Diffusion of Innovations Revisited: From Social Network to Innovation Network

Xin Rong
School of Information
University of Michigan
ronxin@umich.edu

Qiaozhu Mei
School of Information
University of Michigan
qmei@umich.edu

ABSTRACT

The spreading of innovations among individuals and organizations in a social network has been extensively studied. Although the recent studies among the social computing and data mining communities have produced various insightful conclusions about the diffusion process of innovations by focusing on the properties and evolution of social network structures, less attention has been paid to the interrelationships among the multiple innovations being diffused, such as the competitive and collaborative relationships between innovations. In this paper, we take a formal quantitative approach to address how different pieces of innovations "socialize" with each other and how the interrelationships among innovations affect users' adoption behavior, which provides a novel perspective of understanding the diffusion of innovations. Networks of innovations are constructed by mining large scale text collections in an unsupervised fashion. We are particularly interested in the following questions: what are the meaningful metrics on the network of innovations? What effects do these metrics exert on the diffusion of innovations? Do these effects vary among users with different adoption preferences or communication styles? While existing studies primarily address *social* influence, we provide a detailed discussion of how innovations interrelate and influence the diffusion process.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining

Keywords

diffusion of innovations; innovation networks

1. INTRODUCTION

"We don't adopt techniques; techniques adopt us."

The study of the diffusion of innovations is concerned with the adoption and spreading of new products, techniques, algorithms, and ideas via certain communication channels among individuals and organizations, usually in the context of a social network [24]. Having an innovation spread quickly in a social system is not a trivial problem. Many social scientists and economists have developed theories to optimize rival marketing strategies for promoting innovations. Among such studies, three elements of the diffusion process are often considered: the attributes of the innovation, the communication channel, and the social network structure [24, 25].

Considerable effort in diffusion studies has been devoted to both modeling the macro diffusion process and modeling the behavior of individual users. Studies on the macro level usually focus on modeling the growth of a population's collective attention to an innovation [4, 11, 19, 5]. Other works look into the structural characteristics of interpersonal networks and capture the impact of social influence [18, 23, 26, 29]. Diffusion studies regarding individual user's behavior have become increasingly popular by taking advantage of newly emerged social network data, such as Facebook, Twitter and LiveJournal [2, 28, 32], as well as academic collaboration networks, such as co-authorship networks and citation networks [8, 12, 15]. These studies have revealed and reconfirmed the underlying connections between social influence and the outcomes of diffusion.

While the data mining community has extensively explored the impact of social influence on the dynamics of diffusion, less attention has been paid to the interactions between innovations. In contrast, rival marketing studies focus on the diffusion of multiple competing products and model product interactions [1, 13], but such studies typically consider just two or a few products that roll out concurrently, and lack a comprehensive account for the inter-relationships among all innovations in the same industry. We think that a study enabled by big data addressing the interaction among a large group of innovations and its impact on diffusion is urgent.

In this paper, we take a formal quantitative approach to account for how the inter-innovation relationships explain the variance of user adoptions. In many real-life situations, we have access to text content that describes, documents, reviews, and compares the innovations. Innovations most often appear as noun phrases or entities in such textual documents, which allows us to use text mining methods to discover and analyze the relationships among the entities under different statistical association or similarity measures.

Compared to other methods that establish similarity measurement among entities, such as collaborative filtering, extracting relationships from textual content allows the sepa-

CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

http://dx.doi.org/10.1145/2505515.2505587.

ration of innovation-specific information from user-involved information. This separation is especially important in predicting the diffusion pattern at the early stage of diffusion, when few adoption records are available. Given these motivations, we choose to establish a network of computer algorithms and statistical models (i.e., the network of innovations) for our study. We build this network by mining a large collection of journal articles and conference papers in computer science. Within the network, a link exists between two innovations if their similarity or association is above a threshold. In our analysis, we also include a citation network of authors, which helps us distinguish and compare the factors related to the social network and factors related to the innovation network in predicting users' adoption behavior and innovation diffusion patterns.

We are particularly interested in the following research questions: (1) What are the meaningful metrics on the network of innovations? (2) How do innovations with different network attributes (measured under the above metrics) differ in terms of their diffusion process? (3) How do such effects relate and compare to social influence through social networks? (4) Do these effects vary among users with different adoption preferences or communication styles?

To address these research questions, we transfer concepts from socioeconomic literature, such as exclusiveness, perceived advantage, prestige and social influence, to quantitative measures on innovation networks and social networks. We also introduce multiple concepts on innovation networks according to the analogy that an innovation and a user are symmetric in an adoption record (i.e., the expression that "user u has adopted innovation a" is symmetric to that "innovation a has reached user u"). An example of such concepts is the peer influence on innovation networks, which imitates the real social influence in social networks, but instead measures how prior adoption history of a specific user affects her later adoption decisions.

To evaluate the predictive value of the innovation networks, we design a real task of predicting a user's adoption decision about a specific innovation. We are especially interested in the contribution of innovation-network-related features on the improvement of performance.

The contribution of this paper can be summarized as follows:

1. We provide a novel perspective of the study of the diffusion of innovations, by investigating how the innovations compete and collaborate with each other. These relationships can be identified through an automated text mining process, which results in large scale networks of innovations.
2. On top of the features from social networks, features extracted from the innovation networks significantly improve the prediction of the adoption of innovations.
3. The study of innovation networks provides new insights on the variance and categorization of adopters, which could not be obtained from previous studies on social influence.

The rest of this paper is organized as follows. We start with a brief introduction of the related work in Section 2. Section 3 elaborates on the process of building networks of innovations and explores various network metrics. Section 4 provides a formal quantitative study of the predictive value of innovation networks, followed by Section 5, which concludes our findings and discusses the implications of our results on the social science research.

2. RELATED WORK

To the best of our knowledge, this is the first study of the network of innovations and its predictive power on the adoption of innovation based on large-scale text data. Similar intuitions of studying the interaction between innovations in diffusion have been seen in recent literature [27, 30, 1, 13]. These works consider competition as the only relationship between the entities that are diffused, and do not address collaboration in the context of innovation networks.

The most related work that addresses both competition and collaboration between items being diffused is presented by Myers and Leskovec [20]. They study the interaction between memes in diffusion through social media by constructing a model that quantifies the degree to which different clusters of memes compete or cooperate with each other. They conclude that stronger (more infectious) memes enhance the diffusion of relevant-weaker memes, but prohibit the diffusion of irrelevant-weaker memes. The relevance is measured by the cosine similarity between the language models of the memes. Although their characterization of competition and cooperation is similar to our setting, their approach relies on the availability of large-scale user adoption history for training the infection model. In contrast, the presented innovation networks are directly extracted from text data in an unsupervised fashion, which captures the interactions between innovations without the requirement of user adoption records.

