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 vison, 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

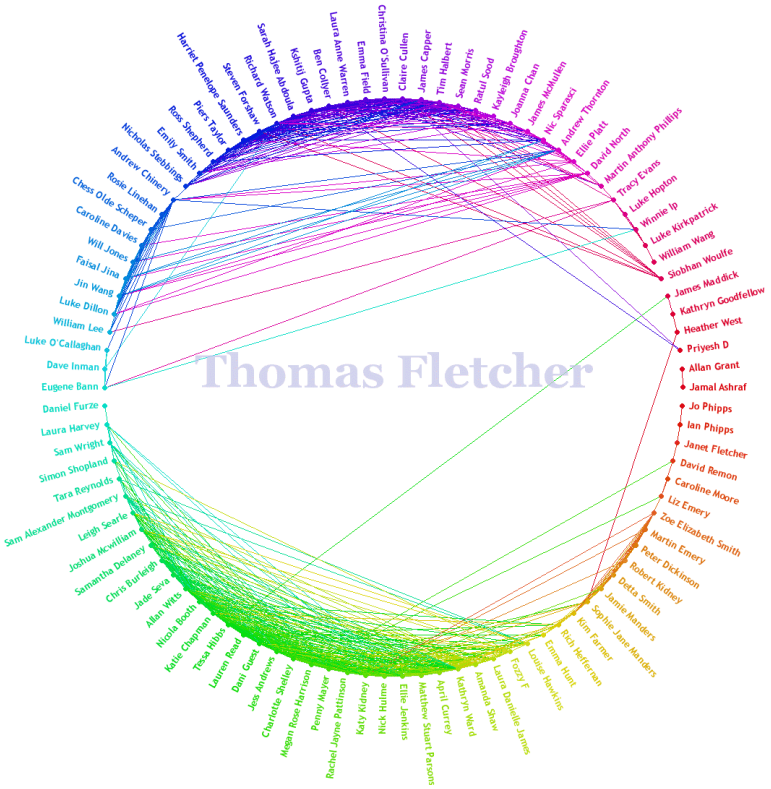
- Clustering seems to be a very difficult/ill-posed problem.
- Many heuristics seem to work well in practice.
- Question: can we show that we can cluster “clusterable” data? How do we model clusterable data?

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

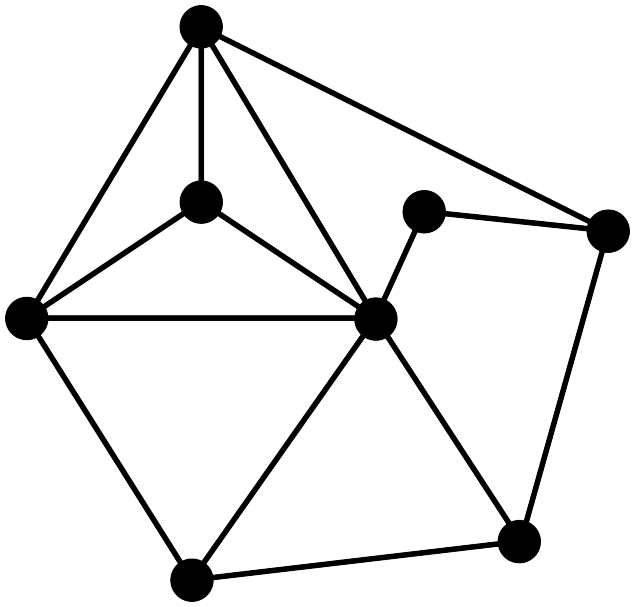
Example: Communities in Social Networks

- Nodes = users
- Edges = “friendship”.
- Densely connected groups = communities



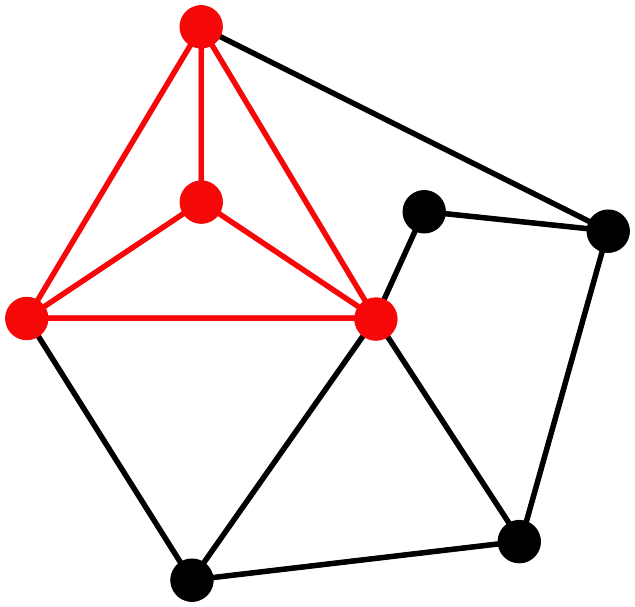
Cliques of a graph

- 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

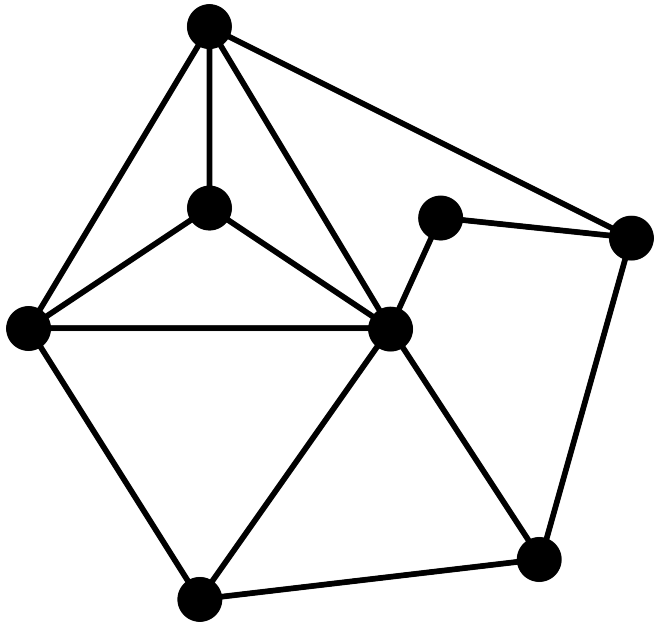
- Complexity results are worst case.
- There are some graphs for which finding the maximum clique is as difficult as hardest problems known.
- Not necessarily true for all graphs.
- We focus on identifying trade-offs between clique size and density of nonclique edges that ensure we can solve Clique efficiently.

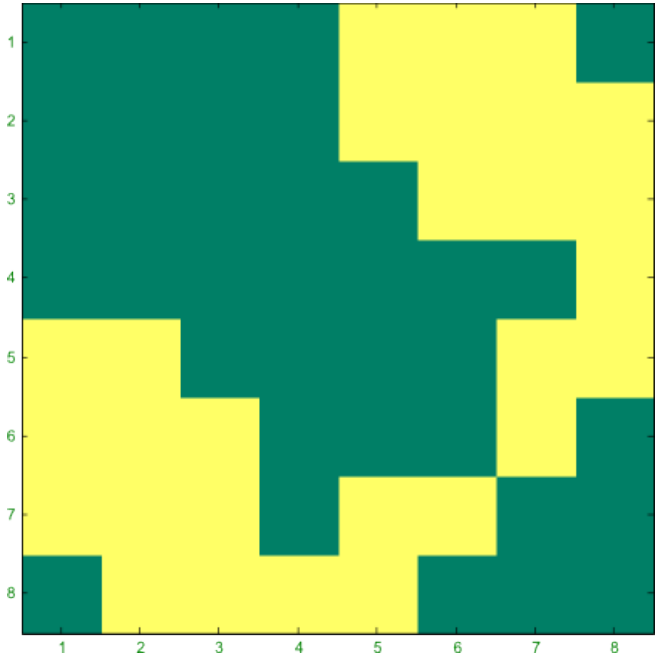
Matrix representation of cliques

- Characteristic vector of C: vector $\mathbf{v} \in \{0,1\}^V$ 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 $\mathbf{v}$, let $X = \mathbf{v}\mathbf{v}^T$.
- Nonzero entries of X form a $|C| \times |C|$ all-ones block in $A_G + I$.



