Noname manuscript No.
(will be inserted by the editor)

Computational Models of Technology Adoption at the Workplace

Charalampos Chelmis · Ajitesh Srivastava · Viktor K. Prasanna

Received: date / Accepted: date

Abstract Popular social networking sites have revolutionized the way people interact on the Web, enabling rapid information dissemination and search. In an enterprise, understanding how information flows within and between organizational levels and business units is of great importance. Despite numerous studies in information diffusion in online social networks, little is known about factors that affect the dynamics of technological adoption at the workplace. Here, we address this problem, by examining the impact of organizational hierarchy in adopting new technologies in the enterprise. Our study suggests that middle-level managers are more successful in influencing employees into adopting a new microblogging service. Further, we reveal two distinct patterns of peer pressure, based on which employees are not only more likely to adopt the service, but the rate at which they do so quickens as the popularity of the new technology increases. We integrate our findings into two intuitive, realistic agent-based computational models that capture the dynamics of adoption at both microscopic and macroscopic levels. We evaluate our models in a real-world dataset we collected from a multinational Fortune 500 company. Prediction results show that our models provide great improvements over commonly used diffusion models. Our findings provide significant insights to managers seeking to realize the dynamics of adoption of new technologies in their company, and could assist in designing better strategies for rapid and efficient technology adoption and information dissemination at the workplace.

Keywords Agent based computational models · Adoption dynamics · Diffusion of innovations · Diffusion models · Dynamic systems · Evolutionary models · Influence · Technology adoption · Social networks

C. Chelmis
Department of Computer Science, University of Southern California, Los Angeles, USA
Tel.: +1213-740-9130
E-mail: chelmis@usc.edu

A. Srivastava
Department of Computer Science, University of Southern California, Los Angeles, USA

V. Prasanna
Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, USA

1 Introduction

The importance of social networks on information spread has been well studied [28, 11, 6, 4], emphasizing particularly on information dissemination. Traditionally, diffusion and cascading behavior have been formalized as transmission of infectious agents in a population, where each individual is either infected or susceptible, and infected nodes spread the contagion along the edges of the network. There are, however, differences between the way information flows, and the spread of viruses. While virus transmission is an indiscriminate process, information transmission is a selective process. Information is passed by its host only to individuals the host thinks would be interested in it. Diffusion models heavily rely on the premise that contagion propagates over an implicit network, the structure of which is assumed to be sufficient to explain the observed behavior. However, the structure of the underlying network has to be learned [11] from a plethora of historical evidence, i.e. cascades. Although diffusion theory brings up the importance of friendship relations, adoption behavior is instead examined on the premises of the behavior of the entire population [6].

In online social networks in particular, where individuals tend to organize into groups based on their common activities and interests, it has been hypothesized that the network structure (friendship or interaction) affects the way information spreads, and that adoption quickens as the number of adopting friends increases [3]. However, many times a node activation is not just a function of the social network but also depends on many other factors like imitation [28]. This has lead to the development of epidemiology models [14] and computational approaches that are based on thresholds models [12], deterministic or stochastic [24]. Each agent has a threshold that, when exceeded, leads the agent to adopt an activity. When the threshold is applied within a local neighborhood [25, 23], local models emerge [17]. Instead, global diffusion models perform thresholding to the whole population [6].

Unlike online social networks where users create links to others who are similar to them (a phenomenon known as homophily [20]), or whose contributions they find interesting [27, 19], in a corporate environment, employees form "bonds" not because of similar "tastes" but due to tasks at hand (i.e. a function to be completed or an organizational need) or because of reporting-to relationships (i.e. team members reporting to their supervisor). In this sense, there is no explicit "social network", however, formal structures such us the organizational hierarchy may provide hints of influence at the workplace. As illustrated in Figure 1, the formal organization structure may constrain influence patterns, but informal communication outside the boundaries and restrictions of this formal "backbone" may also affect how users behave and ultimately how the diffusion network changes and grows.

The dynamics of information diffusion on a corporate environment are yet unknown and may be entirely different from online social networks. The interplay between formal structure and information propagation at the enterprise has been recently examined [26]. The authors found that social and organizational structure significantly impacts the spreading process of emails, while at the same time indicating context independence. In our study, on the contrary, we do not know the chain of infections, i.e. we do not observe who influences whom (i.e., middle layer in Figure 1). Instead, we empirically quantify the role of reporting-to relationships and local behavior (teammates), as well as the effect of global influence (overall popularity) in the spread of technology adoption at the workplace. Prior work that quantifies influential users within a social network includes [10]. Influence models typically do not take the topology of the network into account, and when they do, they make assumptions about the details of the underlying dynamic process taking place on the network. In our empirical study, we characterize individual dynamics and influence, and examine the spread of adoption through the formal organizational hierarchy.

Contrary to online social networks, microblogging services for enterprises are primarily designed to improve intra-

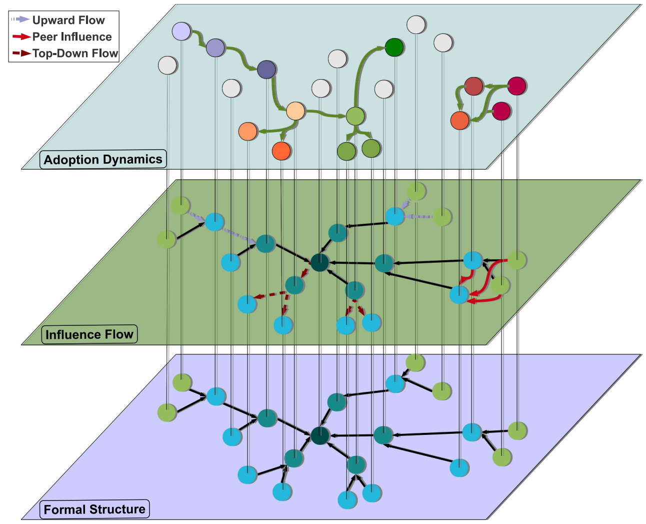


Fig. 1 Technology adoption dynamics at the workplace. Dynamics on and of the formal network structure are strongly coupled. The bottom layer illustrates the formal organization hierarchy, where black arrows represent "reporting-to" relationships between employees. The directionality of edges go from lower level employees up to the company CEO. The middle layer depicts the flow of influence between people in the same group (red arrows), top-down influence from supervisors to team members (dashed, dark red arrows) and vice versa, bottom up team members' influence on their supervisors (dashed purple arrows). The upper layer, depicts observed adoption dynamics, i.e., a potential propagation tree.

firm transparency and knowledge sharing. However, the adoption of such collaborative environments presents certain challenges to enterprises [13]. [30] provided a case study on the perceived benefits of corporate microblogging and barriers to adoption. Key factors influencing microblogging systems adoption in the workplace include: privacy concerns, communication benefits, perceptions regarding signal-to-noise ratio, and codification effort, reputation, expected relationships, and collaborative norms [13]. The work, closest to ours, [26] examined email threads and the formal network (e.g. hierarchical structure) imposed by a large technology firm. They argued that the spreading process (to whom and how fast people forward information) can be well captured by a simple stochastic branching model. In our study, on the contrary, we do not know the chain of infections (i.e. we do not observe who influences whom). Instead, we use the outcome of our empirical study to quantify influence as a result of individual pressure from supervisors towards their team members, as well as an effect of global popularity.

To characterize the adoption mechanism of new technologies at the workplace, we propose two simple and intuitive agent-based computational models with the least possible number of parameters. We emphasize on accurately modeling the cumulative number of adoptions over time, rather than trying to predict which node in the network will infect which other nodes. In this sense, we not only model the influence each node has on the diffusion (*microscopic* *modeling*), permitting user behavior to vary according to the behavior of the general crowd, but we also provide a simple mechanism by which adoption rate rises and decays over time (*macroscopic dynamics*). For our study, we have acquired the organizational hierarchy of a Fortune 500 multinational company. In addition, we gathered adoption logs of the internal microblogging service, which resembles Twitter, during the first two years of adoption of the service in the enterprise. This dataset allows us to empirically characterize individual dynamics and influence, and examine the spread of adoption through the hierarchy. The company did not officially initiate usage of the microblogging service. Rather, it was independently initiated by an employee, in the begging of July, 2010. It was not promoted or even mentioned in any formal corporate communications. Our dataset does not contain information with respect to growth and invitations. We can only speculate that growth was achieved through email and word of mouth invitations.

The rest of the paper is organized as follows. Sect. 2 provides an overview of the most relevant related work which has been undertaken in this area. We describe our dataset in Sect. 3. We study the impact of hierarchical structure on the way adoption spreads in Sect. 5, and we examine employees behavior with respect to overall popularity of the microblogging service in Sect. 6. To capture the macroscopic, temporal dynamics of adoption at the workplace, we propose two novel models that effectively model user behavior with respect to the entire population and individual influence in Sect. 4. In Sect. 7, we provide extended social simulation results of our agent-based computational models of adoption at the workplace. Finally, we discuss the findings of our work and draw our conclusions in Sect. 8.

2 Related Work

The importance of social networks in information dissemination has been thoroughly investigated [11, 4, 6]. In online social networks in particular, where individuals tend to organize into groups based on their common activities and interests (a phenomenon known as homophily [20]), it has been hypothesized that the network structure (friendship or interaction) affects the way information spreads, and that adoption quickens as the number of adopting friends increases [3]. However, many times a node activation is not just a function of the social network but also depends on many other factors like imitation [28]. This has lead to the development of epidemiology models [14] and computational approaches that are based on thresholds models [12], deterministic or stochastic [24]. Each agent has a threshold that, when exceeded, leads the agent to adopt an activity. When the threshold is applied within a local neighborhood [25, 23], local models emerge [17]. Instead, global diffusion models perform thresholding to the whole population [6].

