Dependent Stotchastic Blockmodels

Eunsil Gim, Sojeong Ha, and Seungjin Choi

Abstract— **A stochastic blockmodel is a generative model for blocks, where a block is a set of coherent nodes and relations between the nodes are explained by the corresponding pair of blocks. Most existing methods make use of both the presence and the absence of links between nodes, encoded by the adjacency matrix, to learn the corresponding models. In this paper, we present a new method in which we use only the presence of links to learn the model, exploiting the dependency between source and destination nodes, leading to a *dependent stochastic blockmodel*. We allow for mixed membership and the degrees of nodes in our dependent stochastic blockmodel. Experiments on the political books network and Twitter social network indicate that the behavior of our dependent stochastic blockmodel is superior to that of existing methods.**

I. INTRODUCTION

A network is an efficient way to describe relational data in the real world. Entities and their relations are conceptualized as nodes and edges in networks. Networks are often simplified by using only a few essential characteristics of nodes and edges. In this way, although some subsidiary information is lost, we get valuable insights about networks, model them in a manageable form, and prevent overfitting.

Clustering is one such approach. It concentrates on the fact that nodes are often divided into subsets with similar characteristics. For example, in social networks, we can know much about a person if it is known about his/her job, religion, family, hobby, or location. These pieces of information would help us to determine clusters. Clustering creates proper subsets of nodes, based on the algorithm. Classical clustering is based on the assumption that similiar nodes are more likely to be connected. Some clustering methods are extended to let nodes belong to multiple clusters using a mixture model or allow both assortative and disassortative relations between clusters [1] (see Fig. 1). Still, clustering does not learn about the interactions in and between the clusters.

A blockmodel is a model that describes subsets of nodes with similar characteristics [2], [3]. It does not matter if network is assortative, disassortative, or even any mixture of them, because blockmodel also models relations in and between the subsets. This results in higher accuracy of prediction. The subsets of nodes in blockmodel are called blocks. They are mutually exclusive and collectively exhaustive subsets of a complete node set, where nodes in identical blocks have homogeneous patterns of link generation. Using

Eunsil Gim is with Samsung, Korea (email: fortru0@gmail.com). Sojeong Ha and Seungjin Choi are with the Department of Computer Science and Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang 790-784, Korea (email: {hsi0317,seungjin}@postech.ac.kr).

a blockmodel, we can predict unobserved links or generate a synthetic network.

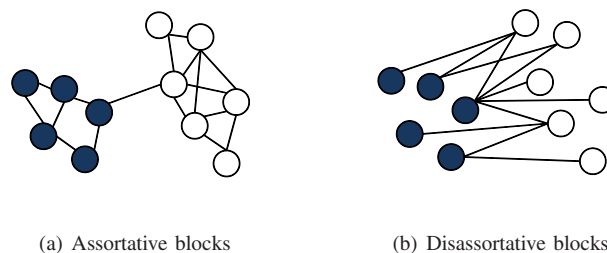(a) Assortative blocks   (b) Disassortative blocks

Fig. 1.   Block structures with two blocks

A blockmodel learns two kinds of parameters: the affiliation of nodes to blocks and the relationship between blocks. Relationships between blocks determine the edges between the nodes affiliated to each block. These parameters are matrices with an element for each node-node and node-block pair, i.e. if there are N nodes and K blocks, there is a $K \times N$ matrix for affiliation of nodes to blocks, and a $K \times K$ matrix for relations between blocks. Each element of the $K \times K$ matrix indicates the probability of link existence in and between the blocks, and thus it captures any relation between blocks with almost no limitation: not only assortativity and disassortativity, but also any combination of them. It makes the relations between blocks flexible, and improves the representation of each block's characteristics. Furthermore, it is efficient because it does not require knowledge of each node-to-node interaction pattern, but only block-to-block and node-to-block interactions.

Initially, a blockmodel was hypothesized as a deterministic relation between blocks [4]. The relationship between blocks was represented as either 1 or 0, where 1 signified links and 0 the absence of links. In this model, each node belonged to exactly one block. Therefore, any two blocks with 1 as their relationship had edges on every possible node pair between them.

Stochastic blockmodels were subsequently introduced [5], [6]. They changed the deterministic relationship between blocks and adopted the probability to generate links. In a stochastic blockmodel, links are drawn from Bernoulli distribution with the relations between blocks as parameters. However, each node still belonged to one block deterministically.

In a further step, variance of degrees was introduced [7], [8], [9], [10]. Variance of degrees is an important feature of networks, and has even been highlighted for its scale-free characteristic and link prediction accuracy [11]. Blockmodels

that have no degree variance sometimes clusters the nodes according to their degrees in networks with high degree variance of nodes. This results in an erroneous community structure and, of course, poor prediction results. Blockmodels that have degree variance treat both block structure and degrees with equal importance. Additional care for degrees prevents the clustering according to the degrees of the nodes.

Node to blocks mixed-membership was introduced to describe the phenomena of people often having multiple roles. The mixed-membership stochastic blockmodel [12] has enabled nodes to affiliate number of blocks with weights on blocks. In this model, the blocks for each node are selected independently, for $N \times N$ of every node pair. On the basis of the selected blocks of each node pair, a link is drawn using Bernoulli distribution.

