Who to Follow and Why: Link Prediction with Explanations

Nicola Barbieri
Yahoo Labs
Barcelona, Spain
barbieri@yahoo-inc.com

Francesco Bonchi
Yahoo Labs
Barcelona, Spain
bonchi@yahoo-inc.com

Giuseppe Manco
ICAR-CNR
Rende, Italy
manco@icar.cnr.it

ABSTRACT

User recommender systems are a key component in any on-line social networking platform: they help the users growing their network faster, thus driving engagement and loyalty.

In this paper we study *link prediction with explanations* for user recommendation in social networks. For this problem we propose WTFW ("Who to Follow and Why"), a stochastic topic model for link prediction over directed and nodes-attributed graphs. Our model not only predicts links, but for each predicted link it decides whether it is a "topical" or a "social" link, and depending on this decision it produces a different type of explanation.

A topical link is recommended between a user interested in a topic and a user authoritative in that topic: the explanation in this case is a set of binary features describing the topic responsible of the link creation. A social link is recommended between users which share a large social neighborhood: in this case the explanation is the set of neighbors which are more likely to be responsible for the link creation.

Our experimental assessment on real-world data confirms the accuracy of WTFW in the link prediction and the quality of the associated explanations.

Categories and Subject Descriptors: H.2.8 [Database Management]: Database Applications - *Data Mining*
Keywords: social networks; link prediction

1. INTRODUCTION

Link prediction is the task of estimating the likelihood of the existence of an unobserved link between two nodes, based on the other observable links around the two nodes and, when available, the attributes of the nodes [8]. It finds application in any context in which the network is only partially observable and we want to guess the unobserved part. A typical setting is when we consider the network evolving along time, so that the unobservable part of the network is the set of links which are not yet created: given the graph observed at time t, we want to predict the set of links which will be created in the time interval $[t, t+1]$[17].

KDD'14, August 24–27, 2014, New York, NY, USA.

http://dx.doi.org/10.1145/2623330.2623733 .

Link prediction has been applied in a variety of domains, ranging from bioinformatics to web sites management, from bibliography to e-commerce [12, 18, 5]. However, the most immediate and prominent application of link prediction is the recommendation of users to other users of a social network. This is one of the most fundamental functionalities common to all on-line social networking platforms[1]: it helps the users having a quicker start in building their network, thus driving engagement and loyalty. It is a key component for growth and sustenance of a social network: for instance, the WTF ("Who to Follow") service at `Twitter` is claimed to be responsible for millions of new links daily [11]. Given that growing the user base and maintaining a high level of engagement are key factors for the success (or the death) of these billion-dollar businesses, one can easily figure out the importance of user recommendation systems.

In this paper we study *link prediction with explanations for user recommendation systems* in on-line social networks. Enriching recommendations with explanations has the benefit to increase the trust of the user in the recommendation, and thus the likelihood that the recommendation is adopted. While these benefits are well understood in classic collaborative-filtering recommender systems [14, 25, 30], providing explanations in the context of user recommendation systems is still largely underdeveloped: in fact, in most of the real-world systems, the unique explanations given for user recommendations are of the type "you should follow user Z because your contacts X and Y do the same".

Our starting observation is that a link creation is usually explainable by one of two main reasons: interest identity or personal social relations. This observation is rooted in sociology, where it goes under the name *common identity and common bond theory* [24, 26]. Identity-based attachment holds when people join a community based on their interest in a well-defined common theme shared by all of the members of that community. The goal in this case is information collecting and sharing in the specific theme of interest. People joining a community through identity-based links may not even directly participate, e.g., by producing content or by engaging with other members, and instead only passively consume information.

Conversely, bond-based attachment is driven by personal social relations with other specific individuals (e.g., family, friends, colleagues), and thus it does not require a common theme of interest to be justified. Bond-based links are usu-

[1] E.g., *"People You May Know"* in `Facebook` and `LinkedIn`, *"Recommended Blogs"* in `Tumblr`, or *"Who to Follow"* in `Twitter`, just to mention a few.

ally reciprocated, while identity-based links are much more directional, where the direction is given by the level of authoritativeness of the user on the theme. The two types of links create two different types of communities, that for simplicity we name "topical" for identity-based and "social" for bond-based [10].

Based on this observation we define a stochastic model, dubbed WTFW ("Who to Follow and Why"), which not only predicts links, but *for each predicted link it decides whether it is a topical or a social link, and depending on this decision it produces a different type of explanation.*

A topical link $u \rightarrow v$ (u should follow v) is usually recommended to u when v is authoritative in a topic in which u has demonstrated interest. In this case the explanation is a set of the top-k binary features (e.g., tags in `Flickr` or hashtags in `Twitter`) describing the topic of authoritativeness of v, which makes v a potential source of interesting information for u. A social link $u \rightarrow v$ instead is recommended when u and v are already part of the same social community, i.e., they have many contacts in common. In this case the explanation is the set of the top-k common neighbors w.r.t. the likelihood of being responsible for the link creation. As an important by-product, WTFW also *implicitly detects communities and their type (social or topical).*

More in details WTFW is a bayesian topic model defined over directed and nodes-attributed graphs. In WTFW each link creation and each attribute adoption by a node are explained w.r.t. a finite number of latent factors. These latent factors can be abstractly thought as *topics* or *communities*: in the rest of the paper we will use the three terms (latent factor, topic, and community) interchangeably. Each community is characterized by a level of sociality/topicality: social communities are characterized by high density and reciprocity of links, whereas topical communities are characterized by low entropy in the features and by the presence of authoritative users on the relevant topic. Each user tend to be involved in different communities to different extent and with different roles. These components are modeled by three different multinomial distributions over the set of users, modeling their sociality, authoritativeness and interest in each topic. Finally, each topic is characterized by a multinomial distribution over the feature set, which provide a semantic interpretation of the topic.

Paper contributions. The contributions of this paper are summarized as follows:

- We study for the first time the problem of *link prediction with explanations*, which is motivated by the real-world application of user recommender systems in online social networks.
- We introduce WTFW ("Who to Follow and Why") a stochastic topic model which not only predicts links, but for each predicted link it decides whether it is a topical or a social link, and depending on this decision it produces a different type of explanation.
- As a by-product, WTFW also implicitly extract communities that can be labeled as either topical or social.
- Our experimental assessment on two real-world datasets (`Twitter` and `Flickr`) confirms that our model is very accurate in link prediction and in labeling the predicted link as social or topical. The experiments also highlight the high quality of the topics extracted and their coherence with the topical Vs. social labeling.

2. RELATED WORK

Link prediction has attracted a great deal of attention in the last decade (the interested reader may refer to [12, 18] for a comprehensive survey): however, to the best of our knowledge, no previous work has studied *link prediction with explanations for user recommendation systems.* Our proposal can also be collocated in the literature on relational learning methods that are able to leverage attribute information on nodes [29, 34, 19]. The main drawback of those approaches is scalability, which seriously prevents their application on real-world networks.

The *Supervised Random Walk* algorithm for link prediction [1], exploits edge features to learn the edge strength that is then used random walk transition probability. Alternative random-walk approaches rely on merging the social graph and node attributes in a unique graph with person-nodes and attribute-nodes linked among them [33, 9].

The joint factorization of social links and node attributes is closely related to the task of detecting communities in nodes-attributed graphs. [35] uses node attributes to augment the social graph by generating "attribute edges" between nodes that are similar on a given attribute, and then identify communities in the augmented graph. [21] introduces the problem of *finding cohesive patterns*, defined as connected subgraphs whose density exceeds a given threshold, and with homogeneous values on node-attributes. [31] proposes a co-clustering framework based on users and tags. Users are implicitly connected by their common interests, as expressed by the tags they use. [23] studies the problem of finding communities with *concise descriptions* based on the nodes attributes.

Several stochastic models for community detection in networks with node attributes have been proposed in the literature. In *Link-LDA* [6] social connections and user attributes are generated by a mixture of user-specific distributions over topics. In [22, 15, 32] the community-membership vectors are used to factorize both links and the attribute-profile of each user. [27] extends the author-topic model to communication networks in which the sender and recipient of each post are known. [2] proposes a generative stochastic model to detect communities from the social graph and a database of information propagations over the social graph.

