An End-to-End Community Detection Model: Integrating LDA into Markov Random Field via Factor Graph

Dongxiao He[1], **Wenze Song**[1], **Di Jin**[1,*], **Zhiyong Feng**[1], **Yuxiao Huang**[2]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
[2]Data Science, George Washington University, Washington, D.C., United States
{hedongxiao, swz, jindi, zyfeng}@tju.edu.cn, yuxiaohuang@gwu.edu

Abstract

Markov Random Field (MRF) has been successfully used in community detection recently. However, existing MRF methods only utilize the network topology while ignore the semantic attributes. A straightforward way to combine the two types of information is that, one can first use a topic clustering model (e.g. LDA) to derive group membership of nodes by using the semantic attributes, then take this result as a prior to define the MRF model. In this way, however, the parameters of the two models cannot be adjusted by each other, preventing it from really realizing the complementation of the advantages of the two. This paper integrates LDA into MRF to form an end-to-end learning system where their parameters can be trained jointly. However, LDA is a directed graphic model whereas MRF is undirected, making their integration a challenge. To handle this problem, we first transform LDA and MRF into a unified factor graph framework, allowing sharing the parameters of the two models. We then derive an efficient belief propagation algorithm to train their parameters simultaneously, enabling our approach to take advantage of the strength of both LDA and MRF. Empirical results show that our approach compares favorably with the state-of-the-art methods.

1 Introduction

Networks such as social and biological networks often contain abundant topological and attribute information. Detecting communities in such networks can help people understand the organization structures and function modules underlying the networks.

Many community detection methods using different theories and techniques have been proposed (see a nice review in [Fortunato and Hric, 2016]). They include hierarchical clustering [Girvan and Newman, 2002], modularity-based methods [Newman and Girvan, 2004], heuristic methods [Ruan *et al.*, 2013], spectral optimization [Chen and Li, 2010], and statistical modeling [He *et al.*, 2015]. Among these methods, statistical modeling has been actively studied in community detection due to the solid theories and good performance. The method was first used on networks with topology structure alone [Brian and Newman, 2011], and later extended to networks with both topological and attribute information [Tao *et al.*, 2019].

Existing community detection methods mainly belong to the directed graphical model. That is, they typically formulize the generative process of networked data as a sequence of rigorous probability distributions. In contrast, Markov Random Field (MRF), which is a type of undirected graphical model, has also been applied to community detection very recently [He *et al.*, 2018; Jin *et al.*, 2019]. They obtained satisfactory results because, 1) the field structure of MRF more naturally describes the neighborhood information in networks, and 2) MRF is more flexible than directed graphical model since the probabilistic constrains for MRF are often relaxed when defining energy functions.

However, existing MRF methods focus on the network topology alone while ignore the semantic attributes of nodes (which are also important to community detection). A straightforward way to combine these two sources of information (also as discussed in the above MRF works) is that, one can first use a topic clustering model (e.g. LDA [Blei *et al.*, 2003]) to derive the group membership of nodes by using semantic attributes, and then take this result as a prior to define the MRF model (which utilizes topological information).

However, in this two-stage approach, the modeling of network topology in MRF has no effect on the clustering of topics in LDA. Moreover, the influence of LDA's result on MRF is largely limited since it is fixed after training LDA and only taken as a prior of MRF. That is, the parameters of these two models (LDA and MRF) cannot be adjusted by each other. This prevents the two-stage approach from really realizing the complementation of the advantages of these two models in dealing with network structure and node semantics.

To address this problem, one can integrate LDA into MRF to form an end-to-end model, where the parameters of these two sub-models can be learned jointly. However, LDA is a directed graphical model (which is modeled by the conditional dependence between probability distributions) whereas MRF is an undirected graphical model (which is modeled in the form of energy functions). The different nature of the two mod-

*Corresponding author.

els make their integration challenging for at least two reasons. First, it is difficult to derive the parameter sharing mechanism between directed and undirected models, which is necessary to bridge the two models. Second, the training mechanism of directed and undirected models are essentially different, making it difficult to propagate information between parameters of the two models.

In this paper we integrate the topic model LDA [Blei *et al.*, 2003] into a network-specific MRF [He *et al.*, 2018] to form an end-to-end learning system, named attrMRF, for community detection in attributed networks. To address the first challenge (parameter sharing mechanism between directed and undirected models), we first unify the likelihoods of LDA and MRF using Gibbs distribution, and then transform them into a unified factor graph framework based on the new likelihoods. To handle the second challenge (training mechanism of directed and undirected models), we propose an efficient belief propagation algorithm under the unified factor graph, to achieve an end-to-end learning of the two models (such that their parameters can be trained jointly).