However, it is a questionable idea that the blocks that appear on an edge are chosen independently. When one meets one's colleagues, one does not act as their family member, but as a colleague in the same field. In all the multiple social roles one plays, one's selection of a role is affected by the other person's role in the interaction. That is, one's social behavior pattern in each interaction is correlated with other's social behavior pattern in that interaction.

In this paper, we propose dependent stochastic blockmodels in which we exploit the dependencies between the chosen blocks by selecting the block of the destination node based on the block of the source node. To encompass this idea, we restructure our blockmodel to correspond to a sequential process: from drawing the source node to drawing the destination node, where each step is dependent on the previous step. This process generates directed links, as the source node actively reaches out, where as the destination node passively receives links. The Twitter network is a good example of such a process: a click by a follower is enough to generate a link, without the followee taking any action to generate that link. The process for voting, citation, and blog hyperlink networks is similar.

As each process, from drawing the source node to the drawing destination node, makes an edge, our model scales linearly in the number of edges, for both learning and generating networks.

Moreover, the process we propose naturally adapts both degree-correction and mixed-membership concepts during the drawing of nodes and blocks. Describing both characteristics of a network not only results in a simple merging, but also creates synergy in modeling networks. We describe our experiments and results that demonstrate this in Section IV.

The main contributions of this paper are summarized as follows.

- We exploit source-to-destination dependency, allowing the model to learn from only the presence of edges, while most existing models make use of both the presence and the absence of links between nodes.
- Our model allows for both degree-correction and mixed membership, leading to *degree-corrected mixed mem-*

bership stochastic blockmodels, which describe networks more accurately.

The rest of this paper is organized as follows. Section II reviews existing blockmodels. In Section III, we propose concepts and algorithms for dependent stochastic blockmodels; in this section, the model's extension and issues are also covered. Section IV presents three experiments using real networks to verify our proposed dependent stochastic blockmodel.

II. Related Work

In this section, we review existing blockmodels. The standard stochastic blockmodel is presented first, and then its extensions are presented with highlights on the additional properties.

A. Standard Blockmodel

A blockmodel partitions the nodes in network into blocks that are mutually exclusive and collectively exhaustive subsets of nodes [4]. It assumes that the characteristics of each node are well represented by the characteristics of the block to which the node belongs. In a network, the only characteristic a node has is its links. Therefore, we assume that nodes in an identical block have a homogeneous link generation pattern. Thus, in a blockmodel, node-to-block affiliation and the interaction between blocks describes all node-to-node interaction. The interactions between blocks are recorded deterministically in standard blockmodels, with 1 signifying the presence of links, and 0 signifying the absence of links. In the standard model, nodes in the same block are perfectly homogeneous: if there is a node s that has a link with one of the nodes in the block, then all the other nodes in the block have a link with node s, and vice versa.

Stochastic blockmodels [5], [6] change the deterministic relationship stochastically. The relationships between blocks ($\boldsymbol{B}$) are presented as an $N \times N$ matrix in which each element signifies the probability of a link between two blocks. Bernoulli distribution is used to draw links between nodes. In a stochastic blockmodel, nodes in the same block are stochastically equivalent. That is, if s and s' are two nodes in the same block, the probability of a link existing between s and d is the same as that between s' and d, for any node d. However, node-to-block affiliation is deterministic. Affiliation vector $\boldsymbol{z}_s$ is a binary vector that denotes the block membership of each node s with one element that states that the affiliated block is 1, while other elements are 0.

In this model, we represent relational data as a network $\boldsymbol{G} = (\boldsymbol{N}, \boldsymbol{A})$, where $\boldsymbol{A}$ is an adjacency matrix. $\boldsymbol{Z}$ is a $K \times N$ matrix in which each column is an indicator vector for the block to which each node belongs. The probability of the network is calculated by drawing a link for every node pair.

$$p(\boldsymbol{A}|\boldsymbol{B}, \boldsymbol{Z}) = \prod_{s=1}^{N} \prod_{d=s}^{N} (\boldsymbol{z}_s^T \boldsymbol{B} \boldsymbol{z}_d)^{A_{s,d}} (1 - \boldsymbol{z}_s^T \boldsymbol{B} \boldsymbol{z}_d)^{1-A_{s,d}}. \quad (1)$$

B. Degree-Corrected Stochastic Blockmodel

Degree variance is one of the most important features of networks, including social networks [11]. Some of the degree variations are explained by the relationship between blocks, but there are variances even in the nodes in the same block. Furthermore, the blockmodels without degree variance sometimes results fallacious blocks that were classified by their nodes' degrees.

A degree-corrected stochastic blockmodel [10] is a model that adopts degree variance to blockmodels. This model was developed for undirected networks with multi-edges. Degree correction for each node is exploited by a new parameter θ. It is used in calculating the number of links between nodes s and d from Poisson distribution, where the expected number of links is $\theta_s\theta_d z_s^T B z_d$. It is assumed that each node is deterministically affiliated in one block, just as in standard blockmodels. Therefore, z_s and z_d are binary vectors with 1 for the element that corresponds to the block to which nodes s and d belongs. The probability of a link is not just a bare relationship between groups, but is multiplied by θ of each node, to encompass each node's expected degree.

