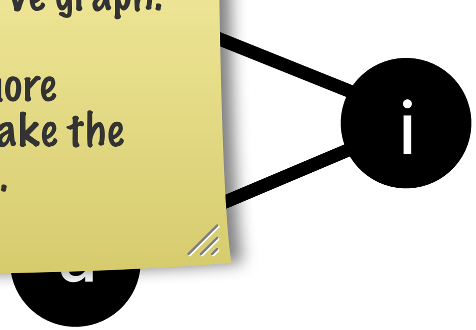
The blessing of dimensionality for sparse Stochastic Blockmodels.

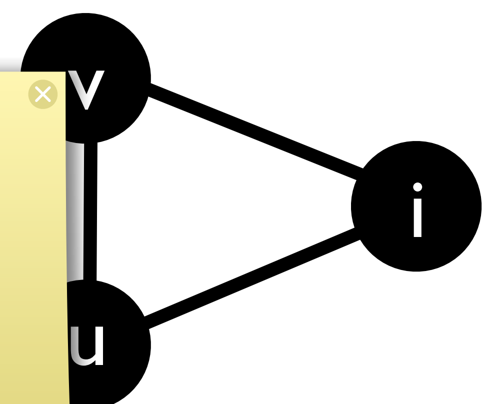
Karl Rohe, Tai Qin, Haoyang Fan
UW Madison, Statistics

graph model: high degree hard.

inference: observe graph. edges are like observations, more observations make the problem easier....



work in progress. comments please!



Disclaimer:
This is work
in progress.

Abstract

- What is the model? What are we estimating?
- Current literature
  - Low-dimensional
  - Networks not transitive
- "Highest dimensional model"
  - Transitive
  - Easier estimation
  - Faster, local algorithms.

Inference for clustering

- Presume that graph is observed.
- Sampled from a distribution in a known class.
- What is estimable?
- Certain models lead to clusters. Can we estimate the corresponding parameters?

Exchangeable Random Graphs and Latent space models (Hoff et al 2002)

Observe: $A = \{0,1\}^{n\times n}$ undirected, unweighted.

Edges conditionally independent.

$$P(A|Z) = \prod_{u<v} P(A|Z_u, Z_v)$$

Wish to estimate latent variables

$$Z_1, \ldots, Z_n$$

Stochastic Blockmodels (Holland et al 1983) are a type of exchangeable random graph.

Stochastic Blockmodels (SBMs) say class memberships

$Z_i \in \{1, \ldots, K\}$ yield

$$P(A_{uv} = 1 | Z_u, Z_v) = \Theta_{Z_u, Z_v}$$

for $\Theta \in [0,1]^{K \times K}$

Large diagonal elements in Theta makes communities.
Estimating Z = clustering

Several have studied estimators for Z under SBM

A nonparametric view of network models and Newman–Girvan and other modularities

Peter J. Bickel[a,1] and Aiyou Chen[b]

SPECTRAL CLUSTERING AND THE HIGH-DIMENSIONAL STOCHASTIC BLOCKMODEL

By Karl Rohe, Sourav Chatterjee and Bin Yu

University of California Berkeley

Stochastic blockmodels with a growing number of classes

By D. S. Choi

School of Engineering and Applied Sciences, Harvard University, Cambridge, Ma 02138, U.S.A.

dchoi@seas.harvard.edu

THE METHOD OF MOMENTS AND DEGREE DISTRIBUTIONS FOR NETWORK MODELS

By Peter J. Bickel[1], Aiyou Chen[2] and Elizaveta Levina[3]

University of California, Berkeley, Google Inc. and University of Michigan

Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model *

Alain Celisse[†] et al

A consistent adjacency spectral embedding for stochastic blockmodel graphs

Daniel L. Sussman, Minh Tang, Donniell E. Fishkind, Carey E. Priebe

Johns Hopkins University, Applied Math and Stati... Department

CONSISTENT BICLUSTERING

By Cheryl J. Flynn and Patrick O. Perry

New York University

ASYMPTOTIC NORMALITY OF MAXIMUM LIKELIHOOD AND ITS VARIATIONAL APPROXIMATION FOR STOCHASTIC BLOCKMODELS

By Peter Bickel , David Choi , Xiangyu Chang , and Hai Zhang

University of California, Berkeley
University of California, Berkeley
Xi'an Jiaotong University
Northwest University