3. WHO TO FOLLOW AND WHY

In this section we introduce the WTFW model for link prediction with explanations. Our application scenario is that of online social networking platforms, where users build and maintain social connections, share information, and follow updates from other users. We represent this as a directed graph, where each node is a user and it has associated a set of binary features, representing the interests of the user.

More formally, let $G = (V, E)$ be the social graph where V is a set of n users, $E \subseteq V \times V$ is a set of m *directed* arcs, and (u, v) indicates that u follows v and hence he is notified of v's activities. We also denote the neighborhood of a node u as $N(u) = \{v \in V : (u, v) \in E \vee (v, u) \in E\}$. Moreover let $\mathcal{F}$ denote a set of h binary features. We are given a binary $n \times h$ matrix F such that $F_{u,f} = 1$ when user u is interested in the feature f. For simplicity we denote this case also as $(u, f) \in F$. Finally, we denote all the features of the node u as $F(u) = \{f \in \mathcal{F} : (u, f) \in F\}$ and the set of all the nodes having attribute f as $V(f) = \{u \in V : (u, f) \in F\}$.

1. sample $\mathbf{\Pi} \sim Dir\left(\vec{\xi}\right)$
2. For each $k \in \{1, \ldots, K\}$ sample

$$\delta_k \sim Beta(\delta_0, \delta_1) \qquad \tau_k \sim Beta(\tau_0, \tau_1)$$
$$\mathbf{\Phi}_k \sim Dir\left(\vec{\gamma}\right) \qquad \boldsymbol{\theta}_k \sim Dir\left(\vec{\alpha}\right)$$
$$\boldsymbol{A}_k \sim Dir\left(\vec{\beta}\right) \qquad \boldsymbol{S}_k \sim Dir\left(\vec{\eta}\right)$$

3. For each link $l \in \{l_1, \ldots, l_m\}$ to generate:
   (a) Choose $k \sim Discrete(\mathbf{\Pi})$
   (b) Sample $x_l \sim Bernoulli(\delta_k)$
   (c) if $x_l = 1$
      - sample source $u \sim Discrete(\boldsymbol{\theta}_k)$
      - sample destination $v \sim Discrete(\boldsymbol{\theta}_k)$

   (d) else
      - sample source $u \sim Discrete(\mathbf{S}_k)$
      - sample destination $v \sim Discrete(\mathbf{A}_k)$
4. For each feature pair $a \in \{a_1, \cdots, a_t\}$ to associate
   (a) sample $k \sim Discrete(\Pi)$
   (b) Sample $y_a \sim Bernoulli(\tau_k)$:
      - if $y_a = 1$ then $u_a \sim Discrete(\boldsymbol{A}_k)$
      - otherwise $u_a \sim Discrete(\boldsymbol{S}_k)$

   (c) sample $f_a \sim Discrete(\mathbf{\Phi}_k)$

Figure 1: Generative process for the WTFW model

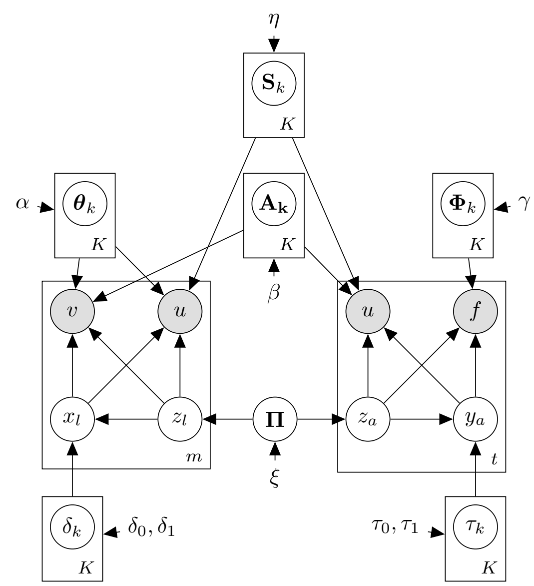


Figure 2: The WTFW model in plate notation.

Following the *common identity and common bond theory* discussed in Section 1, we assume two main types of behavior in creating connections in a social network. The "topical" behavior, in which a user u decides to follow another user v because of u's interest in a topic in which v is authoritative; and the "social" behavior in which u follows v because they know each other in the real world, or they have many common contacts in the social network. In the topical behavior case we can further identify two distinct roles for a user, either as authoritative ("influential") for the topic or just interested ("susceptible") in the topic. In the social case instead there are no specific roles, but a generic tendency to connect among the users of a close-knit circle.

Following these considerations, we propose to explain the structure of the network (the links) and the features of the nodes, by introducing a set of latent factors representing users' interests, and by labeling the links as either social or topical. This is done by means of a unique stochastic topic model, which is based on the following assumptions:

- Links can be explained by different latent factors (overlapping communities);
- Social links tend to be reciprocal and communities characterized by a high level of sociality exhibit *high density*;
- Topical links tend to exhibit a clear directionality and communities that are highly topicality have *low entropy* on the set of features assigned to nodes.

More in details, the degree of involvement and role of user u in the community/topic k is governed by three parameters: (1)$A_{k,u}$ which measures the degree of the authoritativeness of u in k; (2)$S_{k,u}$ which measures the degree of interest u in the topic k, or in other terms, the likelihood of following users that are authoritative in k (susceptibility to social influence); and (3)$\theta_{k,u}$ denotes the social tendency of u, i.e., her likelihood to connect to other social peers within community k. Moreover, each latent factor k is characterized by a propensity to adopt certain features in $\mathcal{F}$ over others. We can formalize such a propensity by means of a weight $\Phi_{k,f}$, denoting the importance of feature f within k.

All these components are accommodated in a mixture membership model expressed in a Bayesian setting [4], to define distributions governing the stochastic process, given some prior hypotheses. Bayesian modeling is better suited when the underlying data is characterized by high sparsity (like in our case), as it allows a better control of the priors which govern the model and it prevents overfitting.

In particular, we directly model each *observed* social link $(u, v) \in E$ or adoption of feature by a node $(u, f) \in F$ and introduce random variables on the source/destination of these observations. That is, for each link $(u, v) \in E$ we model the likelihood that there exists a latent factor k, such that u has high probability of being a source, while v has high probability of being a destination. We further introduce a latent variable $x_{u,v}$, which encodes the (social/topical) nature of an existing link. Analogously, the adoption of an observed feature association $(u, f) \in F$ will be explained by a latent factor k and by the status of the latent variable $y_{u,f}$ which represent the role of the user u, either as authoritative or just interested, when adopting the feature f.

The underlying generative process for social links and adoption of features depends jointly on the components $\boldsymbol{\theta}$, $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{\Phi}$, as described in Figure 1 and depicted in plate notation in Figure 2. The overall generative process is governed by the following components:

- A multinomial distribution $\mathbf{\Pi}$ over a fixed number of K latent factors, which generate latent community-assignments z_l and z_a, for each link $l \in E$ and for each adoption of feature $a \in F$;
- The multinomial distributions $\boldsymbol{\theta}_k, \mathbf{A}_k$ and $\mathbf{S}_k$ over the set of user V, which specify, respectively, the degree of sociality, authority and susceptibility of each user within k;
- The multinomial probability $\mathbf{\Phi}_k$ over $\mathcal{F}$ which specify the likelihood of observing each feature within the community k.

- The degree of "sociality" δ_k (or "topicality", $1-\delta_k$) which measures the likelihood of observing social/topical connections within each community k;
- The "authoritative attitude" τ_k of observing the adoption of an attribute by authoritative subject in k (or, dually, the "susceptible attitude", $1-\tau_k$).

Since the whole model relies on multinomial and Bernoulli distributions, a full Bayesian treatment can be obtained by adopting Dirichlet and Beta conjugate priors.