A degree-corrected stochastic blockmodel generates networks by drawing a number of links for every node pair. It assumes that the number of links is independently Poisson distributed, where the expected number of links between nodes s and d is $\theta_s\theta_d z_s^T B z_d$. For convenience, calculation of the probability of a network G under the degree-corrected stochastic blockmodel is divided into two parts: probability of links between blocks and probability of links within blocks. The following is the probability of the entire network:

$$
\begin{aligned}
&p(\boldsymbol{A}|\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{B}) \\
&= \prod_{s=1}^{N} \prod_{d=s+1}^{N} \frac{(\theta_s\theta_d z_s^T B z_d)^{A_{s,d}}}{A_{s,d}!} exp(-\theta_s\theta_d z_s^T B z_d) \\
&\times \prod_s \frac{(\frac{1}{2}\theta_s^2 z_s^T B z_s)^{A_{s,s}/2}}{(A_{s,s}/2)!} exp(-\frac{1}{2}\theta_s^2 z_s^T B z_s).
\end{aligned}
$$

(2)

Although the degree-corrected stochastic blockmodel draws a link for each node pair, its algorithm for learning scales in number of nodes. Thanks to deterministic block affiliation, the learning is done by choosing the most probable block for each node, given blocks of all other nodes. Since it chooses the best one, it converges quickly, but it might get stuck in local maxima. Like the previous blockmodels, the blocks of nodes on interaction are independent.

C. Mixed-Membership Stochastic Blockmodel

People belong to several groups - family, school, work, friends in various fields, hobby, etc. This contradicts the assumption of the standard blockmodel that each node is affiliated in exactly one block. Adopting this idea, the mixed-membership stochastic blockmodel [12] extends classical stochastic blockmodel by allowing nodes to be affiliated with multiple blocks. Affiliation between nodes and blocks ($\boldsymbol{\Pi}$)
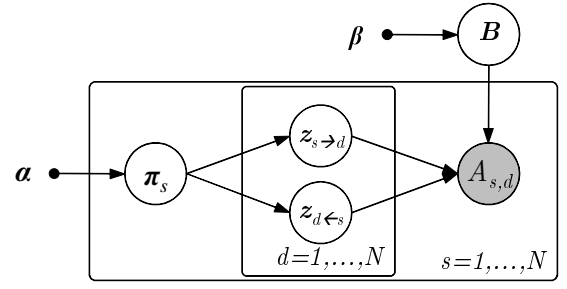

Fig. 2. Graphical representation of the mixed-membership stochastic blockmodel.

does not deterministically record the affiliated block, but it provides each node's probability of affiliation to each block.

The mixed-membership stochastic blockmodel draws a link for each node pair. The procedure begins with the drawing of the affiliated blocks of two nodes. Given those blocks, we draw link existence from Bernoulli distribution on association between blocks ($\boldsymbol{B}$). $z_{s\to d}$ and $z_{d\leftarrow s}$ are indicator vectors for the chosen source and destination blocks on the interaction between nodes s and d. In the original mixed-membership stochastic blockmodel, $\boldsymbol{B}$ is given as a parameter. We smoothed it to be drawn from beta distribution, as in [13]. A graphical model representation is in Figure 2. The corresponding procedure is given as:

- Choose $\boldsymbol{\Pi} \sim \text{Dir}(\boldsymbol{\alpha})$
- Choose $\boldsymbol{B} \sim \text{Beta}(\boldsymbol{\beta})$
- For each node pair $(s,d) \in N \times N$
  1) Choose a membership indicator for the source node,
     $z_{s\to d} \sim \text{Mult}(\boldsymbol{\pi}_s)$
  2) Choose a membership indicator for the destination node,
     $z_{d\leftarrow s} \sim \text{Mult}(\boldsymbol{\pi}_d)$
  3) Choose the existence of link,
     $A_{s,d} \sim \text{Bernoulli}(z_{s\to d}^T B z_{d\leftarrow s})$

There are $K \times K$ ways of block selection for each node pair. If the link existence is right, different block selection does not make any difference on the observed data. Accordingly, the probability of each node pair is the summation of the probability through all $K \times K$ block pairs. The resultant probability of a network is,

$$
\begin{aligned}
&p(\boldsymbol{A}, \boldsymbol{\Pi}, \boldsymbol{B}, \boldsymbol{Z}_\to, \boldsymbol{Z}_\leftarrow|\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&= \prod_{s=1}^{N} p(\boldsymbol{\pi}_s|\boldsymbol{\alpha}) \times p(\boldsymbol{B}|\boldsymbol{\beta}) \\
&\times \prod_{s,d=1}^{N} p(z_{s\to d}|\boldsymbol{\pi}_s)p(z_{d\leftarrow s}|\boldsymbol{\pi}_d)p(A_{s,d}|z_{s\to d}, z_{d\leftarrow s}, \boldsymbol{B}).
\end{aligned}
$$

(3)

