Hierarchical Solutions for the Stochastic Block Model

Aditi Nair and Akash Shah

May 11, 2017

The Stochastic Block Model

- Let G be a graph with k communities of m nodes each
- For nodes (i, j) in the same community, let the p be the probability that they share an edge on G
- For nodes (i, j) *not* in the same community, let the q be the probability that they share an edge on G
- Edges are drawn independently, and $p > q$
- Goal: recover the original partition

The Semidefinite Program for $k = 2$

Let A be the adjacency matrix of G. Then we need to solve:

$$\max \sum_{i,j} A_{ij} x_i x_j$$

$$\text{s.t. } x_i = \pm 1 \ \forall i$$

$$\sum_j x_j = 0$$

The Semidefinite Program for $k = 2$

Let $B = 2A - (\mathbf{1}\mathbf{1}^T - I)$. Then the original problem is equivalent to

$$\max \sum_{i,j} B_{ij} x_i x_j$$

$$\text{s.t. } x_i = \pm 1 \ \forall i$$

$$\sum_j x_j = 0$$

The Semidefinite Program for $k = 2$

By dropping the constraint $\sum_j x_j = 0$ and with a convex relaxation, we have the following semidefinite program:

$$\max Tr(BX)$$

$$\text{s.t. } X_{ii} = 1 \ \forall i$$

$$X \succeq 0$$

Under the right conditions, with high probability the solution coincides with $X = gg^T$ where g is the true partition. We obtain g by finding the leading eigenvector of X.

Conditions for Recovery

	Exact	Partial
Formula	$p = \frac{\alpha \log n}{n},\ q = \frac{\beta \log n}{n}$	$p = \frac{a}{n},\ q = \frac{b}{n}$
$k = 2$	$\sqrt{\alpha} - \sqrt{\beta} \geq \sqrt{2}$	$(a - b)^2 > 2(a + b)$
$k > 2$	$\sqrt{\alpha} - \sqrt{\beta} \geq \sqrt{k}$	$\frac{(a-b)^2}{k(a+(k-1)b)} > 1$

Hierarchical Approach for $k > 2$

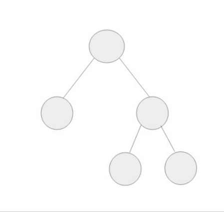


Figure 1: Hierarchy for k = 3

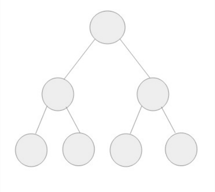
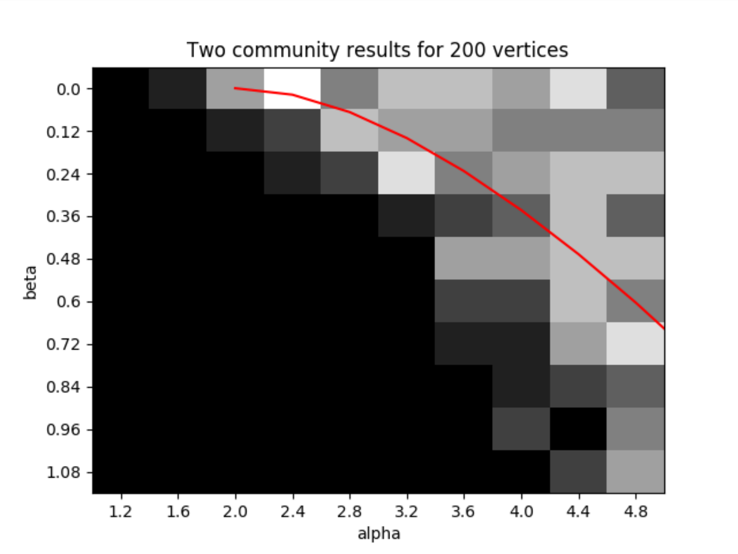
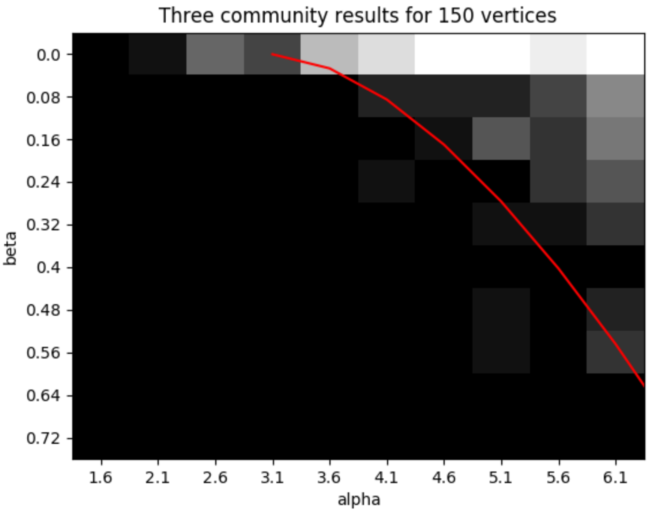


