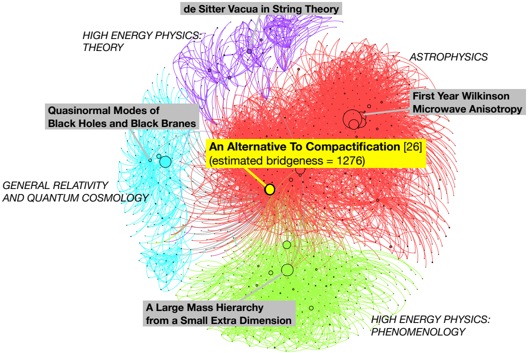
Efficient Discovery of Overlapping Communities in Large Networks

Prem Gopalan, David M. Blei.

Department of Computer Science, Princeton University

Motivation

- Community detection is important to understand the structural and functional properties of real networks.
- Communities are known to overlap significantly.
- Real networks are large.

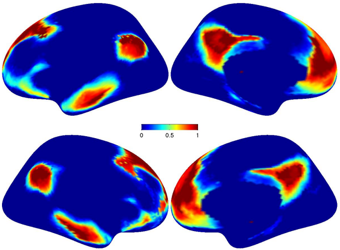


Community overlap around an article in a citation network from arXiv.org.

Motivation

A Bayesian approach to community detection

- Posit a probabilistic model of networks where each node can belong to multiple communities.
- Analyze a network by computing the posterior, the conditional distribution of the hidden communities given the observed network.



Approximate posterior assignment of neurons to a community of neuronal activity.

Motivation

- Bayesian Inference provides a powerful, principled approach to fitting sophisticated probabilistic models of networks to data.
- Unfortunately, inference algorithms for finding overlapping communities often scale quadratic in the number of nodes.
- **This talk**: Algorithms to scale approximate posterior inference to large networks by subsampling the data.
- This is a principled variational inference algorithm for a mixed-membership stochastic blockmodel.

A model of communities in assortative networks

The generative process of the assortative mixed-membership stochastic blockmodel is

1. For each community k, draw community strength $\beta_k \sim \mathrm{Beta}(\eta)$.
2. For each node a, draw community memberships $\theta_a \sim \mathrm{Dirichlet}(\alpha)$.
3. For each pair of nodes a and b,
   - 3.1 Draw interaction indicator $z_{a \rightarrow b} \sim \theta_a$.
   - 3.2 Draw interaction indicator $z_{a \leftarrow b} \sim \theta_b$.
   - 3.3 Draw link $y_{ab} \sim \mathrm{Bernoulli}(r)$, where

$$r = \begin{cases} \beta_k & \text{if } z_{a \rightarrow b,k} = z_{a \leftarrow b,k} = 1, \\ \epsilon & \text{if } z_{a \rightarrow b} \neq z_{a \leftarrow b}. \end{cases}$$

A model of communities in assortative networks

The conditional probability of a connection is

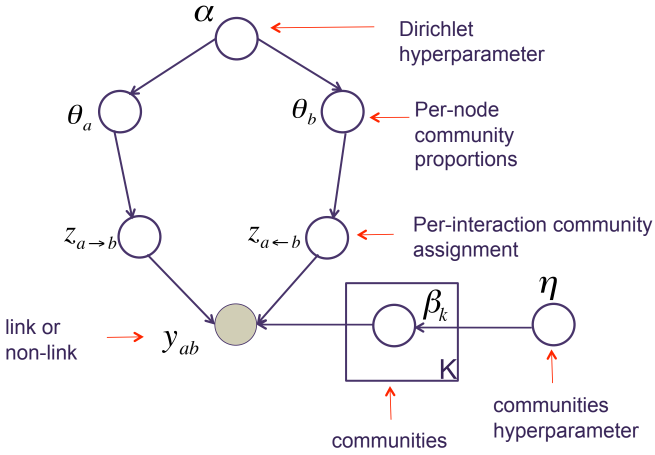
$$p(y_{ij} = 1 | \theta_a, \theta_b, \beta_k) = \sum_{k=1}^{K} \theta_{ak} \theta_{bk} \beta_k \tag{1}$$

θ_a: [princeton:0.7, tennis:0.1, networks:0.2]
θ_b: [princeton:0.8, tennis:0.1, networks:0.1]

- The model is based on classical stochastic blockmodel (Nowicki et al., 2001).
- A type of mixed-membership stochastic blockmodel (Airoldi et al., 2008) that captures assortativity.

