Nuclear Norm Minimization for the Planted Clique and Biclique Problems

Brendan P.W. Ames   Stephen A. Vavasis

Department of Combinatorics and Optimization
University of Waterloo

Continuous Optimization Seminar
Thursday, February 12, 2009

Outline

Outline

1. Nuclear Norm Minimization
2. Maximum Clique
   - The Adversarial Case
   - The Random Case

Outline

Affine Rank Minimization

- **Affine rank minimization** refers to optimization problems of the form

$$\begin{array}{ll} \min & \operatorname{rank}(X) \\ \text{s.t.} & \mathcal{A}(X) = b \end{array}$$

  for linear $\mathcal{A} : \mathbf{R}^{m \times n} \rightarrow \mathbf{R}^p$, $b \in \mathbf{R}^p$.
- Known to be **NP-hard**.

Nuclear Norm Relaxation

- Solution suggested by **Recht-Fazel-Parrillo (2007)**.
- Replace $\text{rank}\,(X)$ with $\|X\|_* = \sigma_1(X) + \sigma_2(X) + \cdots + \sigma_{\min m,n}(X)$.
- **R-F-P** and **Candès-Recht (2008)** show that the relaxation

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & \mathcal{A}(X) = b \end{array}$$

returns the optimal solution to the original problem provided that $\mathcal{A}$ samples "enough" of the entries of X.

Compressive Sensing

- Rank minimization can be thought of as a generalization of **vector cardinality minimization**:

$$\begin{array}{ll} \min & \#\text{ nonzero entries of } x \\ \text{s.t.} & Ax = b \end{array}$$

  for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$.
- Problem is NP-hard.
- Can be relaxed using the l_1-norm: $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$.
- For certain classes of random A, then the l_1-relaxation recovers exact solution to the original problem.

Properties of the Nuclear Norm

- $\|\cdot\|_*$ is convex. In fact, $\|\cdot\|_*$ is the **convex envelope** or **convex hull** of $\text{rank}(\cdot)$ on $\{X \in \mathbf{R}^{m\times n} : \|X\| \leq 1\}$.
- $\|\cdot\|_*$ can be calculated by solving the following SDP:

$$\|X\|_* = \begin{array}{ll} \max_{Y} & \langle X, Y\rangle \\ \text{s.t.} & \begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0 \end{array} \qquad (P)$$

$$= \begin{array}{ll} \min_{W_1,W_2} & \frac{1}{2}(\text{tr}(W_1) + \text{tr}(W_2)) \\ \text{s.t.} & \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0. \end{array} \qquad (D)$$

Subdifferential of $\|\cdot\|_*$

Lemma (Watson (92))

- *Let X have rank r and SVD $X = \sum_{k=1}^{r} \sigma_k u_k v_k^T$.*
- *Then ϕ is a subgradient of $\|\cdot\|_*$ at X if and only if*

$$\phi = \sum_{k=1}^{r} u_k v_k^T + W$$

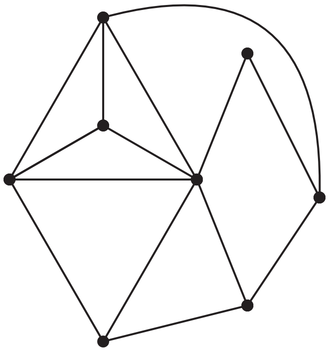
where

$$\|W\| \leq 1, \quad W^T u_k = 0 \; \forall \, k = 1, \ldots, r, \quad W v_k = 0 \;\; \forall \, k = 1, \ldots, r.$$

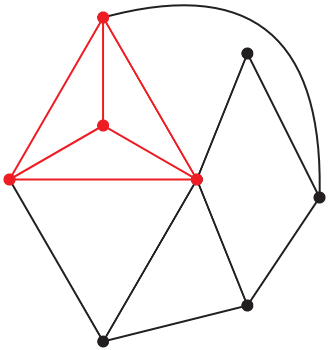
Maximum clique problem

- Given graph $G = (V, E)$, want to find the largest complete subgraph (clique) of G.
- One of the original NP-hard problems.
- Can be used as a simple model for data mining problems.

