Non-parametric Overlapping Community Detection

Nishma Laitonjam and Neil Hurley

Abstract In this paper, we present a non-parametric overlapping community detection method based on the affiliation graph model using an Indian Buffet Process to determine the number of communities. We compare this model with a full stochastic blockmodel using the same prior, as well as two other models, a blockmodel and a community model, that employ non-parametric priors based on a Gamma Process. We ask two questions; firstly, whether community models are sufficient to model the overlapping structure of real networks, without resorting to blockmodels that entail significantly more parameters; secondly, which is the better non-parametric approach of the two analysed? Measuring performance in terms of predicting missing links, we find that all models obtain similar performance, but in general, the Indian Buffet Process prior results in simpler models, with fewer blocks or communities. We argue that, when obtaining the latent structure is the purpose of the analysis, the simpler affiliation graph model, with Indian Buffet Process is preferred.

Keywords Affiliation graph model · Generative model · Indian buffet process · Non-parametric model · Overlapping community detection

1 Introduction

Much work has been carried out in the identification of community structure in social networks. Communities are sub-sets of highly interconnected vertices. Two broad and distinct categories of community identification are studied, namely non-overlapping and overlapping community detection. In the latter case, which is the focus of this paper, vertices are allowed to belong to more than one community and this would seem the best model of real-world social communities. After all,

N. Laitonjam (✉) · N. Hurley
Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
e-mail: nishma.laitonjam@insight-centre.org

N. Hurley
e-mail: neil.hurley@insight-centre.org

people tend to belong to many social groups such as leisure and work groups. A wide range of approaches to overlapping community detection have been taken in the state-of-the-art such as local seed-expansion [10], global optimization [3] and other heuristic approaches, including an extension of the label propagation method to the overlapping case [7]. Statistical generative approaches propose a model through which the graph is generated, by initially selecting the community labels for the nodes, followed by the generation of edges, whose probability of existence depends on the community labels of their end-points. Computational inference techniques are used to determine the model parameters from an observed graph and hence to infer the community labels. In this paper, we focus on such generative methods.

We adopt the following notation. Let $G = (V, E)$ be an undirected graph, such that V is a set of nodes or vertices and $E \subseteq V \times V$ is a set of edges. An edge can be represented as the undirected tuple (v, w), where $v, w \in V$. Let, $C = \{1, \ldots, K\}$ be a set of community labels, where $K > 0$ is the total number of communities. Write $n = |V|$. Assuming some numbering such that the vertices can be written as $\{v_1, \ldots, v_n\}$, we will sometimes refer to a vertex by its index i under this numbering. An unweighted graph may be represented by its binary *adjacency matrix* $A = \{a_{ij}\}$ where $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. We sometimes consider integer-valued positive *weights* associated with an edge (i, j), and we write $X = \{x_{ij}\}$ for the weighted graph with $x_{ij} \in \mathbb{Z}^+$. A *community-assignment* is an association of a set of communities $C_i \subseteq C$, with each $v_i \in V$. It is useful to describe the community assignment as the binary $n \times K$ matrix $Z = \{z_{ik}\}$, such that $z_{ik} = 1$ if $i \in C_k$ and $z_{ik} = 0$ otherwise. When a positive real value is used to indicate the affinity of a node i to a community k, $\Phi = \{\phi_{ik}\}$ is used with $\phi_{ik} \in \mathbb{R}^+$.

2 Models

From the ground truth of real world networks, it has been found that many real world network communities exhibit *pluralistic homophily* [18], that is, that the likelihood of an edge existing between any pair of nodes is directly proportional to the number of shared communities. The focus of this paper is on models that adhere to pluralistic homophily. In the following sub-sections we gather such models into a framework which we refer to as "Overlapping Weighted Stochastic Blockmodels" and from this general framework, we define four models that are of particular interest in our analysis. Two of these models are full blockmodels that allow for any type of block structure in the network, while the other two are constrained to just community structure. Our hypothesis is that community models, containing significantly fewer parameters than full blockmodels, should be easier to learn and sufficiently accurate to model real-world overlapping community structure. As part of this analysis, we combine for the first time, the *affiliation graph model* (AGM) with an Indian Buffet Process and compare this with the other three variants from the state-of-the-art.

