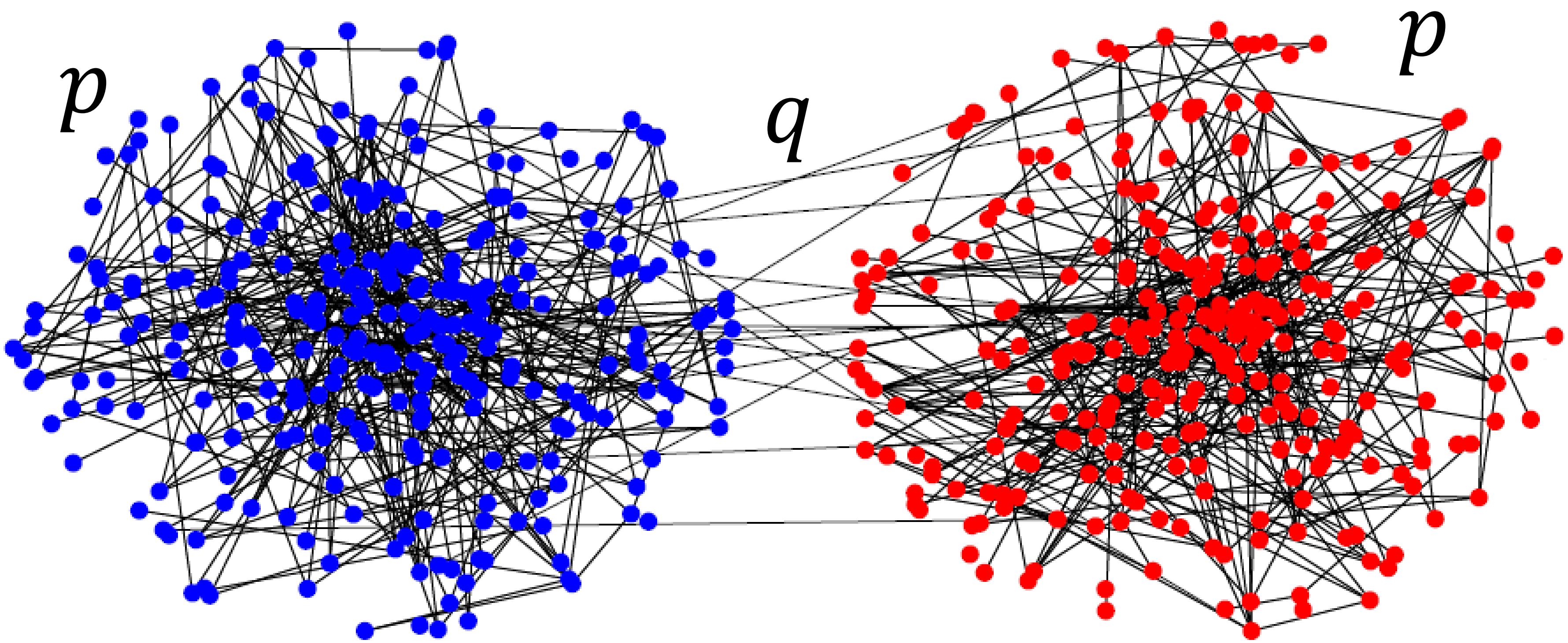
Low-rank approaches for SDP's in community detection

Afonso Bandeira (MIT→Courant)

