

NONPARAMETRIC BAYESIAN MATRIX FACTORIZATION FOR ASSORTATIVE NETWORKS

Mingyuan Zhou

The University of Texas at Austin
IROM Department, McCombs School of Business
Austin, TX 78712, USA

ABSTRACT

We describe in detail the gamma process edge partition model that is well suited to analyze assortative relational networks. The model links the binary edges of an undirected and unweighted relational network with a latent factor model via the Bernoulli-Poisson link, and uses the gamma process to support a potentially infinite number of latent communities. The communities are allowed to overlap with each other, with a community's overlapping parts assumed to be more densely connected than its non-overlapping ones. The model is evaluated with synthetic data to illustrate its ability to model assortative networks and its restriction on modeling disassortative ones.

Index Terms— Gamma process, factor analysis, Bernoulli-Poisson link, overlapping community detection, link prediction

1. INTRODUCTION

For an assortative relational network, a subset of nodes (or vertices) that are densely connected to each other but sparsely to the others are often considered to belong to the same community. For example, in a social network, a community may consist of a group of closely related friends. Whereas for a disassortative relational network, a subset of nodes that are sparsely connected to each other but densely connected to another subset of nodes are often considered to belong to the same community. For example, in a predator-prey network, a community may consist of a group of animals that play similar roles in the ecosystem but not necessarily prey on each other. A relational network may exhibit both assortativity and disassortativity, which are also known as homophily and stochastic equivalence, respectively [1]. In this paper, we focus on the study of assortative networks with dense intra-community and sparse inter-community connections, and we assume that a community's overlapping parts are more densely connected than its non-overlapping ones.

While there is a wide variety of network algorithms for community detection based on various effective heuristics, see [2] and references therein, we are interested in construct-

ing a generative model that describes how the edges of an undirected and unweighted relational network are generated in a probabilistic manner. With a generative model, one would be able to not only detect latent communities, but also simulate random networks and predict missing edges. One could further introduce an appropriate nonparametric Bayesian prior to avoid the need of model selection, allowing the number of latent communities to be automatically inferred from the observed network.

A probabilistic model for network analysis can often be considered either as a latent class model or a latent factor model. The stochastic blockmodel (SBM) [3] and its nonparametric Bayesian version based on the Chinese restaurant process, the infinite relations model (IRM) [4], are popular latent class network models. A SBM assigns each node to a single community and models the probability for an edge to exist between two nodes solely based the two nodes' community assignments. It has low computational complexity and can model both assortativity and disassortativity, but is restrictive in that the latent communities are not allowed to overlap. To generalize the SBM by allowing a node to belong to multiple communities, one may consider the mixed-membership stochastic blockmodel (MMSB) [5]. Yet for the the MMSB, the computation grows quadratically as a function of the number of nodes N , a community's overlapping parts are assumed to be less densely connected than its non-overlapping parts [6, 7], and the number of communities is a model parameter that needs to be carefully selected. The Eigenmodel of [1] is a typical latent factor model that uses the logit link to connect the binary edges to a factor model in the latent Gaussian space, and the infinite latent feature relational model (LFRM) of [8] can be considered as a nonparametric Bayesian generalization of the Eigenmodel. These latent Gaussian factor models can be used for link prediction but not necessarily community detection, as their latent representations are often not easily interpretable.

In this paper, we consider the gamma process edge partition model (GP-EPM) that is well suited for analyzing assortative networks. The GP-EPM can be considered as both a latent factor model and a latent class model: as a latent factor model, different from the Eigenmodel and LRFM that links a

binary edge to a latent Gaussian random variable via the logit link, the GP-EPM links a binary edge to a latent count via a Bernoulli-Poisson link, and factorizes the latent random count matrix under the Poisson likelihood; as a latent class model, different from the SBM that clusters nodes and the MMSB that clusters all possible edges, the GP-EPM partitions only the observed edges, which further leads to the partition of nodes. The GP-EPM supports a potentially infinite number of clusters, allows the communities to overlap, and its computation grows as a linear function of the number of nodes N for sparse networks commonly observed in practice. We mention that the GP-EPM can be considered as a special case of the hierarchical gamma process edge partition model (HGP-EPM) proposed in [9], which can be used to model both assortativity and disassortativity. In this paper, we discuss the GP-EPM in detail and demonstrate its ability to model assortative networks and its restriction on modeling disassortative ones.