Our main contributions are as follows.

1. Existing MRF-based community detection methods consider the network topology alone while ignore the semantic attributes of nodes (which are also important to community detection). This is the first time to consider these two sources of data together under the MRF framework in an end-to-end way.
2. We integrate LDA (which is good at modeling the semantic attributes of nodes and a directed graphical model) into MRF (which is good at describing the network structures and an undirected one) to form a unified model. The challenge in parameter sharing and joint training are handled by unifying the two models using a factor graph and propagating belief on the factor graph.
3. Empirical results on 6 real-world datasets show that the proposed approach usually compares favorably with the state-of-the-art methods.

2 Preliminaries

Here we introduce the notations used and the Markov Random Field, which is the base of the proposed approach.

2.1 Notations

Let $G = (A, W)$ be an attribute network with N nodes and e edges. Here, A is a $N \times N$ adjacent matrix, where entry a_{ij} is 1 if the i-th and j-th nodes are connected in G and 0 otherwise. Matrix W, on the other hand, is a $N \times M$ attribute (content) matrix (where M is the number of attributes in each node). We use $X = \{x_1, \ldots, x_N\}$ to represent the nodes in the network and $C = \{c_1, \ldots, c_N\}$ to represent a community partition of network G, where c_i denotes the community node x_i belongs to.

2.2 Markov Random Field

Markov Random Field (MRF), which is an undirected graphical model, has been widely used in many areas, such as image segmentation and network analysis [Blake *et al.*, 2011]. The energy function of a general MRF model is often defined as

$$E(C; A, W) = \sum\nolimits_i \varphi_i + \sum\nolimits_{i \neq j} \theta_{ij}.$$

Here φ_i is the unary potential of node x_i (e.g. a pixel in the image or a node in the network), measuring the cost for the difference between the source priori and value of x_i; θ_{ij} is the pairwise potential of nodes x_i and x_j (defined on the neighborhood systems of the data), describing the costs across all the possible combinations of values of the nodes.

3 The Method

We will first provide an overview of the method, then propose the unified model, and last introduce the inference algorithm.

3.1 Overview

The proposed model, attrMRF, consists of two parts, the LDA layer and MRF layer. The LDA layer uses the attribute information (as the unary potentials) to find communities by extracting features on the global level. The MRF layer, on the other hand, uses the topology information (as the pairwise potentials) to find communities smoothly in neighbor systems. The key of attrMRF is integrating LDA (a directed graphical model) into MRF (an undirected one) by formulizing them using a unified factor graph framework in an end-to-end way (such that their parameters can be shared and trained jointly).

To build the unified factor graph, we first unify the likelihoods of LDA and MRF, making them both suitable for factor graph. We then combine them into a new likelihood function via Gibbs distribution. Last, we formulize them into a unified factor graph framework using the new likelihoods.

To make attrMRF a real end-to-end learning process, we design a set of message passing rules under the framework of factor graph, so that the parameters of LDA and MRF can be adjusted by each other. To be specific, in the process of message passing, every node in LDA receives messages from its neighbors (in LDA) to obtain a rough solution. The nodes then send this rough solution as a message to MRF. Consequently, every node in MRF uses this message as unary potentials, and collects messages from its neighbors (in MRF) to refine this solution. Then the nodes send this solution back to LDA to start the next iteration. The above message passing (starting from LDA, passing to MRF, then passing back to LDA) repeats until the model has converged (or reaching the maximum number of iterations). In this way, we are able to integrate and train these two types of models jointly to develop a true end-to-end method for community detection.

3.2 Building the Unified Model

In order to integrate LDA (a directed model) into MRF (an undirected one), we need to transform them into a unified form. As mentioned previously, in this paper the transformation is done using factor graph due to the fact that it is a general and flexible probabilistic graphical model. To be specific, factor graph can not only describe the conditional dependence of probability distribution, but also describe the constraint between variables by using energy functions. More importantly, both directed and undirected graphical models can be converted into a factor graph without losing their inherent characteristics [Yedidia *et al.*, 2003]. If we want to build a factor graph, we need to first get the joint probability distribution of the model and then decompose it into the products of a set of factor nodes. We will explain how this is done in the rest of this section.