Overlapping Weighted Stochastic Blockmodel: Many statistical methods for overlapping community detection proposed to date are based on a common over-

lapping weighted stochastic blockmodel (OWSB). The general model assumes that edges in a graph G are generated independently at random, given the community assignment parameters, which we write as Ψ. The likelihood of an observed weighted graph is

$$P(X|\Psi) = \prod_{i=1}^{n}\prod_{j=1}^{n} P(X_{ij} = x_{ij}|\Psi) \tag{1}$$

where $P(X_{ij} = w)$ is the probability that the edge weight w is observed on edge (i, j). In particular, $P(X_{ij} = w)$ is Poisson distributed with rate depending on the community assignments of nodes i and j, such that, the rate may be written as $\sum_{k=1}^{K}\sum_{\ell=1}^{K} \phi_{ik}\phi_{j\ell}\lambda_{k\ell}$, where $\lambda_{k\ell}$ represents the rate of edge generation between nodes in communities k and ℓ and $\phi_{ik} \in \mathbb{R}^{+}$ is a real-valued *affiliation* strength between node i and the community k. Thus, the community assignment parameters consist of the set $\Psi = (\phi_{ik}, \lambda_{k\ell}, K)$. The rates between communities are symmetric i.e. $\lambda_{k\ell} = \lambda_{\ell k}$. We refer to the above model as a *blockmodel* because between-community edges as well as within-community edges are modelled, i.e. we have $\lambda_{k\ell} > 0 \forall k, \ell$. It is referred to in [20] as an Edge Partition Model (EPM).

The weighted model may be reduced to an unweighted model, with likelihood

$$P(A|\Psi) = \prod_{i=1}^{n}\prod_{j=1}^{n} p_{ij}^{a_{ij}}(1 - p_{ij})^{1-a_{ij}} , \tag{2}$$

where $p_{ij} = P(X_{ij} > 0)$. Hence, $p_{ij} = 1 - \prod_{k}\prod_{\ell}(1 - \pi_{k\ell})^{\phi_{ik}\phi_{j\ell}}$ and $\pi_{k\ell} \equiv 1 - \exp(-\lambda_{k\ell})$. Note that $\pi_{k\ell} \in [0, 1]$ and represents the probability that nodes in communities k and ℓ are joined by an edge. Thus the unweighted model may be generated, by starting with rates $\lambda_{k\ell}$, or directly from probabilities $\pi_{k\ell}$.

For a fully Bayesian treatment of this model, we require the posterior distribution, which is proportional to the product of the likelihood and prior probabilities, $P(\Psi)$, i.e. $P(\Psi|X) \propto P(X|\Psi) \times P(\Psi)$. Different specialisations of the model differ in their choice of prior. In particular, the value of K may be given as an input parameter, in which case a model selection method must be applied to choose among a range of K's. A more sophisticated approach is to use a *non-parametric* prior.

Affiliation Graph Model: The Affiliation Graph Model (AGM) [18] is the specialisation of the OWSB in which the between-community edge rate $\lambda_{k\ell}$ for $k \neq \ell$, is set to zero and the edges are unweighted, with ϕ_{ik} constrained as $\phi_{ik} \equiv z_{ik} \in \{0, 1\}$. In this case, the $K(K + 1)/2$ rate parameters reduce to just K parameters which we write as $r_k \equiv \lambda_{kk}$, with $\pi_k = 1 - \exp(-r_k)$. To allow for the possibility of an edge between any pair of nodes, a *null* community is introduced, consisting of *all* the nodes in the network, to capture any noisy edges that are not otherwise explained by community membership. We refer to this as the ϵ community, with associated parameters r_ϵ and π_ϵ. State-of-the-art methods that follow this model, and fit it using a heuristic optimization technique include [11, 18, 19].

Non-parametric Models: Non-parametric models do not assume that the structure of the model is fixed. In the context of overlapping community detection, this corresponds to allowing the number of communities K to be inferred from the data and potentially grow to infinity. Two approaches are proposed in the state-of-the-art. The Infinite Edge Partition Model [20] uses a Hierarchical Gamma Process on $\lambda_{k\ell}$, and hence we refer to it as HGP-EPM. This model has been constrained to the AGM likelihood, in which case Gamma Process priors on r_k suffice and we refer to this constrained model as GP-AGM.

