Predicting Quitting on Yammer: An Online Social Enterprise Space

Gaurang Gavai, Sricharan Kumar*, Juan Liu, Oliver Brdiczka, and John Hanley
Palo Alto Research Center, Palo Alto, CA, USA
{ggavai,skumar,jjliu,obrdiczka,jhanley}@parc.com

ABSTRACT
The profusion of social media in daily life has accelerated the co-evolution of the online space and the real-world environment. Given this, we try to understand the manner in which negative real-world events manifest into destructive dynamics in online social behavior, and whether it is possible to infer the latter from the former. Modeling destructive dynamics is important for organizations, since unsatisfied employees can exhibit various types of dysfunctional behavior.

To investigate this, we take the novel approach of analyzing a corporate social network, the Yammer data set of a large corporation. In particular, we study how active employees are in using the social network, and in turn determine if people are likely to (i) quit the yammer space, (ii) quit the parent company or, (iii) quit groups within Yammer. In this paper, we take a data-driven approach to identify quitters. To this end, a rich set of features including graph based features, content features and, work practice features are derived from the data set and subsequently used to model quitting behavior. Our experiments show that our proposed data-driven approach can be used to predict user engagement levels within Yammer and the probability of users quitting the corporation and Yammer with fair accuracies of approximately 68% and 82% respectively. The corresponding recall rates are also moderately high at 62% and 87%. Given the difficulty of the quitting prediction problem, these accuracies and recall rates are fairly encouraging.

1. INTRODUCTION
User behavior in online social networks often exhibits a high degree of dynamism. How to describe and model such dynamism is an interesting and important research topic. We refer readers to Easley and Kleinberg’s book, Networks, Crowds, and Markets: Reasoning About a Highly Connected World [1] for an excellent overview. Our research interest is focused on understanding on-line social space dynamics in relation to real-world events. Social dynamics refer to not only the interactions of individuals amongst themselves, but also their interactions with the groups created as a result of these individual interactions being used as building blocks. We question how real-world events manifest into dynamics in online social behavior, and whether it is possible to infer and predict real-world activities from online social dynamics.

1.1 Quitting in online social spaces
In contrast, much of the research on social network evolution so far has focused on user engagement and growth of online communities to come up with models such as social diffusion [2] in which social ties influence people to join a group. Our approach focuses on a complementary problem of user disengagement and groups in decline. The effectiveness of a group can be undermined when members depart. Groups may exhibit various degrees of robustness. Some groups quickly recover from member loss, while others may collapse and vanish over time. The ability to model destructive dynamics and predict quitting behavior is important, as it offers insights on factors that affect online group effectiveness. It also provides practical guidance for modeling risk management and retention strategies.

In this paper we focus on quitting analysis within the Yammer space. Yammer is an enterprise social space, with functionalities similar to Facebook and Twitter, but tailored to the needs of social communication in corporate environments. In this space, we analyze the behavior of users quitting their social groups and/or stopping social activity.

1.2 Quitting in the real-world
The equivalent of users quitting their online social groups in the real-world environment is quitting one’s job, which is a common behavior in today’s workplace. In this paper we are interested in studying the problem of predicting quitting behavior of employees in an organization. In particular, we investigate users behavior in social networks in relation with their real-world quitting events. We are interested in (1) identifying the level of engagement employees exhibit while interacting in enterprise social networks, and (2) identifying users who are likely to quit the organization.

Prediction of quitting may help an organization achieve better employee retention. It is widely accepted that the success
of an organization is crucially predicated on the happiness and well-being of its employees [3]. Unsatisfied employees can result in several problems for the organization ranging from dysfunction of day-to-day operations, to employees taking the extreme measure of quitting which results in dilution of the work-force. Quitting prediction may also inform risk management. Infosecurity Magazine [4] reports that “50% of job leavers are likely to steal confidential company data”. The ability to guard against unwarranted data/information leaks is particularly important in certain sensitive industries where employees handle sensitive, proprietary, trade-secret, personally identifying, or security related information.