CO-CLUSTERING FOR DIRECTED GRAPHS; THE STOCHASTIC CO-BLOCKMODEL AND A SPECTRAL ALGORITHM

KARL ROHE[1] AND BIN YU[2]

[1] *UNIVERSITY OF WISCONSIN MADISON*

[2] *UNIVERSITY OF CALIFORNIA BERKELEY*

Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model

Kamalika Chaudhuri KAMALIKA@CS.UCSD.EDU
Fan Chung FAN@CS.UCSD.EDU
Alexander Tsiatas ATSIATAS@CS.UCSD.EDU
Department of Computer Science and Engineering
University of California, San Diego

Classification and estimation in the Stochastic Block Model based on the empirical degrees

Antoine Channarond
Jean-Jacques Daudin
and Stéphane Robin

Abstract

- What is the model? What are we estimating?
- **Previous parameterizations:**
    - Low-dimensional
    - Networks not transitive
- "Highest dimensional model"
    - Transitive
    - Easier estimation
    - Faster, local algorithms.

In previous work, the asymptotic settings have captured two features of empirical networks.

1) Sparse edges...
   expected degree grows like polylog(n).
   would like to have constant expected degree

2) Fixed number of blocks or growing slowly.
   best result: K = o(sqrt(n)).

Easier to consider the planted partition (aka four parameter SBM)

Recall that the Stochastic Blockmodel make clusters when

Theta has large diagonal elements.

and small off-diagonal elements.

Four parameter model,

constant diagonal

constant off-diagonal

equally sized blocks.

Easier to consider the planted partition (aka four parameter SBM)

- K = number of blocks
- s = population of each block. Condition on Z. All blocks have equal population.
- r = probability of an out-of-block connection
- p = probability of an in-block connection

So, $$n = Ks$$

Easy expression for expected degree.

$$\text{Expected degree} = \underbrace{(\overbrace{s}^{\text{block population}}-1)\,p}_{\text{in-block prob}} + \underbrace{(\overbrace{n}^{\text{number of nodes}}-s)\,r}_{\text{out-of-block prob}}$$

Easy expression for expected degree.

$$\text{Expected degree} = \overbrace{(s-1)}^{\text{block population}}\ \underbrace{p}_{\text{in-block prob}} + \overbrace{(n-s)}^{\text{number of nodes}}\ \underbrace{r}_{\text{out-of-block prob}}$$

If the expected degree is polylog(n), then

$$p = O\left(\frac{\log^\alpha n}{s}\right) = O\left(\frac{K \log^\alpha n}{n}\right)$$

In previous work, p --> 0.

$$p = O\left(\frac{\log^{\alpha} n}{s}\right) = O\left(\frac{K \log^{\alpha} n}{n}\right)$$

Shrinking p has a dramatic effect on the clusters in the network

A similar assumption is made with more general models.

does not change with n

$$P(A_{uv} = 1 | Z_u, Z_v) = \rho_n w(Z_u, Z_v)$$

“edge density” converges to zero

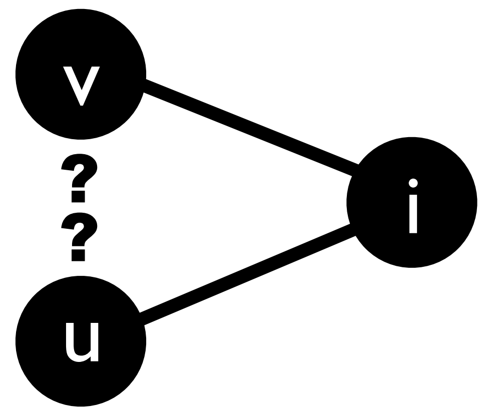
$$\rho_n = \frac{\mathrm{E}(\text{number of edges})}{\binom{n}{2}}$$

In previous research, p_max goes to zero.