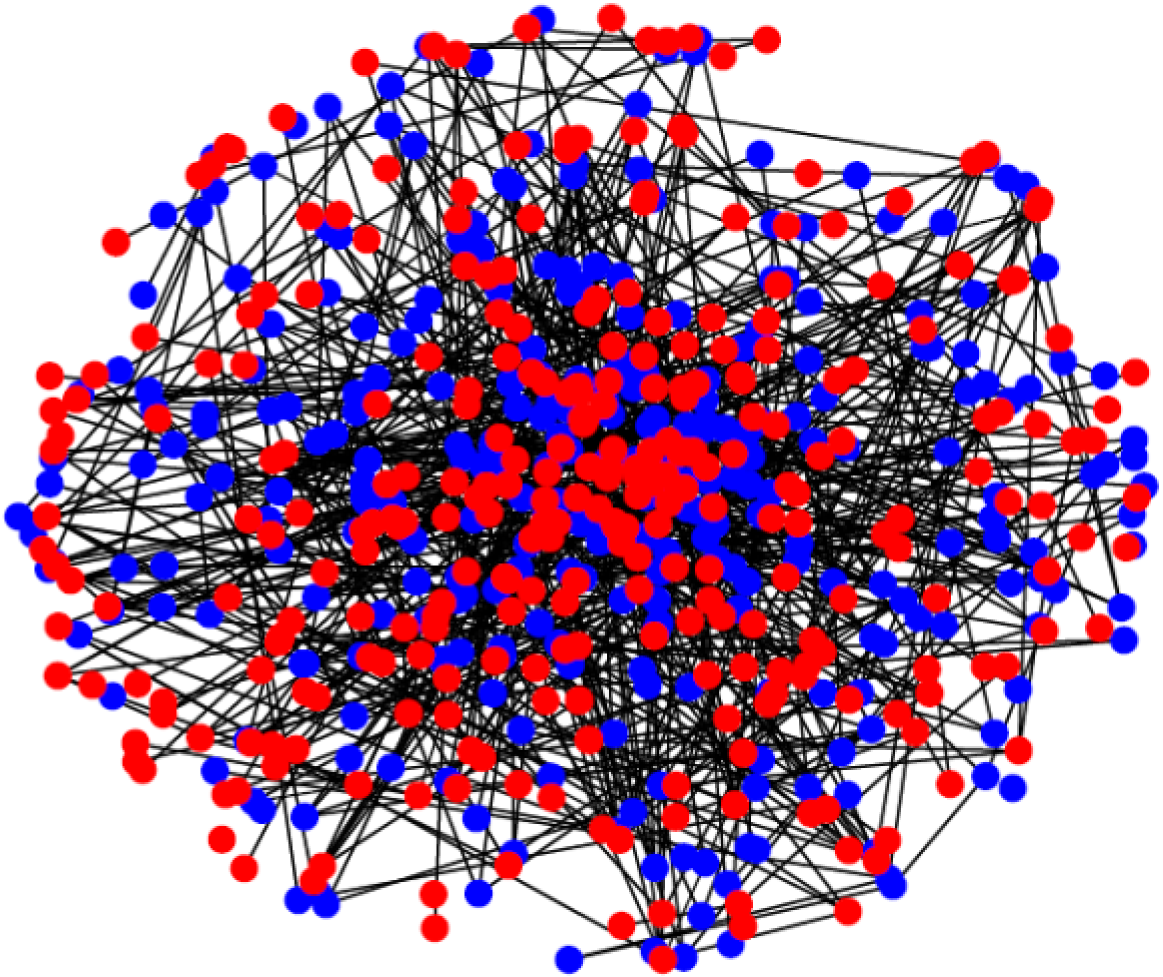
Nicolas Boumal (Princeton University)

Vladislav Voroninski (MIT)