Example



Example



Formulation as affine rank minimization

- The adjacency matrix of the graph $H' = H \cup \{vv : v \in V(H)\}$ has rank equal to one if and only if H is a complete graph.
- Thus, a clique K of G containing n vertices can be found by solving

$$\begin{aligned} \min \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & \sum_{i \in V} \sum_{j \in V} X_{ij} \geq n^2, \\ & X_{ij} = 0 \text{ if } (i,j) \notin E \text{ and } i \neq j, \\ & X \in [0,1]^{V \times V}. \end{aligned}$$

- Still NP-hard.

Nuclear norm relaxation of max clique

- We consider the convex relaxation

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \sum_{i \in V} \sum_{j \in V} X_{ij} \geq n^2, \\ & X_{ij} = 0 \text{ if } (i,j) \notin E \text{ and } i \neq j. \end{aligned}$$

- **Our result:** If G consists of a clique on n vertices plus some diversionary edges, then this convex relaxation recovers the maximum clique of G for sufficiently large n.

Optimality Conditions

Theorem

- *Let* $V^* \subset V$ *such that* $|V^*| = n$. *Let* $\bar{v}$ *be the characteristic vector of* V^* *Let* $X^* = \bar{v}\bar{v}^T$.
- *If there exists* $W \in \mathbf{R}^{V\times V}$ *and* $\lambda \in \mathbf{R}^{V\times V}$, $\mu \in \mathbf{R}_+$ *such that* $W\bar{v} = 0, \quad \bar{v}^T W = 0, \quad \|W\| \leq 1$ *and*

$$\frac{\bar{v}\bar{v}^T}{n} + W = \mu ee^T + \sum_{(i,j)\in\bar{E}} \lambda_{ij} e_i e_j^T .$$

then

  - X^* *is the optimal solution to the nuclear norm relaxation:*

$$\min\{\|X\|_* : \sum X_{ij} \geq n^2, \;\; X_{ij} = 0 \; \forall\, (i,j) \in \bar{E}\}.$$

  - *For every clique* K *of* G, $|V(K)| \leq n$.
  - *If* $\|W\| < 1$, $\mu > 0$, *the maximum clique spanned by* V^* *is unique.*

Proof

- Optimality of X^* follows immediately from the KKT conditions:

$$\partial \|X^*\|_* = \left\{ \frac{\bar{v}\bar{v}^T}{n} + W : Wv = 0, W^T v = 0, \|W\| \leq 1 \right\}.$$

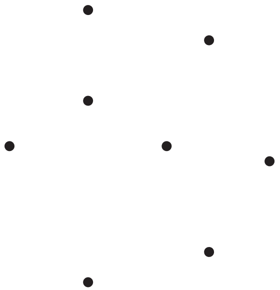
- Suppose V' spans a clique. Let $\bar{v}'$ be the characteristic vector of V'.
- Then $X' = \bar{v}'\bar{v}'^T (n^2/|V'|^2)$ is feasible for the relaxation.
- So

$$\|X'\|_* = \frac{n^2}{|V'|} \geq \|X^*\|_* = n.$$

- Therefore, $|V'| \leq n$.

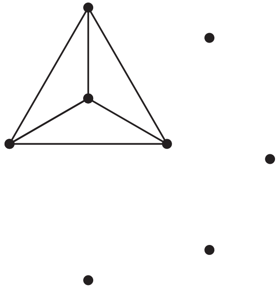
The Adversarial Case

- Suppose $G = (V, E)$ is generated as follows:



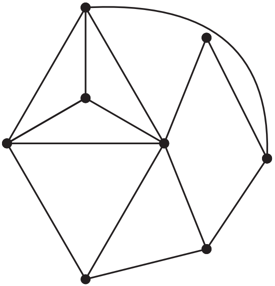
The Adversarial Case

- Suppose $G = (V, E)$ is generated as follows:
  - For $V^* \subset V$, for all $i, j \in V^*$ add (i, j) to E.



