

# Modeling Topic Diffusion in Multi-Relational Bibliographic Information Networks

Huan Gui<sup>†</sup> Yizhou Sun<sup>‡</sup> Jiawei Han<sup>†</sup> George Brova<sup>†</sup>

<sup>†</sup> University of Illinois at Urbana-Champaign, Urbana, IL

<sup>‡</sup> Northeastern University, Boston, MA

<sup>†</sup>{huangui2, hanj, brova2}@illinois.edu <sup>‡</sup>yzsun@ccs.neu.edu

## ABSTRACT

Information diffusion has been widely studied in networks, aiming to model the spread of information among objects when they are connected with each other. Most of the current research assumes the underlying network is homogeneous, i.e., objects are of the same type and they are connected by links with the same semantic meanings. However, in the real word, objects are connected via different types of relationships, forming multi-relational heterogeneous information networks.

In this paper, we propose to model information diffusion in such multi-relational networks, by distinguishing the power in passing information around for different types of relationships. We propose two variations of the linear threshold model for multi-relational networks, by considering the aggregation of information at either the model level or the relation level. In addition, we use real diffusion action logs to learn the parameters in these models, which will benefit diffusion prediction in real networks. We apply our diffusion models in two real bibliographic information networks, DBLP network and APS network, and experimentally demonstrate the effectiveness of our models compared with single-relational diffusion models. Moreover, our models can determine the diffusion power of each relation type, which helps us understand the diffusion process better in the multi-relational bibliographic network scenario.

## Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data mining*

## Keywords

Multi-relational information networks; information diffusion; action prediction

## 1. INTRODUCTION

Information diffusion in networks has been studied widely in different domains, such as social science [9, 18], network science [8, 14], computer science [25], and the medical domain [1, 12, 5]. Many applications have benefited from these studies, such as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM'14, November 3–7, 2014, Shanghai, China.

Copyright 2014 ACM 978-1-4503-2598-1/14/11 ...\$15.00.  
http://dx.doi.org/10.1145/2661829.2662000.

diffusion-based prediction [26, 25], influence maximization [14], word-of-mouth marketing [16], cyber security [25], and social media analysis [8, 10, 23].

### Information Diffusion Models.

The information diffusion process can be largely described as: some *active* nodes will influence their *inactive* neighbors in the network and turn them into active nodes with a certain probability, and the newly activated nodes can progressively trigger their neighbors into becoming active. A node is called active if it has already taken the action related to the information. For example, in an iPhone social marketing scenario, a consumer will be called active if she has already purchased the device. Many diffusion models have been proposed, such as linear threshold model [9, 24, 18], independent cascade model [6], decreasing cascade model [15], general threshold model [14], heat diffusion-based model [17], and so on. In this paper, we will focus on the Linear Threshold Model (LTM).

Like many classical diffusion models, LTM is defined in the setting with discrete timestamps, and the activation probability of a node  $u$  at time  $t + 1$  will be defined according to the activation set at time  $t$ . To be more specific, in LTM, each node  $u$  will be activated if and only if the total weight of its active neighbors is at least  $\theta_u$ :

$$\sum_{v \in \Gamma(u)} w_{v,u} \delta(v, t) > \theta_u \quad (1)$$

where  $\Gamma(u)$  denotes the neighbor set of node  $u$ ,  $w_{v,u}$  denotes the influence strength from  $v$  to  $u$ ,  $\delta(v, t) = 1$  if  $v$  is activated at time  $t$  and 0 otherwise, and  $\theta_u$  is the system-setting threshold.

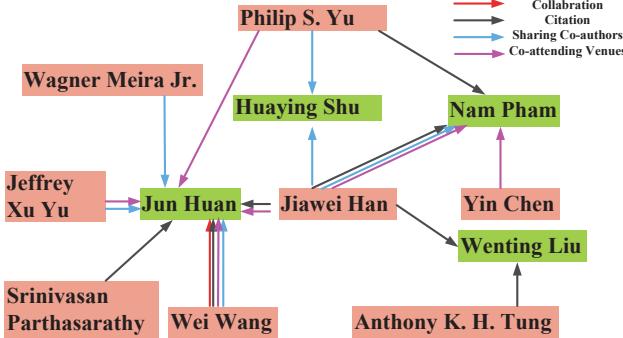
### Multi-relational Information Networks.

For most of current diffusion studies, the diffusion models are built upon the assumption that the underlying network is *homogeneous*, i.e., objects are of the same type and they are connected by links with the same semantic meanings. For example, in a topic diffusion case in academia, the objects in the network are researchers, and the links in the network are collaboration relationships. In a social network case, the objects are people, and they are connected to each other by friendships.

However, in the real word, objects are usually connected via different types of relationships, forming *multi-relational networks* [27, 22]. For example, in the bibliographic information network case, researchers could be linked together via different types of relationships: collaboration relationships, citation relationships, sharing common co-authors, co-attending conferences, etc. In the social network case, people are connected via friendships, colleague relationships, family relationships, etc.

Based on the observation that many different relation types can be derived from different types of interactions, we can generate a

multi-relational bibliographic information network via meta-path-based techniques [28]. To be more specific, researchers can be connected via co-authorship (*author* – *paper* – *author*), co-attending conferences (*author* – *paper* – *venue* – *paper* – *author*), sharing co-authors (*author* – *paper* – *author* – *paper* – *author*), and citation relationships (*author* – *paper* → *paper* – *author*). A real-world example is shown in Figure 1, where authors are connected by four meta-path-based relation types.



**Figure 1:** An illustrative example of a multi-relational network, where topic “Association Rules” is propagating along all the types of relationships among authors. Red nodes are authors who have already adopted the topic, while green nodes represent authors being triggered to adopt the topic at the next timestamp.

Formally, a multi-relational information network  $G = (V, E)$  is a special type of heterogeneous information network, which contains one type of objects, i.e.,  $V$ , and multiple types of relationships between these objects, i.e.,  $E = E_1 \cup E_2 \cup \dots \cup E_K$ , where  $E_k$  represents  $k$ th type of relationships.

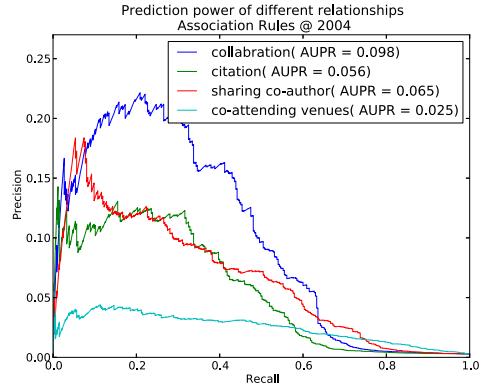
### Information Diffusion in Multi-relational Information Networks.

As multi-relational information network is ubiquitous in the real world, an interesting question is how we can utilize different relation types to make prediction over future diffusion behavior?

We take Figure 1 as an example, studying how the topic “association rule” propagates in this multi-relational network. We apply LTM to each relation type, while ignoring influence from the remaining relation types. The predictions precision about the activated node set at the next timestamp are shown in Figure 2. According to the results, every relation type has positive prediction power, i.e., all these relation types are involved in the diffusion process. In this case, if we only model influence of one relation type, there would be great information loss for the remaining relation types.