Diffusion models heavily rely on the premise that contagion propagates over an implicit network, which has to be learned from a plethora of historical evidence, i.e. cascades. User characteristics such as topical or latent interests have to be considered in user-to-user content transfers, whereas users' homophily shapes the structure of the network through which information flows. In a corporate environment, employees form "bonds" not because of similar "tastes" but due to a task at hand (i.e. a function to be completed or an organizational need) or because of reporting-to relationships (i.e. team members reporting to their supervisor). [11] examined the problem of inferring the unobserved directed network over which cascades propagate in online social networks. Unlike their approach, which requires traces of numerous different explicit cascades to be given as inputs, we solely rely on one *implicit* sample to infer influence between employees at the workplace. In fact, many influence models have been proposed to rank actors within a social network [10]. However, the underlying dynamic process occurring on the network may not be applicable to the organizational hierarchy. Influence models typically do not take the topology of the network into account, and when they do, they make assumptions about the details of the underlying dynamic process tacking place on the network. In our empirical study, we characterize individual dynamics and influence, and examine the spread of adoption through the formal organizational hierarchy.

Even though most prior work has mainly focused on publicly available online social networks, microblogging capabilities have penetrated the enterprise as well [30]. Contrary to online social networks, microblogging services for enterprises are primarily designed to improve intra-firm transparency and knowledge sharing. However, the adoption of such collaborative environments presents certain challenges to enterprises [13]. [30] provided a case study on the perceived benefits of corporate microblogging and barriers to adoption. Key factors influencing microblogging systems adoption in the workplace include: privacy concerns, communication benefits, perceptions regarding signal-to-noise ratio, and codification effort, reputation, expected relationships, and collaborative norms [13]. The work, closest to ours, [26] examined email threads and the formal network (e.g. hierarchical structure) imposed by a large technology firm. They argued that the spreading process (to whom and how fast people forward information) can be well captured by a simple stochastic branching model. In our study, on the contrary, we do not know the chain of infections (i.e. we do not observe who influences whom). Instead, we use the outcome of our empirical study to quantify influence as a result of individual pressure from supervisors towards their team members, as well as an effect of global popularity.

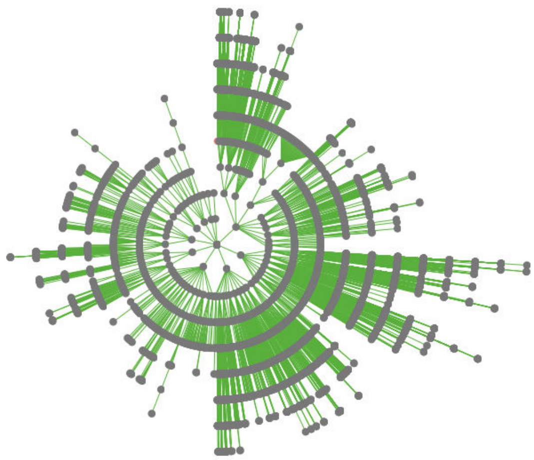


Fig. 2 The organization hierarchy of the company we consider in our experiments presents a tree like structure (CEO, the root of the tree, is at the center).

3 Dataset

The company we studied is a Fortune 500 multinational company, which operates outside the IT-sector. Our dataset consists of a snapshot of the organizational hierarchy, containing over 12000 employees. Figure 2 shows that the organization hierarchy of the company has a tree like structure. Our dataset further contains employees' join logs during the first two years of adoption of a microblogging service from the enterprise (July 2, 2010 to March 22, 2012). During this time period, the number of employees who join the service increases dramatically. Even though, not all employees have joined the microblogging service by the time we obtained the raw data for this paper, a broad spectrum of employees (9,421 users) had joined the microblogging service (77.35% of hierarchy dataset), sharing 19,371 status updates and exchanging 20,370 replies [8]. The functionality of the microblogging service resembles that of Twitter, imposing no restrictions on the way people interact or who they chose to follow. As in Twitter, users author messages in the enterprise microblogging service, and form threaded discussions. The main purpose of the corporate microblogging service is to promote and enable collaboration and sharing within the enterprise. The ultimate goal of the corporate microblogging service is to become the primary platform for asynchronous collaboration and colleagues' communication.

The company did not officially initiate usage of the microblogging service. Rather, it was independently initiated by an employee, in the begining of July, 2010. It was not promoted or even mentioned in any formal corporate communications. Our dataset does not contain information with respect to growth and invitations. We can only speculate that growth was achieved through email and word of mouth invitations. More details on the topological properties of the corporate microblogging service, its dynamics and characteristics, and the interplay between its social and topical components, users' homophily and activity, as well as latent topical similarity and link probability can be found at [8].

4 Modeling Technology Adoption at the Workplace

What is the underlying hidden process that drives adoption of new technologies at the workplace? Our goal here is to find a generative model that generates the observed adoption process of the new microblogging service at the enterprise we are studying, given the organizational hierarchy. We aim for simple and intuitive modeling with the least possible number of parameters. Even though our model is applicable to other social datasets which may exhibit different types of enterprise hierarchies, we restrict our discussion to the setting of microblogging adoption at the workplace, were we track employees joining the service over a period of time.

Prior work on modeling complex networks in social, biological and technological domains has focused on replicating one or more aggregate characteristics of real world networks [21]. Here, we take a different approach. Instead of having a target network to generate, we let individual influence and peer pressure dynamics determine the diffusion process of adoption of the new microblogging service over the formal organization hierarchy. We formally introduce two models that account for influence effects imposed by the formal organizational structure. We compare our results to the true epidemic and we show that the estimates produced by our models are consistent with the real observations (See Section 7).

Model Formulation. The underlying process of influencing employees towards adopting the microblogging service is unknown and non trivial. Here, we assume that when an employee chooses to join the corporate microbloging service, she then has some influence on the employees who directly report to her, according to the formal organizational chart (as shown in Figure 1). Some of these employees will choose to join, which will in turn influence some of their team members into joining themselves and so on. Therefore, we assume that an employee's decision to join depends on: 1) direct influence by her manager, and 2) social influence resulting from the overall popularity of the microblogging service in the enterprise. Here, we assume that employees are not susceptible to peer influence by their teammates (i.e., we assume independence between teammates' choices). We revisit this hypothesis in (Section 6.1).

We study the problem of progressive diffusion, where the employees who adopt the microblogging service become "infected" and do not become "healthy" again (i.e. employees do not unsubsribe the service once they join). As time progresses, employees become "infected" when they adopt (join) the microblogging service. We only observe the time t_u when a particular employee joins the microblogging service. We posit that manager u urges her team members to join the microblogging service. A directed link e_{ju} exists if employee j directly reports to u according to the formal organizational hierarchy. If j joins the microblogging service after u, we call her join an "influenced join". One can think of "influenced joins" as an implicit indicator of the underlying "influence" network. We define n_t as the number of employees that have joined the microblogging service by time t, i.e., the number of infections at time t. We aim to fit the number of infections over time.

A natural question is how to model the number of infections over time, n_t, as a function of individual "influence" functions due to reporting relationships, and general influence as a function of the service popularity. Next, we describe how we incorporate these dynamics into our model.

4.1 Complex Contagion Model

We begin by selecting a single node from the organization hierarchy to start the infection. We chose the seed node to be the exact employee that first registered to the microblogging service according to our dataset. At each time step, the virus can spread as follows. Each node that was infected at time $t-1$ has n chances to infect the n employees that directly report to her, each with probability p, at time t. Once a node is infected, it cannot be infected again. An infected employee is not allowed to infect her direct supervisor, so following this strategy, the virus can only propagate towards the leaves of the hierarchy tree. Once all infected nodes are examined, healthy nodes have the chance to be "randomly" infected by observing the general popularity of the microblogging service up to time $t-1$. For n_{t-1} total infected nodes at time $t-1$, the probability of "random" infection at time t is r_{t-1}.

Our model incorporates the following dynamics:

- Employees are influenced by their managers to join the microblogging service.
- Employees have multiple chances to get infected (join). Once an employee is infected, she cannot recover, i.e. an employee does not unsubscribe from the service.
- As employees observe others adopting the microblogging service, they are not only more likely to adopt the service, but the rate at which they do so quickens as the popularity of the service increases.

4.1.1 Mathematical Formulation

Next, we present an efficient procedure to estimate the expected outcome of the model, given probability p, a formula for the random infection r_t and the initiator of the virus epidemic. Table 1 summarizes the notation used in our modeling.

Table 1 Notation

u_i	Employee at level i
e_i	Number of employees at level i
N_u	Number of employees directly reporting to manager u
$N_u^j \leq N_u$	Number of employees who directly report to manager u and have joined the microblogging service
$\alpha(u) \leq N_u^j$	Number of employees that report to u and have joined after her
$q(u) \leq N_u^j$	Number of team members of u and have joined before her
$\iota(u)$	Influence score of manager u
ι_λ	Aggregated influence score of managers with λ team members
ι_i	Aggregated influence score of level i
n_t	Number of infections at time t
p	Probability of a manager infecting a team member at any time t
r_t	Probability of random infection at time $t+1$
$A_{i,t}$	Probability of employee u_i getting infected at time t
$B_{i,t}$	Probability of employee u_i getting infected at or before time t
$E_{i,t}$	Indicator function, which value is 1 if employee u_i is infected at or before time t
$S^i_{m,\tau}$	Tuple representing the state of employees from m^{th} level to the i^{th} level at time τ: $(E_{m,\tau}, E_{m+1,\tau}, \ldots, E_{i,\tau})$
$\zeta^i_{m,\tau}$	State representing $(E_{m,\tau} = 0, E_{m+1,\tau} = 0, \ldots, E_{i,\tau} = 0)$
$Z^i_{m,\tau}$	Probability of the zero suffix state $\zeta^i_{m,\tau}$

