Statistical Network Modeling

- Quant. of uncertainty still an open area of work

Classes of models

- Regression models (ERGMs)
- Latent variable models
- Mixture models (SBMs)

Let $C_1, \dots, C_K$ be the blocks, then assign different probs for between/within

Generative model

(a) Assign classes

(b) Assign edges | groups with prob π_{rz}

SBM: $z_i \overset{iid}{\sim} \text{Multi}(1, \alpha)$

$A_{ij} \mid z_i = z_i, z_j = z_j \sim \text{Bern}(\pi_{z_i, z_j})$

- Assign $A_{ij} = 0, i = j$ and π symmetric
- Only identifiable up to class labels

Szemerédi Regularity Lemma: SBMs can approximate any random graph.

- Bounds cut metric in an approximation Lemma. in terms of ε

- Need $K = 2^{2/\epsilon^2}$ classes

Conceptually: — Two sets $\{Z, Y\}$ half unobservable data

- Need to try and observe them

$$\ell(y; z_i) = \sum_i \sum_t z_{it} \log \alpha_t + \frac{1}{2} \sum_{i \neq j} \sum_{t \neq r} z_{it} z_{jr} \log \text{bern}(y_{ij}; \pi_{tr})$$

Don't see z so $P(Y=y) = \sum_z P(Y=y, z=z)$

<u>Computational Methods</u>

- EM; iterate between Expectation/maximization; only works for small networks

- Variational EM; Approximation to EM

- Profile likelihoods (Bickel/Chen 2009)

Consistency & other Properties — (See slides)

Goodness of fit: Integrated Classification likelihood

Look at Lei for community detection estimation of Q.