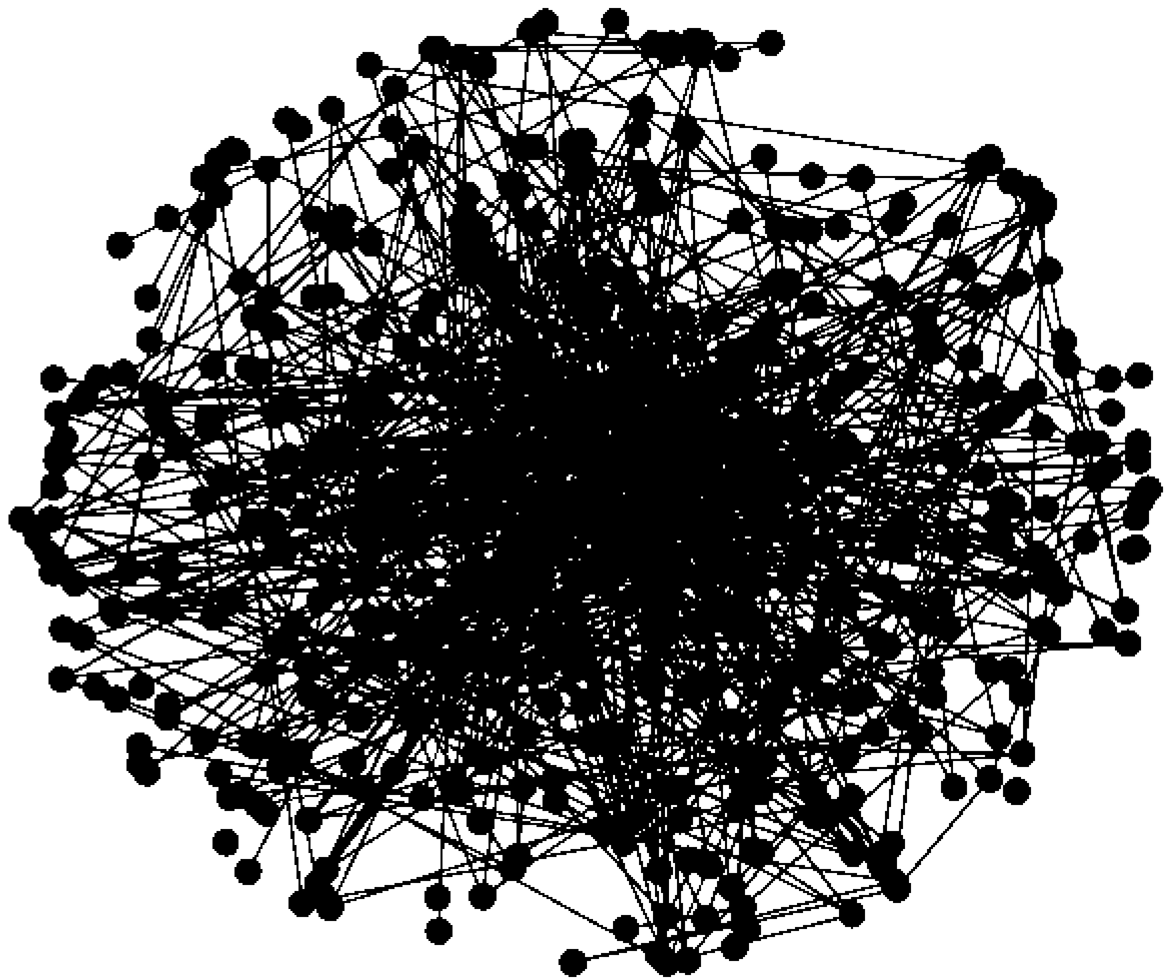
Community detection in the stochastic block model (SBM)



Community detection in the stochastic block model (SBM)



Community detection in the stochastic block model (SBM)



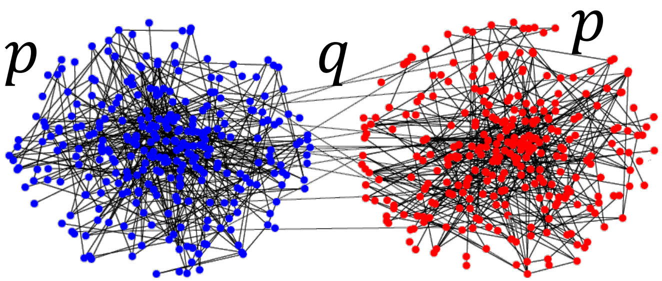
The constant average degree regime

Link within: $p = \frac{a}{n}$

Link across: $q = \frac{b}{n}$

Each community has a giant connected component, whp.

Hope for non-trivial correlation with true partition.



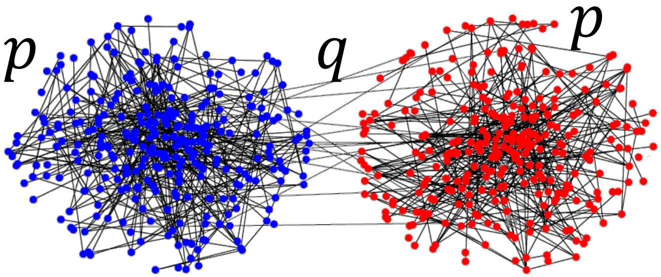
The dense regime

Link within: $p = \alpha \frac{\log n}{n}$

Link across: $q = \beta \frac{\log n}{n}$

Each community is connected, whp.

Hope for exact recovery.