A Bayesian hierarchical model

The graphical model of a 2-node network with K communities.



This is a mixed-membership stochastic blockmodel (Airoldi et al., 2008).

Interpreting the posterior distribution

- The posterior distribution of the community structure,

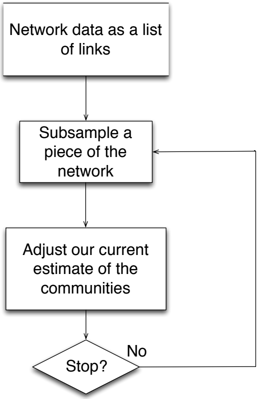
$$p(\boldsymbol{\theta}, \mathbf{z} \mid \mathbf{y}) = p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{y}) / p(\mathbf{y}).$$

- The marginal probability of the data,

$$p(\mathbf{y}) = \int_{\boldsymbol{\theta}} \sum_{\mathbf{z}} p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{y}).$$

- This marginal cannot be easily computed. We approximate the posterior using an efficient algorithm.

Computational structure of our inference algorithm



- This is a *stochastic inference algorithm* (Hoffman et al., 2012, Gopalan et al., 2012).
- Algorithm combines mean-field variational inference and stochastic optimization.

Mean-field variational inference

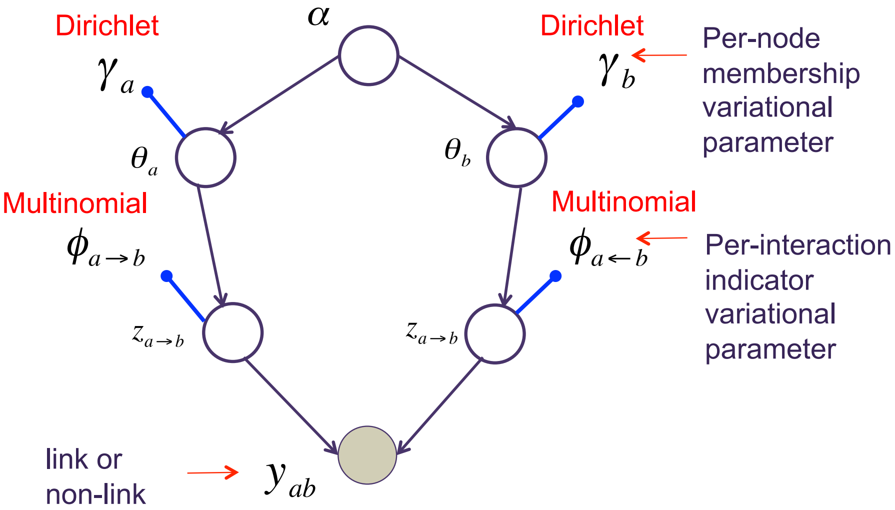
- Approximate the posterior by defining and fitting a parametrized distribution.
- This distribution is close to the posterior in Kullback-Leibler divergence.
- In our inference, the mean-field variational family is,

$$q(\boldsymbol{\theta}, \mathbf{z}) = \prod_{n=1}^{N} q(\theta_n \,|\, \gamma_n) \prod_{i<j} q(z_{i \to j} \,|\, \phi_{i \to j}) \, q(z_{i \leftarrow j} \,|\, \phi_{i \leftarrow j}) \,.$$

- The variational problem is to solve

$$q^*(\boldsymbol{\theta}, \mathbf{z}) = \arg\min_{\boldsymbol{\gamma}, \boldsymbol{\phi}} \mathrm{D}(q(\boldsymbol{\theta}, \mathbf{z}) || p(\boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{y})).$$