Let $\Theta = \{\mathbf{\Pi}, \boldsymbol{\delta}, \boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{A}, \mathbf{S}\}$ denote the status of the distributions described above. Both the probability of observing link $l = (u, v)$ and a feature assignment $a = (u, f)$ can be expressed as mixtures over the latent community assignments z_l and z_a:

$$\Pr(l|\Theta) = \sum_{k=1}^{K} \pi_k \Pr(l|z_l = k, \Theta) \tag{1}$$

$$\Pr(a|\Theta) = \sum_{k=1}^{K} \pi_k \Pr(a|z_a = k, \Theta) \tag{2}$$

The generation of a link changes depending on the status of the latent variable x_l. A social connection $l = (u, v)$ can only be observed if, by picking a latent community k, u and v have high degrees of social attitude $\theta_{k,u}$ and $\theta_{k,v}$, that is

$$\Pr(l|z_l = k, x_l = 1, \Theta) = \theta_{k,u} \cdot \theta_{k,v}.$$

Conversely, a topical connection $l = (u, v)$ can only be observed if, by picking a latent community k, u has a high degree of activeness $A_{k,u}$ and v have a high degree of passive interest $S_{k,u}$, that is

$$\Pr(l|z_l = k, x_l = 0, \Theta) = S_{k,u} \cdot A_{k,v}.$$

Note that the likelihood of observing the reciprocal link (v, u) is equally likely in case of social connection, while it is different in a topical context, and hence reflect our design assumption on the directionality of links in social/topical communities. Each link is finally generated by taking into account the social/topical mixture of each community:

$$\begin{aligned}\Pr(l|z_l = k, \Theta) =& \delta_k \Pr(l|z_l = k, x_l = 1, \Theta) \\ &+ (1-\delta_k) \Pr(l|z_l = k, x_l = 0, \Theta) \\ =& \delta_k \cdot \theta_{k,u} \cdot \theta_{k,v} + (1-\delta_k) \cdot S_{k,u} \cdot A_{k,v}\end{aligned}$$

Similarly, the probability of observing a node-feature pair $a = (u, f) \in F$ depends on the degree of authoritativeness/susceptibility of the user and by the likelihood of observing the attribute f within each latent factor k:

$$\Pr(a|z_a = k, \Theta) = (\tau_k A_{k,u} + (1-\tau_k) \cdot S_{k,u}) \, \Phi_{k,f}.$$

Here, the term $\tau_k \mathbf{A}_k + (1-\tau_k)\mathbf{S}_k$ defines a multinomial distribution over users, which encodes the joint (both susceptible and authoritative) attitude of users within that community.

3.1 Learning

We have described the intuitions behind our joint modeling of links and feature associations and now we focus on defining a procedure for inference and parameter estimation under WTFW.

Let $\Xi = \{\vec{\xi}, \vec{\alpha}, \vec{\beta}, \vec{\gamma}, \vec{\eta}, \vec{\delta} = \{\delta_0, \delta_1\}, \vec{\tau} = \{\tau_0, \tau_1\}\}$ denote the set of hyperparameters of the Dirichlet/Beta priors. Also, let $\mathbf{Z}^e$ represents a binary $m \times K$ matrix where $z_{l,k} = 1$ denotes that link l has been associated with the k-th latent factor (i.e., $z_l = k$). Analogously, $\mathbf{Z}^f$ denotes the $t \times K$ binary matrix where $z_{a,k} = 1$ denotes that feature assignment $a \in F$ is associated with the k-th latent factor ($z_a = k$). Finally, $\mathbf{X}$ and $\mathbf{Y}$ denote the vectors of assignments x_l and y_a. With an abuse of notation, we also introduce the counters described in Tab. 1, relative to these matrices.

The key problem in inference is to compute the posterior distribution of latent variables given the observed data. We start by expressing the joint likelihood as:

$$\begin{aligned}\Pr(E, F, \Theta, \mathbf{Z}^e, \mathbf{Z}^f, \mathbf{X}, \mathbf{Y}|\Xi) = \\ & \Pr(E|\Theta, \mathbf{X}, \mathbf{Z}^e) \Pr(F|\Theta, \mathbf{Y}, \mathbf{Z}^f) \\ & \Pr(\mathbf{Z}^e|\Pi) \Pr(\mathbf{Z}^f|\Pi) \\ & \Pr(\mathbf{X}|\mathbf{Z}^e, \boldsymbol{\delta}) \Pr(\mathbf{Y}|\mathbf{Z}^f, \boldsymbol{\tau}) \Pr(\Theta|\Xi)\end{aligned} \tag{3}$$

where

$$\Pr(E|\Theta, \mathbf{X}, \mathbf{Z}^e) = \prod_u \prod_k \theta_{k,u}^{c_{k,u}^{s,s} + c_{k,u}^{s,d}} S_{k,u}^{c_{k,u}^{t,s}} A_{k,u}^{c_{k,u}^{t,d}}$$

$$\Pr(F|\Theta, \mathbf{Y}, \mathbf{Z}^f) = \prod_u \prod_k A_{k,u}^{d_{k,u}^a} S_{k,u}^{d_{k,u}^s} \prod_f \prod_k \Phi_{k,f}^{d_{k,f}}$$

$$\Pr(\mathbf{Z}^e|\Pi) = \prod_k \pi_k^{c_k}$$

$$\Pr(\mathbf{Z}^f|\Pi) = \prod_k \pi_k^{d_k}$$

$$\Pr(\mathbf{X}|\mathbf{Z}^e, \boldsymbol{\delta}) = \prod_k \delta_k^{c_k^s} (1-\delta_k)^{c_k^t}$$

$$\Pr(\mathbf{Y}|\mathbf{Z}^f, \boldsymbol{\tau}) = \prod_k \tau_k^{d_k^a} (1-\tau_k)^{d_k^s}$$

and $\Pr(\Theta|\Xi)$ represents the product of all the Dirichlet and Beta priors. By marginalizing over Θ, we can obtain a closed form for the joint likelihood $\Pr(E, F, \mathbf{Z}^e, \mathbf{Z}^f, \mathbf{X}, \mathbf{Y}|\Xi)$. The latter is the basis for developing a stochastic EM strategy [3, section 11.1.6], where the E-step consists of a collapsed Gibbs sampling procedure [13, 3] for estimating the matrices $\mathbf{Z}^e, \mathbf{Z}^f, \mathbf{X}$ and $\mathbf{Y}$, and the M-step estimates both the predictive distributions in Θ and the hyperparameters of interest in Ξ. In particular, the sampling step consists of a sequential update of each arc and feature-assignment, of the status of the corresponding latent variables in $\mathbf{Z}^e, \mathbf{Z}^f, \mathbf{X}$ and $\mathbf{Y}$. A possible sampling strategy for each arc $l \in E$ and adoption $a \in F$ is based on the following chain: $\Pr(z_l = k|Rest), \Pr(z_a = k|Rest), \Pr(x_l = 1|Rest)$ and $\Pr(y_a = 1|Rest)$[2]. By algebraic manipulations, we can devise the sampling equations expressed in Tab. 8. The overall learning scheme is shown in Alg. 1. Lines 5-12 of the algorithm represent the Gibbs sampling steps, while line 14 represents the update of the multinomial distributions which are collapsed in the derivation of the sampling equations:

$$A_{k,u} = \frac{c_{k,u}^{t,d} + d_{k,u}^a + \eta_u}{c_k^t + d_k^a + \sum_u \eta_u} \tag{4}$$

$$\theta_{k,u} = \frac{c_{k,u}^{s,s} + c_{k,u}^{s,d} + \alpha_u}{2c_k^s + \sum_u \alpha_u} \tag{5}$$

$$S_{k,u} = \frac{c_{k,u}^{t,s} + d_{k,u}^s + \eta_u}{c_k^t + c_k^s + \sum_u \eta_u} \tag{6}$$

$$\phi_{k,f} = \frac{d_{k,f} + \gamma_f}{d_k + \sum_f \gamma_f} \tag{7}$$

$$\pi_k = \frac{c_k + d_k + \xi_k}{m + t + \sum_k \xi_k} \tag{8}$$

In line 15 we update the Beta ($\vec{\delta}$, $\vec{\tau}$) and Dirichlet $\vec{\xi}$ hyperparameters, according to the fixed point iterative procedure

[2] The term $Rest$ denotes the remaining variables in the set $\{E, F, \mathbf{Z}^e, \mathbf{Z}^f, \mathbf{X}, \mathbf{Y}, \Theta, \Xi\}$ after the explicit variables in both the conditioning and conditioned part have been removed.

