2016 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 13–16, 2016, SALERNO, ITALY

THE INFLUENCE OF HYPER-PARAMETERS IN THE INFINITE RELATIONAL MODEL

Kristoffer J. Albers, Morten Mørup, Mikkel N. Schmidt

Department of applied Mathematics and Computer Science, Section for Cognitive Science
Technical University of Denmark

ABSTRACT

The infinite relational model (IRM) is a Bayesian nonparametric stochastic block model; a generative model for random networks parameterized for unipartite undirected networks by a partition of the node set and symmetric matrix of inter-partion link probabilities. The prior for the node clusters is the Chinese restaurant process, and the link probabilities are, in the most simple setting, modeled as iid. with a common symmetric Beta prior. More advanced priors such as separate asymmetric Beta priors for links within and between clusters have also been proposed. In this paper we investigate the importance of these priors for discovering latent clusters and for predicting links. We compare fixed symmetric priors and fixed asymmetric priors based on the empirical distribution of links with a Bayesian hierarchical approach where the parameters of the priors are inferred from data. On synthetic data, we show that the hierarchical Bayesian approach can infer the prior distributions used to generate the data. On real network data we demonstrate that using asymmetric priors significantly improves predictive performance and heavily influences the number of extracted partitions.

Index Terms— Infinite relational model, hyperparameter inference, link-prediction, Bayesian nonparametrics.

1. INTRODUCTION

Many systems, both naturally occurring and engineered, can be described as complex networks. These include biological systems such as functional and structural brain connectivity, social and economic behaviour as well as infrastructure such as power grids, communication and transport networks.

Network science is concerned with developing theoretical and practical methods for modelling and quantifying hidden structure in complex networks, and plays a prominent role in acquisition of knowledge within many different research areas. One way to extract information from a complex network is to cluster the network into groups of nodes that have similar structural connectivity patterns.

The most prominent statistical tool for clustering network data is the stochastic block model [1, 2], which is a probabilistic generative model for random networks. It models a network using a latent clustering of the network nodes. The probabilities of links between two nodes depend only on their cluster assignments and a link probability parameter which is defined for each pair of clusters. In the infinite relational model (IRM) [3, 4] the prior for the cluster structure is the Chinese restaurant process: A stochastic process which defines a distribution over partitions. The CRP provides a nonparametric Bayesian mechanism for learning the number of clusters that best fit the observed network.

The prior for the link probability parameters are, in the most simple setting, chosen as a symmetic Beta distribution. Without any further information available, a vague symmetric prior such as a Beta$(\frac{1}{2}, \frac{1}{2})$ (arcsine) or Beta$(1, 1)$ (uniform) distribution is suited. If more prior information is available, such as beliefs about the overall link density of the network or belief that the link densities within and between clusters are different, using a more elaborate prior is relevant.

In this paper we investigate how different prior constructions in the IRM influence the learned clustering structure as well as the predictive performance of the fitted model. In particular, we demonstrate that using an asymmetric informative prior leads to superior predictive performance compared to other constructions.

2. METHOD

2.1. Review of the infinite relational model

Let A be the adjacency matrix of a simple graph. Including separate parameters for the Beta priors for links within and between clusters, the infinite relational model (IRM) is [3] is given by the following generative

This project was supported by the Lundbeck Foundation, grant nr. R105-9813

978-1-5090-0746-2/16/$31.00

process:

$$z \sim \text{CRP}(\alpha) \quad \text{Clusters} \tag{1}$$
$$\eta_{\ell\ell} \sim \text{Beta}(\beta_{\text{w}}^{+}, \beta_{\text{w}}^{-}) \quad \text{Link probabilities within} \tag{2}$$
$$\eta_{\ell m} \sim \text{Beta}(\beta_{\text{b}}^{+}, \beta_{\text{b}}^{-}) \quad \text{—} \quad \text{between} \tag{3}$$
$$A_{ij} \sim \text{Bernoulli}(\eta_{z_i,z_j}) \quad \text{Observed links} \tag{4}$$

The prior for the clustering is a Chinese Restaurant Process (CRP), which allows the model to automatically infer an appropriate number of clusters from data. The probability of observing a link between two nodes i and j follows a Bernoulli distribution, where the parameter η_{z_i,z_j} depends on the cluster assignments of the two nodes. In our setup, the link probabilities $\eta_{\ell m}$ within and between clusters follow separate Beta distributions.

