

Modeling Mutual Influence Between Social Actions and Social Ties

Xiaofeng YU Junqing XIE

HP Labs China
Universal Business Park
10 Jiu XianQiao Road, Chaoyang District, Beijing, China
{xiaofeng.yu, jun-qing.xie}@hp.com

Abstract

In online social media, social action prediction and social tie discovery are two fundamental tasks for social network analysis. Traditionally, they were considered as separate tasks and solved independently. In this paper, we investigate the high correlation and mutual influence between social actions (i.e. user-behavior interactions) and social ties (i.e. user-user connections). We propose a unified coherent framework, namely mutual latent random graphs (MLRGs), to flexibly encode evidences from both social actions and social ties. We introduce latent, or hidden factors and coupled models with users, users' behaviors and users' relations to exploit mutual influence and mutual benefits between social actions and social ties. We propose a gradient based optimization algorithm to efficiently learn the model parameters. Experimental results show the validity and competitiveness of our model, compared to several state-of-the-art alternative models.

1 Introduction

With the dramatically rapid growth and great success of many large-scale online social networking services, social media bridge our daily physical life and the virtual Web space. Popular social media sites (e.g., Facebook and Twitter) and mobile social networks (e.g., Foursquare) have gathered billions of acting users and are still attracting millions of newbies everyday. Modeling *social actions* and *social ties* are two fundamental tasks in online social media. Social actions are the users' activities or behaviors in socially connected networks. For example, a social action can be "posting a tweet" on Twitter or the "check-in" behavior on Foursquare. A social tie or social relation is referred to any relationship between two or more individual users in a social network, such as the friend and colleague relationships. By understanding a user's behaviors and accordingly exploiting potentially interesting services to her/him, one can improve the user's experience and boost the revenue of social media sites. Also, precise social tie prediction will help people tap into the wisdom of crowds, to aid in making more informed decisions.

Since individual users are socially connected, social influence occurs through information diffusion in social networks. Social influence happens when one's opinions or behaviors are affected by others. It is well known that different types of social ties have essentially different influence on social actions. Intuitively, a user's trusted friends on the web affect that user's online behavior. Ma *et al.* (2009) and Ma *et al.* (2011) claimed that one user's final behavior decision is the balance between his/her own taste and her/his trusted friends' favors. On the other hand, social actions also have important influence on social ties. Obviously, users with similar preferences or behaviors are more likely to be friends than others in social media. Users with momentous activities will attract many other users to be connected with. On the contrary, no body will be interested in users with trivial or insignificant behaviors.

Consequently, we face some very interesting questions: Is there any dynamics or mutual influence between social actions and social ties? To what extent do they influence each other? A fundamental mechanism that drives the dynamics of networks is the underlying social phenomenon of *homophily* (McPherson *et al.*, 2001): people tend to follow the behaviors of their friends, and people tend to create relationships with other people who are already similar to them. This suggests that both actions and ties

This work is licenced under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

are bi-directionally correlated and mutually influenced in social media, they could be mutually reinforced if modeled jointly.

Inspired by this mechanism, we propose a single unified framework based on exponential-family random graph models ((Frank and Strauss, 1986), (Wasserman and Pattison, 1996)) to exploit homophily for simultaneous social action prediction and social tie discovery. This mutual latent random graph (MLRG) framework incorporates shared latent factors with users, users' behaviors and users' relations, and defines coupled models to encode both social action and social tie information, to capture dynamics and mutual influence between them. We propose a gradient based algorithm for learning and optimization. During the learning procedure, social actions (i.e. user-behavior interactions), social ties (i.e. user-user connections), and deep dependencies and interactions between them could be efficiently explored. Experimental results demonstrate that social actions and social ties are highly correlated and mutually helpful. By coupling actions with ties jointly in a single coherent framework, MLRG achieves significantly better performance on both social action prediction and social tie inference, compared to state-of-the-art systems modeling them independently.

2 Related Work

Social network analysis has attracted much interest in both academia and industry recently. Considerable research and engineering has been conducted for social media modeling, analytics and optimization, including social community detection (Fortunato, 2010), user behavior modeling and prediction ((Benvenuto *et al.*, 2009), (Kwak *et al.*, 2010), (Ma *et al.*, 2009), (Ma *et al.*, 2011)), social tie analysis ((Tang *et al.*, 2011), (Tang *et al.*, 2012)), social sentiment analysis ((Wasserman *et al.*, 1994), (Pang and Lee, 2008)), etc.

Social action prediction and social tie discovery are two fundamental tasks for social media and social network analysis. Traditionally, they were considered as separate tasks and solved independently without considering the bidirectional interactions and interdependencies between them. Social action investigation is essentially important in online social media. Users behaviors could be affected by various kinds of complex factors, such as users' attributes, users's historical behaviors, social influence and social network structures. Based on this motivation, Tan *et al.* (2010) proposed a noise tolerant time-varying model to track social actions. Aiming at modeling user actions more accurately and realistically, Ma *et al.* (2009) and Ma *et al.* (2011) considered connections among users and proposed social trust ensemble to fuse the users' tastes and their trusted friends' favors together. Gao *et al.* (2013) investigated users' social behaviors from a spatio-temporal-social aspect in location-based mobile social networks. In particular, Gao *et al.* (2013) focused on temporal effects in terms of temporal preferences and temporal correlations, and modeled temporal cyclic patterns to capture a user's mobile behavior to investigate correlations to the spatial context and social context in location-based social networks.

Social tie is the most basic unit to form the network structure. Tang *et al.* (2011) proposed a semi-supervised framework, the partially labeled factor graph model to infer the type of social relationships. The task was formulated as a relationship mining problem to detect the relationship semantics in real-world networks. Tang *et al.* (2012) further incorporated social theories and leveraged features based on those social theories to infer social ties across heterogeneous networks via a transfer learning framework. As can be seen, predicting social actions and inferring social ties were modeled as separate and independent tasks in the above-mentioned approaches, deep interactions and mutual influence between them were not taken into consideration. In social media users interact with one another to share the content they both create and consume. According to the homophily phenomenon, exploring bi-directional information and mutual influence between them is intuitively appealing.