2. GAMMA PROCESS EDGE PARTITION MODEL

2.1. Gamma Process

To support a potentially infinite number of latent communities, we first define

$$G \sim \Gamma\text{P}(G_0, 1/c_0) \quad (1)$$

as a gamma process on a product space $\mathbb{R}^+ \times \Omega$, where $\mathbb{R}^+ = \{x : x > 0\}$, Ω is a complete separable metric space, $1/c_0$ is a positive scale parameter, and G_0 is a finite and continuous base measure, such that $G(A) \sim \text{Gam}(G_0(A), 1/c_0)$ for each Borel set $A \subset \Omega$ [10, 11]. The Lévy measure of the gamma process can be expressed as $\nu(dr d\phi) = r^{-1} e^{-c_0 r} dr G_0(d\phi)$. Since $\int_{\mathbb{R}^+ \times \Omega} \nu(dr d\phi) = \infty$ but $\int_{\mathbb{R}^+ \times \Omega} r \nu(dr d\phi)$ is finite, a draw from the gamma process consists of countably infinite atoms, expressed as $G = \sum_{k=1}^{\infty} r_k \delta_{\phi_k}$, where $\phi_k \stackrel{iid}{\sim} g_0$, $g_0(d\phi) = G_0(d\phi)/\gamma_0$ is the base distribution, and $\gamma_0 = G_0(\Omega)$ is the mass parameter.

2.2. Link Probability and Poisson Factor Analysis

With $\phi_k = (\phi_{1k}, \dots, \phi_{Nk})^T$ measuring how strongly the N nodes are affiliated with community k and with r_k measuring the prevalence of community k , we use the product $r_k \phi_{ik} \phi_{jk}$ to measure how strongly nodes i and j are connected due to their affiliations with community k , and use

$$\lambda_{ij} = \sum_{k=1}^{\infty} r_k \phi_{ik} \phi_{jk} \quad (2)$$

to measure the overall connection strength between nodes i and j . To use this positive strength to model the probability for an edge b_{ij} to exist between two nodes, we assume that

$$P(b_{ij} = 1) = 1 - e^{-\lambda_{ij}} = 1 - \exp\left(-\sum_{k=1}^{\infty} r_k \phi_{ik} \phi_{jk}\right). \quad (3)$$

Thus if both nodes i and j are strongly affiliated with a popular community in the network, then there is a high probability for nodes i and j to be connected by an edge.

A key challenge for the likelihood in (3) is to infer the parameters inside the summation under this nonlinear link function. As in [9], the Bernoulli random variable $b \sim \text{Bernoulli}(1 - e^{-\lambda})$ can be equivalently generated by thresholding a Poisson random variable as

$$b = \mathbf{1}(m > 0), \quad m \sim \text{Pois}(\lambda). \quad (4)$$

Using this Bernoulli-Poisson (BerPo) link function, we transform the problem of modeling binary edges into a problem of factorizing latent counts under the Poisson likelihood as

$$m_{ij} \sim \text{Po}\left(\sum_{k=1}^{\infty} r_k \phi_{ik} \phi_{jk}\right), \quad (5)$$

where $m_{ij} = m_{ji}$ is the latent integer-valued weight that links nodes i and j , which is zero almost surely (a.s.) given $b_{ij} = 0$ and is drawn from a truncated Poisson distribution given $b_{ij} = 1$. Clearly, two nodes with similar latent features are encouraged to be linked by an edge with a large latent integer weight and the overlapping parts of two communities would be more densely connected than their non-overlapping ones. Thus this model is well suited to model an assortative relational network exhibiting homophily but not necessarily stochastic equivalence.

We note [12] had examined a model related to (5), but used Poisson distribution to model binary data, did not provide a principled way to set the number of communities, and had to create possibly nonexistent self-edges in order to derive tractable expectation-maximization (EM) inference. This paper will rigorously address all these issues in a nonparametric Bayesian manner and provide efficient Bayesian inference.