So, how about applying LTM to all relationships by ignoring the relation type information? The answer is no. The heterogeneity of the prediction accuracy of different relation types implies that each relation type plays a different role in the diffusion process. For example, in the topic diffusion example, influence from a researcher’s coauthors would be rather different from the ones she met in conferences.

In this paper, we address the problem of modeling information diffusion in multi-relational information networks, using topic diffusion in multi-relational bibliographic networks as a case study, and propose more general but effective models that can leverage the relation type information in such networks. There are two major challenges in providing such models. First, how can we model the diffusion process in multi-relational networks, with heterogeneous diffusion power for each relation type? Second, for a particular



**Figure 2:** Prediction results on diffusion processes of topic “Association Rules” propagating along different types of relationships.

network and a concrete diffusion task, how can we automatically determine the best weight for each relation type?

In order to solve the first challenge on modeling, we extend LTM, the well-known diffusion model in single-relational information networks, into two variational models for multi-relational information networks, which we call *MLTM-M* and *MLTM-R*. *MLTM-M* is a **M**ulti-relational **L**inear **T**hreshold **M**odel that combines each relation type at the **M**odel output level. In other words, the information first diffuses along each type of relationships independently, and whether an object is activated or not is an aggregation of model outputs from all these relation types. *MLTM-R* is a **M**ulti-relational **L**inear **T**hreshold **M**odel that combines each relation type at the **R**elation level. In other words, the multi-relational network is treated as a single-relational network by putting different weights on different types of links, and then apply the single-relational linear threshold model to determine the activation probability.

In order to determine the weights for the models and therefore make the models applicable for diffusion prediction in the real world, we propose to use diffusion action logs to learn the parameters in these models. The diffusion action logs record the object set that is activated at each timestamp. By maximizing the likelihood of observing the action logs, either obtained from one cascade or multiple cascades, we can find the MLE estimators for the parameters using optimization methods. With the learned parameters, either the weight of each relation-based diffusion model or the weight for relationships of each relation type, we can not only understand the role of each relation type in the diffusion process but also predict the diffusion according to given initial set of activated objects.

We apply our diffusion models to two real multi-relational bibliographic information networks, DBLP network and APS networks. It turns out that, by distinguishing the relation types in multi-relational information networks, our diffusion models can not only produce higher prediction accuracy, but also have more interpretability in explaining diffusion processes with regards to heterogeneous diffusion power of different relation types.

In summary, we make the following contributions in this paper:

1. We propose the problem of studying information diffusion in multi-relational information networks, where objects are connected to each other due to different reasons and with different semantic meanings, and the diffusion power is no longer homogeneous on the links from different types.
2. We propose two novel diffusion models that encode the diffusion power for each relation type in the multi-relational networks.

3. Unlike most other diffusion studies, we propose to learn the parameters of our diffusion models from action logs, which can help us to understand how different relation types play different roles in the diffusion process and to predict future diffusion.

4. We test our model in two real-world multi-relational bibliographic networks, DBLP and APS network, which demonstrates the superiority of our model compared with baselines.

The rest of paper is organized as follows. In Section 2, two multi-relational diffusion models are proposed, and the learning algorithms for these two models are provided in Section 3. Experiments are introduced in Section 4. Section 5 presents a brief summary of related work. Finally, Section 6 concludes the work.

## 1.1 Information Diffusion Models in Single-Relational Networks

Information diffusion has been widely studied in single-relational information networks. The information diffusion process can be largely described as: some *active* nodes will influence their *inactive* neighbors in the network and turn them into active nodes with a certain probability, and the newly activated nodes can progressively trigger their neighbors into being active. A node is called active if it has already taken the action related to the information. For example, in the research topic diffusion scenario, an author will be called active if she has already adopted the research topic.

Many diffusion models have been proposed, such as linear threshold model [9, 24, 18], independent cascade model [6], decreasing cascade model [15], general threshold model [14], heat diffusion-based model [17], and so on. Most of these models are defined in the setting with discrete timestamps, and the activation probability of a node  $u$  at time  $t + 1$  will be defined according to the activation set at time  $t$ .

In this paper, we focus on the linear threshold model, and propose two extensions in the multi-relational information network scenario. Linear Threshold Model was proposed by Granovetter in [9], where the activation probability of a node  $u$  is dependent on its activated neighbors  $v$ 's. Each node has a threshold  $\theta_u$  in the range of  $[0, 1]$ .  $u$  is activated if and only if the activation probability is larger than the preset threshold  $\theta_u$ :

$$\sum_{v \in \Gamma(u)} w_{v,u} \delta(v, t) > \theta_u \quad (2)$$

where  $\Gamma(u)$  denotes the neighbor set of node  $u$ ,  $w_{v,u}$  denotes the influence strengths from  $v$  to  $u$ , and  $\delta(v, t) = 1$  if  $v$  is active at time  $t$  and 0 otherwise.

## 2. DIFFUSION MODELS FOR MULTI-RELATIONAL NETWORKS

In this section, we propose diffusion models for multi-relational networks, by extending the single-relational LTM. In these models, we want to explore how relation type information can be encoded in the diffusion models.

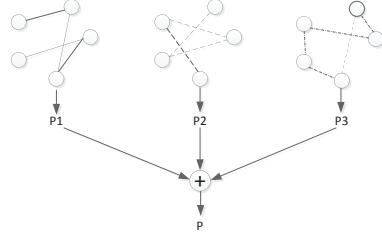
### 2.1 Overview

Unlike single-relational diffusion models, objects are now connected to other objects via *different types of relationships*, and influence propagates to objects by all these relation types simultaneously. We make the assumption that different types of relationships should carry different levels of diffusion power. For example, researchers found that different types of relationships have different impacts on obesity spreading [5].

Therefore, we encode the heterogeneous diffusion power of each relation type in the diffusion process. Moreover, we aim to test the

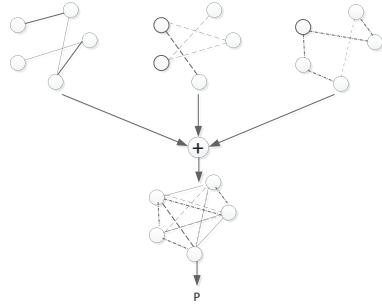
following two assumptions on how different types of relationships work together in propagating information.

- Assumption 1 (Relation Independent Diffusion). Under this assumption, the information is propagated along each relation type independently, and then the activation probability of an object is an weighted aggregation of the activation probabilities for all the types of relationships, as shown in Figure 3.



**Figure 3:** Illustration of Assumption 1, Relation Independent Diffusion

- Assumption 2 (Relation Interdependent Diffusion). Under this assumption, the information is propagated on the mixed set of links from any types of relationships, with the weights of different relation types being treated differently, as shown in Figure 4.



