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
\[ E \left[ \sum_k \sum_i \lambda_{ik} \right] = \frac{1}{\beta} E \left[ \xi G(\Omega) + [G(\Omega)]^2 - \sum_k r_k^2 \right]. \]

Using Campbell’s theorem \cite{Kingman1993}, we have
\[ E \left[ \sum_k r_k^2 \right] = \int_0^\infty \int_0^\infty r^2 r^{-1} e^{-c_0 r} dr G_0(\omega) = \frac{\gamma_0}{c_0}. \]

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

B MCMC Inference for HGP-EPM

Sample \( m_{ij} \). As in Section 2.2 we sample a latent count for each \( b_{ij} \) as
\[ (m_{ij} \mid -) \sim b_{ij} \text{Po} \left( \sum_{k_1} \sum_{k_2} \phi_{ik_1 k_2} \lambda_{k_2} \phi_{jk_2} \right). \]  

Sample \( m_{ik, k_2} \). Using the relationships between the Poisson and multinomial distributions, similar to the derivation in \cite{Zhou2012}, we partition the latent count \( m_{ij} \) as
\[ (\{m_{ik, k_2}\} \mid -) \sim \text{Mult} \left( m_{ij}, \sum_{k_2} \phi_{ik_1 k_2} \lambda_{k_2} \phi_{jk_2} \right). \]  

Note that in each MCMC iteration we store \( m_{ik, k_2} \) and \( m_{k_1, k_2} \), but not necessarily \( m_{ik, k_2} \) in the memory.

Sample \( a_t \). Using \cite{Jiang2012} and the data augmentation technique developed in \cite{Zhou2012} for the negative binomial distribution, we sample \( a_t \) as
\[ (\ell_{ik} \mid -) \sim \sum_{t=1}^{m_{ik}} \text{Ber} \left( \frac{a_t}{a_t + \ell_{ik} - 1} \right), \]
\[ (a_t \mid -) \sim \text{Gam} \left( c_0 + \sum_{k=1}^{K} \ell_{ik}, \frac{1}{f_0 - \sum_k \ln(1 - \hat{p}_{ik})} \right), \]
where with a slight abuse of notation, but for added conciseness, we use \( x \sim \sum_{t=1}^{m} \text{Ber} \left[ a_t/(a + t) \right] \) to represent \( x = \sum_{t=1}^{m} a_t, \ a_t \sim \text{Ber} \left[ a_t/(a + t) \right] \).

Sample \( \phi_{ik} \). Using \cite{Ghahramani2015} and the gamma-Poisson conjugacy, we have
\[ (\phi_{ik} \mid -) \sim \text{Gam} \left( a_i + m_{ik} \cdot, 1/(c_i + \omega_{ik}) \right). \]

Sample \( r_k \). Similar to the inference of \( a_t \), we sample \( r_k \) as
\[ (l_{kk} \mid -) \sim \sum_{t=1}^{m_{kk}} \text{Ber} \left( \frac{r_k \xi_{r_k} (r_{kk})^{1-\delta_{kk}}}{r_k \xi_{r_k} (r_{kk})^{1-\delta_{kk}} + t - 1} \right), \]
\[ (r_k \mid -) \sim \text{Gam} \left( \gamma_0 + \sum_{k=1}^{K} l_{kk}, \frac{1}{c_0 - \sum_{k} \xi_{r_k} (r_{kk})^{1-\delta_{kk}} \ln(1 - \hat{p}_{kk})} \right). \]  

Sample \( \xi \). We resample the auxiliary variables \( l_{kk} \) using the updated \( r_k \) and then sample \( \xi \) as
\[ (\xi \mid -) \sim \text{Gam} \left( c_0 + \sum_{k} l_{kk}, \frac{1}{f_0 - \sum_k r_k \ln(1 - \hat{p}_{kk})} \right). \]  

Sample \( \lambda_{k_1, k_2} \). Using \cite{Ghahramani2015} and the gamma-Poisson conjugacy, we have
\[ (\lambda_{k_1, k_2} \mid -) \sim \text{Gam} \left( r_k \xi_{r_k} (r_{kk})^{1-\delta_{kk}} + m_{k_1, k_2}, 1/\beta + \theta_{k_1, k_2} \right). \]  

