ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Bayesian non Parametric Inference of Discrete Valued Networks

Laetitia Nouedoui and Pierre Latouche

SAMM EA 4543, Université Paris 1 Panthéon-Sorbonne
90 rue de Tolbiac 75634-Paris Cedex 13, France.

Abstract. We present a non parametric bayesian inference strategy to automatically infer the number of classes during the clustering process of a discrete valued random network. Our methodology is related to the Dirichlet process mixture models and inference is performed using a Blocked Gibbs sampling procedure. Using simulated data, we show that our approach improves over competitive variational inference clustering methods.

1 Introduction

Recent years have been the witness of an increasing interest in random network analysis. Networks are now used in many scientific fields from Biology to social sciences, among others. For example, protein-protein interaction networks describe possible physical interactions between proteins [1] while social networks aim at characterizing relational ties between actors [2, 3]. In order to extract information from these structured data sets, a wide range of methods have been proposed. Some of them focus on discovering communities, i.e. regions with high connectivity between nodes [4, 5]. For instance, the network modularity of [6] aims at detecting communities by looking for dense zones in a graph using a modularity score. An alternative approach is the Latent Position Cluster Model proposed by [7]. This model looks for clusters of vertices depending on their positions in an unobserved euclidean latent space. Other models seek disassortative structures in which vertices are more likely to connect with vertices belonging to other clusters [8].

The Stochastic Block Model (SBM) [2, 9] is a more flexible model. Given a network, it assumes that each vertex belongs to a latent class among K classes and uses a K $\times$ K connectivity matrix to describe the connection probabilities [10]. As shown by [2], it can be used to retrieve both communities and disassortative mixing, but not only. However, SBM only deals with binary relational data. To tackle this issue, [11] introduced recently an extension of SBM to deal with discrete valued edges, called the Poisson Mixture model for graphs (PM). In order to perform inference, a Variational Expectation Maximization (VEM) algorithm is applied on a network for various values of the number K of classes. The Integrated Classification Likelihood (ICL) criterion is then computed for each K and $\hat{K}$ is chosen such that the criterion is maximized.

Recently, [12] proposed the Infinite Relational Model (IRM) as well as a Gibbs sampling procedure. Their approach for relation data is non parametric and allows the number of clusters to be estimated automatically while clustering

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

the observations. [13] proposed a similar approach, however, their methodology is restricted to multinomial relations.

In this paper, we propose a bayesian non parametric method for the clustering of discrete valued networks. First, we recall the PM introduced by [11] and show how it can be described in a Bayesian framework. We then consider a Chinese Restaurant Process (CRP) on the latent structure and perform inference using a Blocked Gibbs sampling procedure. The algorithm allows the number of clusters to be automatically obtained during the estimation process. Finally, we apply our methodology on simulated data and compare our results with a competitive approach.

2 Model: Infinite Poisson mixture model

We consider a random graph $\mathcal{G}$ represented by a N $\times$ N adjacency matrix $\mathbf{X}$. Each entry X_{ij} takes its value in $\mathbb{N}$ and describes the relation between vertex i and vertex j. Note that $X_{ij} = 0$ corresponds to the absence of an edge.

2.1 Poisson mixture

The Poisson Mixture model for graphs (PM) introduced by [11] assumes that the vertices of $\mathcal{G}$ are spread into K classes. Thus, the model associates to each vertex a latent variable Z_i drawn from a multinomial distribution:

$$Z_i \sim \text{Mult}\left(1; \alpha\right), \quad \forall \left\{i \in 1, \cdots N\right\}, \tag{1}$$

where α denotes the vector of class proportions ($\sum_{q=1}^{K} \alpha_q$). Given the classes, edges are then sampled from Poisson distributions $\mathcal{P}(\cdot)$. Thus, if i belongs to class q and j to class l, X_{ij} is assumed to be drawn from:

$$X_{ij} | \{Z_{iq} Z_{jl} = 1\} \sim \mathcal{P}\left(\lambda_{ql}\right), \tag{2}$$

where λ_{ql} represents the expected value of the links between nodes from class q to nodes of class l. According to the PM model, the latent variables $Z_1, \cdots, Z_N$ are iid and given the latent structure, all edges are assumed to be independent.

In the following, we assume that $\mathcal{G}$ is an undirected random graph without self-loops. In other words, for all pairs of nodes i and j, $X_{ij} = X_{ji}$ and $X_{ii} = 0$. However, we emphasize that all results presented here can easily be extended to directed and undirected networks, with or without self-loops.