In addition to the parameters above, the sparsity parameter ρ is used to down-weight the probability of links. Some of the absent links actually exist, but are unobserved. These links should not be explained by the model, and the weight of the
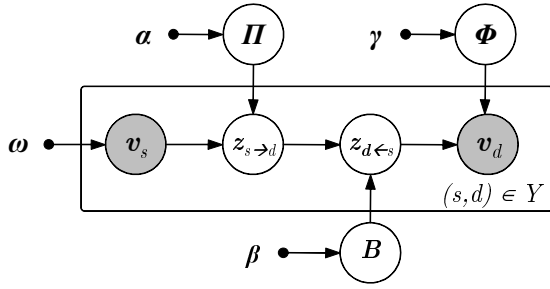
Fig. 3. Graphical representation of the dependent stochastic blockmodel for two-mode network.

unobserved links is captured by ρ. The probability of a link given two membership indicators z_s and z_d is reevaluated as $\rho z_s^T B z_d$.

III. DEPENDENT STOCHASTIC BLOCKMODEL

In this section, we describe the dependent stochastic blockmodel. The new structure we propose for the dependent stochastic blockmodel deals with the source and destination nodes separately, which matches two-mode networks. Thus, first we develop the model for directed two-mode networks; further, in Section III-A, we specialize it with minor changes for one-mode networks. Some issues related to our model are dealt with in Section III-B. In this section, we focus on unweighted and directed networks without multi-edges.

A dependent stochastic blockmodel is a generative probabilistic model that, captures dependencies between blocks. To reflect the dependency between a source and destination node, we restructure the process of generating an edge. In our model, each edge is generated by sequentially drawing the source node, source block, destination block, and then the destination node, conditional on the previous variable. The step of drawing a destination block that is conditional on the source block captures the dependency between the blocks. Moreover, the step of drawing a source node reflects degrees, and drawing a destination node reflects degrees that are conditional on the destination block. In addition, a source block is drawn from a source node using mixed-membership. It is notable that a dependent stochastic blockmodel does not have isolated degree and mixed-membership terms; it naturally mixes them.

Since a dependent stochastic blockmodel deals with source nodes and destination nodes separately, this model provides explanations of two-mode networks. A two-mode network is a network where the vertices are divided into two disjoint subsets, and edges connect vertices in different subsets. Thus, source nodes and destination nodes have different characteristics.

In a dependent stochastic blockmodel, the node sets are further defined as a source node set and a destination node set, where every edge is initiated from the source node set and reaches out to the destination node set. N_1, N_2 and K_1, K_2 are used to denote the number of nodes and blocks in each node set.

The following is the procedure used to generate a network in a two-mode network, with corresponding graphical model in Figure 3. Π is a $K_1 \times N_1$ matrix that represents the probability of each source block being conditional on the source node, and Φ is a $K_2 \times N_2$ matrix that represents the probability of each destination node being conditional on the destination block. B is a $K_1 \times K_2$ matrix that denotes the probability of each destination block given a source block. $\Pi, \Phi,$ and B are drawn from Dirichlet distribution. v_s and v_d are binary indicator vectors for source and destination nodes, and $z_{s \to d}$ and $z_{d \leftarrow s}$ are binary indicator vectors for source and destination blocks on the interaction between nodes s and d. We assume that ω, the out-degree of nodes in the training data, is given.

- Choose $\Pi \sim \text{Dir}(\alpha)$
- Choose $B \sim \text{Dir}(\beta)$
- Choose $\Phi \sim \text{Dir}(\gamma)$
- For each edge $(s, d) \in Y$
  1) Choose a source node indicator,
     $v_s \sim \text{Mult}(\omega)$
  2) Choose a membership indicator for the source node,
     $z_{s \to d} \sim \text{Mult}(\Pi v_s)$
  3) Choose a membership indicator for the destination node,
     $z_{d \leftarrow s} \sim \text{Mult}(z_{s \to d}^T B)$
  4) Choose a destination node indicator,
     $v_d \sim \text{Mult}(z_{d \leftarrow s}^T \Phi)$

The new link generation procedure requires a new definition of the relationships between blocks. In the existing blockmodels, the relationship between blocks is the probability of a link between two blocks. In our model, the relationship between blocks, which is each row of B, is redefined as the probability of choosing each block as a destination block, given the source block. This section illustrates a natural dependency between blocks, that is, each node's choice of block depends on the other node's block.

Degree variance and mixed-membership are required two times each. Degree variance is used to select nodes. In step 1, since we do not have prior knowledge of the link, nodes are drawn completely by their probability to initiate a link ω. In step 4, however, since we know the block to which a selected node belongs, selection is done by the multinomial distribution with $z_{d \leftarrow s}^T \Phi$, which is the probability of each node given a destination block.

Mixed-membership is used on the relationship between nodes and blocks in steps 2 and 4. Step 2 is purely about the mixed-membership of each node using the latent variable Π. However, step 4 indirectly encompasses mixed-membership to give positive probability for all block-node pairs using latent variable Φ.

The proposed procedure naturally scales in the number of edges (E). This is a very useful characteristic because most real networks are very sparse. The algorithm gets much faster as there are fewer links for each node. In the typical case of a sparse network, E is several constant times the number

of nodes. However, we do not lose much information by ignoring the absence of links. Links in sparse matrices embed much information because they are rare with high perplexity. Furthermore, we cannot distinguish between absent links and unobserved links, which leads to erroneous no-link cases.