The key SNR quantity: $\lambda(p,q)$

$$\lambda(p,q) = \frac{p-q}{\sqrt{2(p+q)}}\sqrt{n}$$

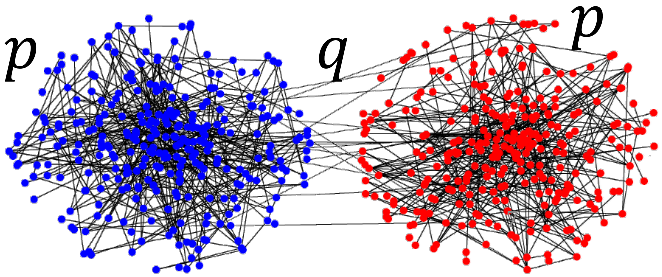
For non-trivial correlation: need $\lambda > 1$

Decelle et al. ‘11, Mossel et al. ‘14, Massoulié et al. ‘14

For exact recovery: “need $\lambda > \sqrt{2\log n}$”

Mossel et al. ‘14, Abbe et al. ’14

(Precise condition: $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$.)



Relaxation of MLE gives SDP for SBM

With A the adjacency matrix and $A' = A - \frac{p+q}{2}\mathbf{1}\mathbf{1}^T$:

$$\max_X \langle A', X\rangle \ \text{s.t.}\ \operatorname{diag}(X) = \mathbf{1}, X \succcurlyeq 0$$

Non-trivial correlation:

$\forall\delta > 0$, if $\frac{p+q}{2}n$ large enough, $\lambda > 1 + \delta$ is enough

Guedon & Vershynin '14, Montanari & Sen '15, Javanmard et al. '15

Exact recovery: SDP is tight at the info limit

Hajek et al. '14, Bandeira '15

The Burer–Monteiro approach

$$\max_X \langle A', X \rangle \text{ s.t. } \operatorname{diag}(X) = \mathbf{1}, X \succcurlyeq 0$$

Parameterize $X = YY^T$ with Y of size $n \times p$:

$$\max_Y \left\langle A', YY^T \right\rangle \text{ s.t. } \operatorname{diag}\left(YY^T\right) = \mathbf{1}$$

The aggressive version: $p = 2$.

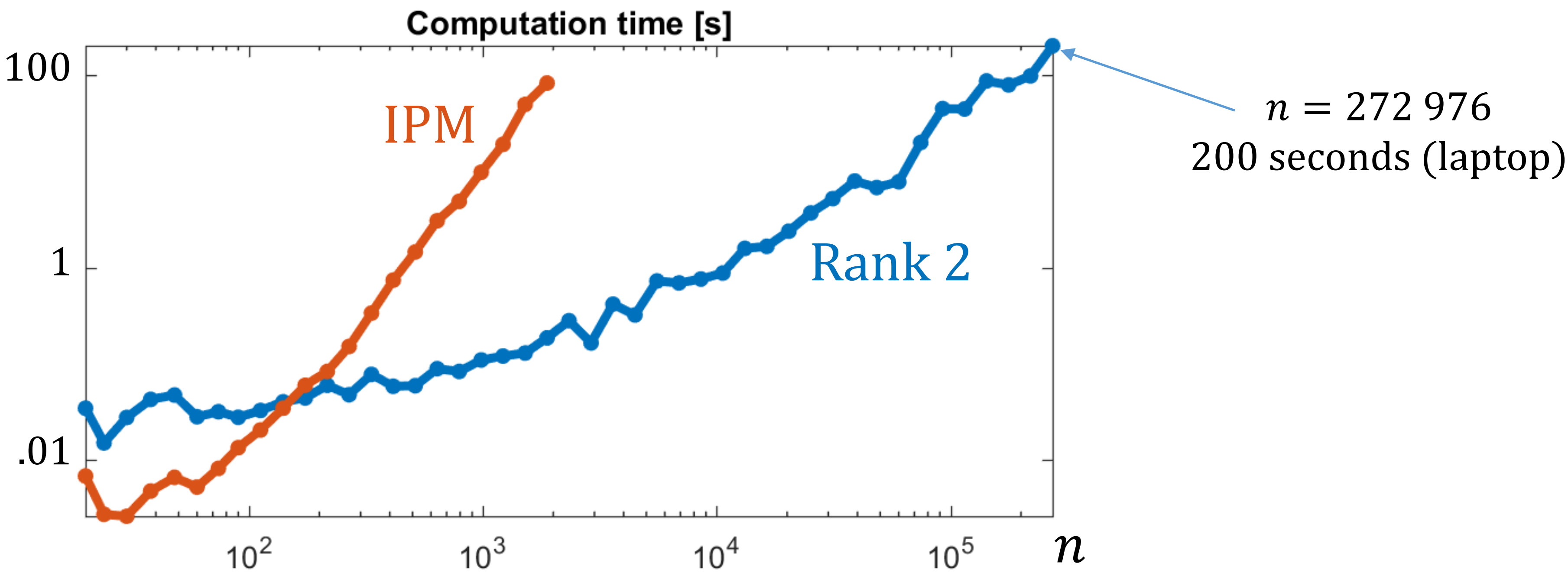
Non-convex optimization on the n-torus

$$\max_{Y \in \mathbf{R}^{n\times 2}} \langle A', YY^T \rangle \text{ s.t. } \operatorname{diag}(YY^T) = \mathbf{1}$$

Low-dimensional, and no conic constraint.