Figure 2: Hierarchy for k = 4

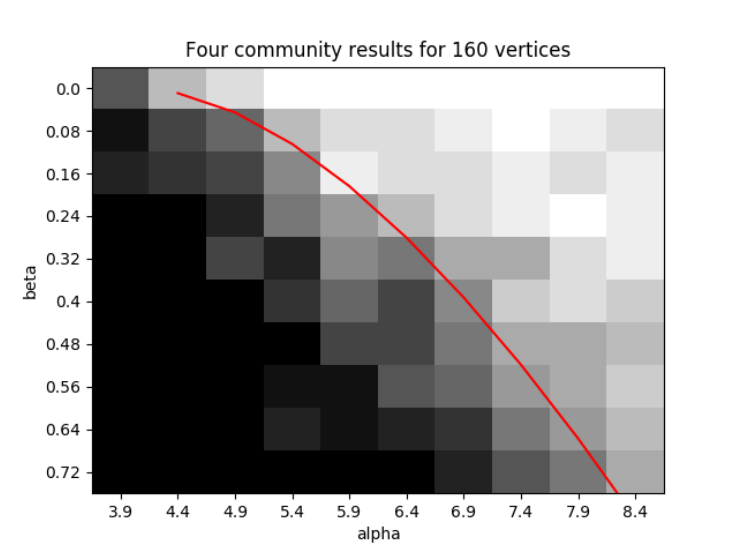
Exact Recovery for $k = 2$



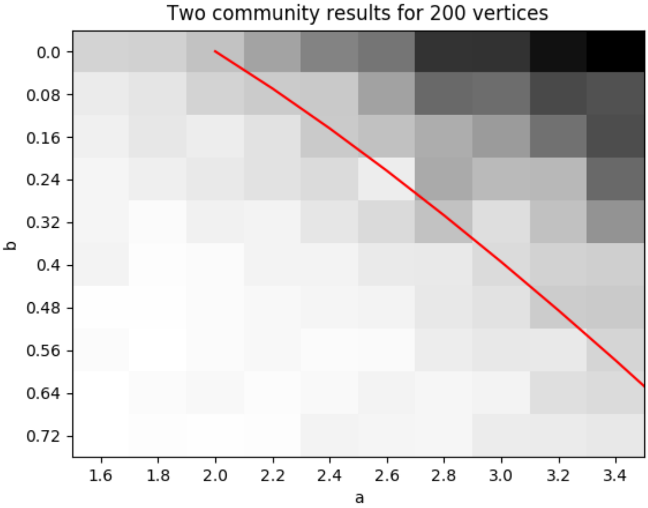
Exact Recovery for $k = 3$



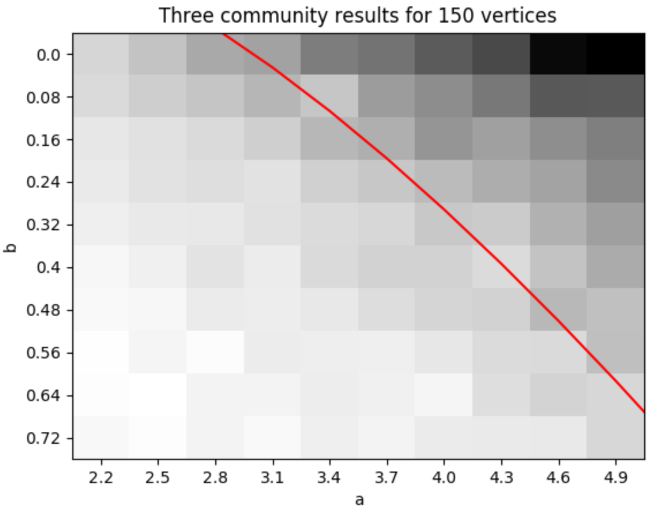
Exact Recovery for $k = 4$



Partial Recovery for $k = 2$



Partial Recovery for $k = 3$



Partial Recovery for $k = 4$