As mentioned previously, the proposed unified model, attrMRF, integrates LDA into MRF. For MRF [He *et al.*, 2018], the pairwise potential between nodes x_i and x_j is defined as

$$\theta_{ij}(c_i, c_j; a_{ij}) = -(-1)^{\delta(c_i,c_j)}(\frac{d_i d_j}{2e} - a_{ij}). \tag{1}$$

Here d_i is the degree of node x_i, and $\delta(c_i, c_j)$ is 1 if c_i=c_j (i.e. x_i and x_j in the same community) and 0 otherwise. Since the MRF in attrMRF uses pairwise potentials alone, the global energy potential is thus the sum of all pairwise potentials:

$$\sum\nolimits_{i \neq j} \theta_{ij}(c_i, c_j; a_{ij}). \tag{2}$$

For the LDA in attrMRF, on the other hand, the joint probability distribution is

$$P(Z, W|\alpha, \beta) \propto \prod_{n=1}^{N}\prod_{k=1}^{K} \frac{\Gamma(\sum_{m=1}^{M} w_{m,n} z_{m,n}^k + \alpha)}{\Gamma[\sum_{k=1}^{K}(\sum_{m=1}^{M} w_{m,n} z_{m,n}^k + \alpha)]} \times \prod_{m=1}^{M}\prod_{k=1}^{K} \frac{\Gamma(\sum_{n=1}^{N} w_{m,n} z_{m,n}^k + \beta)}{\Gamma[\sum_{m=1}^{M}(\sum_{n=1}^{N} w_{m,n} z_{m,n}^k + \beta)]},$$

where $w_{m,n}$ says that word m is in document n and $z_{m,n}^k$ says that word m is in topic k of document n. Particularly, m is the semantic attribute of the n-th node in the model.

In order to represent the probability distribution as the product of a set of factor nodes, we rewrite the above joint probability distribution of LDA as

$$P(Z, W|\alpha, \beta) \propto \prod_{n=1}^{N} f_{\theta_n} \prod_{m=1}^{M} f_{\phi_m} \quad \text{where}$$

$$f_{\theta_n} = \prod_{k=1}^{K} \frac{\Gamma(\sum_{m=1}^{M} w_{m,n} z_{m,n}^k + \alpha)}{\Gamma[\sum_{k=1}^{K}(\sum_{m=1}^{M} w_{m,n} z_{m,n}^k + \alpha)]} \quad \text{and}$$

$$f_{\phi_m} = \prod_{k=1}^{K} \frac{\Gamma(\sum_{n=1}^{N} w_{m,n} z_{m,n}^k + \beta)}{\Gamma[\sum_{m=1}^{M}(\sum_{n=1}^{N} w_{m,n} z_{m,n}^k + \beta)]}.$$

Based on the definition of conditional probability, we have

$$P(Z|W, \alpha, \beta) \propto \prod_{n=1}^{N} f_{\theta_n} \prod_{m=1}^{M} f_{\phi_m}. \tag{3}$$

In order to combine (2) (the energy function of MRF) and (3) (the joint probability distribution of LDA), we convert $P(Z|W, \alpha, \beta)$ into the form of energy function according to Gibbs distribution:

$$E(Z; W, \alpha, \beta) = -\sum_{n=1}^{N} \frac{1}{\beta_1} \ln f_{\theta_n} - \sum_{m=1}^{M} \frac{1}{\beta_1} \ln f_{\phi_m}, \tag{4}$$

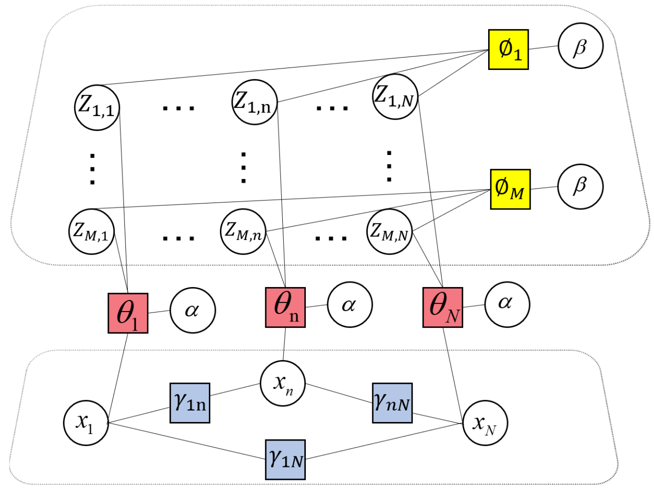

Figure 1: The two-layer structure factor graph of attrMRF. The bottom layer is used to model MRF and describe the structural relationship of the networks, where x_i represents the nodes in the networks. There is a factor node (blue square) between each pair of neighboring nodes and is used to describe the constraint relationship of neighbor systems. The top layer is used to model LDA, where $z_{m,n}$ represents the semantic attribute m of the n-th node. The factor nodes (yellow squares) are used to denote the constraint relationship between different nodes belonging to the same semantics (each row). Last, the factor nodes between the two layers (red squares) are used to represent the constraint relationship between different semantic attributes belonging to the same node (each column).

where β_1 is a temperature coefficient.