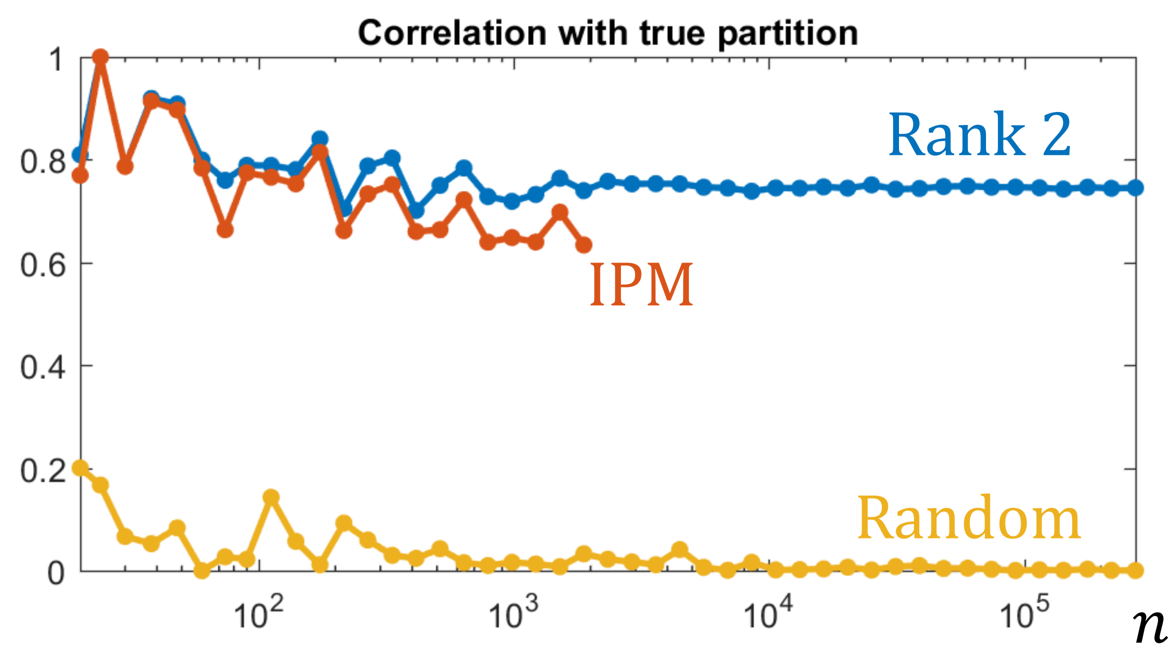
We run Riemannian trust regions via Manopt.

Do KKT points have good statistical properties?



$$p = \frac{10}{n}, q = \frac{2}{n}$$

$$\lambda = 1.63$$



Main result 1: non-trivial correlation

$$\max_{Y \in \mathbf{R}^{n\times 2}} \langle A', YY^T \rangle \text{ s.t. } \operatorname{diag}(YY^T) = \mathbf{1}$$

In the constant average degree regime, for any $\delta > 0$, if $\frac{p+q}{2}n$ is large enough and $\lambda > 8 + \delta$,

Then, there exists $\varepsilon > 0$ such that, whp, all second order KKT points Y correlate non-trivially with the true partition $\boldsymbol{g}$:

$$\frac{1}{n}\left\|Y^T \boldsymbol{g}\right\|_2 \geq \varepsilon.$$

Main result 2: exact recovery

$$\max_{Y \in \mathbf{R}^{n \times 2}} \langle A', YY^T \rangle \text{ s.t. } \operatorname{diag}(YY^T) = \mathbf{1}$$