We investigate the following different prior constructions: A joint symmetric prior with only one parameter, $\beta = \beta_{\text{w}}^{+} = \beta_{\text{w}}^{-} = \beta_{\text{b}}^{+} = \beta_{\text{b}}^{-}$ as proposed in [3], a joint asymmetric prior with two parameters, $\beta = \{\beta_{\text{w}}^{+} = \beta_{\text{b}}^{+}, \beta_{\text{w}}^{-} = \beta_{\text{b}}^{-}\}$ as used for block modeling in [5], and separate asymmetric priors for link probabilities within and between clusters with four parameters $\beta = \{\beta_{\text{w}}^{+}, \beta_{\text{w}}^{-}, \beta_{\text{b}}^{+}, \beta_{\text{b}}^{-}\}$.

Because the Beta prior is conjugate to the Bernoulli likelihood, the link probabilities ($\eta_{\ell m}$-parameters) can be marginalized analytically, revealing the following joint distribution

$$P(A, z|\alpha, \beta) = \text{CRP}(z|\alpha) \prod_{\ell} \frac{\text{B}(N_{\ell\ell}^{+} + \beta_{\text{w}}^{+}, N_{\ell\ell}^{-} + \beta_{\text{w}}^{-})}{\text{B}(\beta_{\text{w}}^{+}, \beta_{\text{w}}^{-})} \prod_{\ell<m} \frac{\text{B}(N_{\ell m}^{+} + \beta_{\text{b}}^{+}, N_{\ell m}^{-} + \beta_{\text{b}}^{-})}{\text{B}(\beta_{\text{b}}^{+}, \beta_{\text{b}}^{-})}.6 \tag{5}$$

Here $N_{\ell m}^{+}$ and $N_{\ell m}^{-}$ are the number of links and non-links between cluster ℓ and m, and $\text{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the Beta function. We can then further place priors on the parameters β in a Bayesian hierarchical manner. In the following we employ improper flat priors, such that the joint distribution can be written as $P(A, z, \beta|\alpha) \propto P(A, z|\alpha, \beta)$.

2.2. Inference using Markov chain Monte Carlo

To solve the clustering problem, we condition on the observed network to find the posterior distribution of the clustering by, $P(z|A, \alpha)$. To infer the clustering we employ two different transition kernels: Gibbs sampling and split-merge sampling. In Gibbs sampling, we loop over each node in the network: For each node we evaluate the posterior distribution when assigning the node to each of the existing clusters or a new empty cluster, conditioned on all the other node assignments $z_{\setminus i}$. The node is then reassigned based on the probability distribution of possible node assignments. The probability of assigning node i to cluster m is then given by:

$$P(z_i = m|A, z_{\setminus i}, \alpha, \beta) = \frac{P(z_i = m, A, z_{\setminus i}|\alpha, \beta)}{\sum_{\ell} P(z_i = \ell, A, z_{\setminus i}|\alpha, \beta)}, \tag{6}$$

where ℓ in the sum ranges over all existing groups and a new empty group.

In split-merge sampling [6], two nodes in the network are selected uniformly at random. If the nodes are in the same cluster it is proposed to be split, otherwise the clusters of the two nodes are proposed to be merged. The proposals are accepted or rejected based on the Metropolis-Hastings acceptance probability:

$$P(z^*|z) = \min\left[1, \frac{P(z^*, A|\alpha, \beta^+, \beta^-) q(z|z^*)}{P(z, A|\alpha, \beta^+, \beta^-) q(z^*|z)}\right], \tag{7}$$

where $q()$ is the transition probability and z^* is the proposed clustering. To generate the proposed split state, the two selected nodes are placed in separate clusters and the remaining nodes in the cluster are allocated randomly between the two. A number of rounds of Gibbs sampling is performed (restricted to the nodes in the two clusters), and the final proposal and its transition probability is then given by the final Gibbs round. For a split configuration, $q()$ is given by the product of the individual transition probabilities of repartitioning each node from the launch state to the split configuration. As there is only one way of merging two clusters, the transition probability for merging clusters is always one.

To infer the parameters of the prior, β, we use a Metropolis-Hastings procedure: We sample each parameter in turn using a Gaussian proposal distribution centered on the current value and with variance $\sigma^2 = 1$.