There is also a body of work that addresses the diffusion of topics or community memberships among scholars through certain kinds of social structure, such as co-authorship networks and citation networks [15, 2, 8, 12, 17]. These studies primarily focus on instantiating and interpreting structural features of the social networks, or developing algorithms that can generate effective and efficient diffusion strategies. Compared to these studies, the proposed framework of innovation networks provides a novel and orthogonal perspective on modeling and interpreting various diffusion processes.

3. ESTABLISHING INNOVATION NETWORKS

In this section, we introduce our approach to establishing innovation networks by analyzing a large collection of text documents. This approach allows us to investigate the attributes of innovations and their inter-relationships separated from the effects of social networks.

3.1 Network Components

Nodes: innovations. We select algorithms in computer science as the nodes of the innovation network. As our target of diffusion analysis, new algorithms are being constantly created by computer scientists every year and spread via multiple communication channels. By regarding citations as indicators of diffusion of innovation, we can create a directed network of users featuring social influence.

We analyze the CiteSeer dataset which was originally used for HCIR 2011 Challenge. The dataset is public on their website. It contains over 800,000 research papers in computer science published by over 2 million authors (among which group there are 36.8 million citation links), with complete meta-data information, including publishing date and

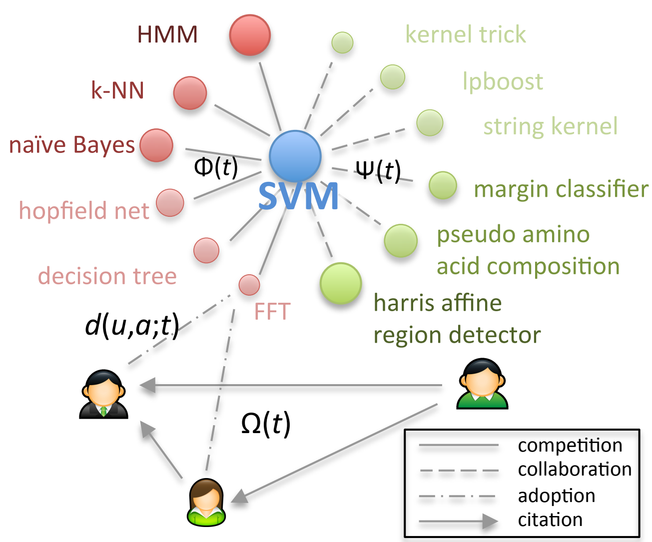

Figure 1: Networks and relationships involved in this paper. The top 6 competitors (left) and collaborators (right) of `support vector machine` extracted from real data are shown. Node sizes correspond to connection strengths. $\Phi(t)$: compepetitor network, $\Psi(t)$: collaborator network, $\Omega(t)$: social network, $d(u, a; t)$: adoption status.

citations. We consider 804,000 research papers in our analysis, which range from the year of 1900 to 2010. We identify the event that a researcher adopts an innovation at the first time she mentions the innovation in one of the papers she authors/coauthors.

To obtain a good list of innovations, we extract computer algorithms from the Wikipedia category lists. It has a relatively complete set of computer algorithms documented in hierarchical categories. We start from three root categories, "algorithms," "statistical models," and "probabilistic models" and extract a total number of 8,500 entities.[1] After deduplication and manual removal of false positives, we identify 1,692 algorithms and models as our final list. Wikipedia redirection links are used to obtain the aliases of the algorithms, so that synonyms can be regarded as one entity.

In sum, by choosing the target of innovation diffusion analysis, we have also identified the community of users (authors in computer science), communication channels (journals and conferences) and the behavior of adoption (writing about the algorithm in one's publications). We believe this setting of diffusion analysis is generalizable to other domains of innovation diffusion.

Edges: competition and collaboration. Social ties are formed partially because of the commonalities between people. Similarly, innovations are bonded to each other due to their similar characteristics. For example, two algorithms are considered to be related, either because they belong to the same category and share similar functions, or because they are often used together. To quantitatively evaluate the relationships among innovations (i.e., entities), we empirically define two types of relationships: competition and collaboration.

Definition 1. If two entities a_i and a_j are *replaceable* in many contexts, define them as *competitors*. A context $c(a_i|d)$ is a part of the sentence d with a certain length L surrounding the position where a_i has appeared. For example, "`quicksort`" and "`merge sort`" share similar contexts in multiple documents, such as "`using the _ algorithm to sort`"; thus they are competitors. Let us denote the network of competing entities as $\Phi(t)$.

Definition 2. If two entities a_i and a_j *co-occur* in multiple sentences $\{d_c\}$, define the two entities as *collaborators*. For example, "`Support Vector Machine`" and "`Kernel tricks`" co-occur in a great number of contexts, because support vector machines can apply kernel tricks; therefore, these two entities are collaborators. Denote the network of collaborating entities as $\Psi(t)$.

The two types of relationships, *competition* and *collaboration*, define two different types of links, and thus two different innovation networks. The subsequent analysis will show that these two networks effectively recover the underlying commonalities and interactions among entities, and contribute to the prediction of the adoption of innovations.

In practice, to extract *competitors*, we extract the contextual words (i.e., neighboring words) of each occurrence for each entity of interest and break them into multiple shingles of 4 to 6 words long. Then we aggregate the shingles for each entity, constituting a context vector ("bag of shingles") and compute the cosine similarity between the context vectors of each pair of entities, and identify those pairs with cosine similarity above a threshold as competitors. This is related to the distributional similarity [16] in the literature of natural language processing, but is defined on bag of shingles instead of bag of words.

To extract *collaborators*, we compute the pointwise mutual information (PMI) [9] between each pair of entities that co-occur for at least once, and identify those pairs with PMI value higher than a threshold as collaborators. Then we use the two types of scores to build two innovation networks for the same set of innovations, and thus obtaining the *competitor network* and the *collaborator network*.

Note that although it may sound absurd, a pair of entities could be identified as *competitors* and *collaborators* at the same time. This is because of the nature of the way we identify such relationships, and the fact that authors do mention *competitors* in the same context occasionally, especially when they make comparisons among a set of alternative algorithms. In the subsequent analysis, we do not perform additional processing for this situation. Although ignoring the existence of competitor-collaborator overlaps may result in undesired correlation between the effects of the competitor network and the collaborator network, such a way of processing will minimize human interference with the data, and as we show later, it will still be capable of revealing the separated effects of the two types of relationships.

