Multi-relational Influence Models for Online Professional Networks

Arti Ramesh
SUNY Binghamton
artir@binghamton.edu

Mario Rodriguez
LinkedIn Corporation
mrodriguez@linkedin.com

Lise Getoor
University of California, Santa Cruz
getoor@soe.ucsc.edu

ABSTRACT

Professional networks are a specialized class of social networks that are particularly aimed at forming and strengthening professional connections and have become a vital component of professional success and growth. In this paper, we present a holistic model to jointly represent different heterogenous relationships between pairs of individuals, user actions and their respective propagations to characterize influence in online professional networks. Previous work on influence in social networks typically only consider a single action type in characterizing influence. Our model is capable of representing and combining different kinds of information users assimilate in the network and compute pairwise values of influence taking the different types of actions into account. We evaluate our models on data from the largest professional network, *LinkedIn* and show the effectiveness of the inferred influence scores in predicting user actions. We further demonstrate that modeling different user actions, node features, and edge relationships between users leads to around *20%* increase in precision at top k in predicting user actions, when compared to the current state-of-the-art model.

1 INTRODUCTION

The last decade has witnessed the rise of social networks and their prevalence in our everyday lives. Users perform several actions (e.g., browsing content, adding connections, joining groups) and interactions (e.g., sharing/commenting on content, following people) in a social network. Multiple factors affect user actions and interactions in social networks: personal interests, popularity of an action, or social contacts performing the action *influencing* them to perform the same action. Several works in the past have studied the effect of users' actions on their connections in the social network, which they refer to as *influence* [3, 6]. For example, a user witnessing her friends perform a certain action on a social networking site might be influenced into performing the same action herself. Detecting and quantifying influence is a hard but a very useful problem having a number of applications, which include personalized recommendations [13, 14], trust modeling [5, 7, 15, 16], feed ranking [1], and viral marketing [4, 8, 11].

In this work, we focus on a particular class of social networks: online professional networks. While influence has been previously studied in the context of social networks, professional networks present a unique set of opportunities for modeling various user actions that are not readily observed in a regular social network. Professional networks are a specialized class of social networks, which users utilize to form, strengthen, and maintain professional relationships. With professional networks, there are a plethora of actions that users can perform—long-term and life-changing actions such as moving jobs and acquiring new professional skills, to daily activities and interests as observed in any social network. These networks also capture other important edge relationships such as organization hierarchy, relationship strength, and individual's seniority in the profession and the network that could affect the presence and amount of influence between individuals. While most previous works only consider a single action type when modeling influence, the presence of multiple actions, edge relationships, and node features in social networks, particularly in professional networks necessitates more sophisticated models that represent and reason about heterogeneous relationships.

To this end, we develop a holistic model based on *hinge-loss Markov Random Fields (HL-MRFs)* that combines different heterogenous relationships between individuals to learn pair-wise influence probabilities. We show that measuring influence between two users in a network involves meticulously taking into account all user actions and interactions. We demonstrate how to encode multiple action propagations, edge relationships, and node features present in professional networks and compute combined values of influence that integrates many different interactions between users. Our framework can easily be extended to add other node features and edge relationships.

Our contributions in this paper are as follows:

(1) We generate features that take into account the richness of professional networks and capture different kinds of user interactions. We identify four different action types relevant to modeling influence in professional networks and their respective propagations: moving jobs (job propagation), adding a new skill (skill propagation), following content (content propagation), and adding oneself to groups (group propagation). Along with this, we also extract other edge relationships such as organizational hierarchy, strength of relationship, and user's seniority in the network.
(2) We then construct a holistic framework using a recently developed statistical relational learning method, Hinge-loss Markov random fields (HL-MRFs) [2]. We demonstrate how to encode different edge and node relationships that exist in graphs in our framework and combine them efficiently to infer influence values between pairs of individuals in the network. We show that our framework is capable of encoding the rich features in this domain as opposed to

WI '17, Leipzig, Germany

DOI: 10.1145/3106426.3106531

previous efforts that can only encode a single action type. We develop two models: 1) *influence model*: for predicting influence between pairs of individuals, and 2) *influential model*: for predicting influential users in a network. We then develop a third richer and more powerful model by combining the influence and influential models and jointly predicting influence and influential users.