The Adversarial Case

- Suppose $G = (V, E)$ is generated as follows:
    - For $V^* \subset V$, for all $i, j \in V^*$ add (i, j) to E.
    - An adversary chooses a number of the remaining potential edges to be added to the graph.



Theorem

- *Suppose that every vertex in* $V - V^*$ *is adjacent to at most* δn *vertices in* V^* *for* $0 < \delta < 1$.
- *Then* G *can contain at most* $O(n^2)$ *edges other than those in* $V^* \times V^*$ *and* K_{V^*} *will still be the maximum clique of* G.

Theorem

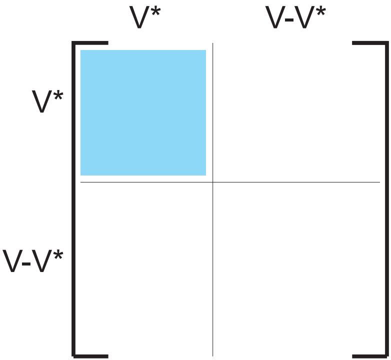
- *Suppose that every vertex in* $V - V^*$ *is adjacent to at most* δn *vertices in* V^* *for* $0 < \delta < 1$.
- *Then* G *can contain at most* $O(n^2)$ *edges other than those in* $V^* \times V^*$ *and* K_{V^*} *will still be the maximum clique of* G.

- To prove we want to find W, λ, μ satisfying the KKT conditions.
- Take $\mu = 1/n$.

Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$



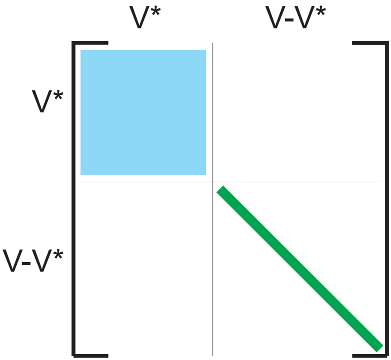
Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

$$W(i,j) = 1/n.$$



Construction of W

(1) $(i,j) \in V^* \times V^*$:

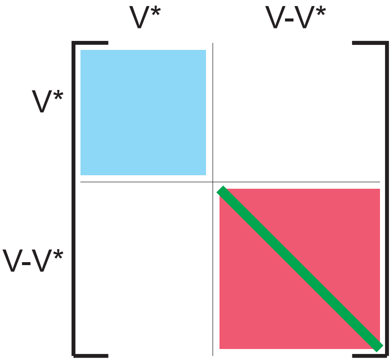
$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

$$W(i,j) = 1/n.$$

(3) $(i,j) \in (V - V^*) \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = 0.$$



Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

$$W(i,j) = 1/n.$$

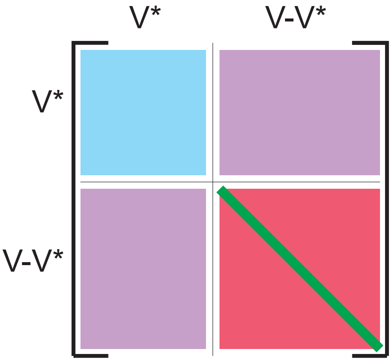
(3) $(i,j) \in (V - V^*) \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = 0.$$

(4) $(i,j) \in V^* \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = -\frac{p_j}{n(n - p_j)};$$

p_j is the # of edges from j to V^*.



Proof: $W\bar{v} = 0$

- First show $W\bar{v} = 0$
- If $i \in V^*$, then $W(i,j) = 0$ for all $j \in V^*$. So

$$W(i,:)\bar{v} = \sum_{j \in V^*} W(i,j) = 0$$

- If $i \in V - V^*$,

$$W(i,:)\bar{v} = p_i \frac{1}{n} - (n - p_i)\frac{p_i}{n(n-p_i)} = 0.$$

- So $W\bar{v} = 0$.

