

# Structural Analysis in Multi-Relational Social Networks

Bing Tian Dai \*      Freddy Chong Tat Chua \*      Ee-Peng Lim \*

## Abstract

Modern social networks often consist of multiple relations among individuals. Understanding the structure of such multi-relational network is essential. In sociology, one way of structural analysis is to identify different positions and roles using blockmodels. In this paper, we generalize stochastic blockmodels to *Generalized Stochastic Blockmodels* (GSBM) for performing positional and role analysis on multi-relational networks. Our GSBM generalizes many different kinds of *Multivariate Probability Distribution Function* (MVPDF) to model different kinds of multi-relational networks. In particular, we propose to use *Multivariate Poisson distribution* for multi-relational social networks. Our experiments show that GSBM is able to identify the structures for both synthetic and real world network data. These structures can further be used for predicting relationships between individuals.

## 1 Introduction

There are often different kinds of interactions co-existing with one another in social networks. For example in **Facebook**<sup>1</sup>, one can write on others' wall, poke her friends, or tag her friends in her own photos. When single relational network model is used to represent such a social network, the different interaction types are treated the same making them indistinguishable. As information is lost in the network representation, we may not be able to model the network accurately.

A multi-relational network, on the other hand, allows multiple relations to exist in the network and even exist between two individuals. A *relation* represents a social connection or a set of interactions of the same kind between two individuals. For example, Figure 1 indicates 5 different relations from Alice to Bob on **Facebook**. The top two are social connections, i.e., Alice is a classmate of Bob, and also a friend of Bob. There are also 5 interactions between Alice and Bob, as shown in the lower part of Figure 1. For instance, Alice poked Bob on Monday and Wednesday, which contributes two

interactions of the same kind. Thus the 5 interactions describe three different relations, i.e, **write-on-wall**, **poke** and **tag**. In other words, each interaction is understood as a relation instance and multiple interactions of the same kind between a pair of individuals represent instances of the same relation.

Each relation defines a single relational network. As shown in Figure 2, the multi-relational network in **Facebook** consists of multiple single relational networks, shown in different planes. The single relational networks correspond to relations **write-on-wall**, **poke** and **tag** respectively. Since Alice wrote on Bob's wall, there is a directed interaction edge from Alice to Bob in the **write-on-wall** network.

We can then define a *multi-relational network* as a merger of multiple single relational networks. Figure 2 represents a multi-relational network constructed by combining the three single relational networks together. Dotted lines in Figure 2 indicate vertices representing the same individuals, which are merged into one vertex in the multi-relational network. An edge between two nodes in a multi-relational network is also called a *relationship*, which consists of all relations and interactions between the two individuals.

Various structural analyses have been studied on single relational network. For example, given a single relational network constructed by relation **is-a-friend-of**, plenty of clustering methods group the individuals based on how densely they are connected, i.e., individuals within the same group are more likely to be friends, while individuals across different groups are less likely. The grouping of the individuals provides a structural summary for the entire network and also determines the communities formed. The community affiliations of the individuals help in learning their behavior patterns, e.g. one would like to buy items which many others in her community have bought. Discovery of structural information is thus essential to better understanding of all kinds of social networks, including multi-relational networks.

However, multi-relational networks are much more difficult to analyze than single relational networks. In a single relational network, grouping of nodes are identified by the density of the connections among them. In a multi-relational networks, such groupings can be

\*Living Analytics Research Centre (LARC), School of Information Systems, Singapore Management University

<sup>1</sup><http://www.facebook.com>

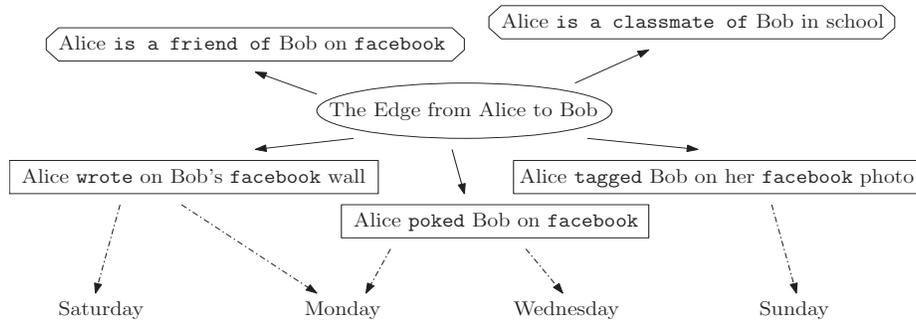


Figure 1: Representing the edge from Alice to Bob by 5 different relations

identified by the different types of the connections, or rather combinations of relations. For example, one may find a group consisting of classmates who attended some course together, whereas another group consisting of colleagues who worked together before. The other key difference between groupings in single relational networks and multi-relational networks is that, individuals across different groups are assumed to be loosely connected in single relational networks, but in a multi-relational networks, the edges across different groups may be as dense as edges within the groups, and these cross-group and within-group edges are of different relations. One example is the antagonistic communities. Within the communities, the edges are positive as opinions are consistent. However, cross-community edges are negative since opinions from one community are against to the opinions from the other community. Therefore, relations play a big role in the discovery of structures in multi-relational networks.

In a multi-relational network, each relation suggests different association semantics between two individuals. In the Facebook example, relation *poke* suggests an informal association which usually happens between close friends; while relation *write-on-wall* suggests a formal or neutral association. Relation *tag* even suggests two people have been physically together for an outing. Such heterogeneity in relation semantics has made analysis of structures in multi-relational networks more challenging than single relational networks, simply because relations should not be treated in the same way.

Although interactions in a multi-relational network are more heterogeneous, they provide additional semantics that helps to identify the social positions (social statuses) of participating parties [8]. This is due to the fact that, the interactions occurred between two people largely depend on their *positions* in the social network and the *roles* they play. The position (or social position) refers to a collection of individuals sharing similarities in their relationships. The role (or social role) refers to the relationship between individuals or between positions.

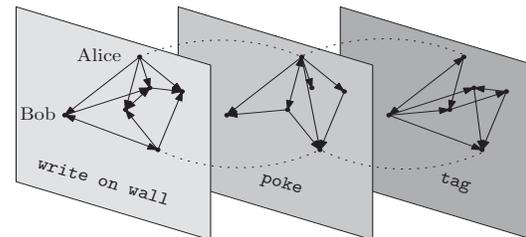


Figure 2: Combining three networks of different relations into one multi-relational network

Each role represents some pattern of relations (interactions) between two individuals in the corresponding positions. We give the following example to illustrate.

**EXAMPLE 1.** Consider the multi-relational network in Figure 2. Alice works in marketing, and uses of social networks to promote her products by posting product recommendations on other people's walls. Bob's hobby is photography, and he tags his friends appearing in his photos. This explains why Alice performs a lot of *write on wall* interactions, whereas Bob performs a lot of *tag* interactions. We say that Alice and Bob have marketer and photographer positions respectively in the network due to their distinct relationship compositions with others. Furthermore, the interactions between Alice, Bob and others are grouped into roles.

Our *structural analysis problem* on a multi-relational social network is thus to discover groupings of individuals and relationships corresponding to positions and roles respectively [21]. In other words, individuals in a multi-relational network should be grouped such that, their relationships are consistent at the group level, i.e. the relationships are grouped by the roles they play. Given two pairs of individuals belonging to the same pair of groups, their relationships are expected to be similar. The relationships between individuals are then described by the relationships between their groups. Ideally, with the rich information enclosed in

a multi-relational network, structural analysis is able to group individuals by their social positions, and the relationships are then determined by the social roles between such social positions. As in the above example, the relationship from Alice to Bob is determined by the role from the position of marketer to the position of photographer. Such positions can be explicit or implicit. For instance, in Example 1, one's occupation is an explicit position while one's hobbies determine an implicit position.