The IMRM [13] is an alternative non-parametric model, that constrains $\phi_{ik} \equiv z_{ik} \in \{0, 1\}$ and imposes an Indian Buffet process (IBP) prior on z_{ik}, allowing for an arbitrary number of communities. The IBP can be defined in terms of a community weight w_k, such that $z_{ik}|w_k \sim \text{Bernoulli}(w_k)$, with $w_k|\alpha \sim \text{Beta}(\alpha/K, 1)$. Allowing the number of communities $K \to \infty$, we obtain $Z \sim \text{IBP}(\alpha)$. The parameter α controls the number of active communities, the number of communities a node belongs to and the expected number of entries in Z. Our contribution is to constrain the IMRM to the AGM, by setting inter-community rates to zero. We refer to this model as an IBP-AGM model and its generative process is given in Algorithm 1.

The Bernoulli-Beta process in the IBP is a rich-get-richer process in which the probability that a node is assigned to any community is proportional to the number of nodes already in that community. On the other hand, the community assignment parameter in the HGP-EPM and GP-AGM depends only on node-specific parameters, so we do not expect to see a strong preferential attachment phenomenon in the community assignments in this case.

3   Related Work

In recent years, a number of generative models of networks with latent structure have been proposed in the state-of-the-art. The stochastic blockmodel (SBM) [8] posits that each node belongs to a single latent cluster and that interactions between pairs of nodes are governed by parameters, conditional on the cluster assignments of the nodes. The mixed membership stochastic blockmodel (MMSB) [1] extends this idea by assuming that each node has a distribution over the set of latent clusters and that pairs sample their cluster labels before selecting their interaction parameters. The assortative MMSB [5], specialises the MMSB by disallowing between-cluster interactions. However, MMSB models do not explicitly model for pluralistic homophily. The infinite relational model [9] is a non-parametric version of the SBM that places a Dirichlet Process prior on the community assignment of nodes. Another work on detecting latent overlapping clusters in networks is the mixtures of Dirichlet Network Distributions (MDND) model [17], that focuses on clustering links, rather than nodes. Other work focuses on the link prediction task, rather than on community identification. The work of [2] for instance, presents a model for sparse network generation, which is not explicitly defined in terms of latent clusters, but which can be related to stochastic blockmodels, in so far as the probability of a link between two

---

Algorithm 1 Generative process model of IBP-AGM

1: **procedure** GENERATENETWORK
2:    $\alpha \sim \text{Gamma}(1, \gamma)$
3:    $Z \sim \text{IBP}(\alpha)$
4:    **for** $k = 1 : K$ **do**
5:        $\pi_k \sim \text{Beta}(a, b)$
6:    **for** $i < j$ such that $j \in 1 : n$ and $i \geq 1$ **do**
7:        $a_{ij} \sim \text{Bernoulli}(p_{ij})$ where $p_{ij} = 1 - (1 - \pi_\epsilon) \prod_{k=1}^{K} (1 - \pi_k)^{z_{ik} z_{jk}}$

---

nodes depends on the nodes' parameters through a link function. These parameters do not correspond to explicit cluster assignments. In summary, while there are many works on stochastic blockmodels or similar models, our focus is on models that are based on the OWSB likelihood that explicitly models pluralistic homophily.

4   Inference Techniques

We have used standard MCMC algorithms for inference in IBP-AGM. The pseudo-code for the MCMC of IBP-AGM is given in Algorithm 2. For sampling the IBP parameter α, we use Gibbs sampling with a Gamma$(1, \gamma)$ prior on α. This results in a Gamma posterior given in step 14 of Algorithm 2. For sampling the community assignment Z, due to the non-conjugacy of the likelihood $P(A|Z, \pi)$ in (2) and its prior $Z \sim \text{IBP}(\alpha)$, we use Auxiliary Gibbs Sampling (AGS) to sample from a non-conjugate IBP [6]. This is similar to AGS [14] for the non-conjugate Dirichlet process mixture model. If $m_{-ik} = \sum_{j \neq i}(z_{jk}) > 0$, we sample z_{ik} from its posterior in step 4 of Algorithm 2. To sample new communities, we consider K^* auxiliary communities with $\pi^* \sim \text{Beta}(a, b)$. A node i can be assigned to any of 2^{K^*} combinations of these K^* communities. With l running over all 2^{K^*} possibilities, each Z_l^* has joint posterior consisting of the Bernoulli probabilities of assigning the node i to each of the auxiliary communities with probability $\frac{\alpha/K^*}{N+\alpha/K^*}$. For sampling the edge probability parameter π, we use Hamiltonian Monte Carlo (HMC) [4]. This method utilizes the gradient of the log posterior to reduce the correlation between successive sampled states, thus targeting higher probability states and resulting in fast convergence. It requires a discretisation of the Hamiltonian equations, for which we use the leapfrog method. Similar to the HMC for IMRM [13], π is sampled using HMC by transforming $\pi_k \in [0, 1]$ to $s_k \in (-\infty, \infty)$ with $\pi_k = \frac{1}{1+\exp(-s_k)}$, $\pi_k \sim \text{Beta}(a, b)$ and computing the log posterior, $\log P(s|A, Z)$ and its gradient, $\nabla \log P(s|A, Z)$.