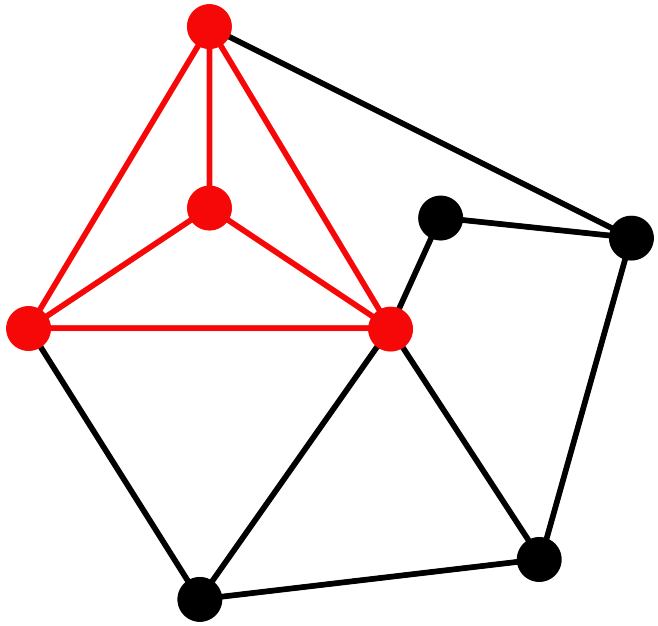


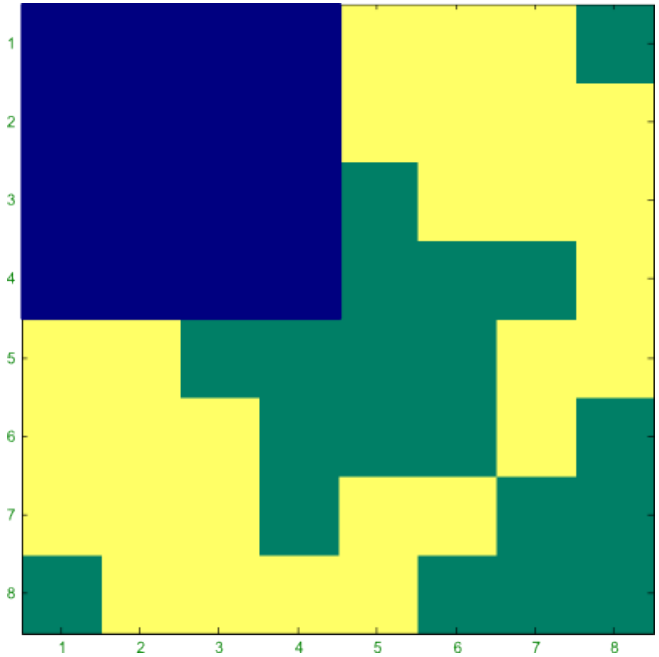
Matrix representation of cliques

- Characteristic vector of C: vector $\mathbf{v} \in \{0,1\}^V$ with

$$v_i = 1 \text{ if } i \in C \quad \text{and} \quad v_i = 0 \text{ otherwise}$$

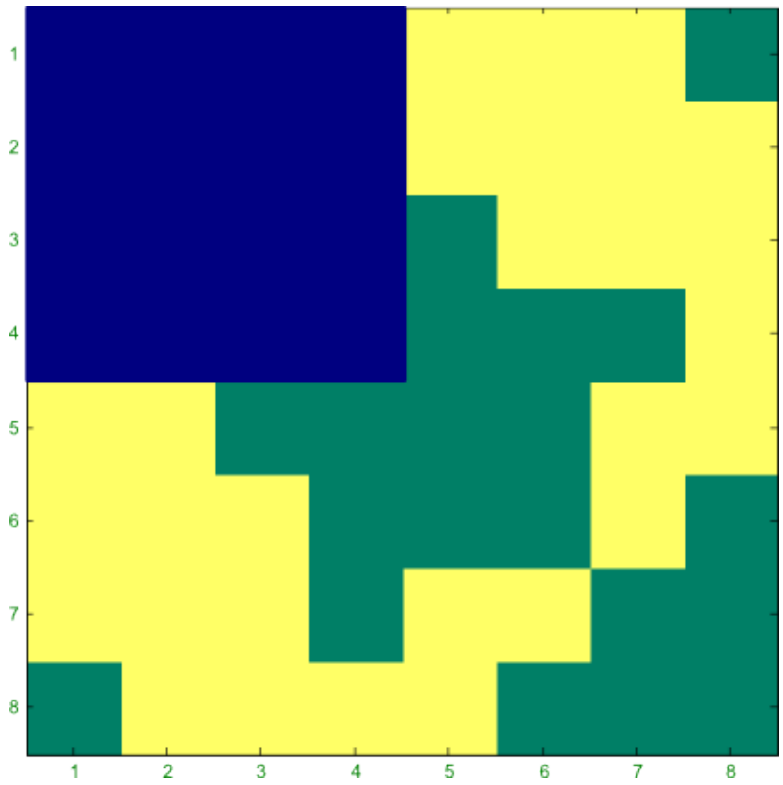
- If C is a clique with characteristic vector $\mathbf{v}$, let $X = \mathbf{v}\mathbf{v}^T$.
- Nonzero entries of X form a $|C| \times |C|$ all-ones block in $A_G + I$.





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
- $\mathrm{rank}(X) = 1$

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:

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

where $\tilde{E} = V \times V - \left\{ E \cup \{(u,u) : u \in V\} \right\}$.

Affine Rank minimization

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

$$\min\{\mathrm{rank}(X) : \mathcal{A}(X) = \mathbf{b}\}.$$

- Well-known to be NP-hard.
- Fazel 2002: Relax using the nuclear norm $\|X\|_*$

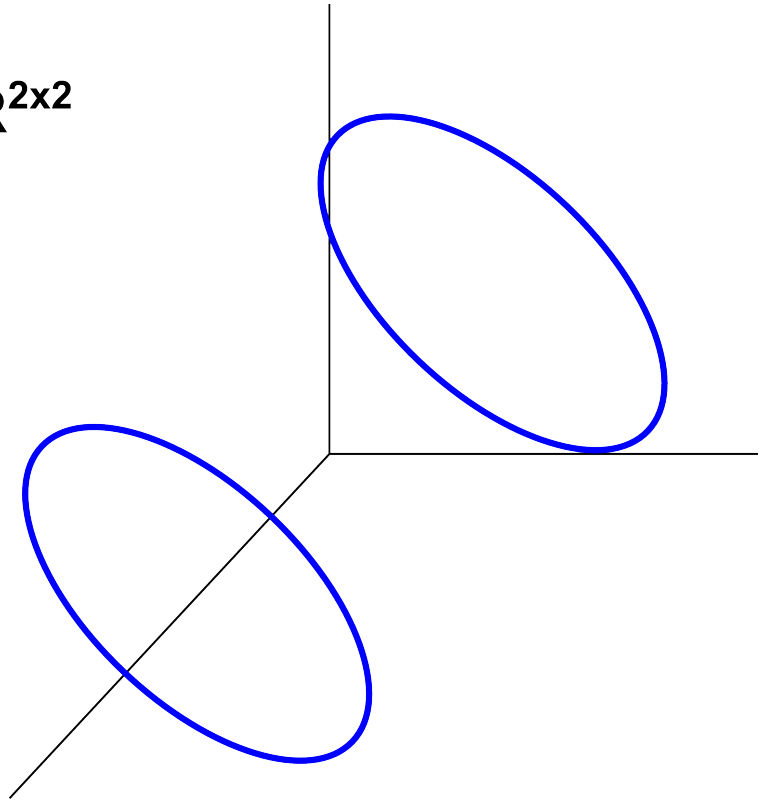
$$\mathrm{rank}(X) = \|\sigma(X)\|_0 \;\Rightarrow\; \|X\|_* = \|\sigma(X)\|_1 = \sum_i \sigma_i(X)$$

Geometry of nuclear norm relaxation

- Relax by replacing $\mathrm{rank}(X)$ with $\|X\|_* = \sum \sigma_i(X)$
- $\|\cdot\|_*$ is the convex envelope of rank on $B = \{X : \|X\| \leq 1\}$.