**Figure 4:** Illustration of Assumption 2, Relation Interdependent Diffusion

We now propose two variations of LTM under the two assumptions, which are introduced in detail in the remainder of this section. We use  $\beta_k$  to denote the hyper-level weight for relation type  $k$ , and  $w_{u,v}^{(k)}$  to denote the weight of the link between  $u$  and  $v$  of relation type  $k$ . More notations are summarized in Table 1.

### 2.2 Multi-Relational Linear Threshold Model (MLTM)

We now extend LTM into the multi-relational case. In particular, we assume the threshold of every object follows a uniform distribution in  $[0, 1]$ , such that the weighted percentage of the activated neighbors determines the object activation probability, where the weight is determined by the weight of the link. We consider the extensions under the two assumptions, when there are multiple types of links involved, in the following sections.

#### 2.2.1 Relationship Weighting

Before proceeding to the proposed models, the diffusion strength (weighting) of every pair of nodes with regards to different relation types needs to be estimated. Remark that our models are orthogonal

Notation	Meaning
$u, v, i, j$	object index
$k$	relation type index
$E_k$	relationship set of $k$ th type
$\Gamma_t(i, k)$	neighbor set of object $i$ with respect to relationship set $E_k$ at timestamp $t$
$w_{ij,t}^{(k)}$	weight of the relationship between $i$ and $j$ of type $k$ at timestamp $t$
$\delta(j, t)$	1, if node $j$ is activated at timestamp $t$ ; 0, otherwise
$p_u(t)$	the activation probability of node $u$ at time $t$
$V(t)$	objects activated at timestamp $t$
$R_t$	objects not activated yet by timestamp $t$ ( $t$ not included)

**Table 1:** Notation Table.

to the weighting schema, so that any reasonable diffusion strength estimation methodologies would fit in the models. In particular, we use the pathsim strategy proposed in [28] to calculate the diffusion strength. Given relation type  $E_k$  (symmetric meta-path), the weight between  $u$  and  $v$  can be defined as

$$w_{uv} = \frac{2|P_{(u,v)}^{E_k}|}{|P_{(u,u)}^{E_k}| + |P_{(v,v)}^{E_k}|},$$

where  $P_{(u,v)}^{E_k}$  is the set of meta-paths (interactions), of relation type  $E_k$ , starting from  $u$  and ending at  $v$ , and  $|\cdot|$  denotes the size of the set. More details on meta-path and pathsim can be found in [28].

### 2.2.2 MLTM-M: Model-Level Aggregation

Under the first assumption, the multi-relational network can be considered as  $K$  independent networks over the same set of objects. We can calculate the activation probability for an object  $i$  under each relation type  $k$ :

$$p_i^{(k)}(t+1) = \frac{\sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} \delta(j, t)}{\sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)}} \quad (3)$$

where  $\Gamma_t(i, k)$  represents the neighbor set of object  $i$  with respect to relationship set  $E_k$  at timestamp  $t$ ,  $w_{ij,t}^{(k)}$  denotes the weight of link between  $i$  and  $j$  for relation type  $k$ , and  $\delta(j, t)$  denotes whether  $j$  is activated at timestamp  $t$ .

Then the activation probability for object  $i$  is an aggregated function of activation probability of  $i$  in each relation type that  $p_i(t+1) \propto \sum_k \beta_k p_i^{(k)}(t+1)$ . Without loss of generality, we choose logistic function as the aggregation function to integrate the propagation factors for all different types of relationships:

$$p_i(t+1) = \frac{e^{\sum_k \beta_k p_i^{(k)}(t+1) + \beta_0}}{1 + e^{\sum_k \beta_k p_i^{(k)}(t+1) + \beta_0}} \quad (4)$$

where  $\beta_0$  represents the weight for the constant factor, which can be considered as the diffusion factor of global effect [20], such as influence from the mass media, that cannot be captured by the network typologies. Under this model, several interesting properties can be derived, which disclose the impact of each individual relation type to the overall diffusion process.

**PROPERTY 1.** Let  $p_i^{(-k)}(t+1)$  be the activation probability of  $i$  at timestamp  $t+1$  without relation type  $k$ , under MLTM-M, we have

- if  $\beta_k > 0$ , then  $p_i(t+1) > p_i^{(-k)}(t+1)$ , that is, a relation type with positive diffusion power will increase the activation probability of  $i$ ;

- if  $\beta_k = 0$ , then  $p_i(t+1) = p_i^{(-k)}(t+1)$ , that is, a relation type with no diffusion power (such as noise) will not change the activation probability of  $i$ ; and

- if  $\beta_k < 0$ , then  $p_i(t+1) < p_i^{(-k)}(t+1)$ , that is, a relation type with negative diffusion power will decrease the activation probability of  $i$ .

PROOF. Proof is omitted due to space limitation.  $\square$

According to this property, adding a new type of relationships always leads to the increase of the chance for an object to be activated, as long as the relation type has positive diffusion power. We can expect the phenomenon in the real world scenarios such as disease propagation: if a disease can be propagated via multiple channels, the exposure of a person to more channels of the pathogens leads to a higher probability for her to get infected.

### 2.2.3 MLTM-R: Relation-Level Aggregation

Unlike the first assumption, Assumption 2 stipulates that the multi-relational network can be converted to a single-relational network by aggregating all the edges together, where links from different types carry different hyper-level weights. Under this assumption, the activation probability of object  $i$  at timestamp  $t+1$  can be defined as:

$$p_i(t+1) = \frac{\sum_k \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} \delta(j, t) + \beta_0 \sum_j \delta(j, t)}{\sum_k \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} + \beta_0 N} \quad (5)$$

where  $\Gamma_t(i, k)$  represents the neighbor set of object  $i$  with respect to relationship set  $E_k$  at timestamp  $t$ ,  $w_{ij,t}^{(k)}$  denotes the weight of link between  $i$  and  $j$  for relation type  $k$ ,  $\delta(j, t) = 1$  if object  $j$  is activated at  $t$ , and 0 otherwise,  $N$  is the total number of objects in the network. Note that we also include a constant factor with weight  $\beta_0$ , which can be viewed as global effect. Furthermore, remark that, in Eq.(5),  $p_i(t+1)$  would only be influenced by the ratios of  $\beta_k$ 's, i.e.,  $\frac{\beta_k}{\beta_0}$ , so that we can simply set  $\beta_0 = 1$  in this case.

In order to ensure  $0 \leq p_i(t+1) \leq 1$ , we require  $\beta_k \geq 0$  for all  $k$ . Under this model, we can get the following properties.

**PROPERTY 2.** Let  $p_i^{(-k)}(t+1)$  be the activation probability of  $i$  at timestamp  $t+1$  without relation type  $k$ , under MLTM-R, we have

- $\min(p_i^{(k)}(t+1), p_i^{(-k)}(t+1)) \leq p_i(t+1) \leq \max(p_i^{(k)}(t+1), p_i^{(-k)}(t+1))$ ;
- when  $\beta_k \rightarrow 0$ ,  $p_i(t+1) \rightarrow p_i^{(-k)}(t+1)$ ; and
- when  $\beta_k \rightarrow \infty$ ,  $p_i(t+1) \rightarrow p_i^{(k)}(t+1)$ .