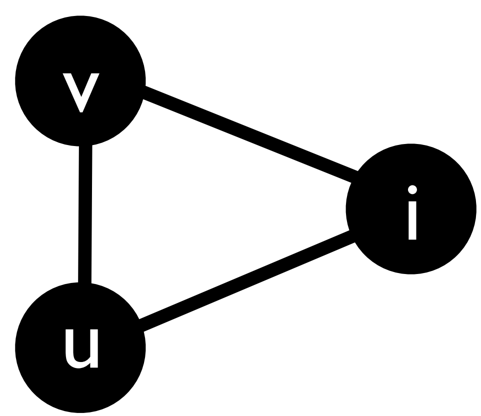
$$p_{\max} = \max_{Z_u, Z_v} P(A_{uv} = 1 | Z_u, Z_v)$$

This ensures sparsity.
It also removes transitivity.

Transitivity

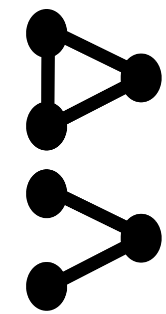


Friends of friends
are friends



Transitivity and the transitivity coefficient

$$\text{transitivity coefficient} = \frac{3 \times \text{number of triangles}}{\text{number of 2 stars}}$$



- "proportion of closed triangles"
- In social networks, this quantity is typically between .3 and .6 (Snijders slides).
- Extremely large networks also have non-negligible coefficients (Leskovec SNAP).

Theorem

Under the exchangeable random graph model, if

$$n\rho_n \to \infty, \;\; \rho_n = o(1), \;\; \text{and } p_{\max} = O(\rho_n)$$

then,

$$\text{Transitivity Coefficient}(A) \xrightarrow{P} 0$$

recall: $$\rho_n = \frac{\mathrm{E}(\text{number of edges})}{\binom{n}{2}}$$

Conclusion: To preserve transitivity, it is necessary for p_max > epsilon > 0.

What parameters ensure a Stochastic Blockmodel with *sparsity and transitivity*?

Abstract

- What is the model? What are we estimating?
- Previous parameterizations:
  - Low-dimensional
  - Networks not transitive
- "Highest dimensional model"
  - Transitive
  - Better theoretical results
  - Faster, local algorithms.

Want expected degree to be constant.

block population

number of nodes

$$\text{Expected degree} < s\,p \; + \; n\,r$$

in-block prob,
must stay fixed.

out-of-block prob

Want expected degree to be constant.

$$\text{Expected degree} < \overbrace{s}^{\text{block population}}\ \underbrace{p}_{\text{in-block prob, must stay fixed.}} + \overbrace{n}^{\text{number of nodes}}\ \underbrace{r}_{\text{out-of-block prob}}$$

$nr = C$ implies $r = O(1/n)$

Want expected degree to be constant.

block population

number of nodes

Expected degree $< s\,p + n\,r$

in-block prob, must stay fixed.

out-of-block prob

sp = C and p = pmax not going to zero together imply that s (block population) does not grow.

Want expected degree to be constant.

block population

number of nodes

Expected degree $< s\,p + n\,r$

in-block prob, must stay fixed.

out-of-block prob

sp = C and p = pmax not going to zero together imply that s (block population) does not grow.

n = Ks implies K must grow proportionally to n!