Then the global energy function of the unified model can be written as the sum of (2) and (4):

$$E(Z, C; A, W, \alpha, \beta) = \sum\nolimits_{i \neq j} \theta_{ij}(c_i, c_j; a_{ij}) - \sum_{n=1}^{N} \frac{1}{\beta_1} \ln f_{\theta_n} - \sum_{m=1}^{M} \frac{1}{\beta_1} \ln f_{\phi_m}. \tag{5}$$

Last we can use Gibbs distribution to convert the global energy function in (5) back into the form of probability:

$$P(Z, C|A, W, \alpha, \beta) = \frac{1}{\mathcal{Z}} \prod_{n=1}^{N} f_{\theta_n} \prod_{m=1}^{M} f_{\phi_m} \prod_{i \neq j} f_{\gamma_{ij}}. \tag{6}$$

Here $\mathcal{Z}$ is a normalization term, f_{θ_n} and f_{ϕ_m} are defined in (3), and $f_{\gamma_{ij}}$ is the pairwise potential of nodes x_i and x_j.

It is worth noting that (6) is the objective function of the unified model, attrMRF, which can be represented as a factor graph, shown in Figure 1. Here x_n (where $n \in [1, N]$) is a node in MRF, $z_{m,n}$ (where $m \in [1, M]$) the semantic attribute m of node x_n, while α and β the hyperparameters of LDA.

3.3 Inferring the Unified Model

With the objective function in (6), the community partition C on N nodes can be estimated as the joint maximum of the posteriori configuration:

$$\hat{C} = \arg\max_{C} P(Z, C|A, W, \alpha, \beta). \tag{7}$$

The equation in (7) says that the estimated community configuration is the one that corresponds to the largest joint probability. Since Loopy Belief Propagation (LBP) [Chorowski *et al.*, 2014; Rosenberg, 2007] allows identifying a configuration (of variables) that contributes to the largest joint probability from factor graph (containing cycles), it is also used here to find the estimated community configuration. Another reason for using LBP is that, the message passing in the method allows the parameters of the two models in attrMRF (LDA and MRF) to influence each other.

Specifically, there are two kinds of messages when using LBP on factor graph. The message from a variable node to a neighboring factor node, and the message from a factor node to a neighboring variable node (for simplicity, in the rest of this section we use VNs to denote variable nodes and FNs to represent factor nodes). These two kinds of messages can be calculated iteratively in our model:

1. The message from VN x_i to neighboring FN f_j:

$$\mu_{x_i \to f_j}(x_i) = \prod_{f_k \in ne(x_m) \setminus f_j} \mu_{f_k \to x_i}(x_i).$$

   This says that message from x_i to f_j is the product of the messages from neighboring FNs of x_i (except for f_j).

2. The message from FN f_j to VN x_i:

$$\mu_{f_j \to x_i}(x_i) = \max_{x_k} \left(f_j \prod_{x_k \in ne(f_j) \setminus x_i} \mu_{x_k \to f_j}(x_k) \right).$$

   This says that the message from f_j to x_i is the maximum of the product of f_j and the product of the messages from neighboring VNs of f_j (except for x_i).

Here are three main steps for passing messages in attrMRF.

1. Initialize each VN with a probability distribution, representing the probabilities of the node belonging to the communities. The probability distributions are then used as the messages from VNs to FNs. Since we have the initial communities allocation of all nodes, the pairwise potentials can then be calculated as

$$f_{\gamma_{ij}} = \exp\left[(-1)^{\delta(c_i, c_j)} \beta_1 \left(\frac{d_i d_j}{2e} - a_{ij}\right)\right].$$

Here d_i is the degree of node x_i, β_1 is a temperature coefficient, and $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise. Note that the pairwise potentials are the same as the factor functions of the FNs in MRF layer.

The value of FNs in LDA layer can be calculated as

$$f_{\theta_d} = \frac{1}{\sum_{c_i} [\mu_{-m,n}(c_i) + \alpha]}, f_{\phi_m} = \frac{1}{\sum_m [\mu_{m,-n}(c_i) + \beta]}.$$

Here f_{θ_d} normalizes incoming messages by the total number of messages for all topics (i.e. communities) associated with node x_n (to make the outgoing messages comparable across nodes). On the other hand, f_{ϕ_m} normalizes incoming messages by the total number of messages for all attributes (to make the outgoing messages comparable across attributes).

2. Update messages from FNs to VNs. Once all messages from VNs and the factor functions are obtained, the messages form FNs to VNs could be updated. In practice, however, the product of multiple incoming messages often leads to a result close to zero [Zeng and Liu, 2008]. To avoid arithmetic underflow, we approximate the product operation by the sum operation of incoming messages (since when the product value increases, the sum value also increases [Zeng *et al.*, 2013]).