2.3. Data and experiments

As a generative model, IRM can be used to generate synthetic data. We use this to investigate how well IRM with the different prior configurations is capable of inferring the underlying true parameters and clustering on a synthetic network. We further investigate IRM with the various prior configurations on three real world network data of various sizes and from different domains. These networks are presented in table 1.

We consider the following three prior constructions:

Prior	Parameter(s)
Joint symmetric [3]	One: $\beta = \beta_{\text{w}}^{+} = \beta_{\text{w}}^{-} = \beta_{\text{b}}^{+} = \beta_{\text{b}}^{-}$.
Joint asymmetric [5]	Two: $\beta^{+} = \beta_{\text{w}}^{+} = \beta_{\text{b}}^{+}$, $\beta^{-} = \beta_{\text{w}}^{-} = \beta_{\text{b}}^{-}$.
Separate asymmetric	Four: β_{w}^{+}, β_{w}^{-}, β_{b}^{+}, β_{b}^{-}.

Network	Nodes J	Links L	Link ratio $a^+ = \frac{L}{P}$	Nonlink ratio $a^- = \frac{P-L}{P}$	Description
USAir	332	2123	0.0387	0.9613	Traffic network of airlines [7], binarized as in [8].
Hagmann	998	37,926	0.0762	0.9238	Average of five brain connectivity networks in [9].
Facebook	4,039	88,234	0.0108	0.9892	Social circles from Facebook [10, 11].

Table 1: Topology for the examined networks. P denotes the total possible links, computed as $P = J(J-1)/2$.

We compare sampling for these constructions using the following fixed symmetric, uninformed priors:
$\beta^+ = \beta^- = \{\ 0.05\ ,\ 0.5\ ,\ 1\ ,\ 5\ \}$,
and using fixed values based on the link- and nonlink-ratio found empirically in the network data (shown in table 1):
$\beta^+ = c \cdot a^+ \quad , \quad \beta^- = c \cdot a^- \ , \quad \text{for } c = \{0.1, 1, 2, 10\}.$

We use the following MCMC sampling procedure for 1000 iterations, where the first 750 iterations are discarded as burn in. Each iteration consists of: One Gibbs sweep over all nodes followed by 10 split-merge proposals, each with three restricted Gibbs sweeps. When sampling the hyper-parameters, 10 proposals for each of the sampled parameters are then performed in each iteration. We consider the concentration parameter α of the CRP fixed at $\log(J)$, where J is the number of nodes in the network. For sampling the β parameters we use a Gaussian distribution with variance 1.

To compare the clustering found by IRM with the ground truth of synthetic data, we use normalized mutual information, $\mathrm{NMI}(z, z') = \frac{2\mathrm{I}(z,z')}{\mathrm{H}(z)+\mathrm{H}(z')}$, where $\mathrm{H}(z)$ is the entropy of the clustering z.

To compare the sampling procedures on real world networks, we evaluate the predictive performance based on the inferred clusterings. When sampling a real world network, we exclude 10 percent of the links as hold out data and measure the predictive performance by evaluating the area under the receiver operating characteristic curve (AUC) when predictions are made for the hold out data [12]. For a given clustering, we compute the expected probability of a link between two clusters as:

$$\langle \eta_{\ell m} \rangle = \frac{N^+_{\ell m} + \beta^+_{\mathrm{b}}}{N^+_{\ell m} + N^-_{\ell m} + \beta^+_{\mathrm{b}} + \beta^-_{\mathrm{b}}} \tag{8}$$

$$\langle \eta_{\ell \ell} \rangle = \frac{N^+_{\ell \ell} + \beta^+_{\mathrm{w}}}{N^+_{\ell \ell} + N^-_{ll} + \beta^+_{\mathrm{w}} + \beta^-_{\mathrm{w}}} \tag{9}$$

The expected probability of a link between two nodes is considered the link probability between the two clusters, the nodes belongs to: $\langle A_{ij} \rangle = \langle \eta_{z_i, z_j} \rangle$. When examining the AUC, we compare averaging over the last 250 iterations of the MCMC sampling and using the estimate for the last iteration only.

For fixed asymmetric priors we use the a^+ and a^- ratio based on the entire network. Instead of modelling the hold out data as missing [13], we treat it as non-existing links in the network [14, 15]. This is a more conservative link prediction strategy that is more prone to overfitting and can hence easier show whether IRM will exhibit overfitting issues when sampling the hyper-parameters.