The joint probability of a network $G = \{V, Y\}$, where Y is a set of edges, and the latent variable $\{\Pi, B, \Phi, Z_\rightarrow, Z_\leftarrow\}$ is written in the following form:

$$
\begin{aligned}
&p(Y, \Pi, B, \Phi, Z_\rightarrow, Z_\leftarrow | \alpha, \beta, \gamma, \omega) \\
&= \prod_{s=1}^{N} p(\pi_s | \alpha) \prod_{k=1}^{K} p(b_k | \beta) \prod_{k=1}^{K} p(\phi_k | \gamma) \\
&\times \prod_{(s,d) \in Y} \Big\{ p(v_s | \omega) p(z_{s \rightarrow d} | v_s, \Pi) \\
&\times \ p(z_{d \leftarrow s} | z_{s \rightarrow d}, B) p(v_d | z_{d \leftarrow s}, \Phi) \Big\}. \quad (4)
\end{aligned}
$$

We use collapsed Gibbs sampling algorithm [14], [15] to learn the dependent stochastic blockmodel. Gibbs sampling is an Markov chain Monte Carlo algorithm that simulates approximate posterior distribution. After a sufficient number of steps, it is expected that the states of the variables are distributed according to the desired true posterior distribution. Therefore, the result asymptotically converges to the true distribution, as it directly mirrors the algorithm.

Gibbs sampling algorithms require ergodicity, which is satisfied by ensuring positive probability to get to any state. This is ensured by the positive priors for Dirichlet distribution. The prior values have one more function: proper diversity of latent variables is tuned by their priors, because larger prior results in higher probability of selecting unlikely blocks.

We use collapsed Gibbs sampling, marginalizing out Φ and B, which is summarized in Algorithm 1. $n_{v \rightarrow k}$ and $n_{v \leftarrow k}$ denote the number of times node v belongs to block k, and $m_{k_1 \rightarrow k_2}$ indicates the number of times destination block k_2 is drawn from source node k_1. S and D represent the blocks indicated by $Z_\rightarrow$ and $Z_\leftarrow$.

Algorithm 1 Learning dependent stochastic blockmodel

repeat
  for i=1:E **do**
    (s, d) is i^{th} edge in Y
    for $k = 1 : K_1$ **do**
      $p(k) \leftarrow (n_{s \rightarrow k}^{\neg i} + \alpha_k) \times \frac{m_{k \rightarrow D}^{\neg i} + \beta_D}{\sum_{k_2=1}^{K} m_{k \rightarrow k_2}^{\neg i} + \beta_D}$
    end for
    $S \leftarrow$ draw from $Mult(p)$
    for $k = 1 : K_2$ **do**
      $p(k) \leftarrow (m_{S \rightarrow k}^{\neg i} + \beta_k) \times \frac{n_{d \leftarrow k}^{\neg i} + \gamma_d}{\sum_{v=1}^{N} n_{v \leftarrow k}^{\neg i} + \gamma_d}$
    end for
    $D \leftarrow$ draw from $Mult(p)$
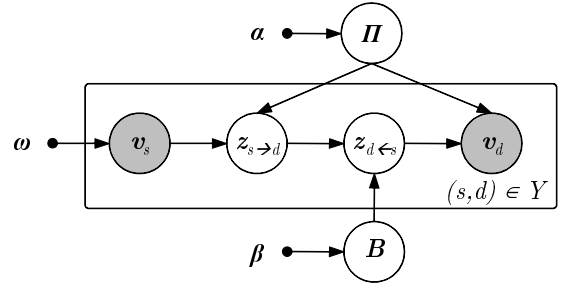 **end for**
until Convergence

Fig. 4. Graphical representation of the dependent stochastic blockmodel for a one-mode network.

A. Special Case: One-Mode Network

In this section, we describe the dependent stochastic blockmodel for a one-mode network. A one-mode network is a network where each node acts as both source and destination. Thus, based on the assumption that each node's interaction as source node and destination node are homogeneous, characteristics of each node are learned from both source and destination node cases, which requires mergence of source and destination cases. A social network between people and a co-authorship network among authors are examples of one-mode networks.

The major difference between the dependent stochastic blockmodel for a two-mode and a one-mode network is the use of latent variables Π and Φ. We assume that the characteristic of each node is homogeneous on both roles as source and destination node, to describe each node consistently. Hence, one latent variable Π in the dependent stochastic blockmodel for a one-mode network replaces both Π and Φ in the dependent stochastic blockmodel for a two-mode network.

Π is a $K \times N$ matrix, and it denotes the relative frequency of affiliation of nodes to blocks, exploiting mixed-membership and the degree of each node. With normalization, each columns of Π represents the probability of each source block conditional on the source node, and each rows of Π represents the probability of each destination node conditional on the destination block. The definition of B remains unchanged.