Consider a hierarchy of employees $\mathbf{U} = (u_1, u_2, \ldots, u_l)$, where the subscript denotes the level of an employee, i.e., u_i reports to u_{i-1}, and l is the bottom level. Let $A_{i,t}$ denote the probability of employee u_i being infected at time t. Also, let $B_{i,t}$ be the probability of u_i being infected at or before time t. Then the following relation holds:

$$B_{i,t} = \sum_{\tau=0}^{t} A_{i,\tau}\,. \tag{1}$$

Suppose the probability of random infection at time t is r_{t-1}, where r_t denotes the probability of random infection as a function of number of infections at time $t-1$. Let $E_{i,\tau}$ be a random variable, which is 1 if u_i is infected at or before time τ, and 0 otherwise. For a given level i, we define state $S^i_{m,n} = (E_{m,\tau}, E_{m+1,\tau}, \ldots, E_{i,\tau})$, for $1 \leq m \leq i$. We use these states to study the spread of infection and find the probability of infection of u_i. Note that,

$$B_{i,t} = P(E_{i,t} = 1) = 1 - P(E_{i,t} = 0) = 1 - P(S^i_{i,t} = (0))\,. \tag{2}$$

The probability $P(E_{i,t} = 0)$ of u_i not being infected till time t, depends on the state of infection of its manager at

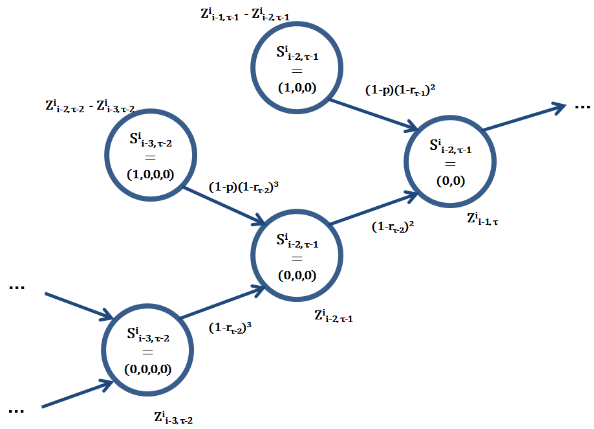


Fig. 3 Zero-Suffix State Transitions.

time $t-1$. Intuitively, $(E_{i,t}=0)$ can be reached only from states $(E_{i-1,t-1}=0, E_{i,t-1}=0)$ and $(E_{i-1,t-1}=1, E_{i,t-1}=0)$. Observe that,

$$
\begin{aligned}
P(E_{i-1,t-1}=1, & E_{i,t-1}=0) \\
&= P(E_{i,t-1}=0) - P(E_{i-1,t-1}=0, E_{i,t-1}=0) \\
&= P(S^i_{i,t}=(0)) - P(S^i_{i-1,t-1}=(0,0)) \,. \qquad (3)
\end{aligned}
$$

In order to find $B_{i,t}$, it is sufficient to look only at specific states. We define these states below:

Definition 1 *A Zero-Suffix State $\zeta^i_{m,\tau}$ is the state that represents $(E_{m,\tau}=0, E_{m+1,\tau}=0, \dots, E_{i,\tau}=0)$.*

Zero-Suffix State $\zeta^i_{m,\tau}$ is reachable either from $\zeta^i_{m-1,\tau-1}$ or from $S^i_{m-1,\tau-1}=(E_{m-1,\tau-1}=1, \zeta^i_{m,\tau-1})$, as demonstrated in Figure 3. Particularly, we have the following transitions:

- From $\zeta^i_{m-1,\tau-1}$: All random infection attempts on employees $\{u_m, \dots, u_i\}$ failed. This happens with probability $(1-r_{\tau-1})^{i-m+1}$.
- From $(E_{m-1,\tau-1}=1, \zeta^i_{m,\tau-1})$: All random infection attempts on employees $\{u_m, \dots, u_i\}$ failed. The attempt of u_{m-1} to infect u_m also failed. This happens with probability $(1-p)(1-r_{\tau-1})^{i-m+1}$.

Therefore,

$$
\begin{aligned}
P(\zeta^i_{m,\tau}) &= P(\zeta^i_{m,\tau}|\zeta^i_{m-1,\tau-1})P(\zeta^i_{m-1,\tau-1}) + \\
&\quad P(\zeta^i_{m,\tau}|(E_{m-1,\tau-1}=1, \zeta^i_{m,\tau-1}))P((E_{m-1,\tau-1}=1, \zeta^i_{m,\tau-1})) \,. \qquad (4)
\end{aligned}
$$

Let $Z^i_{m,\tau}=P(\zeta^i_{m,\tau})$. Then, Equation 4 becomes

$$
\begin{aligned}
Z^i_{m,\tau} =& Z^i_{m-1,\tau-1}(1-r_{\tau-1})^{i-m+1} + \\
& (Z_{m,\tau-1}-Z_{m-1,\tau-1})(1-p)(1-r_{\tau-1})^{i-m+1} \,. \qquad (5)
\end{aligned}
$$

Equation 5 is valid for $n>1$. However, the same equation applies for $Z^i_{0,\tau}$, such that $Z^i_{0,\tau}=Z^i_{1,\tau}$. Finally,

$$B_{i,t} = 1 - Z^i_{i,t} \,. \qquad (6)$$

To calculate $B_{i,t}$ for all i, we recursively evaluate $Z^i_{i,t}$ using Equation 5. Then r_t is updated based on $B_{i,t}$ values, after which we can proceed for time $t+1$, and so on.

One important observation is that the same set of equations apply to all top-down paths $(u_1, u_2, \dots, u_l)$ in the employee hierarchy. Assuming that all paths have the same initial conditions, all employees at a given level are equivalent. Let e_i denote the number of employees at level i. Then the number of infections at time t is given by the following formula:

$$n_t = \sum_i e_i B_{i,t} \,. \qquad (7)$$

Recall, that the probability of random infection r_t, is a function of the number of infections at time $t-1$.

4.1.2 Initial Conditions

A difference in $B_{i,t}$ values can arise for two different paths, which have different initial conditions. The following cases are possible depending on who is infected at $t=0$: (i) no infection at all, (ii) the employee at the topmost level (root) is infected, or (iii) a non-root employee is infected. Next, we elaborate on each case.

Infection at the topmost level (root). If the root of the tree is infected, then all paths share the same initial condition, i.e., $B_{1,0}=1$ and $B_{i,0}=0, \forall i \neq 1$. This is due to the fact that the root is the first node in every top-down path $(u_1, u_2, \dots, u_l)$. Therefore, in this case, all employees at level i are governed by the same probability of infection. Hence, $\forall i$,

$$Z^i_{1,0}=0, \text{ and } Z^i_{k,0}=1, \forall k \text{ such that } 2 \leq k \leq i \,.$$

No infection. If no employee is infected at the beginning, the same initial condition applies to every top-down path, i.e., $B_{i,0}=0, \forall i$. Hence, the probability of infection of any employee is dependent only on the level. Therefore, $\forall i$,

$$Z^i_{k,0}=1, \forall k \text{ such that } 1 \leq k \leq i \,.$$

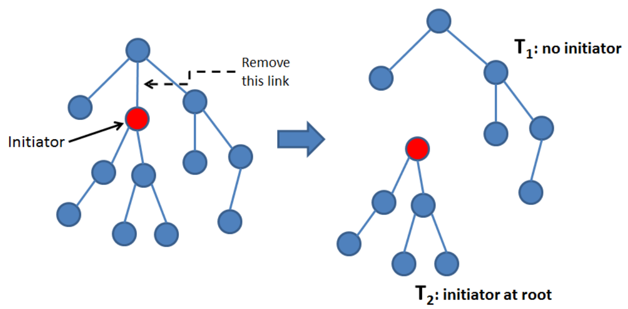


Fig. 4 When the initiator is not at the root, we break the tree into two trees that we know how to deal with.

A non-root employee is infected. In this case two different paths can have different initial conditions depending on whether the path contains the infected employee or not. Since the infected employee can no longer be infected by her manager, we cut the link between this employee and her manager to get two trees T_1 and T_2 (See Figure 4), which resemble the previous two conditions (i.e., infection at the topmost level, and no infection). Tree T_1 is rooted at the original root, whereas T_2 is rooted at the infected employee. Our analytical model applies to each tree, hence, we compute the model in parallel for each tree, i.e., solve Equation 5 for both trees at time t. Note that we calculate r_{t-1} based on the total number of infections in both trees. To summarize, the following formulas hold for each tree:

$$T_1:$$
$$(Z^i_{1,0})_1 = 0, (Z^i_{k,0})_1 = 1, \forall k, 2 \leq k \leq i$$
$$T_2:$$
$$(Z^i_{1,0})_2 = 0, (Z^i_{k,0})_2 = 1, \forall k, 1 \leq k \leq i$$

, where $(\cdot)_1$ and $(\cdot)_2$ denote quantities calculated over trees T_1 and T_2 respectively.

For initial conditions (i) and (ii), the number of infected nodes at time t is given by Equation 7. In this case however, because of the differentiation between the two trees, the number of infected nodes at time t is calculated using the following formula:

$$n_t = \sum_{i=1}^{maxdepth(T_1)} (e_i)_1 (B_{i,t})_1 + \sum_{i=1}^{maxdepth(T_2)} (e_i)_2 (B_{i,t})_2 \quad (8)$$

4.2 Complex Cascade Model

The Complex Contagion model spreads the adoption of the microblogging service over the formal organization hierarchy as a virus, which leaves a trail whenever employees are infected by their supervisors (i.e., local influence), or when employees are influenced by the overall popularity of the microblogging service (See Section 4.1). To model this we used parameter p, which measures how infectious managers are, and parameter r_t, which controls the effect of overall growing popularity of the microblogging service over time. Here we take an alternate approach based on which, nodes “choose” to become infected after examining their immediate neighborhood (which includes both the manger and employees directly reporting to them) or after examining the overall growing popularity of the microblogging service over time.