(3) We test our models on data from the professional network, LinkedIn. Our dataset consists of millions of users and millions of actions comprising of four different types of actions: moving jobs, adding skills to LinkedIn profile, following content, and joining groups.

(4) We construct a predictive modeling setup to predict user actions using the influence scores and compare our approach to the state-of-the-art model for inferring influence values. We evaluate precision at top k for predicting user actions and demonstrate that our models are capable of predicting user actions better than the existing approaches for inferring influence values.

2 RELATED WORK

Influence in social networks has mostly been studied in the context of influence maximization. The influence maximization problem is as follows: given a social network with edge influence probabilities of influence, how to select the k set of users that maximize the spread of information in the network? Viral marketing is the most prevalent application of influence maximization where determining the k set of nodes is crucial to maximize marketing. Domingos and Richardson [4], Richardson and Domingos [11] were the first to consider the problem of finding influential users in the network. They follow a data mining approach to understand influence propagation and use that to identify influential users.

Kempe et al. [8] show that the influence maximization problem is NP-complete and derive approximation guarantees for the problem. They obtain provable approximation guarantees on two fundamental propagation models, namely *Linear Threshold Model* and *Independent Cascade Model*. They also prove the equivalence of the Linear Threshold and Independent Cascade models, and propose a generalized framework called the *General Threshold Model (GTM)*. They then develop a greedy approximation algorithm to calculate the spread of influence spread by exploiting the monotonic and submodular nature of influence maximization.

Leskovec et al. [10] study a problem very similar to viral marketing—*outbreak detection*: how to select nodes in a network to detect the spread of a virus? They employ the ideas in viral marketing and the submodular nature of the influence spread to construct an optimization framework to effectively select seed nodes. All the papers discussed above assume the basic framework and propagation models of [8], where the influence probabilities $p_{v,u}$ on the edges are given as input.

Our work is closest to Goyal et al. [6] and Saito et al. [12]'s work on labeling pairs of users with influence probabilities. Goyal et al. focus on the GTM, while Saito et al. focus on the Independent Cascade model of propagation. Goyal et al. use the action log and the connection graph to learn pairwise influence probabilities between users. However, their model for calculating influence probabilities is only capable of taking a single action type into account. In this work, we build on Goyal et al.'s approach to design a holistic model that takes into account various action propagations, other edge relationships between individuals, and node features to compute pairwise influence scores.

3 PROBLEM DEFINITION

Consider a graph G, of the form $G = (V, E)$, where nodes V are users, with time-stamped edges E between pairs of users. $e(u, v, t) \in E$ between users u and v represents the presence of a friendship link in the network between u and v, time-stamped with time t when the connection was made. We also construct an action log by observing the various actions performed by users. The action log is represented by $Action\text{-}log(User, Action, Action\text{-}Type, Timestamp)$, each tuple in the relation representing a user action in the four action types—1) moving jobs, 2) adding a new skill, 3) following content, and 4) joining groups. Note that the first two action types are unique to professional networks. The last two action types are present across the breadth of social networks, including professional networks. Using the action log and the connection graph, we construct an action propagation graph, to capture how users' react to actions performed by their friends in the network. We extend the definition of action propagation in Goyal et al. [6] to account for the different types of actions.

Definition 3.1. An action $a \in A$ of type $\gamma \in \Gamma$ propagates from user u to v, iff: (i) $e(u, v, t) \in E$; (ii) $\exists(u, a, \gamma, \tau_1)$, $(v, a, \gamma, \tau_2) \in$ Action-log with $\tau_1 < \tau_2$; and (iii) $t \le \tau_1$. We refer to the action propagation as prop(a, γ, u, v, $\triangle\tau$), where $\triangle\tau = \tau_2 - \tau_1$.

Here, we define action propagation as two users acting on the same action displaced by time $\tau_2 - \tau_1$. Both users have to perform the action after they connect in the network, given by $t < \tau_1$, for it to be considered an action propagation. Note that users u and v should be connected in the network before either of them perform the action, for it to be considered an action propagation. For skills, content, and groups, same action constitutes adding the same skill, reacting to the same post/article, and joining the same LinkedIn group, respectively. For jobs, we treat same action as joining the same company.