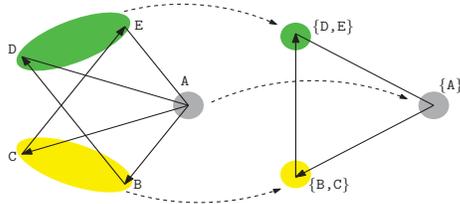


Figure 3: By clipping vertices B, C together, vertices D, E together, the left figure is reduced to its core structure on the right, which preserves all edges

One direct application of our structural analysis is to provide a structural view of the entire social network, which describes how these positions are related to one another. For example, in Figure 3, by merging B and C to  $\{B,C\}$ , and D and E to  $\{D,E\}$ , we have a 3-node structural view over the 5-node graph ABCDE, where all directed and undirected edges among ABCDE are preserved by the edges between the groups of the vertices. We often call such a reduced graph a *core structure* of the original graph. We can apply our structural analysis to identify the *core structure* where each node in the core structure represents a particular position and each edge represents the role from one position to the other. Core structure provides a good visualization of the entire network, as well as a tool to interpret relationships between individuals.

Other applications of structural analysis include the use of core structures to predict relationships. For example, we can predict a negative link exists between two individuals who join two different groups having an antagonistic relationship. It is also interesting to study key changes in the graph that cause changes to the core structure. With the core structure, we can generate synthetic graphs that up-size or downsize the original graph easily for performance study, or anonymize data through structural similarity [11].

Although structural analysis is widely applicable, it is however challenging due to the following three reasons. First, individuals in social networks do not always have a definite position. For instance, one can be a sales executive at work, a part-time student at school,

and a husband or even a father at home. Thus, it is not reasonable to restrict one to have a single position in the core structure. Instead, an individual should have a mixture of the positions. This general challenge of mixed membership on single relational networks has been solved in [1], and we will extend the solution further for multi-relational networks.

The second challenge is brought by the multi-relational aspect. Single relational networks are mostly modeled as a binary graph where an edge either exists or not, or a weighted graph where an edge is associated with a value. For multi-relational networks, the challenge is therefore to model multiple relations simultaneously as a relationship between nodes may be characterized by a combination of relations. For instance, relation *is-a-friend-of* is a binary relation, the relation *tag* on Facebook is a weighted relation where the number of times one tags another is a signal on how strong the relation is between them. When these two relations co-exist in the multi-relational network, our model should be able to model them together, not separately.

The third challenge is that, relations are often correlated, especially in social networks. Modeling of a multi-relational network by directly combining single relational networks without considering the correlation among relations will likely compromise the accuracy of modeling network positions and roles. For example, relation *dine-together* and relation *drink-together* are very likely to be correlated between two friends who like to hang out together. Therefore, our model ought to handle the challenging task of modeling the correlations among relations.

The contributions of this paper are summarized as follows.

- We propose *Generalized Stochastic Blockmodels (GSBM)* as a novel framework to perform structural analysis on multi-relational networks. The framework is designed to model multiple relations corresponding to different types of networks and to capture the correlations among relations. The framework can also accommodate different *multivariate probability distribution functions (MVPDF)* for generating interactions of different relations between a pair of individuals based on their positions.
- *Multivariate Poisson Distribution (MVPois)* is introduced as a MVPDF in the GSBM framework so as to derive network structures from interactions in social networks, since MVPois incorporates the correlations between relations. We estimate the parameters of the GSBM with MVPois using Maximum Likelihood Estimation (MLE) for learning the model parameters.

- Finally, we demonstrate that our blockmodel can discover the ground truth grouping in synthetic network and predict relationships in real (IMDb) networks. We also show the results of structural analysis using a case study.

The remainder of the paper is organized as follows. We will first give a review on blockmodels in Section 2. The framework of our GSBM is then introduced in Section 3.1, which is followed by the discussion about connections between our GSBM and other various blockmodels in Section 3.2. A basic MVPDF, multiple Poisson (mPois) distribution, is examined first with GSBM to model multi-relational social networks in Section 4.1. It is then extended to multivariate Poisson (MVPois) distribution to capture correlations among relations in Section 4.2. Together with this particular modeling of social networks, we then apply MLE and EM algorithm to build a blockmodel from a given multi-relational social network in Section 5. Finally, we experimentally validate our GSBM in Section 6.

## 2 Related Works

Positional and role analysis receives a lot of attention from sociology researchers. White, Boorman and Breiger [22] introduced the concept of *blockmodel* to study roles and positions as parts of social structure. They also developed blockmodels for networks with multiple ties. Wasserman and Faust [21] defined a blockmodel as a partition of individuals into  $k$  positions, and the blockmodel also specifies  $m$  different ties between and within the positions. A blockmodel is represented by a tensor  $\mathbf{B} = \{b_{ruv}\}$ ,  $m \times k \times k$ , with each entry  $b_{ruv}$  (called a *role block*) indicating the presence or absence of tie  $r$  from position  $u$  to position  $v$ .  $b_{ruv} = 1$  represents a *oneblock*, meaning there exists tie  $r$  from each individual in position  $u$  to everyone in position  $v$ .  $b_{ruv} = 0$  indicates a *zeroblock*, meaning there is no connection between individuals in positions  $u$  and  $v$ .

Doreian, Batagelj and Ferligoj [7] generalized the blockmodel in [22]. Instead of allowing only oneblocks and zeroblocks, they also considered modeling networks by other types of role blocks, e.g., dominant blocks where at least one individual from one position is connected to everyone in the other position, and regular blocks where everyone in one position is connected to at least one individual from the other position. Žiberna [19] further generalized this idea to valued networks, which aims to minimize the inconsistency among edges grouped into each role block.

Another direction to generalize the original blockmodels is to make blockmodels stochastic, i.e., associate one or more probability distributions to each block as studied in [12, 20]. However, most stochastic block-

models assign every individual to one position only, except Airoldi, Blei, Fienberg and Xing’s [1] mixed membership blockmodels (MMB) on single relational binary networks. MMB describes (i) each individual by a mixed membership probabilistic distribution indicating her different position with probabilities; and (ii) a stochastic blockmodel by a probability matrix where each probability induces a Bernoulli process on the existence of the single relation for its corresponding role block. In MMB, the existence of the edge from one individual to another is then modeled as a mixture on all Bernoulli processes via position variables following both their membership distributions.

However, the basic MMB only considers binary graphs and does not capture the “relationship strength” in social networks [23]. Gallagher [10] further extended the stochastic blockmodel from binary graphs to weighted networks. In his model, the weight of an edge is the number of observations of such edge, e.g., the number of messages. Thus instead of modeling each role block using a Bernoulli distribution, it is modeled as a Poisson distribution. We shall see that the two aforementioned blockmodels are special cases of our generalized framework in Section 3.1.

Multi-relational networks have attracted much attention in recent years [5, 16, 3]. Many works focus on finding communities in multi-relational networks [6, 18, 2, 17]. However, these works revolve around modeling the relations within the communities and neglect the relations between different communities. Although blockmodels on multi-relational networks have been discussed in [22, 12, 9], mixed membership stochastic blockmodel has not been studied on multi-relational networks. In this paper, our main focus is to generalize stochastic blockmodels to multi-relational networks.

## 3 Generalized Stochastic Blockmodel

### 3.1 The Framework

Relationships in a multi-relational network are vectors of occurrences of a set of ordered relations. Based on the stochastic blockmodels [1] mentioned in Section 2, we propose the *generalized stochastic blockmodels* (GSBM):  $\theta \triangleq (\mathbf{\Pi}, \mathbf{B})$  on a multi-relational network  $G = (X, Y)$ .