PROOF. Proof is omitted due to space limitation.  $\square$

From these properties, we can see that under this model when a new relation type is added, it might cause a decrease of the activation probability, if the activation probability for this single relation type is smaller than the current overall activation probability. We can also expect this phenomenon in some real scenarios. For example, in the iPhone social marketing case, a consumer would be distracted from purchasing an iPhone (becoming activated) by information, of much higher influence strength level, about Android phones from channels of the newly added relation types.

### 3. DIFFUSION MODEL LEARNING FROM ACTION LOG

In the last section, we have proposed two variations of the linear threshold model in multi-relational networks. A natural question then arises: for a concrete diffusion scenario, which model should we use? To answer this question, we have to learn the best parameters associated with each model, and test the fitness of the two models over real data. In this section, we address the issue about how to learn the parameters in the models using action logs of diffusion history.

#### 3.1 The Learning Framework

For each diffusion process, or *cascade*, an *action log* is a sequence of object set recording when an object is activated:  $\mathcal{A} = \{V_t\}_{t=1}^T$ . In our setting,  $t$ 's are collected from discrete timestamps.

The general learning framework is then to find the best parameters, i.e., the weight  $\beta_k$  for each relation type  $E_k$ , in the diffusion models that can maximize the likelihood of observing these actions recorded by the action logs. At a timestamp  $t$ , the activation probability  $p_u(t)$  of every inactive object  $u$ ,  $u \in R_t = V \setminus \bigcup_{t'=1}^{t-1} V_{t'}$ , follows the Bernoulli distribution, where  $R_t$  denotes the object set that has not been activated by  $t - 1$ .

In other words, at timestamp  $t$ , for any object  $u \in R_t$ , it would be activated with probability  $p_u(t)$ , and stay inactive with probability  $1 - p_u(t)$ . Therefore, the probability of observing the set of objects  $V_t$  activated at time  $t$ , and the set of objects  $R_t \setminus V_t$  not activated at time  $t$ , can be calculated as follows:

$$p(V_t, R_t \setminus V_t) = \prod_{u \in V_t} p_u(t) \prod_{u \in R_t \setminus V_t} (1 - p_u(t)) \quad (6)$$

The probability of observing the action log of a cascade is then:

$$L = \prod_{t=1}^T p(V_t, R_t \setminus V_t) \quad (7)$$

If multiple cascades are available, the likelihood is then the product of probabilities of each of these cascades:

$$L = \prod_{c \in \mathcal{C}} \prod_{t=1}^{T(c)} p(V_t^{(c)}, R_t^{(c)} \setminus V_t^{(c)}) \quad (8)$$

where  $c$  represents one cascade from the cascade set  $\mathcal{C}$ .

In practice, as there are many more inactive objects than active objects in the diffusion process, we sample the same number of inactive objects as active ones for any timestamp  $t$ , and treat each sampled inactive object with weight equal to the inverse of the sample rate  $\gamma_t$ , where  $\gamma_t = \frac{|V_t|}{|R_t \setminus V_t|}$ .

Now we introduce the learning algorithms for the two models. The goal is to find the best  $\beta$  that can maximize the likelihood, i.e., the MLE estimators, when plugging  $p_u(t)$  into the likelihood function. Without loss of generality, we only give the algorithms when one cascade is observed, as the extension to cases of multiple cascades being observed is straightforward.

#### 3.2 Learning Algorithm for MLTM-M

Under MLTM-M model, by plugging Eq.(4) into the likelihood function Eq.(7), we get the log-likelihood function as:

$$\begin{aligned} \log L_1 &= \sum_t \left\{ \sum_{u \in V_t} \log p_u(t) + \sum_{u \in R_t \setminus V_t} \log(1 - p_u(t)) \right\} \\ &= \sum_t \left\{ \sum_{u \in V_t} \log \frac{e^{\sum_k \beta_k p_u^{(k)}(t) + \beta_0}}{1 + e^{\sum_k \beta_k p_u^{(k)}(t) + \beta_0}} \right. \\ &\quad \left. + \sum_{u \in R_t \setminus V_t} \log \left(1 - \frac{e^{\sum_k \beta_k p_u^{(k)}(t) + \beta_0}}{1 + e^{\sum_k \beta_k p_u^{(k)}(t) + \beta_0}}\right) \right\} \end{aligned}$$

where  $p_u^{(k)}(t)$  is the single relation-based activation probability, and is calculated according to Eq.(3).

The problem now is reduced to a logistic regression problem, where each data point corresponds to an object at some timestamp, and the features are single relation-based activation probabilities,  $p_u^{(k)}(t)$ . The goal is to find the best weight  $\beta_k$  associated with each relationship set  $E_k$ .

We can use standard Newton-Raphson method to derive the best  $\beta$  by calculating the first and second derivative of  $\log L_1(\beta)$ . More specifically,  $\beta$  can be updated iteratively via:

$$\beta^{new} = \beta^{old} - \mathbf{H} \log L_1(\beta^{old})^{-1} \nabla \log L_1(\beta^{old}) \quad (9)$$

where  $\nabla \log L_1(\beta^{old})$  is the first derivative:

$$\frac{\partial \log L_1}{\partial \beta_k} = \sum_t \sum_{u \in R_t} p_u^{(k)}(t) (\mathbf{1}_{\{u \in V_t\}} - p_u(t; \beta^{old})) \quad (10)$$

and  $\mathbf{H} \log L_1(\beta^{old})$  is the Hessian matrix, i.e., the second derivative matrix:

$$\begin{aligned} \frac{\partial^2 \log L_1}{\partial \beta_k \partial \beta_l} &= - \sum_t \sum_{u \in R_t} \{p_u^{(k)}(t) p_u^{(l)}(t) p_u(t; \beta^{old}) \\ &\quad (\mathbf{1}_{\{u \in V_t\}} - p_u(t; \beta^{old}))\} \end{aligned} \quad (11)$$

#### 3.3 Learning Algorithm for MLTM-R

For MLTM-R model, by plugging Eq.(5) into the likelihood function Eq.(7), we can get the log-likelihood function as:

$$\begin{aligned} \log L_2 &= \sum_t \left\{ \sum_{u \in V_t} \log p_u(t) + \sum_{u \in R_t \setminus V_t} \log(1 - p_u(t)) \right\} \\ &= \sum_t \left\{ \sum_{u \in V_t} \log \frac{\sum_k \beta_k \sum_v w_{uv}^{(k)} \delta(v, t)}{\sum_k \beta_k \sum_v w_{uv}^{(k)}} \right. \\ &\quad \left. + \sum_{u \in R_t \setminus V_t} \log \frac{\sum_k \beta_k \sum_v w_{uv}^{(k)} (1 - \delta(v, t))}{\sum_k \beta_k \sum_v w_{uv}^{(k)}} \right\} \end{aligned} \quad (12)$$

Unlike MLTM-M, where  $\beta_k$  can be any real number, we have constraints  $\beta_k \geq 0$  for all  $k$  to guarantee that the activation probability lies in  $[0, 1]$ . We apply coordinate descent method to find the best non-negative  $\beta_k$ 's that maximizes that log-likelihood, by setting the learning step  $\eta$  smartly.