Definition 3.2. For each action type γ, we define an action propagation graph $G_\gamma = (V_\gamma, E_\gamma)$ with unidirectional edges, where $V_\gamma = \{v \mid \exists\,(v, a, \gamma, \tau) \in \textit{Action-log}\}$; there is a directed edge from u to v, $e_\gamma(u, v)$ in E_γ, whenever prop(a, γ, u, v, $\triangle\tau$).

Using the definition of action propagation (3.1), we construct action propagation graphs where the individual edges capture action propagations between pairs of users. Note that the action propagation graphs are directed acyclic graphs (DAGs), due to the strict time constraint in Definition 3.1. We generate four propagation graphs, for the four types of actions. We refer to our propagation graphs as JOB-PROP(u,v), SKILL-PROP(u,v), CONTENT-PROP(u, v), and GROUP-PROP(u, v) to the capture the four different action propagations, respectively.

The problem we address in this work is—how can we combine information from the social connection graph, action propagation graphs, other edge relationships, and node features such as user seniority in the network, and strength of social connection, to create rich models of influence and learn influence values between pairs of users. Section 4 gives more details about our framework and features we use in our models.

4 INFLUENCE PREDICTION MODELS

In this section, we develop rich, multi-relational models using HL-MRFs for modeling influence. We first present an overview of the state-of-the-art models based on GTM and then show how we extend these models by encoding multi-relational edge relationships including the influence values predicted by GTM in our HL-MRF framework.

4.1 General Threshold Model (GTM)

The GTM formulates any user u as either active (already has performed the action), or inactive. Time unfolds in discrete steps and when user u activates, u further can activate other connections of u that are not active yet. Equation 1 gives probability of user u performing an action ($P_u(S)$), using influence values $P_{v,u}$, where $v \in S$, the set of users connected to u, who have already performed the action.

$$P_u(S) = 1 - \prod_{v \in S} (1 - P_{v,u}) \tag{1}$$

Goyal et al. compute the influence values, $P_{v,u}$, via the following three approaches: 1) using maximum likelihood estimation (MLE), 2) using Jaccard index (Jaccard), and 3) using a discrete time variation model (DTM). The MLE model estimates $P_{v,u}$ by calculating the maximum likelihood estimates: ratio of number of successful attempts at influencing over total number of trials. The Jaccard model takes into account users' similarity in calculating the influence probabilities. The discrete time variation model assumes that influence of an active user v on its neighbor remains constant at $P_{v,u}$ for time window of τ after the v performs the action, and drop to 0 after time τ. We defer the reader to [6] for more details.

4.2 Hinge-loss Markov Random Fields (HL-MRFs)

The GTM-based model proposed by Goyal et al., is capable of only examining the effect of a single action type on users. To represent and combine different heterogenous relationships between users, we propose a more powerful approach using HL-MRFs. HLMRFs are a scalable class of continuous, conditional graphical models [2]. HL-MRF models can be specified using *Probabilistic Soft Logic (PSL)* [2], a weighted first order logical templating language. An example of a PSL rule is

$$\lambda : P(a) \wedge Q(a,b) \rightarrow R(b),$$

where P, Q, and R are predicates, a and b are *variables*, and λ is the weight associated with the rule. The weight of the rule indicates its importance in the HL-MRF probabilistic model, which defines a probability density function of the form

$$P(\mathbf{Y}|\mathbf{X}) \propto \exp\Big(- \sum_{r=1}^{M} \lambda_r \phi_r(\mathbf{Y},\mathbf{X}) \Big)$$
$$\phi_r(\mathbf{Y},\mathbf{X}) = (\max\{l_r(\mathbf{Y},\mathbf{X}), 0\})^{\rho_r} \quad , \tag{2}$$

where $\phi_r(\mathbf{Y},\mathbf{X})$ is a *hinge-loss potential* corresponding to an instantiation of a rule, and is specified by a linear function l_r and optional exponent $\rho_r \in \{1,2\}$.

For example, in our influence model, if u and v denote users, predicate JOB-PROP(u, v) denotes the propagation of job from user u to user v in the action propagation graph, and the predicate INFLUENCE(u, v) denotes the target variable denoting the probability of influence of u on v. A PSL rule to encode that job propagation from u to v implies that u influences v is

$$\lambda : \textsc{job-prop}(u,v) \rightarrow \textsc{influence}(u,v).$$

We can generate more complex rules connecting the different features and target variables, e.g.

$$\lambda : \textsc{job-prop}(u,v) \wedge \textsc{manages}(u,v) \rightarrow \textsc{influence}(u,v).$$

This rule encodes that if u propagates job to v and u is the manager of v, then u influences v.