We first define the following notations. **1.**  $X \triangleq \{x_i: i = 1, 2, \dots, n\}$ : the set of vertices, where  $n$  is the total number of vertices. **2.**  $Y \triangleq \{\vec{Y}_{ij}: i \neq j\}$ : the observed relationships; each  $\vec{Y}_{ij}$  is a  $m$ -dimensional vector representing the relationship from  $x_i$  to  $x_j$ , where  $m$  is the number of relations in the multi-relational network. The  $r^{th}$  element of  $\vec{Y}_{ij}$ ,  $Y_{ijr}$ , denotes

the interaction count of the  $r^{th}$  relation from  $x_i$  to  $x_j$ . For example, in Figure 1, Alice wrote on Bob's wall twice, poked him twice and tagged him once, so the edge from Alice to Bob,  $\vec{Y}_{\text{Alice,Bob}}$ , is (2, 2, 1), in the order of **write-on-wall** (W), **poke** (P) and **tag** (T). **3.**  $\mathbf{\Pi} \triangleq \{\vec{\pi}_i: i = 1, 2, \dots, n\}$ :  $\vec{\pi}_i$  is a  $k$ -dimensional vector, representing the mixed membership probabilistic distribution of  $x_i$ , where  $k$  is the number of positions.  $\pi_{iu}$  is the  $u^{th}$  element of  $\vec{\pi}_i$ , the probability of  $x_i$  belonging to the  $u^{th}$  position.

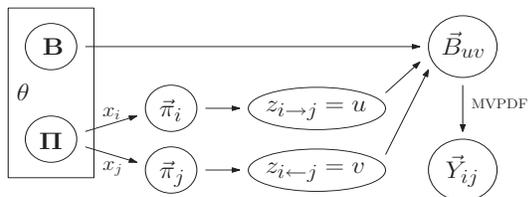


Figure 4: An Illustration of GSBM parameters

The model generates each relationship  $\vec{Y}_{ij}: x_i \rightarrow x_j$  as shown in Figure 4. **1.** We define a pair of random latent variables  $z_{i \rightarrow j}$  and  $z_{i \leftarrow j}$  which take values in  $\{1, 2, \dots, k\}$  and follow multinomial distributions with parameters  $\vec{\pi}_i$  and  $\vec{\pi}_j$  respectively. That is,  $z_{i \rightarrow j} \sim \text{mNom}(\vec{\pi}_i)$  and  $z_{i \leftarrow j} \sim \text{mNom}(\vec{\pi}_j)$ , where  $\text{mNom}$  stands for a multinomial distribution. **2.**  $\mathbf{B} \triangleq \{\vec{B}_{uv}: u, v = 1, 2, \dots, k\}$ :  $\mathbf{B}$  is a  $s \times k \times k$  dimensional tensor (see Figure 5).  $\vec{B}_{uv}$  (each gray pillar) is a  $s$  dimensional vector of parameters, describing a multivariate probability density function  $\text{MVPDF}_{s \rightarrow m}(\vec{B}_{uv})$  which takes the  $s$  parameters and generates  $m$ -dimensional relationships. **3.** Suppose  $z_{i \rightarrow j} = u$ ,  $z_{i \leftarrow j} = v$ , we determine the  $(u, v)$ -column of tensor  $\mathbf{B}$ ,  $\vec{B}_{uv}$  (the shaded column in Figure 5), and generate  $\vec{Y}_{ij}$  using the corresponding MVPDF, i.e.,  $\vec{Y}_{ij} \sim \text{MVPDF}_{s \rightarrow m}(\vec{B}_{uv})$ .

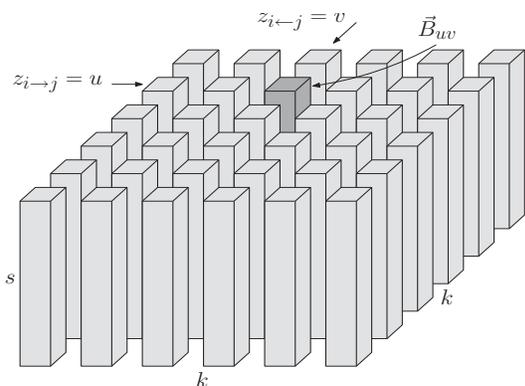


Figure 5: An Illustration of blockmodels  $\mathbf{B}$ , which is a tensor of dimension  $s \times k \times k$ ; when  $z_{i \rightarrow j} = u$  and  $z_{i \leftarrow j} = v$ , column  $\vec{B}_{uv}$  estimates  $\vec{Y}_{ij}$

The relationship  $\vec{Y}_{ij}$  is estimated by  $\vec{B}_{uv}$  with probability

$$\begin{aligned} \Pr(\vec{Y}_{ij}, z_{i \rightarrow j}, z_{i \leftarrow j} | \theta) &= \Pr(u | \vec{\pi}_i) \Pr(v | \vec{\pi}_j) \Pr(\vec{Y}_{ij} | \vec{B}_{uv}) \\ (3.1) \qquad \qquad \qquad &= \pi_{iu} \pi_{jv} \text{MVPDF}(\vec{Y}_{ij} | \vec{B}_{uv}) \end{aligned}$$

The probability of  $\vec{Y}_{ij}$  is thus

$$\Pr(\vec{Y}_{ij} | \theta) = \sum_{u,v} \pi_{iu} \pi_{jv} \text{MVPDF}(\vec{Y}_{ij} | \vec{B}_{uv})$$

Therefore, relationship  $\vec{Y}_{ij}$  is modeled as a mixture of all columns  $\vec{B}_{uv}$  via  $z_{i \rightarrow j} = u$  and  $z_{i \leftarrow j} = v$ .

### 3.2 GSBM Variants

The multi-relational networks studied in [22, 12] directly combine multiple single relational networks together, where each block is modeled by one parameter on each relation. As relations are modeled independently of one another, the MVPDF in this context is specifically a joint probability distribution over all relations, i.e., the probability of each relationship is the product of the probabilities of the values on all relations, each determined by the parameter on the corresponding relation. Hence, our GSBM is also considered as a generalization of the blockmodels on multi-relational networks, by incorporating more general MVPDF to cater for the correlations among relations. This will be elaborated in Section 4.

As the MMB introduced in [1] models each block by a Bernoulli process on a single relation, our GSBM reduces to MMB with  $m = 1$  and the MVPDF being a Bernoulli distribution on that relation.

Another example of MVPDF is a multinomial distribution,  $\text{MVPDF}_{m+1 \rightarrow m}$  takes an integer  $w$  and  $m$  probabilities  $p_1, p_2, \dots, p_m$  with sum 1, and the probability of a relationship  $(w_1, w_2, \dots, w_m)$  with sum  $w$  is  $\frac{n!}{w_1! w_2! \dots w_m!} p_1^{w_1} p_2^{w_2} \dots p_m^{w_m}$ .

Our GSBM does provide solutions to address the second challenge mentioned in Section 1, by adopting a MVPDF which is able to handle multiple types of relations in a multi-relational network, including binary relations from social connections and weighted relations from social interactions. For an example, suppose the relation **is-a-friend-of** and the relation **tag** are modeled by Bernoulli distributions and multinomial distributions respectively. If these two relations are independent, the MVPDF for each block can be taken as the product of the two corresponding distributions. If there exists dependency, the MVPDF for each block can be a combination of two multinomial distributions for **tag**, depending on relation **is-a-friend-of** is present or not.

In general, MVPDF can be any probability distribution whose outcome is a  $m$ -dimensional count vector. We can relax even further to allow each different  $\vec{B}_{uv}$  to have a different MVPDF. We shall leave this for future research. In the following sections, we will discuss further on MVPDFs for general social networks, to tackle the third challenge in Section 1.

## 4 GSBM with Poisson Distributions

### 4.1 mPois: Basic MVPDF

Although our GSBM is general enough to accommodate many different kinds of MVPDF, we still have to select an appropriate MVPDF for modeling relationships in social networks. For example, the multinomial distribution above is not appropriate for multi-relational networks as it requires all relationships modeled by the same column  $\vec{B}_{uv}$  to have the same number of total occurrences (or interaction count).