Specifically, there are three types of FNs: f_{θ_i}, f_{ϕ_w}, and γ_{ij}. First, the messages from f_{θ_i} to VNs are updated as

$$\mu_{\theta_i \to z_{m,i}} = \frac{\mu_{-m,i}(c_i) + \mu_{x_i \to \theta_i}(c_i) + \alpha}{\sum_{c_i} [\mu_{-m,i}(c_i) + \mu_{m_i \to \theta_i}(c_i) + \alpha]}, \quad (8)$$

where $\mu_{-m,i}(c_i)$ represents the sum of all messages from VNs connected to θ_n (except for $z_{m,i}$). Particularly, the message from f_{θ_i} to VN x_i is updated as

$$\mu_{\theta_i \to x_i} = \frac{\mu_{\cdot,i}(c_i) + \alpha}{\sum_{c_i} [\mu_{\cdot,i}(c_i) + \alpha]}, \quad (9)$$

where $\mu_{\cdot,i}(c_i)$ represents the sum of all messages from VNs connected to θ_n (except for x_i). It is worth noting that while (8) and (9) look different, they actually work in the same way.

Next, the messages from f_{ϕ_w} (the second kind of FNs) to VNs are updated as

$$\mu_{\phi_m \to z_{m,i}} = \frac{\mu_{m,-i}(c_i) + \beta}{\sum_{c_i} [\mu_{m,-i}(c_i) + \beta]}, \quad (10)$$

where $\mu_{m,-i}(c_i)$ represents the sum of all messages from VNs connected to ϕ_m (except for $z_{m,i}$).

Last, the messages from the third kind, γ_{ik}, are updated as

$$\mu_{\gamma_{ik} \to x_i}(c_i) = \max_{c_k} \left(\exp((-1)^{\delta(c_i, c_k)} \beta_1 (\tfrac{d_i d_k}{2e} - a_{ik})) + \mu_{x_k \to \gamma_{ik}}(c_k) \right), \quad (11)$$

where $\mu_{x_k \to \gamma_{ik}}(c_k)$ is the message from VN x_k to FN γ_{ik}.

3. Update messages from VNs to FNs. Once all messages from FNs to VNs have been updated, the messages from VN

Algorithm 1 The Inference Process for attrMRF

Input: The topology and attribute matrices of the network, A, W. The hyperparameters of attrMRF, α, β, and β_1.
Output: Community partition C

```
Initialize the messages from VNs to FNs.
do t = 0 ; conv = ε + 10.
while conv > ε and t < T do
    for every FN in attrMRF do
        Update messages from FN to VNs via (8) to (11).
    end for
    for every VN in in attrMRF do
        Update messages from VN to FNs via (12) to (15).
    end for
    do t = t + 1 ; conv = |μ_new − μ_old|.
end while
Compute Ĉ from the max-beliefs according to (16).
```

Datasets	n	e	w	c
Cornell	195	283	283	5
Washington	217	366	1,578	5
Cora	2,708	5,278	1,432	7
Citeseer	2,559	3,182	3,698	6
UAI2010	3,061	28,308	4,973	19
Pubmed	19,717	44,338	500	3

Table 1: Dataset descriptions. Here n is the number of nodes, e the number of edges, w the number of attributes, and c the number of communities.

x_i to FN can update again as

$$\mu_{x_i \to \gamma_{ij}}(c_i) = \mu_{\theta_i \to x_i} + \sum_{k \in ne(i) \backslash j} \mu_{\gamma_{ik} \to x_i}(c_i), \tag{12}$$

$$\mu_{x_i \to \theta_i} = \sum_{k \in ne(i)} \mu_{\gamma_{ik} \to x_i}. \tag{13}$$

Here $\mu_{\gamma_{ik} \to x_i}$ represents the message from FN γ_{ik} to VN x_i.

Since there are only two FNs (red and yellow squares in Figure 1) connected to each VN (circles in the figure) in the LDA layer, the message can be updated as

$$\mu_{z_{m,i} \to \theta_i} = \mu_{\phi_m \to z_{m,i}}, \tag{14}$$

$$\mu_{z_{m,i} \to \phi_m} = \mu_{\theta_i \to z_{m,i}}. \tag{15}$$

With the above three steps we can update the messages passing in the unified model. When the algorithm converges, the variable max-belief, $B_i(c_i)$, can be calculated as

$$B_i(c_i) = \mu_{\theta_i \to x_i}(c_i) + \sum_{k \in ne(i)} \mu_{\gamma_{ik} \to x_i}(c_i), \tag{16}$$

which is the sum of all the messages sending to node x_i

The inference process discussed above is summarized in Algorithm 1.