Algorithm 2 MCMC algorithm for IBP-AGM

    Input: A, K^*
    Initialize: a, b, π, Z, α
    For each iteration, repeat :
1: **for** $i = 1 : n$ **do** {Sample Z using **Auxiliary Gibbs Sampling**}
2:     **for** $k = 1 : K$ **do**
3:         **if** $m_{-ik} > 0$ **then**
4:             Sample z_{ik} from $P(z_{ik} = z|A, Z_{-ik}, \pi) \propto ((1 - z)n - (-1)^z m_{-ik})P(A|Z, \pi)$
5:         **else**{$m_{-ik} = 0$}
6:             Set one of the auxiliary community edge probability π^* to be π_k and $z_{ik} = 0$
7:     Sample the remaining π^* from Beta(a, b)
8:     **for** $l = 1 : 2^{K^*}$ **do**
9:         Compute the posterior $P(Z_l^*|A, Z, \pi, \pi^*) \propto P(Z_l^*)P(A|Z_l^*, Z, \pi, \pi^*)$
10:     Sample Z_l^* from 2^{K^*} possibilities with probabilities $P(Z_l^*|A, Z, \pi, \pi^*)$
11:     Set $\pi = \{\pi, \pi^*\}$ and $Z = \{Z, Z_l^*\}$
12:     Remove the zero columns from Z and its corresponding π
13: Sample π using **Hamiltonian Monte Carlo** using $\nabla \log P(s|A, Z)$, where $\pi = \frac{1}{1+\exp(-s)}$
14: Sample α using **Gibbs Sampling** from its posterior $P(\alpha|Z) = \text{Gamma}(1 + K, \gamma + \sum_{j=1}^{n} \frac{1}{j})$

---

5 Experiments

The main purpose of our empirical analysis is to compare the blockmodels with the community models, with the two choices of non-parametric prior (GP and IBP). The inference algorithm for IBP-AGM is described in Sect. 4. For IMRM, we use HMC for IMRM [13] and Auxiliary Gibbs sampling for IBP. For EPM models, we used Gibbs sampling described in [20], using a truncated Gamma Process and hence requires to input a maximum number of communities. On the other hand, inference with the IBP can generate any number of communities (truly non-parametric).

The following settings are used throughout the experiments, unless otherwise specified. In IBP-AGM and IMRM, in each Gibbs update for AGS, we choose the number of auxiliary communities $K^* = 3$ and for the HMC, a step size of 0.01 and a number of leapfrogs of 10 and set $\pi_\epsilon = 0.00005$. For IBP-AGM, we consider $\gamma = 100$ and $a = b = 1$. Similarly to HGP-EPM, we choose the prior on the edge probability for IMRM in such a way that interaction within a community is greater than between communities, i.e. $a = 1, b = 5 \forall k \neq \ell$ and $a = 5, b = 1 \forall k = \ell$ as given in [13] and consider $\gamma = 1$. For EPM, we have considered the same settings and initialization as given in [20]. In all the four non-parametric models, the number of communities is initialised, similarly to [20], with the condition $K = \min(100, n)$ if $n < 2,000$ and $K = \min(256, n)$, otherwise. We have taken the same initial settings for IBP-AGM and GP-AGM in terms of initial community assignments and edge probability parameters i.e. $z_{ik} = 1 \forall i, k$ and $\pi_k = 1 - \exp(-1/K)$ for IBP-AGM or $r_k = 1/K$ for GP-AGM. For IMRM, each edge probability is assigned to $1 - \exp(-1/K)$ initially. All the experiments are done on Intel core i5, 4 cores.

