Convexified Modularity Maximization for Degree-corrected Stochastic Block Models

Jiaming Xu

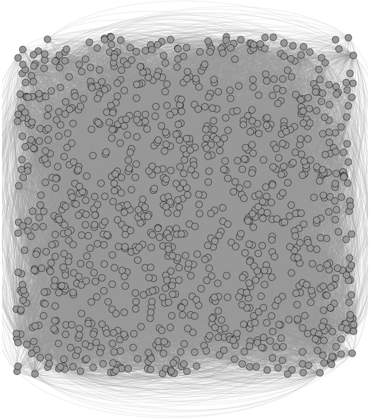
Simons Institute

Joint work with Yudong Chen (Cornell) and Xiaodong Li (UC Davis)

February 3, 2016

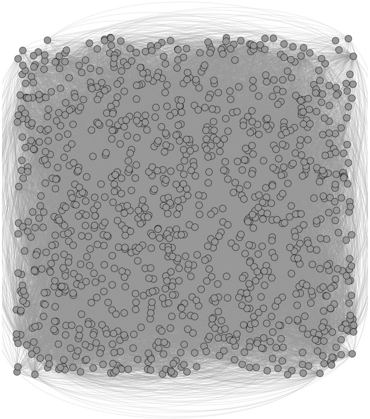
An example: Facebook friendship network

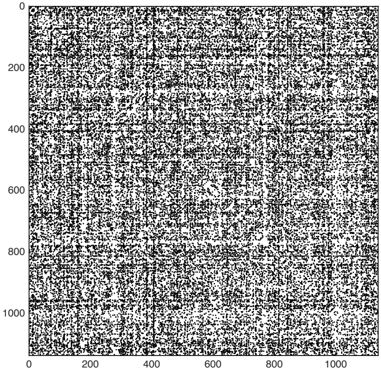
Simmons College network: 1137 students; 24257 undirected friend links [Traud-Mucha-Porter '12]



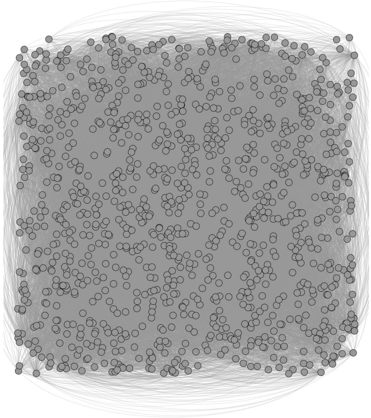
An example: Facebook friendship network

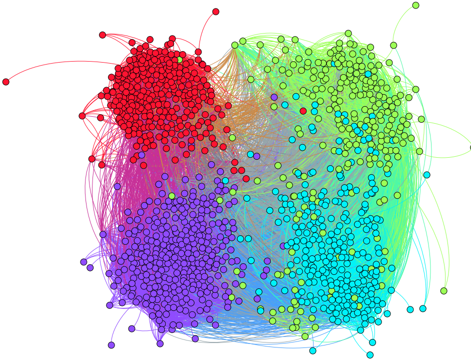
Simmons College network: 1137 students; 24257 undirected friend links [Traud-Mucha-Porter '12]