Symbol	Description	Expression
c_k	Number of links associated with community k	$\sum_{l\in E} z_{l,k}$
c^s	Number of social links	$\sum_{l\in E} x_l$
c^t	Number of topical links	$\sum_{l\in E}(1-x_l)$
c^s_k	Number of social links associated with community k	$\sum_{l\in E} x_l \cdot z_{l,k}$
c^t_k	Number of topical links associated with community k	$\sum_{l\in E}(1-x_l)\cdot z_{l,k}$
$c^{s,s}_{k,u}$	Number of social links associated with community k where u is the source	$\sum_{l=(u,\cdot)\in E}\{x_l \cdot z_{l,k}\}$
$c^{s,d}_{k,u}$	Number of social links associated with community k where u is the destination	$\sum_{l=(\cdot,u)\in E}\{x_l \cdot z_{l,k}\}$
$c^{t,s}_{k,u}$	Number of topical links associated with community k where u is the source	$\sum_{l=(u,\cdot)\in E}\{(1-x_l)\cdot z_{l,k}\}$
$c^{t,d}_{k,u}$	Number of topical links associated with community k where u is the destination	$\sum_{l=(\cdot,u)\in E}\{(1-x_l)\cdot z_{l,k}\}$
d_k	Number of feature-assignments associated with community k	$\sum_{a\in F} z_{a,k}$
d^a	Number of authoritative feature-assignments	$\sum_{a\in F} y_a$
d^s	Number of susceptible feature-assignments	$\sum_{a\in F}(1-y_a)$
d^a_k	Number of feature-assignments within community k on authoritative users	$\sum_{a\in F} y_a \cdot z_{a,k}$
d^s_k	Number of feature-assignments within community k on susceptible users	$\sum_{a\in F}(1-y_a)\cdot z_{a,k}$
$d_{k,f}$	Number of recipients associated with community k relative to feature f	$\sum_{a=(\cdot,f)\in F}\{(1-y_a)\cdot z_{a,k}\}$
$d^a_{k,u}$	Number of features associated with community k where u is the authoritative source	$\sum_{a=(u,\cdot)\in F}\{y_a \cdot z_{a,k}\}$
$d^s_{k,u}$	Number of features associated with community k where u is the susceptible source	$\sum_{a=(u,\cdot)\in F}\{(1-y_a)\cdot z_{a,k}\}$

Table 1: Counters adopted in the Gibbs Sampling and their meaning.

described in [20]. The final predictive distributions $\mathbf{A}$, $\mathbf{S}$, $\boldsymbol{\theta}$ and $\boldsymbol{\Pi}$, $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$ are averaged along all the steps of the Gibbs sampling procedure.

A single iteration of the sampler performs $\mathcal{O}((m+t)\cdot K)$ computations and hence it is linear on the size of observed data. In Alg. 1 we assume that the number K of topics is given as input; typically this value is determined experimentally as the number of topics that maximizes the predictive performances. However, it is possible to automatically devise the number of topics by relying on Bayesian nonparametrics. In fact, as shown in [7], it is possible to adapt the sampling equations in order to make explicit the annihilation of some topics as well the generation of new ones, according to the *Chinese Restaurant Process* principle.

Algorithm 1 Gibbs-sampling with parameter estimation

```
Require: G and F, the number of latent features K,
         initial hyperparameter set Ξ.
Random initialization for the matrices Z^e, Z^f, X and Y;
it ← 0
converged ← false
while it < nMaxIt and ¬converged do
   for all observed link l do
      Sample z_l according to Eq. 13 and 14
      Sample x_l according to Eq. 15 and 16
   end for
   for all observed attribute-assignment a do
      Sample z_a according to Eq. 17 and 18
      Sample y_a according to Eq. 19 and 20
   end for
   if (it > burn-in) and (it%sampleLag = 0) then
      Sample A (Eq. 4), θ (Eq. 5), S (Eq. 6), Φ (Eq. 7), and Π
      (Eq. 8);
      Update hyperparameters δ⃗, τ⃗ and ξ⃗;
   end if
   it ← it + 1
end while
```

3.2 Producing explanations

The success of a recommender system does not only depend on its accuracy in inferring and exploiting users' interests, but it also relies on how the deployed recommendations are perceived by the users. Explanations increase the transparency of the recommendation process and may positively contribute in gaining users' trust and satisfaction.

When generating explanations for social recommendations, the first step is to understand if the proposed connection $l=(u,v)$ is social (i.e., such that $x_l=1$) or topical (i.e., $x_l=0$). WTFW provides a natural way to do this:

$$\Pr(x_l=1|l,\Theta)\propto\sum_k \pi_k\delta_k\theta_{k,u}\theta_{k,v} \tag{9}$$

$$\Pr(x_l=0|l,\Theta)\propto\sum_k \pi_k(1-\delta_k)S_{k,u}A_{k,v}, \tag{10}$$

Social connections have a natural explanation in terms of close-knit circles. Thus, for a given link $l=(u,v)$ predicted as social (i.e., such that $x_l=1$), we can provide an explanation as the set of the *most prospective common neighbors*, ranked according to the following score:

$$\mathit{rank}(w;l)=\sum_k \pi_k\delta_k\theta_{k,u}\theta_{k,v}\theta_{k,w}. \tag{11}$$

This rank promotes common neighbors that have high degree of involvement in social communities where both u and v are involved as well. Interestingly, the score finds an explanation in terms of the probability of observing a social triangle among u, v and w. In fact, the joint probability of observing $(u,v),(u,w)$ and (v,w) within community k is proportional to $\theta_{k,u}\theta_{k,v}\theta_{k,w}$. And, since by definition both (u,w) and (v,w) hold in the data, the score explain the prospective new link (u,v) in terms of the common neighbors which are more likely to devise a triangle in the data.

Conversely, topical links can be explained through a list of attributes which are representative of the topics of interest by the current user and for which the recommended connection has high authority. For each feature common to the two nodes, we define the following score:

$$\mathit{rank}(f;l)=\sum_k \pi_k(1-\delta_k)\Phi_{k,f}A_{k,v}\cdot \left(\tau_k A_{k,u}+(1-\tau_k)S_{k,u}\right). \tag{12}$$

Here, the latter term represents the topical involvement of the user u within community k. Again, the score has an interpretation in terms of the prospective triangle among $(u,v),(u,f)$ and (v,f). Notice, however, that the directionality plays a role here, since we are only interested in those features for which v is authoritative.

The procedure for producing explanations for a recommended link is summarized in Alg. 2. In short, the procedure predicts the nature (either social or topical) of the prospective link, hence providing the list of most prominent neighbors/common features.

Algorithm 2 Producing explanations

Require: The social network G, the WTFW model, a recommended link $l = (u, v)$ and the number of explanations L;
Ensure: a list $\mathcal{L}$ of either social or topical explanations for the link.

```
ℒ ← ∅
Compute x_l according to equations 9 and 10
if x_l = 1 then
    ℒ_N ← ∅
    for all w ∈ N(u) ∩ N(v) do
        Compute rank(w, l) according to Eq. 11
        L_N ← L_N ∪ (w, rank(w, l))
    end for
    Sort L_N and compute ℒ = top(L_N, L)
else
    L_F ← ∅
    for all f ∈ F(u) ∩ F(v) do
        Compute rank(f, l) according to Eq. 12
        L_F ← L_F ∪ (f, rank(f, l))
    end for
    Sort L_F and compute ℒ = top(L_F, L)
end if
```

4. EXPERIMENTAL EVALUATION

In this section we report the empirical assessment of the proposed WTFW model on real networks. The experimentation is aimed at assessing the following:

- The accuracy of the model for what concerns both *link prediction* and *label prediction*, where the latter refers to the classification of a link as either social or topical.
- The scalability and stability of the learning procedure, by studying learning time and performance varying the number of iterations of the Gibbs sampler.
- The quality of the associations between links and features, that we show by means of anecdotal evidence in the reconstruction of the data through the model.