We start with the organization hierarchy, and two colors. Let red represent employees who have joined the microblogging service and blue those that have not. We choose a single node to be the seed user, i.e. have color red. All other users are painted blue. As before, we chose the seed node to be the exact employee that first registered to the microblogging service according to our dataset. At each time step, nodes painted blue (not infected), calculate the payoff of picking the color red over blue, and decide their color $f(\text{color})$ as follows:

$$f(\text{color}) = \begin{cases} \text{red}, & \alpha \frac{n_{red}}{n} > \beta \frac{n_{blue}}{n} \\ \text{blue}, & \text{otherwise} \end{cases}, \quad (9)$$

where n_{blue} denotes the number of blue neighbors, n_{red} denotes the number of red neighbors and $n = n_{blue} + n_{red}$ is total number of neighbors. Parameters α and $\beta = 1-\alpha$ denote the rewards for choosing red and blue accordingly. Once a node is painted red, it cannot change color again. Finally, nodes have the chance to be “randomly” infected by observing the general popularity of the microblogging service up to time $t-1$. As in our contagion model, for n_{t-1} infected nodes at time $t-1$, the probability of “random” infection at time t is r_{t-1}.

5 Influence Estimation

Let N_u be the total number of employees directly reporting to manager u. Among these N employees, let K be the number of employees who joined the microblogging service after their manager u, and k be the total number of employees who joined the microblogging service after their manager u within the first ‘n draws’. We counted the number of employees who joined the microblogging service after their manager and found that there are three classes of employees: (i) employees who did *not* join the microblogging service even if their manager did (10.94%), (ii) employees who *did join* the microblogging service *before* their manager (36.04%), and (iii) employees who *did join* the microblogging service *after* their manager (53.01%).

The stochastic process according to which employees directly reporting to u choose to join the microblogging service is described by the “urn model” [10], in which n balls are drawn without replacement from an urn containing N

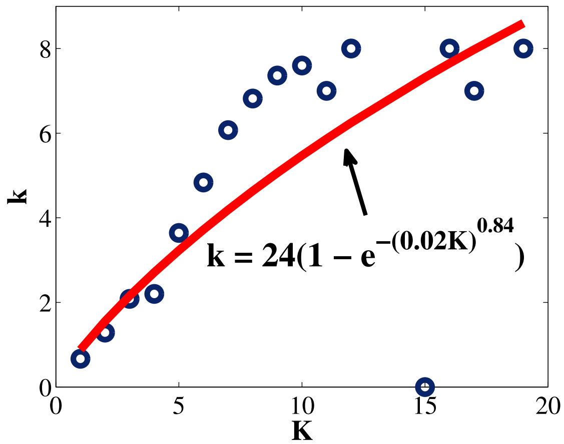


Fig. 5 Average number k of employees that joined the microblogging service after their manager, within the first n samples vs the total number K of employees that joined the microblogging service after their manager, and approximation.

balls in total, of which K are white. The probability $P(X = k|K,N,n)$ that k of the first n employees reporting to manager u, joined the microblogging service after their manager purely by chance is equivalent to the probability that k of the n balls drawn from the urn are white. We set $n = 8$, calculating the number of employees that joined the microblogging service after their manager within the first 8 draws. This probability is given by the hypergeometric distribution:

$$P(X = k|K,N,n) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}. \tag{10}$$

We plot the average number of employees that joined the microblogging service after their manager during the first n samples as a function of the number of employees that joined the microblogging service after their manager. Figure 5 shows the result. The scatter plot is approximatedby the Weibull cumulative distribution ($\hat{k} = 24(1-e^{-(0.02K)^{0.84}})$ [10]. We use this expression to estimate the expected number $\hat{k}$ of employees to join the microblogging service after their manager within the first n joins for a manager with K employees reporting to her that joined the microblogging service after her. Using Equation 10, we calculate the probability that $\hat{k}$ employees joined after their manager purely by chance. We found that for $K > 3$, this probability is exceedingly small. Since it is exceedingly highly unlikely for employees to adopt the microblogging service after their manager purely by chance, we conclude that the number of employees who joined after their manager u is a prominent indicator of u's influence.

5.1 Influence Score

Let N_u^j denote the number of employees who directly report to u and have joined the microblogging service. Let $\alpha(u) \leq N_u^j$ be the number of employees that report to u and have joined the microblogging service after u, and let $q(u) \leq N_u^j$ be the number of employees that report to u and have joined the microblogging service before u. While a high number of employees reporting to u that have joined the microblogging service after u implies that u has high influence, a high q value is an indicator that one lacks influence. We propose an adaptation of the z-score [29], as a measure that combines the number of employees that have joined before and after their supervisor. Influence score ("ι-score") measures how different this behavior is from a user with "random" influence, i.e. a manager the employees reporting to whom join after him with probability $p = 0.5$ and before him with probability $(1-p) = 0.5$. We would expect such a random influencer to have $N_u^j * p = N_u^j/2$ team-members who joined after their supervisor with a standard deviation of $\sqrt{N_u^j * p * (1-p)} = \sqrt{N_u^j}/2$ [29]. The ι-score measures how many standard deviations above or below the expected "random" value a manager u lies:

$$\iota(u) = \frac{\alpha - N_u^j/2}{\sqrt{N_u^j}/2} = \frac{\alpha(u) - q(u)}{\sqrt{\alpha(u) + q(u)}}. \tag{11}$$

If the employees reporting to manager u have joined the microblogging service after u about half of the time, u's ι-score will be close to 0. If they join after u more often than not, u's ι-score will be positive, otherwise, negative. We also calculate the time-independent ι-score of employees using Equation 11, with the difference that $\alpha(u) \leq N_u^j$ is the number of employees that have joined the microblogging service (irrespective of time) and $q(u) \leq N_u^j$ is the number of employees that have *not* joined the microblogging service. Above, we measured influence at the level of individual employees, assuming that influence scores are fixed in time, but that they differ from employee to employee. A more sophisticated model of influence might include some small increase (similarly for decrease) in influence score as a function of time. We stick to the simpler model for simplicity, and because our fundamental result is not sensitive to such details (see Section 7).

Next, we examine the correlation between ι-score of managers and the number of employees reporting to them (team size), hoping to get a clearer picture of the relationship between the two quantities.

Definition 2 *We define the average ι-score of managers with λ employees reporting to them as*

$$\iota_\lambda = \frac{1}{|u : \lambda_u = \lambda|} \sum_{u:\lambda_u=\lambda} \iota(u), \tag{12}$$

where $\iota(u)$ is the influence score of manager u.

We now turn our attention to the impact of organizational levels. Here, we assume that influence scores are characteristic of a particular level at the organization hierarchy tree, are fixed in time, and are the same for all employees at that particular level. To compute the average influence score for hierarchy level i, we first find employees m that belong to level i. We then find the total number of employees N that directly report to managers in level i. Quantities α and q are defined as before, with the difference that they now operate on the total number of employees N that directly report to managers in E_l. Finally, we use Equation 11 to calculate the influence score for each level.

Definition 3 *We define the average ι-score of managers at level i as*

$$\iota_i = \frac{\sum_{level(u)=i}(\alpha(u)-q(u))}{\sqrt{\sum_{level(u)=i}(\alpha(u)+q(u))}}\,. \tag{13}$$

Levels are ascending from the CEO (level 1) to lower levels. Level 13, which represents bottom level employees in our dataset, contains employees with no team members reporting to them.

5.1.1 Relation to Complex Contagion Model

The probability of a randomly selected team member v being infected after her manager u is $\sum_k P(v \text{ infected at time } t > k | u \text{ infected at time } t = k) P(u \text{ infected at time } t = k)$. Assuming that all employees are equivalent, i.e., they are described by the same probabilities $A_{i,t}$, and applying our mathematical formalism for the team members of u, we get:

$$\alpha(u) = N_u \sum_{k_2 > k_1} A_{i,k_1} A_{i+1,k_2}\,. \tag{14}$$

Similarly, for $q(u)$, we get:

$$q(u) = N_u \sum_{k_1 > k_2} A_{i,k_1} A_{i+1,k_2}\,. \tag{15}$$

The above formulas are valid only when there is no initial infection, or if the initiator is the root (see Section 4.1.2). When the initiator is not the root, we cut the link between the initiator and its manager, thus splitting the tree into two trees. The $\alpha(u)$ and $q(u)$ values for each node are then calculated separately for both trees.

It is straightforward to show that,

$$\frac{\alpha(u)}{N_u} = \sum_k A_{i,k}(B_{i+1,T} - B_{i+1,k})\,, \tag{16}$$

and

$$\frac{q(u)}{N_u} = \sum_k A_{i,k} B_{i+1,k-1}\,, \tag{17}$$

where T is the total time. Hence, the influence score of employee u only depends on u's level i and the number of u's team members. This is shown in Equation 18, where the fraction in the right hand side depends only on level i (see Equation 16 and Equation 17), and so we denote it by $f(i)$:

$$\begin{aligned}\iota(u) &= \frac{\alpha(u)-q(u)}{\sqrt{\alpha(u)+q(u)}} = \sqrt{N_u}\frac{(\alpha(u)-q(u))/N_u}{\sqrt{(\alpha(u)+q(u))/N_u}} \\ &= \sqrt{N_u} f(i)\,.\end{aligned} \tag{18}$$

The average influence score of managers with λ employees reporting to them is then computed as follows:

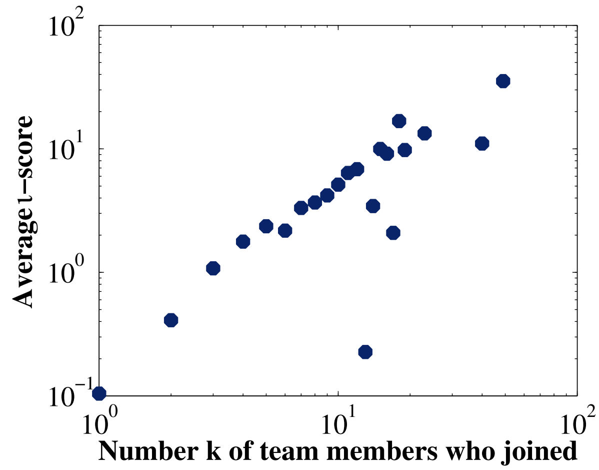
$$\begin{aligned}\iota_\lambda &= \frac{1}{|\{u : N_u = \lambda\}|}\sum_{u:N_u=\lambda}\iota(u) \\ &= \frac{1}{|\{u : N_u = \lambda\}|}\sum_i \sum_{\substack{level(u)=i,\\ N_u=\lambda}} \sqrt{\lambda} f(i) \\ &= \sqrt{\lambda}\frac{\sum_i N_{i,\lambda} f(i)}{\sum_i N_{i,\lambda}}\,,\end{aligned} \tag{19}$$

, where $N_{i,\lambda}$ denotes the number of employees at level i who manage λ employees. Finally, the average influence score of managers at level i is given by the following formula:

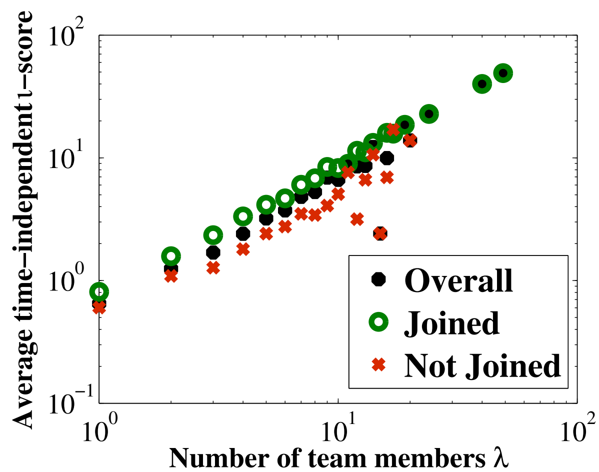
$$\iota_i = \left(\sqrt{\sum_{level(u)=i} N_u}\right) f(i) = \left(\sqrt{\sum_\lambda \lambda N_{i,\lambda}}\right) f(i)\,. \tag{20}$$

5.1.2 Empirical Estimation of Influence

Figure 6(a) shows the average ι-score of managers with λ employees reporting to them, that have joined the microblogging service. Here, we focus on managers that have themselves joined the microblogging service, so that a time comparison of joining times is meaningful. A clear increasing trend is evident, providing a supporting evidence on top-down influential flow through the formal organizational hierarchy. Figure 6(b) shows the average time-independent ι-score of managers with λ employees reporting to them. Figure 6(b) further shows different plots of the average time-independent ι-score of managers based on the premise that they have joined the microblogging service themselves or not. The average time-independent ι-score of managers that have not joined the microblogging service exhibits more fluctuations due to greater data sparsity. In every case, the average time-independent ι-score of managers that have joined the microblogging service is slightly higher than for managers that have not joined the service. Even though we cannot at the time explain the reasons why this effect appears, the average time-independent ι-score increases for both classes



(a)



(b)

Fig. 6 (a) Average ι-score of managers with λ team members that have joined the microblogging service. (b) Average time-invariant influence of managers, who have themselves joined the microblogging service (similarly for those who have not joined), with λ team members.

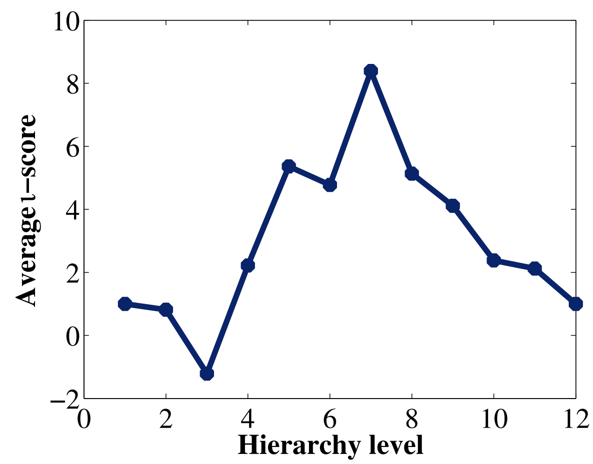


Fig. 7 Average influence score as a function of hierarchy level.

as the team size λ increases, clearly indicating a strong correlation between the two quantities. We explain this trend as a prominent indicator of influence imposed by managers to employees reporting directly to them.

The organizational levels impact is shown in Figure 7. Level 13 has no influence score, thus it does not appear in Figure 7. Most levels exhibit positive influence scores, with the exception of higher levels, that are closest to the CEO. Particularly, level 3, exhibits negative influence on average. As before, we measured influence at the granularity of hierarchical levels, assuming that influence scores are fixed in time, but that they differ from level to level. A more sophisticated model of influence might include some small increase (similarly for decrease) in influence score as a function of time, and also introduce a balancing factor based on the number of total employees at a level and the number of total employees reporting to them. While it is intuitive to assume that higher levels in the organization would have higher impact due to the report-to relationships involved, our study suggests that middle levels are more successful in influencing employees lying lower in the hierarchy. Even though we do not have supporting evidence from other use-cases, we conjecture that middle-level managers are the most influential with respect to "convincing" others to adopt new technologies (in this case the new microblogging service). This assumption can be further supported with evidence from other datasets. We intend to experiment with more datasets in future work.

6 Accounting for Peer Pressure

So far, we have assumed that an employee can be infected either by her direct supervisor or randomly, as a result of the overall popularity of the microblogging service in the enterprise. Classic models of social and biological contagion (e.g. [12,22]) and observational studies of online contagion [2,3,7,18] predict that the likelihood of infection increases with the number of infected contacts. However, recent studies suggest that this correlation can have multiple causes that might be unrelated to social influence processes [4]. In our observational study of microblogging service adoption at the workplace, this assumption suggests two alternative modeling scenarios. According to the first scenario, an employee is more likely to adopt the microblogging service if more of her teammates join the service (Section 6.1). According to the second scenario, an employee is more likely to adopt the microblogging service as its popularity increases (Section 6.2). Our goal in this section is to estimate the probability of adoption for each user given the actions of their teammates (local neighborhood) or overall popularity (global influence).

6.1 Independent Cascade Model

Influence of friends is generally modeled to be additive. For instance, the independent cascade model (ICM) [17] states that a node has n independent chances to become infected, where n is the number of infected “friends”. In our case, every node can be infected only once, and once infected, it stays infected. Because of the structure of the organizational hierarchy, employee u's “friends” may include either (i) her teammates alone, or (ii) her teammates and her direct supervisor. Starting with a single employee who has joined the microblogging service, employees *susceptible* to infection, decide to join the microblogging service with some probability that depends on the number of their infected “friends”. We model the influence employees receive by their “friends” as multiple exposures to an infection according to ICM [17] as $p_{ICM} = 1-(1-\lambda)^n$.

We measured this quantity on our dataset, by isolating the employees in two classes: a) those who had exactly n “friends” joining the microblogging service and did not join, and b) those who had exactly n “friends” joining the microblogging service before they themselves joined. We found that the likelihood of adoption when no “friends” have joined is remarkably high (0.7581 when considering teammates only and 0.6807 when the supervisor is also considered). In both cases, the likelihood of adoption becomes 1 when at least one “friend” has joined the service. We conclude that the relationship between the number of “friends” that have joined and likelihood of joining most probably reflects heterogeneous popularity of the microblogging service across teams [4]. Therefore, the naive conditional probability does not directly give the probability increase due to influence via multiple joining “friends” [4].

6.2 Exponential Growth Model

We studied earlier the effect of multiple teammates and neighbors of an employee u on the probability of u to join the microblogging service. Even though we discovered a positive correlation, we argued that this correlation might be an effect of multiple causes. We hypothesized that the more popular the microblogging service is for a team, the more likely it is for multiple team members to adopt it. Further, as employees observe others adopting the microbloggig service, they may not only be more likely to adopt the service, but the rate at which they do so may quicken as the popularity of the service increases. Here, we explore this hypothesis further.

We start by splitting the employee population in two pools: those who have already joined the microblogging service and those who have not. Assuming an exponential growth model, the rate at which employees join the service should follow an increasing trend. Intuitively, as more people adopt

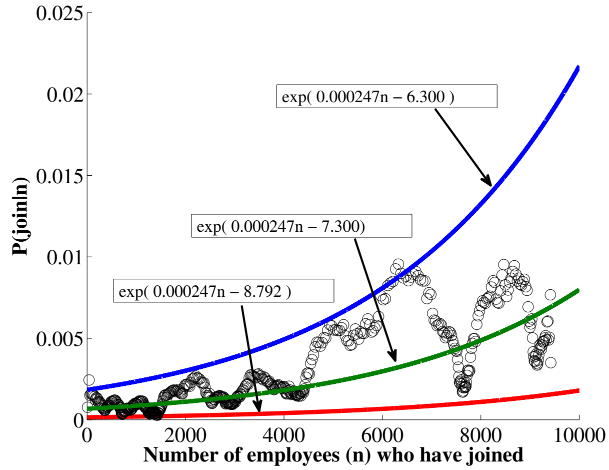


Fig. 8 Probability (calculated empirically from our dataset) an employee joins the microblogging service given that n employees have adopted the service before. Solid lines depict probability estimates calculated with the exponential growth model.

the microblogging service, a certain “buzz” around the service begins to unfold, increasing the probability of others joining the service as well. Figure 8 shows the probability that an employee will join the microblogging service as a function of the service popularity. Interestingly, Figure 8 reveals that the probability of employees joining the microblogging service is in fact neither constant nor monotonically increasing (similarly decreasing). Instead it exhibits increasing and decreasing regimes over time. This observation suggests that more complex dynamics take place over the organizational hierarchy. One possible explanation of this phenomenon is that whenever influential managers join the microblogging service, a period of “influenced joins” follows (Section 4). In essence, this provides a hint that the adoption mechanism follows a snowball effect propagating the epidemic in a top to bottom fashion, followed by a random infection that exposes new portions of the population.