We are also aware of several research work attempting to explore joint models to capture mutual benefits and deep dependencies between different tasks in NLP, data mining and information extraction research communities ((Ko *et al.*, 2007), (Yu and Lam, 2008), (Yu *et al.*, 2009), (Liu *et al.*, 2009), (Yu and Lam, 2010b), (Yu and Lam, 2010a), (Yu *et al.*, 2011), (Yu and Lam, 2012), (Zeng *et al.*, 2013)). In general, joint models aim to handle multiple hypotheses and uncertainty information and to predict many variables simultaneously such that subtasks can aid each other to boost the performance.

Ko *et al.* (2007) proposed a joint answer ranking framework based on probabilistic graphical models for question answering. Yu and Lam (2008) proposed an integrated probabilistic and logic approach based on Markov Logic Networks (MLNs) (Richardson and Domingos, 2006) to encyclopedia relation extraction. However, this modeling only captures single relation extraction task. Liu *et al.* (2009) developed a Bayesian hierarchical approach, the topic-link LDA, to perform topic modeling and author community discovery for large-scale linked documents in one unified framework. Yu *et al.* (2009) integrated two sub-models in a unified framework via Markov chain Monte Carlo (MCMC) sampling based inference algorithms. This is a loosely coupled model since parameter estimation is performed separately for the two sub-models. Yu and Lam (2012) further proposed a joint model incorporating probabilistic graphical models and first-order logic for information extraction. This joint model exploits structured variational approximation for tractable parameter learning. Zeng *et al.* (2013) presented a semi-supervised graph-based approach to joint Chinese word segmentation and POS tagging. However, none of these models has been investigated or applied to social media and social network analysis. We believe that one major reason could be the problem of high computational complexity, such as Yu *et al.* (2009) and Yu and Lam (2012). Since many social network sites contain millions of users, exploiting such models could be very challenging. Currently, research on building joint approaches is still in the infancy stage. To the best of our knowledge, there is few systematically study on building joint models to explore mutual influence for social actions and social ties.

3 Model

In this section we consider both social action prediction and social tie inference in the context of social media, where evidences for both actions and ties are available. We begin by necessary description of preliminaries and notations, we then present the mutual latent random graphs (MLRGs) model, upon which both sources of evidence could be exploited simultaneously to capture their mutual influence. We also discuss the major difference and superiority of this model against several alternative models.

3.1 Preliminaries and Notations

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a social network graph, where $\mathbf{V} = \{v_1, v_2, \dots, v_N\}$ is the set of $|\mathbf{V}| = N$ users and $\mathbf{E} = \{e_{11}, e_{12}, \dots, e_M\} \subset \mathbf{V} \times \mathbf{V}$ is the set of $|\mathbf{E}| = M$ connections between users. Let $\mathbf{y} = \{y_1, y_2, \dots, y_N\} (y_i \in \mathcal{Y})$ be the set of actions associated with N users, and $\mathbf{s} = \{s_{11}, s_{12}, \dots, s_M\} (s_{ij} \in \mathcal{S})$ be the set of corresponding social tie labels associated with M connections. The connection $e_{ij} (1 \leq i, j \leq N, i \neq j)$ between v_i and v_j might be directed or undirected. To be consistent, both $s_{ij} \neq s_{ji}$ and $s_{ij} = s_{ji}$ are valid settings. Given the observed social network data \mathcal{D} constructing the graph \mathcal{G} , our goal is to simultaneously detect the most likely types of actions \mathbf{y}^* and ties \mathbf{s}^* such that both of them are optimized.

The exponential-family random graph models (ERGMs) (Frank and Strauss, 1986), (Wasserman and Pattison, 1996) take the form of an exponential family as $P_{y_i|\mathcal{G}} = \prod_{y_i \in \mathcal{Y}} \phi(y_i) = \frac{\exp\{\sum_{y_i \in \mathcal{Y}} \eta \xi(y_i)\}}{\kappa_\eta}$ for the social action y_i in the social network graph \mathcal{G} , where $\phi(\cdot)$ is a factor, η is a vector of parameters, $\xi(\cdot)$ is a p-vector of sufficient statistics, which captures network features of interest, its postulated dependence structure, or both. Lastly, κ_η is a normalization function to make all probabilities sum to one. The class of ERGMs is a popular framework for social network modeling to capture global network characteristics.

3.2 Modeling Social Actions

To characterize the user action y_i , we assume that for the user v_i there exist observable attributes or properties \mathbf{m}_i , such as the user's registered information and historical actions. Without loss of generality, we further assume that there exist some hidden, or latent properties \mathbf{x}_{ij} for v_i . These properties are implicit and cannot be observed directly, such as the influence from social ties. Consequently, we denote the observable factor $\phi(y_i, v_i, \mathbf{m}_i)$ for observable properties and latent factor $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ for hidden properties, respectively. Given the graph \mathcal{G} , the probability distribution of y_i depends on both observable and latent factors as:

$$P_{y_i|\mathcal{G}} \sim \phi(y_i, v_i, \mathbf{m}_i), \quad P_{y_i|\mathcal{G}} \sim \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}), \quad P_{y_i|\mathcal{G}} \sim \phi(y_i, v_i, \mathbf{m}_i) \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}). \quad (1)$$