Proof : $\|W\| \leq 1$

- Write $W = W_D + W_0$. Then

$$\|W\|^2 \leq 2(\|W_D\|^2 + \|W_0\|^2) \leq 2(1/n^2 + \|W_0\|^2).$$

- $\|W_0\|^2 \leq \|W_0\|_F^2 = \|W_0(V - V^*, V - V^*)\|_F^2 + 2\|W(V^*, V - V^*)\|_F^2$.
- At most r edges are not in the clique so

$$\|W_0(V - V^*, V - V^*)\|_F^2 \leq 2r/n^2.$$

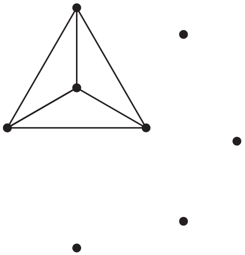
- Since $n - p_j \geq (1 - \delta)n$,

$$\|W(V^*, V - V^*)\|_F^2 = \sum_{j \in V - V^*} \frac{p_j}{n^2} + (n - p_j)\frac{p_j^2}{(n - p_j)^2}$$
$$\leq const \cdot \sum_{j \in V - V^*} \frac{p_j}{n^2} \leq const \cdot \frac{r}{n^2}.$$

- So $\|W\|^2 \leq 1$ if $r \leq const \cdot n^2$. □

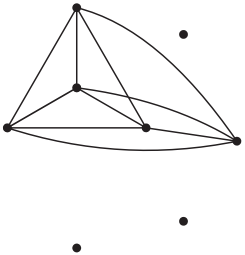
Example

- Suppose an adversary can add n edges from $v \in V - V^*$ to K_{V^*}:



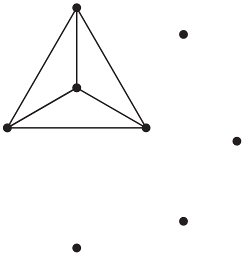
Example

- Suppose an adversary can add n edges from $v \in V - V^*$ to K_{V^*}:
- Then $K_{V^*} \cup \{v\}$ is a bigger clique.



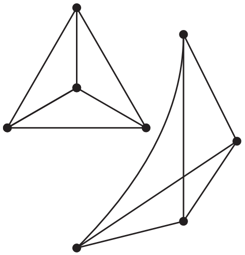
Example

- Suppose adversary can add n^2 edges not in the clique.



Example

- Suppose adversary can add n^2 edges not in the clique.
- Can form an additional clique of size n^2.



The Random Case

- Form $G = (V, E)$ as follows:
  - For $V^* \subset V$, for all $i, j \in V^*$ add (i, j) to E.
  - Add each remaining potential edge to the graph independently with probability p.

Feige-Krauthgamer Algorithm

- Suppose $G = (V, E)$ is randomly generated as on last slide with a clique of size $k > c\sqrt{|V|}$ for sufficiently large c.
- Then, with probability exponentially close to 1, $\vartheta(\overline{G}) = k$ and

$$\vartheta(\overline{G \setminus v}) = \begin{cases} k - 1 & \text{if } v \text{ is in the clique,} \\ k & \text{otherwise} \end{cases}$$

  for every $v \in V(G)$.
- **Algorithm:** Check every node v in G:
  - If $\vartheta(\overline{G \setminus v}) < \vartheta(\overline{G}) - 1/2$, add v to K.
  - When done, K is the maximum clique of G with high probability.

Theorem

- *Suppose* $n > \alpha\sqrt{|V|}$ *for sufficiently large* $\alpha > 0$.
- *Then* K_{V^*} *is the unique maximum clique of* G *and* X^* *is the optimal solution of the nuclear norm relaxation with probability exponentially close to 1.*

Theorem

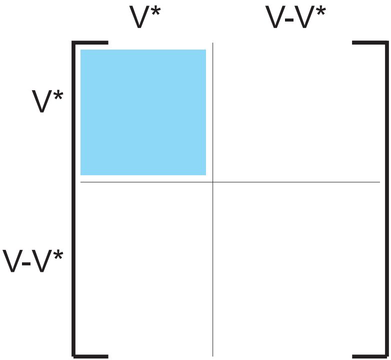
- *Suppose* $n > \alpha\sqrt{|V|}$ *for sufficiently large* $\alpha > 0$.
- *Then* K_{V^*} *is the unique maximum clique of* G *and* X^* *is the optimal solution of the nuclear norm relaxation with probability exponentially close to 1.*

- Take $\mu = 1/n$. Want to find W, λ that satisfy the optimality conditions with high probability.

Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$



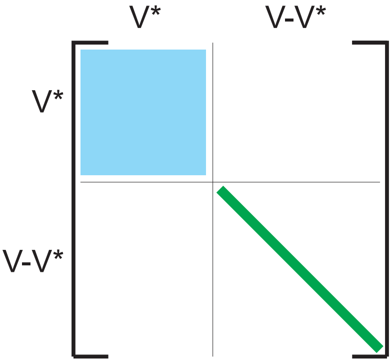
Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

$$W(i,j) = 1/n.$$



Construction of W

(1) $(i,j) \in V^* \times V^*$:

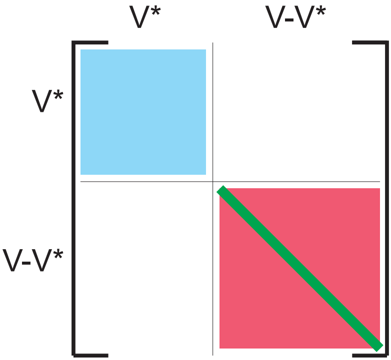
$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

$$W(i,j) = 1/n.$$

(3) $(i,j) \in (V - V^*) \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = -\frac{p}{n(1-p)}.$$



Construction of W

(1) $(i,j) \in V^* \times V^*$:

$$W(i,j) = 0.$$

(2) $(i,j) \notin V^* \times V^*$, $ij \in E$ or $i = j$:

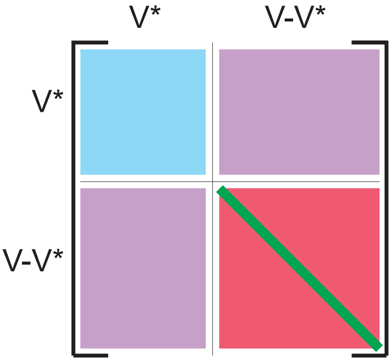
$$W(i,j) = 1/n.$$

(3) $(i,j) \in (V - V^*) \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = -\frac{p}{n(1-p)}.$$

(4) $(i,j) \in V^* \times (V - V^*)$, $ij \notin E$:

$$W(i,j) = -\frac{p_j}{(n - p_j).}$$



Proof

- Write $W = W_1 + W_2 + \cdots + W_5$.
- If $(i,j) \in V^* \times V^*$ or $i = j$ choose

$$W_1(i,j) = \begin{cases} 1/n & \text{with prob } p \\ -p/n(1-p) & \text{with prob } 1-p \end{cases}$$

- Otherwise,

$$W_1(i,j) = \begin{cases} 1/n & \text{if } (i,j) \in E \\ -p/n(1-p) & \text{otherwise} \end{cases}$$

Lemma (Füredi-Komlós)

- *Suppose a random symmetric matrix* $A = [A_{ij}] \in \Sigma^n$ *is defined as follows.*
  - *Take* $A_{ij} = 1$ *with probability* p, $-p/(1-p)$ *with probability* $1-p$ *for all* $1 \leq i \leq j \leq n$.
  - *Define* $A_{ji} = A_{ij}$ *for all* $j > i$.
- *Call this distribution* Ω.
- *Then*

$$\|A\| \leq 3\sigma\sqrt{n}$$

with probability $\geq 1 - \exp(-cn^{1/6})$ *for some* c *depending on* σ.