$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$ in $\mathbf{R}^{2\mathrm{x}2}$

rank(X) = 1
|x + z| =1

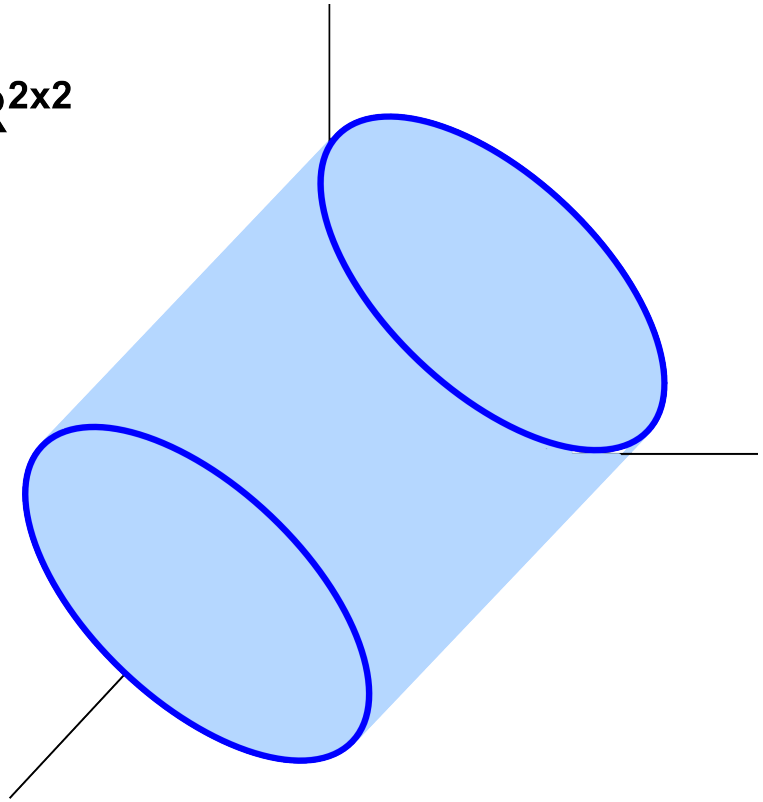


Geometry of nuclear norm relaxation

- Relax by replacing $\mathrm{rank}(X)$ with $\|X\|_* = \sum \sigma_i(X)$
- $\|\cdot\|_*$ is the convex envelope of rank on $B = \{X : \|X\| \leq 1\}$.

$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$ in $\mathbf{R}^{2\mathrm{x}2}$

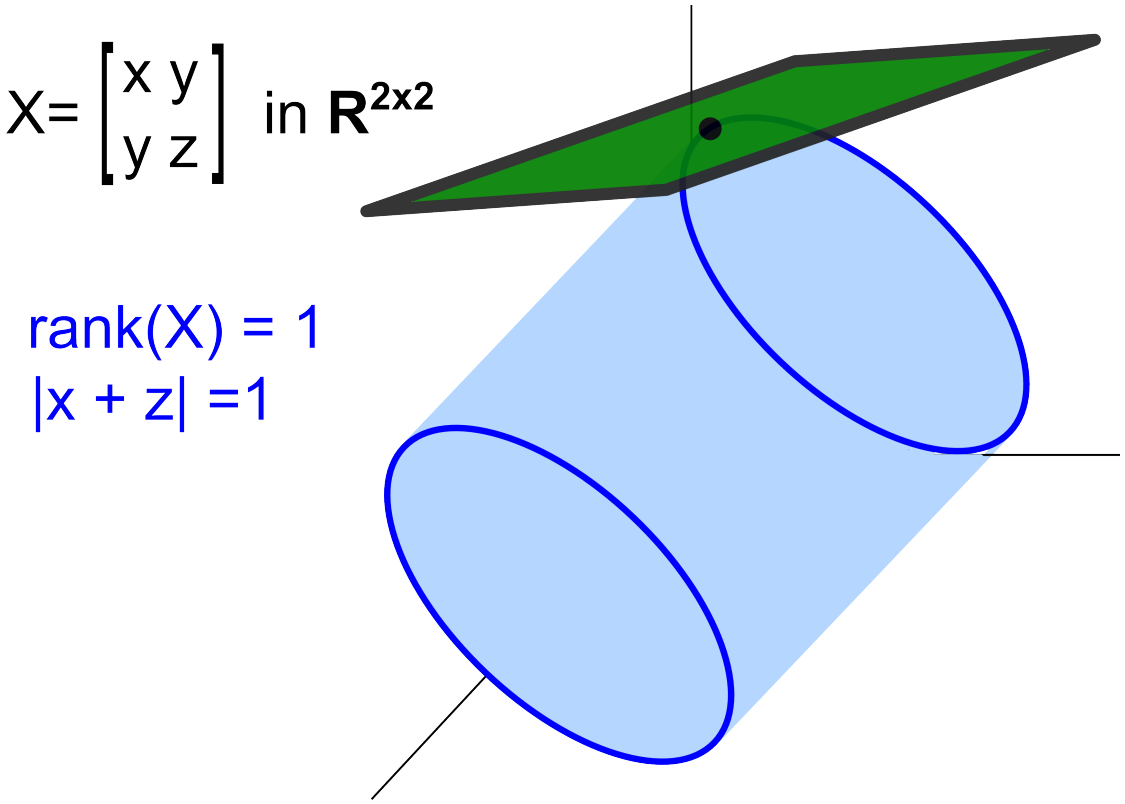
rank(X) = 1
|x + z| =1



Guaranteed low-rank recovery

- Recht, Fazel, Parrilo 2007, Candés and Recht 2008:
  If $\mathcal{A}$ and **b** are given by sufficiently many ($m = \Omega(rn \log n)$ random observations of a rank-r $n \times n$ matrix then

$$\arg\min\{\mathrm{rank}(X) : \mathcal{A}(X) = \mathbf{b}\} = \arg\min\{\|X\|_* : \mathcal{A}(X) = \mathbf{b}\}$$



Nuclear norm relaxation of Clique

- We have the rank minimization problem:

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

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

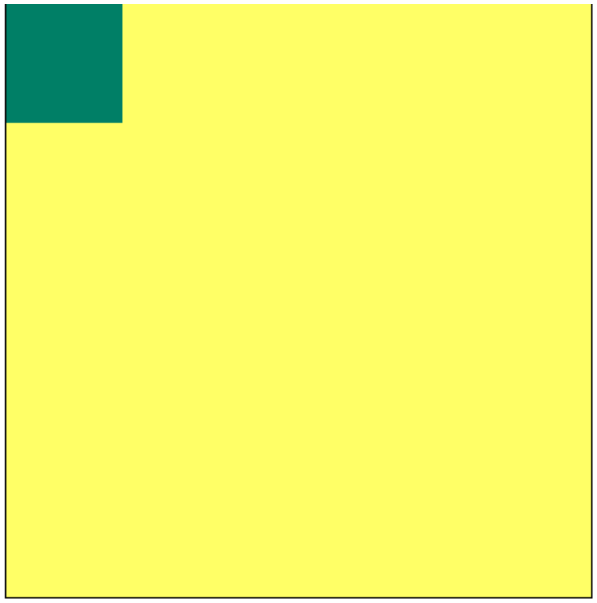
$$\min \left\{ \|X\|_* : \mathbf{e}^T X \mathbf{e} = k^2,\ X_{ij} = 0 \text{ if } (i,j) \in \tilde{E} \right\} \quad (\mathbf{NNR})$$

- When does the solution of the relaxation coincide with that of the rank minimization problem?

The planted case

Construction:

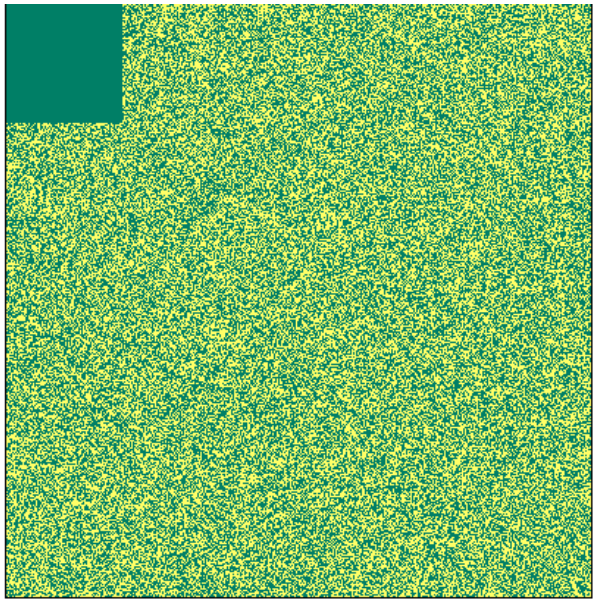
- Add edges between nodes in vertex set V^* of size k.