Since, the probability of joining given the number of total infections incorporates the probability of an “influenced join”, we fit three exponential growth models. The first model (blue line) provides an “*optimistic*” expected probability of adoption. Contrary, the “*pessimistic*” model (red line) yields a probability of adoption that increases marginally as the total number of people who join the service increases over time. Finally, the average fit (green line) shows how the probability of adoption follows on average an increasing trend as a function of previous adoptions.

7 Experiments

In this section, we validate our models by extensive numerical simulations. We begin with the organization hierarchy of 12,170 employees, and infect the true initiator of the epidemic (the employee who first joined the microblogging service). Each time step represents a day. We let our models run for 600 steps, or until all employees are infected. We compare the obtained epidemics against the real cumulative number of adoptions extracted from our dataset. We experimented with various values of infection probability for our contagion model and parameters α and β for our cascade model. In the end, we decided to use $p = 0.023$ for our contagion model, and $\alpha = 0.82$ and $\beta = 0.18$ in our cascade model. We simulated our models 10 times and report our findings. We compare three properties of the simulated epidemics as opposed to the true number of adoptions over time: (i) overall number of infections, (ii) cumulative number of infections over time, and (iii) total time required to infect N employees. We find that our models' estimates are consistent with the real observations.

7.1 Baselines

We compare our proposed models' ability to approximate the true cumulative distribution of infected users with three models, which have shown superior performance in the task of information and innovation diffusion in social networks. Particularly, we consider the Susceptible-Infected Model [15], the Independent Cascade Model [17], and the Diffusion Model proposed by [1,5,9].

- **Susceptible-Infected Model (SI)** [15]: According to the SI model, each node can infect her neighbors, each with probability p_{SI}. We considered the Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Resistant (SIR) models [14], as well as the Susceptible-Infected-Dead (SID) model [16] as alternatives to model social contagion, as these models are widely used in prior work. These models however do not appropriately capture the semantics of adoption, according to which, an employee that joins the microblogging service does not unsubscribe, thus returning to the susceptible state, or becoming resistant. Further, our analysis did not provide any supporting evidence for the hypothesis that infected employees do not infect others, thus modeling them as "dead" is not appropriate in this case.
- **Independent Cascade Model** [17] (see Section 6).
- **Diffusion Model (DM)** [1,5,9]: Each individual's willingness to adopt the microblogging service at time t, U_u^t, is modeled by three main elements: the service's stand-alone benefit, network effects, and the idiosyncratic reservation utility. Formally, $U_u^t = Q_u + \gamma N_u^{(t-1)} - R_u$, where, Q_u represents the service's intrinsic value perceived by employee u, which is not affected by whether other people adopt it or not. $N_u^{(t-1)}$ represents the proportion of adopters in u's neighborhood at time $t-1$, and γ denotes the relative importance against stand-alone benefits. R_u indicates u's inherent reluctance or reservation about adopting the new service.

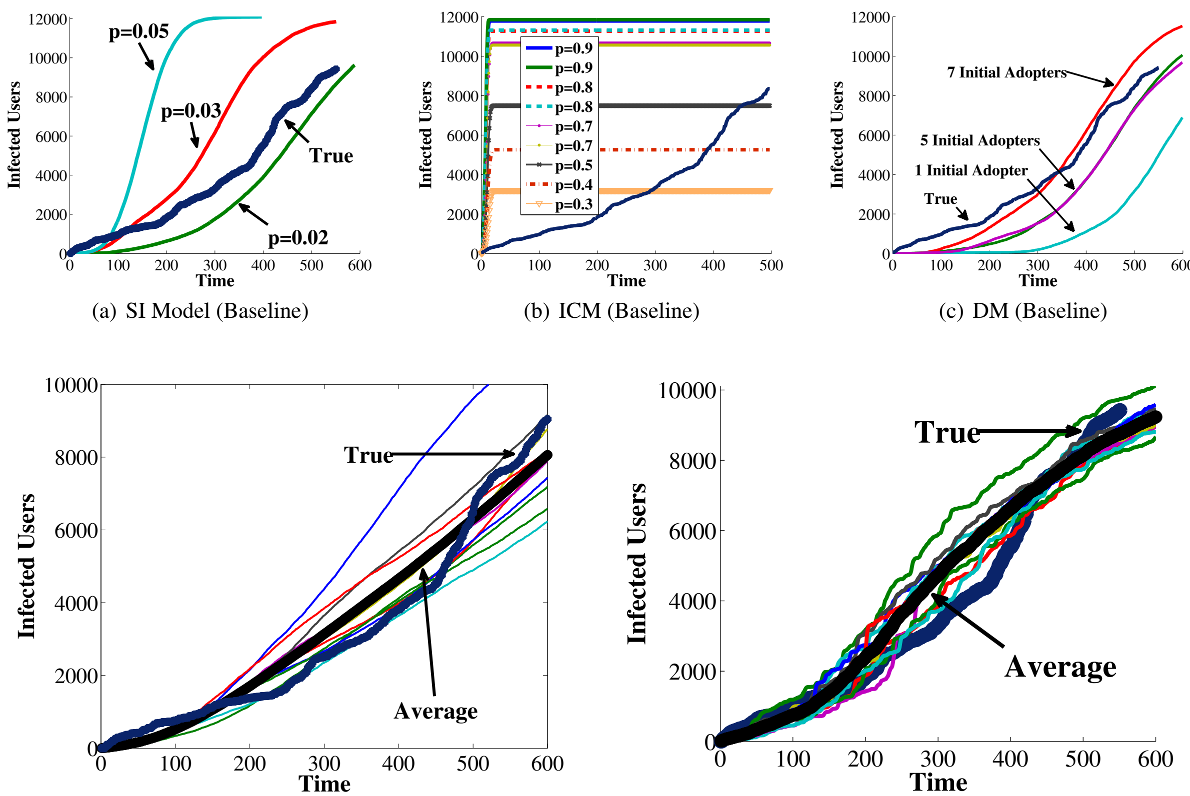
7.2 Empirical Evaluation

First, we study simulation results produced by the baselines, i.e. the SI, ICM and DM models. Figure 9(a) shows the true user adoption curve, compared to simulation results produced by the SI model, for varying infection probability values. We notice that simulation models do not fit the real cumulative number of adoptions over time. High infection probability values result in sudden outbreaks, whereas very small probability values result in smooth cumulative distributions that do not exhibit the statistical properties of the true cumulative number of infected users. The total number of infections and the time required to infect the whole body of employees is also inconsistent with the observed adoption curve.

Figure 9(b) shows simulation results produced by the ICM model, for varying infection probability values. Clearly, the simulation results do not fit the real cumulative number of adoptions over time. In fact, this model results in sudden epidemics, which also fail to cover the entire population, and eventually come to a halt. No new infections are achieved due to the fact that each exposure has a single chance of success. If the result of an exposure is no infection, that connection is not examined again. Hence if the root of a subtree is not infected, the infection cannot proceed further down the subtree. The simulation results corroborate our conjecture that the naive conditional probability does not directly give the probability increase due to influence via multiple joining "friends" [4] (see Section 6.1).

Figure 9(c) shows simulation results produced by the DM model, for varying numbers of initial adopters. When the first true adopter is selected to start the infection, the epidemic progresses slowly. Instead, when five true adopters are used, the epidemic is substantially speeded up. When the seed set contains seven of the true adopters, the simulation result adequately fits the observed adoption curve, without however exhibiting the statistical properties of the true cumulative number of infected users. Overall, this model too fails to capture the hidden dynamics of technology adoption at the workplace.

Next, we show the outcome of ten runs of our complex contagion model (see Section 4.1) in Figure 9(d). The figure also shows the average of the ten runs. Notice a very good alignment between the reality and simulated epidemics in all cases. Not all runs result in the total number of true infections by the time threshold. Further, a few runs overestimate the cumulative number of infections, resulting in rapid epidemics. Unlike the baselines, our complex contagion model



(a) SI Model (Baseline) (b) ICM (Baseline) (c) DM (Baseline)

(d) Complex Contagion Model (e) Complex Cascade Model

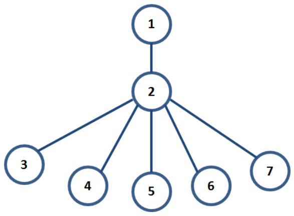
Fig. 9 True and predicted cumulative number of employees who have adopted the microblogging service (i.e. infected users). Time is measured in days. Solid line curves represent the outcome of (a) the SI model for various probabilities of infection, (b) the ICM model for various probabilities of infection, (c) the DM model for various numbers of initial adopters, (d) ten runs of our complex contagion model (see Section 4.1), and (e) ten runs of our complex cascade model (see Section 4.2).

fits more naturally the true cumulative number of infected users in all cases. Particularly, the simulation results remarkably follow the speedups and slowdowns of adoption over time, exhibiting non-linear characteristics as the true adoption curve. Some runs diverge from the true curve after about 400 days. However, running the model numerous times and averaging the results seems to adequately approximate the statistical properties of the true cumulative number of infected users. We conclude that this is a direct result of the asymmetric contagion due to the hierarchical influence to adoption and the integration of peer pressure due to growing popularity of the microblogging service at the enterprise.