As a stochastic process representing the total number of some event occurred independently in a given time interval follows a Poisson distribution [15], the count of interactions of a particular relation within a given time interval is assumed to follow a Poisson distribution. In the blockmodels proposed by Gallagher [10], the weight of a relationship is the count of interactions of some relation, e.g., the number of messages seen in an email network, which is modeled by a Poisson distribution. Here, we extend this blockmodel by considering multiple relations, i.e., there is one Poisson distribution associated with each relation for every role block.

Given a set of relations  $\mathbf{R} \triangleq \{R_1, R_2, \dots, R_m\}$  and relationship  $\vec{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \dots, Y_{ijm})$ , where each  $Y_{ijr}$  ( $1 \leq r \leq m$ ) is the interaction count of relation  $R_r$  following a Poisson distribution  $\text{Pois}(\lambda_r)$ . Let  $\vec{B}_{uv}$ , a column in  $\mathbf{B}$ , be a  $m$ -dimensional vector  $(\lambda_1, \lambda_2, \dots, \lambda_m)$ ; as relations are assumed to be independent, we can define the MVPDF as a *multiple Poisson distribution (mPois)* on  $m$  relations, i.e.

$$\text{mPois}_{m \rightarrow m}(\vec{Y}_{ij} | \vec{B}_{uv}) = \prod_{r=1}^m \text{Pois}(Y_{ijr}; \lambda_r)$$

The blockmodel for a weighted single relation network in [10] now becomes a special case where each role block  $B_{uv}$  is modeled by a single Poisson parameter  $\lambda_{uv}$ .

Our GSBM combines  $m$  weighted networks, each corresponding to a relation, into one multi-relational network. It therefore has  $m$  Poisson parameters to model one role block. The probability of a relationship is thus the product of the probabilities of having the interaction count of  $Y_{ijr}$  for each relation  $R_r$ .

### 4.2 MVPois: Multivariate Poisson MVPDF

Simply combining relations together without considering their correlations is far from modeling real networks. Relations are often correlated. For example, if a user is tagged in another user's photo, the former is also more likely to comment on the latter's photo.

In order to capture such correlations, we consider the co-occurrences of the relations. Define  $\mathcal{P}(\mathbf{R})$  the powerset of  $\mathbf{R}$  excluding the empty set, i.e.  $\forall S \in \mathcal{P}(\mathbf{R}), S \subseteq \mathbf{R}$  and  $S \neq \emptyset$ . For each set  $S$ , we model the correlation of relations in  $S$  by a Poisson distribution  $\text{Pois}(\lambda_S)$ . For instance, if  $S = \{\mathbf{W}, \mathbf{P}\}$ ,  $\text{Pois}(\lambda_S)$  models the event that the relations  $\mathbf{W}$  and  $\mathbf{P}$  happen together. The correlations among relations are thus captured through the subsets of relations. As the size of set  $\mathcal{P}(\mathbf{R})$  is exponential to the number of relations, we may restrict the size of set  $S$  to reduce the total number of subsets to be considered. For example, we can restrict  $S$  to consist of at most two relations. We define subsets of relations of not more than size 2 as  $\mathbf{S} = \{S \in \mathcal{P}(\mathbf{R}) : |S| \leq 2\}$  to capture the correlations with the minimum number of subsets,  $m + \binom{m}{2}$ .

$\{\mathbf{W}\}$	$\{\mathbf{P}\}$	$\{\mathbf{T}\}$	$\{\mathbf{W}, \mathbf{P}\}$	$\{\mathbf{P}, \mathbf{T}\}$	$\{\mathbf{W}, \mathbf{T}\}$
2	2	1	0	0	0
2	1	0	0	1	0
1	2	0	0	0	1
1	1	1	1	0	0
1	0	0	1	1	0
0	1	0	1	0	1
0	0	1	2	0	0

Table 1: All combinations of assignments from the relationship (2,2,1): the sums of  $\mathbf{W}$ ,  $\mathbf{P}$  and  $\mathbf{T}$  on each row yield 2, 2, and 1 respectively.

With the set  $\mathbf{S}$ , we propose using *Multivariate Poisson Distribution (MVPois)* as the MVPDF capturing correlations. Kawamura [14] studied trivariate Poisson distribution as a marginal probability on all possible combinations of subsets. Given a relationship  $\vec{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \dots, Y_{ijm})$ , there exists a combination of assignments  $w_S$  for all  $S \in \mathbf{S}$  satisfying  $m$  constraints,  $\sum_{R_r \in S \in \mathbf{S}} w_S = Y_{ijr}, \forall 1 \leq r \leq m$ .

Consider the previous example. Alice has a relationship (2, 2, 1) to Bob on relations  $(\mathbf{W}, \mathbf{P}, \mathbf{T})$ , and let  $\mathbf{S} = \{\{\mathbf{W}\}, \{\mathbf{P}\}, \{\mathbf{T}\}, \{\mathbf{W}, \mathbf{P}\}, \{\mathbf{P}, \mathbf{T}\}, \{\mathbf{W}, \mathbf{T}\}\}$ , each row in Table 4.2 is then a possible combination of assignments for  $S \in \mathbf{S}$ . As  $\{\mathbf{W}\}, \{\mathbf{W}, \mathbf{P}\}, \{\mathbf{W}, \mathbf{T}\}$  are the sets containing relation  $\mathbf{W}$ , the constraint that their assignments sum up to be 2 is satisfied for every row in Table 4.2. Same applies for relation  $\mathbf{P}$  and  $\mathbf{T}$ . We further denote all such combinations of assignments (e.g. all rows in Table 4.2) by  $\{\vec{W}_{\mathbf{S}} : C(\vec{Y}_{ij})\}$  where  $W_{\mathbf{S}} = \{\langle S, w_S \rangle : S \in \mathbf{S}\}$  and  $C(\vec{Y}_{ij})$  is the set of  $m$  constraints imposed from  $\vec{Y}_{ij}$ .

The probability of each combination of assignments can be obtained by the mPois distribution on set  $\mathbf{S}$ , since correlations among relations are now captured by sets in  $\mathbf{S}$  and the independence is assumed among sets in  $\mathbf{S}$ . So the probability of an observed relationship  $\vec{Y}_{ij}$  is the sum of the probabilities of all combinations  $W_{\mathbf{S}}$  with constraint  $C(\vec{Y}_{ij})$ . For example, the probability of the relationship (2, 2, 1) can be obtained by summing up all probabilities of the rows in Table 4.2.

Hence, we formulate the MVPois as follows.

$$(4.2) \quad \text{MVPois}_{|\mathbf{S}| \rightarrow m}(\vec{Y}_{ij} | \vec{B}_{uv}) = \sum_{\{W_{\mathbf{S}}: C(\vec{Y}_{ij})\}} \prod_{S \in \mathbf{S}} \text{Pois}(w_S; \lambda_S)$$

Each column  $\vec{B}_{uv}$  is a vector of  $|\mathbf{S}|$  Poisson parameters, each  $\lambda_S$  models the correlation of relations in set  $S$ . Our GSBM can therefore incorporate the MVPois to model correlations among relations in a multi-relational network. The choice of MVPois is our belief for social networks with correlated relations. There may be other more complicated multivariate functions that better describe multi-relational networks. A detailed study of these MVPDF functions will be a topic for our future work. In the next section, we will turn our attention to GSBM model estimation for a given multi-relational social network.

## 5 GSBM Model Estimation

We now describe how to build a good GSBM for social networks to discover roles and positions. We use Maximum Likelihood Estimation (MLE), a popular statistical method, to learn the parameters of the mixture model [13].