Hyperparameters	NMI	NOCs
Fixed, symmetric		
$\beta^+ = 0.05\ ,\ \beta^- = 0.05$	0.9502 ± 0.0083	20
$\beta^+ = 0.1\ ,\ \beta^- = 0.1$	0.9789 ± 0.0017	26
$\beta^+ = 0.5\ ,\ \beta^- = 0.5$	0.9502 ± 0.0083	20
$\beta^+ = 1\ ,\ \beta^- = 1$	0.9532 ± 0.0076	20
$\beta^+ = 5\ ,\ \beta^- = 5$	0.9502 ± 0.0083	20
Fixed, empiric		
$\beta^+ = 0.1 \cdot a^+\ ,\ \beta^- = 0.1 \cdot a^-$	0.9903 ± 0.0014	34
$\beta^+ = 1 \cdot a^+\ ,\ \beta^- = 1 \cdot a^-$	0.9913 ± 0.0017	33
$\beta^+ = 2 \cdot a^+\ ,\ \beta^- = 2 \cdot a^-$	0.9875 ± 0.0012	30
$\beta^+ = 10 \cdot a^+\ ,\ \beta^- = 10 \cdot a^-$	0.9652 ± 0.0030	24
Fixed, ground truth		
$\beta^+ = 0.1\ ,\ \beta^- = 1.5$	0.9923 ± 0.0000	35
Inferred		
$\beta = \beta^+_{\mathrm{w}} = \beta^-_{\mathrm{w}} = \beta^+_{\mathrm{b}} = \beta^-_{\mathrm{b}}$	0.9778 ± 0.0016	28
$\beta^+ = \beta^+_{\mathrm{w}} = \beta^+_{\mathrm{b}}, \beta^- = \beta^-_{\mathrm{w}} = \beta^-_{\mathrm{b}}$	0.9921 ± 0.0005	35
$\beta^+_{\mathrm{w}},\ \beta^-_{\mathrm{w}},\ \beta^+_{\mathrm{b}},\ \beta^-_{\mathrm{b}}$	0.9919 ± 0.0005	35

Table 2: Normalized Mutual Information (NMI) and number of components (NOCs) found by sampling IRM on a synthetic network with $J = 500$ nodes and 17.324 links, generated from an IRM with $\alpha = log(J)$, $\beta^+ = \beta^+_{\mathrm{w}} = \beta^+_{\mathrm{b}} = 0.1$ and $\beta^- = \beta^-_{\mathrm{w}} = \beta^-_{\mathrm{b}} = 1.5$. The true clustering contains 35 components. The results are based on five random restarts each averaged over the last 250 iterations of the sampling procedure.