Degree distributions. The constructed innovation networks are undirected networks with the density depending on the selected thresholds of similarity scores. Figure 2 illustrates the distribution of network measures for the two types of innovation networks. The weights on the links (i.e.,

[1] We choose these two extra root categories in order to enrich the set of innovations with models such as `latent semantic analysis` and `kernel trick`.

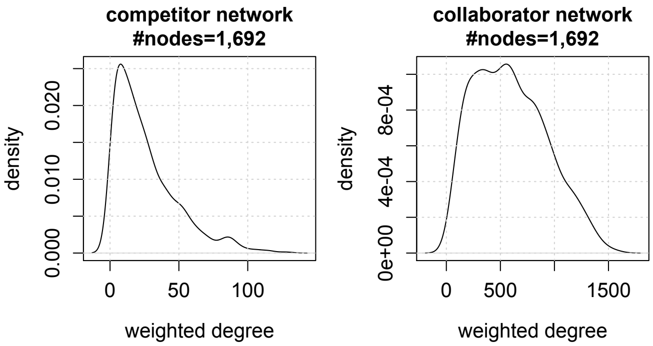

Figure 2: Weighted degree distributions on innovation networks.

the similarity scores between innovations) on a network approximately follows a log-normal distribution. Interestingly, the weighted degrees of both networks follow log-normal distributions, comparing to social networks which are usually featured with a power-law degree distribution.

3.2 Interplay between Innovation Network and Social Network

If we want to model diffusion using innovation networks, we have to ask an important question: how do innovation networks interplay with the social network? This question can be addressed by revisiting two common factors on social networks that are frequently considered in diffusion literature, social influence and prestige [31, 2, 7, 24].

Social Influence. Social influence theories typically posit that the probability of an individual user adopting an innovation increases with the number of friends having adopted the same innovation [24, 26, 29, 2]. What would be the equivalent hypothesis on innovation networks? If we switch the roles of the user and the innovation in the above proposition about social influence, then we obtain the following hypothesis: the probability of an innovation being adopted by a user increases with the number of its neighbors (i.e., its competitors and collaborators) that have been adopted by the same user. We denote this effect as the *innovation influence.*

Prestige. Diffusion studies also suggest that one of the motivations of innovation adoption is to increase the adopter's prestige [24]. Innovations with high likelihood of increasing the user's social prestige are more likely to be adopted than ordinary innovations. A robust measurement of a user's prestige on citation networks is PageRank [22], which evaluates both the number of followers of the unit and the prestige of those followers.[2]

By a similar analogy to the one used in defining social influence, we define the *innovation's prestige* as the weighted degree of the innovation on the innovation network. This definition is consistent with social prestige because innovation networks are undirected and PageRank on undirected networks converges to weighted degree.

In addition to measuring the prestige of users and innovations in their networks respectively, we can also measure the prestige of an innovation by looking at the prestige of its adopters, and measure the prestige of a user by looking at the prestige of her adopted innovations. Here we define *innovation's user prestige* as the average (logarithm of) PageRank score p_u of the users u having adopted this innovation [31], and define *user's innovation prestige* as the average (logarithm of) weighted degree of the innovations that have been adopted by this user. These two additional metrics serve as alternatives to the original prestige definition, and may provide richer information for studying diffusion. To be specific, a higher value of the former metric, *innovation's user prestige*, indicates that the innovation tends to be adopted by more prestigious users, and a higher value of latter metric, *user's innovation prestige*, may indicate that the user has more strict criteria for innovations.

In sum, we extend the concepts of *social influence* and *prestige* from social networks to innovation networks by switching the roles of the *user* and the *innovation* in the concepts. The analogies we have used here between users and innovations are solely for the purpose of finding a way to model the new interactions. The conjugate pairs of metrics, such as *social influence* vs. *innovation influence*, or *social prestige* vs. *innovation prestige*, may possess very distinct meanings. In the next section, we perform drill-down analysis on these features to find out their meanings and their impact on diffusion.

4. THE PREDICTIVE POWER OF INNOVATION NETWORKS

In this section we provide empirical evidence to show that the innovation network is a new and valuable perspective for studying diffusion of innovations. Note that our goal here is by no means to do feature engineering or proposing a new computational algorithm (for which there has been plenty of good effort, such as [2, 8, 12, 15, 17, 27]), but to answer the research questions asked in the first section, which also can be phrased as: how different is this new perspective from the old ones? What are the relations between the two? What new conclusions can be made from this new perspective?

We answer these questions by designing a prediction task regarding the individual user's adoption behavior and performing regression analysis on the new family of features. These features can only be derived from innovation networks. We show that this family of features do provide meaningful interpretations of the diffusion process that could not be obtained before.

4.1 Predicting adoption behavior

Notations. For every year in our data set, t, we construct a snapshot of the innovation networks taken at t (e.g., $t = 2005$) which reflects all the activities from the earliest record of the data set until the end of the year t. Let $t_0(u)$ be the year when a user $u \in U$ publishes her first paper. Let $t_0(a)$ be the year when an entity $a \in A$ is first adopted by any $u \in U$. Let $d(u, a; t)$ be an indicator function of the status of adoption: $d(u, a; t) = 1$ if user u has adopted innovation

[2]In addition to PageRank, a widely-used prestige metric for scholars is the *h-index*, which considers both the number of publications and the number of citations. Recent studies suggest that in real citation networks, *h-index* is not significantly correlated with PageRank or other centrality measures [6, 10]. To preserve generalizability to domains outside of scientometrics, we use the PageRank as the measurement of social prestige for the authors.