Given the blockmodel  $\theta = (\mathbf{\Pi}, \mathbf{B})$  and relationship  $\vec{Y}_{ij}$ , suppose the two latent variables  $z_{i \rightarrow j}$  and  $z_{i \leftarrow j}$  follow distribution  $q_{ij}$ , i.e.,  $q_{ij}(u, v | \vec{Y}_{ij}, \theta)$  is the probability of  $z_{i \rightarrow j} = u$  and  $z_{i \leftarrow j} = v$ . The *expected complete log likelihood* for  $\vec{Y}_{ij}$  is

$$\langle l_c(\theta | \vec{Y}_{ij}) \rangle_{q_{ij}} \triangleq \sum_{u,v} q_{ij}(u, v | \vec{Y}_{ij}, \theta) \log \Pr(\vec{Y}_{ij}, u, v | \theta)$$

where  $\sum_{u,v} q_{ij}(u, v | \vec{Y}_{ij}, \theta) = 1$ . An EM algorithm is then adopted to maximize the log likelihood in the above equation. In E-step, by *Jensen's Inequality* [4], the choice of  $q_{ij}(u, v | \vec{Y}_{ij}, \theta) = \Pr(u, v | \vec{Y}_{ij}, \theta)$  maximizes the above log likelihood. As in Equation 3.1,  $\Pr(\vec{Y}_{ij}, u, v | \theta) = \pi_{iu} \pi_{jv} \text{MVPDF}(\vec{Y}_{ij} | \vec{B}_{uv})$ ,

$$\begin{aligned} \Pr(u, v | \vec{Y}_{ij}, \theta) &= \frac{\Pr(\vec{Y}_{ij}, u, v | \theta)}{\Pr(\vec{Y}_{ij} | \theta)} \\ &= \frac{\pi_{iu} \pi_{jv} \text{MVPDF}(\vec{Y}_{ij} | \vec{B}_{uv})}{\Pr(\vec{Y}_{ij} | \theta)} \end{aligned}$$

In M-step, we optimize the expected complete log likelihood for all relationships, i.e.  $l_c(\theta | \mathbf{Y}) \triangleq \sum_{\vec{Y}_{ij} \in \mathbf{Y}} \langle l_c(\theta | \vec{Y}_{ij}) \rangle_{q_{ij}}$  with respect to model  $\theta$ . By setting  $\frac{\partial l_c(\theta | \mathbf{Y})}{\partial \pi_{iu}} = 0$  with the constraint  $\sum_u \pi_{iu} = 1$  for all  $i$ , we have

$$\pi_{iu} = \frac{\sum_{\vec{Y}_{ij}, v} q_{ij}(u, v | \vec{Y}_{ij}, \theta) + \sum_{\vec{Y}_{ji}, v} q_{ji}(v, u | \vec{Y}_{ji}, \theta)}{|\{\vec{Y}_{ij} \in \mathbf{Y}\}| + |\{\vec{Y}_{ji} \in \mathbf{Y}\}|}$$

To optimize the blockmodel  $\mathbf{B}$ , we first prove:

**THEOREM 5.1.** *A multiple Poisson distribution is equivalent to a joint probability distribution of a univariate Poisson distribution and a multinomial distribution, where the univariate Poisson distribution models the sum of all variables, and the multinomial distribution models the variables, i.e.*

$$\text{mPois}(\lambda_1, \lambda_2, \dots, \lambda_m) = \text{Pois}(\lambda) \times \text{mNom}(p_1, p_2, \dots, p_m)$$

where  $\lambda_i = \lambda p_i \quad \forall 1 \leq i \leq m$ .

*Proof.* Given  $\vec{w} = (w_1, w_2, \dots, w_m)$  and  $n = \sum_i w_i$

$$(5.3) \quad \begin{aligned} &\text{Pois}(n; \lambda) \times \text{mNom}(\vec{w}; p_1, p_2, \dots, p_m) \\ &= \frac{\lambda^n e^{-\lambda}}{n!} \cdot \frac{n!}{\prod_i w_i!} \prod_i p_i^{w_i} = \prod_i \frac{(\lambda p_i)^{w_i} e^{-\lambda p_i}}{w_i!} \\ &= \text{mPois}(\vec{w}; \lambda p_1, \lambda p_2, \dots, \lambda p_m) \end{aligned}$$

Equation 5.3 makes use of the condition  $\sum_i p_i = 1$ .

By the above theorem, Equation 4.2 can be revised to an equivalent definition. Define each block  $\vec{B}_{uv}$  by  $|\mathbf{S}| + 1$  parameters, with one Poisson parameter  $\lambda_{uv}$  and  $|\mathbf{S}|$  probabilities  $p_{uvS}$  satisfying  $\sum_{S \in \mathbf{S}} p_{uvS} = 1$ .

$$\begin{aligned} &\text{MVPois}_{|\mathbf{S}|+1 \rightarrow m}(\vec{Y}_{ij} | \vec{B}_{uv}) \\ &= \sum_{\{W_{\mathbf{S}}: C(\vec{Y}_{ij})\}} \text{Pois}\left(\sum_{S \in \mathbf{S}} w_S; \lambda_{uv}\right) \cdot \text{mNom}(W_{\mathbf{S}}; p_{uvS}) \end{aligned}$$

We now optimize the expected complete log likelihood  $l_c(\theta | \mathbf{Y})$  with respect to the above formulation. By setting  $\frac{\partial l_c(\theta | \mathbf{Y})}{\partial \lambda_{uv}} = 0$ ,  $\lambda_{uv}$  is the root to the polynomial

$$\sum_{\vec{Y}_{ij} \in \mathbf{Y}} q_{ij}(u, v | \vec{Y}_{ij}, \theta) \frac{\sum_{\alpha} \frac{1}{\alpha!} \cdot (\beta_{\alpha+1} - \beta_{\alpha}) \cdot \lambda_{uv}^{\alpha}}{\sum_{\alpha} \frac{1}{\alpha!} \cdot \beta_{\alpha} \cdot \lambda_{uv}^{\alpha}} = 0$$

where  $\beta_{\alpha} = \sum_{\{W_{\mathbf{S}}: C(\vec{Y}_{ij})\}} \sum_{S \in \mathbf{S}} w_S = \alpha$   $\text{mNom}(W_{\mathbf{S}}; p_{uvS})$  is a parameter with respect to a pair of  $(\vec{Y}_{ij}, \vec{B}_{uv})$ .  $\lambda_{uv}$  is thus obtained by Newton-Raphson method.

Setting  $\frac{\partial l_c(\theta | \mathbf{Y})}{\partial p_{uvS}}$  to a Lagrange multiplier [4] since  $\sum_{S \in \mathbf{S}} p_{uvS} = 1$ , for a particular  $S_0 \in \mathbf{S}$ , we have  $p_{uvS_0}$  proportional to

$$\sum_{\vec{Y}_{ij} \in \mathbf{Y}} q_{ij}(u, v | \vec{Y}_{ij}, \theta) \frac{\sum_{\{W_{\mathbf{S}}: C(\vec{Y}_{ij})\}} w_{S_0} \prod_{S \in \mathbf{S}} \frac{(\lambda_{uv} p_{uvS})^{w_S}}{w_S!}}{\sum_{\{W_{\mathbf{S}}: C(\vec{Y}_{ij})\}} \prod_{S \in \mathbf{S}} \frac{(\lambda_{uv} p_{uvS})^{w_S}}{w_S!}}$$