The planted case

Construction:

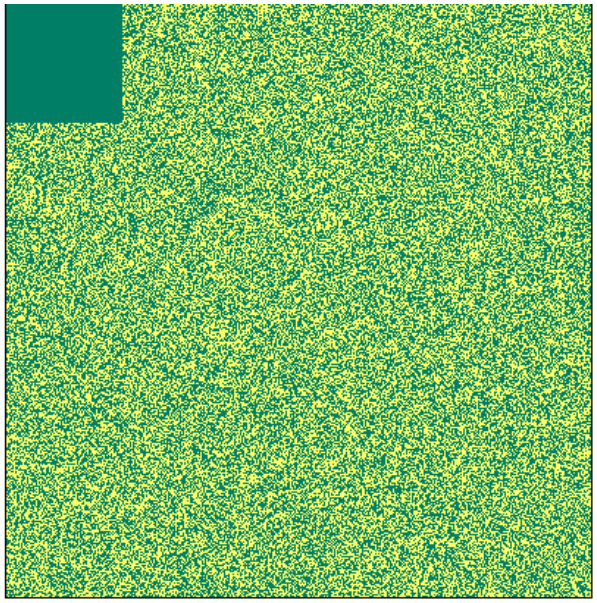
- Add edges between nodes in vertex set V^* of size k.
- Each remaining edge is added to E independently with fixed probability p.



The planted case

Construction:

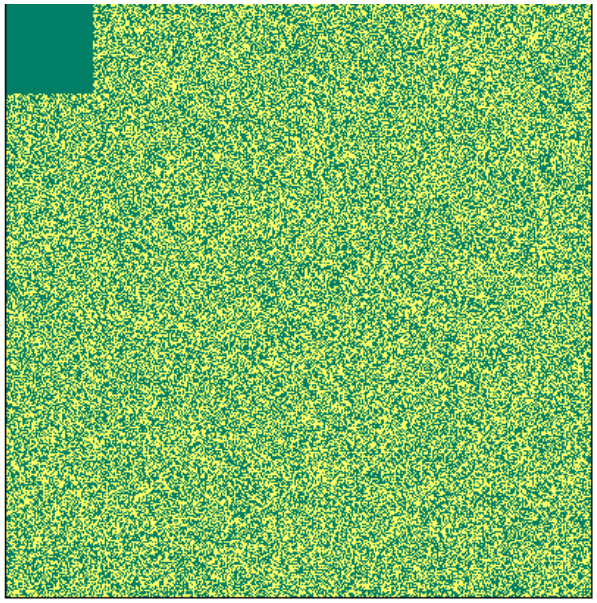
- Add edges between nodes in vertex set V^* of size k.
- Each remaining edge is added to E independently with fixed probability p.
- V^* is a clique of G (called a planted or hidden clique).



The planted case

Construction:

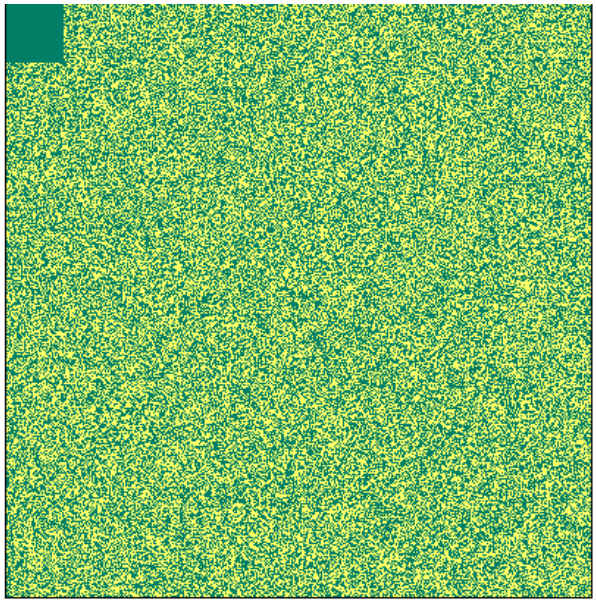
- Add edges between nodes in vertex set V^* of size k.
- Each remaining edge is added to E independently with fixed probability p.
- V^* is a clique of G (called a planted or hidden clique).



The planted case

Construction:

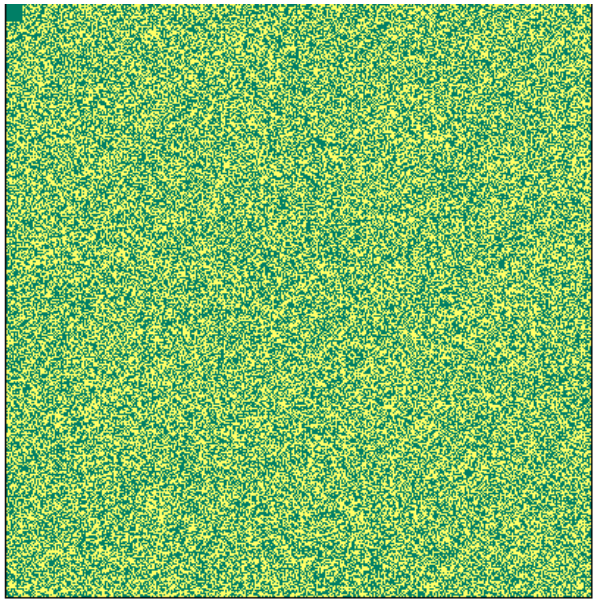
- Add edges between nodes in vertex set V^* of size k.
- Each remaining edge is added to E independently with fixed probability p.
- V^* is a clique of G (called a planted or hidden clique).



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*

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

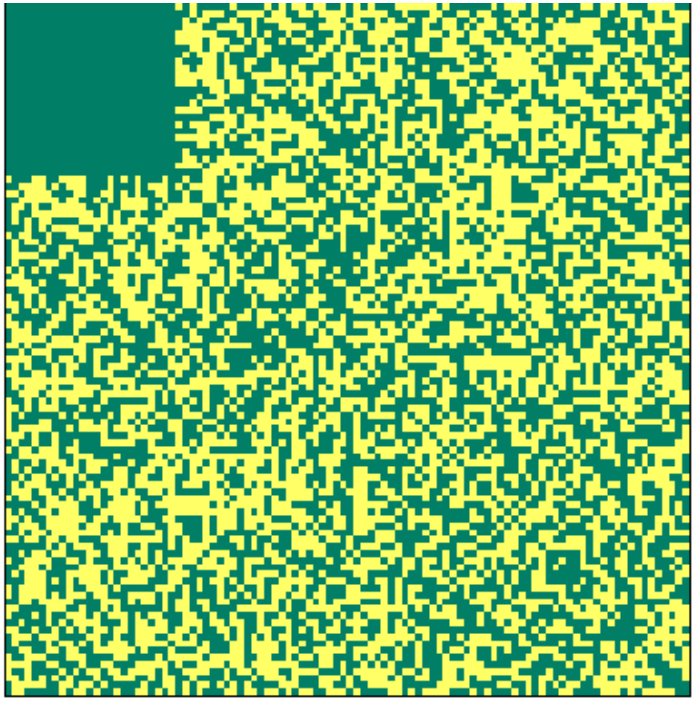
then

- V^* *is the unique maximum clique of* G*, and*
- $X^* = \mathbf{v}\mathbf{v}^T$ *is the unique optimal solution of* $(\mathbf{NNR})$

with high probability.