According to gradient descent method, a local maximum of  $\beta_k$  can be derived by iteratively updating the following formula:

$$\beta_k^{new} = \beta_k^{old} + \eta \frac{\partial \log L_2}{\partial \beta_k} \quad (13)$$

where  $\frac{\partial \log L_2}{\partial \beta_k}$  is the first derivative of function  $\log L_2$ :

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_k} &= \sum_t \sum_{u \in V_t} \frac{\sum_v w_{uv}^{(k)} \delta(v, t)}{\sum_{k'} \beta_{k'} (\sum_v w_{uv}^{(k')} \delta(v, t))} \\ &+ \sum_t \sum_{u \in R_t \setminus V_t} \frac{\sum_v w_{uv}^{(k)} (1 - \delta(j, t))}{\sum_{k'} \beta_{k'} (\sum_v w_{uv}^{(k')} (1 - \delta(j, t)))} \\ &- \sum_t \sum_u \frac{\sum_v w_{uv}^{(k)}}{\sum_{k'} \beta_{k'} \sum_v w_{uv}^{(k')}} \end{aligned} \quad (14)$$

In order to get non-negative updates for  $\beta_k$ 's, let  $\eta = \frac{\beta_k^{old}}{\sum_t \sum_u \frac{\sum_v w_{uv}^{(k)}}{\sum_{k'} \beta_{k'}^{old} \sum_v w_{uv}^{(k')}}}$ , then we have updating formula:

$$\begin{aligned} \beta_k^{new} &= \sum_t \left\{ \sum_{u \in V_t} \frac{\sum_v w_{uv}^{(k)} \delta(v, t)}{\sum_{k'} \beta_{k'}^{old} \sum_v w_{uv}^{(k')} \delta(v, t)} \right\} \beta_k^{old} \\ &+ \sum_{u \in R_t \setminus V_t} \frac{\sum_v w_{uv}^{(k)} (1 - \delta(v, t))}{\sum_{k'} \beta_{k'}^{old} \sum_v w_{uv}^{(k')} (1 - \delta(v, t))} \beta_k^{old} \\ &/ \left\{ \sum_t \sum_u \frac{\sum_v w_{uv}^{(k)}}{\sum_{k'} \beta_{k'}^{old} \sum_v w_{uv}^{(k')}} \right\} \end{aligned} \quad (15)$$

From the updating formula Eq.(15), we can see that if for a relation type  $k$ , most of the active (inactive) objects have a higher weighted percentage of active (inactive) neighbors defined by relation type  $k$ ,  $\beta_k^{new}$  will have a higher value than  $\beta_k^{old}$ , which is consistent with our intuition.

For the coordinate descent method, we update  $\beta$  on one dimension  $\beta_k$  each time, and then update  $\beta_{k+1}$  using the updated  $\beta_k$ , repeat this process cyclically until  $\log L_2$  converges.

By setting  $\beta_0 = 1$ , it can be shown that eigenvalues of the Hessian Matrix are all negative, i.e., the Hessian Matrix is negative-definite, and the objective function  $\log L_2(\beta)$  is concave. The learned  $\hat{\beta}$  that achieves a local maximum of log-likelihood function actually also achieves the global maximum, so that  $\hat{\beta}$  learned by the coordinate descent method is the global solution.

Note that if we sample the non-activated objects instead of using all of them, we need to consider the weights of the negative samples accordingly, i.e., the inverse of the sample rate  $\gamma_t$ .

### 3.4 Action Prediction

Once we have learned the parameters  $\hat{\beta}$  in the models, we can utilize these parameters to predict the future action of objects given an initial set of active objects.

These probabilities can be used to do (1) ranking: who are most likely to be activated at the next timestamp, and (2) prediction: predict the total number of activated objects at a future timestamp. Note that the total expected number of activated objects at timestamp  $t$  can be calculated as  $E(\sum_{u \in R_t} Y_u(t))$ , which is equal to  $\sum_{u \in R_t} EY_u(t)$ , assuming the independence of the activation behavior among inactive objects. For Bernoulli distribution,  $EY_u(t) = p_u(t)$ , therefore  $E(\sum_{u \in R_t} Y_u(t)) = \sum_{u \in R_t} p_u(t)$ .

## 4. EXPERIMENTS

In this section, we evaluate the performance of the proposed diffusion models in two multi-relational bibliographic information networks. We first list all the research questions that we intend to answer, followed by the experimental setup and performance analysis. Major research questions include:

- Q1. How much improvement can we achieve by modeling information diffusion in multi-relational networks, compared with single-relational networks?
- Q2. Is there heterogeneity among diffusion power of different relation types?

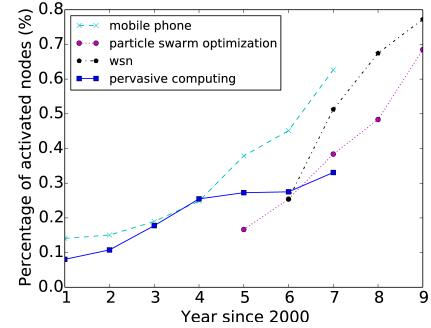


Figure 5: Trends of example topics during selected periods.

- Q3. Which one is a better choice for diffusion models in real networks, MLTM-M or MLTM-R?
- Q4. Does the diffusion power of each relation type remain the same at different stages of diffusion?

### 4.1 Datasets

We apply our models in two real-world multi-relational bibliographic information networks: DBLP (computer science domain) and APS (physics domain), which are constructed as following:

1. DBLP. Objects in the network are authors, and the relation types are constructed using meta-path-based techniques, including coauthor (APA), citation (APPA), sharing coauthors (APAPA), and co-attending conferences (APVPA).
2. APS. APS network has objects of authors, and relation types of coauthor (APA), citation (APPA), sharing coauthors (APAPA), and co-affiliation (AOA).

We study topic diffusion in the bibliographic networks, with topics extracted from papers' titles and abstracts using a NLP tool [21]. We extract 79 topics in DBLP dataset, and 30 topics in APS dataset, and we study diffusion of these topics during selected periods when these topics have increasing popularity trends. The trends of some example topics are shown in Figure 5. We also define one prediction period as prediction of one topic at a single timestamp. There are 520 prediction periods in the DBLP network, and 340 prediction periods in APS network.

The network statistics of DBLP<sup>1</sup> [32] and APS<sup>2</sup> networks are listed in Table 2. This DBLP dataset contains publication records from year 1937 to year 2009, including titles, authors, venues, and partial citations and abstracts information. APS contains journal publications from year 1893 to 2009, including titles, authors, authors' affiliations, and partial citations information.

	Authors	Papers	Venues	A-P	Citation
DBLP	916,988	1,572,278	6,713	4,135,188	2,083,947
	Authors	Papers	Org.	A-P	Citation
APS	323,675	463,348	41,411	2,471,474	4,710,547

Table 2: Statistics of real world two bibliographic networks.

