Infinite Edge Partition Models for Overlapping Community Detection and Link Prediction: Appendix

A Proof for Lemma 1

Using the law of total expectation, we have

$$\mathbb{E}\left[\sum_{k_1}\sum_{k_2}\lambda_{k_1k_2}\right] = \frac{1}{\beta}\mathbb{E}\left[\xi G(\Omega) + [G(\Omega)]^2 - \sum_k r_k^2\right].$$

Using Campbell's theorem (Kingman, 1993), we have

$$\mathbb{E}\left[\sum_k r_k^2\right] = \int_\Omega\int_0^\infty r^2 r^{-1}e^{-c_0 r}drG_0(d\omega) = \frac{\gamma_0}{c_0^2}.$$

The proof is completed by further using $\mathbb{E}[G(\Omega)] = \gamma_0/c_0$ and $\mathbb{E}[G^2(\Omega)] = \gamma_0^2/c_0^2 + \gamma_0/c_0^2$. □

B MCMC Inference for HGP-EPM

Sample m_{ij}***.*** As in Section 2.2, we sample a latent count for each b_{ij} as

$$(m_{ij}|-) \sim b_{ij}\mathrm{Po}_+\left(\sum_{k_1=1}^K\sum_{k_2=1}^K \phi_{ik_1}\lambda_{k_1k_2}\phi_{jk_2}\right). \quad (18)$$

Sample $m_{ik_1k_2j}$***.*** Using the relationships between the Poisson and multinomial distributions, similar to the derivation in Zhou et al. (2012), we partition the latent count m_{ij} as

$$(\{m_{ik_1k_2j}\}|-) \sim \mathrm{Mult}\left(m_{ij}; \frac{\{\phi_{ik_1}\lambda_{k_1k_2}\phi_{jk_2}\}}{\sum_{k_1}\sum_{k_2}\phi_{ik_1}\lambda_{k_1k_2}\phi_{jk_2}}\right). \quad (19)$$

Note that in each MCMC iteration we store $m_{ik\cdot\cdot}$ and $m_{\cdot k_1k_2\cdot}$ but not necessarily $m_{ik_1k_2j}$ in the memory.

Sample a_i***.*** Using (16) and the data augmentation technique developed in Zhou and Carin (2012, 2015) for the negative binomial distribution, we sample a_i as

$$(\ell_{ik}|-) \sim \sum_{t=1}^{m_{ik\cdot\cdot}} \mathrm{Ber}\left(\frac{a_i}{a_i+t-1}\right),$$
$$(a_i|-) \sim \mathrm{Gam}\left(e_0 + \sum_k \ell_{ik}, \frac{1}{f_0 - \sum_k \ln(1-p'_{ik})}\right), \quad (20)$$

where with a slight abuse of notation, but for added conciseness, we use $x \sim \sum_{t=1}^m \mathrm{Ber}[a/(a+t)]$ to represent $x = \sum_{t=1}^m u_t,\ u_t \sim \mathrm{Ber}[a/(a+t)]$.

Sample ϕ_{ik}***.*** Using (14) and the gamma-Poisson conjugacy, we have

$$(\phi_{ik}|-) \sim \mathrm{Gam}\big[a_i + m_{ik\cdot\cdot}, 1/(c_i + \omega_{ik})\big]. \quad (21)$$

Sample r_k***.*** Similar to the inference of a_i, using (17), we sample r_k as

$$(l_{kk_2}|-) \sim \sum_{t=1}^{m_{\cdot kk_2\cdot}} \mathrm{Ber}\left(\frac{r_k\xi^{\delta_{kk_2}}(r_{k_2})^{1-\delta_{kk_2}}}{r_k\xi^{\delta_{kk_2}}(r_{k_2})^{1-\delta_{kk_2}} + t - 1}\right),$$
$$(r_k|-) \sim \mathrm{Gam}\Bigg[\frac{\gamma_0}{K} + \sum_{k_2} l_{kk_2}, \frac{1}{c_0 - \sum_{k_2}\xi^{\delta_{kk_2}}(r_{k_2})^{1-\delta_{kk_2}}\ln(1-\tilde{p}_{kk_2})}\Bigg]. \quad (22)$$

Sample ξ***.*** We resample the auxiliary variables l_{kk} using the updated r_k and then sample ξ as

$$(\xi|-) \sim \mathrm{Gam}\Bigg[e_0 + \sum_k l_{kk}, \frac{1}{f_0 - \sum_k r_k \ln(1-\tilde{p}_{kk})}\Bigg]. \quad (23)$$

Sample $\lambda_{k_1k_2}$***.*** Using (15) and the gamma-Poisson conjugacy, we have

$$(\lambda_{k_1k_2}|-) \sim \mathrm{Gam}\Big[r_{k_1}\xi^{\delta_{k_1k_2}}(r_{k_2})^{1-\delta_{k_1k_2}} + m_{\cdot k_1k_2\cdot}, 1/(\beta + \theta_{k_1k_2})\Big]. \quad (24)$$