4 Experiments

We first validate why attrMRF (the proposed method) works and then compare it with some state-of-the-art methods.

A summary of the datasets used is shown in Table 1. Since all the methods require the number of communities to be given, we set it the same as that of the ground truth. We used Accuracy (AC) [Liu *et al.*, 2012] and Normalized Mutual Information (NMI) [Danon *et al.*, 2005] as accuracy metrics for performance evaluation.

4.1 Why attrMRF Works

In this section we analyze why attrMRF works from both quantitative and qualitative perspectives.

Quantitative Analysis

Here we compare attrMRF with NetMRF (which is a MRF-based method using network topology alone and the base of the MRF part of our method) and the two-stage approach (where we first use LDA on attributes to derive an initial solution and then use MRF for refinement using topology) in terms of AC and NMI.

Datasets	AC (%)			NMI (%)		
	MRF	two-stage	attrMRF	MRF	two-stage	attrMRF
Cornell	40.00	41.02	**47.69**	11.68	19.43	**21.38**
Washington	39.17	46.54	**66.36**	12.22	21.31	**26.06**
Cora	48.89	51.05	**52.65**	32.97	35.51	**36.08**
Citeseer	32.59	50.21	**52.71**	11.98	25.16	**26.25**
UAI2010	32.86	36.77	**37.24**	27.90	37.32	**38.37**
Pubmed	52.39	50.65	**55.48**	14.87	15.73	**15.99**

Table 2: Comparisons of NetMRF, the two-stage approach (two-stage for short), and attMRF in terms of AC and NMI obtained on 6 networks. The best results are in bold.

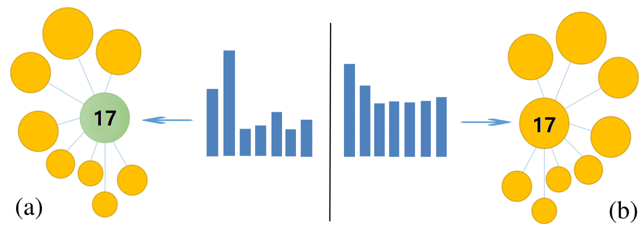

Figure 2: An illustrative example on node id17 in the Cora detest. (a) is the result of the two-stage approach and (b) the result of attrMRF.

As shown in Table 2, attrMRF performs better than the two-stage approach which performs better than NetMRF. To be specific, attrMRF is on average 11.03% and 5.98% more accurate than the two-stage approach and NetMRF in terms of AC. We get similar result in terms of NMI. It validates that the proposed end-to-end model, attrMRF, indeed better integrates the advantages of MRF and LDA, and consequently derives better results.

Qualitative Analysis

We demonstrate how the end-to-end mechanism in attrMRF leads to better results. This can be done by taking a closer look at the results of the two-stage approach and attrMRF. As an example Figure 2 shows node id17 and its neighbors in the Cora dataset, where (a) is the result of the two-stage approach and (b) the result of attrMRF. Specifically, the two-stage approach wrongly finds the community node id17 belongs to (denoted by color green), which is different from the community its neighbors belong to (color orange). This error is due to the fact that LDA and MRF are trained separately in the two-stage approach. Thus, without the correction from MRF during training, the probability for the wrong community produced by LDA is too large and stubborn to be overturned by the following MRF.

While the two-stage approach finds the wrong community for node id17, attrMRF detects the correct one. The reason why attrMRF is able to do so is due to the fact that LDA and MRF are trained jointly in attrMRF. Thus, with the correction from MRF during training, the probability for the wrong community produced by LDA is small enough to be overturned by MRF (as shown in the histogram of Figure 2 (b)). This theoretically explains what we experimentally demonstrated in the quantitative analysis (where attrMRF has the highest AC and NMI values).

Metrics (%)	Datasets	Methods							
		DCSBM	MRF-N	BPLDA	PCLDC	B-LDA	SCI	TLSC	attrMRF
AC	Cornell	39.95	42.05	45.14	30.26	46.15	45.64	**47.69**	**47.69**
	Washington	31.08	44.23	54.83	29.95	39.17	51.15	51.61	**66.36**
	Cora	38.48	42.72	48.00	34.08	25.52	40.62	47.62	**52.65**
	Citeseer	26.57	25.55	39.03	24.85	24.35	27.98	35.74	**52.71**
	UAI2010	2.60	34.83	34.11	28.82	16.04	30.94	29.37	**37.24**
	Pubmed	53.64	47.00	45.04	63.55	49.10	N/A	**61.38**	55.48
NMI	Cornell	9.69	6.23	21.07	7.23	6.81	11.44	13.16	**21.38**
	Washington	9.87	7.68	**31.30**	5.66	3.69	12.37	19.63	26.06
	Cora	17.07	27.27	24.83	17.54	1.41	19.26	33.20	**36.08**
	Citesser	4.13	9.33	16.15	2.99	2.42	4.87	23.16	**26.25**
	UAI2010	31.21	32.92	35.47	26.92	5.70	24.80	20.68	**38.37**
	Pubmed	12.28	7.63	12.53	**26.84**	6.58	N/A	19.63	15.99