The link generation procedure in the dependent stochastic blockmodel for a one-mode network remains similar to that of the two-mode network case: we start with the drawing of the source nodes, and end up with the drawing of the destination nodes. The procedure, probability, and algorithm of the dependent stochastic blockmodel for a one-mode network requires mere changes. The dependent stochastic blockmodel for a one-mode network assumes the following generative process (Figure 4):

- Choose $\Pi \sim \text{Dir}(\alpha)$
- Choose $B \sim \text{Dir}(\beta)$
- For each edge (s, d) $\in Y$
  1) Choose a source node indicator,
     $v_s \sim \text{Mult}(\omega)$
  2) Choose a membership indicator for the source

node,
$$\boldsymbol{z}_{s \to d} \sim \text{Mult}(\boldsymbol{\Pi}\boldsymbol{v}_s)$$
3) Choose a membership indicator for the destination node,
$$\boldsymbol{z}_{d \leftarrow s} \sim \text{Mult}(\boldsymbol{z}_{s \to d}^T \boldsymbol{B})$$
4) Choose a destination node indicator,
$$\boldsymbol{v}_d \sim \text{Mult}(\boldsymbol{z}_{d \leftarrow s}^T \boldsymbol{\Pi})$$

The joint probability of the model is

$$
\begin{aligned}
& p(\boldsymbol{Y}, \boldsymbol{\Pi}, \boldsymbol{B}, \boldsymbol{Z}_\to, \boldsymbol{Z}_\leftarrow | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}) \\
= \quad & p(\boldsymbol{\Pi}|\boldsymbol{\alpha}) \prod_{k=1}^{K} p(\boldsymbol{b}_k|\boldsymbol{\beta}) \\
\times \quad & \prod_{(s,d) \in Y} \Big\{ p(\boldsymbol{v}_s|\boldsymbol{\omega}) p(\boldsymbol{z}_{s \to d}|\boldsymbol{v}_s, \boldsymbol{\Pi}) \\
\times \quad & p(\boldsymbol{z}_{d \leftarrow s}|\boldsymbol{z}_{s \to d}, \boldsymbol{B}) p(\boldsymbol{v}_d|\boldsymbol{z}_{d \leftarrow s}, \boldsymbol{\Pi}) \Big\}. \quad (5)
\end{aligned}
$$

B. Side Issues

In this section, we deal with three side issues arising from the dependent stochastic blockmodel. The issues are 1) application of the dependent stochastic blockmodel to undirected graphs, 2) possibility of generating duplicate nodes, and 3) learning from absent edges.

The first issue is how to use dependent stochastic blockmodels for the undirected networks. The dependent stochastic blockmodel was developed for directed networks: to describe dependency between blocks. However, we can easily adopt this model to undirected networks. This is because an undirected link is also defined as a bilateral link. By replacing each undirected edge $\{s, d\}$ with two directed edges $\{(s, d), (d, s)\}$, an undirected network is changed into a directed network to which the dependent stochastic blockmodel can be applied.

The second issue is the fact that we may encounter duplicate links when the dependent stochastic blockmodel generates a network. As each edge independently selects its source and destination nodes, there is the possibility of duplicate links. However, the probability of such a duplicate link (s, d) is negligible. By rough calculation, the expected probability of drawing a source node from $\boldsymbol{\omega}$ is about $1/N_1$, and that of drawing a destination node from $\boldsymbol{z}_{d \leftarrow s}\boldsymbol{\Pi}$ is also about $1/N_2$. The number of edges (E) of a sparse matrix is expected to increase by about the order of the number of nodes, thus the expected number of duplicate links in the scale of $N \times 1/N^2 = 1/N$, which linearly decreases as the number of nodes increases. For large-scale networks, the probability of such a situation occurring is low enough to ignore [10], [16].

IV. EXPERIMENTS

In this section, we present the experiments conducted on three sets of data to verify the efficacy of the dependent stochastic blockmodel for link prediction and node clustering. In Section IV-A, the model is tested to detect blocks. An application to the Twitter network, described in Sectin IV-B, tests the model on prediction links.

A. Books about US politics

In this experiment, our objective was to detect the political preferences expressed in the political books sold online by the bookseller Amazon (Amazon.com) around the time of the 2004 presidential election in the United States[17]. Each book was in one of three categories: "conservative," "liberal," or "neutral." Accordingly, we grouped the books into three ("conservative," "liberal," or "neutral") or two ("conservative" or "liberal") categories. To compare the results of the dependent stochastic blockmodel (DSB) with existing blockmodels, we also ran the degree-corrected stochastic blockmodel (DCB) and the mixed-membership stochastic blockmodel (MMSB). Furthermore, we compared the latent Dirichlet allocation with threshold noise filtering after two-step labeling (filter-2step LDA) which improved accuracy of topic identification over standard LDA [18].

The number of blocks (K) was set at 2 or 3. The parameter values were $\alpha = 1, \beta = 2$. For MMSB, we set ρ as $1 - 100E/N^2$. For filter-2step LDA, we considered nodes with more than 10 degree as popular books, and filtered out less than 0.05 probability as noise.

The expected blocks and topics that resulted from these models after 200 iterations were compared with the truth blocks to verify the reliability of each model. Though the links in this network are undirected, which is equivalent to two directed links that share end nodes with a different direction, DSB and LDA learned each edge as a directed edge to avoid DSB having an advantage.