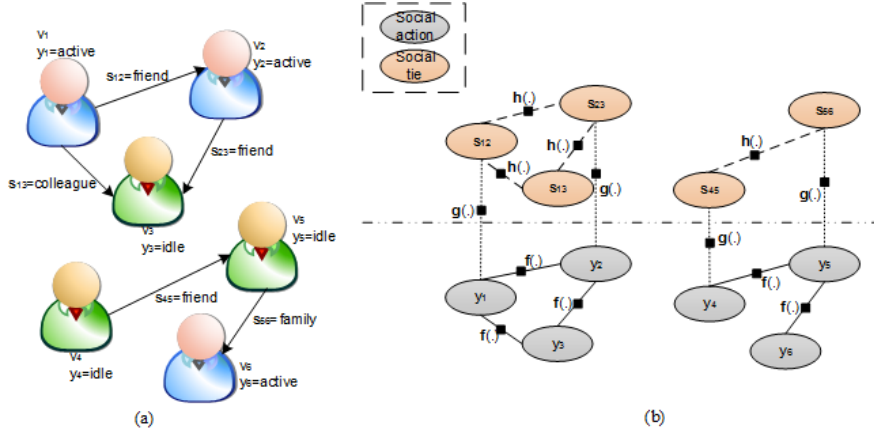


Figure 1: (a) A social network containing 6 users and 5 social ties. The social action can be active or idle, and the social tie can be friend, colleague, or family. (b) The three-dimensional graphical representation of the corresponding MLRG model. We use different lines to represent functions $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$.

The Mutual Latent Random Graph (MLRG) model	
$\forall y_i \in \mathcal{Y}$	$P_{y_i \mathcal{G}} \sim \phi(y_i, v_i, \mathbf{m}_i) \phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$
$\forall s_{ij} \in \mathcal{S}$	$P_{s_{ij} (y_i, \mathcal{G})} \sim \phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) \phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$
$\forall y_i \in \mathcal{Y}, \forall s_{ij} \in \mathcal{S}$	$P_{(y_i, s_{ij}) \mathcal{G}} \sim \phi(y_i, v_i, \mathbf{m}_i) \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}) \phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij})$

This modeling integrates two types of factors for both observable and latent properties. It captures not only the user-behavior dependencies, but also the influence from social ties, for exploring social actions.

3.3 Modeling Social Ties

To characterize the social tie s_{ij} between user pair (v_i, v_j) , we also assume that there exist observable properties \mathbf{w}_{ij} , such as the posterior probability of the social tie s_{ij} assigned to (v_i, v_j) . We denote the observable factor $\phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij})$ for \mathbf{w}_{ij} . Similarly, we further assume that there exist some latent properties to incorporate the social action influence on social ties. To be consistent, we still use the vector \mathbf{x}_{ij} to represent the latent properties and the latent factor $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ to capture the social action influence on social ties. Note that both \mathbf{x}_{ij} and $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ now play double duties in encoding social action dependency and social tie connection simultaneously. On the one hand, $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ exploits influence from social ties for modeling social actions. On the other hand, this factor exploits influence from social actions for modeling social ties. By doing so, the latent factor $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ is bi-directionally coupled, encoding both sources of evidence and exploring mutual influence and dynamics between social actions and social ties. Such mutual influence and dynamics are crucial and modeling them often leads to improved performance. Given the user action y_i and the graph \mathcal{G} , we devise the following model for the probability distribution of s_{ij} depending on both observable and latent factors as:

$$P_{s_{ij}|(y_i, \mathcal{G})} \sim \phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij}), \quad P_{s_{ij}|(y_i, \mathcal{G})} \sim \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}), \quad P_{s_{ij}|(y_i, \mathcal{G})} \sim \phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}). \quad (2)$$

3.4 Modeling Mutual Influence

The mutual correlation between social actions and social ties advocates joint modeling of both sources of evidence in a single unified framework. Based on the above descriptions, we define our mutual latent random graph (MLRG) based on exponential-family random graph models (ERGMs) (Frank and Strauss, 1986), (Wasserman and Pattison, 1996)), which have gained tremendous successes in social network analysis and have even become the current state-of-the-art (Robins *et al.*, 2007). To design a concrete model, one needs to specify distributions for the dependencies for MLRGs. According to the celebrated Hammersley-Clifford theory, the joint conditional distribution $P_{(y_i, s_{ij})|\mathcal{G}}$ is factorized as a product of potential functions over all cliques in the graph \mathcal{G} and we summarize the MLRG in the above table. In summary, our model consists of three factors: the factor $\phi(y_i, v_i, \mathbf{m}_i)$ measuring dependencies

of the social action y_i conditioned on \mathcal{G} , the factor $\phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij})$ measuring the social tie s_{ij} between two arbitrary users v_i and v_j in \mathcal{G} , and the latent factor $\phi_h(y_i, s_{ij}, \mathbf{x}_{ij})$ exploiting mutual influence between the social action y_i and social tie s_{ij} .

The three factors $\phi(\cdot)$, $\phi_h(\cdot)$, and $\phi'(\cdot)$ can be instantiated in different ways. In this paper, each factor is defined as the exponential family of an inner product over sufficient statistics (feature functions) and corresponding parameters. Each factor is a clique template whose parameters are tied. More specifically, we define these factors as

$$\begin{aligned} \phi(y_i, v_i, \mathbf{m}_i) &= \frac{1}{Z_\alpha} \exp\left\{ \sum_{y_i \in \mathcal{Y}} \alpha \mathbf{f}(y_i, v_i, \mathbf{m}_i) \right\}, & \phi_h(y_i, s_{ij}, \mathbf{x}_{ij}) &= \frac{1}{Z_\beta} \exp\left\{ \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \beta \mathbf{g}(y_i, s_{ij}, \mathbf{x}_{ij}) \right\}, \\ \phi'(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) &= \frac{1}{Z_\gamma} \exp\left\{ \sum_{s_{ij} \in \mathcal{S}} \gamma \mathbf{h}(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) \right\}, \end{aligned} \quad (3)$$

where α, β , and γ are real-valued weighting vectors and $\mathbf{f}(\cdot)$, $\mathbf{g}(\cdot)$, and $\mathbf{h}(\cdot)$ are corresponding vectors of feature functions.

