Finding hidden cliques and clusters by convex optimization

Brendan Ames

Computing + Mathematical Sciences
California Institute of Technology

CUNY College of Staten Island
March 20, 2014
Agenda

• Present a convex relaxation for the maximum clique problem based on nuclear norm relaxation for rank minimization.

• Extend this relaxation to one for the densest $k$-subgraph and graph clustering problems.

• Give a probabilistic model for “clusterable” data and graphs, and theoretical recovery guarantees.

• Experimental results and open problems.

• Joint work with Stephen Vavasis, University of Waterloo, and Ting-Kei Pong, University of British Columbia.
Clustering

- **Clustering**: partition data so that items in each cluster are similar to each other and items not in the same cluster are dissimilar.

- Fundamental problem in statistics and machine learning:
  - pattern recognition, computational biology, image processing/computer vision, network analysis.

- No consensus on what constitutes a *good* clustering; depends heavily on application.

- **Intractable**: usually modeled as some NP-hard problem (e.g. clique, normalized cut, k-means).
A sanity check

- Clustering seems to be a very difficult/ill-posed problem.

- Many heuristics seem to work well in practice.

- **Question:** can we show that we can cluster “clusterable” data? How do we model clusterable data?
Graph clustering

- **Similarity Graph**: represent data set as a graph
  - items = nodes
  - edges indicate similarity

- Cluster the data set by dividing the graph into dense subgraphs.

- Dense = large average degree
Example: Communities in Social Networks

- Nodes = users
- Edges = “friendship”.
- Densely connected groups = communities
Cliques of a graph

- Given graph $G = (V, E)$, a clique of $G$ is a pairwise adjacent subset of $V$.

- $C \subseteq V$ is a clique of $G$ if $uv \in E$ for all $u, v \in C$.

- The subgraph $G(C)$ induced by $C$ is complete.
Cliques of a graph

- Given graph $G = (V, E)$, a clique of $G$ is a pairwise adjacent subset of $V$.

- $C \subseteq V$ is a clique of $G$ if $uv \in E$ for all $u, v \in C$.

- The subgraph $G(C)$ induced by $C$ is complete.
The Maximum Clique problem

- **Optimization version**: Find largest clique of $G$; size of the largest clique is the clique number $\omega(G)$.

- **Decision version**: Given graph $G$, integer $k$: does $G$ contain a clique of cardinality at least $k$?

- **Complexity**: NP-complete, NP-hard to approximate $\omega(G)$ within a ratio of $N^{1-\epsilon}$ for any $\epsilon > 0$, ($N = |V|$)

- **Many applications**: communication, power, and social networks, mathematical biology, cryptography.
Worst case vs planted case

- Complexity results are **worst case**.
- There are some graphs for which finding the maximum clique is as difficult as hardest problems known.
- Not necessarily true for all graphs.
- We focus on identifying trade-offs between clique size and density of nonclique edges that ensure we can solve **Clique** efficiently.
Matrix representation of cliques

- Characteristic vector of $C$: vector $v \in \{0, 1\}^V$ with
  $$v_i = 1 \text{ if } i \in C \text{ and } v_i = 0 \text{ otherwise}$$
- If $C$ is a clique with characteristic vector $v$, let $X = vv^T$.
- Nonzero entries of $X$ form a $|C| \times |C|$ all-ones block in $A_G + I$. 
Matrix representation of cliques

- Characteristic vector of $C$: vector $v \in \{0, 1\}^V$ with
  \[
  v_i = 1 \text{ if } i \in C \quad \text{and} \quad v_i = 0 \text{ otherwise}
  \]
- If $C$ is a clique with characteristic vector $v$, let $X = vv^T$.
- Nonzero entries of $X$ form a $|C| \times |C|$ all-ones block in $A_G + I$. 
Matrix representation of cliques (2)