Highest dimensional model yields transitivity & sparsity

- K grows proportional to n.
- "Highest dimensional" because number of blocks cannot grow faster than number of nodes. Each block needs a member!
- Best results in the current literature require that K = o(sqrt(n)) (Choi et al 2012)

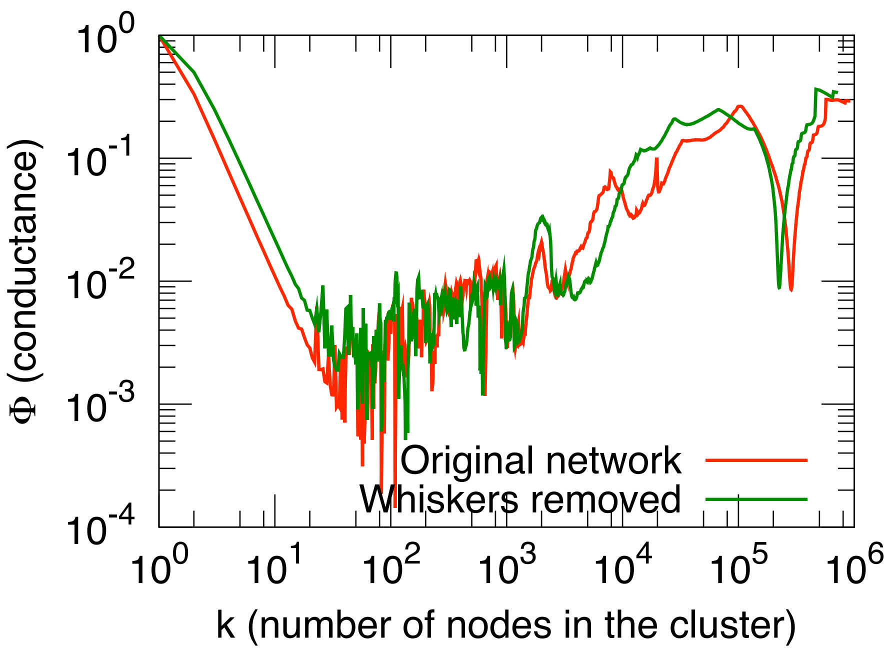
Highest dimensional model is asymptotically transitive.

Proposition:
Under four-parameter SBM, with fixed s, fixed p>0 , and r = c/n,

$$\text{TransitivityCoefficient(A)} \xrightarrow{P} \frac{E(\text{number of triangles in } A|Z)}{E(\text{number of two stars in } A|Z)} > 0$$

Empirical networks have small communities.

- Leskovec, Lang, Dasgupta, Mahoney 2008. In large empirical networks, “best clusters” are no larger than 100 nodes.