Inference of the most probable explanation in HL-MRFs is a convex optimization problem, which makes working with PSL very efficient in comparison to many relational modeling tools that use discrete representations.

4.3 Feature Engineering

In this section, we develop the node features and pairwise interactions between users in the network.

4.3.1 Action Propagations. We derive four action propagation graphs corresponding to the four different user actions using Definitions 3.1 and 3.2: job propagation (JOB-PROP), skill propagation (SKILL-PROP), content propagation (CONTENT-PROP), and group propagation (GROUP-PROP). For content propagation, we capture if two people act on the same article, and weight the strength of the propagation according to different sub-actions such as viewing, liking, sharing, commenting on content, with commenting or sharing having more weight than liking/viewing. Also note that job propagation and skill propagation are very unique to professional networks, as users tend to specify details related to their professional career on their profile.

4.3.2 Relationship Strength (People You May Know score). We capture the strength of relationship between two users using the *People You May Know* score [9]. The score is part of the people recommendation framework at LinkedIn. This score is a unidirectional score in [0, 1]. In our models, we refer to this score by STRENGTH(u, v).

4.3.3 Manager-managee Relationship. We capture the organization hierarchy information for users within the LinkedIn organization in MANAGES(u, v), where user u is the manager of user v.

4.3.4 Member Seniority score. The predicate SENIORITY(u) captures the reputation of user u within the social network. This is a continuous score in [0, 1].

4.3.5 Content Follower-Followee Score. By considering and appropriately weighting all content-related interactions between pairs of users according to their importance, we generate a score that captures the content-following relationship between individuals.

PSL-Influence Rules
Rules combining action propagations
JOB-PROP(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
JOB-PROP(USER-A, USER-B) $\wedge$ GROUP-PROP(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
GROUP-PROP(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining seniority and action propagation
SENIORITY(USER-A) $\wedge$ SKILL-PROP(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining user influenceability and action propagation
GROUP-PROP(USER-A, USER-B) $\wedge$ INFLUENCEABILITY(USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining propagation and manager-managee relationship
JOB-PROP(USER-A, USER-B) $\wedge$ MANAGES(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining propagation and content follower-followee relationship
CONTENT-PROP(USER-A, USER-B) $\wedge$ CONTENT-FOLLOW(USER-B, USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining GTM influence values
GTM_{group}(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
$\text{GTM}_{group\text{-}mle}$(USER-A, USER-B) $\wedge$ $\text{GTM}_{group\text{-}jaccard}$(USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
GTM_{group}(USER-A, USER-B) $\wedge$ $\text{GTM}_{content}$(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Transitive Rules
GROUP-PROP(USER-A, USER-B) $\wedge$ INFLUENCE(USER-B, USER-C) $\rightarrow$ INFLUENCE(USER-A, USER-C)
CONTENT-PROP(USER-A, USER-B) $\wedge$ INFLUENCE(USER-B, USER-C) $\rightarrow$ INFLUENCE(USER-A, USER-C)

Table 1: Representative rules from PSL-Influence model

PSL-Influential Rules
Rules combining action propagations
JOB-PROP(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
JOB-PROP(USER-A, USER-B) $\wedge$ GROUP-PROP(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
GROUP-PROP(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENTIAL(USER-A)
Rules combining seniority and action propagation
SENIORITY(USER-A) $\wedge$ SKILL-PROP(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
Rules combining propagation and manager-managee relationship
JOB-PROP(USER-A, USER-B) $\wedge$ MANAGES(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
Rules combining propagation and content follower-followee relationship
CONTENT-PROP(USER-A, USER-B) $\wedge$ CONTENT-FOLLOW(USER-B, USER-A) $\rightarrow$ INFLUENTIAL(USER-A)
Rules combining GTM influence values
GTM_{group}(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENTIAL(USER-A)
$\text{GTM}_{group\text{-}mle}$(USER-A, USER-B) $\wedge$ $\text{GTM}_{group\text{-}jaccard}$(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
GTM_{group}(USER-A, USER-B) $\wedge$ $\text{GTM}_{content}$(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)
Rules combining propagation and manager-managee relationship
JOB-PROPAGATION(USER-A, USER-B) $\wedge$ MANAGES(USER-A, USER-B) $\rightarrow$ INFLUENTIAL(USER-A)

Table 2: Representative rules from PSL-Influential model

Likes are weighted less than *comments*, which are in turn weighted less than *shares*. This score is also continuous in [0, 1].