There exists c (universal) such that, if

$$\lambda \geq cn^{1/3},$$

then, whp, all second-order KKT points Y are optimal and correspond to $\boldsymbol{g}$:

$$YY^T = \boldsymbol{g}\boldsymbol{g}^T.$$

1. $\operatorname{Hess} f(Y) \succcurlyeq 0 \Leftrightarrow \operatorname{ddiag}(A'YY^T) \succcurlyeq A' \circ YY^T$

$$\langle \operatorname{ddiag}(A'YY^T), \boldsymbol{g}\boldsymbol{g}^T \circ YY^T \rangle \geq \langle A' \circ YY^T, \boldsymbol{g}\boldsymbol{g}^T \circ YY^T \rangle$$

2. Link A' to the signal: $A' \propto \boldsymbol{g}\boldsymbol{g}^T + \frac{n}{\lambda} E + D$

$$\left\langle \boldsymbol{g}\boldsymbol{g}^T + \frac{n}{\lambda} E, YY^T \right\rangle \geq \left\langle \boldsymbol{g}\boldsymbol{g}^T + \frac{n}{\lambda} E, \boldsymbol{g}\boldsymbol{g}^T \circ YY^T \circ YY^T \right\rangle$$

3. Use noise property: $\max_{X \succcurlyeq 0, \operatorname{diag}(X) = \mathbf{1}} \langle E, X \rangle \leq (2 + o_d(1))n$

$$\|Y^T \boldsymbol{g}\|^2 \geq \|YY^T\|_F^2 - \frac{2n^2}{\lambda}(2 + o_d(1)) \geq n^2 \left(\frac{1}{2} - \frac{4 + o_d(1)}{\lambda} \right)$$

Note: Sufficient to ensure $\operatorname{Hess} f(Y) \succcurlyeq -\varepsilon \cdot \mathrm{I}$

More of this in statistics?

It’s a tempting family of estimators.

Toward computation bounds: see global rates of convergence to KKT points on manifolds: 1605.08101

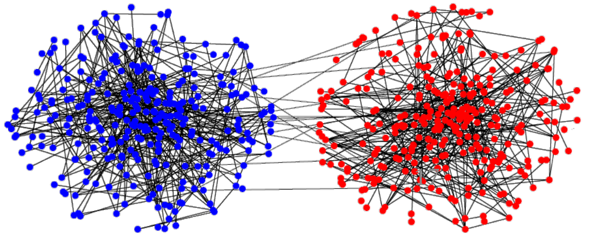
More guarantees for Burer–Monteiro approach to ‘smooth’ SDP’s: 1606.04970

Maximum likelihood estimation for SBM is combinatorial optimization

$A \in \mathbf{R}^{n\times n}$ is the observed adjacency matrix:

$$\max_{\boldsymbol{z}} \langle A, \boldsymbol{z}\boldsymbol{z}^T \rangle \quad \text{s.t.} \quad \boldsymbol{z} \in \{\pm 1\}^n \ \text{ and } \ \mathbf{1}^T\boldsymbol{z} = 0$$

Hard problem for general A: need to relax.



Step 1: remove the linear constraint

$$\max_{\boldsymbol{z}} \langle A, \boldsymbol{z}\boldsymbol{z}^T \rangle \quad \text{s.t.} \quad \boldsymbol{z} \in \{\pm 1\}^n \ \text{ and } \ \mathbf{1}^T\boldsymbol{z} = 0$$

If $\boldsymbol{g} \in \{\pm 1\}^n$ is the true partition, then

$$\mathbf{E}\{A\} = \frac{p+q}{2}\mathbf{1}\mathbf{1}^T + \frac{p-q}{2}\boldsymbol{g}\boldsymbol{g}^T.$$

Remove the bias toward $\mathbf{1}$: $A' = A - \frac{p+q}{2}\mathbf{1}\mathbf{1}^T$

Step 2: relax to a semidefinite program

$$\max_{\boldsymbol{z}} \left\langle A', \boldsymbol{z}\boldsymbol{z}^T \right\rangle \ \text{ s.t. } \ \boldsymbol{z} \in \{\pm 1\}^n$$

Binary constraints $\boldsymbol{z} \in \{\pm 1\}^n$
are equivalent to $\text{diag}\left(\boldsymbol{z}\boldsymbol{z}^T\right) = \mathbf{1}$.

Introduce $X = \boldsymbol{z}\boldsymbol{z}^T$. Equivalent problem:

$$\max_X \langle A', X \rangle \ \text{ s.t. } \text{diag}(X) = \mathbf{1}, X \succcurlyeq 0, \text{rank}(X) = 1$$