2.3. Overlapping Community Structures

Note that (5) can be augmented as

$$m_{ij} = \sum_k m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}). \quad (6)$$

where m_{ijk} represents how often nodes i and j interact due to their affiliations with community k . Similar to [9], we may consider that the model is partitioning the count m_{ij} into $\{m_{ijk}\}_k$, and hence we call the Poisson factor model in (5) together with the BerPo link as an edge partition model (EPM), in which each edge is partitioned according to all possible intra-community interactions, and how strongly node i is affiliated with community k can be measured with $r_k \phi_{ik} \sum_{j \neq i} \phi_{jk}$, which represents how strongly node i interacts with all the other nodes through its affiliation with community k . We further introduce the latent count

$$m_{i \cdot k} := \sum_{j=i+1}^N m_{ijk} + \sum_{j=1}^{i-1} m_{jik} \quad (7)$$

to represent how often node i is connected to the other nodes due to its affiliation with community k . We can then assign node i to multiple communities in $\{k : m_{i,k} \geq 1\}$, or (hard) assign it to a single community using either $\operatorname{argmax}_k(r_k \phi_{ik} \sum_{j \neq i} \phi_{jk})$ or $\operatorname{argmax}_k(m_{i,k})$. By hard assigning each node to a single community and ordering the nodes from the same community to be adjacent to each other, we expect the ordered adjacency matrix to exhibit a block structure, with the blocks along the diagonal represent the intra-community connections.

2.4. Hierarchical Model and Gibbs Sampling

In this paper, we consider an unweighted undirected network, where $b_{ji} = b_{ij}$ and self-links b_{ii} are not defined. Thus we only consider b_{ij} for $j > i$. We truncate the number of atoms of the gamma process to be K and construct the (truncated) gamma process EPM as

$$\begin{aligned} b_{ij} &= \mathbf{1}(m_{ij} \geq 1), \\ m_{ij} &= \sum_{k=1}^K m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}), \\ \phi_{ik} &\sim \text{Gam}(a_i, 1/c_i), \quad a_i \sim \text{Gam}(e_0, 1/f_0), \\ r_k &\sim \text{Gam}(\gamma_0/K, 1/c_0), \quad \gamma_0 \sim \text{Gam}(e_1, 1/f_1), \end{aligned} \quad (8)$$

where the $\text{Gam}(1, 1)$ prior is also imposed on c_0 and c_i . We usually set K to be large enough to ensure a good approximation to the truly infinite model. Note that if we marginalize out both m_{ij} and m_{ijk} , then we have

$$b_{ij} \sim \text{Bernoulli} \left[1 - \prod_{k=1}^K \exp(-r_k \phi_{ik} \phi_{jk}) \right].$$

We exploit the augmented representation in (8) to derive Gibbs sampling update equations, as described below.

Let the latent counts $m_{..k}$ be defined as

$$m_{..k} := \sum_{i=1}^N \sum_{j=i+1}^N m_{ijk} = \frac{1}{2} \sum_{i=1}^N m_{i..k}.$$

Using the Poisson additive property, we have

$$m_{i..k} \sim \text{Po} \left(r_k \phi_{ik} \sum_{j \neq i} \phi_{jk} \right), \quad (9)$$

$$m_{..k} \sim \text{Po} \left(r_k \frac{\sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}{2} \right). \quad (10)$$

Marginalizing out ϕ_{ik} from (9), we have

$$m_{i..k} \sim \text{NB}(a_i, p'_{ik}), \quad (11)$$

where

$$p'_{ik} := \frac{r_k \sum_{j \neq i} \phi_{jk}}{c_i + r_k \sum_{j \neq i} \phi_{jk}}.$$

Marginalizing out r_k from (10), we have

$$m_{..k} \sim \text{NB}(\gamma_0/K, \tilde{p}_k), \quad (12)$$

where

$$\tilde{p}_k := \frac{\sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}{2c_0 + \sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}.$$

Using (9)-(12), we can develop closed-form Gibbs sampling update equations for all model parameters, as described below.