### 4.2 Experimental Setup

For both datasets, we define each year as a timestamp. Given network structures  $G$  and the action log  $\mathcal{A}$  of all timestamps, we

<sup>1</sup>Download DBLP from <http://arne-miner.org/download>.

<sup>2</sup>Download APS from <https://publish.aps.org/datasets/inquiry>.

construct the training and testing data set for diffusion prediction of topic  $T_c$  at timestamp  $t + 1$  as follows:

- **Training dataset.** Label authors that are newly activated at  $t$  as positive, and the remaining inactive ones as negative. Sample the same number of negative authors as positive authors. We aggregate network features from  $(t - 3)$  to  $(t - 1)$  of these labeled authors for learning.
- **Testing dataset.** Label authors that are newly activated at  $t + 1$  as positive, and the remaining inactive ones as negative. The labeled authors compose the candidate set on which we perform prediction. Also, we aggregate the network features from  $(t - 2)$  to  $t$  of these labeled authors for prediction.

#### 4.2.1 Baselines

In order to verify that the proposed diffusion models in multi-relational networks outperform classical models in single-relational networks, we compare our method with the following baselines under different settings of single-relational networks:

1. Single Relational Network (Homo- $E_k$ ). Homo- $E_k$  is the single-relational network of relation type  $E_k$ .
  - LTM-Homo- $E_k$ . (Homo- $E_k$  for simplicity.) Apply classical LTM model to Homo- $E_k$  network. The influence strength level from node  $j$  to node  $i$  is  $w_{ij,t} = w_{ij,t}^{(k)}$ , and the activation probability for  $i$  at timestamp  $(t + 1)$  is  $p_i(t + 1) = p_i^{(k)}(t + 1) = \frac{\sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} \delta(j,t)}{\sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)}}$ .
  - MLTM-M-Homo- $E_k$ . (MH- $E_k$ ) Apply the proposed MLTM-M model to Homo- $E_k$  network. The two relation types to be aggregated at model level is  $E_k$  and global effect. Note that global effect is determined by  $\beta_0$ .
  - MLTM-R-Homo- $E_k$ . (MR- $E_k$ ) Apply the proposed MLTM-R model to Homo- $E_k$  network, by aggregating the relation type of  $E_k$  and global effect at relation level.
2. Projected Network (Project). Project is the network constructed by aggregating all relationships while ignoring relation type information, i.e., project the multi-relational network into a single-relational network with only “projected relation type”.
  - LTM-Project. Apply classical LTM model to Project network. The influence strength level from node  $i$  to node  $j$  is  $w_{ij,t} = \sum_k w_{ij,t}^{(k)}$ , leading to the activation probability of  $i$  at  $(t + 1)$ :  $p_i(t + 1) = \frac{\sum_{j \in \Gamma_t(i)} w_{ij,t} \delta(j,t)}{\sum_{j \in \Gamma_t(i)} w_{ij,t}}$ , where  $\Gamma_t(i)$  is  $i$ 's neighbor set on all relation types.
  - MLTM-M-Project. Apply the MLTM-M model to Project network, by aggregating the global effect and “projected relation type” at the model level.
  - MLTM-R-Project. Apply the MLTM-R model to Project network, by aggregating the global effect and “projected relation type” at the relation level.

#### 4.2.2 Evaluation Methods

We compare our models with baselines quantitatively from the following three different perspectives:

- **Local prediction** evaluates the ranking accuracy of who are most likely to be activated at the next timestamp, by Area Under Precision and Recall Curve (AUPR), where authors are ranked by predicted activation probability.

• **Perplexity/Likelihood** is widely used in the evaluation of language models and speech recognition [4], which is defined as  $\log L = \sum_{u \in V_{t+1}} \log p_u(t + 1) + \sum_{u \in R_{t+1} \setminus V_{t+1}} \log (1 - p_u(t + 1))$ . It is applied to validate predicted probability of each object to be activated at the next timestamp.

• **Global prediction** measures precision of predicted overall popularity. Predicted global popularity is defined as  $N_p = EY_i = \sum_i p_i(t + 1)$ . Define global prediction error rate as  $\varepsilon = \frac{1}{2} \cdot (\frac{N_p}{N_0} + \frac{N_0}{N_p}) - 1$ , where  $N_0$  is ground truth. The smaller  $\varepsilon$  is, the more accurate the global prediction is.

### 4.3 Model Effectiveness Comparison

In this section, we will verify the effectiveness of our model by answering the questions posed at the beginning of this section.

#### 4.3.1 Global prediction

Global prediction estimates how popular the specific topic would be at the next timestamp. As we can see from Table 3, both MLTM-M (MH-\*) and MLTM-R (RH-\*) models can give very accurate global prediction compared with LTM in single-relational network (Homo-\*). The reason why MLTM-M and MLTM-R can make better prediction is that the activation probability of each user is adjusted by global effect, while global effect plays an important role in the diffusion process [20]. Another observation is that MLTM-M models have better performance than MLTM-R models, though they are close.

Dataset	Model	APA	APPA	APAPA	APVAP	Full
DBLP	Homo-	0.654	0.287	1.005	1.269	N/A
	MH-	<b>0.033</b>	<b>0.04</b>	<b>0.034</b>	<b>0.041</b>	<b>0.031</b>
	RH-	0.072	0.07	0.092	0.128	0.125
Dataset	Model	APA	APPA	APAPA	APOAP	Full
APS	Homo-	0.249	0.398	0.107	0.144	N/A
	MH-	<b>0.045</b>	<b>0.052</b>	<b>0.039</b>	<b>0.068</b>	<b>0.052</b>
	RH-	0.073	0.082	0.076	0.079	0.11

**Table 3:** Global prediction of different network settings and models. The numbers are average values of global prediction error rates of all prediction periods under each setting.

#### 4.3.2 Single-Relational v.s. Multi-relational networks

To answer Question 1, we conduct experiments in both multi-relational networks and single-relational networks, as shown in Figure 6. By modeling diffusion in multi-relational network with all relation types included (Full), we can make better prediction than models in single-relational networks which contain only one relation type, e.g., APA, APPA, APAPA. The experimental results suggest that diffusion models in multi-relational networks model the real world diffusion processes better than models in single-relational networks. In single-relational networks, only influence from one relation type is considered, while influence from links of the remaining relation type can not be captured in the models, resulting in great information loss.

To answer Question 2, we compare experimental results on two networks, one that ignores the diffusion power of different relation types (Project), and another one that learns the diffusion power of all the relation types (Full). As shown in Figure 6, models in Full setting makes better prediction than Project setting, confirming that different relation types have different diffusion power, and

the difference should not be ignored. When projecting the multi-relational network into single-relational network, we drop the semantic information contained in the network topology, leading to great information loss and noisy models.

Therefore, it has theoretical value and practical significance to model information diffusion processes in multi-relational networks, while weighting the diffusion power of different relation types differently based on real action logs.