Feature Category	Depend.	Feature: Explanation
Basic feature set (non-network-involved)	U-spec.	User's age of activeness: $t_1 - t_0(u)$
		User's "popularity": number of innovations the user has adopted
	I-spec.	Innovation's active years: $t_1 - t_0(a)$
		Innovation's popularity: number of adopters (users that have adopted the innovation)
	UI-dep.	Relevance: cosine similarity between the n-gram models of $D_u(t)$ and $D_a(t)$, n=1, 2, 3
Social-network-involved (Ω)	U-spec.	User's weighted degree, in-degree, out-degree on the social network
		User's social prestige $p_\Omega(u,t)$: user's log-PageRank score on the social network
	I-spec.	Innovation's user prestige $p_{\Omega'}(a,t)$: average log-PageRank score of the adopters
	UI-dep.	Social influence $\eta_\Omega(u,a;t)$: number of user's followees that have adopted the innovation
		Prestige difference on the social network: $p_{\Omega'}(a,t) - p_\Omega(u,t)$
Innovation-network-involved ($I \in \{\Phi, \Psi\}$)	I-spec.	Innovation's weighted degree on the innovation network
		Innovation's prestige $p_I(a,t)$: innovation's weighted degree on the innovation network
	U-spec.	User's innovation prestige $p_{I'}(u,t)$: average weighted degree of innovations the user adopted
	UI-dep.	Innovation influence $\eta_I(u,a;t)$: number of the innovation's neighbors adopted by the user
		Preference $\eta'_I(u,a;t)$: average weight of the innovation's neighbors adopted by the user
		Prestige difference on the innovation network: $p_I(a,t) - p_{I'}(a,t)$

Table 1: Description of features. Each of the features involved with the innovation network are instantiated on both the competitor network and the collaborator network, i.e., $I \in \{\Phi, \Psi\}$.

a by year t and $d(u,a;t) = 0$ otherwise. For the two types of innovation networks, let $\Phi(t)$ be a snapshot of the *competitor network* taken at t, $\Psi(t)$ be the *collaborator network*, and the $\Omega(t)$ be the snapshot of the *social network*. By the year t, denote the collection of documents published by u as $D_u(t)$ and the collection of documents mentioning a as $D_a(t)$.

Task setting. Our task is: at a given year t_1, for a user-innovation pair (u,a) where $u \in U$, $a \in A$, $t_0(u), t_0(a) \leqslant t_1$, $d(u,a;t_1) = 0$, to predict whether u will adopt a within Δt years ($\Delta t > 0$). This is equivalent to estimating the value of $d(u,a;t_1 + \Delta t)$. The prediction is based on the adoption history of all users and all innovations in the past, i.e., $d(\mu,\alpha;t)$ for any $\mu \in U$, $\alpha \in A$, and $t \leqslant t_1$, the snapshots of three networks $\Phi(t_1)$, $\Psi(t_1)$, $\Omega(t_1)$, and all text information available at time t_1, namely $D_u(t_1)$ and $D_a(t_1)$.

We construct two samples for training and testing purposes. The training sample is selected at $t_1 = 1995$ and $\Delta t = 5$. There are 120,411 positive cases, including all (u,a) pairs that satisfy the above constraint plus $d(u,a;t_1+\Delta t) = 1$. The same number of negative cases are randomly sampled from all other (u,a) pairs where $d(u,a;t+\Delta t) = 0$. The test sample is selected at $t_2 = 2000$ and $\Delta t = 5$, which yields 402,911 positive examples and the same number of negative examples.

Following the methods described in Section 3.1, we construct the snapshots of the competitor networks $\Phi(t)$ and collaborator networks $\Psi(t)$ at $t = 1995, 2000$. To minimize human interference, no threshold is used to filter edges in $\Phi(t)$, and the threshold for $\Psi(t)$ is simply set to 0. Accordingly, all the features related to the degree of nodes are weighted by the strengths of the links.

Features. Following the intuitions explained in Section 3.2, we formally instantiate the features of innovation networks now. We also include baseline measurements of the innovation (e.g., popularity) and metrics related to social influence, all presented in Table 1.

All of the features defined in Table 1 have their corresponding concepts in classical diffusion studies. For example, the degree of an innovation in Φ corresponds to the *uniqueness* of the innovation, and the degree in Ψ corresponds to *compatibility*. The prestige difference corresponds to the user's *perceived benefits* of the innovation. These connections are explained in greater details in Section 5.

Among all new features, *innovation influence* is one of the most important. We use three ways to compute it in practice. The first way is just as defined in Table 1:

$$\eta_I(u,a;t) = \sum_{i \in N_I(a)} w_{ia} \cdot d(u,i;t), \qquad (1)$$

which equals the *number* of neighbors of the innovation a that have been adopted by u, where $N_I(a)$ refers to the set of neighboring nodes of a on network I. The measure relates to the standard *threshold model* of the diffusion of information [14, 24]. The second way is to calculate the *proportion* of neighbors, i.e., to normalize Eq.(1) by $|N_I(a)|^{-1}$, which is a variation of the threshold model that is also widely discussed. The third way of calculating the innovation influence is:

$$\eta'_I(u,a;t) = \left(\sum_{i \in N_I(a)} w_{ia} \right)^{-1} \cdot \eta_I(u,a;t). \qquad (2)$$

We denote Eq.(2) as *preference* (see Table 1), because it measures the average closeness (or intimacy) between what user u has adopted to what she is considering adopting. The third metric is conceptually different from the first two metrics.

Note that for either the user or the innovation, there are two types of prestige metrics defined on the innovation networks and the social network respectively, so there are four prestige definitions in total (i.e., $p_\Omega(u,t)$, $p_{\Omega'}(a,t)$, $p_I(a,t)$, $p_{I'}(u,t)$, $I, I' \in \{\Phi, \Psi\}$), which all have different semantics. In addition, two types of *difference of prestige* metrics are defined by subtracting the corresponding user-innovation prestige scores on different networks (i.e., $p_{\Omega'}(a,t) - p_\Omega(u,t)$ and $p_I(a,t) - p_{I'}(a,t)$). They characterize the user's mental process of comparing the innovation with her personal criteria or past adoption history. This metric also has an interpretation related to the theory of *social status*.

Figure 3 compares the distribution of innovation-network-involved features on positive cases and negative cases. All the present features show some level of predictive power, indicating that they may contribute to modeling the user's adoption behavior. In general, a user is more likely to adopt

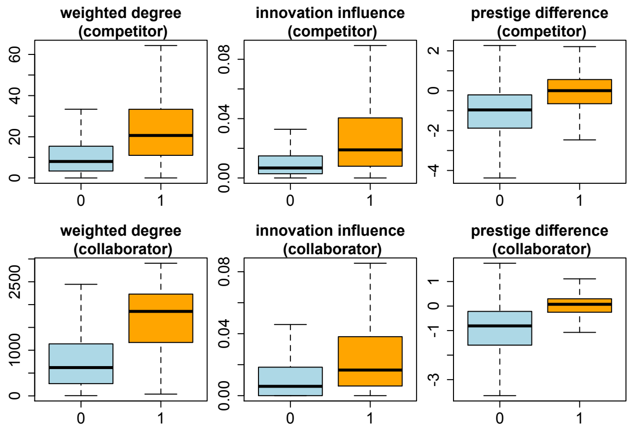

Figure 3: Distribution of measurements on innovation networks on the training set factorized by user's adoption decision. 0: non-adopted, 1: adopted.