Sample m_{ij} . We sample a latent count for each b_{ij} as

$$(m_{ij}|-) \sim b_{ij} \text{Po}_+ \left(\sum_{k=1}^K r_k \phi_{ik} \phi_{jk} \right). \quad (13)$$

Sample m_{ijk} . Using the relationship between the Poisson and multinomial distributions, similar to the derivation in [13], we have

$$(\{m_{ijk}\}_{k=1:K}|-) \sim \text{Mult} \left(m_{ij}; \frac{\{r_k \phi_{ik} \phi_{jk}\}_{k=1:K}}{\sum_{k'} r_{k'} \phi_{ik'} \phi_{jk'}} \right). \quad (14)$$

Note that in each MCMC iteration we can store $m_{i..k}$ but not necessarily m_{ijk} in the memory.

Sample a_i . Using (11) and the data augmentation technique developed in [14, 15] for the negative binomial distribution, we sample a_i as

$$(\ell_{ik}|-) \sim \sum_{t=1}^{m_{i..k}} \text{Ber} \left(\frac{a_i}{a_i + t - 1} \right), \quad (15)$$

$$(a_i|-) \sim \text{Gam} \left(e_0 + \sum_k \ell_{ik}, \frac{1}{f_0 + \sum_k \ln(1 - p'_{ik})} \right). \quad (16)$$

Sample ϕ_{ik} . Using (9) and the gamma-Poisson conjugacy, we have

$$(\phi_{ik}|-) \sim \text{Gam} \left(a_i + m_{i..k}, \frac{1}{c_i + r_k \sum_{j \neq i} \phi_{jk}} \right). \quad (17)$$

Sample γ_0 . Similar to the inference of a_i , using (12), We sample γ_0 as

$$(\ell_k|-) \sim \sum_{t=1}^{m_{..k}} \text{Ber} \left(\frac{\gamma_0/K}{\gamma_0/K + t - 1} \right), \quad (18)$$

$$(\gamma_0|-) \sim \text{Gam} \left(e_0 + \sum_k \ell_k, \frac{1}{f_0 - \frac{1}{K} \sum_k \ln(1 - \tilde{p}_k)} \right). \quad (19)$$

Sample c_i and c_0 . They can be sampled from gamma distributions using the conjugacy between gamma distributions, omitted here for brevity.

Sample r_k . Using (10) and the gamma-Poisson conjugacy, we have

$$(r_k|-) \sim \text{Gam}\left(\frac{\gamma_0}{K} + m_{\cdot,k}, \frac{1}{c_0 + \sum_i \sum_{j \neq i} \frac{1}{2} \phi_{ik} \phi_{jk}}\right). \quad (20)$$

Each Gibbs sampling iteration for the gamma process EPM proceeds from (13) to (20).

2.5. Gamma Process AGM

We notice an interesting connection to the community-affiliation graph model (AGM) of [6, 7]. In fact, a slightly modified GP-EPM as

$$b_{ij} \sim \text{Ber}\left[1 - e^{-\epsilon} \prod_k \exp(-r_k \phi_{ik} \phi_{jk})\right],$$

where $\epsilon \in \mathbb{R}^+$ and $\phi_{ik} \in \{0, 1\}$, could be considered as a gamma process AGM. It is argued in [6, 7] that all previous community detection methods, including clique percolation and MMSB, would fail to detect communities with dense overlaps, because they all had a hidden assumption that a community's overlapping parts are less densely connected than its non-overlapping ones. The same as the AGM, the gamma process EPM does not make such a restrictive assumption and allows overlaps of communities to be denser than communities themselves; beyond the AGM, we do not restrict ϕ_{ik} to be either zero or one, and our generative model is built under a rigorous nonparametric Bayesian framework with efficient Bayesian inference.

Closely related to the gamma process EPM, the hierarchical model for the (truncated) gamma process AGM can be expressed as

$$\begin{aligned} b_{ij} &= \mathbf{1}(m_{ij} \geq 1), \\ m_{ij} &= u_{ij} + \sum_{k=1}^K m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}), \\ u_{ij} &\sim \text{Po}(\epsilon), \quad \epsilon \sim \text{Gam}(a_0, 1/b_0), \\ \phi_{ik} &\sim \text{Ber}(\pi_i), \quad \pi_i \sim \text{Beta}(a_1, b_1), \\ r_k &\sim \text{Gam}(\gamma_0/K, 1/c_0), \quad \gamma_0 \sim \text{Gam}(e_1, 1/f_1). \end{aligned} \quad (21)$$

We sample r_k , γ_0 and c_0 in the same way we sample them in the gamma process EPM. Below we describe the other Gibbs sampling update equations.