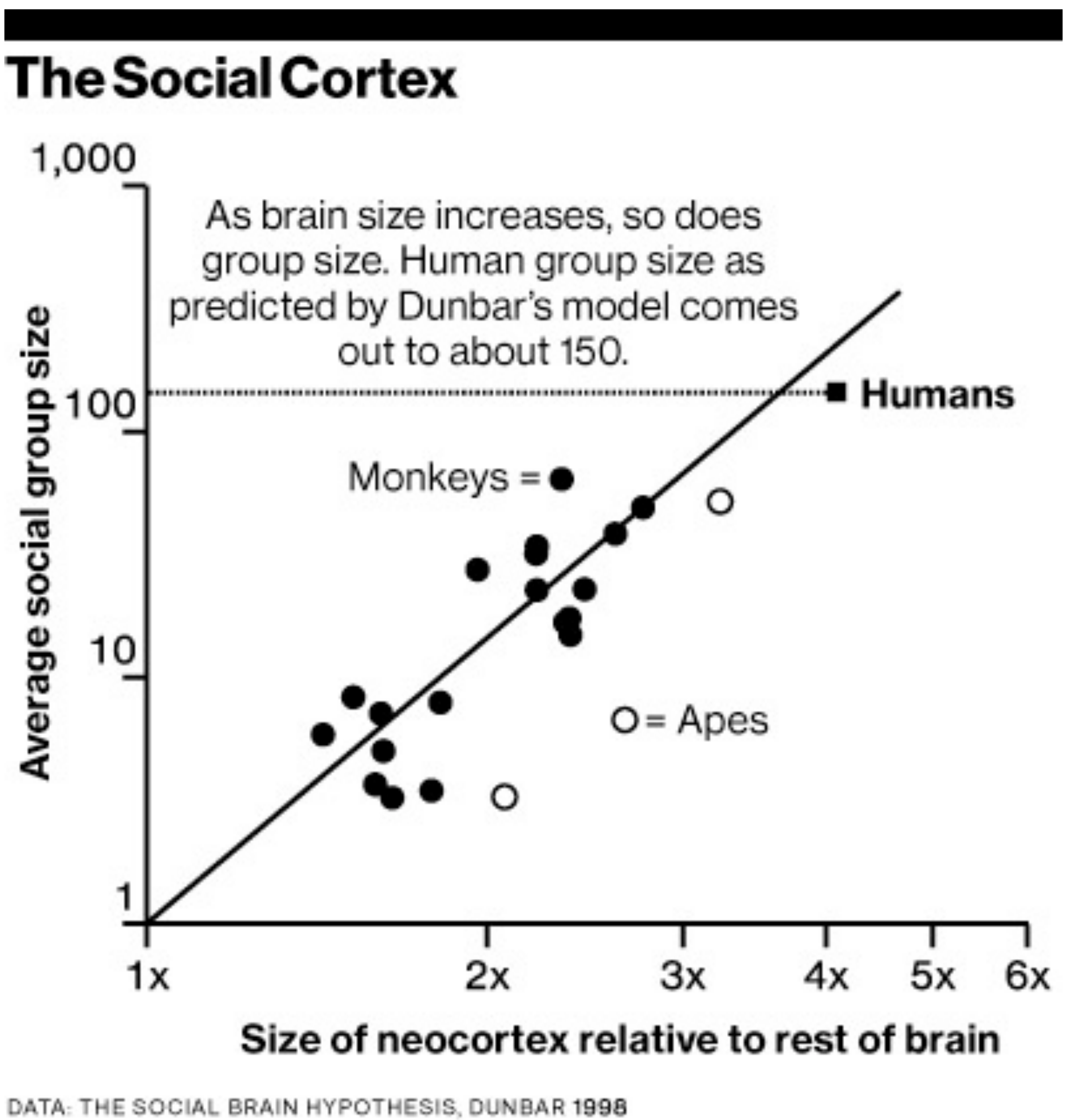


(a) LiveJournal01

Biological evidence for small communities.

38 primates

Biological evidence for small communities.



38 primates

Graphic from: "The Dunbar Number, From the Guru of Social Networks" Businessweek Jan 2013

Abstract

- What is the model? What are we estimating?
- Previous parameterizations:
  - Low-dimensional
  - Networks not transitive
- "Highest dimensional model"
  - Transitive
  - Easier estimation
  - Faster, local algorithms.

Recall . . .

The Stochastic Blockmodel sets

$$P(A_{uv} = 1 | Z_u, Z_v) = \Theta_{Z_u, Z_v}$$

for $Z_i \in \{1, \ldots, K\}$ and $\Theta \in [0, 1]^{K \times K}$

A regularized estimator

Log-likelihood:

$$L(A; z, \theta) = \sum_{i<j} \{A_{ij} \log \theta_{z_i z_j} + (1 - A_{ij}) \log(1 - \theta_{z_i z_j})\}.$$

standard MLE: $$\hat{z} = \arg\max_{z} \max_{\theta \in [0,1]^{K \times K}} L(A; z, \theta)$$

A regularized estimator

Log-likelihood:

$$L(A; z, \theta) = \sum_{i<j} \{A_{ij} \log \theta_{z_i z_j} + (1 - A_{ij}) \log(1 - \theta_{z_i z_j})\}.$$

standard MLE: $$\hat{z} = \arg\max_z \max_{\theta \in [0,1]^{K \times K}} L(A; z, \theta)$$

restricted MLE: $$\hat{z}^R = \arg\max_z \max_{\theta \in R_K} L(A; z, \theta)$$

A regularized estimator

Log-likelihood:

$$L(A; z, \theta) = \sum_{i<j} \{A_{ij} \log \theta_{z_i z_j} + (1 - A_{ij}) \log(1 - \theta_{z_i z_j})\}.$$

standard MLE: $$\hat{z} = \arg\max_{z} \max_{\theta \in [0,1]^{K \times K}} L(A; z, \theta)$$

restricted MLE: $$\hat{z}^R = \arg\max_{z} \max_{\theta \in R_K} L(A; z, \theta)$$