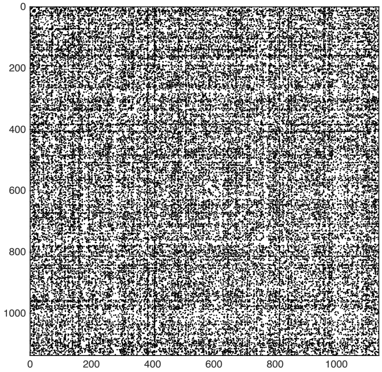


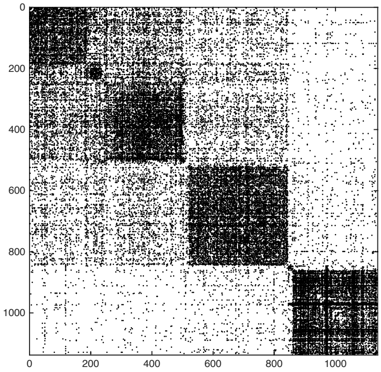
Apply our community detection algorithm



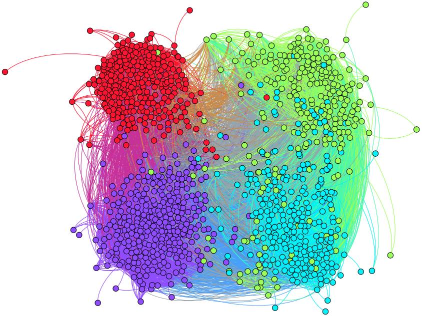


Sort adjacency matrix according to clustering result

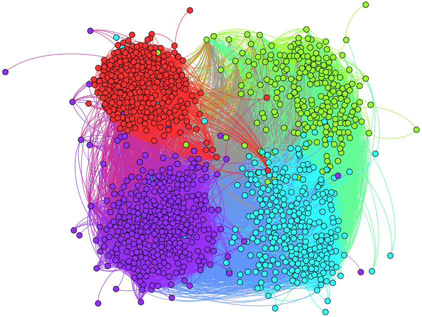




Clustering result has strong correlation with graduation year

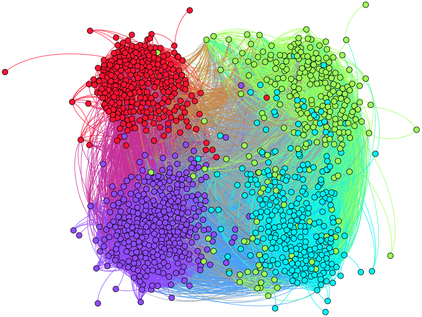


Color: clustering result

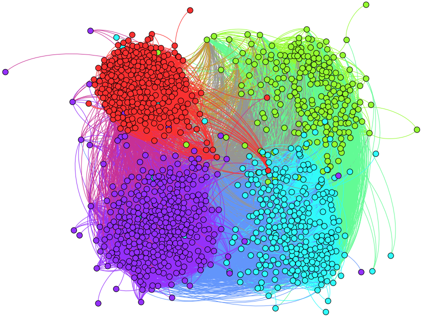


Color: graduation year

Clustering result has strong correlation with graduation year



Color: clustering result



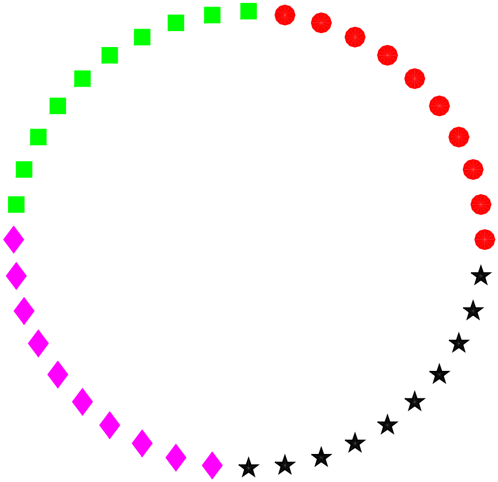
Color: graduation year

Our method misclassifies 12% of nodes
SCORE [Jin '15] and Regularized Spectral [Zhang-Levina-Zhu '14] misclassify 24% of nodes

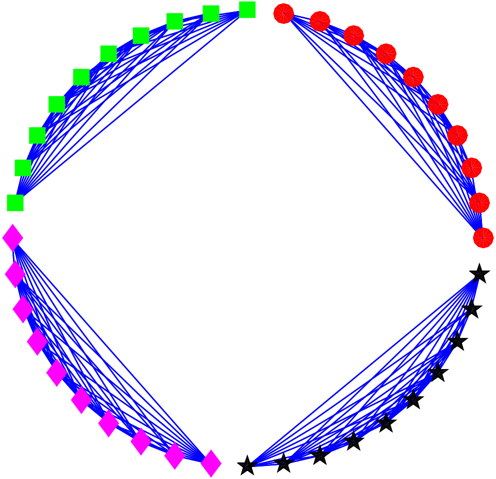
Outline of the talk

1. Models and previous work
2. Our algorithm
3. Theoretical guarantee
4. Empirical performance
5. Conclusions

Stochastic block model [Holland-Laskey-Leinhardt '83]
Planted partition model [Condon-Karp 01']

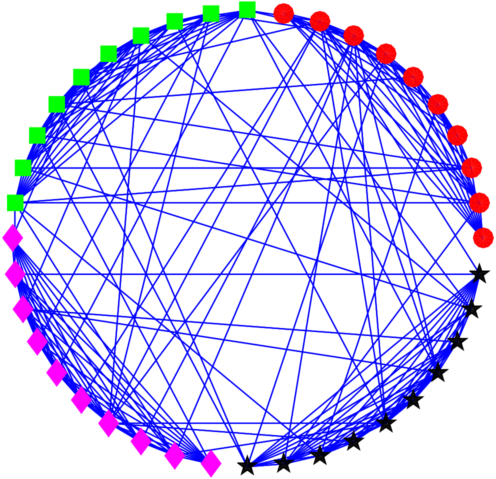


Stochastic block model [Holland-Laskey-Leinhardt ’83]
Planted partition model [Condon-Karp 01’]