- $X \in \{0, 1\}^{V \times V}$ corresponds to a $k$-clique if and only if
  
  - it is the adjacency matrix of a subgraph of $G$:
    
    $$X_{ij} = 0 \text{ if } i \neq j \text{ or } ij \notin E,$$
  
  - it contains $k^2$ nonzero entries ($\binom{k}{2}$ edges):
    
    $$\sum_{i,j \in V} X_{ij} = k^2,$$
  
  - it is symmetric, and
  
  - $\text{rank}(X) = 1$
Clique as rank minimization

- $G$ has a $k$-clique if and only if there exists rank-one symmetric binary matrix $X$ such that

$$
\sum \sum X_{ij} = k^2
$$

$$
X_{ij} = 0 \quad \forall \; ij \notin E, \; i \neq j.
$$

- Clique is equivalent to the rank minimization problem:

$$
\min_{X \in \{0,1\}^{V \times V}} \left\{ \text{rank}(X) : e^T X e = k^2, X_{ij} = 0 \; \text{if} \; (i, j) \in \tilde{E} \right\}
$$

where $\tilde{E} = V \times V - \{E \cup \{(u, u) : u \in V\}\}$. 
Affine Rank minimization

- **Affine rank minimization problem**: find minimum rank solution of a given set of linear equations:

  \[ \min \{ \text{rank}(X) : A(X) = b \} \]

- Well-known to be NP-hard.

- **Fazel 2002**: Relax using the nuclear norm \( \| X \|_* \)

  \[ \text{rank}(X) = \| \sigma(X) \|_0 \implies \| X \|_* = \| \sigma(X) \|_1 = \sum_i \sigma_i(X) \]
Geometry of nuclear norm relaxation

- Relax by replacing $\text{rank}(X)$ with $\|X\|_* = \sum \sigma_i(X)$

- $\| \cdot \|_*$ is the **convex envelope** of $\text{rank}$ on $B = \{ X : \|X\| \leq 1 \}$.

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \text{ in } \mathbb{R}^{2 \times 2}$$

$\text{rank}(X) = 1$

$|x + z| = 1$
Geometry of nuclear norm relaxation

- Relax by replacing $\text{rank}(X)$ with $\|X\|_* = \sum \sigma_i(X)$

- $\| \cdot \|_*$ is the convex envelope of $\text{rank}$ on $B = \{ X : \|X\| \leq 1 \}$.

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \text{ in } \mathbb{R}^{2 \times 2}$$

- $\text{rank}(X) = 1$
- $|x + z| = 1$
Guaranteed low-rank recovery

- Recht, Fazel, Parrilo 2007, Candés and Recht 2008:
  If $A$ and $b$ are given by sufficiently many ($m = \Omega(r n \log n)$) random observations of a rank-$r$ $n \times n$ matrix then
  \[
  \arg\min\{\text{rank}(X) : A(X) = b\} = \arg\min\{\|X\|_* : A(X) = b\}
  \]

\[
X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \text{ in } \mathbb{R}^{2 \times 2}
\]

\[
\text{rank}(X) = 1 \quad |x + z| = 1
\]
Nuclear norm relaxation of Clique

• We have the rank minimization problem:

\[
\min_{X \in \{0,1\}^{V \times V}, X \in \Sigma^V} \left\{ \text{rank}(X) : \mathbf{e}^T X \mathbf{e} = k^2, \ X_{ij} = 0 \text{ if } (i, j) \in \tilde{E} \right\}
\]

• Relax rank with the nuclear norm, drop binary and symmetry constraints:

\[
\min \left\{ \|X\|_\ast : \mathbf{e}^T X \mathbf{e} = k^2, \ X_{ij} = 0 \text{ if } (i, j) \in \tilde{E} \right\} \quad \text{(NNR)}
\]

• When does the solution of the relaxation coincide with that of the rank minimization problem?
The planted case

Construction:
- Add edges between nodes in vertex set $V^*$ of size $k$. 
The planted case

Construction:

• Add edges between nodes in vertex set $V^*$ of size $k$.

• Each remaining edge is added to $E$ independently with fixed probability $p$. 

The planted case

Construction:

- Add edges between nodes in vertex set $V^*$ of size $k$.

- Each remaining edge is added to $E$ independently with fixed probability $p$.