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

Sample γ_0***.*** As show in Lemma 1, the mass parameter γ_0 plays an important role in determining the total sum of the infinite rate matrix $\{\lambda_{k_1k_2}\}$. Our experiments show that it could be used as a tuning parameter to impose one's prior preference on the number of active communities to be discovered. In this paper, we impose a gamma prior as $\gamma_0 \sim \mathrm{Gam}(1,1)$ to let the data infer the posterior of γ_0. We employ an independence chain Metropolis-Hastings algorithm to sample γ_0, with the proposal distribution constructed as

$$Q(\gamma_0^*) = \mathrm{Gam}\left(1 + \sum_k \tilde{l}_k, \frac{1}{1 - \frac{1}{K}\sum_k \ln(1-\tilde{\tilde{p}}_k)}\right), \quad (25)$$

where $(\tilde{l}_k|-) \sim \mathrm{Gam}\left(\sum_{k_2} l_{kk_2}, \gamma_0/K\right)$ and

$$\tilde{\tilde{p}}_k := \frac{-\sum_{k_2}\xi^{\delta_{kk_2}}(r_{k_2})^{1-\delta_{kk_2}}\ln(1-\tilde{p}_{kk_2})}{c_0 - \sum_{k_2}\xi^{\delta_{kk_2}}(r_{k_2})^{1-\delta_{kk_2}}\ln(1-\tilde{p}_{kk_2})}.$$

We accept γ_0^* with probability $\min\{1,\pi\}$, where π is

$$\frac{\prod_{k=1}^K \mathrm{Gam}(r_k;\gamma_0^*/K, 1/c_0)\mathrm{Gam}(\gamma_0^*;1,1)Q(\gamma_0)}{\prod_{k=1}^K \mathrm{Gam}(r_k;\gamma_0/K, 1/c_0)\mathrm{Gam}(\gamma_0;1,1)Q(\gamma_0^*)},$$

which is usually greater than 50% for the networks considered in this paper.

Each iteration of the MCMC for the HGP-EPM proceeds from (18) to (25).

C Gamma Process EPM

The gamma process EPM differs from the HGP-EPM in that it omits inter-community interactions, which leads to a simpler hierarchical model and faster computation at the expense of reduced ability to model stochastic equivalence. It is found to have good performance on assortative networks but not necessarily on disassortative ones.

C.1 Hierarchical Model

The (truncated) gamma process EPM is expressed as

$$\begin{aligned}
&b_{ij} = \mathbf{1}(m_{ij} \geq 1),\\
&m_{ij} = \sum_{k=1}^{K} m_{ijk},\ m_{ijk} \sim \mathrm{Po}\left(r_k \phi_{ik} \phi_{jk}\right),\\
&\phi_{ik} \sim \mathrm{Gam}(a_i, 1/c_i),\ a_i \sim \mathrm{Gam}(e_0, 1/f_0),\\
&r_k \sim \mathrm{Gam}(\gamma_0/K, 1/c_0),\ \gamma_0 \sim \mathrm{Gam}(e_1, 1/f_1),. \qquad (26)
\end{aligned}$$

where the $\mathrm{Gam}(1,1)$ prior is also imposed on c_0 and c_i. As $K \to \infty$, we recover the (exact) gamma process with a finite and continuous base measure. 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

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

C.2 Gibbs Sampling

Let the latent counts $m_{i \cdot k}$ and $m_{\cdot \cdot k}$ be defined as

$$m_{i \cdot k} := \sum_{j=i+1}^{N} m_{ijk} + \sum_{j=1}^{i-1} m_{jik},$$

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

Using the Poisson additive property, we have

$$m_{i \cdot k} \sim \mathrm{Po}\left(r_k \phi_{ik} \sum_{j \neq i} \phi_{jk}\right), \qquad (27)$$

$$m_{\cdot \cdot k} \sim \mathrm{Po}\left(r_k \frac{\sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}{2}\right). \qquad (28)$$

Marginalizing out ϕ_{ik} from (27), we have

$$m_{ik \cdot \cdot} \sim \mathrm{NB}\left(a_i, p'_{ik}\right), \qquad (29)$$

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 (28), we have

$$m_{\cdot \cdot k} \sim \mathrm{NB}\left(\gamma_0/K,\ \tilde{p}_k\right), \qquad (30)$$

where

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

Similar to the inference techniques used in Appendix B, one may exploit (27)-(30) to derive closed-form Gibbs sampling update equations for all model parameters, omitted here for brevity.

D Gamma Process AGM

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

$$\begin{aligned}
&b_{ij} = \mathbf{1}(m_{ij} \geq 1),\\
&m_{ij} = u_{ij} + \sum_{k=1}^{K} m_{ijk},\ m_{ijk} \sim \mathrm{Po}\left(r_k \phi_{ik} \phi_{jk}\right),\\
&u_{ij} \sim \mathrm{Po}(\epsilon),\ \epsilon \sim \mathrm{Gam}(a_0, 1/b_0),\\
&\phi_{ik} \sim \mathrm{Ber}(\pi_i),\ \pi_i \sim \mathrm{Beta}(a_1, b_1),\\
&r_k \sim \mathrm{Gam}(\gamma_0/K, 1/c_0),\ \gamma_0 \sim \mathrm{Gam}(e_1, 1/f_1). \qquad (31)
\end{aligned}$$

We sample r_k, γ_0 and c_0 in the same way we sample them in the gamma process EPM. To sample ϕ_{ik}, one may use (27) as the likelihood, under which ϕ_{ik} is equal to one a.s. if $m_{i \cdot k} > 0$ and is drawn from a Bernoulli distribution if $m_{i \cdot k} = 0$. Gibbs sampling update equations for the other model parameters can be conviniently derived by exploiting conditional conjugacies, omitted here for brevity.