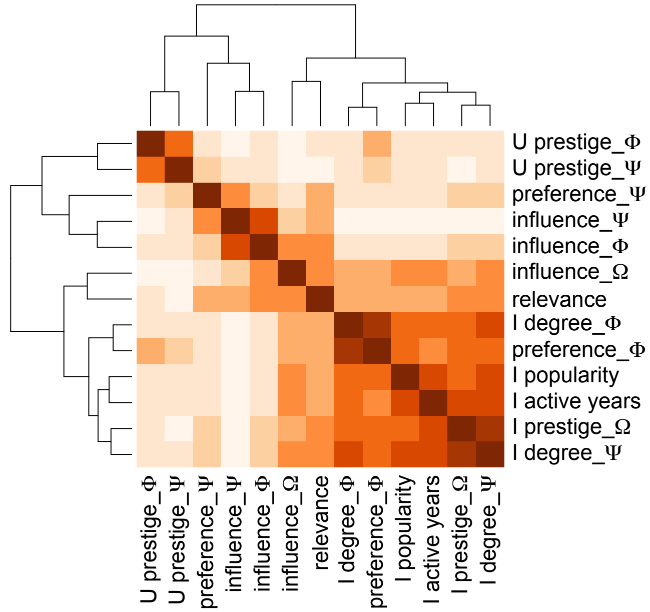

Figure 4: Correlation analysis between selected features. Darkness of cells corresponds to *Person's r* correlation. U: user, I: innovation, Ω: social network, Φ: competitor network, Ψ: collaborator network.

an innovation if the innovation has more competitors or collaborators, or if she has already adopted more competitors or collaborators of the innovation, or the prestige of the innovation is higher than the average prestige of the innovations she adopted. Although, by definition, some of the metrics are correlated with the popularity and active years of innovations to some degree, they still provide extra information for modeling diffusion that may be valuable.

In addition, among the present features, the ones associated with the collaborator network (Ψ) appear to be more distinctive than those associated with the competitor network (Φ), though this distinction is subject to parameterization during the establishment of the networks.

In short, the innovation-network-involved features are informative of predicting the user's adoption behavior.

Correlation analysis. To understand how the features inter-relate and the structure of the feature set, we construct the correlation matrix between the aforementioned three groups of features. A sample of 220,412 cases in which the user-innovation relevance is greater than 0.1 is selected for analysis. *Pearson's r* is reported, and the features are reordered based on hierarchical clustering analysis for illustration. Figure 4 presents a heat map and a dendrogram.

An important observation is that several pairs of features instantiated on the competitor network (Φ) and the collaborator network (Ψ) are correlated, including user's innovation prestige (on Φ vs. on Ψ) and innovation influence, with $r = 0.450$ and 0.771 respectively. This is not surprising because the local structure of the competitor and collaborator networks are highly dependent on the popularity of the innovation. This dependency makes the features that are closely related to the local structure of both Φ and Ψ interrelated. Despite the pairs of correlated features of Φ and Ψ, there is also substantial difference between the two types of innovation networks. For example, the *preference* features on Φ and Ψ present very weak correlations.

From Figure 4, we can also observe that a number of features defined on the innovation networks are not strongly correlated with features defined on the social network. This implies that the features from innovation networks add substantial new information to the features from the social network. We may utilize this new information in improving the performance of prediction in the subsequent study.

Regression analysis. To further evaluate the predictive value of the features, we want to quantify the impact of each individual feature on increasing the probability of adoption. Logistic regression is able to provide an answer. Note that we have observed some features that are inter-correlated, such as the user's popularity and innovation influence. Therefore, to prevent feature collinearity from impairing the interpretation of the regression results, we select multiple subgroups of features by constraining the maximum correlation between any features in each subgroup to be less than 0.3. For each of the subgroup, a logistic regression model is trained accordingly (called "a run"), and the coefficients and the significance levels of the independent variables are recorded for each run. Then we horizontally compare the coefficients for each feature across multiple runs, and we have observed that all features exhibit high significance levels, and the signs and magnitudes of the coefficients remain stable and robust across multiple runs.

Table 2 summarizes the regression analysis. The magnitudes of the coefficients are not comparable because the independent variables are not normalized, but the signs of the coefficients tell a story that is good enough. From Table 2, we can observe that the features defined on the competitor network and the collaborator network have similar effects in predicting adoption decisions. Generally, the probability of adoption increases with the innovation's degree on innovation networks, the user's preference over the innovation, and the innovation influence. While a high value of the user's innovation prestige (i.e., a high criterion for adoption) will result in a low probability of adoption, enlarging the difference between the innovation's prestige and the user's prestige criterion (i.e., prestige difference Φ, Ψ) will increase the probability of adoption. The result shown in Table 2 is promising. It indicates that the features we extract are all strong predictors of the user's adoption behavior and the effects are significant.

The results of baseline features and social-network-related features are also shown in Table 2. The results match general intuition. We can observe that the probability of adoption increases with user's popularity (number of innovations she has adopted), innovation's popularity, relevance between the user and innovation, and social influence. Interestingly, the coefficients of prestige difference on the social network are negative. This could be explained as an effect of *regression to the mean*—if the user has already adopted a number of innovations that are highly prestigious, then the next innovation she adopts is more likely to be less prestigious, which explains the negative coefficient.

Category	Feature	Ave. Coefficient
Baseline	User's popularity	5.3E-05 ***
	Inno's popularity	2.5E-03 ***
	Relevance	6.5E-01 ***
Social Net Ω	Inno's user prstg Ω	1.7E-01 ***
	User's social infl Ω	1.7E-03 ***
	Prestige diff Ω	-6.5E-02 ***
Inno Nets Φ, Ψ	Innovation's deg Φ	2.7E-03 ***
	Innovation's deg Ψ	2.5E-04 ***
	User's inno prstg Φ	-8.4E-03 **
	User's inno prstg Ψ	-4.8E-02 ***
	Preference Φ	2.6E+00 ***
	Preference Ψ	4.3E-02 ***
	Inno influence Φ	2.0E+00 ***
	Inno influence Ψ	1.6E+00 ***
	Prestige diff Φ	1.3E-01 ***
	Prestige diff Ψ	1.9E-01 ***

Significant at the: *** 0.01, ** 0.05, * 0.1 level.