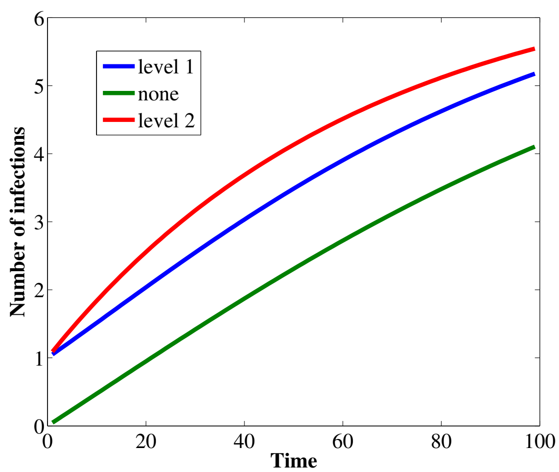
Finally, we present the outcome of ten runs of our complex cascade model (see Section 4.2), and their average, in Figure 9(e). In this case too, simulated epidemics match the reality very well. Similar to the epidemics produced by our contagion model, not all runs result in the total number of true infections by the time threshold. Further, smooth regimes of adoption, speedups and slowdowns of the acceptance of the microblogging service from employees is apparent. Unlike our contagion model, this model slightly overestimates the cumulative number of infections. In all cases however, we find that this model too fits rather closely to the true cumulative number of infected users, replicating the statistical properties of the empirical epidemic.

7.3 Analysis

Using our complex contagion mathematical formulation, we now study the expected behavior of the model. Our goal is to provide insights on the model and the expected average behavior of a technological adoption "epidemic" at the workplace. In Figure 10, we compare the average of 50 simulations results produced by our model in NetLogo, with our theoretical expected outcome. To cover the three cases stated in Section 4.1.2, we do the comparisons with three different initial conditions: initiator at the topmost level (Figure 10(a)), no initiator (Figure 10(b)), and non-root initiator (Figure 10(c)). Note how the theoretically predicted behavior matches the simulation results in all the three cases. When the initiator is at the topmost level, the number of infections rises quickly. On the other hand, when the initiator is at level 5, the adoption curve seems to be very similar to the case of no initiator. This is due to the fact that the adoption caused by the initia-



(a) The tree structure



(b) Adoption curve for the toy example

Fig. 11 An example where infecting the CEO may not lead to the fastest adoption. The parameters are $p = 0.01$ and $r_t = \exp(0.001n_t - 5)$. Here, due to a small value of p, if one starts from the level 1, it takes some time to reach level 2. On the other hand node '2' at level 2 has many children (team members), and so it is more likely to infect one of its children.

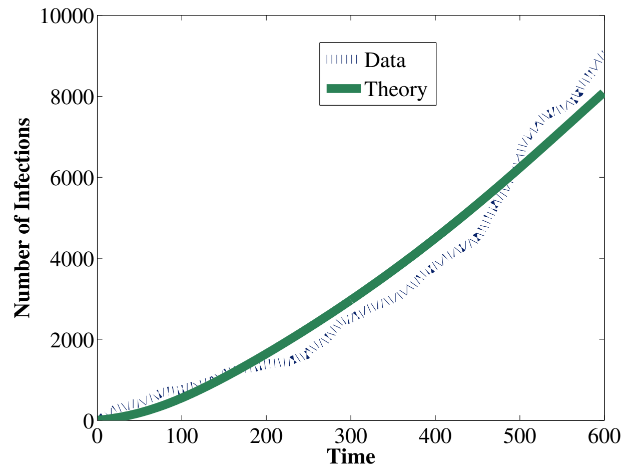


Fig. 12 True vs. expected number of infections.

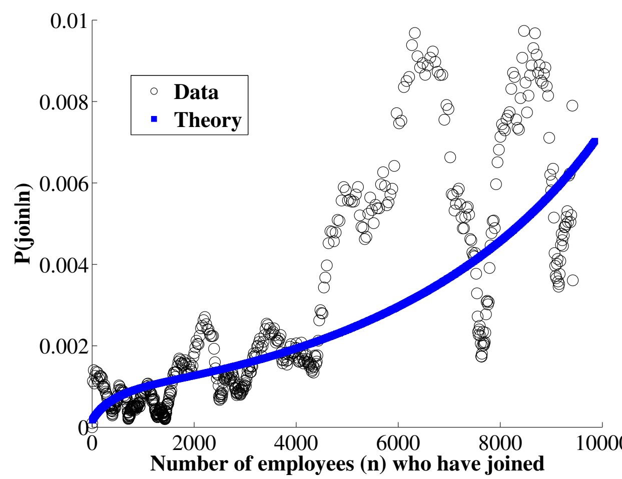
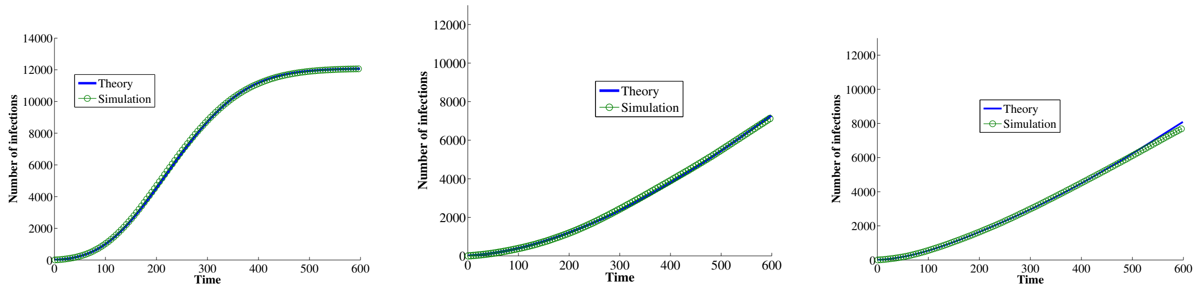


Fig. 13 Probability of infection when n employees are already infected.

tor propagates only in her social capital, i.e., her subtree in the hierarchy. The size of the subtree of the initiator at level 5 is very small compared to the total size of the organization ($\sim 5.6\%$), and hence the behavior is very similar to the case of no iniatator. However, this does not necessary imply that starting with the CEO will always lead to the fastest adoption in the organization. Figure 11 shows a toy example where initiating the infection through CEO may not lead to the fastest adoption.

Figure 12 compares the true number of infections over time with the epidemic predicted in theory ($p = 0.023$, $r_t = \exp^{(}0.000247n_t - 8.792))$. In the theoretical framework the original initiator is infected at $t = 0$. Unlike the baselines (see Section 7.2), our theoretical complex contagion model nicely fits, on average, the true cumulative number of infections over time. We also compare the probability of adoption given n people have already joined (Figure 13). The curve from the data has noticeable peaks which are determined by the specific employee who got infected. These peaks could have occured anywhere else depending on the actual adoption sequence. Since our formulation captures the average behavior, the curve obtained by the theory lacks these peaks, but is still able to capture the general increasing trend.

To study the initiator's effect on the spread of infection we fixed parameters p and r_t and varied the first infected individual. In our "what-if" analysis, we selected for each level the employee with the greatest social capital, i.e., the node with the largest subtree, as the most prominent node to have the greatest effect with respect to influence. Figure 14 shows the result. As expected, starting from the topmost level leads to the fastest spread of infection. This is because apart from the random infection, which is equally likely to infect any employee at a given time, an infected node can influence only its subtree. The effect of level of initiator is not very prominent at the lower levels. The adoption curves, when the initiators are in level 6 through level11 (not shown in the figure), all lie in the small region between the adoption curves of level 5 and level 12 initiator.



(a) Initiator at the topmost level (b) No initiator (c) Initiator at level 5

Fig. 10 Comparison of Complex Contagion Model simulation results (averaged over 50 runs), with theoretically predicted outcome.

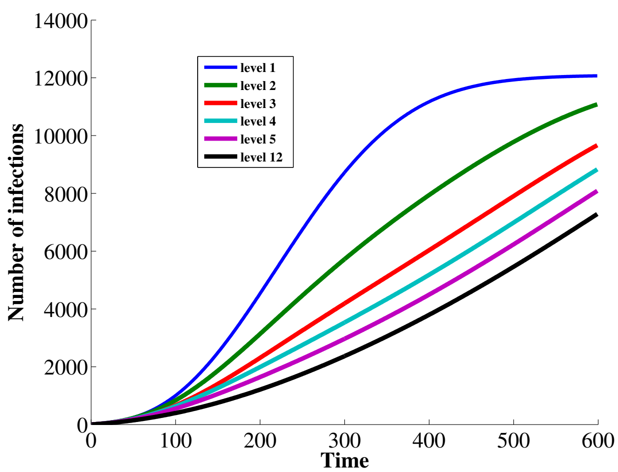


Fig. 14 Effect of level of initiator on the number of infections over time.

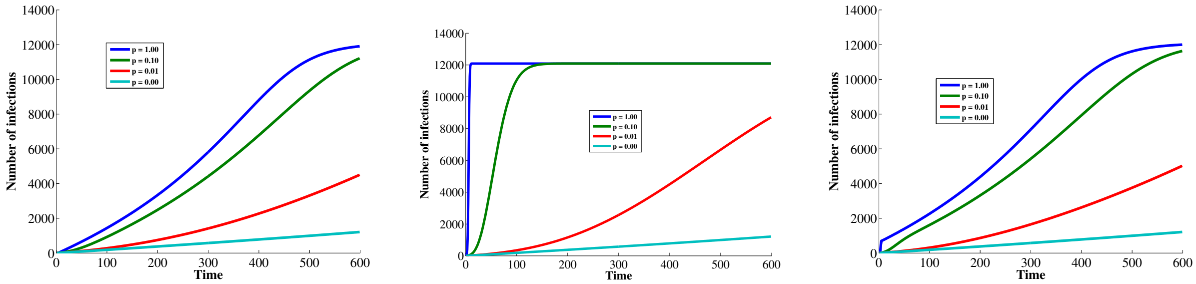
Next, we study the effects of parameters p and r_t on the spread of infections. First, we varied the value of parameter p, while calculating r_t according to the formula $r_t = \exp^{(}0.000247 n_t - 8.792)$. Then, we varied r_t while keeping the value of p constant ($p = 0.023$). The difference in infection for varying p is most pronounced when the initiator is at level 1. However, for the case of varying r_t, it is least prominent when the initiator is at level 1. This shows that if the CEO is the initiator, the influence is more important, in the sense that it causes faster infection spread. Otherwise if the initiator is at lower level, random infection is more important. This behavior is expected because if we have a high value of p with the CEO being the initiator, the infection will reach all the employees very quickly. But if the initiator is at a lower level it can affect only its subtree irrespective of the value of p, and further adoption relies on the random infection.