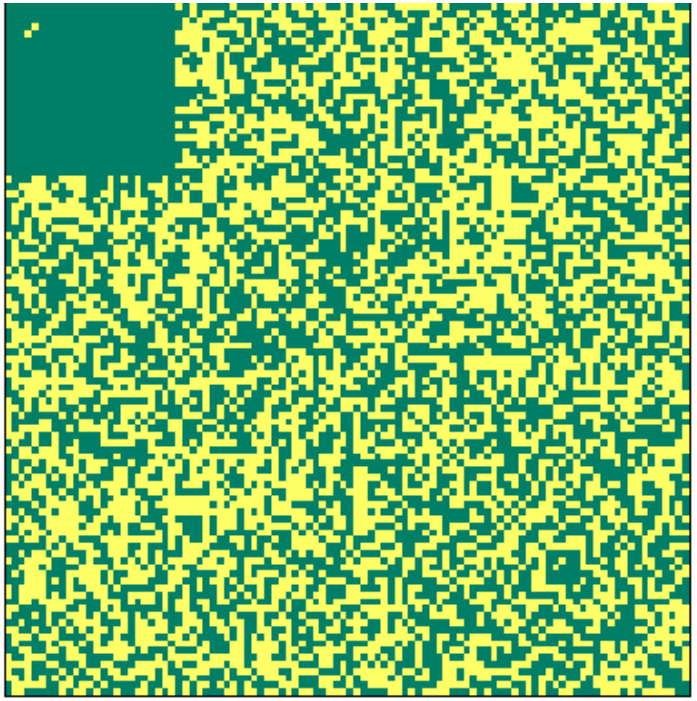
What happens if edges are deleted?

- 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 (**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 (**NNR**).



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 $d(H)$ = 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 $\mathbf{v}$.
- Introduce a correction Y for entries of $X = \mathbf{v}\mathbf{v}^T$ 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 $G(V^*)$:

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

Relaxation as Principal Component Pursuit

- Can relax (**DKS**) as

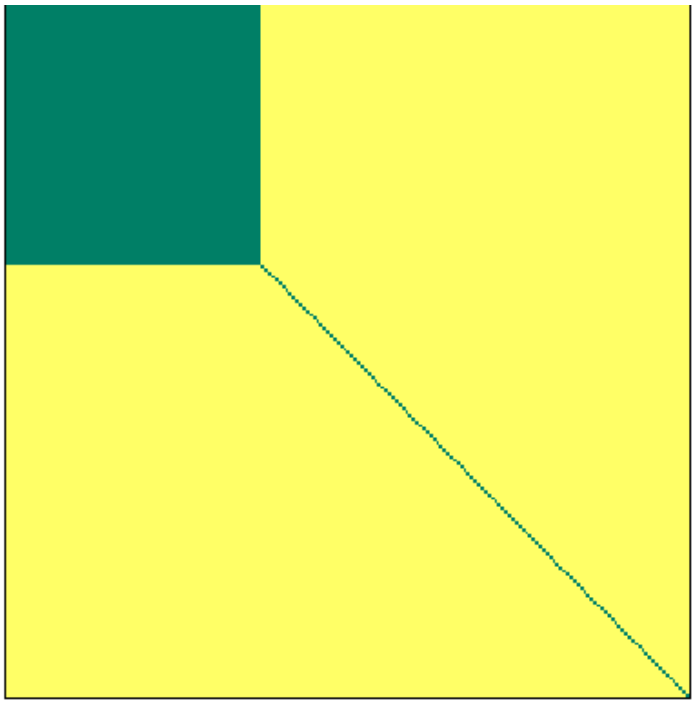
$$\begin{aligned} \min \quad & \|X\|_* + \gamma \|Y\|_1 \\ \text{st} \quad & \mathbf{e}^T X \mathbf{e} = k^2 \\ & X_{ij} + Y_{ij} = 0 \text{ if } ij \in \tilde{E} \\ & X \in [0,1]^{V \times V} \end{aligned} \qquad (\textbf{DKSR})$$

  where γ is a regularization parameter.

- Relax $\|Y\|_0$ using the ℓ_1-norm $\|Y\|_1$, $\mathrm{rank}(X)$ with the nuclear norm $\|X\|_*$.

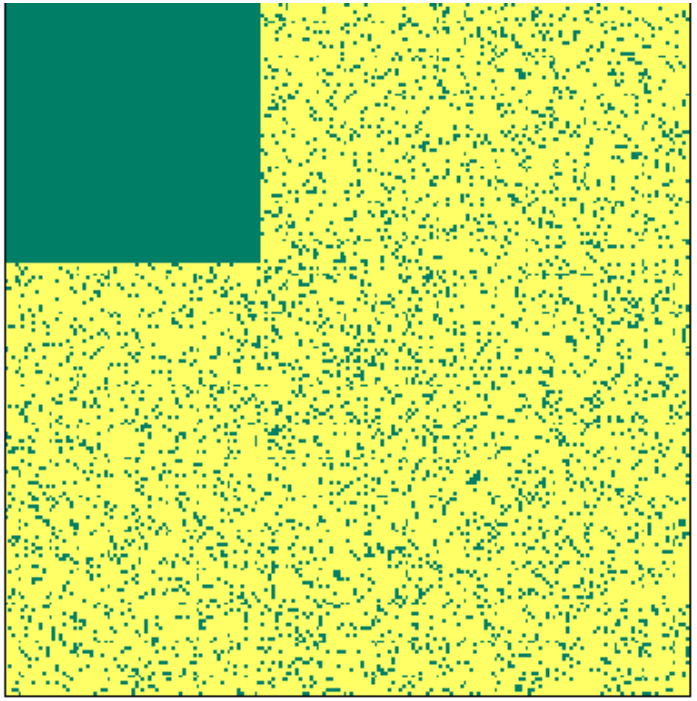
Planted case

- 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



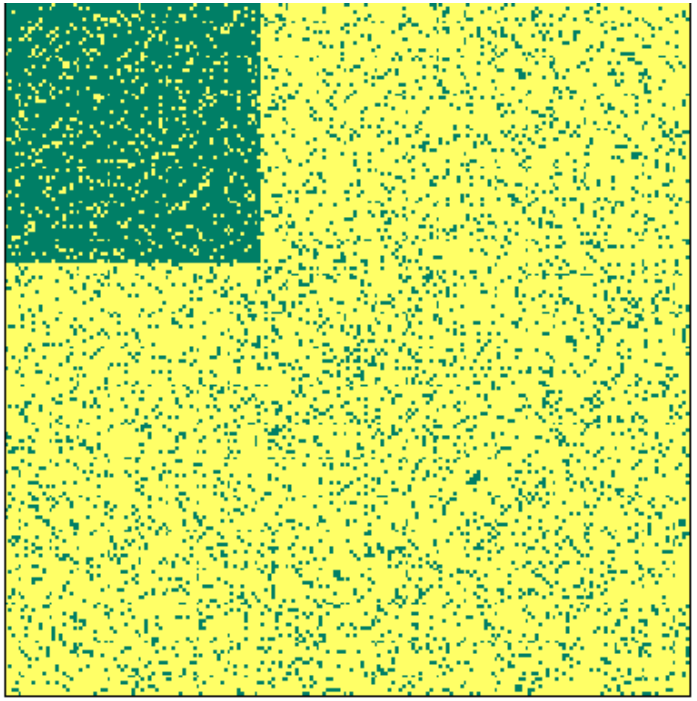
Planted case

- 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



Recovery guarantee

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*
- $X^* = \mathbf{v}\mathbf{v}^T$ *is the unique optimal solution of* (**DKSR**)

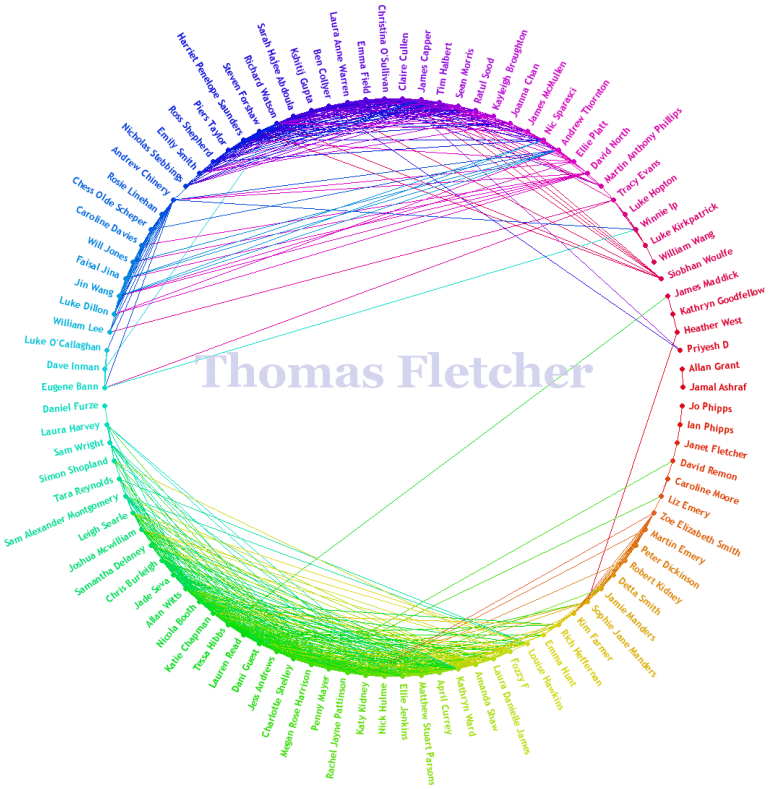
with high probability.