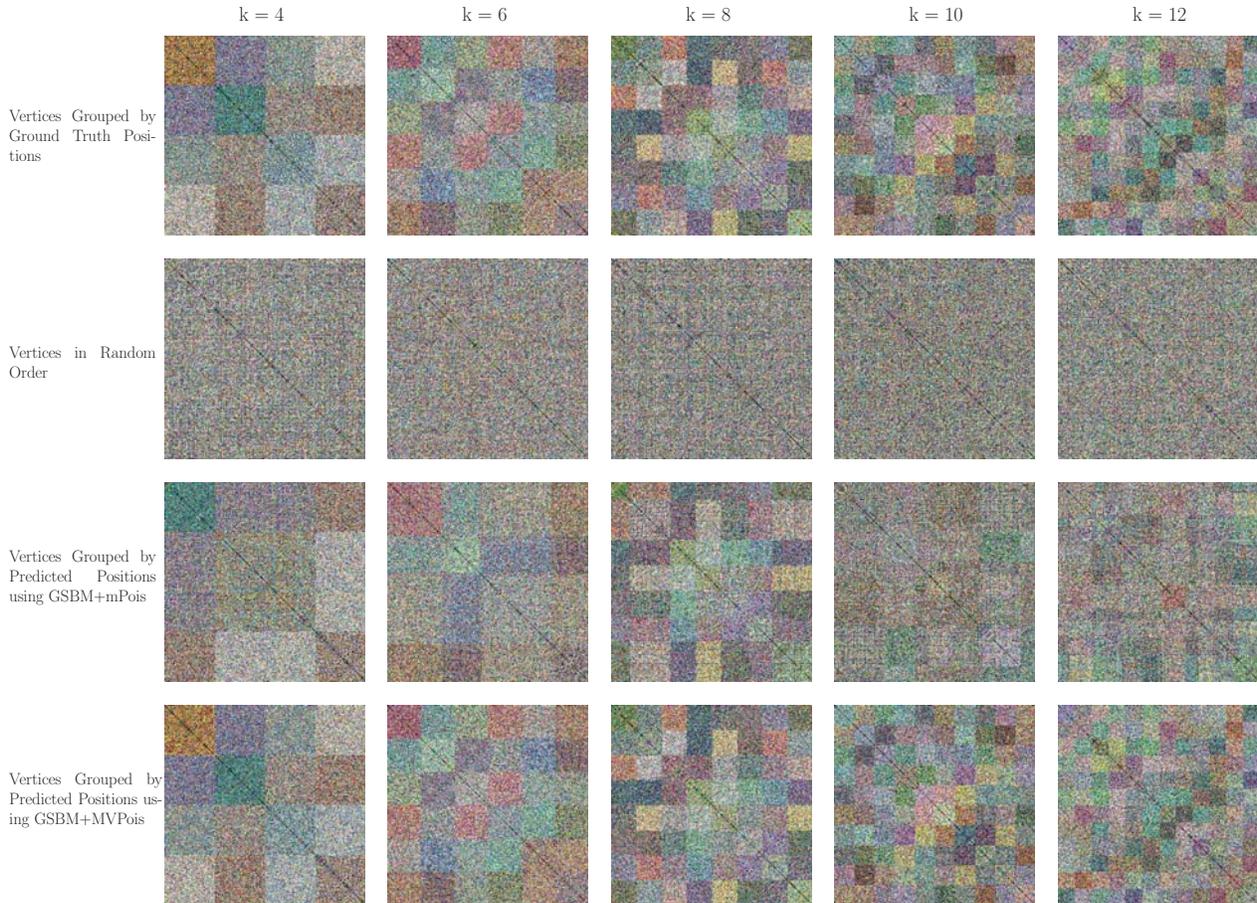


Figure 6: Blockmodels on synthetic data (if not printed in color, please refer to the electronic version)

With all the above formulae, EM algorithm iterates for the MLE, which will be validated in the later experimental study.

## 6 Experimental Study

In this section, we evaluate our proposed GSBM by using both synthetically generated networks and multi-relational networks extracted from the IMDb dataset. In Section 6.1, we aim to show that GSBM together with MVPois can perform well on recovering the hidden positions by visualizing such positions. We use GSBM with mPois as the baseline method to demonstrate the different results produced by ignoring the correlations among relations. Section 6.2 then focuses on validating our GSBM by relationship prediction on real world networks. A GSBM with MVPois is trained from a network in an earlier period, and then predicts relationships for a later period. Finally, we present a case study that reveals interesting positions discovered from the IMDb network in Section 6.3.

### 6.1 Evaluation using Synthetic Data

We first developed a strategy to generate synthetic multi-relational networks with vertices of known positions. By hiding the position labels, we performed position and role analysis using GSBM with mPois and MVPois, which were denoted by **GSBM+mPois** and **GSBM+MVPois** respectively. The predicted positions by GSBM+mPois and GSBM+MVPois were then compared with the ground truth positions and their accuracies were measured.

**Synthetic network generation.** For easy visualization, we assumed the multi-relational networks involved 3 relations represented by primary colors, i.e., red(R), green(G) and blue(B). These networks had 1000 vertices divided into  $k$  positions. Our experiment used  $k = 4, 6, 8, 10$  and  $12$ . We only considered correlation for at most 2 relations and this gave rise to  $\mathbf{S} = \{R, G, B, RG, RB, GB\}$  as the possible relation co-occurrences.

For each of the  $k \times k$  role blocks of relationships, we generated the Poisson parameter for each  $S \in \mathbf{S}$ . Each

block contained about  $\frac{1000}{k}$  vertices<sup>2</sup>. We randomly generated a Poisson parameter for each  $S$  from the uniform distribution  $U[0.2, 1.0)$ . Hence, each role block had a unique set of Poisson parameters. Next, we determined the relation occurrences of the relationship between every vertex  $i$  and vertex  $j$  in role block  $(u, v)$  respectively by first generating an integer value for each  $S \in \mathbf{S}$  using the Poisson parameters assigned to  $(u, v)$ . The red relation occurrence count of relationship  $(i, j)$  was the sum of the values on R, RG and RB; the green relation occurrence count was the sum of the values on G, RG and GB; the blue relation occurrence count was the sum of the values on B, RB and GB. The bitmaps in the first row of Figure 6 depicted the colors generated for each of the  $k \times k$  role blocks based on ground truth positions of vertices when the vertices were sorted by position. Note that each bitmap image had blocks assigned with distinctive color and block boundaries.

**Position prediction.** To ensure that vertices of different positions were well mixed before applying GSBM, we randomly shuffled the vertices as shown in the second row, leading to indistinguishable positions. We then applied the GSBM+mPois and GSBM+MVPois on the multi-relational networks of shuffled vertices. Once the positions of vertices were learned, we grouped them by position as shown in the last two rows in Figure 6. Note the positions were rearranged to make the bitmap most similar to that of the ground truth.

**Results.** Visually, vertices with correctly predicted positions when grouped together should have the bitmap of relationships look identical to that of relationships of vertices grouped by ground truth positions. GSBM+mPois which did not consider correlations was built on the three relations only. The bitmaps generated by GSBM+mPois had some resemblance with those of ground truth especially when  $k$  was small. As  $k$  increased, the resemblance became less distinct.

GSBM+MVPois that considered the correlations up to two relations, i.e. R, G, B, RG, RB and GB, demonstrated significantly better results. It generated bitmaps very similar to those of ground truth. This showed that the GSBM with MVPois can predict the positions with good accuracy.

## 6.2 Evaluation Using Real Dataset

**IMDb network.** We applied our GSBM+MVPois on IMDb<sup>3</sup> dataset. The IMDb dataset had several types of nodes and relations. We included

**actors/actresses, directors and producers.** Three relations were included, i.e., **collaborate** (among actors/actresses, among directors, or among producers), **direct** (from directors to actors/actresses) and **work\_for** (from actors/actresses to producers and from directors to producers). We extracted a connected IMDb Network as follows.

We started from a set of movies between 2003 and 2006 (inclusive) which are directed by eleven directors, James Cameron, Chris Columbus, Jon Favreau, Ron Howard, Doug Liman, Christopher Nolan, Guy Ritchie, Martin Scorsese, Steven Soderbergh, Steven Spielberg and David Yates. We then expanded the set of movie twice by the following way. Among the set of movies, we selected actors, actresses and producers who participated in at least two movies. Then, we obtained a new set of movies in which at least two of the above actors, actresses and producers participated. Within this set of movies, we chose those directors and those actors, actresses or producers who participated in at least two movies. We thus obtained a set of 1530 users, and a network was build from this set of users with the relations mentioned above.