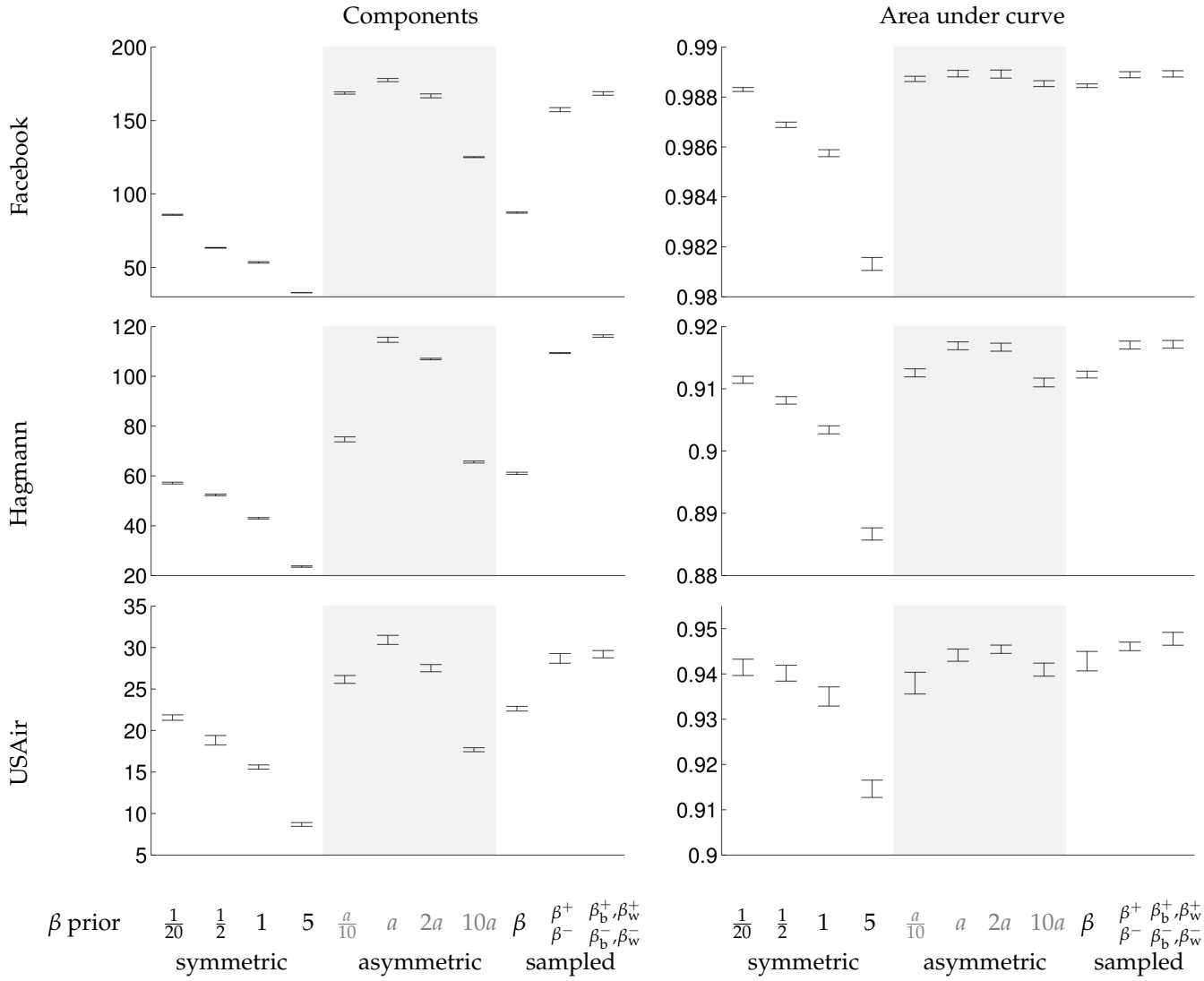

Fig. 1: Number of components and AUC for the real world networks for the different prior constructions on β. The sampling procedures were performed with 1000 sweeps for five restarts on five different hold out set for each network. The number of components was averaged over the last 250 sweep. AUC is computed from the averaged clustering over the last 250 sweeps. Errorbars indicate the standard deviation of the mean over five restarts, i.e. std/$\sqrt{5}$.

Network	Density	Joint asymmetric		Separate asymmetric			
		Mean cluster density	$\rho = \frac{\beta^+}{\beta^+ + \beta^-}$	Mean cluster density within	Mean cluster density between	$\rho_w = \frac{\beta^+_w}{\beta^+_w + \beta^-_w}$	$\rho_b = \frac{\beta^+_b}{\beta^+_b + \beta^-_b}$
USAir	0.0348	0.2100	0.1787	0.4556	0.1676	0.3819	0.1451
Hagmann	0.0686	0.0788	0.0780	0.8336	0.0689	0.8200	0.0674
Facebook	0.0097	0.0253	0.0231	0.4968	0.0202	0.4639	0.0197

Table 3: The inferred values of the hyperparameters compared to mean cluster densities of the inferred clustering and the density of the training network. Results are averaged for the last 100 sweeps of the sampling procedures for a single run.

3. RESULTS AND DISCUSSION

3.1. Synthetic data

Table 2 shows the average normalized mutual information as well as average number of inferred components when sampling with the different prior constructions in a synthetic network generated from an IRM. The network is generated with a joint assymetric prior for $\beta^+ = 0.1$, $\beta^- = 1.5$ and the generated clustering contains 35 components. When using a fixed symmetric prior the model under-estimates the number of components, while using fixed asymmetric priors allows the model to better adapt to the network. The model can correctly identify values for β that corresponds to a high NMI if the hyperparameters are fixed appropriately.