Similarly to the evaluation in [12, 13, 20], we have used missing link prediction as a measure to compare the performance between methods. The test set consists of 20% of the available pairs of nodes, whose between-edge existence or otherwise is withheld from the training data. AUC-ROC values are calculated for this test set.

While the AUC-ROC is a useful measure to detect convergence of the models, it is worth noting, as seen in the experiments later in the paper, that the different models are able to predict missing edges with similar performance. We emphasise that the purpose of our work is to detect the community structure as an output, and therefore, given similar AUC-ROC performance, we contend that simpler latent structures are preferable over more complex ones. This implies a preference towards community models, rather than blockmodels, since overlapping block structure is difficult to comprehend and a preference towards fewer, rather than more communities. Initially, we test their ability to recover the communities from networks generated from the AGM. These generated networks are simple with a known number of communities yet sufficiently complex to allow for different community structures to be found. Then, we test their performance in detecting real world communities.

Networks Generated by AGM: Networks with two communities are generated using the generative process of AGM. We choose the network size from $n = \{30, 100, 500\}$ and set $K = 2$ with $\pi_k = 0.8 \forall k$ and $\pi_\epsilon = 0.00005$. We choose community assignment Z, such that in each network, 20% of the nodes belong to the overlapping region of the two communities and 40% of the nodes belong to each community only. An edge between nodes i and j is generated with probability $p_{ij} = 1 - (1 - \pi_\epsilon) \prod_k (1 - \pi_k)^{z_{ik} z_{jk}}$. We run the different models with these networks as input, expecting to find $K = 2$ communities. After burn in of 15,000 iterations, 15,000 samples are collected, and the average result of 5 random runs is reported.

From Table 1, we can see that GP-AGM converges to a model with greater than two communities, whereas IBP-AGM converges to 2 communities. HGP-EPM and IMRM also fit the networks with more complex block structure rather than having two blocks with $\pi_{k\ell} \approx 0$, $\forall k \neq \ell$. As the number of edges increases in the network, the blockmodels scale less well (in time) compared with the community models. From the trace plots of the number of communities and log likelihood in Figs. 1 and 2, respectively, IBP-AGM converges in fewer iterations to GP-AGM. Moreover, as GP-AGM converges to a higher number of communities, this results in a smaller likelihood than IBP-AGM, see Fig. 2. In terms of predicting the missing links, all models seem to perform equally.

Real World Networks: We take 3 small networks ($n < 250$) i.e. Football [15] with a ground truth of 12 communities, Protein230 [20], NIPS234 [20], and 4 large networks ($n > 2,000$) i.e. NIPS12 [20], Yeast [13], UsPower [13], Erdos [13]. 1,500 samples are collected after burn-in of 1,500 iterations and the average result of 5 random runs is reported. The first notable finding of Table 2 is that all models achieve similar AUC-ROC scores, with HGP-EPM and IBP-AGM marginally outperforming the others. Examining the found latent structure, we see that generally, the models based on the Gamma Process prior, HGP-EPM and GP-AGM, sample a greater number of blocks or communities than IMRM and IBP-AGM. The shrinkage