#### 4.3.3 MLTM-M v.s. MLTM-R

Another important target of our work is to answer Question 3, i.e., how objects are activated when there is influence from different types of relationships simultaneously. We answer the question by comparing the prediction results of MLTM-M and MLTM-R in DBLP and APS real networks. As shown in Figure 6, MLTM-R outperforms MLTM-M consistently. Regarding prediction in real world networks, MLTM-R seems to be a better choice. A possible explanation is as follows: under MLTM-R assumption, when an author makes decisions on topics adoption, she will consider the weighted influence sum from links of all relation types; while under MLTM-M assumption, the author first make topic adoption decision on each single relation type, and then weight all these decisions to make a final one; however, in reality, MLTM-R is a simpler and more intuitive way for people to think and make decisions, leading to its better performance.

### 4.4 Difference in Diffusion Power

With regards to Question 4, we use the submodule method to estimate the significance of each relation type. We use  $M$  to denote the set of all relation types,  $M = \{E_i | i = 1, 2, \dots, n\}$ , where  $n$  is the total number of relation types. Remark that global effect is included in  $M$ . To study the significance of relation type  $E_k$ , we construct a submodule of  $M$  as  $N_k = \{E_i | i \neq k\}$ ,  $N_k \subsetneq M$ . The significance of relation type  $E_k$  can defined as

$$S_{E_k} = \frac{\log L_{N_k}}{\log L_M}, \quad (16)$$

where  $\log L_{N_k}$  ( $\log L_M$ ) is the log likelihood of MLTM-R model applied to a multi-relational network containing relation type set of  $N_k$  ( $M$ ). Note that, in order to apply the submodule method, any relation type pairs in  $M$  should not be strongly correlated. Based on this observation, we first perform relation types selection to remove strongly correlated relation type pairs, which are identified by Pearson correlation scores.

#### 4.4.1 Relation Type Selection

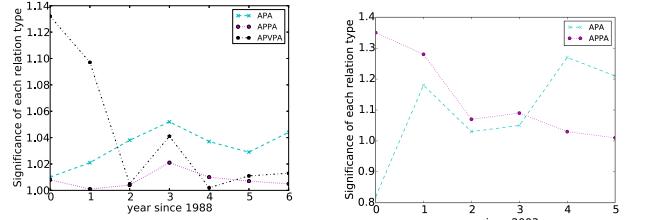
Pearson correlation scores of strongly correlated relation type pairs are listed in Table 4. For each pair of strongly correlated relation type pairs, we remove the one with less prediction power. As the removed relation types are strongly correlated with the remaining relation types, the prediction results change little after performing relation type selection, as shown in Table 5.

Dataset	relationship 1	relationship 2	correlation coefficient
DBLP	APA	APAPA	0.808
APS	APA	APAPA	0.66
	APA	APOPA	0.523

**Table 4:** Pearson correlation scores of the strongly correlated relationship type pairs.

Dataset	Methods	AUPR	logL
DBLP	MLTM-R(APA, APPA, APVPA)	<b>0.133</b>	-2013.8
	MLTM-R(Full)	0.131	-2015.3
APS	MLTM-R(APA, APPA)	<b>0.158</b>	-1385.8
	MLTM-R(Full)	<b>0.158</b>	-1391.8

**Table 5:** Prediction results after relation type selection.



(a) Diffusion of "support vector machine" in DBLP bibliographic network  
(b) Diffusion of "ultracold atom" in APS bibliographic network

**Figure 7:** The relative importance of each relationships of topic diffusion in bibliographic networks change at different stages

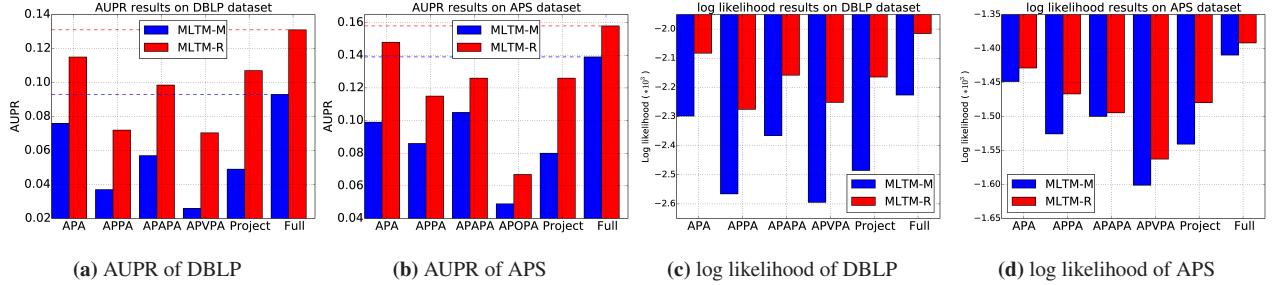
#### 4.4.2 Significance of Each Relation Type

We use significance value, defined in Eq.(16), as proxy variable for diffusion power. To study the diffusion power of each relation type in concrete diffusion cases, we choose one topic in each network, and study the significance of each relation type at different stages of the diffusion process. Note that the current relation type set is obtained by performing relation type selection on the original relation type set, i.e., there are no two relation types that are strongly correlated.

From the DBLP dataset, we select the topic of "Support Vector Machine" for a case study. The significance of each relation type at different timestamps is shown in Figure 7a. In the beginning, when the topic SVM starts to appear, the co-attending conference relation type (APVPA) is most important, as people are likely to learn this new topic from conferences. In the later stage, the coauthor relationship is becoming more and more important, compared with the other relation types. This can be possibly explained by the fact that the study on SVM is becoming more narrow and diverse, so that closer relation types, such as APA, have stronger diffusion power to propagate this topic around. Also, along the co-attending conference relation type, authors would pay more attention to other newly emerged topics, leading to its weak diffusion power for "Support Vector Machine", a relatively old topic.

From the APS network, we select the topic "ultracold atom" for a case study. As APS is a journal dataset, in the early stage, authors are more likely to learn this new topic by reading, and cite the papers they read, leading to strong diffusion power (significance) of the citation relationship, as shown in Figure 7b. In the following stages, coauthor relation type plays a more important role in propagating the topic around, as this topic is becoming more narrow and diverse. Another observation is that the same relation type of citation (APPA) plays quite different roles in the two bibliographic networks, which can be explained by the difference in the network topology of the two datasets.

To sum up, we find that the same relation type plays different roles at different stages of the diffusion process and in different networks, while the significance / diffusion power of relation types can be explained by the properties of topics being propagated and the typologies of the multi-relational networks.



**Figure 6:** Experimental results on DBLP and APS dataset. The x-axis represents the diffusion relation set included in each setting. “Full” standing for the multi-relational networks with all relation types included, while others stand for single-relational networks of specified relation types.