Sample m_{ij} . We sample a latent count for each b_{ij} as

$$(m_{ij}|-) \sim b_{ij} \text{Po}_+\left(\epsilon + \sum_{k=1}^K r_k \phi_{ik} \phi_{jk}\right). \quad (22)$$

Sample u_{ij} and m_{ijk} . We first sample u_{ij} as

$$(u_{ij}|-) \sim \text{Binomial}\left(m_{ij}; \frac{\epsilon}{\epsilon + \sum_{k=1}^K r_k \phi_{ik} \phi_{jk}}\right) \quad (23)$$

and then sample m_{ijk} as

$$(\{m_{ijk}\}_{k=1:K}|-) \sim \text{Mult}\left(m_{ij} - u_{ij}; \frac{\{r_k \phi_{ik} \phi_{jk}\}_{k=1:K}}{\sum_{k'} r_{k'} \phi_{ik'} \phi_{jk'}}\right). \quad (24)$$

Sample ϵ . We sample ϵ as

$$(\epsilon|-) \sim \text{Gam}\left(a_0 + \sum_i \sum_{j>i} u_{ij}, \frac{1}{b_0 + \frac{N(N-1)}{2}}\right). \quad (25)$$

Sample ϕ_{ik} . We sample ϕ_{ik} as

$$(\phi_{ik}|-) \sim \text{Ber}\left(\frac{\pi_i \exp(r_k \sum_{j \neq i} \phi_{jk})}{\pi_i \exp(r_k \sum_{j \neq i} \phi_{jk}) + 1 - \pi_i}\right) \quad (26)$$

if $m_{i \cdot k} = 0$ and let $\phi_{ik} = 1$ if $m_{i \cdot k} > 0$.

Sample π_i . We sample π_i as

$$(\pi_i|-) \sim \text{Beta}\left(a_1 + \sum_k \phi_{ik}, b_1 + K - \sum_k \phi_{ik}\right). \quad (27)$$

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

We study a synthetic network of 70 nodes to illustrate the properties of the proposed and related models. Fig. 1 (a) shows the ground-truth link probabilities and Fig. 1 (b) shows a simulated network adjacency matrix using these probabilities. This assortative network consists of four communities, all of which have dense intra-community connections and the second community overlaps with both the first and third communities. We randomly select 80% of the pairs of nodes and use them to estimate the ground-truth link probabilities. As shown in Fig. 1 (c)-(f), the IRM accurately captures the community structures but produces cartoonish blocks, the Eigenmodel somewhat overfit the data, the AGM produces some undesired artifacts, and the GP-EPM provides a reconstruction that looks most similar to the ground truth. As shown in Table 1, both the GP-EPM and Eigenmodel perform well and clearly outperform the IRM and AGM in missing link prediction, measured by both the area under the ROC curve and the area under the precision-recall (PR) curve.

We further consider a dissortative network, whose ground-truth link probabilities and simulated network adjacency matrix are shown in Figs. 2 (a) and (b), respectively. This dissortative network consists of four communities: the first and second communities have dense intra-community connections and they partially overlap with each other, the third and fourth communities have no intra-community connections but dense inter-community connections, and the third community also partially overlaps with the second community. We randomly select 80% of the pairs of nodes and use them to estimate the ground-truth link probabilities. As shown in Figure 2 (c)-(f), the IRM accurately captures the community structures but produces cartoonish blocks, Eigenmodel still

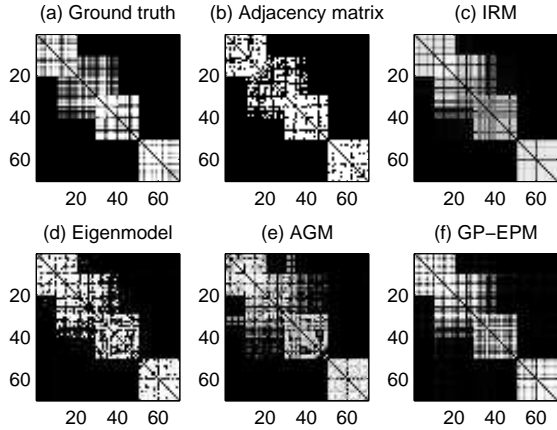


Fig. 1. Comparison of four algorithms' abilities to recover the ground-truth link probabilities using 80% of the pairs of nodes randomly selected from a simulated unweighted undirected assortative relational network. The number of features for the Eigenmodel is set as $K = 4$.

