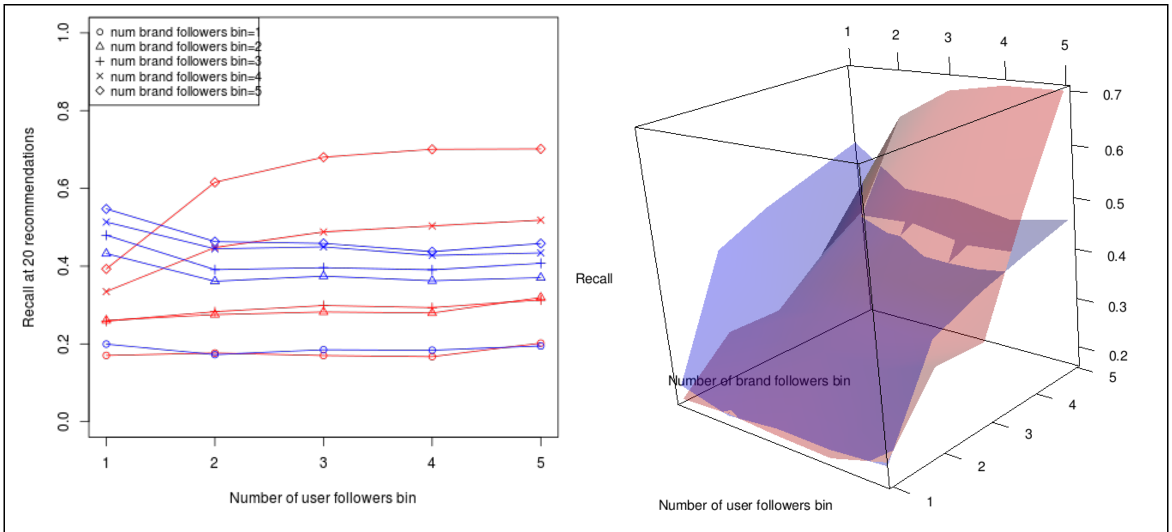

Figure 4. Performance of the social network and the collaborative filtering methods as a function of the held-out brand's popularity and the test user's number of followers. Users and brands were placed in equal-sized bins based on the number of followers. The number of recommendations was fixed to 20. (Left) The lines correspond to the different brands' bins. (Right) The blue plane corresponds to the collaborative filtering system, and the red to the social network-based system.

Secondly, we compare their performance by the nicheness of the held-out brand. These results are the most exciting, because they demonstrate that the performance of the social network-based approach varies among different types of brands, differing by brand "nicheness". We evaluate performance by the held-out brand's demographic skew. Figure 5 displays the difference in recall between the two systems as a function of the demographic skew and the popularity of the predicted held-out brand (horizontal axis). Each plot has two lines: one line (red) represents the brands that have higher than average skew, and the second line (blue) the brands that have lower than average skew. The results suggest that the social network-method performs better when the predicted brand has an atypical age and/or education bias or an audience of a specific age. A gender bias does not necessarily yield better performance from the social network method. However, it is important to note that the social network-based approach does well with popular brands with both low and high demographic skews.