4.3.6 User Influenceability Score. We construct user influenceability score, denoted by INFLUENCEABILITY(u), by examining how easily users are influenced by their connections. This is calculated by taking the ratio of number of actions that were propagated to the user to the total number of actions performed by the user.

4.3.7 GTM Features. We use the influence values computed by Goyal et al. [6] as features in our model. We refer to influence scores obtained using maximum likelihood estimation as GTM_{mle}, using Jaccard index as $GTM_{Jaccard}$, and the discrete time variation of maximum likelihood estimation as GTM_{DT}.

4.4 PSL Influence Models

We present three models of influence: 1) *PSL-Influence*, that combines the action propagation graphs, edge relationships and node features to learn influence values, 2) *PSL-Influential*, that predicts influential users in a network, and 3) *PSL-Combine*, that jointly predicts both influential users and influence values.

PSL-Combine Rules
Rules combining action propagations and influential
JOB-PROP(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
JOB-PROP(USER-A, USER-B) ∧ GROUP-PROP(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
GROUP-PROP(USER-A, USER-B) ∧ SENIORITY(USER-A) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
Rules combining seniority, influential, and action propagation
SENIORITY(USER-A) ∧ SKILL-PROP(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
Rules combining action propagation, user influenceability, and influential
GROUP-PROP(USER-A, USER-B) ∧ INFLUENCEABILITY(USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
Rules combining propagation, manager-managee relationship, and influential
JOB-PROP(USER-A, USER-B) ∧ MANAGES(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
Rules combining propagation and content follower-followee relationship
CONTENT-PROP(USER-A, USER-B) ∧ CONTENT-FOLLOW(USER-B, USER-A) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-B)
Rules combining GTM influence values and influential
GTM_{group}(USER-A, USER-B) ∧ SENIORITY(USER-A) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-C)
$\text{GTM}_{group\text{-}mle}$(USER-A, USER-B) ∧ $\text{GTM}_{group\text{-}jaccard}$(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-C)
Rules combining propagation, manager-managee relationship, and influential
JOB-PROPAGATION(USER-A, USER-B) ∧ MANAGES(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ⟶ INFLUENCE(USER-A, USER-C)
Transitive Rules
GROUP-PROP(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ∧ INFLUENCE(USER-B, USER-C) ⟶ INFLUENCE(USER-A, USER-C)
CONTENT-PROP(USER-A, USER-B) ∧ INFLUENTIAL(USER-A) ∧ INFLUENCE(USER-B, USER-C) ⟶ INFLUENCE(USER-A, USER-C)

Table 3: Representative rules from PSL-Combine model

4.4.1 PSL-Influence. We construct weighted logical rules to combine the different features and encode dependencies among them to infer influence values. Table 1 gives representative rules from six different combinations of predicates in our PSL-Influence model. INFLUENCE(u, v) captures the value of influence for pairs of users. The weights are manually encoded, taking into account the importance of the feature or combination of features. The rules combine various edge relationships and node features together to predict influence. The rules fall under the six categories mentioned below.

Combining Action Propagations. Here, we combine various action propagations to infer influence values between pairs of users in the network. We capture that each propagation signifies the presence of influence between two individuals in the network. We take advantage of the possibility to capture complex dependencies in HL-MRFs and encode that combination of action propagations between the same two individuals leads to a stronger influence between them. For example, the first rule specifies that if USER-A propagates job to USER-B, then USER-A influences USER-B. The second rule builds on the first rule by combining group propagation and job propagation.

Combining Node Features and Action Propagations. Combining node features such as SENIORITY and INFLUENCEABILITY with action propagations provide a stronger signal for influence between individuals. For example, in the second set of rules in Table 1, we capture that a person more senior in the network has a higher possibility of influence on users she has propagated action(s) to in the network. Similarly, combining user influenceability score and action propagations, we encode that influenceable users are more susceptible to action propagations from their connections.