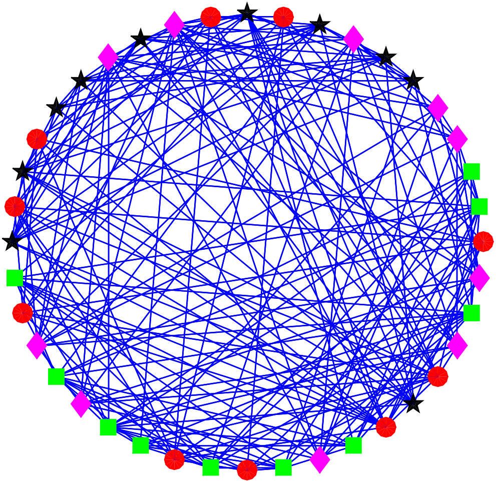
$p = 0.8$

Stochastic block model [Holland-Laskey-Leinhardt '83]
Planted partition model [Condon-Karp 01']



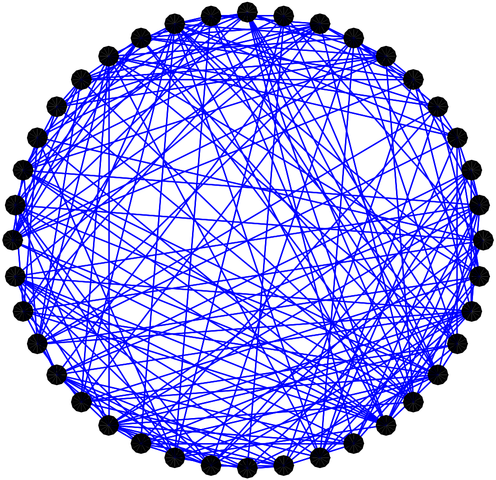
$p = 0.8 \quad q = 0.09$

Stochastic block model [Holland-Laskey-Leinhardt '83]
Planted partition model [Condon-Karp 01']



$p = 0.8 \quad q = 0.09$

Stochastic block model [Holland-Laskey-Leinhardt '83]
Planted partition model [Condon-Karp 01']



$$p = 0.8 \qquad q = 0.09$$

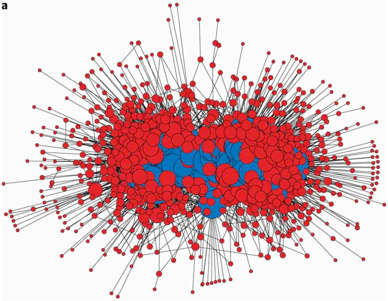
Main restriction of SBM

- All nodes in the same community are statistically equivalent
- Degrees are often highly inhomogeneous across nodes

Main restriction of SBM

- All nodes in the same community are statistically equivalent
- Degrees are often highly inhomogeneous across nodes

Political blog network [Adamic and Glance '05] [Karrer-Newman '11]:
Max degree 351, mean degree 27

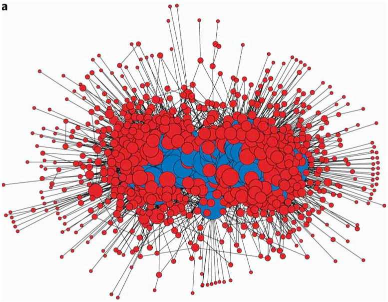


Fit SBM

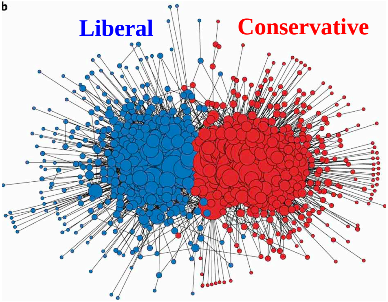
Main restriction of SBM

- All nodes in the same community are statistically equivalent
- Degrees are often highly inhomogeneous across nodes

Political blog network [Adamic and Glance '05] [Karrer-Newman '11]:
Max degree 351, mean degree 27



Fit SBM



True partition

Degree-corrected SBM [Karrer-Newman '11]
Extended PPM [Dasgupta-Hopcoft-McSherry '04]

