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

Our method misclassifies 12% 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 \cite{Holland-Laskey-Leinhardt '83}

Planted partition model \cite{Condon-Karp 01'}

\[ p = 0.8 \quad q = 0.09 \]
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

Liberal Conservative
Liberal Conservative
True partition
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 heterogeneity parameter $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n_+$

- $n$ nodes partitioned into $k$ groups
- $i \sim j$ independently w.p. $\begin{cases} \frac{p}{q} \theta_i \theta_j & \text{if } i \text{ and } j \text{ in the same group} \\ \frac{q}{q} \theta_i \theta_j & \text{otherwise} \end{cases}$
Degree-heterogeneity parameter $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n_+$

- $n$ 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
- $\theta$ is unknown
- $\theta_{\text{min}} = \min_{1 \leq i \leq n} \theta_i$ could be small
Existing community detection algorithms for DCSBM

- Likelihood or modularity maximization
  - Statistically efficient
  - \cite{Zhao-Levia-Zhu '12, Amini-Chen-Bickel-Levina '13}
  - Computationally intractable
  - Efficient algorithms in restricted settings
  - \cite{Amini-Chen-Bickel-Levina '13, Le-Levia-Vershynin '15}

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

\[
\max_Y \sum_{1 \leq i, j \leq n} \left( A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}
\]
• Modularity maximization (close to MLE under DCSBM) [Newman ’06]

\[
\max_Y \sum_{1 \leq i, j \leq n} \left( A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}
\]

A problem: fails to identify small communities
Modularity maximization

- **Modularity maximization** (close to MLE under DCSBM) [Newman '06]
  \[
  \max_Y \sum_{1 \leq i, j \leq n} \left( A_{ij} - \frac{d_i d_j}{\sum_i d_i} \right) Y_{ij}
  \]

  A problem: fails to identify small communities

- **Generalized modularity maximization**
  [Reichartd-Bornholdt '06] [Lancichinetti-Fortunato '11]:
  \[
  \max_Y \sum_{1 \leq i, j \leq n} \left( A_{ij} - \lambda d_i d_j \right) Y_{ij}
  \]
SDP relaxations of modularity maximization

Generalized modularity maximization

$$\max_Y \sum_{1 \leq i, j \leq n} (A_{ij} - \lambda d_i d_j) Y_{ij}$$
Generalized modularity maximization

\[
\max_Y \sum_{1 \leq i,j \leq n} (A_{ij} - \lambda d_i d_j) Y_{ij}
\]

SDP relaxations

\[
\begin{align*}
\max \quad & \langle Y, A - \lambda dd^\top \rangle \\
\text{s.t.} \quad & Y \succeq 0 \\
& Y_{ii} = 1 \quad i \in [n] \\
& 0 \leq Y \leq J
\end{align*}
\]
Weighted $k$-median clustering

Step 1: Defined weighted feature vectors

$$\hat{W} = \hat{Y} \text{diag}\{d\}$$

Step 2: Clustering rows of $\hat{W}$:

$$\min \sum_{1 \leq \ell \leq k} \sum_{i \in C_\ell} d_i \| \hat{W}_{i\bullet} - x_\ell \|_1$$

s.t. $x_\ell \subseteq \text{Rows}(\hat{W})$
Weighted $k$-median clustering

Step 1: Defined weighted feature vectors

$$\hat{W} = \hat{Y} \text{diag}\{d\}$$

Step 2: Clustering rows of $\hat{W}$:

$$\min \sum_{1 \leq \ell \leq k} \sum_{i \in C_\ell} d_i \| \hat{W}_{i\bullet} - x_\ell \|_1$$

s.t. $x_\ell \subseteq \text{Rows} (\hat{W})$

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 \( \sum_{i \in C^*_l} \theta_i \equiv g \)
Why we expect SDP to work?

SDP shall succeed if no noise

\[ Y^* = \arg \max \langle Y, E[A] - \lambda E[d]E[d]^\top \rangle \]

s.t. SDP constraints
Why we expect SDP to work?

SDP shall succeed if no noise

\[ Y^* = \arg \max \langle Y, E[A] - \lambda E[d] E[d]^\top \rangle \]

s.t. SDP constraints

Density gap condition

\[ \frac{p + 3q}{4} \leq \lambda \cdot c < \frac{3p + q}{4} \]

\[ c = [p + (k - 1)q]^2 g^2 \]
Why we expect SDP to work?

SDP shall succeed if no noise

\[ Y^* = \arg \max Y, E[A] - \lambda E[d] E[d]^\top \]

s.t. SDP constraints

Density gap condition

\[
\frac{p + 3q}{4} \leq \lambda \cdot c < \frac{3p + q}{4}
\]

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

Signal-noise decomposition

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

$$
(p - q)g \gtrsim \sqrt{nq} + \sqrt{\frac{pg \log n}{\theta_{\min}}}
$$
Empirical performance
Experiment on synthetic networks

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

$k = 4$ 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

<table>
<thead>
<tr>
<th></th>
<th>CMM</th>
<th>SCORE</th>
<th>ReguarlizedSpectral</th>
</tr>
</thead>
<tbody>
<tr>
<td>mis. frac.</td>
<td>21%</td>
<td>31%</td>
<td>32%</td>
</tr>
</tbody>
</table>
Conclusion

Convexified modularity maximization

- 😊 Statistically and computationally efficient
- 😊 Provably works well even in \textit{sparse} graphs
- 😊 Good empirical performance
Conclusion

Convexified modularity maximization

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

Reference

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