While the HDPR is more broadly applicable, our focus in this paper is on *assortative* models for undirected networks, which assume that the probability of linking distinct communities is small. This modeling choice is appropriate for the clustered relationships found in friendship and collaboration networks. Our work builds on stochastic variational inference methods developed for the assortative MMSB (aMMSB) [6], but makes three key technical innovations. First, adapting work on HDP topic models [19], we develop a nested family of variational bounds which assign positive probability to dynamically varying subsets of the unbounded collection of global communities. Second, we use these nested bounds to dynamically prune unused communities, improving computational speed, predictive accuracy, and model interpretability. Finally, we derive a structured mean field variational bound which models dependence among the pair of community assignments associated with each edge. Crucially, this avoids the expensive and inaccurate local optimizations required by naive mean field approximations [1, 6], while maintaining computation and storage requirements that scale linearly (rather than quadratically) with the number of hypothesized communities.

In this paper, we use our assortative HDPR (aHDPR) model to recover latent communities in social networks previously examined with the aMMSB [6], and demonstrate substantially improved perplexity scores and link prediction accuracy. We also use our learned community structure to visualize business and governmental relationships extracted from the LittleSis database [13].

2 Assortative Hierarchical Dirichlet Process Relational Models

We introduce the assortative HDP relational (aHDPR) model, a nonparametric generalization of the aMMSB for discovering shared memberships in an unbounded collection of latent communities. We focus on undirected binary graphs with N nodes and $E = N(N-1)/2$ possible edges, and let $y_{ij} = y_{ji} = 1$ if there is an edge between nodes i and j. For some experiments, we assume the y_{ij} variables are only partially observed to compare the predictive performance of different models.

As summarized in the graphical models of Fig. 1, we begin by defining a global Dirichlet process to capture the parameters associated with each community. Letting β_k denote the expected frequency of community k, and $\gamma > 0$ the concentration, we define a stick-breaking representation of β:

$$\beta_k = v_k \prod_{\ell=1}^{k-1}(1 - v_\ell), \qquad v_k \sim \mathrm{Beta}(1, \gamma), \qquad k = 1, 2, \ldots \tag{1}$$

Adapting a two-layer hierarchical DP [18], the mixed community memberships for each node i are then drawn from DP with base measure β, $\pi_i \sim \mathrm{DP}(\alpha\beta)$. Here, $\mathbb{E}[\pi_i \mid \alpha, \beta] = \beta$, and small precisions α encourage nodes to place most of their mass on a sparse subset of communities.

To generate a possible edge y_{ij} between nodes i and j, we first sample a pair of indicator variables from their corresponding community membership distributions, $s_{ij} \sim \mathrm{Cat}(\pi_i)$, $r_{ij} \sim \mathrm{Cat}(\pi_j)$. We then determine edge presence as follows:

$$p(y_{ij} = 1 \mid s_{ij} = r_{ij} = k) = w_k, \qquad p(y_{ij} = 1 \mid s_{ij} \neq r_{ij}) = \epsilon. \tag{2}$$

For our assortative aHDPR model, each community has its own self-connection probability $w_k \sim \mathrm{Beta}(\tau_a, \tau_b)$. To capture the sparsity of real networks, we fix a very small probability of between-community connection, $\epsilon = 10^{-30}$. Our HDPR model could easily be generalized to more flexible likelihoods in which each pair of communities k, ℓ have their own interaction probability [1], but motivated by work on the aMMSB [6], we do not pursue this generalization here.

3 Scalable Variational Inference

Previous applications of the MMSB associate a pair of community assignments, s_{ij} and r_{ij}, with each potential edge y_{ij}. In assortative models these variables are strongly dependent, since present edges only have non-negligible probability for consistent community assignments. To improve accuracy and reduce local optima, we thus develop a structured variational method based on joint configurations of these assignment pairs, which we denote by $e_{ij} = (s_{ij}, r_{ij})$. See Figure 1.

Given this alternative representation, we aim to approximate the joint distribution of the observed edges y, local community assignments e, and global community parameters π, w, β given fixed