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

Sample \( \gamma_0 \). As show in Lemma 4 the mass parameter \( \gamma_0 \) plays an important role in determining the total sum of the infinite rate matrix \( \{\lambda_{k_1, k_2}\} \). 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 \( \gamma_0 \sim \text{Gam}(1, 1) \) to let the data infer the posterior of \( \gamma_0 \). We employ an independence chain Metropolis-Hastings algorithm to sample \( \gamma_0 \), with the proposal distribution constructed as
\[ Q(\gamma_0') = \text{Gam} \left( 1 + \sum_k \tilde{t}_k, \frac{1}{1 - \pi} \sum_k \ln(1 - \tilde{p}_{kk}) \right), \]  
where \( (\tilde{t}_k \mid -) \sim \text{Gam} \left( \sum_{k} l_{kk}, \gamma_0/K \right) \) and \( \tilde{p}_{kk} := \frac{- \sum_{k_2} \xi_{r_k} \xi_{r_{kk}} (r_{kk})^{1-\delta_{kk}} \ln(1 - \hat{p}_{kk})}{c_0 - \sum_{k} \xi_{r_k} (r_{kk})^{1-\delta_{kk}} \ln(1 - \hat{p}_{kk})} \).

We accept \( \gamma_0' \) with probability \( \min(1, \pi) \), where \( \pi \) is
\[ \prod_{k=1}^{K} \text{Gam}(r_{kk}; \gamma_0/K, 1/c_0) \text{Gam}(\gamma_0'; 1, 1) Q(\gamma_0'), \]  
which is usually greater than 50% for the networks considered in this paper.

Each iteration of the MCMC for the HGP-EPM proceeds from \cite{Jiang2012} to \cite{Ghahramani2015}.
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

\[ b_{ij} = 1(m_{ij} \geq 1), \]

\[ m_{ij} = \sum_{k=1}^{K} m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}), \]

\[ \phi_{ik} \sim \text{Gam}(a_i, 1/c_i), \quad a_i \sim \text{Gam}(\epsilon_0, 1/f_0), \]

\[ r_k \sim \text{Gam}(\gamma_0/K, 1/c_0), \quad \gamma_0 \sim \text{Gam}(\epsilon_1, 1/f_1). \tag{26} \]

where the 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

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

C.2 Gibbs Sampling

Let the latent counts \( m_{i,k} \) and \( m_{.,k} \) be defined as

\[ m_{i,k} := \sum_{j=i+1}^{N} m_{ijk} + \sum_{j=1}^{i-1} m_{ijk}, \]

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

Using the Poisson additive property, we have

\[ m_{i,k} \sim \text{Po} \left( r_k \phi_{ik} \sum_{j \neq i} \phi_{jk} \right), \tag{27} \]

\[ m_{.,k} \sim \text{Po} \left( r_k \sum_{i} \sum_{j \neq i} \phi_{ik} \phi_{jk} / 2 \right). \tag{28} \]

Marginalizing out \( \phi_{ik} \) from \( \phi_{ik} \)

\[ m_{ik} \sim \text{NB}(a_i, p_{ik}'), \tag{29} \]

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 \( \phi_{ik} \)

\[ m_{.,k} \sim \text{NB}(\gamma_0/K, \tilde{p}_k), \tag{30} \]

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 \( \phi_{ik} \)

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

\[ b_{ij} = 1(m_{ij} \geq 1), \]

\[ m_{ij} = u_{ij} + \sum_{k=1}^{K} m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}), \]

\[ u_{ij} \sim \text{Po}(\epsilon), \quad \epsilon \sim \text{Gam}(a_0, 1/b_0), \]

\[ \phi_{ik} \sim \text{Ber}(\pi_i), \quad \pi_i \sim \text{Beta}(a_1, b_1), \]

\[ r_k \sim \text{Gam}(\gamma_0/K, 1/c_0), \quad \gamma_0 \sim \text{Gam}(\epsilon_1, 1/f_1). \tag{31} \]

We sample \( r_k, \gamma_0 \) and \( \epsilon \) in the same way we sample them in the gamma process EPM. To sample \( \phi_{ik} \), one may use \( \phi_{ik} \)

to as the likelihood, under which \( \phi_{ik} \)
is equal to one a.s. if \( m_{i,k} > 0 \) and is drawn from a Bernoulli distribution if \( m_{i,k} = 0 \). Gibbs sampling update equations for the other model parameters can be conveniently derived by exploiting conditional conjugacies, omitted here for brevity.