The restricted set,

$$R_K = \left\{ \theta \in [0, 1]^{K \times K} : \theta_{ab} = c, \ \forall\, a \neq b \ \text{ and for } c \in [0, 1] \right\}.$$

Highest dimensional SBM

1) Small Blocks

The smallest block size b satisfies $$\frac{\log^4 n}{b} = o(1)$$

2) Large in block prob, small out block prob

except for a small "exceptional set" Q with large probabilities.

$$\Delta < \Theta_{ij} < 1 - \Delta \quad \text{if } i = j \text{ or } (i,j) \in Q$$

$$1/n^2 < \Theta_{ij} < C \log^3(n)/n \quad \text{otherwise}$$

Delta and C are constants.

Theorem

Under the Highest Dimensional SBM with

1) Not too many "exceptional" connections,

$$\left|\left\{(i,j) : (Z_i, Z_j) \in Q\right\}\right| = o(ns)$$

2) Sufficient separation between blocks,

For $a \neq b$, there exists c such that

$$D\left(\theta_{ac} \| \frac{\theta_{ac} + \theta_{bc}}{2}\right) + D\left(\theta_{bc} \| \frac{\theta_{ac} + \theta_{bc}}{2}\right) \geq C \frac{MK}{n^2}$$

where D(||) is the KL divergence and M is the expected number of edges in the graph,

Under these conditions, the proportion of nodes misclustered by RMLE converges to zero:

$$\frac{N_e(\hat{z}^R)}{N} = o_p(1)$$

Proof extended from Choi et al 2012

Definition of misclustered

A node is *correctly clustered* if its true class is in the majority within the estimated class.

$N_e(\hat{z}^R)$

The number of nodes whose true class is not the majority within its estimated class.

Off diagonal terms are all equal.

The restricted set,

$$R_K = \left\{ \theta \in [0,1]^{K \times K} : \boxed{\theta_{ab} = c,\ \forall\ a \neq b \ \text{ and for } c \in [0,1]} \right\}.$$

Off diagonal terms are all equal.

The restricted set,

$$R_K = \left\{ \theta \in [0,1]^{K \times K} : \boxed{\theta_{ab} = c, \ \forall \ a \neq b \ \text{ and for } c \in [0,1]} \right\}.$$

- All out-of-block probabilities estimated equal.
- Does not "fit to" between-block-edges.
  - doesn't move a node from block 9 to 10 based on it's connections to blocks 1, 2, 3, 4.
- Considers "in-block" connections, i.e. local connections, connections that reflect a high level of transitivity.

Note: Newman-Girvan modularity only measures local connections.