- $V^*$ is a clique of $G$ (called a planted or hidden clique).
The planted case

Construction:

- Add edges between nodes in vertex set \( V^* \) of size \( k \).
- Each remaining edge is added to \( E \) independently with fixed probability \( p \).
- \( V^* \) is a clique of \( G \) (called a planted or hidden clique).
The planted case

Construction:

- Add edges between nodes in vertex set $V^*$ of size $k$.
- Each remaining edge is added to $E$ independently with fixed probability $p$.
- $V^*$ is a clique of $G$ (called a planted or hidden clique).
The planted case

Construction:

• Add edges between nodes in vertex set $V^*$ of size $k$.

• Each remaining edge is added to $E$ independently with fixed probability $p$.

• $V^*$ is a clique of $G$ (called a planted or hidden clique).
Recovery guarantee in the planted case

Theorem (Ames-Vavasis 2009)
There exists scalar $c > 0$ (depending only on $p$) such that if

$$k \geq c\sqrt{N}$$

then

- $V^*$ is the unique maximum clique of $G$, and
- $X^* = vv^T$ is the unique optimal solution of $(NNR)$

with high probability.
What happens if edges are deleted?

- Guarantee does not tolerate edge deletion noise.
- Suppose edge $uv$ is deleted for some $u, v \in V^*$.
- Then $V^*$ is not a clique and $X^*$ is not feasible for $(\text{NNR})$. 
What happens if edges are deleted?

- Guarantee does not tolerate edge deletion noise.
- Suppose edge $uv$ is deleted for some $u, v \in V^*$.
- Then $V^*$ is not a clique and $X^*$ is not feasible for (NJR).
The densest k-subgraph problem

- Want a **dense** subgraph of size $k$, not necessarily a clique.

- **Densest $k$-subgraph problem (DKS):** find subgraph $H \subseteq G$ on $k$ nodes with maximum density:

  $$d(H) = \frac{|E(H)|}{|V(H)|} = \frac{|E(H)|}{k}.$$

- **NP-hard:** proof is by reduction to Clique; hard to approximate.

- Maximizing $d(H) = \text{maximizing } |E(H)|$ over all $k$-node subgraphs.
Duality of density and number of missing edges

- Let $V^* \subseteq V$ be a $k$-subset with characteristic vector $v$.

- Introduce a correction $Y$ for entries of $X = vv^T$ that should be 0:

  $$Y_{ij} = \begin{cases} -X_{ij}, & \text{if } ij \in \tilde{E} \\ 0, & \text{otherwise.} \end{cases}$$

- If $V^*$ is almost a clique then $G(V^*)$ should be very dense and $Y$ should be very sparse.

- Cardinality of $Y$ acts as a dual of density of $G(V^*)$:

  $$|E(G(V^*))| = \binom{k}{2} - \frac{\|Y\|_0}{2}$$
Can relax (DKS) as

$$\min \| X \|_* + \gamma \| Y \|_1 \quad \text{st} \quad e^T X e = k^2$$

$$X_{ij} + Y_{ij} = 0 \text{ if } ij \in \tilde{E}$$

$$X \in [0, 1]^{V \times V}$$

(DKSR)

where $\gamma$ is a regularization parameter.

Relax $\| Y \|_0$ using the $\ell_1$-norm $\| Y \|_1$, rank$(X)$ with the nuclear norm $\| X \|_*$. 
Planted case

- Start with set of $N$ nodes $V$.
- Add all edges between nodes in $V^* \subseteq V$.
- Add noise:
  - Add potential edges with probability $p$
  - Delete some edges in $V^* \times V^*$ with probability $q$
Planted case

• Start with set of $N$ nodes $V$.

• Add all edges between nodes in $V^* \subseteq V$.

• Add noise:
  • Add potential edges with probability $p$
  • Delete some edges in $V^* \times V^*$ with probability $q$
Planted case

- Start with set of $N$ nodes $V$.
- Add all edges between nodes in $V^* \subseteq V$.
- Add noise:
  - Add potential edges with probability $p$
  - Delete some edges in $V^* \times V^*$ with probability $q$
Theorem (Ames 2013)
There exist constants $c_1, c_2, c_3 > 0$ depending on $p, q$ such that if

$$k \geq c_1 \sqrt{N}$$

and $\gamma \in (c_2/k, c_3/k)$ then

- $G(V^*)$ is the unique densest $k$-subgraph of $G$, and
- $X^* = vv^T$ is the unique optimal solution of $(\text{DKSR})$ with high probability.
The Weighted Similarity Graph

- Given data and affinity function $f$ indicating similarity between any two items.