Compared to the traditional approach of conducting employee satisfaction surveys, predicting quitting from an online social space offers the benefit of continuous and in-time evaluation. It also circumvents the various issues plaguing the traditional approach such as, the significant overhead of administering an employee satisfaction survey and the problem of surveys being too infrequent. The questions here are rather, how strongly a real-world quitting event correlates with social space data traces, and whether the prediction is reliable.

1.3 Related Work
Modeling destructive dynamics and understanding whether, why, and when a person will depart from an online social group can be important in a number of social applications. Our earlier work [5] examines group stability for two categories of social networks. In the first category, an individual can belong to only one group at any given point of time. For instance, this is the case in World of Warcraft (WoW), a popular massive multi-player online game (MMOG) [6]. An individual who is not satisfied with his/her group (guild in WoW) will quit the group to join another one. Many other MMOGs also adopt a similar exclusive group structure. The second category imposes no restriction on the number of groups an individual belongs to. Facebook and Twitter fall in this category. Individuals do not quit groups even though they may not be active in those groups. For this category we devised a membership score that will reflect participation level of individual members and prolificness/ranking of individual members. Our paper [5] analyzes user quitting behavior in DBLP, a large co-authorship network. We define quitting in this instance to be an author abandoning a conference venue. The same analysis can be easily generalized for a host of social networks with non-exclusive membership group structures.

Attrition has been under intensive investigation by business management, economists, psychologists, and social scientists. Ample broad findings can be found in the literature. For instance, company size, industry and pay scales play a key role in determining attrition rate. Industries that largely employ unskilled labor have a higher rate of attrition as compared to those that largely require skilled labor. Attrition rate is also highest amongst the lowest paying jobs. While these findings provide many qualitative insights, they have not yet reached the level of mathematical predictors that can be deployed to perform churn analysis and prediction. Our earlier work [7] has made a first attempt at using a data-driven approach to study the problem of predicting attrition within an organization. We have constructed models to predict if and when an employee is likely to quit the company using email features. In this paper we extend the analysis to a broader set of data: posts, messages, and group conversations within a much more comprehensive social space.

Another piece of related work is described in [8]. The authors use LinkedIn data to provide job-switching recommendations to users. Their work models a user’s tenure in their current employment with a proportional hazard model, and uses it to decide when to provide job recommendations. In a sense this is the opposite of what we are trying to do.

1.4 Contribution
To the best of our knowledge, this paper is the first attempt to model quitting behavior in social space in relation with real-world events. We use Yammer as the social platform for our investigation to model people’s behavior within the Yammer space, within social groups, and in the real world corporation. We have designed a threading mechanism to reconstruct network structure (described in Section 3). We have developed a privacy preserving feature extraction approach to capture a rich online behavioral dataset free of personally identifiable information.

The feature extraction method is described in Section 4. Our analysis shows that the feature set is well correlated with quitting behaviors in Yammer and in a real world corporation. Furthermore, over a large dataset in Yammer, we have built a model predicting incipient quitting events in the company and in Yammer with moderate accuracies and moderate to high recall rates.

2. YAMMER
Yammer [9] is an enterprise social network service designed to support communication needs within organizations. Like many other social network services such as Facebook and Twitter, users can post a message, “like” other people’s posts, and follow friends. It also supports group communications where a user posts messages only viewable to a group of users. There are several distinct privacy settings. Groups can be public or private, as are messages. Any user can join a public group, but private groups require a group admin to ratify a user before they can access it. Public messages are viewable by anyone within the same Yammer space, while private messages are only viewable by the sender and receiver(s). Compared to Facebook, Twitter, and other popular social network services, Yammer has the key difference that it is organizational, not personal. User participation is synchronized with employment. Creation of a new user profile is validated by the user’s email address within the company. Once an employee quits the company, the user profile is removed from Yammer through a synchronization process.

Figure 1 provides a high-level conceptual sketch of the entities in the Yammer space. The outermost circle is the real-world corporation. Many employees join the corporate Yammer space (the second largest circle) and become Yammer users. Furthermore, users may form groups and participate in group discussions. The Yammer dataset is rich and opens doors to a number of research questions. For instance, from an individual user’s perspective, one can ask: how does a user join, quit, and influence a group? Taking