Understanding Influence in Online Professional Networks

Arti Ramesh[1], **Mario Rodriguez**[2], **Lise Getoor**[3]
[1]University of Maryland, College Park
[2]LinkedIn Corporation [3]University of California, Santa Cruz
artir@cs.umd.edu, mrodriguez@linkedin.com, getoor@soe.ucsc.edu

Abstract

Social networks have become part and parcel of our lives. With social networks, users have access to tremendous amount of information that *influence* many aspects of their lives such as daily activities, habits, and decisions. Recently, there has been a growing interest in understanding influence in social networks. Previous work in this area characterize influence as propagation of actions in the social network. However, typically only a single action type is considered in characterizing influence. In this paper, we present a holistic model to jointly represent different user actions and their respective propagations in the social network. Our model captures node features such as user seniority in the social network, and edge features such as connection strength to characterize influence. Our model is capable of representing and combining different kinds of information users assimilate in the social network and compute pairwise values of influence taking the different types of actions into account. We evaluate our models on data from *LinkedIn* and show the effectiveness of the inferred influence scores in predicting user actions. We further demonstrate that modeling different user actions, node and edge relationships between people leads to around *20 %* increase in precision at top *k* in predicting user actions, when compared to a model based only on General Threshold 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* [7, 3]. 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 [18, 19], trust modeling [8, 21, 6, 20], feed ranking [1], and viral marketing [4, 16, 11].

Our work is closest to Goyal et al. [7], who use the action log and the connection graph to learn pairwise influence probabilities between users. Their model is an instance of the General Threshold Model (GTM) [11] for modeling influence propagation in networks. However, their model for calculating influence probabilities only takes a single action type into account. For example, in their evaluation on Flickr social network, they consider only the action of users joining groups. They also do not consider other edge relationships such as organization hierarchy, relationship strength, and individual's seniority in the network that could affect the presence and amount of influence between individuals. Therefore, in this work, we build on Goyal et al.'s approach to design a holistic model that takes into account various action propagations, and other edge relationships between individuals to compute pairwise influence scores.

Our framework based on *hinge-loss Markov Random Fields (HL-MRFs)* combines different heterogenous relationships between individuals to learn influence probabilities. We demonstrate how to encode multiple action propagations, edge relationships, and node features in social networks and use that to learn a combined value of influence that integrates many different interactions between users. We show that influence probabilities between users is a measure of social influence a person exerts on another person in the network and calculating them involves meticulously taking into account all user actions and interactions. Our framework can easily be extended to add other node and edge relationships.

Our main contributions are as follows:

1. We construct a holistic framework capable of encoding multiple pairwise interactions between individuals using a recently developed statistical relational learning method, Hinge-loss Markov random fields (HL-MRFs). We demonstrate how to encode different edge and node relationships that exist in graphs and combine them efficiently to infer influence.
2. We test our models on data from the professional social network, LinkedIn. We generate features that take into account the richness of the dataset and capture different kinds of user interactions. We show that our framework is capable of encoding the rich features in this domain as opposed to previous efforts that can only encode a single action type. Our dataset consists of millions of users and millions of actions comprising of four different types of actions: joining groups, following content, moving jobs, and adding skills to LinkedIn profile.
3. We compare our approach to the state-of-the-art for inferring influence values that extend GTM, using a predictive modeling setup for predicting user actions. 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 [16] 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. [11] 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 influence spread by exploiting the monotonic and submodular nature of influence spread.

Leskovec et al. [13] 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 [11], where the influence probabilities $p_{v,u}$ on the edges are given as input.

Goyal et al. [7] and Saito et al. [17]'s work on labeling pairs of users with influence probabilities are most similar to our work. Goyal et al. focus on the GTM, while Saito et al. focus on the Independent Cascade model of propagation. Goyal et al. derive influence probabilities using the action log and the connection graph. In our work, we extend Goyal et al.'s model by combining multiple action propagations to infer influence probabilities.

3 Problem Definition

Consider a graph G, of the form $G = (V, E, T)$, 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 social network link between u and v, time-stamped with time t when the connection was made. In addition to the social network, we construct an action log by observing the various actions performed by users. Each entry in the action log identifies a single action by the user. We classify user actions into four broad types—1) joining groups, 2) following content, 3) moving jobs, and 4) adding a new skill. The action log is a relation $Actions(User, Action\text{-}Type, Action, \tau)$, each tuple in the relation representing a user action in the four categories mentioned above. For instance, (u, $group$, $group\text{-}id$, τ) captures that user u joined group $group\text{-}id$ at time τ.

Using the action log and the connection graph, we can construct an action propagation graph, to capture propagation of actions in the network. The action propagation graphs capture how users' react to actions performed by their connections. Our definition of action propagation is very similar to Goyal et al. [7], except that we add an additional term a_t to identify the action-type.