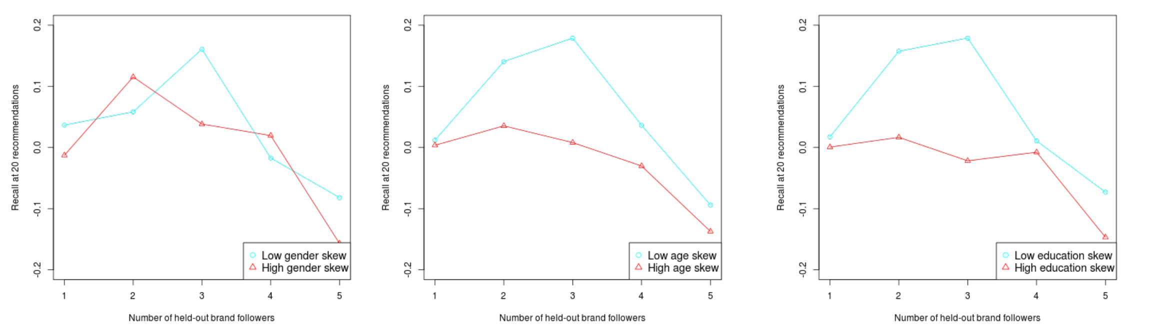

Figure 6. Difference in recall of the two systems as a function of the held-out brand's (left) gender skew, (center) age skew, and (right) education level skew. Negative values represent cases in which the social network system tends to outperform the collaborative filtering system and vice-versa.

Discussion and next steps

A number of recent studies in information systems and business intelligence have investigated whether both social and product networks can be used to derive value for firms. Many popular companies also use social network features to predict their customers' outcomes and for better target marketing. However, there is little understanding of in what circumstances social networks can aid prediction at both the product and the user level, and when they cannot. In this study, we demonstrate that performance varies across brands and product types, something that to our knowledge no earlier research had shown. We then explore possible reasons for the difference in performance by brands by investigating two features associated with brands (number of followers and demographic skew) and one feature associated with users (number of followers). Our results demonstrate that, consistent with sociological literature on homophily, social network-based targeting works best when the brand audience has a demographic skew. Additionally, both the brand and user must have a significant number of connections for social network-based targeting to be feasible.

Our research presents many possible next steps. First, our prediction tasks have many moving parts: users and their social networks, and brands and their networks. One important feature for future work is to investigate which products and product types people tend to follow and why. We would like to understand whether true purchase preferences can be evaluated from Twitter following data via online surveys of brand followers. Our data shows that a relatively large number of people follow expensive luxury brands, but it is unlikely that all followers are actually able to purchase their products. On the other hand, common, everyday brands like soap and laundry detergent, whose use does not necessarily add anything to a person's reputation, are not followed by a significant number of users on Twitter.

It is important to acknowledge here that in this paper we are making a number of assumptions, including that people actually like the brands they follow and that following brands on Twitter is reflective of the types of things people actually purchase.

In addition, we are assuming that Twitter's recommendation engine is not biased in a way that hurts or helps our predictions. This work cannot therefore be taken as a final statement on social network-based prediction, and these assumptions will have to be tested and validated by further research. Nevertheless, despite these assumptions, this paper is an important first step at understanding the differences in performance between product network and social network based predictions.

References

Aral, S., Muchnik, L., and Sundararajan, A. 2009a. "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences* (106:51), pp. 21544-21549.

Aral, S., Muchnik, L., and Sundararajan, A. 2009b. " Distinguishing Influence Based Contagion from Homophily Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences (PNAS)* (106:51).

Domingos, P., and Richardson, M. 2001. "Mining the Network Value of Customers," in: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. San Francisco, California: ACM, pp. 57-66.

Goldenberg, J., Oestreicher-Singer, G., and Reichman, S. 2012. "The Quest for Content: The Integration of Product Networks and Social Networks in Online Content Exploration," *Journal of Marketing Research* (49:4), pp. 452-468.

Hill, S., Bell, R., Agarwal, D., and Volinsky, C. 2006a. "Building an Effective Representation for Dynamic Networks," *Journal of Computational and Graphical Statistics* (15:3), pp. 584 - 608.

Hill, S., Provost, F., and Volinsky, C. 2006b. "Network-Based Marketing: Identifying Likely Adopters Via Consumer Networks," *Statistical Science* (21:2), May, pp. 256-276.

Provost, F., Dalessandro, B., Hook, R., Zhang, X., and Murray, A. 2009. "Audience Selection for on-Line Brand Advertising: Privacy-Friendly Social Network Targeting," *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*: ACM, pp. 707-716.

Shalizi, C.R., and Thomas, A.C. 2012. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research* (40:2), pp. 211-239.