Table 1 Comparison between HGP-EPM, IMRM, GP-AGM, IBP-AGM on generated networks

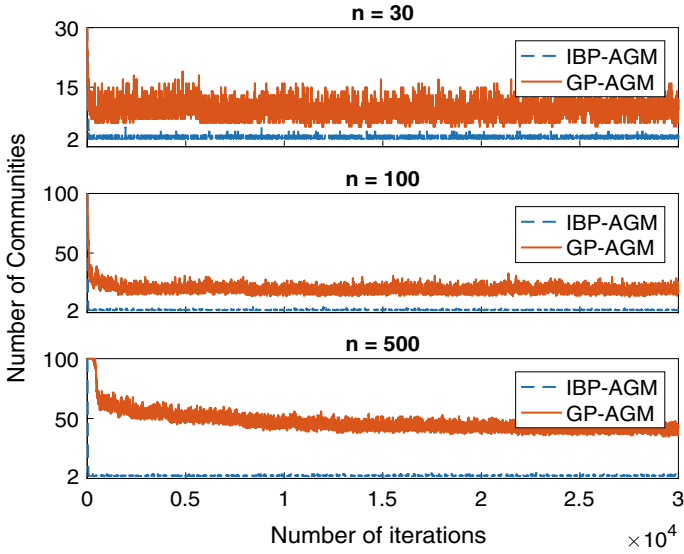| #Nodes	HGP-EPM			IMRM			GP-AGM			IBP-AGM		
	AUC-ROC	Time (h)	#Block	AUC-ROC	Time (h)	#Block	AUC-ROC	Time (h)	#Comm	AUC-ROC	Time (h)	#Comm
n = 30	0.9166	0.0279	3.1508	0.9054	0.0952	2.5641	0.8608	0.0104	7.4751	0.8968	0.0605	2.0596
n = 100	0.8629	0.2907	5.5419	0.8697	0.5634	2.4125	0.8650	0.0604	21.8374	0.8732	0.1084	2.0593
n = 500	0.8643	6.9062	8.0207	0.8635	9.8146	4.5546	0.8656	1.2149	43.5966	0.8649	0.6952	2.0829

Fig. 1 Trace plots of the number of communities for a single run of the IBP-AGM and GP-AGM on generated networks
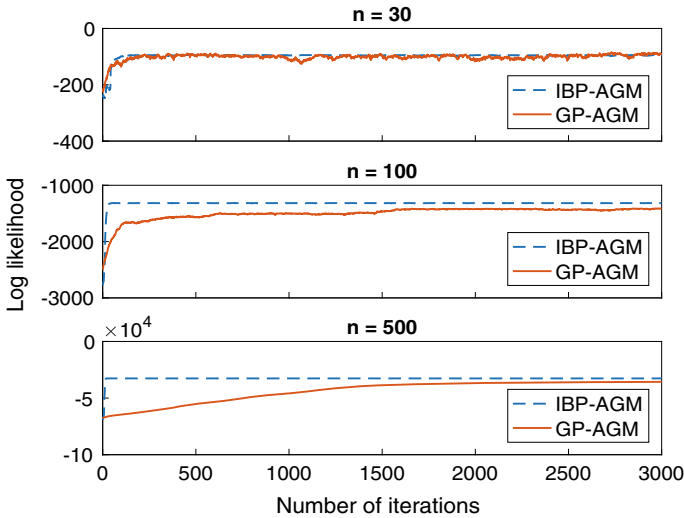

Fig. 2 Trace plots of log likelihood of first 3000 iterations for a single run of the IBP-AGM and GP-AGM on generated networks

Table 2 Comparison between HGP-EPM, IMRM, GP-AGM, IBP-AGM on real world networks

Real networks	HGP-EPM			IMRM			GP-AGM			IBP-AGM		
	AUC-ROC	Time (h)	#Block	AUC-ROC	Time (h)	#Block	AUC-ROC	Time (h)	#Comm	AUC-ROC	Time (h)	#Comm
Football	0.8411	0.0157	22.8625	0.8082	0.0335	7.9711	0.7651	0.0052	49.6084	0.8357	0.0167	11.2261
Protein230	0.9532	0.0251	31.9112	0.8958	0.0558	6.5621	0.8919	0.0119	44.4279	0.9190	0.0250	7.6931
NIPS234	0.9444	0.0282	60.0447	0.8284	0.0521	6.3777	0.8961	0.0127	63.3365	0.9522	0.0295	14.1815
NIPS12	0.9763	1.0088	242.7089	0.7635	0.9446	6.9140	0.9395	2.5234	234.2423	0.9611	1.4542	67.1507
Yeast	0.9329	1.5899	205.0551	0.8889	1.6602	8	0.8956	4.4883	238.5441	0.9178	1.3865	41.1652
UsPower	0.6175	4.2743	117.2415	0.5634	3.2992	3.9888	0.5278	22.8958	33.7568	0.6581	6.6673	35.0416
Erdos	0.9721	5.3497	91.0867	0.9601	7.5348	8.6455	0.9638	29.9171	190.1051	0.9749	11.1775	60.2116

problem of Gamma Process prior on EPM model resulting into more number of communities has been studied in [16]. This suggests that the IBP prior can recover simpler latent structure, particularly for the larger networks. For example, the GP-AGM finds 190 communities, while IBP-AGM finds 60 in the Erdos network.