Sampling two or four parameters both correctly identify the number of components. The inferred clusterings further seem to have a similar NMI, just as good as found when using the ground truth hyperparameter values for the generated network. Sampling the hyperparameters, on average gives the following values for the three hyperparameter settings:

A:	$\beta = 0.106 \pm 0.003$
B:	$\beta^+ = 0.096 \pm 0.003$, $\beta^- = 1.628 \pm 0.043$
C:	$\beta_{\mathrm{b}}^+ = 0.097 \pm 0.004$, $\beta_{\mathrm{b}}^- = 1.718 \pm 0.076$ $\beta_{\mathrm{w}}^+ = 0.132 \pm 0.005$, $\beta_{\mathrm{w}}^- = 1.437 \pm 0.100$.

Thus, hyperparameter values fairly close to those used to generate the network are identified when sampling two or four parameters.

3.2. Real world networks

Figure 1 shows the results with the various prior configurations on the three real world networks. Ten percent of the networks were omitted as holdout data for computing the AUC. Supporting the findings from sampling on synthetic data, the model performs better when based on asymmetric priors. It identifies more components with a higher AUC. While using fixed asymmetric priors based on network topology can perform well it requires an appropriate scale of the parameters is chosen, which might depend on the particular network. When sampling the asymmetric hyper parameters separately between and within components, the model is capable of identifying more components retaining the same high AUC as sampling the parameters as joint asymmetric.

Figure 2 shows the accumulative sizes of the inferred clusterings when sampling with the different prior configurations. The additional clusters found when using asymmetric priors are not dominated by small or singleton clusters, suggesting that they contain relevant

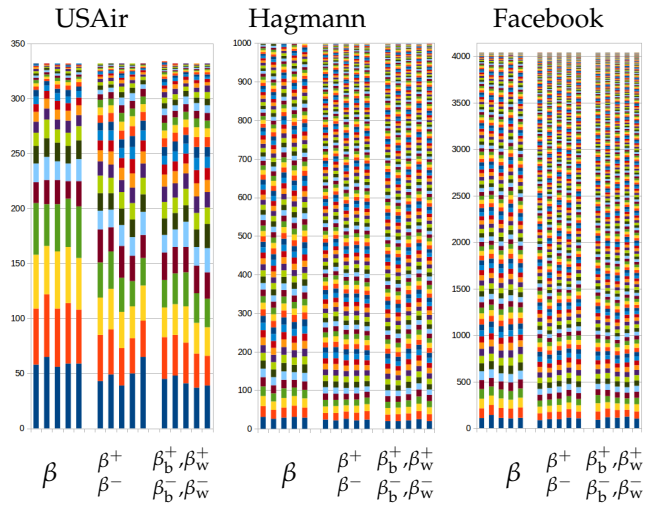

Fig. 2: Accumulative cluster sizes for five restarts for the three sampling procedures on the same training data. The results are based on the clustering inferred after the last sweep of the sampling procedure.

structural information about the network. This further strengthens the indication that IRM describes the structural properties of the network on a more detailed block-level without introducing additional overfitting to the training data.

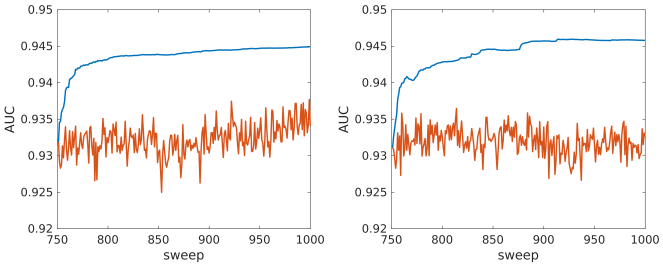

Fig. 3: The progress of AUC for the USAir network. For sweep i, the blue line shows AUC based on the average clustering for sweep 750 to i (Bayesian average). The red line shows AUC for the particular clustering at sweep i. Results are averaged for five restarts on the same hold out data.

Table 4 shows the average AUC computed when basing the link probabilities on the final clustering versus on the average of the last 250 sweeps of the sampling procedure. Using the Bayesian average performs significantly better. The effect is further illustrated in figure 3 for the USAir network.