Table 2: Logistic regression on adoption decisions. For each feature, the average coefficient and the minimum significance level of multiple runs are reported. The signs of the coefficients of all features remain stable across multiple runs.

Classification results. To further illustrate the predictive power of innovation networks, we compare the classification performance of different feature combinations on the test set (see task setting), although feature engineering is not our major goal here. Our task is still to predict whether a user will adopt an innovation within five years. Table 3 summarizes the result. To investigate how the integration of new features improves the performance, we separate a subsample NEW of the test set where $t_1 - t_0(a) <= 1$, which means these innovations are no more than 1 year old. The NEW set includes 8,018 cases. Conversely, the OLD set includes all innovations that satisfy $t_1 - t_0(a) > 10$, including 784,365 cases. All models are trained on the same training set. From Table 3, it can be seen that innovation-network-involved features can improve the performance beyond social-network-involved features. This improvement is especially significant on the NEW sample where less information about the innovation is available. This implies that while the social network is sufficient for predicting the adoption of older innovations, it is insufficient in predicting "new" innovations. In each innovation set, we also examine the performance of the features related to the competitor network (Φ) and the collaborator network (Ψ) separately. The results show that the collaborator network contributes the most to the improvement of the performance. Utilizing more sophisticated machine learning methods, such as the Support Vector Machines, will very likely further improve the prediction accuracy, which, however, is not our intent here.

Inno Set	Feature Set	$F_{0.5}$ %	Precision %	Recall %
ALL	Basic features (B)	81.88	80.25	89.11
	B+Ω	84.12	82.46	91.50
	B+$\Omega + \Phi$	83.47	81.52	**92.31**
	B+$\Omega + \Psi$	**86.49**	**86.02**	88.43
	B+$\Omega + \Phi + \Psi$	86.44	85.98	88.31
	B+$\Phi + \Psi$	81.02	78.90	90.80
NEW	B	36.48	36.23	**37.48**
	B+Ω	41.34	44.10	33.06
	B+$\Omega + \Phi$	40.95	43.34	33.55
	B+$\Omega + \Psi$	**51.43**	**58.81**	34.24
	B+$\Omega + \Phi + \Psi$	50.84	57.65	34.53
	B+$\Phi + \Psi$	27.37	30.07	20.13
OLD	B	82.93	81.15	90.90
	B+Ω	85.01	83.18	93.20
	B+$\Omega + \Phi$	84.39	82.29	**93.96**
	B+$\Omega + \Psi$	**87.20**	**86.48**	90.24
	B+$\Omega + \Phi + \Psi$	87.17	**86.48**	90.05
	B+$\Phi + \Psi$	81.98	79.66	92.80

Table 3: Prediction accuracy on the held-out dataset and sub-datasets with different combinations of features. Innovation networks are especially useful to predict the adoption of NEW innovations. NEW includes only innovations whose active years $\leqslant 1$. OLD includes innovations whose active years > 10.

In sum, our major finding in this section is that the features derived from the innovation networks are strong and robust predictors of users' adoption decision, and they significantly improve the prediction on top of all features extracted from the social network.

4.2 Effects of Adopter Variance

A very interesting question to ask about the above regression study is: do adopters (i.e., users) that possess different characteristics and communication styles exhibit different adoption patterns? Diffusion studies give an answer of "yes" to the question by confirming the distinction between the roles played by different subgroups of population (e.g., early adopters, late adopters) [24]. In this section, we are interested in what new interpretations of this distinction can be brought by looking at the diffusion problem from the perspective of innovation networks.

Classifying adopters. To study the effects of the variance of adopters, we first need to define the variance, which, in this case, is equivalent to finding a way to classify the users into different categories based on their earliness or eagerness of adoption. In the existing diffusion literature, adopter category is defined in the context of diffusion of a single innovation. For example, Rogers [24] uses the four quantiles of earliness of adoption to categorize adopters into five categories: *innovators*, *early adopters*, *early majority*, *late majority*, and *laggards*. To make full use of our dataset, we need to make an extension to Rogers's adopter categorization to account for a user's behavior in the diffusion of multiple innovations concurrently.

Intuitively, to obtain a user's general eagerness of adoption, we can evaluate the eagerness of the user in the diffusion of each individual innovation in which she is involved, and combine the eagerness by taking the average. For a given time t, a user u, and an innovation of that time $a \in A^t$, we characterize the eagerness of u in the diffusion of a using the position of u among all users who have adopted a ranked by the order of adoption. This value is denoted as the *Earliness Index*, given by

$$EI(u,a,t) = \frac{1}{|U_a^t|} \sum_{u' \in U_a^t} d(u', a; t(u;a) - 1), \tag{3}$$

where U_a^t is the set of users that have adopted a by time t, and $t(u,a)$ is the time when u adopts innovation a. $|U_a^t|$ is the number of users in the set U_a^t. As before, $d(u,a;t)$ means that u has adopted a by time t. A lower value of EI indicates that u is more eager in adopting innovation a compared to the other users. Taking the average of EI for all innovations $a \in A^t$, we obtain the general eagerness of user u, given by

$$AEI(u,t) = \frac{1}{|A^t|} \sum_{a \in A^t} EI(u,a,t). \tag{4}$$

For any given u, $AEI(u,t) \in [0,1)$. A lower value of AEI implies that the user generally tends to adopt any innovation at an earlier stage of the innovation's diffusion process.

Adopter Category	Theoretical Parition	*AEI* Range
Innovators	$(-\infty, \mu - 2\sigma)$	$[0, 0.027)$
Early Adopters	$[\mu - 2\sigma, \mu - \sigma)$	$[0.027, 0.177)$
Early Majority	$[\mu - \sigma, \mu)$	$[0.177, 0.430)$
Late Majority	$[\mu, \mu + \sigma)$	$[0.430, 0.686)$
Laggards	$[\mu + \sigma, \infty)$	$[0.686, 1)$

Table 4: Categorization of users using quantiles of *Average Earliness Index.*

We calculate the AEI scores of 46,069 unique users based on their prior adoption history prior in the training set (see Section 4.1). The AEI approximates a normal distribution $N(\mu, \sigma^2)$, with mean $\mu = 0.421$, and standard deviation $\sigma = 0.216$. Following Rogers [24], we categorize all users into five adopter categories according to the four quantiles of AEI distribution (see Table 4).