Datasets. For our purposes we need datasets coming from social networking platforms in which links creation can be explained in terms of interest identity and/or personal social relations. This requirement is satisfied, among the others, by two popular social networking platforms, namely `Twitter` and `Flickr`. On both platforms, the underlying network is inherently directed to reflect interest of users towards important, and authoritative, information sources. Moreover, in these systems the role of users may naturally change with respect to different topics. The `Twitter` dataset we use is publicly available[3] and it includes information from 973 ego-networks crawled from the public API. The resulting network contains roughly 80 thousand nodes and 1.7 million directed links. Attribute information consists in all the hashtags (e.g. #sanfrancisco) and mentions (e.g. @BarackObama), used by those users.

`Flickr` data has been obtained by querying `Flickr` public API in the time window $2004-2008$ and then by performing forest fire sampling [16] on the resulting network. Features are generated by crawling all the tags used by each users. `Flickr` also contains a form of ground-truth for the label prediction task. Specifically, for each link in the dataset there are two flags, namely *friend* and *family*, that a user can specify. We naturally interpret these flags as follows: a link is labeled as "social" if it is either marked as family or friend. Conversely, a link is "topical" if none of the two flags are set. It is important to stress that this ground-truth is expectedly very noisy as it is any user-declared information on the internet. As such, it is likely to produce an underestimation of the accuracy in the label prediction task.

In order to keep the experimental setting as close as possible to the original data (high dimensionality and exceptional sparsity), no further pre-processing has been performed. Basic statistics about these two datasets are given in Table 2. These datasets are characterized by different properties. The social graph in `Twitter` is much more directed and sparse than in `Flickr`, while the number of attributes per user is much higher in `Flickr`.

	`Twitter`	`Flickr`
Number of nodes	$81,306$	$80,000$
Number of links	$1,768,149$	$14,036,407$
Number of one-way links	$1,342,311$	$9,604,945$
Number of bidirectional links	$425,838$	$4,431,462$
Number of social links	-	$6,747,085$
Number of topical links	-	$7,289,322$
Avg in-degree	21	175
Avg out-degree	25	181
Number of features	$211,225$	$819,201$
Number of feature assignments	$1,102,000$	$37,316,862$
Avg. features per user	15	613
Avg. users per feature	5	45

Table 2: Datasets statistics.

Experimental setting. In all the experiments we assume a partial observation of the network and a complete set of user features.[4] The learning algorithm starts with a random assignments to latent variables, it performs a burn-in phase (burn-in=500) to stabilize the Markov chain, and the parameters of the model are updated at regular intervals (sampling lag=20) for the next 2000 iterations. We initialize hyperparameters with the following (symmetric) values: $\alpha = \beta = \eta = \frac{1}{n}$, $\gamma = \frac{1}{h}$, $\tau_0 = \tau_1 = \delta_0 = \delta_1 = \xi = 2$.

4.1 Model Assessment

Evaluation on link prediction. In a first set of experiments, we measure the accuracy of the model in predicting new links. On `Twitter`, we perform a Monte Carlo Cross-Validation in 5 folds, by randomly splitting the network into training and test data. We also measure the accuracy of the learned models for different proportions of training/test, namely 60/40, 70/30, 80/20. This allows us to stress the robustness of the link prediction task for different proportions, and to mitigate the effects of the random splits. In `Flickr` instead the dataset contains the timestamp of creation of the link, allowing us to perform a chronological split, where older links (70% of the data) are used for learning the model, while the most recent 30% are used as prediction target.

The accuracy of link prediction is measured by computing the area under the ROC curve (AUC) over a set of positive and negative examples drawn from the test set. In principle, we can consider all links in the test-set as positive examples, and all non-existing links as negative example. However, the sparsity of the networks poses two major issues: (i) the number of non-existing links can be enormous, thus making the computation of the AUC infeasible; (ii) missing links do not necessarily represent negative information, but rather unseen information [28]. Following [1], we thus limit the negative examples to all the 2-hops non-existing links.

[3] `http://snap.stanford.edu/data/egonets-Twitter.html`

[4] The task of predicting/recommending missing features is not investigated here and it is left as future work.

		Number of latent factors					
Method	Split	8	16	32	64	128	256
WTFW	60/40	0.567	0.615	0.667	0.707	0.739	**0.792**
	70/30	0.565	0.631	0.680	0.713	0.749	**0.798**
	80/20	0.586	0.639	0.692	0.732	0.760	**0.812**
JSVD	60/40	0.439	0.471	0.525	0.588	0.660	0.768
	70/30	0.446	0.48	0.537	0.602	0.679	0.744
	80/20	0.454	0.495	0.545	0.617	0.693	0.763
CNF	0.7025/0.7125/0.7199						
AA-NF	0.7301/0.7397/0.7472						

Table 3: AUC on link prediction - `Twitter`

	Number of latent factors					
Method	8	16	32	64	128	256
WTFW	0.6467	0.6488	0.6534	0.6576	0.661	**0.677**
JSVD	0.598	0.596	0.597	0.609	0.619	0.624
CNF	0.53					
AA-NF	0.58					

Table 4: AUC on link prediction - `Flickr`

	Number of latent factors					
Method	8	16	32	64	128	256
WTFW	0.7393	0.7548	**0.7603**	0.6883	0.6618	0.6582
Baseline	0.6545					

Table 5: AUC on link labeling - `Flickr`.

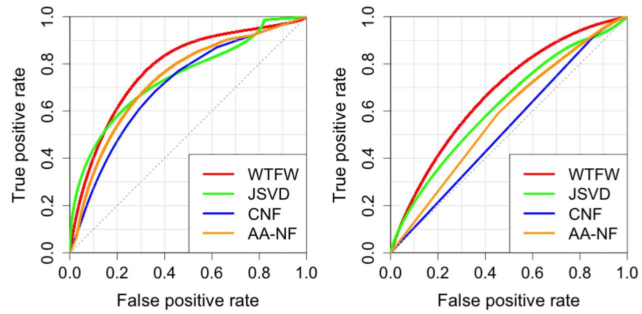
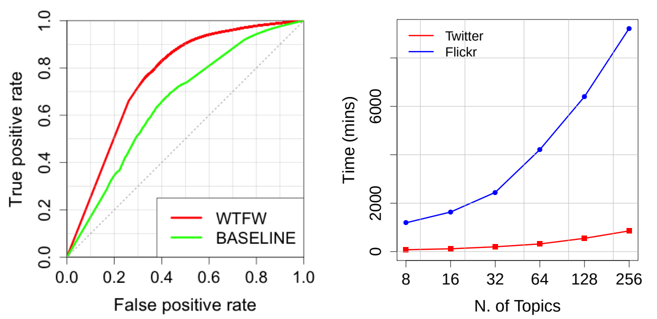


Figure 3: Link prediction: `Twitter` (left) and `Flickr` (right).



(a) (b)

Figure 4: (a) Accuracy of link labeling on `Flickr`. (b) Learning times on the 70/30 split for both `Flickr` and `Twitter`.