Degree heterogeneity parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{R}_+^n$

- n nodes partitioned into k groups
- $i \sim j$ independently w.p. $\begin{cases} p\ \theta_i\theta_j & \text{if } i \text{ and } j \text{ in the same group} \\ q\ \theta_i\theta_j & \text{otherwise} \end{cases}$

Degree-corrected SBM [Karrer-Newman '11]
Extended PPM [Dasgupta-Hopcoft-McSherry '04]

Degree heterogeneity parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{R}_+^n$

- n nodes partitioned into k groups
- $i \sim j$ independently w.p. $\begin{cases} p\ \theta_i\theta_j & \text{if } i \text{ and } j \text{ in the same group} \\ q\ \theta_i\theta_j & \text{otherwise} \end{cases}$

Main challenges

- $\boldsymbol{\theta}$ is unknown
- $\theta_{\min} = \min_{1 \le i \le n} \theta_i$ could be small

Existing community detection algorithms for DCSBM

- Likelihood or modularity maximization

- Spectral methods

Existing community detection algorithms for DCSBM

- Likelihood or modularity maximization
  - Statistically efficient
    [Zhao-Levia-Zhu '12] [Amini-Chen-Bickel-Levina '13]
  - Computationally intractable
  - Efficient algorithms in restricted settings
    [Amini-Chen-Bickel-Levina '13] [Le-Levia-Vershynin '15]

- Spectral methods

Existing community detection algorithms for DCSBM

- Likelihood or modularity maximization
  - Statistically efficient
    [Zhao-Levia-Zhu '12] [Amini-Chen-Bickel-Levina '13]
  - Computationally intractable
  - Efficient algorithms in restricted settings
    [Amini-Chen-Bickel-Levina '13] [Le-Levia-Vershynin '15]

- Spectral methods
  - Statistically efficient
    [Qin-Rohe '13] [Jin '15] [Lei-Rinaldo '15]...
  - Computationally efficient
  - Inconsistent in sparse graphs [Krzakala et al. '13]
  - Sensitive to outliers [Cai-Li '15]

SDP relaxations of MLE under SBM

- Optimal recovery [Hajek-Wu-X. '14], [Bandeira '15]...
- Robust to adversaries [Feige-Kilian '01] [Cai-Li '15]
- Consistent in sparse graphs

  [Guedon-Vershynin '15] [Sen-Montanari '15]
- Computationally efficient [Javanmard-Montanari-Ricci-Tersenghi '15]

SDP relaxations of MLE under SBM

- Optimal recovery [Hajek-Wu-X. '14], [Bandeira '15]...
- Robust to adversaries [Feige-Kilian '01] [Cai-Li '15]
- Consistent in sparse graphs
  [Guedon-Vershynin '15] [Sen-Montanari '15]
- Computationally efficient [Javanmard-Montanari-Ricci-Tersenghi '15]

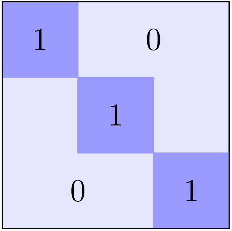
Does SDP relaxation also work well under DCSBM?

Convexified modularity maximization algorithm

Modularity maximization

- Modularity maximization (close to MLE under DCSBM) [Newman '06]

$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}$$



Modularity maximization

- Modularity maximization (close to MLE under DCSBM) [Newman '06]

$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}$$

A problem: fails to identify small communities

Modularity maximization

- Modularity maximization (close to MLE under DCSBM) [Newman '06]

$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}$$

A problem: fails to identify small communities

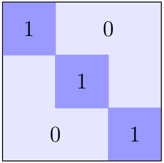
- Generalized modularity maximization [Reichartd-Bornholdt '06] [Lancichinetti-Fortunato '11]:

$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \lambda d_i d_j \right) Y_{ij}$$

SDP relaxations of modularity maximization

Generalized modularity maximization

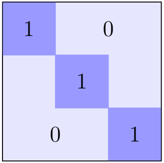
$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \lambda d_i d_j \right) Y_{ij}$$



SDP relaxations of modularity maximization

Generalized modularity maximization

$$\max_{\boldsymbol{Y}} \sum_{1 \le i,j \le n} \left(A_{ij} - \lambda d_i d_j\right) Y_{ij}$$