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.

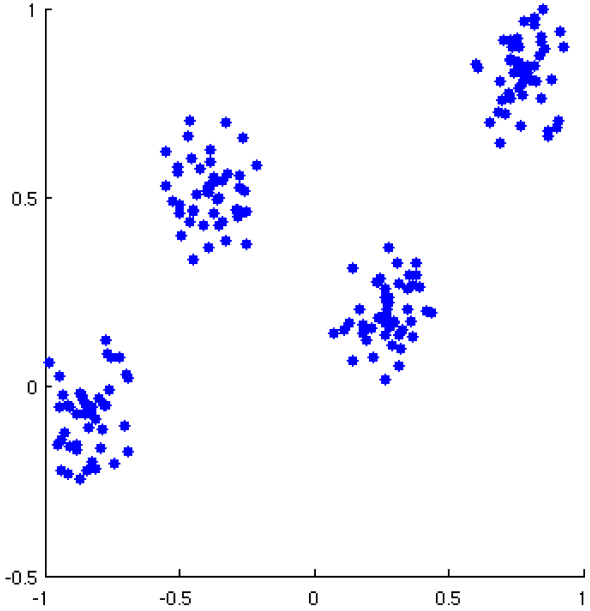
Example: Communities in Social Networks

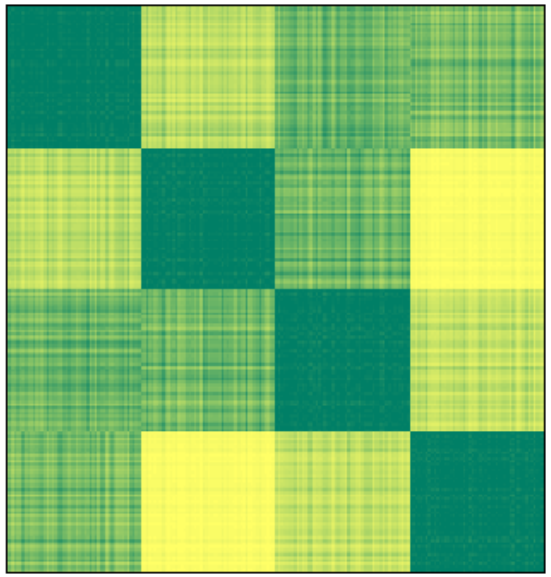
- Nodes = users
- Edges = “friendship”.
- Densely connected groups = communities



Example: Clustered Euclidean data

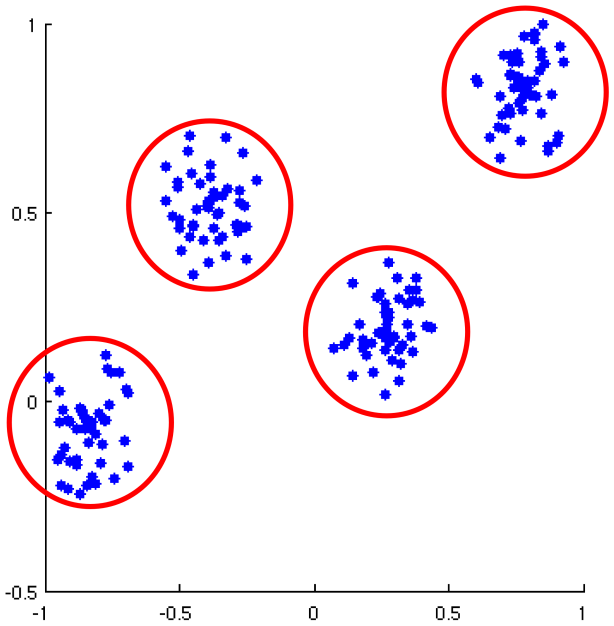
- Suppose each data point in the ith cluster C_i is placed uniformly at random in a ball centered at $c_i \in \mathbf{R}^d$.
- Distance within clusters will be small compared to the distance between clusters if centers are well-separated.
- Choose $W_{ij} = \exp(-\|\mathbf{x}^i - \mathbf{x}^j\|^2)$.

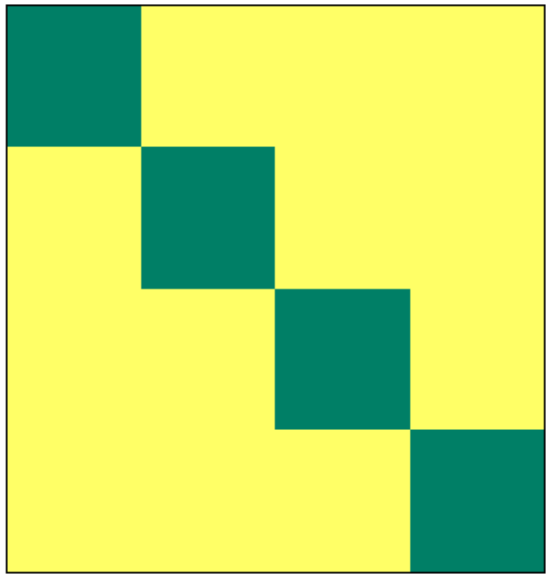




Example: Clustered Euclidean data

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





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:

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

where $\mathbf{v}$ 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 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$
- Lift the r characteristic vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ to the rank-r matrix variable X:

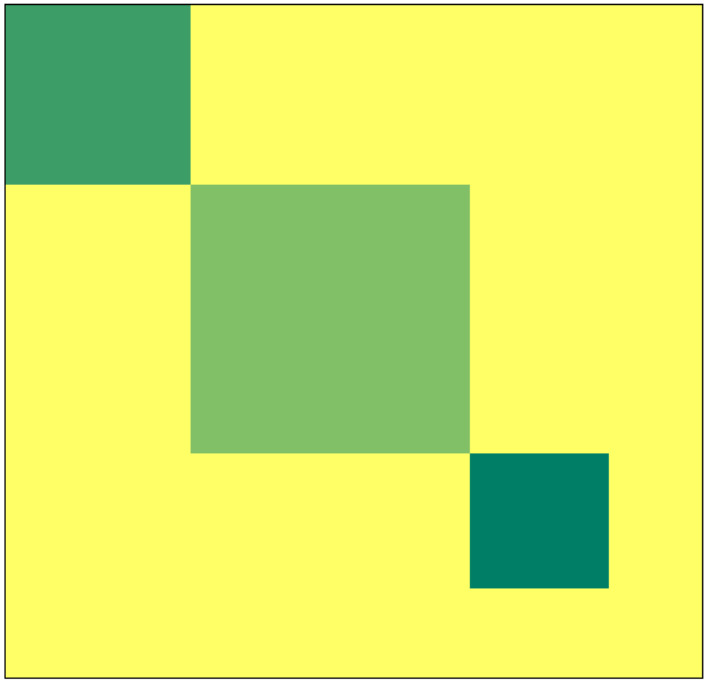
$$X = \sum \frac{\mathbf{v}_i \mathbf{v}_i^T}{\|\mathbf{v}_i\|^2}.$$

- X has trace exactly equal to r and satisfies

$$\mathrm{Tr}(WX) = \sum \frac{\mathbf{v}_i^T W \mathbf{v}_i}{\|\mathbf{v}_i\|^2}$$