We denote $\Theta = \{\alpha, \beta, \gamma\}$ as the set of model's parameters, and concatenate all factor functions as $\Theta \mathbf{q}(y_i, s_{ij}) = \alpha \mathbf{f}(y_i, v_i, \mathbf{m}_i) + \beta \mathbf{g}(y_i, s_{ij}, \mathbf{x}_{ij}) + \gamma \mathbf{h}(s_{ij}, v_i, v_j, \mathbf{w}_{ij})$, the joint probability distribution shown in the above table can be rewritten as

$$P_{(\mathbf{y}, \mathbf{s})|\mathcal{G}} = \prod_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \Phi(y_i, s_{ij}) = \frac{1}{Z} \exp\left\{ \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \Theta \mathbf{q}(y_i, s_{ij}) \right\}, \quad (4)$$

where $Z = Z_\alpha Z_\beta Z_\gamma$ is the partition function of our MLRG model.

Figure 1 shows an example social network ($\mathbf{V} = \{v_1, v_2, v_3, v_4, v_5, v_6\}$, $\mathcal{Y} = \{\text{active, idle}\}$, $\mathcal{S} = \{\text{friend, colleague, family}\}$) and the corresponding 3D graphical representation of the MLRG model. The functions $\mathbf{f}(\cdot)$ model dependencies of social actions in the bottom part, and the functions $\mathbf{h}(\cdot)$ model dependencies of social ties in the upper part. More importantly, the functions $\mathbf{g}(\cdot)$ capture mutual influence and dependencies between social actions and social ties. As we will see, this modeling offers a natural formalism for exploiting bi-directional dependencies and interactions between social actions and social ties to capture their mutual influence, as well as a great flexibility to incorporate a large collection of arbitrary, overlapping and nonindependent features.

3.5 Discussion

Noticeably, our proposed MLRG model is essentially different from the standard exponential-family random graph models (ERGMs) and the prior models discussed in Section 2 mainly in two aspects. Firstly, compared to the standard ERGMs, the MLRG model defines latent factors to assume mutual and dynamical interaction between social ties and social actions. Secondly, compared to the prior models such as (Ma *et al.*, 2009) and (Tang *et al.*, 2011), MLRG provides a single unified framework to address both social action prediction and social tie inference simultaneously while enjoying the benefits of both sources of evidence.

Importantly, we give an analytical explanation on the mutual nature of our model in terms of a random walk (Lovász, 1996) perspective. A random walk on the graph \mathcal{G} is a reversible Markov chain on the vertexes \mathbf{V} . The social influence propagation procedure occurs through information diffusion in the social graph \mathcal{G} . More specifically, a user v_i will propagate his/her influence to other related users, and will propagate more to the user which has a stronger relation (e.g., friendship) with v_i . The influence propagation will stop when the social graph reaches an equilibrium state, in which both social actions and social ties are mutually reinforced. Interestingly, this process is consistent with the homophily phenomenon that a user in the social network tends to be similar to his/her connected neighbors.

4 Learning and Inference

4.1 Mutual Optimization

The goal of learning MLRG model is to estimate a parameter configuration $\Theta = \{\alpha, \beta, \gamma\}$ such that the log-likelihood of observation is maximized. We define the log-likelihood objective function $\mathcal{O}(\Theta)$ of the

Algorithm 1: The Mutual Gradient Descent (MGD) algorithm

Input: The social graph \mathcal{G} , number of iterations n , and the learning rate η .

Output: Optimized parameters $\Theta^* = \{\alpha^*, \beta^*, \gamma^*\}$.

while equilibrium states or a threshold number of iterations are not reached **do**

repeat

 Choose a random example $(y_i, s_{ij}) \in \mathcal{G}$ as a sample;

Optimize social action parameters α and β :

 Compute the approximated gradients $\frac{\partial \mathcal{O}'}{\partial \alpha}$ and $\frac{\partial \mathcal{O}'}{\partial \beta}$ according to Eq. (6), Eq. (7) and stochastic approximation;

 Update α and β with learning rate η : $\alpha \leftarrow \alpha - \eta \cdot \frac{\partial \mathcal{O}'}{\partial \alpha}$, $\beta \leftarrow \beta - \eta \cdot \frac{\partial \mathcal{O}'}{\partial \beta}$.

 // Explore social tie influence

Optimize social tie parameters γ and β :

 Compute the approximated gradients $\frac{\partial \mathcal{O}'}{\partial \gamma}$ and $\frac{\partial \mathcal{O}'}{\partial \beta}$ according to Eq. (8), Eq. (7) and stochastic approximation;

 Update γ and β with learning rate η : $\gamma \leftarrow \gamma - \eta \cdot \frac{\partial \mathcal{O}'}{\partial \gamma}$, $\beta \leftarrow \beta - \eta \cdot \frac{\partial \mathcal{O}'}{\partial \beta}$.

 // Explore social action influence

until converge;

end

return α^* , β^* , and γ^*

observation given the graph \mathcal{G} as

$$\mathcal{O}(\Theta) = \log P_{(\mathbf{y}, \mathbf{s})|\mathcal{G}} - \log \Omega(\Theta) = \log[\exp\{\sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \Theta \mathbf{q}(y_i, s_{ij})\}] - \log Z - \log \Omega(\Theta), \quad (5)$$

where $\Omega(\Theta)$ is regularization to reduce over-fitting and a common choice is a spherical Gaussian prior with mean 0 and covariance $\delta^2 I$. $\Omega(\Theta) = \sum_{y_i \in \mathcal{Y}} \frac{\alpha^2}{2\sigma^2} + \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \frac{\beta^2}{2\sigma^2} + \sum_{s_{ij} \in \mathcal{S}} \frac{\gamma^2}{2\sigma^2}$.