## 5. RELATED WORK

The study of disease propagation in contact networks, which is analogous to the diffusion of news and ideas in formation networks, has long been a base for information diffusion study. In [12], the authors mathematically model the spread of infectious diseases. In [6] and [7], diffusion models are applied to the domain of marketing, aiming to improve viral marketing techniques by better understanding of how information spreads among potential customers. Leskovec *et al.* present how recommendations of products propagate among individuals [16], and mathematically characterize the resulting recommendation networks. Topic diffusion is a special case for diffusion studies. in [10], Gruhl *et al.* study the propagation of topics in the blogspace. By decomposing the topics along two orthogonal axes (chatter and spikes), they develop a model to capture the structures of topics. In [11], the authors make prediction about future trends of events on Twitter based on different characteristics of historical events, such as popularity, retweets count, followers’ size, and relevant URLs outside of Twitter.

In previous work on diffusion models, the probability of activation is usually randomly generated in synthetic data [14, 6]. In contrast, we propose to learn the model parameters based on real action logs, providing prediction power for future diffusion processes. Myers *et al.* proposed to learn the activation probability from two aspects, network influence and global effect [20], which focuses on identifying the influence from outside; however, our work tries to identify different roles of each relation type within the network, while learn the external influence automatically. [26] studies how likely information would propagate from a specified sender to a receiver during a given time period, while our model measures information diffusion from a more macro level: how likely information would propagate along each type of relationships.

Another line of related work lies in the topic of “Multilayer Networks” [19], which are also known as interdependent networks [3], multilevel network [3]. However, along this line of work, there are multiple networks involved. In comparison, we focus on one network, distinguishing between different types of relationships, which makes our work unique. In [2], theoretical analysis of network vulnerability to global cascades under multiplex and simplex network settings is given. However, in this piece of work, the analysis of LTM is based on the assumption that a node activates if a sufficiently large fraction of neighbors in any layer is active, which ignores the different diffusion power along different types of relationships. [33] suffers from this shortcoming. On the other hand, unlike our models which can be directly applied in various applications, these models have no diffusion prediction power in real world scenarios.

Recently, multi-relational networks became an increasingly important topic and there are several works on this new type of network. It has been shown that different types of relationships among the same set of objects could lead to rather different topological properties [29, 13, 30, 31]. [30] shows that the in the online game network with both positive and negative relation types, structures and properties of networks with only positive interactions are quite different from networks with only negative relationships. In [29], Sun *et al.* propose to do clustering based on weighted combination of different types of relationships, while the weights are learned from users’ guidance. Also, based on the intuition that highly ranked items are more likely to be connected, Ji *et al.* propose Rankclass algorithm to combine ranking and classification [13]. In [13], Ji *et al.* assign different weights to different types of links when doing feature selection. There are also some work on analyzing and identifying different types of relationship in networks. In our model, we build relation types from meta-path-based techniques [28], while [31] develops a framework to identify different types of relationships by learning across multiple heterogeneous information networks.

## 6. CONCLUSION

In this work, we proposed the problem of studying information diffusion in multi-relational networks. We found that, in multi-relational networks, information can propagate along every type of relationships, and each relation type plays a different role. We made two assumptions about how different relation types work together to propagate information, based on which we proposed two variations of the linear threshold model to model information diffusion in multi-relational networks. Moreover, we presented corresponding learning algorithms to learn the weight of each relation type from real action logs. In view of real-world experimental results, we discovered that aggregation of influence at the relation level is overall better than at the model level. We also studied the significance of each relation type at different stages of the diffusion processes. Moreover, the two assumptions we made are quite general and can be applied to extend many other classical diffusion models. We believe that our work will not only advance the research on information diffusion in multi-relational networks, but also benefit many real-world applications.

## ACKNOWLEDGEMENTS

Research was sponsored in part by the Army Research Lab. under Cooperative Agreement No. W911NF-09-2-0053 (NSCTA), the Army Research Office under Cooperative Agreement No. W911NF-13-1-0193, National Science Foundation IIS-1017362,

IIS-1320617, and IIS-1354329, HDTRA1-10-1-0120, and MIAS, a DHS-IDS Center for Multimodal Information Access, Synthesis at UIUC, ACE Award and NEU TIER 1 Grant.

## References

- [1] R. M. Anderson and R. M. May. *Infectious Diseases of Humans Dynamics and Control*. Oxford University Press, 1992.
- [2] C. D. Brummitt, K.-M. Lee, and K.-I. Goh. Multiplexity-facilitated cascades in networks. *Physical Review E*, 85(4):045102, 2012.
- [3] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.
- [4] S. F. Chen, D. Beeferman, and R. Rosenfield. Evaluation metrics for language models. 1998.
- [5] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, 357(4):370–379, July 2007. Access to full text is subject to the publisher’s access restrictions.
- [6] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.
- [7] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 2001(9), 2001.
- [8] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’10)*, pages 1019–1028, 2010.
- [9] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [10] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogsphere. In *Proceedings of the 13th international conference on World Wide Web, WWW ’04*, pages 491–501. ACM, 2004.
- [11] M. Gupta, J. Gao, C. Zhai, and J. Han. Predicting future popularity trend of events in microblogging platforms.
- [12] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Rev.*, 42(4):599–653, Dec. 2000.
- [13] M. Ji, J. Han, and M. Danilevsky. Ranking-based classification of heterogeneous information networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1298–1306. ACM, 2011.
- [14] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 2003 ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD’03)*, pages 137–146, Washington, DC, Aug 2003.
- [15] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Automata, Languages and Programming*. Springer Berlin Heidelberg, 2005.
- [16] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), May 2007.
- [17] H. Ma, H. Yang, M. R. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 233–242. ACM, 2008.
- [18] M. W. Macy. Chains of cooperation: Threshold effects in collective action. *American Sociological Review*, 56(6):730–747, Dec. 1991.
- [19] M. Magnani and L. Rossi. The ml-model for multi-layer social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 5–12. IEEE, 2011.
- [20] S. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 33–41. ACM, 2012.
- [21] L.-A. Ratinov, D. Roth, D. Downey, and M. Anderson. Local and global algorithms for disambiguation to wikipedia. In *ACL*, volume 11, pages 1375–1384, 2011.
- [22] M. A. Rodriguez and J. Shinavier. Exposing multi-relational networks to single-relational network analysis algorithms. *Journal of Informetrics*, 4(1):29–41, 2010.
- [23] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704. ACM, 2011.
- [24] T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.
- [25] G. Serazzi and S. Zanero. Computer virus propagation models. In *In Tutorials of the 11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS’03)*. Springer-Verlag, 2003.
- [26] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proceedings of the 16th international conference on World Wide Web*, pages 191–200. ACM, 2007.
- [27] V. Ströele, J. M. Souza, et al. Group and link analysis of multi-relational scientific social networks. *Journal of Systems and Software*, 2013.
- [28] Y. Sun, J. Han, X. Yan, P. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB’11*, 2011.
- [29] Y. Sun, B. Norick, J. Han, X. Yan, P. S. Yu, and X. Yu. Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1348–1356. ACM, 2012.
- [30] M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.
- [31] J. Tang, T. Lou, and J. Kleinberg. Inferring social ties across heterogenous networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 743–752. ACM, 2012.
- [32] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998. ACM, 2008.
- [33] O. Yağan and V. Gligor. Analysis of complex contagions in random multiplex networks. *Physical Review E*, 86(3):036103, 2012.