Hypotheses and experiment design. Classical diffusion models [3, 24] suggest that early adopters are less sensitive to social influence than late adopters, and are more influenced by other communication channels, such as mass media. If this argument is true, then we should observe the same trend in our data: compared to conservative researchers, scholars that are more enthusiastic in experimenting with new ideas should keep their eyes more open to a wider variety of sources of information besides their followed researchers, and thus they should be less subject to social influence, and more influenced by other factors.

To verify this intuition, we formulate the following hypotheses for testing:

(1) Innovators and early adopters are less influenced by *social influence* than majorities.

(2) Innovators and early adopters are more influenced by *innovation influence* (including competitor influence and collaborator influence) than majorities.

While the first hypothesis is supported by classical diffusion studies, the second hypothesis is new and can only be tested by constructing innovation networks. To test the hypotheses, we build multiple logistic regression models, each of which includes only one independent variable out of *social influence*, *competitor influence*, and *collaborator influence*, and is trained on each individual category of adopters separately. Having trained all the regression models, we then examine the coefficients of the independent variables. For each independent variable and each adopter category, a higher coefficient means that a unit increase in the value of the independent variable will cause a more significant increase in the probability that the user adopts the innovation; therefore, horizontal comparison of the coefficients across different adopter categories is informative for investigating the distinction between different groups of users.

Results. Figure 5 summarizes the results of the above experiment. It compares the impact of four distinct features, including social influence[3], popularity of the item, innovation influence on the competitor network, and innovation influence on the collaborator network, on different user categories.

Note that the upper right chart in Figure 5 looks specially at the variable of innovation popularity, which is the number of users that adopted an innovation. This variable is one of the basic features, neither associated with social influence nor innovation influence. We include this feature in the experiment to confirm that our setting of the AEI metric and the computation are correct. Since innovators and early adopters often adopt innovations at the early stage, a unit increase in the popularity of the innovation should not increase the probability of their adoption as much as for the majorities. The distribution of the coefficients in the result matches our expectation, and thus confirms our setting and computation.

The upper left chart in Figure 5 is associated with the first hypothesis. We can observe a very clear contrast between different adopter categories regarding *social influence*. Innovators and early adopters are significantly less prone to social influence than early and late majorities are. In addition, between innovators and early adopters, the former are less affected than the latter. This result supports the first hypothesis, and hence matches the conclusions of the existing diffusion literature [3, 24].

In contrast to social influence, innovation influence (the bottom charts in Figure 5) shows the opposite trend. The decisions made by innovators and early adopters appear to be much more influenced by their prior adoption history of similar items than early and late majorities. This result supports the second hypothesis above, and implies that the first two categories of adopters, compared to the majorities, are more interested in the properties of the innovation itself rather than their peer opinions. In addition, the difference

[3]Here social influence is computed as the *proportion* of friends that have adopted the innovation, instead of the *number* of friends. The latter one is tricky due to collinearity effect.

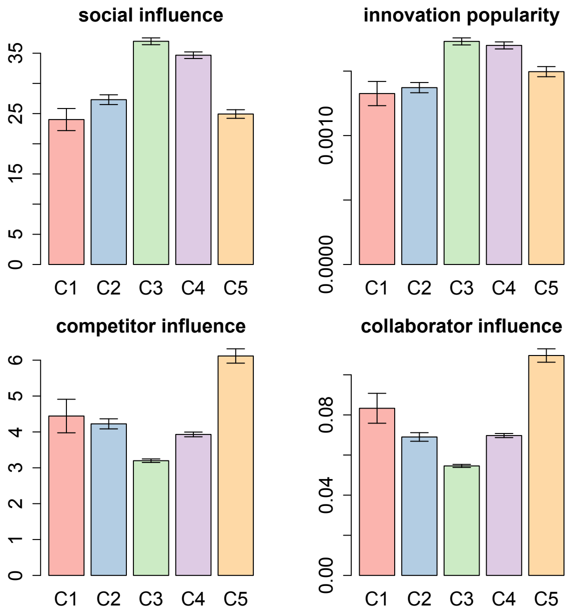

Figure 5: Comparison of predictor coefficients across adopter categories with 90% confidence interval. X-axis: adopter categories. C1: Innovators, C2: Early adopters, C3: Early majority, C4: Late majority, C5: Laggards. Y-axis: coefficient of the predictor in the logistic regression model.

between innovators and early adopters regarding *collaborator influence* is more significant than that regarding *competitor influence*.

It is noteworthy that in all the comparisons, the confidence intervals of the coefficients for the innovators are the largest. This is because the way we partition the data leads to fewer cases for the innovator category than for other categories. In fact, the behavior of innovators are truly the least predictable given the small number of them in society and the wide variety of reasons why they are innovative. In addition, we ignore *laggards* in our analysis because they are the least important in the diffusion process.

5. DISCUSSION AND CONCLUSION

We summarize the four most important observations from the series of analyses above. First, the adoption rate of an innovation will increase if it has either more competitors or more collaborators. This is intuitive because having more competitors and collaborators is likely to increase the exposure of an innovation.

Second, the adoption rate of an innovation increases with the proportion of its competitors or collaborators adopted by the user. Among the users of different adoption styles, the innovators and early adopters are more sensitive to such an influence. Between innovators and early adopters, the former is more likely to be influenced by the collaborators of an innovation than its competitors. (Figure 5). When the diffusion of an innovation become prevalent, the social influence becomes the driven force.

Third, the prestige measurements on innovation networks have stable negative coefficients (Table 2), indicating that the users with higher standards of selecting innovations are less likely to adopt an innovation. This implies a way of defining users' individual thresholds in modeling diffusion. In addition, the difference of prestige on innovation networks characterize how much the innovation supersede the user's standard, and is very discriminative in predicting whether the user will adopt the innovation (Figure 3).

Finally, the features instantiated on innovation networks have a strong predictive power of the adoption of innovation, even when combined with the baseline and social features. In particular, such improvement in performance are more significant for cases where the innovations have just started to diffuse and there exists relatively little social influence with regard to the fresh innovation.

Connection with the social science literature. Classical diffusion studies suggest that the diffusion of innovations is impacted jointly by multiple factors, which include the attributes of the innovation, types of the communication channel, and the social network [24, 21, 25]. The network of innovation can be interpreted as a special attribute of innovations, among which, according to diffusion literature, the five most important ones are: uniqueness, compatibility with the user's past experience, perceived benefits, visibility in the social network, and the cost of adoption.