- Can model the data as weighted similarity graph $G_S = (V, E, W)$ as follows:
  
  - Each item is represented by a node in $V$.
  - We add an edge between each pair of two nodes $i, j$ with edge weight $W_{ij} = f(i, j)$.
  - $W_{ij}$ is large if $i$ and $j$ are highly similar.
Example: Communities in Social Networks

- Nodes = users
- Edges = “friendship”.
- Densely connected groups = communities
Example: Clustered Euclidean data

- Suppose each data point in the $i$th cluster $C_i$ is placed uniformly at random in a ball centered at $c_i \in \mathbb{R}^d$.
- Distance within clusters will be small compared to the distance between clusters if centers are well-separated.
- Choose $W_{ij} = \exp(-\|x^i - x^j\|^2)$. 
Example: Clustered Euclidean data

- Suppose each data point in the $i$th cluster $C_i$ is placed uniformly at random in a ball centered at $c_i \in \mathbb{R}^d$.

- Distance within clusters will be small compared to the distance between clusters if centers are well-separated.

- Choose $W_{ij} = \exp(-\|x^i - x^j\|^2)$. 
The Densest $r$-Disjoint Clique Problem

- To cluster the data we want to partition the graph into cliques with heavy support.

- A $r$-disjoint-clique subgraph of a graph $G$ is a subgraph of $G$ induced by $r$ disjoint cliques.

- Densest $r$-disjoint-clique problem (RDC): find a $r$-disjoint-clique subgraph such that the sum of the densities of the $r$ complete subgraphs induced by the cliques is maximized.

- Density of complete subgraph induced by $C$:

$$d(C) = \frac{1}{|C|} \sum_{i \in C} \sum_{j \in C} W_{ij} = \frac{v^T W v}{v^T v}$$

where $v$ is the characteristic vector of $C$. 

Lifting procedure for RDC

- Let \( \{C_1, \ldots, C_r\} \) define a \( r \)-disjoint-clique subgraph with characteristic vectors \( \{v_1, v_2, \ldots, v_r\} \)

- Lift the \( r \) characteristic vectors \( \{v_1, v_2, \ldots, v_r\} \) to the rank-\( r \) matrix variable \( X \):

\[
X = \sum v_i v_i^T \frac{1}{\|v_i\|^2}.
\]

- \( X \) has trace exactly equal to \( r \) and satisfies

\[
\text{Tr}(WX) = \sum \frac{v_i^T W v_i}{\|v_i\|^2}
\]
Lifted solutions

Lifted solution $X$ must satisfy:

- Inlier rows sum to 1. Outlier rows equal 0:
  \[ X \mathbf{e} \leq \mathbf{e} \]

- $X$ is symmetric doubly nonnegative:
  \[ X_{ij} \geq 0, \quad X \succeq 0 \]

- \( \text{rank}(X) = \text{Tr}(X) = r \)

- other combinatorial constraints
Relaxation to SDP

- Ignoring rank constraint and relaxing combinatorial
consstraints on $X$ gives the semidefinite program:

$$\begin{align*}
\text{max} & \quad \text{Tr}(WX) \\
\text{st} & \quad Xe \leq e \\
& \quad \text{Tr}(X) = r \\
& \quad X_{ij} \geq 0 \quad \forall \ i, j \in V \\
& \quad X \succeq 0.
\end{align*}$$

- $\text{Tr}(X) = \|X\|_* = r$ acts a surrogate for $\text{rank}(X) = r$
The Planted cluster model

Randomly generate weights $W \in [0, 1]^{N \times N}$ according to the following model:

- Start with clusters $C_1, \ldots, C_r$ of sizes $k_1, \ldots, k_r$.
- Sample entries of $W(C_i, C_i)$ i.i.d. from probability distribution $\Omega_1$ with mean $\alpha$.
- Sample remaining entries of $W$ i.i.d. from distribution $\Omega_2$ with mean $\beta << \alpha$. 