Table 1. Comparison of five algorithms' abilities to predict missing edges of a synthetic assortative network. The number of features for the Eigenmodel is set as $K = 4$.

Model	AUC-ROC	AUC-PR
IRM	0.9680 ± 0.0073	0.8636 ± 0.0448
Eigenmodel	0.9746 ± 0.0066	0.9073 ± 0.0236
AGM	0.9291 ± 0.0184	0.8166 ± 0.0470
GP-EPM	0.9746 ± 0.0056	0.9042 ± 0.0270

performs well but somewhat overfits the data, and both the AGM and GP-EPM combine the third and fourth communities into a single block, indicating their limited abilities to model dissortative relational networks.

In conclusion, the gamma process edge partition model provides an efficient and effective solution to model assortative relational networks but has limited ability to model dissortative ones. As in [9], to model dissortative relational networks, one may further introduce an additional latent matrix to capture the interactions between different communities.

REFERENCES

- [1] P. Hoff, "Modeling homophily and stochastic equivalence in symmetric relational data," in *NIPS*, 2008.
- [2] S. Fortunato, "Community detection in graphs," *Physics Reports*, pp. 75–174, 2010.
- [3] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic blockstructures," *JASA*, pp. 1077–1087, 2001.
- [4] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *AAAI*, 2006.

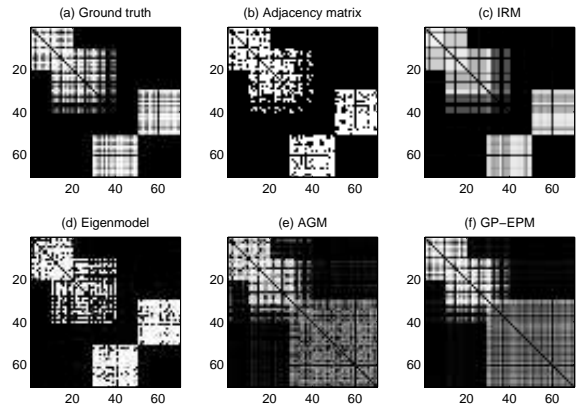


Fig. 2. Comparison of four algorithms' abilities to recover the ground-truth link probabilities using 80% of the pairs of nodes randomly selected from a simulated unweighted undirected relational network that exhibits clear dissortativity. The number of features for the Eigenmodel is set as $K = 4$.

- [5] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, pp. 1981–2014, 2008.
- [6] J. Yang and J. Leskovec, "Community-affiliation graph model for overlapping network community detection," in *ICDM*, 2012.
- [7] J. Yang and J. Leskovec, "Structure and overlaps of ground-truth communities in networks," *ACM Trans. Intell. Syst. Technol.*, pp. 26:1–26:35, 2014.
- [8] K. Miller, M. I. Jordan, and T. L. Griffiths, "Nonparametric latent feature models for link prediction," in *NIPS*, 2009.
- [9] M. Zhou, "Infinite edge partition models for overlapping community detection and link prediction," in *AIS-TATS*, 2015.
- [10] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, 1973.
- [11] J. F. C. Kingman, *Poisson Processes*, Oxford University Press, 1993.
- [12] B. Ball, B. Karrer, and M. E. J. Newman, "Efficient and principled method for detecting communities in networks," *Physical Review E*, 2011.
- [13] M. Zhou, L. Hannah, D. Dunson, and L. Carin, "Beta-negative binomial process and Poisson factor analysis," in *AISTATS*, 2012.
- [14] M. Zhou and L. Carin, "Augment-and-conquer negative binomial processes," in *NIPS*, 2012.
- [15] M. Zhou and L. Carin, "Negative binomial process count and mixture modeling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2015.