We propose a mutual gradient descent (MGD) algorithm based on the stochastic gradient descent (SGD) (Lecun *et al.*, 1998), (Bottou, 2004) framework, for estimating the parameters efficiently in a mutual and collaborative manner. Once we have optimized the social action parameters α and β , the influence and hypotheses of social action can aid the learning of the social tie parameters γ and β , and vice versa. As shown in Algorithm 1, β is coupled parameter vector for both actions and ties, and is updated twice in each iteration of MGD. By doing so, MGD not only allows learning of social action parameters to capture social tie influence, but it also optimizes social tie parameters to alleviate social action influence. This training procedure runs iteratively until converge to boost both the optimization of social actions and social ties.

Each iteration of the MGD algorithm consists of drawing an example at random and applying parameter updates by moving in the direction defined by the stochastically approximated gradient of the loss function (e.g., $\frac{\partial \mathcal{O}'}{\partial \alpha}$). We update each parameter with a learning rate η . Ideally, each parameter should have its own learning rate. If shared parameter weights are used, the best learning rate of a weight should be inversely proportional to the square root of the number of connection sharing that weight (Bottou, 2004). In our MGD implementation, for simplicity we use the same learning rate for all the parameters. We select a small subset of training data and try various learning rates on the subset, then pick the one that most reduces the loss and use it on the full dataset. We summarize the partial derivatives of the log-likelihood function \mathcal{O} with respect to the parameter vectors α , β and γ as follows:

$$\frac{\partial \mathcal{O}}{\partial \alpha} = \sum_{y_i \in \mathcal{Y}} \mathbf{f}(y_i, v_i, \mathbf{m}_i) - \sum_{y_i \in \mathcal{Y}} \mathbf{f}(y_i, v_i, \mathbf{m}_i) \times P_{(\mathbf{y}, \mathbf{s})|\mathcal{G}} - \sum_{y_i \in \mathcal{Y}} \frac{\alpha}{\sigma^2}, \quad (6)$$

$$\frac{\partial \mathcal{O}}{\partial \beta} = \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \mathbf{g}(y_i, s_{ij}, \mathbf{x}_{ij}) - \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \mathbf{g}(y_i, s_{ij}, \mathbf{x}_{ij}) \times P_{(\mathbf{y}, \mathbf{s})|\mathcal{G}} - \sum_{y_i \in \mathcal{Y}, s_{ij} \in \mathcal{S}} \frac{\beta}{\sigma^2}, \quad (7)$$

$$\frac{\partial \mathcal{O}}{\partial \gamma} = \sum_{s_{ij} \in \mathcal{S}} \mathbf{h}(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) - \sum_{s_{ij} \in \mathcal{S}} \mathbf{h}(s_{ij}, v_i, v_j, \mathbf{w}_{ij}) \times P_{(\mathbf{y}, \mathbf{s})|\mathcal{G}} - \sum_{s_{ij} \in \mathcal{S}} \frac{\gamma}{\sigma^2}. \quad (8)$$

It is worth noting that the MGD algorithm computes approximations of the gradients, due to the intractability of the normalizing constant Z in the log-likelihood of our MLRG model. Our proposed

MGD algorithm is a generalized extension and it distinguishes from the standard SGD algorithm in two aspects: (1) MGD optimizes three types of parameters simultaneously, thus MGD is much more general than SGD, and it is more scalable and applicable to real-world problems. (2) MGD performs mutual and collaborative optimization to enable mutual influence between social actions and social ties, whereas SGD does not take such influence into account.

4.2 Complexity Analysis

Given several conditions including a suitable choice of the learning rate and a convex or pseudo-convex objective function, the MGD algorithm converges almost surely to a global optimum, otherwise it converges almost surely to a local optimum. In our experiments, this algorithm has good performance even if it does not reach the global optimum. Let D be the number of samples in the social graph \mathcal{G} , n be the number of iterations, and \bar{p} be the average number of non-zero attributes (features) per sample, the computational complexity of our MGD algorithm takes $O(nD\bar{p})$. As can be seen, this algorithm is computationally efficient, and convergence is very fast when the training examples are redundant since only a few examples are needed to sample. Furthermore, this algorithm is online and scale sub-linearly with the amount of training data, making it very attractive for large-scale datasets.

4.3 Inference

The objective of inference is to find the most likely types of actions \mathbf{y}^* and corresponding social tie labels \mathbf{s}^* , that is, to find $(\mathbf{y}^*, \mathbf{s}^*) = \arg \max_{(\mathbf{y}, \mathbf{s})} P_{(\mathbf{y}, \mathbf{s} | \mathcal{G})}$. The inference procedure is straightforward. Based on the learned parameters $\Theta^* = \{\alpha^*, \beta^*, \gamma^*\}$, we firstly predict the label of each social action y_i by finding a labeling assignment that maximizes $P_{y_i | \mathcal{G}}$ as $y_i^* = \arg \max_{y_i \in \mathcal{Y}} P_{y_i | \mathcal{G}}$. We then infer the social tie label s_{ij} such that $s_{ij}^* = \arg \max_{s_{ij} \in \mathcal{S}} P_{s_{ij} | (y_i, \mathcal{G})}$.

5 Experiments

5.1 Foursquare Data

We crawled one dataset from Foursquare¹, a popular location-based mobile social networking site for mobile devices (e.g., smartphones) for our experimental evaluation. Foursquare allows a user to check in at a physical location via his cellphone, and then let his online friends know where he is by publishing such check-in action online. Users check-in at venues using a mobile website, text messaging or a device-specific application by selecting from a list of venues the application locates nearby. Location is based on GPS hardware in the mobile device or network location provided by the application, and the map is based on data from the OpenStreetMap project. Each check-in awards the user points and sometimes badges. Figure 2 illustrates a snapshot of the Foursquare application interface on smartphones.