SDP relaxations

$$\begin{aligned} \max \ & \langle \boldsymbol{Y}, \boldsymbol{A} - \lambda \boldsymbol{d}\boldsymbol{d}^\top \rangle \\ \text{s.t.} \ & \boldsymbol{Y} \succeq \boldsymbol{0} \\ & Y_{ii} = 1 \quad i \in [n] \\ & \boldsymbol{0} \le \boldsymbol{Y} \le \boldsymbol{J} \end{aligned}$$

Weighted k-median clustering

Step 1: Defined weighted feature vectors

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{Y}} \text{diag}\left\{\boldsymbol{d}\right\}$$

Step 2: Clustering rows of $\widehat{\boldsymbol{W}}$:

$$\begin{aligned} \min \quad & \sum_{1 \le \ell \le k} \sum_{i \in C_\ell} d_i \|\widehat{\boldsymbol{W}}_{\boldsymbol{i}\bullet} - \boldsymbol{x}_{\boldsymbol{\ell}}\|_1 \\ \text{s.t.} \quad & \boldsymbol{x}_{\boldsymbol{\ell}} \subseteq \text{Rows}(\widehat{\boldsymbol{W}}) \end{aligned}$$

Weighted k-median clustering

Step 1: Defined weighted feature vectors

$$\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{Y}} \, \text{diag}\{\boldsymbol{d}\}$$

Step 2: Clustering rows of $\widehat{\boldsymbol{W}}$:

$$\begin{aligned} \min \quad & \sum_{1 \le \ell \le k} \sum_{i \in C_\ell} d_i \| \widehat{\boldsymbol{W}}_{\boldsymbol{i} \bullet} - \boldsymbol{x}_{\boldsymbol{\ell}} \|_1 \\ \text{s.t.} \quad & \boldsymbol{x}_{\boldsymbol{\ell}} \subseteq \text{Rows}(\widehat{\boldsymbol{W}}) \end{aligned}$$

Remarks

- New feature: weighing by degrees
- Exists polynomial-time $\frac{20}{3}$-factor approximation algorithm [Charikar-Guha-Tardos-Shmoys '99]

Theoretical guarantee

Focus on DCSBM with $\sum_{i \in C_\ell^*} \theta_i \equiv g$

Why we expect SDP to work?

SDP shall succeed if no noise

$$\boldsymbol{Y}^* = \arg\max \ \langle \boldsymbol{Y}, \mathbb{E}[\boldsymbol{A}] - \lambda \mathbb{E}[\boldsymbol{d}]\, \mathbb{E}[\boldsymbol{d}]^\top \rangle$$
$$\text{s.t.} \quad \text{SDP constraints}$$

Why we expect SDP to work?

SDP shall succeed if no noise

$$\boldsymbol{Y}^* = \arg\max \ \langle \boldsymbol{Y}, \mathbb{E}[\boldsymbol{A}] - \lambda \mathbb{E}[\boldsymbol{d}] \mathbb{E}[\boldsymbol{d}]^\top \rangle$$

s.t. SDP constraints

Density gap condition

$$\frac{p+3q}{4} \le \lambda \cdot c < \frac{3p+q}{4}$$

$$c = [p + (k-1)q]^2 g^2$$

Why we expect SDP to work?

SDP shall succeed if no noise

$$\boldsymbol{Y}^* = \arg\max \ \langle \boldsymbol{Y}, \mathbb{E}[\boldsymbol{A}] - \lambda \mathbb{E}[\boldsymbol{d}]\, \mathbb{E}[\boldsymbol{d}]^\top \rangle \quad \text{s.t.} \quad \text{SDP constraints}$$

Density gap condition

$$\frac{p+3q}{4} \le \lambda \cdot c < \frac{3p+q}{4}$$

$$c = [p + (k-1)q]^2 g^2$$

Signal-noise decomposition

$$\langle \boldsymbol{Y} - \boldsymbol{Y}^*, \boldsymbol{A} - \lambda \boldsymbol{d}\boldsymbol{d}^\top \rangle = \langle \boldsymbol{Y} - \boldsymbol{Y}^*, \mathbb{E}[\boldsymbol{A}] - \lambda \mathbb{E}[\boldsymbol{d}]\, \mathbb{E}[\boldsymbol{d}]^\top \rangle + \text{noise part}$$

Approximate and exact recovery

Assume the *density gap condition* holds

Theorem (Approximate recovery)

Let S denotes the set of misclassified nodes.

$$\frac{1}{kg}\sum_{i\in S}\theta_i \lesssim \frac{n/g + k\sqrt{np}}{(p-q)g}$$

Approximate and exact recovery

Assume the *density gap condition* holds

Theorem (Approximate recovery)