**Performance measures.** We measured the accuracy of relationship prediction by precisions. For a pair of individuals  $(x_i, x_j)$ ,  $\vec{Y}_{ij}$  denote the actual relationship from  $x_i$  to  $x_j$ . Let the set of relationships in the test data be  $\mathbf{E} = \{\vec{Y}_{ij} : \exists r, Y_{ijr} > 0\}$ , i.e.,  $\mathbf{E}$  is the set of individual pairs where a certain relationship exists. Let  $\hat{Y}_{ij}$  be the predicted relationship from  $x_i$  to  $x_j$  that maximizes the probability, i.e.,  $\hat{Y}_{ij} = \arg \max_{\vec{e} \neq \mathbf{0}} p(\vec{e} | \boldsymbol{\theta}, i, j)$ . The precision is measured by

$$\text{Precision} = \frac{|\{\vec{Y}_{ij} : \hat{Y}_{ij} = \vec{Y}_{ij} \wedge (x_i, x_j) \in \mathbf{E}\}|}{|\mathbf{E}|}$$

However, if there is only one edge type in the test edge set  $\mathbf{E}$ , prediction  $\hat{Y}_{ij}$  is trivial. Thus edge types which appear more frequently should be penalized, compared to edge types that are less frequent. We have therefore considered the *soft precision*, which is the precision weighted by the negative log of its probability, i.e.

$$\text{Soft Precision} = \frac{-\sum_{\hat{Y}_{ij} = \vec{Y}_{ij} \wedge (x_i, x_j) \in \mathbf{E}} \log p(\vec{Y}_{ij})}{|\mathbf{E}| \cdot H(\mathbf{E})}$$

Given edge type  $\vec{e}$  appears with probability  $p(\vec{e})$  in the test set  $\mathbf{E}$ , every correct prediction  $\hat{Y}_{ij} = \vec{Y}_{ij}$ , where  $\vec{Y}_{ij}$  is of type  $\vec{e}$ , will carry a score of  $[-\log p(\vec{e})]$  to the soft precision. The maximum is thus  $|\mathbf{E}| \cdot H(\mathbf{E})$ , where  $|\mathbf{E}|$  is the total number of the edges and  $H(\mathbf{E})$  is the entropy of the edges. Soft precision is thus normalized by the maximum, as given in the above equation.

<sup>2</sup>The bitmaps were drawn based on  $\frac{1}{3}$  of the vertices due to the large file size of  $1000 \times 1000$  bitmaps

<sup>3</sup><http://www.imdb.com>

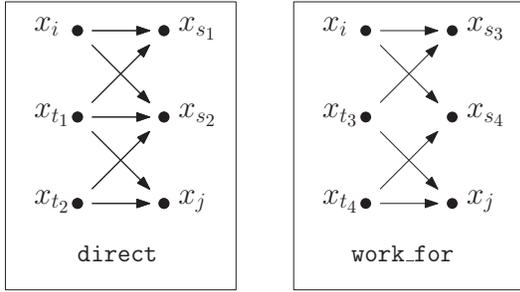


Figure 7: An illustration of modified common neighbor method on relations `direct` and `work_for` from  $x_i$  to  $x_j$

**Baseline Method.** As there hardly exists a method that predicts relationships between two individuals, we modified the well-known common neighbor method in the context of directed network. When common neighbor method is applied on an undirected network, a pair of vertices with more common neighbors is more likely to form a link than a pair with less common neighbors. In a directed network, if the relation is not transitive, there may not be any common neighbor between a pair of potential vertices. For example, assume a network derived by `direct` relation is a bipartite graph with one side all directors and the other side all actors or actresses (i.e., assume no overlapping between directors and actors/actresses); between a director and an actress, which potentially forms a `direct` edge, there is no common neighbor since all edges from the director are connected to actors/actresses, while all edges from the actress are connected to directors. Therefore, we define the number of paths of length 3 as the *modified common neighbor* in a directed network constructed by relation  $R_r$ , i.e.,

$$MCN(i, j; R_r) = |\{(s, t) : \exists Y_{isr} > 0 \wedge Y_{tsr} > 0 \wedge Y_{tjr} > 0\}|$$

The definition of modified common neighbor can be considered as an analogue of the original definition, in the sense that, the number of common neighbors in the original definition is the number of paths of length 2 connecting  $i$  and  $j$ , while MCN is the number of paths of length 3 with the middle edge having the opposite direction in the definition of modified common neighbor.

As shown in Figure 7, there are three pairs of indices served as “common neighbor” from  $x_i$  to  $x_j$  on relation `direct`, namely,  $MCN(i, j; \text{direct}) = \{(s_1, t_1), (s_2, t_1), (s_2, t_2)\}$ . But there are only two pairs between  $x_i$  and  $x_j$  on relation `work_for`:  $MCN(i, j; \text{work_for}) = \{(s_3, t_3), (s_4, t_4)\}$ .

The baseline method assumes the relationship from  $i$  to  $j$  is a pair  $\{(R_0, 1)\}$ , where  $R_0$  is the relation which gives maximum  $|MCN|$ , i.e.,  $|MCN(i, j; R_0)| \geq$

$|MCN(i, j; R_r)| \forall 1 \leq r \leq m$ . In Figure 7, the baseline method will not predict  $\{\{\text{work\_for}, 1\}\}$  since  $|MCN(i, j; \text{direct})| > |MCN(i, j; \text{work\_for})|$ . As common neighbor method is pretty accurate for link prediction, the precision for modified common neighbor method on relationship prediction can reach more than 80% as we shall see in the next paragraph.

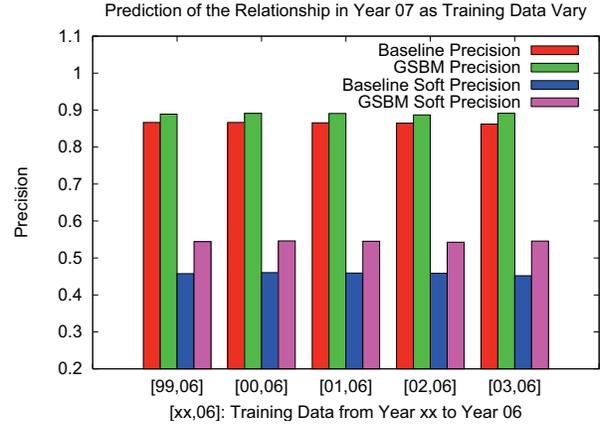


Figure 8: Relationship predictions in IMDb network for the year 07 with different sets of training data

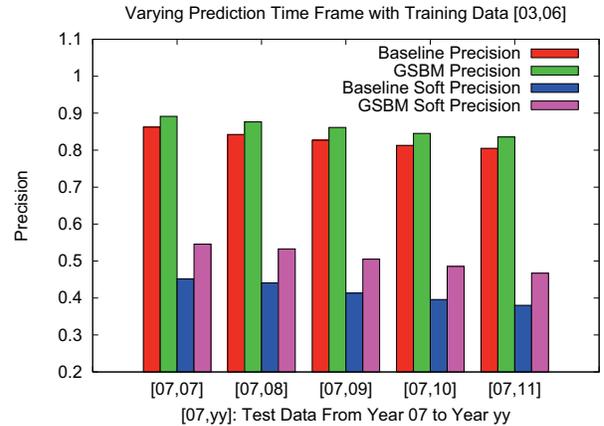


Figure 9: Relationship predictions in IMDb network over different time frames with the same set of training data

**Results.** In Figure 8 and Figure 9, we compared our GSBM model (using MVPois) applied on the IMDb network with the baseline method. As most of the edges have the count of 1 on one relation and 0 on the other two relations, the baseline method can reach more than 80% precision. Figure 8 showed the precision results on Task 1, i.e., predicting the relationship in the year of 2007 with training data on various time frames before 2007. Our model improved the precision

to 90%, almost one quarter of the maximum possible improvement. On the measure of soft precision, GSBM outperformed the baseline by about 20% in all cases, as the additional edges predicted correctly by GSBM carried heavier weights than those predicted correctly by the baseline method. The precision do not always improve with more training data. This could be due to outdated data for a long period of training data. For task 2, as shown in Figure 9, our GSBM model again outperformed baseline in all the test periods. We also observed the decrease in both precision and soft precision as longer time period of test data was involved. This was expected since the information obtained from the training data became more outdated for the recent relationships.