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:

$$X_{ij} \geq 0, \quad X \succeq 0$$

- $\mathrm{rank}(X) = \mathrm{Tr}(X) = r$
- other combinatorial constraints

Relaxation to SDP

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

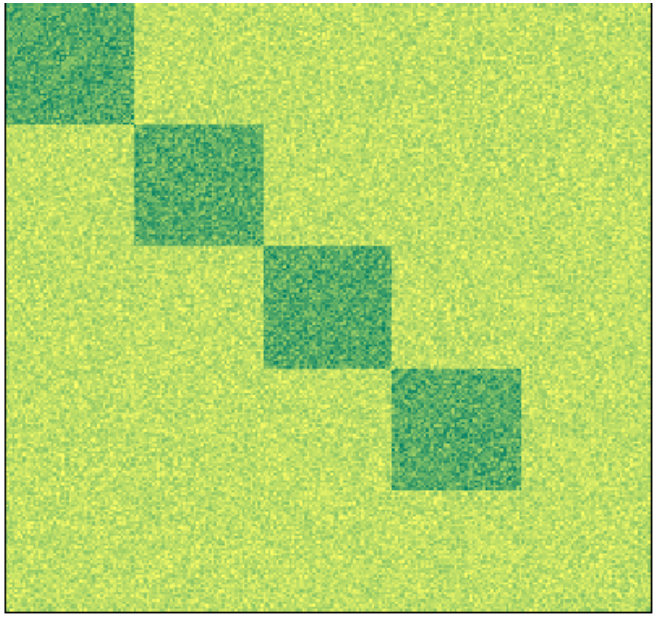
$$\begin{aligned} \max \quad & \mathrm{Tr}(WX) \\ \text{st} \quad & X\mathbf{e} \leq \mathbf{e} \\ & \mathrm{Tr}(X) = r \\ & X_{ij} \geq 0 \quad \forall\, i, j \in V \\ & X \succeq 0. \end{aligned}$$

- $\mathrm{Tr}(X) = \|X\|_* = r$ acts a surrogate for $\mathrm{rank}(X) = r$

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 Ω_1 with mean α.
- Sample remaining entries of W i.i.d. from distribution Ω_2 with mean $\beta << \alpha$.



Guaranteed recovery

- Let $X^* = \sum_{i=1}^{r} \frac{\mathbf{v}_i \mathbf{v}_i^T}{k_i}$ 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

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

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:

	St	**Br**	**Gi**	**So**	**Oc**	**Ke**	**Re**	**Sc**	**Th**
St	1	0.62	0.66	0.63	0.33	0.36	0.25	0.14	0.15
Br	0.62	1	0.72	0.71	0.55	0.47	0.43	0.25	0.24
Gi	0.66	0.72	1	0.78	0.47	0.49	0.43	0.28	0.26
So	0.63	0.71	0.78	1	0.55	0.5	0.44	0.31	0.29
Oc	0.33	0.55	0.47	0.55	1	0.67	0.71	0.54	0.54
Ke	0.36	0.47	0.49	0.5	0.67	1	0.77	0.58	0.59
Re	0.25	0.43	0.43	0.44	0.71	0.77	1	0.66	0.68
Sc	0.14	0.25	0.28	0.31	0.54	0.58	0.66	1	0.79
Th	0.15	0.24	0.26	0.29	0.54	0.59	0.68	0.79	1

Rehnquist Supreme Court (2)

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

1: "Liberal"	2: "Conservative"
Stevens (St)	O'Connor (Oc)
Breyer (Br)	Kennedy (Ke)
Ginsberg (Gi)	Rehnquist (Re)
Souter (So)	Scalia (Sc)
	Thomas (Th)

Rehnquist Supreme Court (3)

	St	Br	Gi	So	Oc	Ke	Re	Sc	Th
St	**1**	**0.62**	**0.66**	**0.63**	0.33	0.36	0.25	0.14	0.15
Br	**0.62**	**1**	**0.72**	**0.71**	0.55	0.47	0.43	0.25	0.24
Gi	**0.66**	**0.72**	**1**	**0.78**	0.47	0.49	0.43	0.28	0.26
So	**0.63**	**0.71**	**0.78**	**1**	0.55	0.5	0.44	0.31	0.29
Oc	0.33	0.55	0.47	0.55	**1**	**0.67**	**0.71**	**0.54**	**0.54**
Ke	0.36	0.47	0.49	0.5	**0.67**	**1**	**0.77**	**0.58**	**0.59**
Re	0.25	0.43	0.43	0.44	**0.71**	**0.77**	**1**	**0.66**	**0.68**
Sc	0.14	0.25	0.28	0.31	**0.54**	**0.58**	**0.66**	**1**	**0.79**
Th	0.15	0.24	0.26	0.29	**0.54**	**0.59**	**0.68**	**0.79**	**1**

Rehnquist Supreme Court (4)

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

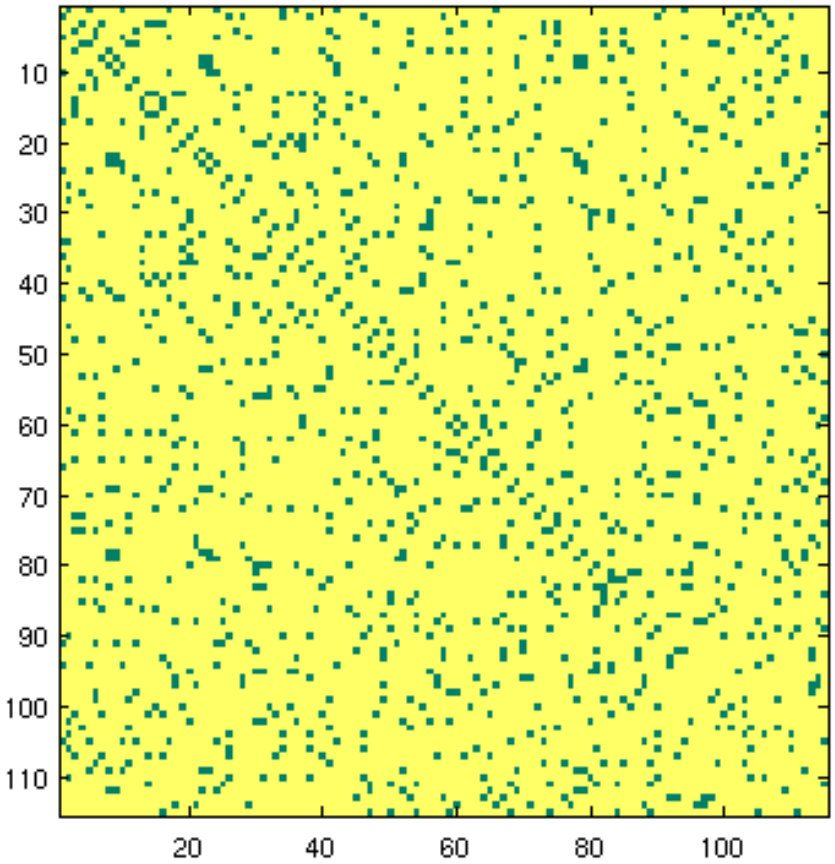
Cluster 1	Cluster 2	Cluster 3
Thomas (Th)	O'Connor (Oc)	Stevens (St)
Scalia (Sc)	Kennedy (Ke)	Breyer (Br)
	Rehnquist (Re)	Ginsberg (Gi)
		Souter (So)

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.