Combining Action Propagations and Edge Relationships. In the fourth set of rules in Table 1, we capture that a user's manager is more likely to wield an influence on her reports. In the fifth set of rules, we encode the dependence between action propagations and content follower-followee relationship. While the content following relationship is more relevant for content propagations, we also capture its dependence with other propagations, as given by the second rule in the fifth set. It is important to note that the rules are weighted. By weighting these rules appropriately, we encode their respective effects of propagation on influence.

Combining GTM Influence scores. In the sixth set of rules, we combine inferred influence values from GTM_{MLE}, $GTM_{Jaccard}$, and GTM_{DT} to eliminate uncertainty and strengthen the GTM scores. For example, in the second rule we combine $GTM_{group\text{-}mle}$ and $GTM_{group\text{-}jaccard}$ to infer influence. We also combine the GTM with other edge relationships such as MANAGES and STRENGTH, and node features such as SENIORITY to infer influence.

Transitive Rules. In the seventh and final set of rules, we capture the propagation of influence values using the transitive property. The first rule that if USER-A propagates an action to USER-B and USER-B influences USER-C, then USER-A influences USER-C. These transitive rules help predict influence values between pairs of users without directly observing propagations between them.

4.4.2 PSL-Influential. The PSL-Influential model summarizes the edge scores for influencer nodes to measure how influential a person is in the network. This is particularly useful in determining the top influencers in the network, which has many uses in viral marketing and information diffusion. The predicate to determine if a user is influential is given by *Influential(user).* Table 2 gives the rules in the PSL-Influential model for inferring *infuential* users. The rules are grouped into five categories. The categories are similar to the PSL-Influence model, except for rules involving INFLUENCEABILITY and transitive rules, which are not relevant for the PSL-Influential model. For example, consider rule 1 in Table 2. This captures that if user A propagates job to user B, A is an influential user. When this rule is grounded using data from the network, all users whom A has propagated a job to are considered. With the effect, the more number of users A propagates job to, the more influential A is in the network. Similarly, if a user propagates multiple actions to other users, then the user is more influential. Also, it is important to note that apart from action propagations, features such as hierarchical relationship between users inside an organization, their connection strength and seniority play an important role in determining influential users, which are captured in the following sets of rules in Table 2.

4.4.3 PSL-Combine. The PSL-Combine model combines both the PSL-Influence models and PSL-Influential models and uses that to jointly infer both influence values and influential users in the network. Table 3 gives the rules that combine influential and influence variables. In addition to these rules, PSL-Combine also has rules from PSL-Influence and PSL-Influential models for inferring influence and influential values, respectively. As can be evidenced in Table 3, the rules capture dependencies between other features and influential variable to infer influence values. The rules are grouped into the same seven categories as the PSL-Influence model. For example, the first rule in Table 3 captures that if A propagates a job to B and A is an influential person in the network, then A has a higher influence on B. It is important to note that influential scores, together with the influenceability scores create possibilities for modeling characteristics of both influencer and the person influenced to create richer and more meaningful influence models, as captured in the third set of rules in Table 3.

5 EXPERIMENTAL RESULTS

In this section, we conduct experiments to: 1) evaluate the effectiveness of the computed influence values, and 2) interpret influence values and use them to understand social interactions in the network.

5.1 Dataset

We test our models on data from the professional social networking site, LinkedIn. LinkedIn is the world's largest professional networking site, enabling users to make professional connections and search for jobs. LinkedIn users have a profile page, where they can enlist their education, professional experiences, and skills. LinkedIn also has a feed customized for each user, which captures the highlights of their connections' activities. LinkedIn users can also create and join groups.

5.2 Predicting Actions using Influence scores

First, we evaluate the the effectiveness of the influence scores by using them to predict user actions of joining groups and following content. Due to the unavailability of labeled influence scores, we devise a prediction task using the influence scores inferred by our models. In order to compare to the GTM models, we use Equation 1 used by Goyal et al. to calculate the probability of user performing an action. As the influence scores given by the PSL models are in (0, 1), they can be substituted in place of $P_{v,u}$ in Equation 1 to calculate P_u, the probability of user u performing an action. We compare our models: PSL-Influence and PSL-Combine to models based only on GTM. In the sections below, we furnish results from two prediction tasks: 1) predicting *joining group* action, and 2) predicting *following content* action. For both these tasks, we consider the subset of users comprising of employees at LinkedIn and their social connections. We split the data into training and test based on actions and use 90% of data for training and 10% for testing. Table 4 gives the group and content action prediction results. Our test dataset (10% of data) has user-action pairs in the order of millions, around hundreds of thousands of users and tens of thousands of actions for both these actions. Statistically significant differences, evaluated using a paired t-test with a rejection threshold of 0.01, are typed in bold in all tables.