The variational family for the 2-node network



Batch variational inference for the MMSB

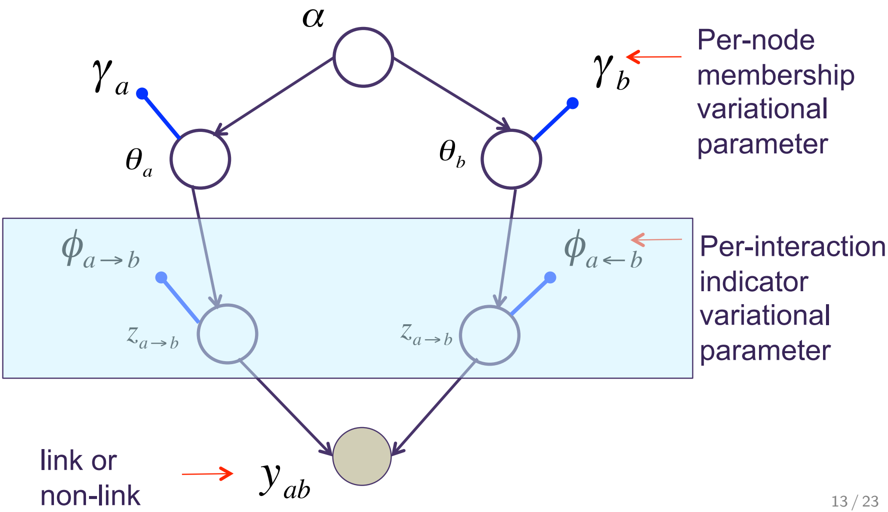
The variational objective is

$$\mathcal{L}(\gamma, \phi) = \mathrm{E}\left[\log p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{y})\right] + \mathrm{H}\left[q(\boldsymbol{\theta}, \mathbf{z})\right],$$

- Maximizing this objective is equivalent to minimizing KL-divergence.
- First term captures how well $q(\boldsymbol{\theta}, \mathbf{z})$ is likely under the model.
- Second term encourages the variational distribution to be entropic.
- Traditional variational inference uses coordinate ascent which scales as quadratic in the number of nodes.

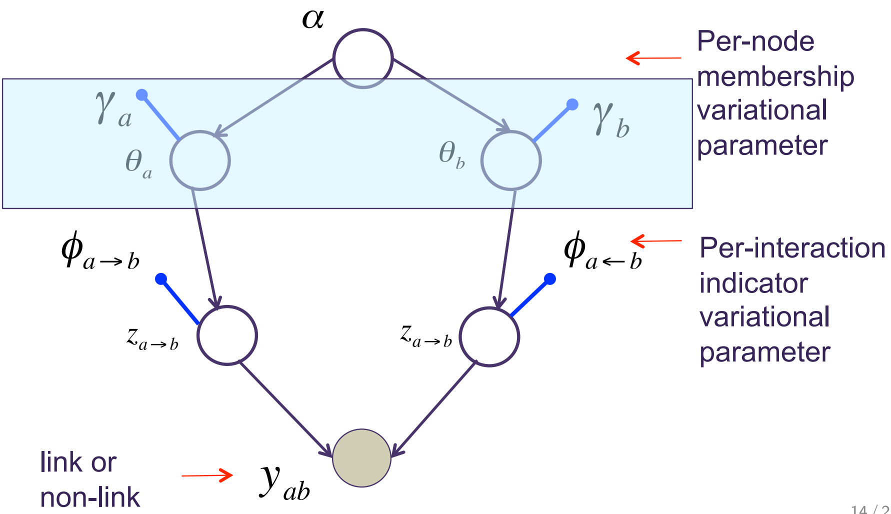
Batch variational inference: local step

In the local step, we iteratively update the parameters for each node pair, holding the per-node parameters fixed.



Batch variational inference: global step

In the global step, we aggregate the parameters computed from the local step and update the per-node parameters.



Stochastic variational inference

- Under conjugacy assumptions, the coordinate ascent is equivalent to taking a natural gradient (Amari, 1998) of step size 1.
- Why compute the real gradients, when a cheaper noisy estimate of the gradient will do (Robbins and Monro, 1951)?
- batch: Compute update from all pairs in the network.

$$\gamma_{a,k}^{t} = \alpha_k + \sum_{(a,b)} \phi_{a \to b,k}. \tag{2}$$

- stochastic: Sample a pair (a, b) from the network. Compute noisy gradient and update. s is a scaling factor.

$$\begin{aligned} \partial\gamma_{a,k}^{t} &= \alpha_k + s\phi_{a \to b,k} - \gamma_{a,k}^{t-1} \\ \gamma_a^t &\leftarrow \gamma_a^{t-1} + \rho_t \partial\gamma_a^t \end{aligned} \tag{3}$$

Sampling methods

- Sample uniformly from the set of all pairs.
- Sample a node and select all its pairs (links and non-links).
- Sample a node and select its linked pairs.
  - We specify a different variational family with assumptions on non-link variational indicators.
- Why does this work? How to ensure unbiased posterior estimates?