To alleviate the data sparsity problem for better evaluation, we selected check-in venues which have been visited by at least two distinct users, and users who have checked in at least 10 distinct venues. The resulting dataset contains 12,368 distinct users, 186,745 venues, 1,425,664 check-in behaviors and 56,395 social connections from January 2012 to December 2012. Table 1 lists the more detailed statistical information on our dataset, where the ‘‘Avg. Num. of Check-ins’’ is the average number of check-ins per user, and ‘‘Max. Num. of Check-ins’’ is the maximal number of check-ins among users (similarly for ‘‘Avg. Num. of Friendships’’ and ‘‘Max. Num. of Friendships’’). The ‘‘Average Clustering Coefficient (ACC)’’ is a measure of the degree to which users in the Foursquare network tend to cluster together, and ‘‘Diameter’’ is the longest shortest path in the network. All user and venue information has been anonymized. Each check-in has a unique id as well as the user id and the venue id, and each social connection consists of two users represented by two unique ids.

5.2 Task

Mobile phones have become an important tool for communication and they are an ideal platform for understanding social influence and social dynamics. Using our Foursquare dataset, we can investigate the mutual influence between social actions and social ties. More specifically, we can investigate how

¹<https://foursquare.com/>



Figure 2: A snapshot of the Foursquare application interface on smartphones.

Duration	Jan 2012 to Dec 2012	Num. of Users	12,368
Num. of Check-ins	1,425,664	Num. of Friendships	56,395
Avg. Num. of Check-ins	115.27	Max. Num. of Check-ins	657
Avg. Num. of Friendships	4.56	Max. Num. of Friendships	265
Average Clustering Coefficient (ACC)	0.42	Diameter	12

Table 1: Statistical information of our Foursquare dataset.

the friendship relations affect users' check-in behaviors, and how users' check-in behaviors affect their friendships. Figure 3 gives an illustrative example of social action prediction and social tie discovery tasks in our Foursquare dataset. Given an unseen Foursquare social network dataset, our objective is to predict whether the users have check-in behaviors and whether there are friendship relations between these users. In the right figure, we list the predicted check-in behaviors (in red color) and the inferred friendship relations between users (in green color). The probabilities associated with the predictions represent corresponding confidence scores.

5.3 Evaluation Methodology

We exploited a wide range of important features to define the factors $\phi(\cdot)$, and $\phi'(\cdot)$, including temporal and social features such as the number of check-ins and number of new check-ins in a user's history, number of friends of a user, the check-in information from a user's friends, etc. For the coupled latent factor $\phi_h(\cdot)$, we incorporated social tie evidences and hypotheses as features to capture social actions, and we also incorporated social action evidences and hypotheses as features to leverage social ties.

For quantitative performance evaluation, we used the standard measures of Precision (P), Recall (R), and F-measure (the harmonic mean of P and R: $\frac{2PR}{P+R}$) for both social action prediction and social tie inference. We performed four-fold cross-validation on this dataset, and took the average performance. We compared our approach with the following alternative methods for predicting social actions and inferring social ties:

–**SVM**: This model views social action prediction and social tie inference as two separate classification problems, and solves them independently. We used the SVM-light² package for this model.

–**ERGM**: This is the traditional exponential-family random graph model without the latent factor $\phi_h(\cdot)$ incorporated for social action prediction and social tie inference. Similar to SVM, this model also performs them separately.

–**DCRF**: This model is a dynamical and factorial CRF (Sutton *et al.*, 2007) used to jointly solve the two tasks. This model was originally proposed for labeling and segmenting sequence data, and we directly

²<http://svmlight.joachims.org/>

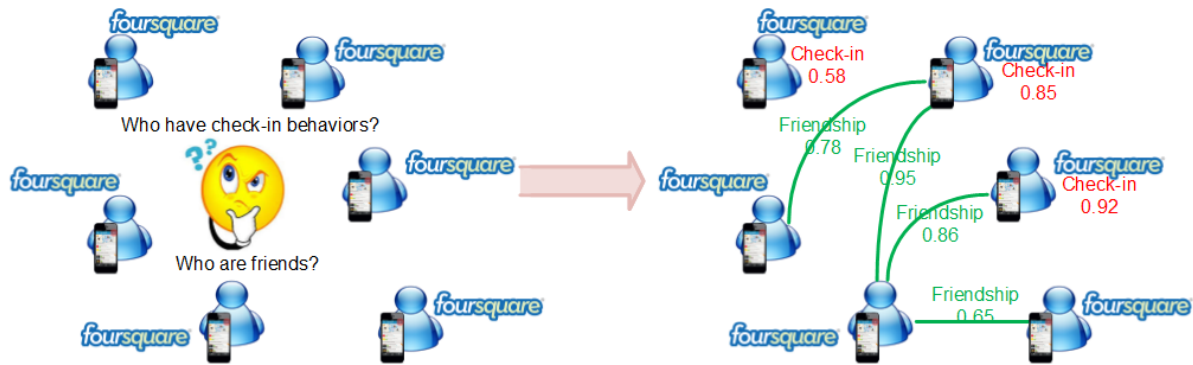


Figure 3: An illustrative example of social action prediction and social tie discovery tasks in our Foursquare dataset. The left is the input of our problem, and the right is the output of the two tasks.

Models	Precision	Recall	F-measure
SVM	73.75	64.54	68.84
ERGM	80.69	79.70	80.19
DCRF	89.45	82.32	85.74
MLRG	89.03	87.89	88.46

Table 2: Comparative performance of different models for social action prediction. The best results are printed in boldface.

Models	Precision	Recall	F-measure
SVM	70.75	61.57	65.84
ERGM	78.85	77.39	78.11
DCRF	82.45	76.56	79.40
MLRG	84.33	83.89	84.11

Table 3: Comparative performance of different models for social tie inference. The best results are printed in boldface.

applied it for our tasks in social network analysis.

All these models exploited standard parameter learning and inference algorithms in our experiments. To avoid over-fitting, penalization techniques on likelihood were also performed. All experiments were performed on the Linux workstation, with 24 2.5GHz Intel Xeon E5-2640 CPUs and 16 GB of memory.