True positive (TP) and false positve (FP) values are used to check the accuracy of each model. For example, true positive value of conservative category is the number of books where both expected and true categories are conservative, while false positive value of conservative category is the number of books where the expected category is conservative, but the true category is not.

DSB outperformed the other models as shown in Table I. In the K=3 cases, DSB resulted in a total TP of 88. DCB and filter-2step LDA resulted in lower values of 85 and 83. However, MMSB grouped 82 out of 105 nodes into one block, resulting in a total TP of 42. This is due to the degree divergence of the nodes, which is not modeled by MMSB.

When we learned the network with K=2, DSB and LDA still clustered the network well with total TP of 87, by allowing neutral books to have a similar tendency of belonging to the conservative and liberal category. On the other hand, there was a significant difference between DSB and DCB. DCB became confused by the neutral books that have links from and to both conservative and liberal books, yielding a result that resembles a random result.

Furthermore, as DSB does not select the block of each node deterministically and instead learns its tendency as probability, it is possible to find neutral books using DSB with K=2, by grouping the nodes with a tendency lower than 80% to either block as neutral. The scheme results were slightly less accurate than those of DSB for K=3, but were better than those of DCB and LDA for K=3. Therefore, we

can conclude that DSB describes the degree divergence and the vague difference between blocks well, even with a non-optimal K value.

TABLE I

Model	K		Conservative	Liberal	Neutral
Truth	3		49	43	13
1-mode	3	TP	41	38	9
DSB		FP	3	2	12
	2*	TP	45	38	5
		FP	8	1	8
	2	TP	48	39	-
		FP	14	4	-
DCB	3	TP	38	38	9
		FP	2	3	15
	2	TP	22	25	-
		FP	25	33	-
MMSB	3	TP	36	5	1
		FP	46	9	8
	2	TP	40	8	-
		FP	46	11	-
filter-2step	3	TP	39	35	9
LDA		FP	2	3	17
	2	TP	44	43	-
		FP	13	5	-

B. Twitter Networks

Let us now look at the problem of predicting unobserved links in a Twitter network. We crawled 1,500 twitter users in July of 2012, using breadth-first-search on the following lists. The crawling was restricted to the users who were following less than 300 people, in order to except commercial accounts. The data were refined by removing the users who were following no one in this network. The resulting Twitter network consisted of 1,385 users with 21,114 directed following links. The links were randomly divided into 90% learning data (19,003 links), and 10% test data (2,111 links).

We trained four network models and evaluated their prediction accuracy. The four models were 1) dependent stochastic blockmodel for one-mode network (1-mode, DSB), 2) dependent stochastic blockmodel for two-mode (2-mode, DSB), 3) degree-corrected stochastic blockmodel (DCB), 4) mixed-membership stochastic blockmodel (MMSB), 5) Latent Dirichlet allocation with threshold noise filtering after two-step labeling (filter-2step, LDA) [18], and 6) preferential attachment (PA) [11]. The models ran until 50 iterations after convergence with both α (if it existed) and γ as 0.1, β for the dependent stochastic blockmodel as 30, and for mixed-membership stochastic blockmodel as $\{30 \times E/N^2, 30 \times (1 - E/N^2)\}$. The users were grouped into K = 10 blocks. For the mixed-membership stochastic blockmodel, we put $\rho = 1 - 100E/N^2$ to compensate for sparsity. For the filter-2step LDA, we considered nodes with more than 30 degree as popular users, and filtered out less than 0.05 probability as noise. We included preferential attachment to check the predictive power of the pure degree information, where as the mixed-membership stochastic blockmodel shows the predictive power of the blockmodel with mixed-membership.

Each model chose the most probable 4,222 candidate links (two times the number in the test data), and we compared those links with the links in test data. The links that appeared in the training data were excluded from candidate links. For the models in undirected form, each candidate link was chosen up to two times minus the number of times it appeared in the training data. Fig. 5 illustrates how the network learned for each model, with the maximum and average number of edges predicted. Average number is calculated for 50 iterations after convergence.

The dependent stochastic blockmodel for one-mode network outperformed the other models by predicted 387 links in average. The dependent stochastic blockmodel for two-mode network predicted 313 links in average, as it does not use the assumption that each node's property as source and destination node are consistent. Degree-corrected blockmodel and fiter-2step latent Dirichlet distribution predicted less number of links, as they cannot simulate mixed-membership or block structure.

The results for the mixed-membership stochastic blockmodel and the preferential attachment were similar. The maximum number of predicted links was higher for the mixed-membership stochastic blockmodel, but the preferential attachment was better on average. This shows that the pure mixed-membership blockmodel and degree heterogeneity has similar predictive power in this network.