Guaranteed recovery

- Let $X^* = \sum_{i=1}^{r} \frac{v_i v_i^T}{k_i}$ be the lifted solution corresponding to the planted clusters $C_1, \ldots, C_r$.

Theorem (Ames 2012)

There exist scalars $c_1, c_2, c_3, c_4 > 0$ such that if

$$c_1 \sqrt{N} + c_2 \sqrt{rk_{r+1}} + c_3 k_{r+1} \leq c_4 \hat{k},$$

where $\hat{k} = \min\{k_1, k_2, \ldots, k_r\}$ and $k_{r+1}$ is the number of outliers, then

- $X^*$ is the unique optimal solution of SDP relaxation, and
- $\{C_1, \ldots, C_r\}$ is the unique densest $r$-disjoint-clique subgraph

with high probability.
Rehnquist Supreme Court

- Data drawn from U.S. Supreme Court decisions (from 1994-95 to 2003-04).
- First consider by Hubert and Steinley 2005.
- Assign edge-weights corresponding to fraction of decisions on which Justices agreed:

<table>
<thead>
<tr>
<th></th>
<th>St</th>
<th>Br</th>
<th>Gi</th>
<th>So</th>
<th>Oc</th>
<th>Ke</th>
<th>Re</th>
<th>Sc</th>
<th>Th</th>
</tr>
</thead>
<tbody>
<tr>
<td>St</td>
<td>1</td>
<td>0.62</td>
<td>0.66</td>
<td>0.63</td>
<td>0.33</td>
<td>0.36</td>
<td>0.25</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>Br</td>
<td>0.62</td>
<td>1</td>
<td>0.72</td>
<td>0.71</td>
<td>0.55</td>
<td>0.47</td>
<td>0.43</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>Gi</td>
<td>0.66</td>
<td>0.72</td>
<td>1</td>
<td>0.78</td>
<td>0.47</td>
<td>0.49</td>
<td>0.43</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>So</td>
<td>0.63</td>
<td>0.71</td>
<td>0.78</td>
<td>1</td>
<td>0.55</td>
<td>0.5</td>
<td>0.44</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Oc</td>
<td>0.33</td>
<td>0.55</td>
<td>0.47</td>
<td>0.55</td>
<td>1</td>
<td>0.67</td>
<td>0.71</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Ke</td>
<td>0.36</td>
<td>0.47</td>
<td>0.49</td>
<td>0.5</td>
<td>0.67</td>
<td>1</td>
<td>0.77</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>Re</td>
<td>0.25</td>
<td>0.43</td>
<td>0.43</td>
<td>0.44</td>
<td>0.71</td>
<td>0.77</td>
<td>1</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>Sc</td>
<td>0.14</td>
<td>0.25</td>
<td>0.28</td>
<td>0.31</td>
<td>0.54</td>
<td>0.58</td>
<td>0.66</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>Th</td>
<td>0.15</td>
<td>0.24</td>
<td>0.26</td>
<td>0.29</td>
<td>0.54</td>
<td>0.59</td>
<td>0.68</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>
Rehnquist Supreme Court (2)

- Solve RDC with $r = 2$ to get the following partition of the Supreme court:

<table>
<thead>
<tr>
<th>1: “Liberal”</th>
<th>2: “Conservative”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stevens (St)</td>
<td>O’Connor (Oc)</td>
</tr>
<tr>
<td>Breyer (Br)</td>
<td>Kennedy (Ke)</td>
</tr>
<tr>
<td>Ginsberg (Gi)</td>
<td>Rehnquist (Re)</td>
</tr>
<tr>
<td>Souter (So)</td>
<td>Scalia (Sc)</td>
</tr>
<tr>
<td></td>
<td>Thomas (Th)</td>
</tr>
</tbody>
</table>
Rehnquist Supreme Court (3)