Lemma (Füredi-Komlós)

- *Suppose a random symmetric matrix* $A = [A_{ij}] \in \Sigma^n$ *is defined as follows.*
  - *Take* $A_{ij} = 1$ *with probability* p, $-p/(1-p)$ *with probability* $1-p$ *for all* $1 \leq i \leq j \leq n$.
  - *Define* $A_{ji} = A_{ij}$ *for all* $j > i$.
- *Call this distribution* Ω.
- *Then*

$$\|A\| \leq 3\sigma\sqrt{n}$$

with probability $\geq 1 - \exp(-cn^{1/6})$ *for some* c *depending on* σ.

- So $\|W_1\| \leq 3\sigma\sqrt{N}/n$ with probability at least $1 - \exp(c_1 N^{1/6})$

Proof (Cont'd)

- Let $W_2(i,j) = W(i,j) - W_1(i,j)$ for all $(i,j) \in V^* \times V^*$, $W_2(i,j) = 0$ otherwise.
- Applying the lemma again, $\|W_2\| \leq 3\sigma/\sqrt{n}$ with probability at least $1 - \exp(c_1 n^{1/6})$.

Proof (Cont'd)

- Let $W_2(i,j) = W(i,j) - W_1(i,j)$ for all $(i,j) \in V^* \times V^*$, $W_2(i,j) = 0$ otherwise.
- Applying the lemma again, $\|W_2\| \leq 3\sigma/\sqrt{n}$ with probability at least $1 - \exp(c_1 n^{1/6})$.

- Let $W_3(i,i) = W(i,i) - W_1(i,i)$ for all $i \in V - V^*$ and $W_3(i,j) = 0$ otherwise.
- Then $W_3 \leq 2/n$.

Lemma

- *Suppose* $A \in \mathbf{R}^{M \times N}$ *with entries chosen according to* Ω.
- *Define* $\tilde{A} \in \mathbf{R}^{M \times N}$ *by*

$$\tilde{A}_{ij} = \begin{cases} 1 & \text{if } A_{ij} = 1, \\ -n_j/(n - n_j) & \text{if } A_{ij} = -p/(1-p) \end{cases}$$

where n_j *is the # of 1's in column* j *of* A.

- *Then there exists* $c_1 > 0, 0 < c_2 < 1$ *depending on* p *such that*

$$\|A - \tilde{A}\|_F^2 \leq c_1 N \quad \text{and} \quad n_j < M \;\; \forall j = 1, 2, \ldots, N$$

with probability at least $1 - (2/3)^N - N c_2^M$.

Proof (Conclusion)

- Take $W_4(V^*, V - V^*) = W(V^*, V - V^*) - W_1(V^*, V - V^*)$, W_4 zero in all other entries.
- So $W_4(V^*, V^*)$ is in the form $A - \tilde{A}/n$ from the lemma:

$$\|W_4\|^2 \leq \|W_4\|_F^2 = \|W_4(V - V^*)\|_F^2 \leq \alpha N/n^2$$

  for some $\alpha > 0$ with probability tending exponentially to 1 as $N \to \infty$ if $n = \Omega(N)$.
- Let $W_5 := W_4^T$. Then

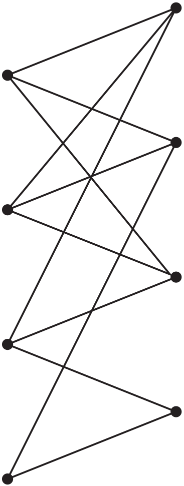
$$\|W_5\|^2 \leq \alpha N/n^2$$

  with the same probability.

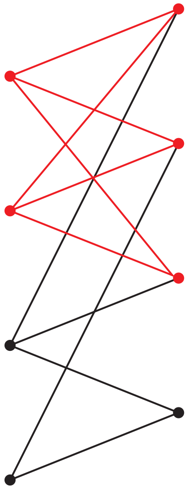
The Maximum Edge Biclique Problem

- Given a bipartite graph $G = (U, V, E)$, the **maximum edge biclique problem** is to find a complete bipartite subgraph of G with maximum number of edges.
- **Peeters:** MEBP is NP-hard.