We compare the performance of the WTFW model with some popular baseline approaches from the literature, which perform well on a range of networks [18, 9]: *Common Neighbors/Features (CNF)* and *Adamic-Adar (AA-NF)*. CNF is a local similarity index that produces a score for each link (u, v), which is given by the number of common neighbors/features:

$$score(u, v) = |N(u) \cap N(v)| + |F(u) \cap F(v)|.$$

AA-NF represents a refinement of the simple counting of common neighbors/features, which is achieved by assigning more weight to less-connected components.

$$score(u, v) = \sum_{w \in N(u) \cap N(v)} \frac{1}{|N(w)|} + \sum_{f \in F(u) \cap F(v)} \frac{1}{|V(f)|}$$

In addition, we compare WTFW with a matrix factorization approach based on SVD, dubbed *Joint SVD (JSVD)* [9]. In practice, the approach computes a low-rank factorization of the joint adjacency/feature matrix $\mathbf{X} = [E\ F]$ as $\mathbf{X} \approx \mathbf{U} \cdot diag(\sigma_1, \ldots, \sigma_K) \cdot \mathbf{V}^T$, where K is the rank of the decomposition and $\sigma_1, \ldots, \sigma_K$ are the square roots of the K greatest eigenvalues of $\mathbf{X}^T\mathbf{X}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ provide substantial interpretation in terms of connectivity of both E and F. The term $U_{u,k}$ can be interpreted as the tendency of u to be either a source in E or an adopter in F, relative to factor k. Analogously, $V_{u,k}$ represents the tendency of u to appear as a destination in E, and $V_{f,k}$ represents the likelihood that item f is adopted in k. The link prediction score can hence be computed as:

$$score(u, v) = \sum_{k=1}^{K} U_{u,k} \sigma_k V_{v,k}.$$

Tables 3 and 4 summarize the results of the evaluation, for increasing values of the number of latent topics/factors. On `Twitter` data, both WTFW and JSVD underperform when the number of latent factors is limited, but exhibit a competitive advange over the baselines for higher values of K. WTFW outperforms the other considered approaches and these results are stable on different training/test set proportions. The prediction on `Flickr` is in general weaker for all methods. However, the results seems stabler, since the difference with regards to JSVD remains constant for increasing values of K. The standard baselines perform poorly on this dataset. Figure 3 shows the slope of the ROC curves on both datasets for $K = 256$. On `Twitter`, there are some limited areas where the JSVD is skewed but, in general, WTFW clearly outperforms the other methods. This is even more evident on the `Flickr` dataset.

Evaluation on link labeling. We next turn our attention to the task of discriminating between social and topical links, thanks to the ground truth that we have in the `Flickr` dataset. Again, we measure the accuracy by computing the AUC on the prediction, and by comparing the result with a baseline based on common neighbors/features. That is, a link $l = (u, v)$ is deemed social if the weight of the common neighbors is higher than those of the common features, and topical otherwise. Formally:

$$\Pr(x_l = 1|l) = \frac{|N(u) \cap N(v)|}{|N(u) \cap N(v)| + |F(u) \cap F(v)|}.$$

Table 5 reports the results for increasing values of K. The best results are obtains on a lower number of topics, and in particular for $K = 32$. This is somehow surprising if we compare this results with the results on link prediction discussed above. In an attempt to explain such a behavior, we analysed the values of $\mathbf{\Pi}$ and $\boldsymbol{\delta}$ in the model, and we noticed that all models exhibit a strongly dominant latent factor. We will discuss this component also in the next subsection: it is worth mentioning, however that the associated probability δ_k leans towards 0.5 (a clear sign that the community tends to mix topical and social contributions). Clearly, the balanced value of δ_k does not affect the performance in link prediction (as it only depends on whether any of the social/topical components is strong enough to trigger

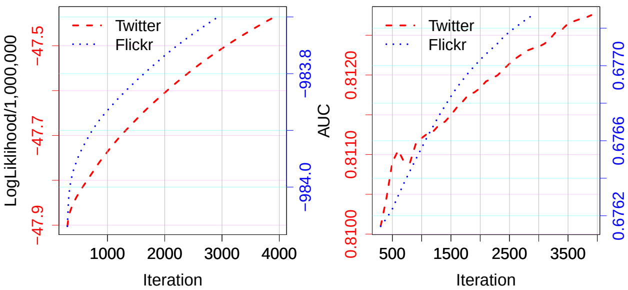


Figure 5: log likelihood and accuracy along the iterations.

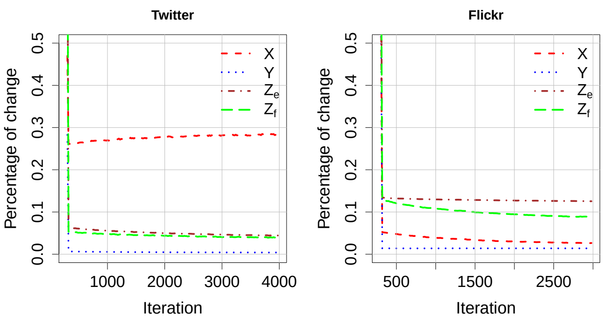


Figure 6: Changes in the matrices $\mathbf{Z}^e$, $\mathbf{Z}^f$, $\mathbf{X}$ and $\mathbf{Y}$.

the link), but it can negatively affect the label prediction. The degradation of the performance for higher values of K can find a justification in the split of this giant component: apparently, the splitting seems to produce a reallocation of the links in the other communities, thus causing the overfitting. Besides this anomalous behavior, WTFW outperforms the baseline prediction for each considered value of K, as shown in Fig. 4(a).

Scalability and sensitivity analysis. We next discuss how the model parameters affect the learning time and the quality of the results. As already mentioned, we chose to perform $2,000$ Gibbs sampling iterations. Figure 4 shows that the learning is substantially linear in the number of topics. Figure 5 shows the behavior of the log-likelihood (on the left) and the AUC values on the validation data, along the iterations. We can see that, although the likelihood increases substantially, the improvements on the AUC is marginal, and the algorithm tends to converge to a stable accuracy value quite rapidly. Hence, limiting the number of iterations to $2,000$ seems to be a good compromise between learning time and accuracy in prediction.

Finally, Figure 6 shows the percentage of relocations in latent factors assignments in the sampling steps of the Gibbs sampler, along iterations. After an initial phase, the latent factor matrices tend to become stable, except for the matrix $\mathbf{X}$ on `Twitter`, for which there is constant 25% change along all iterations. Also, it is interesting to notice that the changes in $\mathbf{Z}^e$ and $\mathbf{Z}^f$ tend to be higher on `Flickr`.

4.2 Qualitative analysis

We next turn to a more qualitative analysis of the models produced. Here we consider the latent structure corresponding to $K = 64$ on `Twitter` and $K = 32$ on `Flickr`. These values are chosen for presentation readability sake.

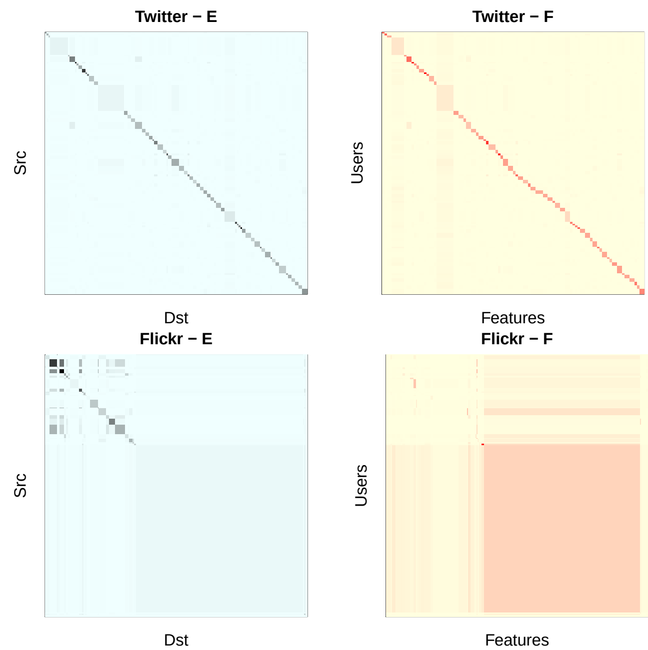


Figure 7: Density plot of the blocks within $\mathbf{E}$ and $\mathbf{F}$.

Figure 7 shows the block adjacency and feature matrices, where both users and features are grouped according to their likelihood to belong to each community. Specifically, for each user, we compute the probability $\Pr(k|u) \propto \sum_k \pi_k \Pr(u|k)$, where $\Pr(u|k)$ is the probability that a user is associated to community k as either source/destination of a connection, or as adopter of a feature. Analogously, $\Pr(k|f) \propto \sum_k \pi_k \Phi_{k,f}$ represents the likelihood that the feature f is observed in the context of the community k. Each user/feature is then associated with the community k for which the probability is higher.

Darker colors in the block matrices denote higher relative density. As witnessed by the diagonal structure on the block adjacency matrix, the discovered communities capture neat patterns of connectivity among users. The joint analysis of the two block matrices supports the following findings.