Let S denotes the set of misclassified nodes.

$$\frac{1}{kg}\sum_{i\in S}\theta_i \lesssim \frac{n/g + k\sqrt{np}}{(p-q)g}$$

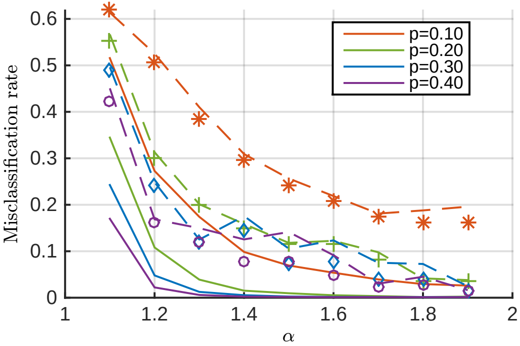
Theorem (Exact recovery)

With high probability $\widehat{\boldsymbol{Y}} = \boldsymbol{Y}^*$, if

$$(p-q)g \gtrsim \sqrt{nq} + \sqrt{\frac{pg\log n}{\theta_{\min}}}$$

Empirical performance

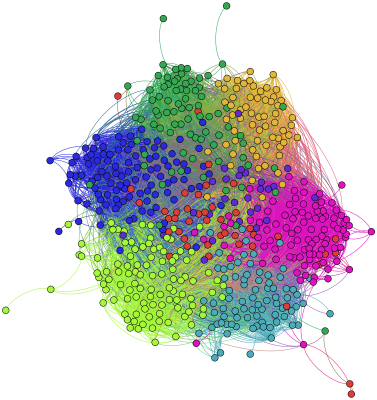
Setup: $\theta_i \overset{\text{i.i.d.}}{\sim}$ power law with exponent α



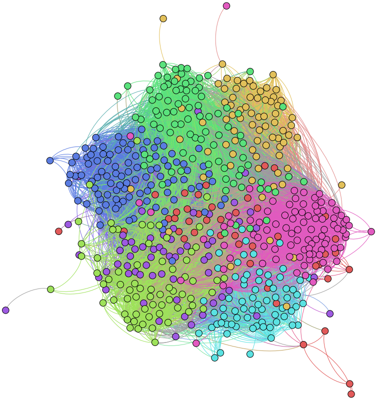
$k = 4$ and $q = 0.3p$

CMM with $\lambda = 1/\sum_i d_i$ (Solid)
SCORE [Jin '15] (Dashed)
Regularized Spectral [Zhang-Levina-Zhu '14] (Markers)

Experiment on Caltech friendship network

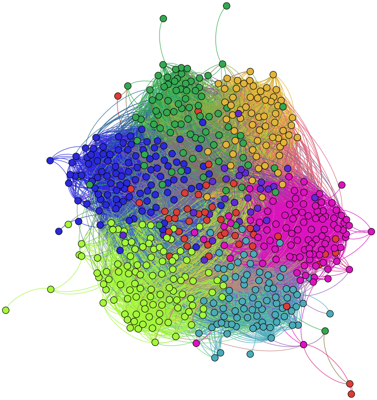


Color: clustering result of CMM

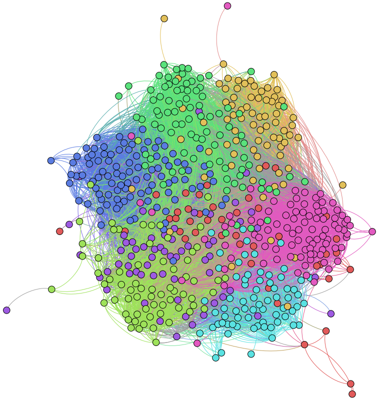


Color: Dorm partition

Experiment on Caltech friendship network



Color: clustering result of CMM



Color: Dorm partition

	CMM	SCORE	RegualarizedSpectral
mis. frac.	21%	31%	32%

Conclusion

Convexified modularity maximization

- Statistically and computationally efficient
- Provably works well even in sparse graphs
- Good empirical performance

Convexified modularity maximization

- ☺ Statistically and computationally efficient
- ☺ Provably works well even in sparse graphs
- ☺ Good empirical performance

Reference

Yudong Chen, Xiaodong Li, and Jiaming Xu, Convexified modularity maximization for degree-corrected stochastic block models. arXiv:1512.08425, Dec. 2015

Code

Available at http://people.orie.cornell.edu/yudong.chen/cmm