While HGP-EPM generally obtains the best AUC-ROC performance, the structure it uncovers, consisting of 91 overlapping blocks in Erdos, is extremely complex. Each edge is explained, either by its membership of multiple blocks or by the fact that it straddles different blocks with non-zero between-block edge probability. It is very difficult to interpret such a model. The corresponding IBP-AGM uncovers just 60 overlapping communities, with between community edges explained entirely as noisy links, and in this case, a marginally higher AUC-ROC score.

6 Conclusion

In this paper, we have presented an overlapping community detection algorithm, which is a non-parametric version of the affiliation graph model (AGM) that uses an Indian Buffet Process (IBP) prior and exhibits pluralistic homophily. We have compared the model with another non-parametric model of AGM that uses a Gamma Process (GP) prior. In terms of convergence, IBP-AGM performs better than the GP-AGM with fewer communities in general, though there is not much difference in AUC scores for missing link prediction. Hence, empirically, IBP-AGM has better shrinkage while GP-AGM tends to overfit the data with larger number of parameters. Comparing with blockmodels, we find that the community models obtain similar performance to blockmodels and generally return simpler structures. Our future work will focus on making IBP-AGM more scalable.

References

1. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic blockmodels. J. Mach. Learn. Res. **9**, 1981–2014 (2008)
2. Caron, F., Fox, E.B.: Sparse graphs using exchangeable random measures. J. R. Stat. Soc. Ser. B Stat. Methodol. **79**(5), 1295–1366 (2017)
3. Crampes, M., Plantié, M.: Overlapping community detection optimization and Nash equilibrium. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, p. 2. ACM (2015)
4. Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D.: Hybrid Monte Carlo. Phys. Lett. B **195**(2), 216–222 (1987)
5. Gopalan, P.K., Gerrish, S., Freedman, M., Blei, D.M., Mimno, D.M.: Scalable inference of overlapping communities. In: Advances in Neural Information Processing Systems, pp. 2249–2257 (2012)
6. Görür, D., Jäkel, F., Rasmussen, C.E.: A choice model with infinitely many latent features. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 361–368. ACM (2006)

7. Gregory, S.: Finding overlapping communities in networks by label propagation. New J. Phys. **12**(10), 103018 (2010)
8. Holland, P.W., Laskey, K.B., Leinhardt, S.: Stochastic blockmodels: first steps. Soc. Netw. **5**(2), 109–137 (1983)
9. Kemp, C., Tenenbaum, J.B., Griffiths, T.L., Yamada, T., Ueda, N.: Learning systems of concepts with an infinite relational model. In: Proceedings of the 21st National Conference on Artificial Intelligence, pp. 381–388 (2006)
10. Lee, C., Reid, F., McDaid, A., Hurley, N.: Seeding for pervasively overlapping communities. Phys. Rev. E **83**, 066107 (2011)
11. McDaid, A., Hurley, N.: Detecting highly overlapping communities with model-based overlapping seed expansion. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 112–119. IEEE (2010)
12. Mørup, M., Schmidt, M.N.: Bayesian community detection. Neural Comput. **24**(9), 2434–2456 (2012)
13. Mørup, M., Schmidt, M.N., Hansen, L.K.: Infinite multiple membership relational modeling for complex networks. In: Proceedings of the International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6. IEEE (2011)
14. Neal, R.M.: Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Stat. **9**(2), 249–265 (2000)
15. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
16. Ohama, I., Sato, I., Kida, T., Arimura, H.: On the model shrinkage effect of gamma process edge partition models. In: Advances in Neural Information Processing Systems, pp. 397–405 (2017)
17. Williamson, S.A.: Nonparametric network models for link prediction. J. Mach. Learn. Res. **17**(1), 7102–7121 (2016)
18. Yang, J., Leskovec, J.: Community-affiliation graph model for overlapping network community detection. In: IEEE 12th International Conference on Data Mining (ICDM), pp. 1170–1175. IEEE (2012)
19. Yang, J., Leskovec, J.: Overlapping community detection at scale: a nonnegative matrix factorization approach. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 587–596. ACM (2013)
20. Zhou, M.: Infinite edge partition models for overlapping community detection and link prediction. In: Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1135–1143 (2015)