Simulation

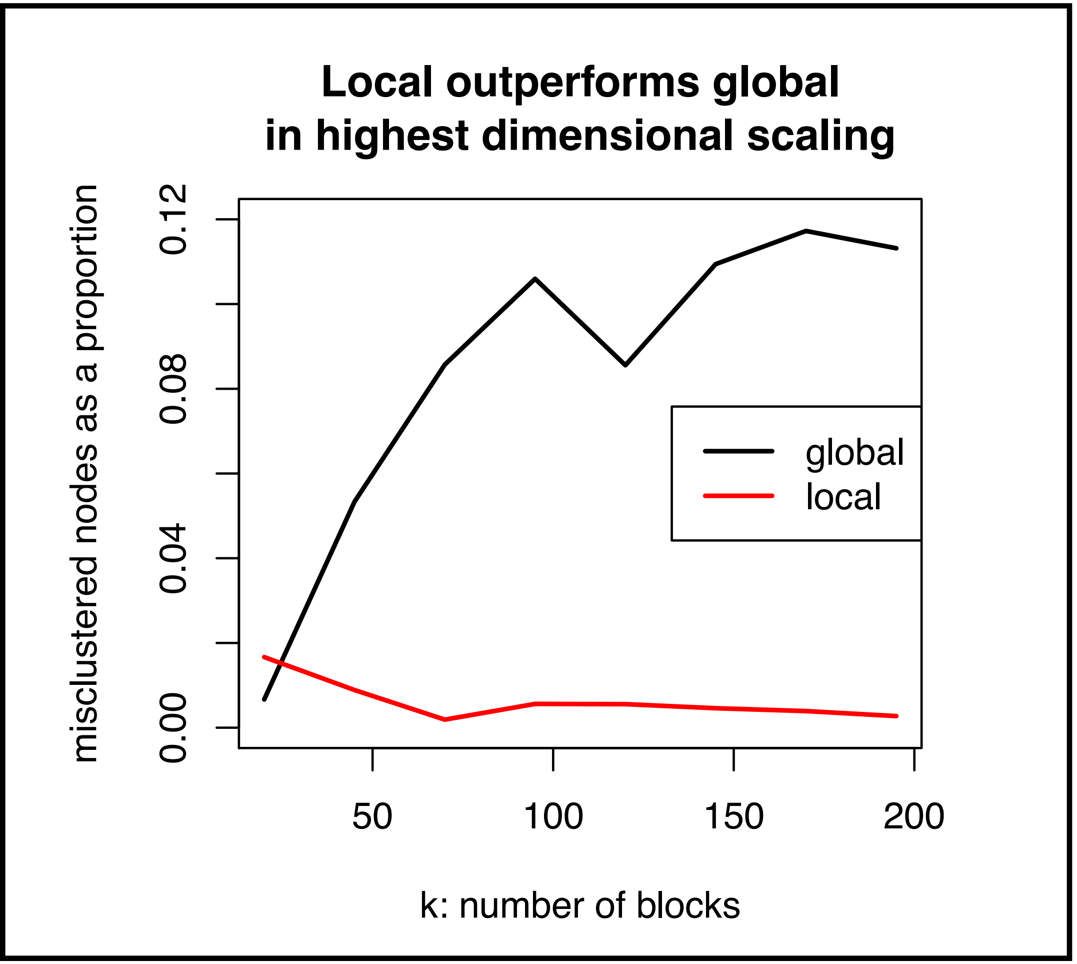
- Computing MLE or RMLE is difficult. Discrete optimization.
- **Key idea of RMLE: ignore the out of block edges. Act “locally” on the “high transitivity” edges.**

Simulation

- This simulation compares
  - "Local" algorithm: agglomerative hierarchical clustering,
    - Dissimilarity measure inversely proportional to number of common neighbors, normalized by degree.
  - "Global" algorithm: spectral clustering
    - k-means on leading eigenvectors of the symmetric graph Laplacian.

Simulation settings

- Each block has 15 members.
- The expected in-block degree is 8 (good edges).
- The expected out-of-block degree is 8 (bad edges).
- K, number of blocks, grows from 20 to 195.



global = spectral clustering with symmetric graph Laplacian.

local = agglomerative hierarchical clustering with dissimilarity

$$d(i,j) = \frac{1}{[LL]_{ij} + .1}$$

symmetric graph Laplacian

LL contains "number of common neighbors," normalized by degree.

Recap

- For sparsity, current exchangeable models send p_max to zero, removing transitivity.
- Empirical networks are transitive. Asymptotics should provide this.
- Transitivity leads to a blessing of dimensionality.
- Weak consistency with parametric regularization. Difficult to fit.
- Regularization through local methods, leveraging transitivity.

Define

$$p_{\max} = \max_{Z_u, Z_v} \; P(A_{uv} = 1 | Z_u, Z_v)$$

$$p_\Delta = P(A_{uv} = 1 | A_{iu} = A_{iv} = 1)$$

So, $\quad p_\Delta \leq p_{\max}$

When, $h_n(u,v) = P(A_{uv} = 1 | Z_u, Z_v) = \rho_n w(u,v)$

$$p_\Delta \leq p_{\max} = O(\rho_n)$$

Sparse graph sends transitivity coefficient to zero.