<table>
<thead>
<tr>
<th></th>
<th>St</th>
<th>Br</th>
<th>Gi</th>
<th>So</th>
<th>Oc</th>
<th>Ke</th>
<th>Re</th>
<th>Sc</th>
<th>Th</th>
</tr>
</thead>
<tbody>
<tr>
<td>St</td>
<td>1</td>
<td>0.62</td>
<td>0.66</td>
<td>0.63</td>
<td>0.33</td>
<td>0.36</td>
<td>0.25</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>Br</td>
<td>0.62</td>
<td>1</td>
<td>0.72</td>
<td>0.71</td>
<td>0.55</td>
<td>0.47</td>
<td>0.43</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>Gi</td>
<td>0.66</td>
<td>0.72</td>
<td>1</td>
<td>0.78</td>
<td>0.47</td>
<td>0.49</td>
<td>0.43</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>So</td>
<td>0.63</td>
<td>0.71</td>
<td>0.78</td>
<td>1</td>
<td>0.55</td>
<td>0.5</td>
<td>0.44</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Oc</td>
<td>0.33</td>
<td>0.55</td>
<td>0.47</td>
<td>0.55</td>
<td>1</td>
<td>0.67</td>
<td>0.71</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Ke</td>
<td>0.36</td>
<td>0.47</td>
<td>0.49</td>
<td>0.5</td>
<td>0.67</td>
<td>1</td>
<td>0.77</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>Re</td>
<td>0.25</td>
<td>0.43</td>
<td>0.43</td>
<td>0.44</td>
<td>0.71</td>
<td>0.77</td>
<td>1</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>Sc</td>
<td>0.14</td>
<td>0.25</td>
<td>0.28</td>
<td>0.31</td>
<td>0.54</td>
<td>0.58</td>
<td>0.66</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>Th</td>
<td>0.15</td>
<td>0.24</td>
<td>0.26</td>
<td>0.29</td>
<td>0.54</td>
<td>0.59</td>
<td>0.68</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>
Rehnquist Supreme Court (4)

- Algorithm is sensitive to choice of $r$.
- Solve with $r = 3$:

<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thomas (Th)</td>
<td>O’Connor (Oc)</td>
<td>Stevens (St)</td>
</tr>
<tr>
<td>Scalia (Sc)</td>
<td>Kennedy (Ke)</td>
<td>Breyer (Br)</td>
</tr>
<tr>
<td></td>
<td>Rehnquist (Re)</td>
<td>Ginsberg (Gi)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Souter (So)</td>
</tr>
</tbody>
</table>
The NCAA Football Network

- First considered in Girvan and Newman 2001.
- Nodes are the 115 Division I football teams.
- Teams are deemed adjacent if they played at least one game against each other in the 2000 season.
- Games are typically scheduled based on membership within athletic conferences and geography.
- Should display community structure based on membership in conferences.
The NCAA Football Network (2)

- The node-node adjacency matrix for the football network.
The NCAA Football Network (3)

- The adjacency matrix with rows/cols reordered according to conferences.
The NCAA Football Network (4)

- Solve instance of RDC with $W = A$, $r = 15$.
  - Use ADMM to solve the SDP
- Obtain optimal solution $X^*$:
The NCAA Football Network (4)