5.4 Performance

Table 2 shows the performance on social action prediction and Table 3 shows the performance on social tie inference of different models, respectively. The best Precision, Recall and F1-measure of these results are highlighted. Our method consistently outperforms other comparative methods on the F-measure. The improvement is statistically and significantly better according to McNemar’s paired tests. These results not only imply that there exists high correlation and mutual influence between social actions and social ties, but also demonstrate the feasibility and effectiveness of our model for exploring them.

The SVM model solves social action prediction and social tie inference independently without considering mutual influence and benefits between them, thus leading to the worst performance. The ERGM outperforms SVM by capturing social network structures. However, the performance of this model is still limited and there is a large room for improving. The DCRF model easily outperforms both SVM and ERGM by modeling social actions and social ties jointly in a single framework. However, compared to our MLRG model, there are still some shortcomings of DCRF. DCRF was proposed to label and segment sequence data, such as POS tagging and NP chunking (Sutton *et al.*, 2007). The graphical structure of DCRF is not well suited for social networks to capture mutual influence. The merits of our proposed MLRG model over other models principally come from (1) appropriate graphical structure for social network modeling, especially the coupled latent factor to exploit mutual influence simultaneously, and (2) the mutual and collaborative learning algorithm MGD to reinforce the optimization of both social actions and social ties.

5.5 Effect of Mutual Influence and Analysis

We also examined the nature and effectiveness of the associated latent factors on the mutual influence, and Figure 4 demonstrates their feasibility in our modeling. Note that if we do not incorporate the latent factors, our MLRG model becomes the traditional ERGM baseline approach. It shows that the

latent factors consistently enhance Precision, Recall, and F-measure for both social action prediction and social tie inference tasks. For example, the latent factors significantly improve the F-measure by 8.27% (from 80.19 to 88.46) for social action prediction, and improve the F-measure by 6.0% (from 78.11 to 84.11) for social tie discovery, respectively. These results not only illustrate that social actions and social ties influence each other to a large extent, but also demonstrate the feasibility and effectiveness of our latent factors for exploring them.

We performed an in-depth error analysis to provide gains of our MLRG model and some insights on the influence between users' check-in behaviors and users' friendship relations. By carefully investigating our Foursquare dataset, we found that approximately 75% users tend to cluster together to create tightly knit groups characterized by a relatively high density of friendship relations or ties, and the remaining 25% users loosely or seldom connect with each other through the friendship relations. In other words, 75% users form high density of relationship ties and the average clustering coefficient (ACC) is high (0.61). However, the tie density of the remaining 25% users is much lower, since the ACC of these users is only 0.18. Compared to the baseline methods (especially the SVM and ERGM methods), the performance improvement of our MLRG model mainly comes from 75% users with high density of friendship ties. In particular, about 20% prediction errors (including social action and social tie prediction errors) of such users made by the SVM model can be corrected by our MLRG model. This finding shows that, the mutual influence between users' check-in behaviors and users' friendship relations increases with the density growth of the friendship relations of these users. This finding is intuitively correct and is consistent with the homophily theory. More interestingly, this finding also implies the gains and merits of our MLRG model for exploiting mutual influence, especially when the users in the Foursquare network cluster together tightly with high density of ties.

5.6 Efficiency

A number of learning algorithms can be applied for parameter optimization of our MLRG model. Table 4 summarizes the efficiency of several alternative optimization algorithms for learning our model's parameters. We compared the learning time (hr.) and inference time (sec.) of the MGD algorithm to loopy belief propagation (LBP), Markov chain Monte Carlo (MCMC) Gibbs sampling (Geman and Geman, 1984), and variational mean-field (VMF) approximation algorithms (Wainwright and Jordan, 2008). Both Sutton *et al.* (2007) and Tang *et al.* (2011) used LBP for parameter estimation. LBP is inherently unstable and may cause convergence problems. When the graph has large tree-width as in our case, the LBP algorithm is inefficient, and is slow to converge. In Gibbs sampling, the candidate sample is always accepted with the probability of 1, lacking the capability of measuring quality of samples and eliminating low grade samples. The VMF approach aims to minimize the Kullback-Leibler (KL) divergence between an approximated distribution Q and the target distribution P by finding the best distribution Q from some family of distributions for which an inference is feasible. The MGD algorithm we proposed is very efficient. It is particularly notable that our MGD algorithm takes much less time than other three algorithms for learning. In particular, our proposed algorithm is over orders of magnitude faster than the LBP for running.

6 Conclusions and Future Work

Finally, we answer the questions in Section 1 to draw the conclusions of this paper as follows:

Is there any dynamics or mutual influence between social actions and social ties? Doubtlessly, social actions and social ties are highly correlated and mutually reinforced. We propose a single unified framework, mutual latent random graph (MLRG), to exploit homophily for simultaneous social action prediction and social tie discovery. The MLRG model incorporates coupled latent factors to capture dynamics and mutual influence between social actions and social ties. Moreover, we propose the mutual gradient descent (MGD) algorithm to perform mutual and collaborative optimization to reinforce both social actions and social ties. By coupling actions with ties jointly in a single coherent framework, MLRG achieves significantly better performance on both social action prediction and social tie inference on our collected Foursquare dataset, compared to several state-of-the-art existing models.

To what extent do they influence each other? We perform an in-depth analysis to show the gains and merits of our MLRG model, as well as some insights on the influence between users’ check-in behaviors and users’ friendship relations. The finding on our real-world Foursquare data demonstrates that social actions (users’ check-in behaviors) and social ties (users’ friendship relations) influence each other to a considerable degree when the users connect each other tightly with high density of ties in the network. Experimental results also illustrate the feasibility and effectiveness of our latent factors for exploring the mutual influence. In particular, the latent factors in our model significantly improve the F-measure by 8.27% (from 80.19 to 88.46) for social action prediction, and improve the F-measure by 6.0% (from 78.11 to 84.11) for social tie discovery, respectively.