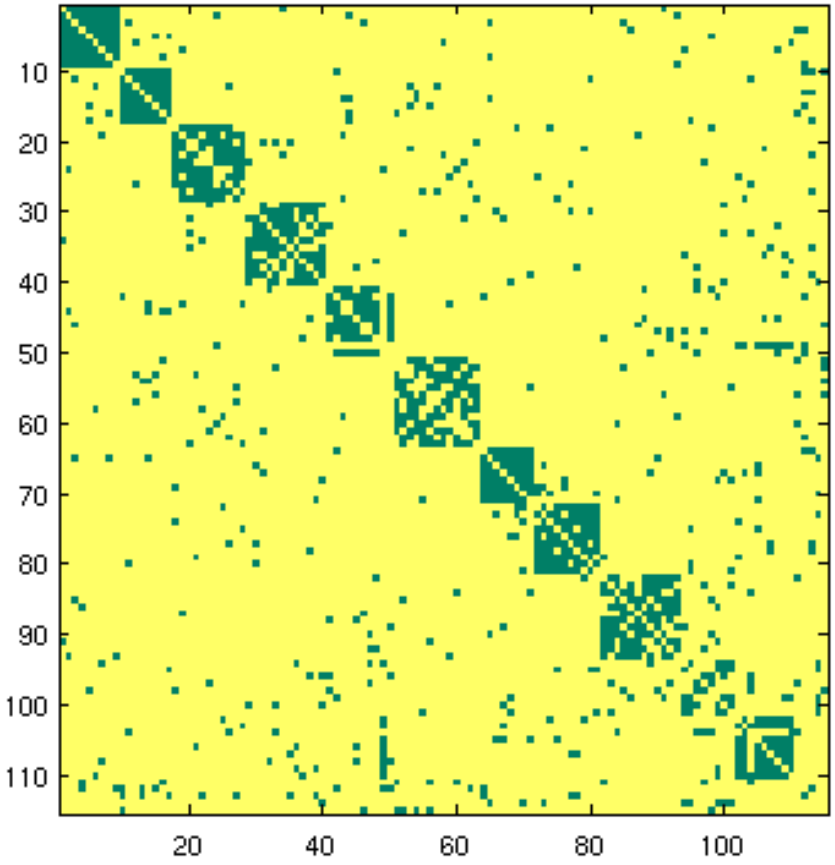
The NCAA Football Network (2)

- The node-node adjacency matrix for the football network.



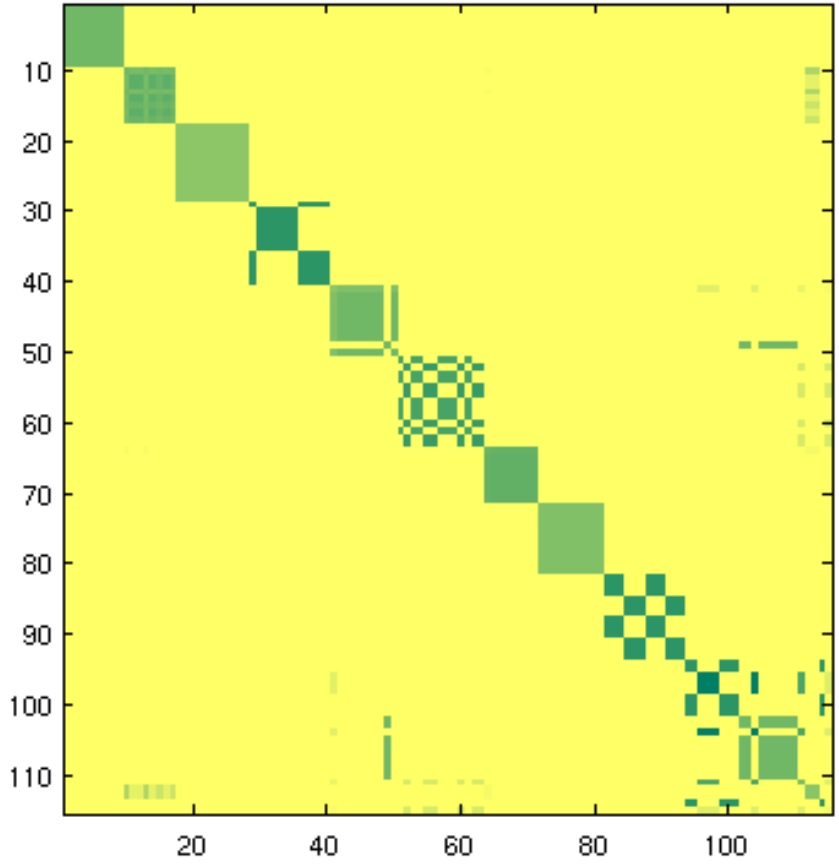
The NCAA Football Network (3)

- The adjacency matrix with rows/cols reordered according to conferences.



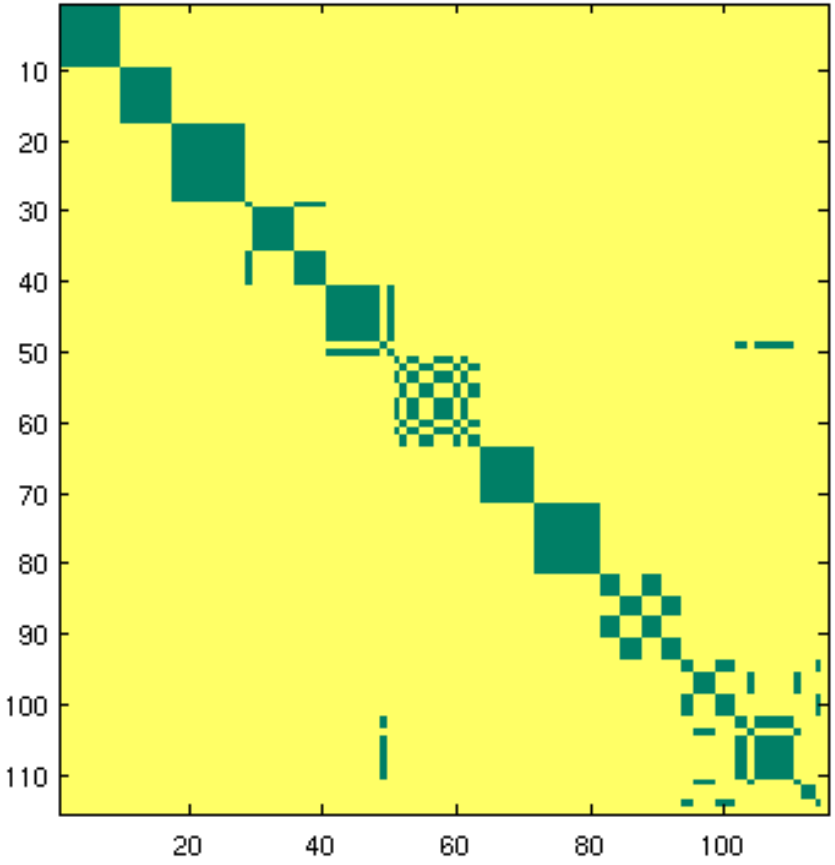
The NCAA Football Network (4)

- 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^*:



The NCAA Football Network (5)

Atlantic Coast
Clemson
Duke
Florida State
Georgia Tech
Maryland
NC State
North Carolina
Virginia
Wake Forest

Big East
Boston College
Miami
Pittsburgh
Rutgers
Syracuse
Temple
Virginia Tech
West Virginia

Big Ten
Illinois
Indiana
Iowa
Michigan
Michigan State
Minnesota
Northwestern
Ohio State
Penn State
Purdue
Wisconsin

Big Twelve 1
Colorado
Kansas
Kansas State
Iowa State
Missouri
Nebraska

Big Twelve 2
Oklahoma State
Oklahoma
Texas Tech
Baylor
Texas A&M
Texas

Conference USA
Alabama-Birm
Army
Cincinnati
East Carolina
Houston
Louisville
Memphis
So. Miss.
Tulane

Outlier
Connecticut

MAC West
Ball State
C. Michigan
E. Michigan
N. Illinois
Toledo
W. Michigan

MAC East
Akron
BGSU
Buffalo
Kent
Marshall
Miami Ohio
Ohio

Mountain West
Air Force
BYU
Colorado St
New Mexico
SDSU
UNLV
Utah
Wyoming

Pacific Ten
Arizona
Arizona State
California
Oregon
Oregon State
Stanford
UCLA
USC
Washington
WSU

SEC West
Alabama
Arkansas
Auburn
LSU
Miss St
Mississippi

SEC East
Florida
Georgia
Kentucky
South Carolina
Tennessee
Vanderbilt

Sun Belt 1
Arkansas State
*Boise State**
Idaho
NMSU
North Texas
*Utah State**

Sun Belt 2
*Central Florida**
L-Lafayette
L-Monroe
*Louisiana Tech**
MTSU

Western Athletic
Fresno St
Hawaii
Nevada
Rice
SJSU
SMU
*Texas Christian**
Texas El Paso
Tulsa

Independents
Navy
Notre Dame

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!

- B. Ames and S. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical Programming*, 129(1):121, 2011.
- B. Ames. Guaranteed clustering and biclustering via semidefinite programming. To appear in *Mathematical Programming*.
- B. Ames. Robust convex relaxation for the planted clique and densest k-subgraph problem. 2013. arxiv.org/abs/1305.4891

- Papers/slides/code: users.cms.caltech.edu/∼bpames