2.2 Infinite Poisson mixture model for random graph

We consider a Chinese Restaurant Process (CRP) as a non parametric prior for the PM model. CRP describes a sampling procedure to generate classes. Thus, starting with a single class with only one observation, observations are added as well as some classes until all observations in the data set are classified. Under CRP, each class attracts new data points depending on its current size (see [12] for instance). Thus, given the classes of the m first observations, a new

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

observation is either assigned to an existing class q with probability $\frac{n_q}{m-1+\eta_0}$ or to a new class with probability $\frac{\eta_0}{m-1+\eta_0}$. The number of data points in class q is denoted n_q while η_0 is an hyperparameter that has to be set in practice. This sampling scheme has two main advantages. First, the distribution of the classes is exchangeable, i.e. changing the order to which the observations are assigned to classes does not change the probability of the corresponding partition. Second, since there is a non-zero probability that a new data point creates a new class, CRP is a prior distribution over partitions with various number K of classes.

CRP is related to Dirichlet processes [14] for which a constructive definition, namely the Stick-Breaking Prior (SBP), exists. Thus, for each class q, a parameter β_q is sampled from a beta distribution:

$$\beta_q \sim \text{Beta}(1; \eta_0).$$

The proportion α_q of the class if then set to β_q if $q = 1$, and

$$\alpha_q = \beta_q \prod_{l=1}^{q-1} (1 - \beta_l), \tag{3}$$

otherwise. Thus, the α_q's are exponentially decreasing, so only a limited number of classes will be involved to model data [15]. In fact, the mean number of classes involved is $\mathcal{O}(\eta_0 \log(N))$. This result underlines the influence of the scalar parameter η_0 on the effective number of classes. A discussion about the choice of η_0 is given in the experiment section.

Finally, in order to obtain analytical expressions for the conditional distributions of the Gibbs sampling procedure, we consider a conjugate prior for the mean intensity λ_{ql}. Since $p(X_{ij}|Z_{iq}Z_{jl} = 1, \lambda_{ql}) = \mathcal{P}(X_{ij}; \lambda_{ql})$ is a Poisson distribution, we rely on a Gamma prior:

$$p(\lambda_{ql}) = \text{Gamma}(\lambda_{ql}; a, b).$$

By construction, the Gamma distribution is informative. In order to limit its influence on the posterior distribution, a common choice in the literature is to set the hyperparameters a and b, controlling the scale and rate respectively, to $a = b = 0.1$. With these choices of priors, the PM model becomes:

$$\begin{aligned} \alpha|\eta_0 \sim SBP(\eta_0) \quad & Z_i|\alpha \sim \text{Mult}(1; \alpha), \\ \lambda_{ql}|a, b \sim \text{Gamm}(a, b) \quad & X_{ij}|Z_{iq}, Z_{jl}, \lambda_{ql} \sim \mathcal{P}(\lambda_{ql}). \end{aligned} \tag{4}$$

We call the later model the Infinite Poisson Mixture model for graphs (IPM).

3 Inference

Given the adjacency matrix $\mathbf{X}$ of a network, the goal is to approximate the posterior distribution $\Pr(Z, \alpha, \lambda|X)$. For this purpose, inference can be performed using the Blocked Gibbs sampling algorithm described in [16]. The Stick-Breaking

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

prior is truncated at an arbitrary value T. This approximation allows the model to be represented with a finite number of random variables. As mentioned in [17], T does not need to be very large to get a good approximation. Actually, T corresponds to a maximum number of classes expected a priori. In practice, T can be set using an expert prior knowledge. During the Blocked Gibbs sampling, some classes tend to get empty and after convergence of the Markov chain, the number of non-empty classes is used as an approximation of K^*, the true number of classes in the data. Algorithm 1 presents the pseudo-code of the Blocked Gibbs sampling algorithm.

Algorithm 1 Gibbs Sampling for the Infinite Poisson mixture model

Require: A value for T and an initial state for Z

1: **repeat**

2: For $q \in \{1, \cdots, T\}$, $\alpha_q = \beta_q \prod_{l<q}(1-\beta_l)$ and $\beta_T = 1$,

where

$$\beta_q \sim \text{Beta}(\gamma_{q1}, \gamma_{q2}) \quad \text{and} \quad \gamma_{q1} = 1 + \sum_{i=1}^{N} Z_{iq} \quad \text{and} \quad \gamma_{q2} = \eta_0 + \sum_{j=q+1}^{T} \sum_{i=1}^{N} Z_{ij}.$$

3: For $i \in \{1, \cdots, N\}$, sample Z_i from $\text{Mult}(1; \tau_i)$, where for all $q \in \{1, \cdots, T\}$,

$$\log(\tau_{iq}) = \log(\alpha_q) + \sum_{l} \sum_{j \neq i} Z_{jl}\{X_{ij} \log(\lambda_{ql}) - \lambda_{ql} - \log(X_{ij}!)\}.$$

4: For $q, l \in \{1, \cdots, T\}$, sample λ such that:

$$\lambda_{ql}|Z, \alpha, X \sim \begin{cases} \text{Gamm}\left(\sum_{i \neq j} X_{ij} Ziq Zjl + a, \sum_{i \neq j} Ziq Zjl + b\right) & \text{if } q < l \\ \text{Gamm}\left(\sum_{i < j} X_{ij} Ziq Zjl + a, \sum_{i < j} Ziq Zjl + b\right) & \text{if } q = l \end{cases}$$

5: **until** Convergence of the Markov chain.