As we explained above, the level of initiator plays an important role in the spread of infection. Therefore, it is natural to wonder if it also affects the influence of a node (manager). The intuition behind this claim is that if the topmost level employee is infected initially, due to top-down nature of influence in our model (parameter p), many nodes may get positive $\iota-$scores. On the other hand, if the initiator is at the lowest level of the hierarchy, it will not be able to directly influence anyone. Since there are many nodes in the middle levels, the random infection is more likely to infect them first, resulting in negative $\iota-$scores for the top and low levels.

Figure 17 shows the effect that the level of the initiator has on the aggregate influence scores ι_i and ι_λ. We find that less levels exhibit positive influence scores as the level of the first infected individual increases. It is also clear that mid level managers are more influential in all cases. The level with the maximum influence sifts slightly as the level of the first infected individual increases, but overall, levels 4 to 8 are the most successful in achieving “influenced joins” in all cases. Finally, the influence scores show an increasing trend on average, as the number of team members increases. This result is consistent with our observations based on the true data (see Figure 6(b) and Figure 7 in Section 5.1.2).

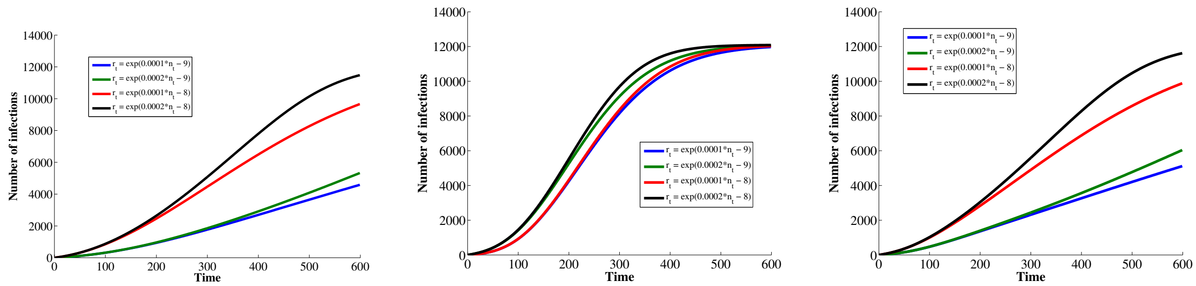
8 Conclusion

In this paper, we studied the effect of the formal organizational structure, to the adoption mechanism of a microblogging service at the enterprise. We addressed the factors that govern the process of adoption at both microscopic and macroscopic levels. We found, microscopically, that employees’ tendency towards adopting the new microblogging service is influenced by their direct supervisors (dependency on the network structure). We proposed ι-score as a prominent indicator of influence imposed by managers on their teams and we demonstrated that middle level managers are on average more successful in promoting the adoption of the new service. Further, we empirically measured employees’ likelihood of adopting the new microblogging service with respect to the behavior of the general crowd. We revealed two distinct patterns, that capture the adoption likelihood incre-



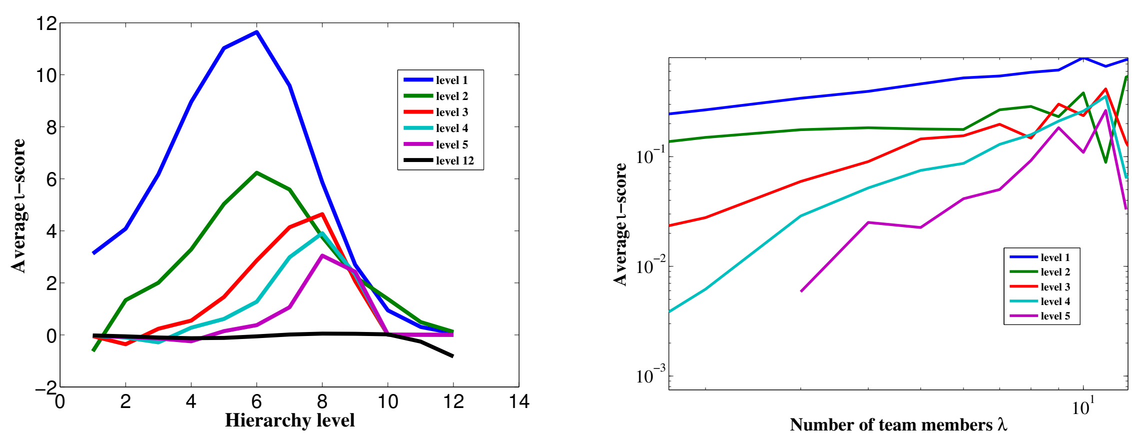
(a) No initiator (b) Initiator at the topmost level (c) Initiator at level 5

Fig. 15 Effect of parameter p on the number of infections over time.



(a) No initiator (b) Initiator at the topmost level (c) Initiator at level 5

Fig. 16 Effect of parameter r_t on the number of infections over time.



(a) ι-score vs level (ι_i) (b) ι-score vs teamsize (ι_λ)

Fig. 17 Average ι-score with different levels of initiator.

ment as a function of the overall service popularity among the employee population. We incorporated our findings into two intuitive and simple adoption mechanisms, which capture both the local and global influence, accurately reproducing the adoption process at the macroscopic level. Prediction results show that our models provide great improvements over commonly used diffusion models. Our findings have important implications to enterprises' understanding of the mechanisms driving adoption of new technologies, and could assist in designing better strategies for rapid and efficient technology adoption and information dissemination within the corporation.

A limitation of our study is that we estimate causal effects only within the formal organizational chart, due to the fact that we are unable to observe the actual adoption "cascade" (i.e. who really influences whom). We are planning to further evaluate our results with extended surveys and targeted interviews, as well as incorporate more datasets in future work. In future work, we plan to enhance our model in various dimensions. First, in the real world topologies other than tree structure may exist. We therefore plan to address the more general problem of influence over an arbitrary graph. Second, we plan to extend our models to allow for influence scores to vary over time, as well as incorporate different roles individual assume in the adoption process, accounting for influence variations as a function of employees' level in the organization hierarchy. Third, we would like to investigate the effect of network evolution (e.g. layoffs, or new hires) on influence, since one's influence may intuitively increase with seniority in the company. Finally, it would be interesting to study adoption dynamics in the presence of competing technologies.

Acknowledgment

This work is supported by Chevron Corp. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

References

1. E. Abrahamson and L. Rosenkopf. Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, 8(3):289–309, 1997.
2. A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15, New York, NY, USA, 2008. ACM.
3. E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 325–334, New York, NY, USA, 2009. ACM.
4. E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528, New York, NY, USA, 2012. ACM.
5. F. M. Bass. A new product growth for model consumer durables. *Manage. Sci.*, 50(12 Supplement):1825–1832, December 2004.
6. C. Budak, D. Agrawal, and A. El Abbadi. Diffusion of information in social networks: Is it all local? In *2012 IEEE 12th International Conference on Data Mining (ICDM)*, pages 121–130, 2012.
7. M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*, pages 721–730, New York, NY, USA, 2009. ACM.
8. C. Chelmis and V. K. Prasanna. Microblogging in the enterprise: A few comments are in order. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 62–70, 2012.
9. H. Choi, S.-H. Kim, and J. Lee. Role of network structure and network effects in diffusion of innovations. *Industrial Marketing Management*, 39(1):170 – 177, 2010.
10. R. Ghosh and K. Lerman. Predicting influential users in online social networks. In *Proceedings of KDD Workshop on Social Network Analysis*, 2010.
11. M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1019–1028, New York, NY, USA, 2010. ACM.
12. M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
13. O. Günther, H. Krasnova, D. Riehle, and V. Schöndienst. Modeling micro-blogging adoption in the enterprise. *Proceedings of the 15th Americas Conference on Information Systems*, 2009.
14. H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.
15. J. A. Jacquez and C. P. Simon. The stochastic si model with recruitment and deaths i. comparison with the closed sis model. *Mathematical Biosciences*, 117(1-2):77–125, 1993.
16. C. Kamp. Untangling the interplay between epidemic spread and transmission network dynamics. *PLoS Computational Biology*, 6(11):e1000984, 11 2010.
17. D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM.
18. J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *Proceedings of the 7th ACM conference on Electronic commerce*, pages 228–237, New York, NY, USA, 2006. ACM.
19. C.-T. Li, T.-T. Kuo, C.-T. Ho, S.-C. Hong, W.-S. Lin, and S.-D. Lin. Modeling and evaluating information propagation in a microblogging social network. *Social Network Analysis and Mining*, pages 1–17.
20. M. McPherson, L. S. Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.
21. M. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
22. M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, Jul 2002.
23. S. Solomon, G. Weisbuch, L. d. Arcangelis, N. Jan, and D. Stauffer. Social percolation models. *Physica A: Statistical Mechanics and its Applications*, 277(12):239 – 247, 2000.
24. D. Strang and M. W. Macy. In search of excellence: Fads, success stories, and adaptive emulation. *American Journal of Sociology*, 107:147–182, 2001.
25. T. W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69–89, 1996.
26. D. Wang, Z. Wen, H. Tong, C.-Y. Lin, C. Song, and A.-L. Barabási. Information spreading in context. In *Proceedings of the 20th international conference on World wide web*, pages 735–744, New York, NY, USA, 2011. ACM.
27. J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270, New York, NY, USA, 2010. ACM.
28. J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, pages 599–608, Washington, DC, USA, 2010. IEEE Computer Society.
29. J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230, New York, NY, USA, 2007. ACM.
30. J. Zhang, Y. Qu, J. Cody, and Y. Wu. A case study of microblogging in the enterprise: use, value, and related issues. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 123–132, New York, NY, USA, 2010. ACM.