What is interesting is that the number of links predicted by the dependent stochastic blockmodel exceeded even the sum of the number of links predicted by the mixed-membership stochastic blockmodel and the preferential attachment. This shows that the mixture of mixed-membership block-structure and degree heterogeneity resulted in a synergy that captured the potential links that were not predicted by either concepts.
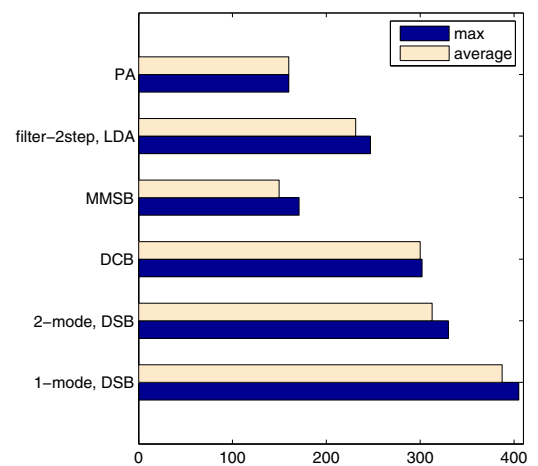

Fig. 5. Number of matched links for the models over multiple iterations. Links are matched between predicted links and test data. Average is caculated on the 50 iterations after convergence.

To investigation further, we generated networks with 19003 links from the models trained in the previous experiment. Figure 6 illustrates the 50 highest in-degrees of real network and four networks generated with one-mode assumption. The values are generally lower than those in the trained network, because in the Twitter network, the in-degree is more diverse than the out-degree. PA is a model that concentrates on degrees, ignoring all other informations. Thus, its degrees are closest to real degrees. On the other hand, DCB, which has degree terms, makes a lower level of in-degrees than PA. This is due to the blockmodel's hypothesis that each node is locked in a block, and thus hardly makes any links to the blocks with which the belonging block rarely links. This limitation forbids modeling hubs, viz., the nodes that have a high degree and connect blocks.

DSB shows a level of in-degrees that is similar to that of PA. This means that the nodes in DSB are not limited by the block structure, because it allows each node to be associated with multiple blocks. Furthermore, the DSB network generates two to five nodes that have similar in-degrees, making the graph step-shaped. As only a real graph and DSB graph show a step-shaped in-degree graph, we can conclude that our model illustrates a real network better than the compared models.
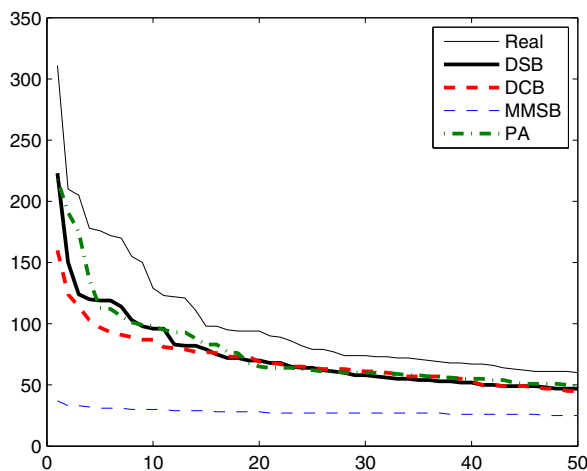

Fig. 6. Number of matched links for four models over multiple iterations. Links are matched between predicted links and test data.

V. CONCLUSIONS

In this paper, we have presented a dependent stochastic blockmodel where only the presence of links are used to learn a model, exploiting the dependency between source and destination nodes. Our model allowed for mixed membership and the degrees of nodes. Experiments on the political books network and Twitter social network demonstrated improvements over existing stochastic blockmodels.

REFERENCES

[1] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, pp. 98–101, 2008.
[2] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.
[3] M. E. J. Newman, "Communities, modules and large-scale structure in networks," *Nature Physics*, vol. 8, no. 1, pp. 25–31, 2011.
[4] H. C. White, S. A. Boorman, and R. L. Breiger, "Social structure from multiple networks. I. Blockmodel of roles and positions," *The American Journal of Sociology*, vol. 81, no. 4, pp. 730–780, 1976.
[5] P. W. Holland, K. Laskey, and S. Leinhardt, "Stochastic blokcmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
[6] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 1077–1087, 2001.
[7] A. Dasgupta, J. E. Hopcroft, and F. McSherry, "Spectral analysis of random graphs with skewed degree distribution," in *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*, 2004, pp. 602–610.
[8] J. Reichardt and D. R. White, "Role models for complex networks," *European Physical Journal B*, vol. 60, no. 2, pp. 217–224, 2007.
[9] M. Mørup, M. N. Schmidt, and L. K. Hanse, "Infinite multiple membership relational modeling for complex network," in *NIPS Workshop on Networks Across Disciplines: Theory and Applications*, 2010.
[10] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community stucture in networks," *Physical Review E*, vol. 83, 2011.
[11] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[12] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, vol. 9, pp. 1981–2014, 2008.
[13] Q. Ho, A. P. Parikh, L. Song, and E. P. Xing, "Multiscale community blockmodel for network exploration," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, 2011.
[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.
[15] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.
[16] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahrmani, "Kronecker graphs: An approach to modeling networks," *Journal of Machine Learning Research*, vol. 11, pp. 985–1042, 2010.
[17] V. Krebs, "Books about US politics," unpublished, http://www.orgnet.com/.
[18] Y. Cha and J. Cho, "Social network analysis using topic models," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Portland, Oregon, USA, 2012.