These results on networks extracted from IMDb dataset confirmed the usefulness of the positions in predicting future relationships.

### 6.3 A Case Study

We also gave a case study to show that the positions found by our GSBM are sound. From the eleven directors mentioned in our IMDb dataset, we selected their 73 movies directed by them from year 2000 to 2010. 486 directors, producers and actors/actresses involved in at least two of these movies. For this smaller IMDb network, we learned the model using GSBM and MVPois with  $k = 6$ .

The clusters formed by the learned positions are shown in Figure 10. The green edges are undirected, representing the `collaborative` relation, whereas the blue edges and the red edges are directed, representing the `direct` and `work_for` relations respectively. The thicker an edge (relation) is, the more interactions it carries.

Out of the 11 directors, 10 directors were together in block A, except James Cameron, who acted in 2 movies, directed 3 movies and also produced 4 movies. His position as a director was not so obvious as other directors, he was thus outside the director block, i.e., block A.

There was also a producer block D, consisting of 6 active producers, who produced up to 10 movies. Other producers did not produce so many movies, and many of them acted in the movies they produced.

There were two interesting actors/actresses blocks, B and F. Block B included the actors/actresses in Ocean's Eleven/Twelve/Thirteen Series, e.g., George Clooney (Danny), Brad Pitt (Rusty), and Matt Damon (Linus). Block F included the actors/actresses in Harry Potter Series, e.g., Daniel Radcliffe (Harry), Emma Watson (Hermione), and Rupert Grint (Ron). As this IMDb network was dominated by the `collaboration`

relation, actors/actresses played in a series of movies are grouped together.

The remaining two blocks, although not as notable as the previous, can also be explained. Block C had 113 people who were mostly producers with relatively fewer movies as compared to block D, which included active producers. Block E has 145 people who were actors/actresses not involved in any special series, or series that did not see many overlapping actors, e.g. the Da Vinci Code and Angles & Demons.

This case study demonstrated that the blocks discovered by our proposed model matches well with the intuition. We therefore conclude that GSBM with multivariate Poisson is an effective method for conducting position and role analysis on a multi-relational network.

## 7 Conclusion

We presented a generalized stochastic blockmodels (GSBM) that discover structures on multi-relational networks. For multi-relational social networks, we proposed to use multivariate Poisson distribution as the MVPDF for our GSBM. Using the sets formed by relations, we are able to accurately capture the correlation among the relations. We then demonstrated an instance of our GSBM using multivariate Poisson modeling and estimated its parameters using EM algorithm. Experimental studies on both synthetic and real data show that our GSBM is effective for discovering structures and predicting the relationship of future links.

## 8 Acknowledgment

We will like to thank Agus Trisnajaya Kwee for helping us with the visualization of multi-relational networks.

This work is supported by Singapore's National Research Foundation's research grant, NRF2008IDM-IDM004-036.

## References

- [1] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, June 2008.
- [2] Michele Berlingerio, Michele Coscia, and Fosca Giannotti. Finding and characterizing communities in multidimensional networks. In *ASONAM*, pages 490–494, 2011.
- [3] Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino Pedreschi. Foundations of multidimensional network analysis. In *ASONAM*, pages 485–489, 2011.
- [4] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

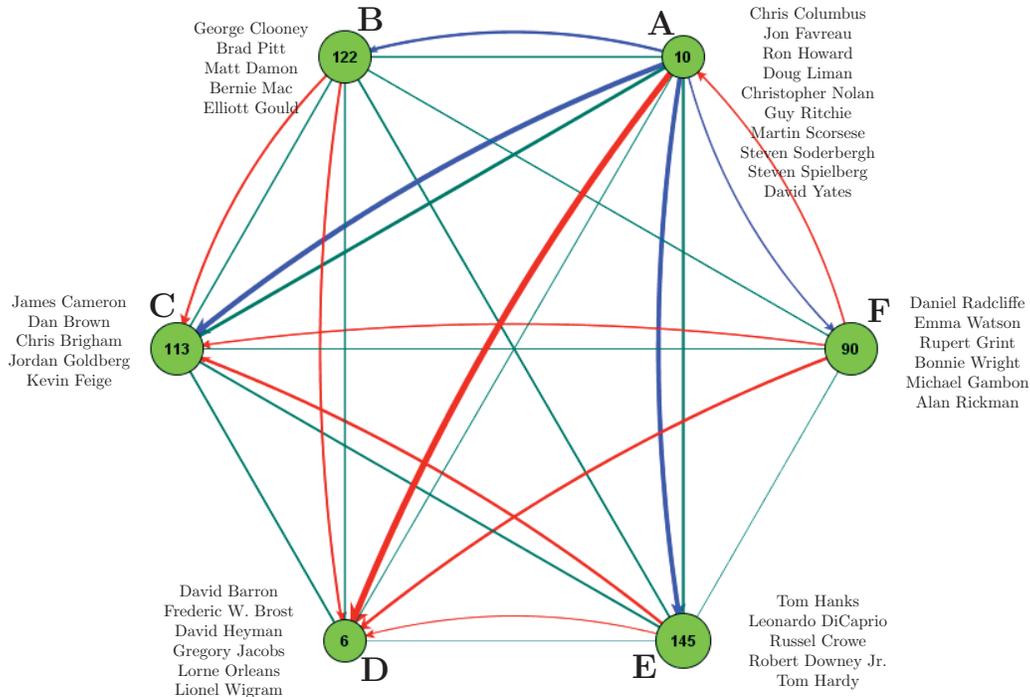


Figure 10: A case study on the blockmodel constructed on 73 movies with  $k = 6$

- [5] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, April 2010.
- [6] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han. Community mining from multi-relational networks. In *PKDD*, pages 445–452, 2005.
- [7] Patrick Doreian, Vladimir Batagelj, and Anuška Ferligoj. *Generalized Blockmodeling*. Cambridge University Press, 2005.
- [8] Joan Ferrante. *Sociology: a global perspective*. Thomson Wadsworth, 2008.
- [9] Stephen E. Fienberg, Michael M. Meyer, and Stanley S. Wasserman. Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389):51–67, March 1985.
- [10] Ian Gallagher. Bayesian block modelling for weighted networks. In *MLG*, pages 55–61, 2010.
- [11] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. Resisting structural re-identification in anonymized social networks. *Proc. VLDB Endow.*, 1:102–114, August 2008.
- [12] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983.
- [13] Michael I. Jordan. *An Introduction to Probabilistic Graphical Models*. in preparation, 2002.
- [14] Kazutomo Kawamura. The structure of trivariate poisson distribution. *Kōtai Mathematical Seminar Reports*, 28(1):1–8, 1976.
- [15] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, 1996.
- [16] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, August 2010.
- [17] Lei Tang, Huan Liu, and Jianping Zhang. Identifying evolving groups in dynamic multimode networks. *IEEE Trans. Knowl. Data Eng.*, 24(1):72–85, 2012.
- [18] Lei Tang, Huan Liu, Jianping Zhang, and Zohreh Nazeri. Community evolution in dynamic multi-mode networks. In *KDD*, pages 677–685, 2008.
- [19] Aleš Žiberna. Generalized blockmodeling of valued networks. *Social Networks*, 29(1):105–126, January 2007.
- [20] Yuchung J. Wang and George Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, March 1987.
- [21] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [22] Harrison C. White, Scott A. Boorman, and Ronald L. Breiger. Social structure from multiple networks. i. blockmodels of roles and positions. *The American Journal of Sociology*, 81(4):730–780, January 1976.
- [23] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *WWW*, pages 981–990, 2010.