DEFINITION 1. *An action $a \in A$ of type $a_t \in A_t$ propagates from user v_i to v_j, iff: (i) (v_i, v_j) $\in$ E; (ii) $\exists$(v_i, a_t, a, τ_i), (v_j, a_t, a, τ_j) $\in$ Actions with $\tau_i < \tau_j$; and (iii) $T_{v_i,v_j} \leq \tau_i$. We refer to the action propagation as prop(a, a_t, v_i, v_j, $\triangle\tau$).*

Note that users v_i and v_j should be connected in the social network before either of them perform the action, for it to be considered an action propagation. Using the action propagations, a propagation graph can be constructed for each of the action types mentioned above.

DEFINITION 2. *For each action a of type a_t we define an action propagation graph PG(a, a_t) = (V, E) with unidirectional edges. V= $\{v \mid \exists\tau$: (v, a_t, a, τ) $\in$ Actions$\}$; there is a directed edge between $v_i \rightarrow v_j$ in E, whenever prop(a, a_t, v_i, v_j, $\triangle\tau$).*

Note that we generate four propagation graphs, for the four types of actions. We refer to our propagation graphs as GROUP-PROP(v_i, v_j), CONTENT-PROP(v_i, v_j), JOB-PROP(v_i, v_j), and SKILL-PROP(v_i, v_j), respectively. We utilize the propagation graphs as features in our model. Section 4 gives more details about the action propagation features.

The problem we address in this work is—how can we combine information from the social connection graph and the action propagation graphs and other node and edge relationships in social graphs such as user seniority in the network, and strength of social connection, to label values of influence for pairs of individuals in the network. For achieving this, we explore HL-MRFs. Section 4 gives more details about our framework and features we use in our models.

4 Influence Prediction Models

In this section, we first present an overview of GTM and then develop an HL-MRF framework incorporating features from GTM for predicting influence.

4.1 General Threshold Model (GTM)

The GTM formulates any user *u* as either active (already an adopter, in the case of actions, already has performed the action), or inactive. The user *u* is more likely to perform an action when more connections become active, given by the monotonic nature of the activation function. 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 $v \in$ set S of users, who have already performed the action.

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

Goyal et al. [7]'s model is an instance of GTM. They compute $P_{v,u}$ via the following three approaches: 1) using maximum likelihood estimation, 2) using Jaccard index, and 3) using a discrete time variation model. 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 $\tau_{v,u}$ after the v performs the action, and drop to 0 after $\tau_{v,u}$. More details are available in [7].

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

The GTM 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-MRFs have achieved state-of-the-art performance in many domains including knowledge graph identification [14], understanding engagement in MOOCs [15], biomedicine and multi-relational link prediction [5], and modelling social trust [9].

4.2.1 Probabilistic Soft Logic

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

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

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, then we have predicates JOB-PROP(U, V) to denote the propagation of job from user *U* to user *V* in the action propagation graph, and INFLUENCE (U, V) is the target variable denoting the probability of influence of *U* on *V*. A PSL rule to encode that job propagation from U to V suggest that U influences V is

$$\lambda : \text{JOB-PROP}(U, V) \rightarrow \text{INFLUENCE}(U, V).$$

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

$$\lambda : \text{JOB-PROP}(U, V) \wedge \text{MANAGES}(U, V) \rightarrow \text{INFLUENCE}(U, V).$$

This rule encodes that if user U propagates job to user V and user U manages user V, then user U influences user V. These rules can be weighted according to their importance using domain knowledge expertise. The HL-MRF model uses these rules to encode domain knowledge about dependencies among the predicates. 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 features in our influence models that capture user pairwise interactions and relationships between individuals in a network.

Action Propagations We derive action propagation graphs according to the definition in Section 3 for four types of user actions on the site: 1) joining groups, 2) following content, 3) moving jobs/companies, and 4) editing profile, particularly updating skills in the profile. We refer to them as *group propagation*, *content propagation*, *job propagation*, and *skill propagation*, respectively. These features are computed using Definitions 1 and 2. We extract features from the action propagation graphs for these four actions as follows.

If there exists an edge in the action propagation graph for users U and V, then, value of PROPAGATION = 1, else 0. Following this, we generate the features: JOB-PROP, GROUP-PROP, CONTENT-PROP, and SKILL-PROP from the action propagation graphs. For content propagation, we only capture if two people act on the same article, and do not differentiate between different kinds of sub-actions such as liking, sharing, commenting on content.