Most of these characteristics are captured and modeled by our proposed innovation networks in a principled way. Among the features we have instantiated, the degree of the innovation in the competitor network corresponds to the *uniqueness*; the degree of the innovation in the collaborator network and innovation influence corresponds to *compatibility*, the prestige of innovations and the prestige difference correspond to the user's *perceived benefits*; the popularity of the innovation and the social influence correspond to the global and local *visibility* of the innovation respectively.

Our results strengthen a series of conclusions made by social scientists with the evidence from large scale empirical analysis. First, through regression analysis we have observed that uniqueness, compatibility, perceived benefits and social influence are all strong and robust predictors of the user's adoption decision, which is consistent with many analyses in diffusion literature [24, 25, 5]. Second, in the analysis of the adopter variance, we have seen a distinct contrast between the effect of social influence and that of innovation influence on early adopters and late adopters, which is also perfectly compatible with existing theories [3, 24]. The matching between our results and the social theories is a very meaningful one, because it provides a plausible way for social scientists to access large-scale text data and make a number of observations that were previously impossible, such as quantitative measurements of the compatibility and the perceived benefits of innovations.

New implications. We have also discovered new interesting results that implies new directions of research. In the adopter variance analysis, the results show that past adoption experience with the collaborators of an innovation exerts more influence on innovators (the earliest category) than on early adopters (the second earliest category). However, with regard to the competitors of the innovation, such a distinction is no longer significant. A possible explanation is that innovators, often being the inventors of the innovations, prefer to devote efforts to the innovations that are more compatible (with more collaborators) than those that are more exclusive (with fewer competitors).

However, as shown in Figure 5, due to the small sample size of innovators , the estimations of the attributes of innovators are uncertain. Given the importance of innovators in the diffusion process and the rare number of them in society, modeling the behavior of innovators remains challenging.

Limitations and future directions. The current work has the following limitations: (1) the construction of innovation networks relies on automated text mining of documents in a specific domain, and thus the conclusions may suffer from lack of generalizability and selection bias introduced by the selected metrics of competition and collaboration, and the corresponding thresholds; (2) the presented analysis is purely empirical. In future work, it will be meaningful to build a statistical model that abstracts the empirical observations and provides a unified theory of diffusion based on the innovation networks; (3) some groups of features are correlated, possibly leading to multicollinearity. In the series of regression analysis, the significance of a feature in each run does not necessarily imply significance in the complete model. Meanwhile, this strong correlation between the features may imply that they correspond to the same set of underlying metrics. In the future, it will be interesting to explore advanced topics such as the evolution patterns of the communities on innovation networks, or the impact of local network structure on the process of diffusion.

6. ACKNOWLEDGMENTS

This work is partially supported by the National Science Foundation under grant numbers IIS-0968489, IIS-1054199, and CCF-1048168, and partially supported by the DARPA under award number W911NF-12-1-0037. The authors want to thank Eytan Adar for meaningful discussions and Yue Wang for help with CiteSeer and Wikipedia data.

7. REFERENCES

[1] N. Alon, M. Feldman, A. D. Procaccia, and M. Tennenholtz. A note on competitive diffusion through social networks. *Information Processing Letters*, 110(6):221–225, 2010.

[2] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *SIGKDD*, pages 44–54. ACM, 2006.

[3] F. M. Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969.

[4] F. M. Bass. Comments on "a new product growth for model consumer durables the bass model". *Management science*, 50(12 supplement):1833–1840, 2004.

[5] C. Bauckhage, K. Kersting, and F. Hadiji. Mathematical models of fads explain the temporal dynamics of internet memes. 2013.

[6] J. Bollen, M. A. Rodriguez, and H. V. de Sompel. Journal status. *Scientometrics*, 69(3):669–687, 2006.

[7] H.-C. Chang. A new perspective on twitter hashtag use: Diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–4, 2010.

[8] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *SIGKDD*, pages 1029–1038. ACM, 2010.

[9] K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

[10] Y. Ding, E. Yan, A. Frazho, and J. Caverlee. Pagerank for ranking authors in co-citation networks. 2010.

[11] L. A. Fourt and J. W. Woodlock. Early prediction of market success for new grocery products. *The Journal of Marketing*, pages 31–38, 1960.

[12] A. Goyal, F. Bonchi, and L. V. Lakshmanan. A data-based approach to social influence maximization. *Proceedings of the VLDB Endowment*, 5(1):73–84, 2011.

[13] S. Goyal and M. Kearns. Competitive contagion in networks. In *Proceedings of the 44th symposium on Theory of Computing*, pages 759–774. ACM, 2012.

[14] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.

[15] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *SIGKDD*, pages 137–146. ACM, 2003.

[16] L. Lee. Measures of distributional similarity. In *ACL*, pages 25–32. Association for Computational Linguistics, 1999.

[17] C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky. The joint inference of topic diffusion and evolution in social communities. In *ICDM*, pages 378–387. IEEE, 2011.

[18] V. Mahajan, E. Muller, and F. M. Bass. New product diffusion models in marketing: A review and directions for research. *The Journal of Marketing*, pages 1–26, 1990.

[19] N. Meade and T. Islam. Modelling and forecasting the diffusion of innovation–a 25-year review. *International Journal of Forecasting*, 22(3):519–545, 2006.

[20] S. A. Myers and J. Leskovec. Clash of the contagions: Cooperation and competition in information diffusion. In *ICDM*, pages 539–548. IEEE, 2012.

[21] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[22] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[23] E. M. Rogers. New product adoption and diffusion. *Journal of Consumer Research*, pages 290–301, 1976.

[24] E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2003.

[25] R. Salickaitė and J. Banytė. Successful diffusion and adoption of innovation as a means to increase competitiveness of enterprises. *Engineering economics*, (1 (56)):48, 2008.

[26] D. Strang and S. A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, pages 265–290, 1998.

[27] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *SIGKDD*, pages 807–816. ACM, 2009.

[28] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[29] T. W. Valente. Network models of the diffusion of innovations. *Computational & Mathematical Organization Theory*, 2(2):163–164, 1996.

[30] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2, 2012.

[31] L. Yang, T. Sun, M. Zhang, and Q. Mei. We know what@ you# tag: does the dual role affect hashtag adoption? In *WWW*, pages 261–270. ACM, 2012.

[32] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su. Understanding retweeting behaviors in social networks. In *CIKM*, pages 1633–1636. ACM, 2010.