Table 3 compares the inferred hyperparameters with the network density and the link density of the inferred

	Sampled parameters	AUC Sweep 1000	Averaged
USAir	β	0.9341 ± 0.005	0.9449 ± 0.002
	β^+, β^-	0.9286 ± 0.007	0.9467 ± 0.003
	$\beta_b^+, \beta_b^-, \beta_w^+, \beta_w^-$	0.9327 ± 0.004	0.9458 ± 0.004
Hagmann	β	0.9112 ± 0.001	0.9127 ± 0.001
	β^+, β^-	0.9162 ± 0.001	0.9184 ± 0.001
	$\beta_b^+, \beta_b^-, \beta_w^+, \beta_w^-$	0.9164 ± 0.001	0.9186 ± 0.001
Facebook	β	0.9876 ± 0.000	0.9885 ± 0.000
	β^+, β^-	0.9878 ± 0.000	0.9890 ± 0.000
	$\beta_b^+, \beta_b^-, \beta_w^+, \beta_w^-$	0.9876 ± 0.000	0.9891 ± 0.000

Table 4: Comparing AUC, computed for the averaged clustering of the last 250 sweeps and computed for the last sweep. The results are the average for five different runs using a single hold out data set.

clusterings. This clearly indicates that sampling the hyperparameters reflects learning block level cluster densities, rather than simply reflecting the overall network link density.

4. CONCLUSION

We have investigated the influence of various hyperparameter constructions in the infinite relational model for clustering complex real world networks. We find that the hyper-parameter construction significantly influences the number of inferred components as well as the predictive performance of the model. We have demonstrated that using informed asymmetric priors can improve predictive performance compared to uninformed symmetric priors, and that the approach proposed in [3] assuming a symmetric prior $\beta = \beta^+ = \beta^-$ that is inferred was outperformed in link prediction by inferred asymmetric priors, providing a more refined block-structure. Separately sampling parameters for within and between components allowed the model to account for even more components without indications of overfitting to the training data. For the examined networks, sampling asymmetric hyper-parameters in IRM performs on par with using joint assymetric priors fixed to reflect the network density for an adequately chosen scale c. However, we find that inferring the hyper-parameters does not simply reflect the density of the network, but reflects the average link densities at the levels of the identified blocks which cannot be estimated in advance from the network.

5. REFERENCES

[1] Harrison C White, Scott A Boorman, and Ronald L Breiger, "Social structure from multiple networks. i. blockmodels of roles and positions," *American journal of sociology*, pp. 730–780, 1976.

[2] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[3] Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda, "Learning systems of concepts with an infinite relational model," in *Proceedings of the national conference on artificial intelligence*, 2006, vol. 21, p. 381.

[4] Zhao Xu, Volker Tresp, Kai Yu, and Hans-Peter Kriegel, "Learning infinite hidden relational models," *Uncertainity in Artificial Intelligence (UAI2006)*, 2006.

[5] Tue Herlau, Mikkel N Schmidt, and Morten Mørup, "Infinite-degree-corrected stochastic block model," *Physical Review E*, vol. 90, no. 3, pp. 032819, 2014.

[6] Sonia Jain and Radford M Neal, "A split-merge markov chain sampling algorithm for bayesian mixture models," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 158–182, 2004.

[7] Vladimir Batagelj and Andrej Mrvar, "Pajek datasets," *http://vlado.fmf.uni-lj.si/pub/networks/data/*, 2006.

[8] Linyuan Lü, Liming Pan, Tao Zhou, Yi-Cheng Zhang, and H Eugene Stanley, "Toward link predictability of complex networks," *Proceedings of the National Academy of Sciences*, vol. 112, no. 8, pp. 2325–2330, 2015.

[9] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biology*, vol. 6, no. 7, pp. 1479–1493, 2008.

[10] Julian J McAuley and Jure Leskovec, "Learning to discover social circles in ego networks.," in *NIPS*, 2012, vol. 2012, pp. 548–56.

[11] Jure Leskovec and Andrej Krevl, "Snap datasets: Stanford large network dataset collection," *https://snap.stanford.edu/*.

[12] Jin Huang and Charles X Ling, "Using auc and accuracy in evaluating learning algorithms," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 3, pp. 299–310, 2005.

[13] Kurt Miller, Michael I Jordan, and Thomas L Griffiths, "Nonparametric latent feature models for link prediction," in *Advances in neural information processing systems*, 2009, pp. 1276–1284.

[14] David Liben-Nowell and Jon Kleinberg, "The link-prediction problem for social networks," *Journal of the American society for information science and technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[15] Aaron Clauset, Cristopher Moore, and Mark EJ Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, no. 7191, pp. 98–101, 2008.