Example



Example



Nuclear norm relaxation of MEBP

- A biclique G containing mn edges can be found by solving

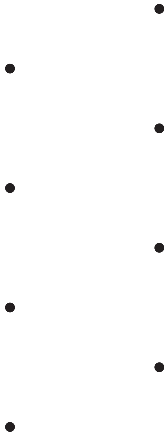
$$\begin{aligned} \min \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & \sum_{i \in U} \sum_{j \in V} X_{ij} \geq mn, \\ & X_{ij} = 0 \text{ if } (i,j) \notin E \\ & X \in [0,1]^{U \times V}. \end{aligned}$$

- We consider the convex relaxation

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \sum_{i \in U} \sum_{j \in V} X_{ij} \geq mn, \\ & X_{ij} = 0 \text{ if } (i,j) \notin E. \end{aligned}$$

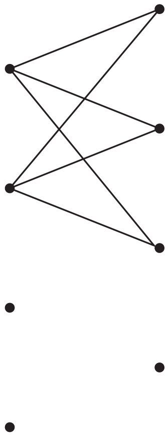
The Adversarial Case

- Suppose bipartite $G = (U, V, E)$ is generated as follows:



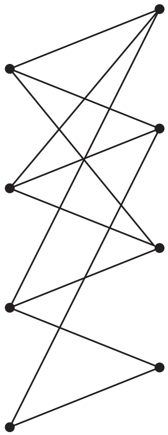
The Adversarial Case

- Suppose bipartite $G = (U, V, E)$ is generated as follows:
  - First add a biclique $U^* \times V^*$ with $U^* \subseteq U$, $V^* \subseteq V$. Let $|U^*| = m$, $|V^*| = n$.



The Adversarial Case

- Suppose bipartite $G = (U, V, E)$ is generated as follows:
  - First add a biclique $U^* \times V^*$ with $U^* \subseteq U$, $V^* \subseteq V$. Let $|U^*| = m$, $|V^*| = n$.
  - An adversary chooses a number of the remaining potential edges to be added to the graph.



Theorem

Suppose that

- *G contains at most r edges not in* $U^* \times V^*$,
- *each vertex of* $V - V^*$ *is adjacent to at most* αm *vertices of* U^* *for* $\alpha \in (0, 1)$, *and*
- *each vertex of* $U - U^*$ *is adjacent to at most* βn *vertices of* V^* *for* $\beta \in (0, 1)$.

Then there exists constant $c \in (0, 1)$ *depending on* α, β *such that* $U^* \times V^*$ *is the unique maximum edge biclique of G if* $r < cmn$.

- As with the maximum clique problem, this bound is the best possible.

The Random Case

- Suppose bipartite $G = (U, V, E)$ such that $|V| = N$, $|U| = \lceil yN \rceil$ for some $y > 0$ is generated as follows:
  - First add biclique $U^* \times V^*$ with $|V^*| = n$, $|U^*| = \lceil zn \rceil$ for some $z > 0$
  - Add each remaining potential edge to the graph with probability p (independently).

The Random Case

- Suppose bipartite $G = (U, V, E)$ such that $|V| = N$, $|U| = \lceil yN \rceil$ for some $y > 0$ is generated as follows:
  - First add biclique $U^* \times V^*$ with $|V^*| = n$, $|U^*| = \lceil zn \rceil$ for some $z > 0$
  - Add each remaining potential edge to the graph with probability p (independently).

Theorem

If $n = \Omega(\sqrt{N})$ then $U^* \times V^*$ is the unique maximum edge biclique of G with probability tending exponentially to 1 as $N \to \infty$ and $X^* = \bar{u}\bar{v}^T$ is the unique solution to the nuclear norm relaxation.

Conclusions and Future Work

- Have shown that for two hard combinatorial problems, nuclear norm minimization returns the exact solution in poly-time in certain cases.
- Future work:
  - Apply the same technique to other hard problems (NMF, clustering).
  - Specialized algorithms.