We determine the sequential nature of the actions, by looking at the time difference between the users making the same action. For jobs, we use the date in users' profile associated with the job rather than using the timestamp when the update was made as users sometimes do not update their positions exactly when they start. To eliminate any uncertainty around propagations, we measure the time difference in months for job propagation. For groups and skills, we measure the time difference in days/minutes, and for content, the time difference is measured in minutes/seconds.

Relationship Strength (People You May Know score) We capture the strength of relationship between two users using the *People You May Know* score [10, 12]. 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).

Manager-managee Relationship For employees within LinkedIn, we have the manager-managee relationships available via an internal portal. The predicate MANAGES(U, V) captures the manager-managee relationship in the model, where user U is the manager of user V.

Member Seniority score We use member seniority scores indicating the popularity and reputation of the member in LinkedIn. The predicate SENIORITY(U) captures the seniority of user U within the social network. This is a continuous score in [0, 1].

Content Follower-Followee Score Similar to the relationship score, we can also generate a score for a user following another user's content. This is done by weighting all interactions involving content between two users. Each action has a score according to its importance. For example, *likes* are weighted less than *comments*, which are in turn weighted less than *shares*. This score also is a continuous score in [0, 1]. The People You May Know score, Seniority score and the Content Follower-Followee score are scores part of existing prediction models at LinkedIn.

User Influenceability Score Following Goyal et al. [7], we construct user influenceability score INFL(USER) for users based on how easily they can be influenced by their connections. This is calculated by taking the ratio of number of actions that were propagated to the user and total number of actions performed by the user.

GTM Features We use the influence values computed by Goyal et al. [7] in their GTM framework 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

4.4.1 PSL-Influence

We construct weighted logical rules to encode dependencies between the features described in Section 4.3 to infer influence. INFLUENCE(U, V) gives the value of influence for pairs of users. The weights in our models are manually specified, taking into account the importance of the feature or combination of features. Table 1 gives some representative rules from our PSL-Influence model. The table gives six different combinations of predicates from our PSL-Influence model. The rules combine various edge and node features together to reason about influence. 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 adding group propagation to job propagation. It specifies that if USER-A propagated both job and group to USER-B, then, USER-A influences USER-B. By weighting these rules appropriately, we can combine the effects of propagation on influence.

Similarly, we use seniority of user along with the propagation graphs to encode that a user who is senior is more likely to influence other users. For employees within LinkedIn, we also have the MANAGES relationship and we use that along with the propagation graphs to encode that managers usually have an influence on their reports. Since the rule is weighted, it does not mandate that influence relationships should follow manager-managee relationship, but helps to also identify influence that flows from employees to their managers.

Combining user influenceability score and action propagations, we can model that influenceable users are more susceptible to action propagations from their connections. We also incorporate the influence scores from Goyal et al.'s model (GTM features), and combine them with seniority scores to infer influence. Also, our framework combines together different inferred influence values from $GTM_{group\text{-}mle}$ and $GTM_{group\text{-}jaccard}$, to eliminate uncertainty and strengthen the scores. The last two rules in our model capture propagation of influence—if USER-A propagates an action to USER-B and USER-B influences USER-C, then USER-A influences USER-C.

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 user influenceability and action propagation
GTM$_{group}$(USER-A, USER-B) $\wedge$ INFL(USER-B) $\rightarrow$ 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)
GTM$_{group\text{-}mle}$(USER-A, USER-B) $\wedge$ GTM$_{group\text{-}jaccard}$(USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
GTM$_{group}$(USER-A, USER-B) $\wedge$ GTM$_{content}$(USER-A, USER-B) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining seniority and relationship strength
RELATIONSHIP-STRENGTH(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENCE(USER-A, USER-B)
Rules combining propagation and manager-managee relationship
GROUP-PROP(USER-A, USER-B) $\wedge$ MANAGES(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

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 social 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 model for inferring *infuential* users. If a user propagated multiple actions to other users, then the user is more influential. Also, it is important to notice that apart from action propagations, features such as hierarchical relationship between users inside organization, their connection strength and seniority play an important role in determining influential users. In Section 5, we show how we use the influential scores to filter users and improve the influence scores to make more informed predictions. Influential scores, together with the influenceability scores create possibilities for modeling characteristics of both influencer and the person influenced to create more meaningful influence models.

5 Experimental Results

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

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 GTM influence values
GTM_{group}(USER-A, USER-B) $\wedge$ SENIORITY(USER-A) $\rightarrow$ INFLUENTIAL(USER-A)
$\mathrm{GTM}_{group\text{-}mle}$(USER-A, USER-B) $\wedge$ $\mathrm{GTM}_{group\text{-}jaccard}$(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

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, which enables making professional connections and also aids in job search. 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 run experiments to evaluate the effectiveness of the influence scores. Since there are no true labels for evaluating the influence values, we use them to predict user actions of joining groups and following content. We compare PSL-Influence models to a model based only on GTM.