Two directions of future work appear attractive: Inferring fine-grained and multiple relationships between users (such as friendship, family, colleague, and advisor-adviser, etc.) on complex social networks and extending our established optimization algorithms for parallel and distributed learning based on the Hadoop MapReduce framework to handle large scale social networks involving billions of users.

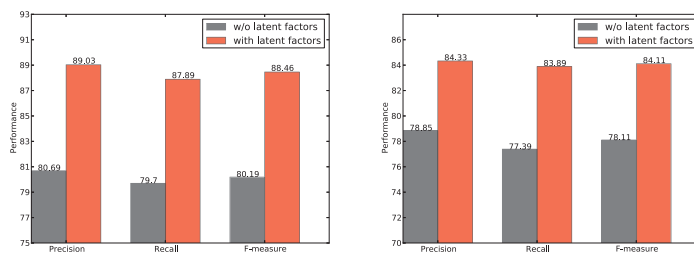


Figure 4: Contribution of latent factors on social action prediction (left) and social tie inference (right).

Algorithms	Learning	Inference
LBP	8.67	8
MCMC	3.45	124
VMF	2.39	7
MGD	0.45	6

Table 4: Efficiency comparison of different optimization algorithms on learning time (hr.) and inference time (sec.).

References

- Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio A. F. Almeida. Characterizing user behavior in online social networks. In *Proceedings of Internet Measurement Conference*, pages 49–62, 2009.
- Léon Bottou. Stochastic learning. In Olivier Bousquet and Ulrike von Luxburg, editors, *Advanced Lectures on Machine Learning*, Lecture Notes in Artificial Intelligence, LNAI 3176, pages 146–168. Springer Verlag, Berlin, 2004.
- Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.
- Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association (JASA)*, 81(395):832–842, 1986.
- Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. Modeling temporal effects of human mobile behavior on location-based social networks. In *Proceedings of CIKM-13*, pages 1673–1678, San Francisco, CA, USA, 2013.
- Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- Jeongwoo Ko, Luo Si, and Eric Nyberg. A probabilistic graphical model for joint answer ranking in question answering. In *Proceedings of SIGIR-07*, pages 343–350, Amsterdam, The Netherlands, 2007.
- Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media? In *Proceedings of WWW-10*, pages 591–600, Raleigh, North Carolina, USA, 2010.
- Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.
- Yan Liu, Alexandru Niculescu-Mizil, and Wojciech Gryc. Topic-link LDA: joint models of topic and author community. In *Proceedings of ICML-09*, pages 665–672, Montreal, Canada, 2009.
- L. Lovász. Random walks on graphs: A survey. *Combinatorics, Paul Erdős is Eighty*, 2:353–398, 1996.

- Hao Ma, Irwin King, and Michael R. Lyu. Learning to recommend with social trust ensemble. In *Proceedings of SIGIR-09*, pages 203–210, Boston, MA, USA, 2009.
- Hao Ma, Irwin King, and Michael R. Lyu. Learning to recommend with explicit and implicit social relations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–19, 2011.
- Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.
- Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.
- Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.
- Garry Robins, Tom A. B. Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. Recent developments in exponential random graph (p^*) models for social networks. *Social Networks*, 29(2):192–215, 2007.
- Charles Sutton, Andrew McCallum, and Khashayar Rohanimanesh. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research*, 8:693–723, 2007.
- Chenhao Tan, Jie Tang, Jimeng Sun, Quan Lin, and Fengjiao Wang. Social action tracking via noise tolerant time-varying factor graphs. In *Proceedings of KDD-10*, pages 1049–1058, Washington, DC, USA, 2010.
- Wenbin Tang, Honglei Zhuang, and Jie Tang. Learning to infer social ties in large networks. In *Proceedings of ECML/PKDD-11*, pages 381–397, Athens, Greece, 2011.
- Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring social ties across heterogeneous networks. In *Proceedings of WSDM-12*, pages 743–752, Seattle, WA, USA, 2012.
- Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008.
- Stanley Wasserman and Philippa Pattison. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika*, 61(3):401–425, 1996.
- Stanley Wasserman, Katherine Faust, Dawn Iacobucci, and Mark Granovetter. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- Xiaofeng Yu and Wai Lam. An integrated probabilistic and logic approach to encyclopedia relation extraction with multiple features. In *Proceedings of COLING-08*, pages 1065–1072, Manchester, United Kingdom, 2008.
- Xiaofeng Yu and Wai Lam. Bidirectional integration of pipeline models. In *Proceedings of AAI-10*, pages 1045–1050, Atlanta, Georgia, USA, 2010.
- Xiaofeng Yu and Wai Lam. Jointly identifying entities and extracting relations in encyclopedia text via a graphical model approach. In *Proceedings of COLING-10*, pages 1399–1407, Beijing, China, 2010.
- Xiaofeng Yu and Wai Lam. Probabilistic joint models incorporating logic and learning via structured variational approximation for information extraction. *Knowledge and Information Systems (KAIS)*, 32:415–444, 2012.
- Xiaofeng Yu, Wai Lam, and Bo Chen. An integrated discriminative probabilistic approach to information extraction. In *Proceedings of CIKM-09*, pages 325–334, Hong Kong, China, 2009.
- Xiaofeng Yu, Irwin King, and Michael R. Lyu. Towards a top-down and bottom-up bidirectional approach to joint information extraction. In *Proceedings of CIKM-11*, pages 847–856, Glasgow, Scotland, UK, 2011.
- Xiaodong Zeng, Derek F. Wong, Lidia S. Chao, and Isabel Trancoso. Graph-based semi-supervised model for joint chinese word segmentation and part-of-speech tagging. In *Proceedings of ACL-13*, pages 770–779, Sofia, Bulgaria, 2013.