5.2.1 Group Action Prediction. For group action prediction, we consider users joining groups in the last five years. We evaluate the models by measuring if the user performs an action in the top k predictions generated by the model. We consider k = *15, 10, 5, and 3* respectively. Table 4a gives the precision at top k for the GTM and PSL models. We observe that both our models: PSL-Influence and PSL-Combine, outperform the GTM models. PSL-Combine achieves the best possible results, outperforming the best GTM model by 20%, which confirms that jointly predicting influential users and influence and incorporating the effect of influential users in predicting influence values helps in improving performance.

5.2.2 Content Action Prediction. For content action prediction, we consider content following actions within the past 100 days. We consider all content-related actions on a single piece of content: like, share, and comment, and treat them equally. Table 4b gives the content action prediction results. Similar to group action prediction, we find that PSL-Influence and PSL-Combine outperform the GTM models in predicting content actions. Again, PSL-Combine achieves the best results, outperforming the best GTM model by 20%.

5.2.3 Highly Influenceable and Influential Users. Our influenceability scores identify how susceptible a user is to influence. Similarly, the influential scores identify how influential a user is in the network. It is evident that both these scores are crucial in modeling influence. In the next set of experiments, we filter users based on the influenceability and influential values and retain only users with values greater than 0.5 for both these predicates. This helps us focus on the key players in the network: highly influential and highly influenceable users, and model influence between these key players in the network. Tables 4c and 4d give the results for predicting group and content actions in these users. We observe that removing less influential and influenceable users from the network helps PSL-Influence and PSL-Combine achieve a higher precision at top k in both prediction tasks. PSL-Combine achieves the best results for both group and content action prediction, significantly outperforming the best GTM model by 33% in group action prediction and 38% in content action prediction.

Models	top 15	top 10	top 5	top 3
GTM-MLE	14.60	14.60	14.53	14.22
GTM-Jaccard	15.30	15.10	14.49	14.10
GTM-DT	15.68	13.56	13.21	13.09
PSL-Influence	16.76	16.67	14.96	13.32
PSL-Combine	**19.01**	**18.89**	**15.83**	**13.33**

(a) Precision at top k for predicting users joining groups

Models	top 15	top 10	top 5	top 3
GTM-MLE	13.45	13.30	12.53	10.90
GTM-Jaccard	15.48	15.09	13.46	13.01
GTM-DT	16.78	15.66	13.45	12.22
PSL-Influence	18.01	17.86	16.65	16.04
PSL-Combine	**20.22**	**20.12**	**17.66**	**17.01**

(b) Precision at top k for predicting users following content

Models	top 15	top 10	top 5	top 3
GTM-MLE	30.96	26.85	18.95	16.11
GTM-Jaccard	35.97	35.78	33.47	27.15
GTM-DT	36.30	36.08	35.50	23.70
PSL-Influence	39.34	39.21	38.72	37.65
PSL-Combine	**48.45**	**46.24**	**45.51**	**45.28**

(c) Precision at top k for predicting users joining groups for INFLUENCEABILITY(U) > 0.5, INFLUENTIAL(U) > 0.5

Models	top 15	top 10	top 5	top 3
GTM-MLE	28.56	23.53	17.61	16.12
GTM-Jaccard	36.97	34.58	33.74	26.95
GTM-DT	35.30	33.08	32.77	31.71
PSL-Influence	41.45	41.02	40.05	37.34
PSL-Combine	**48.93**	**47.85**	**44.23**	**40.81**

(d) Precision at top k for predicting users following content for INFLUENCEABILITY(U) > 0.5, INFLUENTIAL(U) > 0.5

Table 4: Precision at top k for GTM models, PSL-Influence, and PSL-Combine for predicting user actions

5.3 Interpreting Influence scores

The influence scores given by our models help in understanding the influence a person has on others. Our experiments in Section 5.2 demonstrate that the influence scores can be very useful in predicting user actions. However, the scores themselves carry weight, as they bring out the strength of connections in the social network and also can be helpful in a number of applications such as personalization, recommendations, and ranking relevant content. In this section, we present qualitative results of understanding the scores and comparing them to other edge relationships that can exist in the network.