Table 3: Comparisons of eight methods on 6 networks in AC and NMI. B-LDA is short for Block-LDA. The best results are in bold.

4.2 Comparison with Existing Methods

We also compared attrMRF with some state-of-the-art community detection methods, which can be divided into three categories. The first includes DCSBM [Brian and Newman, 2011] and MRF-N [Jin *et al.*, 2019], which use network topology alone. The second includes BPLDA [Zeng *et al.*, 2013], which denotes the LDA model optimized by belief propagation. The third includes PCLDC [Yang *et al.*, 2009], Block-LDA [Balasubramanyan and Cohen, 2011], SCI [Wang *et al.*, 2016], and TLSC [Zhang *et al.*, 2018], which use both the network topology and semantic attributes in the networks.

As shown in Table 3, attrMRF performs the best on 5 and 4 out of the 6 networks in terms of AC and NMI, respectively. On the remaining networks where attrMRF does not perform the best, it is still competitive with the best baselines. To be specific, attrMRF is on average 19.85%, 12.62%, 7.66%, 16.77%, 18.63%, 12.06% and 6.45% more accurate than DCSBM, MRF-N, BPLDA, PCLDC, B-LDA, SCI and TLSC in AC; and 13.31%, 12.17%, 3.79%, 12.82%, 22.92%, 15.22% and 6.11% more accurate than these methods in NMI.

It is not surprising to see attrMRF performs better than D-CSBM, MRF-N and BPLDA since it utilizes more sources of information. It is, however, more interesting to see attrMR-F outperforms the other four methods, which also use both topological and attribute information in networks. This superiority is not by chance. Instead it could be because 1) attrMRF ideally utilizes the advantages of MRF and LDA which are more suitable to describe the topological information and semantic attributes respectively, while 2) existing methods (e.g. TLSC) are mainly the extension of topic models, which incorporate network topology into the original directed graphical model to serve as a role of refinement.

5 Related Work and Discussion

Here we discuss two types of the most related works to state the advantages of the proposed model, attrMRF.

5.1 MRF-Based Community Detection Methods

To our best knowledge, only two MRF methods have been proposed recently for community detection. The first is NetMRF [He *et al.*, 2018]. It uses the MRF field structure to characterize the irregular structure of networks, and then defines the energy function to encode the structure and properties of network communities. This is also the base of the MRF part of our attrMRF approach. The second method is proposed in [Jin *et al.*, 2019]. It makes up for the defect of low coupling of network embedding methods by taking the advantages of MRF in characterizing the relational data, leading to a general MRF framework to better find communities.

However, these existing methods only consider the network topology while ignore the semantic attributes of nodes. Though as discussed by their authors, one can easily incorporate node attributes (by first using LDA on node attributes to get an initial solution and then employing MRF to perform the refinement based on network topology), this is, however, not an ideal way to combine these two sources of information. This claim was experimentally demonstrated by the quantitative and qualitative comparison between NetMRF, the two-stage approach, and attrMRF (Table 2 and Figure 2).

5.2 Community Detection Models on Attributed Networks

Recently, many statistical models for finding communities in attributed networks have been proposed. For example, [Balasubramanyan and Cohen, 2011] proposed the Block-LDA model. It uses the stochastic block model to model the links to assist LDA (which can make good use of the semantic attributes of nodes), allowing it to integrate these two sources of information. [Zhang *et al.*, 2018] proposed a two-level semantic community model. It divides the topics into two levels to solve the problem that topics in the generation of contents are often not from a unique topical level.

However, these methods are typically based on directed graphical model. This may be because the directed graphical model (e.g. topic models or Gaussian mixture model) is good at describing the semantic information of nodes, and the integration of network topology to improve topic models under the original directed graphical model framework is more straightforward to be designed. However, MRF (which is a type of undirected graphical model) could be more suitable to model the topological information. Thus methods such as our attrMRF, which combines MRF and topic models, could be a more suitable way to incorporate these two sources of data. This claim was echoed by the comparison between the state-of-the-art methods and attrMRF (Table 3).