- Connections in `Twitter` tend to be strongly topical; each community is strongly characterized by a corresponding set of features. The only exception is given by two communities (4 and 15) in which associations tend to be looser (and in fact the corresponding features tend to entropically spread along all communities).
- The connectivity behavior of users on `Flickr` is more social. The block adjacency matrix exhibits high density blocks, and by contrast the feature block matrix exhibits higher entropy. Apparently, it seems that there is not a strong characterization in terms of features for the detected communities. While few communities are specialized in a very limited set of features, a larger number of tags is popular in the remaining ones. The dominant community mentioned before, where most of the features association hold is very visible.

The above hypotheses are supported by the analysis in Figure 8, where we report for each community the percentage of users who qualify as either authoritative, susceptible or social. The authoritiveness score for the user u within k

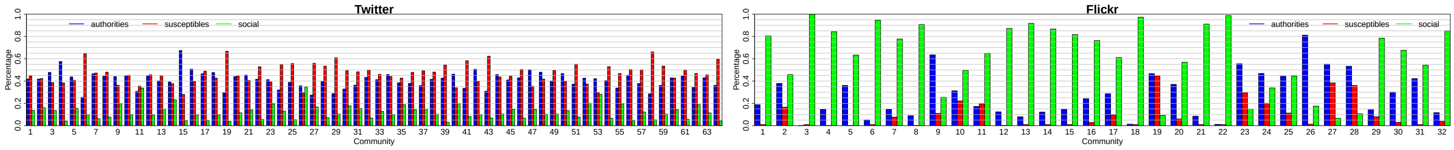


Figure 8: Proportion of authoritative, susceptible and social users in both `Twitter` and `Flickr`.

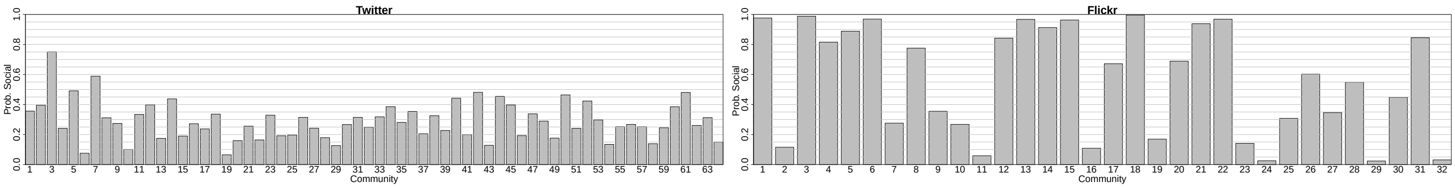


Figure 9: Distribution of δ_k for each community.

is computed as:

$$\Pr(Auth|u,k) = \frac{(1-\delta_k)A_{k,u}}{\delta_k\theta_{k,u} + (1-\delta_k)(A_{k,u}+S_{k,u})}.$$

Scores for measuring the degree of susceptibility and sociality can be computed likewise. Again, `Twitter` tends to exhibit a predominant amount of authorities/followers, whereas the great majority of the communities in `Flickr` tend to include social users. Finally, Figure 9 shows the estimated values of δ_k within the communities. As already observed before, `Flickr` tends to be more clearly social than `Twitter`. Moreover, the distinction between social and topical latent factors is very clear in `Flickr` with almost all communities having probability of being "social" either above 0.9, or below 0.2. This discriminating attitude is even more evident when we increase the number of latent topics.

Finally, a further assessment of the correct identification of social and topical latent factors can be performed by measuring, for a given feature f, the probability of observing it in a social/topical context, computed as follows:

$$\Pr(Social|f) \propto \sum_k \pi_k \delta_k \Phi_{k,f}$$

$$\Pr(Topical|f) \propto \sum_k \pi_k (1-\delta_k) \Phi_{k,f}.$$

In Table 6 we validate the accuracy of the feature-labeling task on two small sets of tags. The first set contains keywords notably associated with social events, such as *family*, and *wedding*. The second set contains keywords specific to photographic techniques, e.g. *hdr* and *polaroid*, which are likely to generate topical interest in the users. The results confirm the capability of the model to discriminate between social and topical features.

Finally, Table 7 summarizes the top keywords detected by our approach on both datasets in some representative communities/topics: highly social communities (large δ) have characteristic features (e.g., *family*, *christmas* on `Flickr`, and *followback* on `Twitter`) which are clearly social.

5. CONCLUSIONS AND FUTURE WORK

This paper introduces WTFW, a novel stochastic generative model that jointly factorizes both social connections and feature associations. The model provides accurate link prediction and contextualized socio/topical explanations to support the predictions. Our approach is based on latent factors which can be interpreted as communities of people sharing a similar behavior, and on the explicit modeling of the underlying latent nature behind each observed connection. The result is a decoupling of social and topical connections to reflect the idea that social communities should have high density and reciprocal connections, whereas topical communities should exhibit clear directionality and a low entropy over user attributes.

Feature	Prob. Social
birthday	0.69
family	0.67
wedding	0.69
party	0.67
puppy	0.69

Feature	Prob. Social
hdr	0.40
vintage	0.29
collage	0.24
nude	0.08
polaroid	0.28

Table 6: Social/Topical connotations of selected tags on `Flickr`.

Flickr			
Topic 1 $\delta = 0.98$	*Topic 5* $\delta = 0.98$	*Topic 18* $\delta = 0.17$	*Topic 22* $\delta = 0.14$
Christmas, esther, passenger, Birthday, eros, party, stories, apple, curling, homemade	family, mom, dog, driving, vitus, bakery, woods, birthday, friends, halloween, shirt, brothers, baby	handmade, warehouse, vintage, knitting, craft, green, pansies, doll, sewing	bird, art, design, illustration, drawing, fotoincatenate, sketch, street, painting, ink, graffiti

Twitter			
Topic 3 $\delta = 0.74$	*Topic 9* $\delta = 0.27$	*Topic 64* $\delta = 0.16$	*Topic 47* $\delta = 0.33$
TeamFollowBack TFB FollowNGain fb InstantFollowBack nowplaying lastfm TeamAutoFollow Follow4Follow 500aDay anime 4sqDay	Autodesk BIM AutoCAD Revit AU2012 Civil3D AEC adsk_sf2012 SWTOR revit CAD au2011 cloud 3dsMax AU2011 C3D2013	ISS space science Discovery Mars nasa spottheshuttle ESA astronomy Enterprise Soyuz	GameofThrones FakeWesteros GoT ooc SXSWesteros TheGhostofHarrenhal GardenofBones asoiaf GRRM GOT

Table 7: Most representative features of selected communities.

Our work can be extended in several directions. First, we deliberately omitted to quantify the quality of the communities that the model produces. Some initial results based on modularity look promising (we measured 0.55 and 0.37 on `Twitter` and `Flickr`, respectively). Also, the qualitative analysis in Section 4 clearly denotes the capability of the model to group users according to both connectivity and common features. However, we plan to devote to future work a more detailed treatment, as well a thorough comparison with other approaches in the literature.

Second, the approach explored in this paper is rooted on mixture membership topic modeling. However, other alternatives are possible, which can be based on probabilistic matrix factorization. We plan to explore and compare these different strategies in a future work.