We consider the subset of users comprising of employees at LinkedIn and their social connections. For groups, we consider actions in the last five years. For content, we consider actions from last 100 days. We split the data into training and test based on actions and use 90% of data for training and 10% for testing. For the groups data, our test dataset has user-action pairs in the order of millions, around hundreds of thousands of users and tens of thousands of actions. For the GTM models, the parameters $P_{v,u}$, and $\tau_{v,u}$ are calculated at training time. At test time, $P_u(S)$ is calculated using Equation 1. For the PSL models, we substitute values from INFLUENCE(V,U) predicate in place of $P_{v,u}$ in Equation 1 to predict user actions. 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. Tables 3 and 4 give the precision at top k for GTM and PSL models. The PSL-Influence model performs better than the GTM models in predicting if the users will perform the action.

Further, we use the *influential* scores given by our models to filter influential users and only consider their influence on users. We rank users' connections using the influential scores and retain influencers with influential score greater than 0.5. This is given by *PSL-Influence (Influential Users)* in Tables 3 and 4. Retaining only influential users in the prediction further improves action prediction scores. Statistically significant differences, evaluated using a paired t-test with a rejection threshold of 0.01, are typed in bold.

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.

Models	*top 15*	*top 10*	*top 5*	*top 3*
GTM-MLE	14.60	14.60	14.53	14.22
GTM-Jaccard	15.30	15.1	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-Influence (Influential users)	**19.01**	**18.89**	**15.83**	**13.33**

Table 3: Precision at top k for GTM models, PSL-Influence, and PSL-Influential 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-Influence (Influential users)	**20.22**	**20.12**	**17.66**	**17.01**

Table 4: Precision at top k for GTM models, PSL-Influence, and PSL-Influential for predicting users following content

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 PYMK 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 $\text{STRENGTH}(\text{A, B}) > \text{STRENGTH}(\text{B, A})$, in 10% of cases, $\text{INFLUENCE}(\text{A, B}) < \text{INFLUENCE}(\text{B, A})$, and vice-versa.

6 Discussion

In this paper, we present preliminary work on understanding influence in rich behavioral settings, such as online social networks, by examining multiple edge and node relationships. 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 influence scores can also be used to recommend feed content for users, which primarily consists of the four action types that we consider. Using our influence and influential models, we can generate more meaningful ranking of feed content, by taking into account the top influencers for each person. 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. Activity ranking in linkedin feed. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, 2014.

[2] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. 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. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, WSDM, 2011.

[4] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, 2001.

[5] Shobeir Fakhraei, Bert Huang, Louiqa Raschid, and Lise Getoor. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014.

[6] Jennifer Golbeck and James Hendler. Inferring binary trust relationships in web-based social networks. *ACM Trans. Internet Technol.*, 6(4):497–529, November 2006. ISSN 1533-5399.

[7] Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the International Conference on Web Search and Data Mining*, WSDM, 2010.

[8] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the International Conference on World Wide Web*, WWW, 2004.

[9] Bert Huang, Angelika Kimmig, Lise Getoor, and Jennifer Golbeck. A flexible framework for probabilistic models of social trust. In *The 2013 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP 2013)*, 2013.

[10] Xinyi (Lisa) Huang, Mitul Tiwari, and Sam Shah. Structural diversity in social recommender systems. In *Proceedings of the RecSys Workshop on Recommender Systems and the Social Web*, 2013.

[11] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, 2003.

[12] Pei Lee, Laks V. S. Lakshmanan, Mitul Tiwari, and Sam Shah. Modeling impression discounting in large-scale recommender systems. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014.

[13] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, 2007.

[14] Jay Pujara, Hui Miao, Lise Getoor, and William Cohen. Knowledge graph identification. In *International Semantic Web Conference (ISWC)*, 2013.

[15] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. Learning latent engagement patterns of students in online courses. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014.

[16] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, 2002.

[17] Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. Prediction of information diffusion probabilities for independent cascade model. In *Proceedings of the International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Part III*, KES, 2008.

[18] Xiaodan Song, Belle L. Tseng, Ching-Yung Lin, and Ming-Ting Sun. Personalized recommendation driven by information flow. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR, 2006.

[19] Xiaodan Song, Yun Chi, Koji Hino, and Belle L. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proceedings of the International Conference on World Wide Web*, WWW, 2007.

[20] Mohsen Taherian, Morteza Amini, and Rasool Jalili. Trust inference in web-based social networks using resistive networks. In *Proceedings of the International Conference on Internet and Web Applications and Services*, ICIW, 2008.

[21] Cai-Nicolas Ziegler and Georg Lausen. Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, 7(4-5):337–358, December 2005.