- Solve instance of RDC with $W = A$, $r = 15$.
- Use ADMM to solve the SDP
- Obtain optimal solution $X^*$:
<table>
<thead>
<tr>
<th>Atlantic Coast</th>
<th>Big Twelve 1</th>
<th>MAC West</th>
<th>Pacific Ten</th>
<th>Sun Belt 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clemson</td>
<td>Colorado</td>
<td>Ball State</td>
<td>Arizona</td>
<td>Arkansas State</td>
</tr>
<tr>
<td>Duke</td>
<td>Kansas</td>
<td>C. Michigan</td>
<td>Arizona State</td>
<td>Boise State*</td>
</tr>
<tr>
<td>Florida State</td>
<td>Kansas State</td>
<td>E. Michigan</td>
<td>California</td>
<td>Idaho</td>
</tr>
<tr>
<td>Georgia Tech</td>
<td>Iowa State</td>
<td>N. Illinois</td>
<td>Oregon</td>
<td>NMSU</td>
</tr>
<tr>
<td>Maryland</td>
<td>Missouri</td>
<td>Toledo</td>
<td>Oregon State</td>
<td>North Texas</td>
</tr>
<tr>
<td>NC State</td>
<td>Nebraska</td>
<td>W. Michigan</td>
<td>Stanford</td>
<td>Utah State*</td>
</tr>
<tr>
<td>North Carolina</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big East</td>
<td>Big Twelve 2</td>
<td>MAC East</td>
<td></td>
<td>Sun Belt 2</td>
</tr>
<tr>
<td>Boston College</td>
<td>Oklahoma State</td>
<td>Akron</td>
<td></td>
<td>Central Florida*</td>
</tr>
<tr>
<td>Miami</td>
<td>Oklahoma</td>
<td>BGSU</td>
<td></td>
<td>L-Lafayette</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>Texas Tech</td>
<td>Buffalo</td>
<td></td>
<td>L-Monroe</td>
</tr>
<tr>
<td>Rutgers</td>
<td>Baylor</td>
<td>Kent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syracuse</td>
<td>Texas A&amp;M</td>
<td>Marshall</td>
<td></td>
<td>Louisiana Tech*</td>
</tr>
<tr>
<td>Temple</td>
<td>Texas</td>
<td>Miami Ohio</td>
<td></td>
<td>MTSU</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td></td>
<td>Ohio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Ten</td>
<td>Conference USA</td>
<td>Mountain West</td>
<td></td>
<td>Western Athletic</td>
</tr>
<tr>
<td>Illinois</td>
<td>Alabama-Birm</td>
<td>Air Force</td>
<td></td>
<td>Fresno St</td>
</tr>
<tr>
<td>Indiana</td>
<td>Army</td>
<td>BYU</td>
<td></td>
<td>Hawaii</td>
</tr>
<tr>
<td>Iowa</td>
<td>Cincinnati</td>
<td>Colorado St</td>
<td></td>
<td>Nevada</td>
</tr>
<tr>
<td>Michigan</td>
<td>East Carolina</td>
<td>New Mexico</td>
<td></td>
<td>Rice</td>
</tr>
<tr>
<td>Michigan State</td>
<td>Houston</td>
<td>SDSU</td>
<td></td>
<td>SJSU</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Louisville</td>
<td>UNLV</td>
<td></td>
<td>SMU</td>
</tr>
<tr>
<td>Northwestern</td>
<td>Memphis</td>
<td>Utah</td>
<td></td>
<td>Texas Christian*</td>
</tr>
<tr>
<td>Ohio State</td>
<td>So. Miss.</td>
<td>Wyoming</td>
<td></td>
<td>Texas El Paso</td>
</tr>
<tr>
<td>Penn State</td>
<td>Tulane</td>
<td></td>
<td></td>
<td>Tulsa</td>
</tr>
<tr>
<td>Purdue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outlier</th>
<th>Connecticu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sun Belt 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun Belt 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Western Athletic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Athletic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independents</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
</tr>
</tbody>
</table>

The NCAA Football Network (5)
Conclusions

- Have presented new convex relaxation based heuristics for the maximum clique and graph clustering problems.

- These relaxations are based on the relationship between cliques and low-rank submatrices of the adjacency graph, and the nuclear norm relaxation for rank minimization.

- If data is sufficiently clusterable (i.e. highly similar within clusters, dissimilar across clusters) then these heuristics successfully recover the correct partition of the data.
Open problems

• Is the lower bound $k = \Omega(\sqrt{N})$ tight?

• Overlapping communities?

• Sparse graphs? Can we improve recovery guarantees?
  • Sparse in the sense that $\alpha > \beta$ are both tending to 0 as $N \to \infty$

• How to choose $r$ in the clustering SDP?

• Algorithmic issues: how to efficiently solve large-scale SDP?
Thank you!


- Papers/slides/code: [users.cms.caltech.edu/~bpames](http://users.cms.caltech.edu/~bpames)