Table 8: Equations for the Gibbs Sampling.

$$Pr(z_l = k|x_l = 1, Rest) \propto (c_k + d_k + \xi_k - 1) \cdot \frac{c_k^s + \delta_0 - 1}{c_k + \delta_0 + \delta_1 - 1} \cdot \frac{\left(c^{s,s}_{k,u_l} + c^{s,d}_{k,u_l} + \alpha_{u_l} - 1\right) \cdot \left(c^{s,s}_{k,v_l} + c^{s,d}_{k,v_l} + \alpha_{v_l} - 1\right)}{\left(2c_k^s + \sum_u \alpha_u - 1\right)\left(2c_k^s + \sum_u \alpha_u\right)} \quad (13)$$

$$\Pr(z_l = k|x_l = 0, Rest) \propto (c_k + d_k + \xi_k - 1) \cdot \frac{c_k^t + \delta_1 - 1}{c_k + \delta_0 + \delta_1 - 1} \cdot \frac{c^{t,s}_{k,u_l} + d^s_{k,u_l} + \eta_{u_l} - 1}{c_k^t + d_k^s + \sum_u \eta_u - 1} \cdot \frac{c^{t,d}_{k,v_l} + d^a_{k,v_l} + \beta_{v_l} - 1}{c_k^t + d_k^a + \sum_u \beta_u - 1} \quad (14)$$

$$\Pr(x_l = 1|Rest) \propto \frac{\left(c^{s,s}_{k,u_l} + c^{s,d}_{k,u_l} + \alpha_u - 1\right) \cdot \left(c^{s,s}_{k,v_l} + c^{s,d}_{k,v_l} + \alpha_v - 1\right)}{\left(2c_k^s + \sum_u \alpha_u - 1\right)\left(2c_k^s + 1 + \sum_u \alpha_u - 1\right)} \cdot \frac{c_k^s + \delta_0 - 1}{c_k^s + c_k^t + \delta_0 + \delta_1 - 1} \quad (15)$$

$$\Pr(x_l = 0|Rest) \propto \frac{c^{t,s}_{k,u_l} + d^s_{k,u_l} + \eta_{u_l} - 1}{c_k^t + d_k^s + \sum_u \eta_u - 1} \cdot \frac{c^{t,d}_{k,v_l} + d^a_{k,v_l} + \beta_{v_l} - 1}{c_k^t + d_k^a + \sum_u \beta_u - 1} \cdot \frac{c_k^t + \delta_1 - 1}{c_k^s + c_k^t + \delta_0 + \delta_1 - 1} \quad (16)$$

$$\Pr(z_a = k|y_a = 1, Rest) \propto (c_k + d_k + \xi_k - 1) \cdot \frac{d_k^a + \tau_0 - 1}{d_k^a + d_k^s + \tau_0 + \tau_1 - 1} \cdot \frac{c^{t,d}_{k,u_a} + d^a_{k,u_a} + \beta_{u_a} - 1}{c_k^t + d_k^a + \sum_u \beta_u - 1} \cdot \frac{d_{k,f_d} + \gamma_{f_a} - 1}{d_k + \sum_f \gamma_f - 1} \quad (17)$$

$$Pr(z_a = k|y_a = 0, Rest) \propto (c_k + d_k + \xi_k - 1) \cdot \frac{d_k^s + \tau_1 - 1}{d_k^a + d_k^s + \tau_0 + \tau_1 - 1} \cdot \frac{c^{t,s}_{k,u_a} + d^s_{k,u_a} + \eta_u - 1}{c_k^t + d_k^s + \sum_u \eta_u - 1} \cdot \frac{d_{k,f_a} + \gamma_{f_a} - 1}{d_k + \sum_f \gamma_f - 1} \quad (18)$$

$$\Pr(y_a = 1|Rest) \propto \frac{c^{t,d}_{k,u_a} + c^a_{k,u_a} + \beta_{u_a} - 1}{c_k^t + d_k^a + \sum_u \beta_u - 1} \cdot \frac{d_k^s + \tau_1 - 1}{d_k^a + d_k^s + \tau_0 + \tau_1 - 1} \quad (19)$$

$$\Pr(y_a = 0|Rest) \propto \frac{c^{t,s}_{k,u_a} + d^s_{k,u_a} + \eta_{u_a} - 1}{c_k^t + d_k^s + \sum_u \eta_u - 1} \cdot \frac{d_k^a + \tau_0 - 1}{d_k^a + d_k^s + \tau_0 + \tau_1 - 1} \quad (20)$$

Acknowledgments. This work was partially supported by MULTISENSOR project, funded by the European Commission, under the contract number FP7-610411.

6. REFERENCES

[1] L. Backstrom and J. Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *WSDM*, 2011.

[2] N. Barbieri, F. Bonchi, and G. Manco. Cascade-based community detection. In *WSDM*, 2013.

[3] C. Bishop. *Pattern Recognition and Machine Learning.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[4] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.

[5] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes. Social network analysis and mining for business applications. *ACM TIST*, 2(3):22, 2011.

[6] E. Erosheva, S. Fienberg, and J. Lafferty. Mixed-membership models of scientific publications. *PNAS*, 101:5220–5227, 2004.

[7] S. Gershman and D. Blei. A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12, 2012.

[8] L. Getoor and C. P. Diehl. Link mining: a survey. *SIGKDD Explorations*, 7(2):3–12, 2005.

[9] N. Z. Gong, A. Talwalkar, L. W. Mackey, L. Huang, E. C. R. Shin, E. Stefanov, E. Shi, and D. Song. Joint link prediction and attribute inference using a social-attribute network. *ACM TIST*, 5(2), 2014.

[10] P. A. Grabowicz, L. M. Aiello, V. M. Eguiluz, and A. Jaimes. Distinguishing topical and social groups based on common identity and bond theory. In *WSDM*, 2013.

[11] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh. Wtf: the who to follow service at twitter. In *WWW*, 2013.

[12] M. A. Hasan and M. J. Zaki. A survey of link prediction in social networks. In *Social Network Data Analytics*, 2011.

[13] G. Heinrich. Parameter Estimation for Text Analysis. Technical report, University of Leipzig, 2008.

[14] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *CSCW*, 2000.

[15] M. Jamali, T. Huang, and M. Ester. A generalized stochastic block model for recommendation in social rating networks. In *RecSys*, 2011.

[16] J. Leskovec and C. Faloutsos. Sampling from large graphs. In *KDD*, 2006.

[17] D. Liben-Nowell and J. M. Kleinberg. The link prediction problem for social networks. In *CIKM*, 2003.

[18] L. Lü and T. Zhou. Link prediction in complex networks: A survey. *Physica A*, 390(6):1150 – 1170, 2011.

[19] K. T. Miller, T. L. Griffiths, and M. I. Jordan. Nonparametric latent feature models for link prediction. In *NIPS*, 2009.

[20] T. Minka. Estimating a Dirichlet distribution. Web, 2000.

[21] F. Moser, R. Colak, A. Rafiey, and M. Ester. Mining cohesive patterns from graphs with feature vectors. In *SDM*, 2009.

[22] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen. Joint latent topic models for text and citations. In *KDD*, 2008.

[23] S. Pool, F. Bonchi, and M. van Leeuwen. Description-Driven Community Detection. ACM TIST 5(2): 28, 2014.

[24] D. A. Prentice, D. T. Miller, and J. R. Lightdale. Asymmetries in Attachments to Groups and to their Members: Distinguishing between Common-Identity and Common-Bond Groups. *Personality and Social Psychology Bulletin*, 20(5):484–493, 1994.

[25] P. Pu and L. Chen. Trust building with explanation interfaces. In *IUI*, 2006.

[26] Y. Ren, R. Kraut, and S. Kiesler. Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3):377–408, 2007.

[27] M. Sachan, D. Contractor, T. A. Faruquie, and L. V. Subramaniam. Using content and interactions for discovering communities in social networks. In *WWW*, 2012.

[28] V. Sindhwani, S. Bucak, J. Hu, and A. Mojsilovic. One-class matrix completion with low-density factorizations. In *ICDM*, 2010.

[29] B. Taskar, M. F. Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In *NIPS*, 2003.

[30] N. Tintarev and J. Masthoff. Effective explanations of recommendations: user-centered design. In *RecSys*, 2007.

[31] X. Wang, L. Tang, H. Gao, and H. Liu. Discovering overlapping groups in social media. In *ICDM*, 2010.

[32] Z. Xu, Y. Ke, Y. Wang, H. Cheng, and J. Cheng, A Model-based Approach to Attributed Graph Clustering. In *SIGMOD*, 2012.

[33] Z. Yin, M. Gupta, T. Weninger, and J. Han. Linkrec: a unified framework for link recommendation with user attributes and graph structure. In *WWW*, 2010.

[34] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu. Stochastic relational models for discriminative link prediction. In *NIPS*, 2006.

[35] Y. Zhou, H. Cheng, and J. X. Yu. Graph clustering based on structural/attribute similarities. *PVLDB*, 2(1):718–729, 2009.