6: Delete each empty class, and actualize T.

We emphasize that our method does not require an initialization algorithm such as kmeans, contrary to most other methods of inference for random graphs, as variational inference for instance.

4 Experiments

4.1 Simulation data

In order to compare our results on model selection to those of [11] on simulated data, we reproduced their example of undirected random graphs without self-loops, with parameters such that the vector α of class proportions is unbalanced and the class intensities satisfy: for all classes p and q, $\lambda_{pp} = \lambda'$ and $\lambda_{pq} = \gamma\lambda'$. Thus, networks with $K = 3$ classes were generated with $\alpha = (57\%, 29\%, 14\%)$

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

and $\gamma = 0.5$. λ' was set such that the mean connectivity equals 2. This induces the following connectivity matrix:

$$\Lambda = \begin{pmatrix} 3 & 1.5 & 1.5 \\ 1.5 & 3 & 1.5 \\ 1.5 & 1.5 & 3 \end{pmatrix}.$$

With those parameters, we simulated 100 networks for each value of N in (50, 100, 500, 1000). Then, we applied our Gibbs inference procedure for IPM to infer the number K^* of classes. For this purpose, we set the maximum number of classes expected a priori to 20. In practice, we tested various values of η_0 but we did not observe any effect on the results presented in the following.

We compared our result in terms of model selection to those of [11]. As mentioned previously, [11] estimate the number of classes using the ICL criterion. Results are shown in table 1.

Network size	Model	$\widehat{K}_n = 3$	$\widehat{K}_n = 2$	$\widehat{K}_n = 4$
$N = 50$	IPM	0.59	0.41	0.00
	PM	0.17	0.82	0.01
$N = 100$	IPM	0.96	0.04	0.00
	PM	0.90	0.07	0.03
$N = 500$	IPM	1.00	0.00	0.00
	PM	1.00	0.00	0.00
$N = 1000$	IPM	1.00	0.00	0.00
	PM	1.00	0.00	0.00

Table 1: Rate at which K is selected for various sizes of networks, with the Infinite Poisson mixture model (IPM) and with the Poisson mixture (PM) model.

Our approach clearly outperforms the results of PM in the case of small networks of size 50 and 100. It gives comparable results for larger networks.

5 Conclusion

In this paper, we proposed a bayesian nonparametric inference of discrete valued random networks. Our methodology is based on Stick-Breaking prior, and uses a Blocked Gibbs sampling algorithm. This method allows to automatically infer the number of classes during the estimation process. The method gives promising results on simulated data, since they are comparable to a competing inference method based on variational inference. Further work should be conducted to assess the quality of parameter estimates as well as clustering, and applied to real data.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

References

[1] A.L. Barabási and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[2] K. Nowicki and T.A.B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[3] T.A.B. Snijders and K. Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100, 1997.

[4] MEJ Newman and EA Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569, 2007.

[5] J.M. Hofman and C.H. Wiggins. Bayesian approach to network modularity. *Physical review letters*, 100(25):258701, 2008.

[6] M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[7] M.S. Handcock, A.E. Raftery, and J.M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

[8] E. Estrada and J.A. Rodríguez-Velázquez. Spectral measures of bipartivity in complex networks. *Physical Review E*, 72(4):046105, 2005.

[9] J.J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and computing*, 18(2):173–183, 2008.

[10] O. Frank and F. Harary. Cluster inference by using transitivity indices in empirical graphs. *Journal of the American Statistical Association*, 77(380):835–840, 1982.

[11] M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, 4(2):715–742, 2010.

[12] C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *Proceedings of the national conference on artificial intelligence*, volume 21, page 381. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[13] Z. Xu, V. Tresp, K. Yu, and H.P. Kriegel. Infinite hidden relational models. *arXiv preprint arXiv:1206.6864*, 2012.

[14] Thomas S. Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, 1(2):209 – 230, 1973.

[15] Y.W. Teh. Dirichlet processes: Tutorial and practical course. *Gatsby Computational Neuroscience Unit, University College London*, 2007.

[16] H. Ishwaran and L.F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

[17] M.I. Jordan. Dirichlet processes, chinese restaurant processes and all that. In *Tutorial presentation at the NIPS Conference*, 2005.