Stochastic optimization of the variational objective

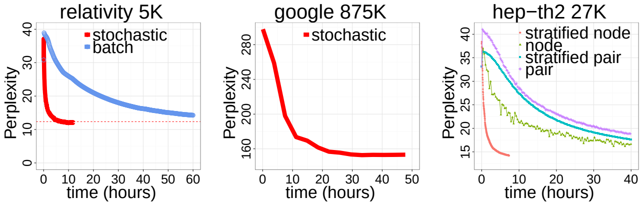
- Optimization follows noisy estimates of the gradient of the objective with a decreasing step-size (Robbins and Monro, 1951).
- We can cheaply compute a scaled stochastic gradient by first subsampling a subset of terms.
- The scaled gradient is a random variable whose expectation is the true gradient. This scaling avoids bias in the estimated posterior.

Stochastic inference algorithm

Stochastic inference for the assortative mixed-membership stochastic blockmodel is more efficient than batch.

1. Subsample a set of pairs of nodes $\mathcal{S}$.
2. For each pair $(i, j) \in \mathcal{S}$, use the current community structure to compute the indicator parameters $\hat{\phi}_{i \rightarrow j}$ and $\hat{\phi}_{i \leftarrow j}$.
3. Adjust the community memberships γ_i and γ_j.
4. Repeat.

Stochastic vs. batch inference on real networks



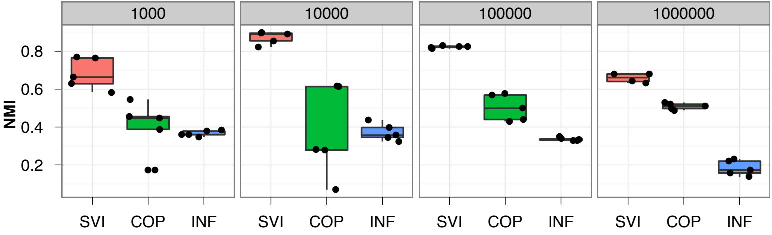
(Plot titles are the data set name and number of nodes.)

- Faster and performs as well as traditional variational inference.
- Better subsampling algorithms let us handle data set sizes not possible with traditional inference.
- More results in Gopalan et al., 2012.

Overview of methods for high accuracy and scalability

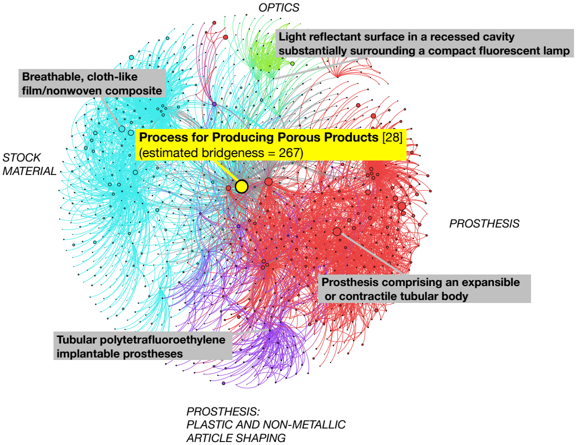
- For sampling methods that only observe links, define the variational family based on the data.
- Use per-community variational annealing to keep the variational mixed-memberships entropic.
- Set the number of communities using either a predictive approach (Geisser et al., 1979) or a fast batch variational inference algorithm.

Accuracy on LFR benchmark networks



- Accuracy is measured using normalized mutual information (Lancichinetti et al., 2009).
- The SVI algorithm outperforms INFOMAP and COPRA on large benchmark networks with overlapping communities.

Exploring 4M US Patents and their citations



The discovered community structure in a subgraph of the U.S. Patents network.

Conclusions

- Our community detection algorithm naturally interleaves subsampling the network and re-estimating its community structure.
- Our approach opens the door to large-scale network analysis with sophisticated probabilistic models.
- Our current work extends inference to more complex models.
  - Bayesian nonparametric assumptions on the number of communities
  - The degree-corrected assortative MMSB