6 Conclusion and Discussion

In this work, we proposed the first MRF approach for community detection in attributed networks in an end-to-end way. We first integrate LDA into MRF to form a unified model via factor graph modeling. We then use belief propagation under this new factor graph model to learn parameters of the two sub-models jointly. Empirical results on 6 real-world datasets show that the proposed approach usually compares favorably with other state-of-the-art methods. While the proposed work focuses on disjoint community, it may be readily extended to find overlapping communities by replacing max-sum rule of BP with sum-product rule.

Acknowledgments

This work was supported by the Natural Science Foundation of China (61876128, 61772361), and the National Key R&D Program of China (2017YFB1401201).

References

[Balasubramanyan and Cohen, 2011] Ramnath Balasubramanyan and William W Cohen. Block-lda: Jointly modeling entity-annotated text and entity-entity links. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 450–461. SIAM, 2011.

[Blake *et al.*, 2011] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov random fields for vision and image processing*. 2011.

[Blei *et al.*, 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[Bojchevski and Günnemann, 2018] Aleksandar Bojchevski and Stephan Günnemann. Bayesian robust attributed graph clustering: Joint learning of partial anomalies and group structure. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Brian and Newman, 2011] Karrer Brian and M E J Newman. Stochastic blockmodels and community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 83(2):016107, 2011.

[Chen and Li, 2010] Dongming Chen and Lili Li. Improved structural algorithm for uncovering community structure in networks. In *International Conference on Computer Application & System Modeling*, 2010.

[Chorowski *et al.*, 2014] Jan Chorowski, Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. End-to-end continuous speech recognition using attention-based recurrent nn: first results. *arXiv preprint arXiv:1412.1602*, 2014.

[Danon *et al.*, 2005] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.

[Fortunato and Hric, 2016] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.

[Girvan and Newman, 2002] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[He *et al.*, 2015] Dongxiao He, Dayou Liu, Di Jin, and Weixiong Zhang. A stochastic model for detecting heterogeneous link communities in complex networks. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[He *et al.*, 2018] Dongxiao He, Xinxin You, Zhiyong Feng, Di Jin, Xue Yang, and Weixiong Zhang. A network-specific markov random field approach to community detection. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Jin *et al.*, 2019] Di Jin, Xinxin You, Weihao Li, Dongxiao He, Peng Cui, Francoise Soulie Fogelman, and Tanmoy Chakraborty. Incorporating network embedding into markov random field for better community detection. In *Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.

[Liu *et al.*, 2012] Haifeng Liu, Zhaohui Wu, Xuelong Li, Deng Cai, and Thomas S Huang. Constrained nonnegative matrix factorization for image representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1299–1311, 2012.

[Newman and Girvan, 2004] M E J Newman and . Girvan, M. Finding and evaluating community structure in networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 69(2):026113, 2004.

[Rosenberg, 2007] Julia Hirschberg Andrew Rosenberg. Joint conference on empirical methods in natural language processing and computational natural language learning. *V-Measure: A conditional entropy-based external cluster evaluation measure. Lecture Notes in Artificial Intelligence. Prague: Association for Computational Linguistics*, 2007.

[Ruan *et al.*, 2013] Yiye Ruan, David Fuhry, and Srinivasan Parthasarathy. Efficient community detection in large networks using content and links. In *International Conference on World Wide Web*, 2013.

[Tao *et al.*, 2019] Zhiqiang Tao, Hongfu Liu, Sheng Li, Zhengming Ding, and Yun Fu. Marginalized multiview ensemble clustering. *IEEE transactions on neural networks and learning systems*, 2019.

[Wang *et al.*, 2016] Xiao Wang, Di Jin, Xiaochun Cao, Liang Yang, and Weixiong Zhang. Semantic community identification in large attribute networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[Yang *et al.*, 2009] Tianbao Yang, Rong Jin, Yun Chi, and Shenghuo Zhu. Combining link and content for community detection: a discriminative approach. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 927–936, Paris, France, 2009.

[Yedidia *et al.*, 2003] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.

[Zeng and Liu, 2008] Jia Zeng and Zhi-Qiang Liu. Markov random field-based statistical character structure modeling for handwritten chinese character recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(5):767–780, 2008.

[Zeng *et al.*, 2013] Jia Zeng, William K Cheung, and Jiming Liu. Learning topic models by belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1121–1134, 2013.

[Zhang *et al.*, 2018] Ge Zhang, Di Jin, Jian Gao, Pengfei Jiao, Francoise Fogelman-Soulié, and Xin Huang. Finding communities with hierarchical semantics by distinguishing general and specialized topics. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3648–3654. AAAI Press, 2018.