Two other edge relationship scores that are worth comparing with the influence scores are *relationship-strength* scores, and *organization hierarchy*. We compare the *influence* scores to both these scores to see how the influence scores between the same pair of individuals are different. Around 12% of times, the influence flows in the reverse direction when compared to the manages relationship, i.e., if User A is User B's manager, then the influence is in the opposite direction User B to User A. In such cases, we find that the employee is often more active in the network, contributing to more actions, which are reciprocated by managers. In around 20% cases, influence between individuals in the same organization is characterized by peers. This verifies how influence relationships do not always flow from top to bottom in an organization.

Comparing *influence* scores to people-you-may-know scores, we find that in about 10% of cases, the influence flows in opposite direction to relationship strength. For example, if User A and User B are connected in a network and STRENGTH(A, B) > STRENGTH(B, A), in 10% of cases, INFLUENCE(A, B) < INFLUENCE(B, A), and vice-versa.

6 CONCLUSION

In this paper, we presented a framework to model influence in rich behavioral settings, such as online social networks, by examining multiple edge and node relationships. We evaluated the model on the LinkedIn professional network and examined the effect of different long-term actions (such as moving jobs) and short-term actions (sharing content) on influence. Our system can be easily extended to more edge relationships, node features and more action types or contexts. There are many exciting directions to go: can we use influence scores in one context to predict influence in other types of actions? Our system can also be extended to combine coarse and fine grained interactions between users and to infer action-specific top influencers in the network to make more personalized recommendations.

REFERENCES

[1] Deepak Agarwal, Bee-Chung Chen, Rupesh Gupta, Joshua Hartman, Qi He, Anand Iyer, Sumanth Kolar, Yiming Ma, Pannagadatta Shivaswamy, Ajit Singh, and Liang Zhang. 2014. Activity Ranking in LinkedIn Feed. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).*

[2] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. 2015. Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. arXiv:1505.04406 [cs.LG] (2015).

[3] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM).*

[4] Pedro Domingos and Matt Richardson. 2001. Mining the Network Value of Customers. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).*

[5] Jennifer Golbeck and James Hendler. 2006. Inferring Binary Trust Relationships in Web-based Social Networks. *ACM Trans. Internet Technol.* 6, 4 (Nov. 2006), 497–529.

[6] Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. 2010. Learning Influence Probabilities in Social Networks. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM).*

[7] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of Trust and Distrust. In *Proceedings of the International Conference on World Wide Web (WWW).*

[8] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the Spread of Influence Through a Social Network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD).*

[9] Pei Lee, Laks V. S. Lakshmanan, Mitul Tiwari, and Sam Shah. 2014. Modeling impression discounting in large-scale recommender systems. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

[10] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective Outbreak Detection in Networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).*

[11] Matthew Richardson and Pedro Domingos. 2002. Mining Knowledge-sharing Sites for Viral Marketing. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

[12] Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. 2008. Prediction of Information Diffusion Probabilities for Independent Cascade Model. In *Proceedings of the International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES)*.

[13] Xiaodan Song, Yun Chi, Koji Hino, and Belle L. Tseng. 2007. Information Flow Modeling Based on Diffusion Rate for Prediction and Ranking. In *Proceedings of the International Conference on World Wide Web (WWW)*.

[14] Xiaodan Song, Belle L. Tseng, Ching-Yung Lin, and Ming-Ting Sun. 2006. Personalized Recommendation Driven by Information Flow. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*.

[15] Mohsen Taherian, Morteza Amini, and Rasool Jalili. 2008. Trust Inference in Web-Based Social Networks Using Resistive Networks. In *Proceedings of the International Conference on Internet and Web Applications and Services (ICIW)*.

[16] Cai-Nicolas Ziegler and Georg Lausen. 2005. Propagation Models for Trust and Distrust in Social Networks. *Information Systems Frontiers* 7, 4-5 (Dec. 2005), 337–358.