# Normalized Matching Property in Random & Pseudorandom Bipartite Graphs

Deepanshu Kush

Department of Mathematics
Indian Institute of Technology Bombay (IITB)

(Joint work with Niranjan Balachandran, IITB)

September 19, 2019

### Definition

*A $k \times n$ star array is a $k \times n$ array $\mathcal{A}$ whose entries are $*$ or blanks.*

**EXAMPLE:**

$$\mathcal{A} := \begin{pmatrix} * & * & & * & & \\ & & * & * & & * \\ & & & * & * & * \end{pmatrix}$$

# 2 Problems

Definition

*A* $k \times n$ *star array is a* $k \times n$ *array* $\mathcal{A}$ *whose entries are* $*$ *or blanks.*

**EXAMPLE:**

$$\mathcal{A} := \begin{pmatrix} * & * & & * & & \\ & & * & * & & * \\ & & & * & * & * \end{pmatrix}$$

**Problem 1:** Given a star array $\mathcal{A}$ when is it possible to replace some of the $*$ by non-negative integers (blanks become zero) s.t. in the resulting integral array, all row sums equal $R$ and all column sums equal $C$ for some $R, C > 0$?

# 2 Problems (contd.)

**Problem 2:** Suppose $q$ is a (large) prime, and suppose $X, Y \subset \mathbb{F}_q$ s.t. $|Y| = 10|X|$, and $|X| \geq q/1000$, is it possible to partition $Y := Y_1 \sqcup \cdots \sqcup Y_{|X|}$ s.t. for each $x \in X$

- $|Y_x| = 10$,
- For each $y \in Y_x$, $x + y$ is a quadratic residue?

# In graph theoretic terms...

# In graph theoretic terms...

Problem 1: If $\mathcal{A}$ is a star array, there is an associated bipartite graph $G_{\mathcal{A}} = G(X, Y, E)$:

- $X$= Set of Rows of $\mathcal{A}$, $Y$= Set of Columns of $\mathcal{A}$,
- For $x \in X, y \in Y$, $(x, y) \in E$ iff $\mathcal{A}(x, y) = *$.

Problem 2: Consider the bipartite graph $G(X, Y, E)$ where for $x \in X, y \in Y$, $(x, y) \in E$ iff $x + y$ is a quadratic residue.

# Perfect Matchings in Bipartite Graphs

Suppose $k = n$. If the associated bipartite graph has a *perfect matching* (PM), i.e., a set of pairwise disjoint edges that span all the vertices then Problem 1 admits an affirmative solution.

# Perfect Matchings in Bipartite Graphs

Suppose $k = n$. If the associated bipartite graph has a *perfect matching* (PM), i.e., a set of pairwise disjoint edges that span all the vertices then Problem 1 admits an affirmative solution.

The converse also holds: If $G_{\mathcal{A}}$ has no PM then the star array does **not** have this property:

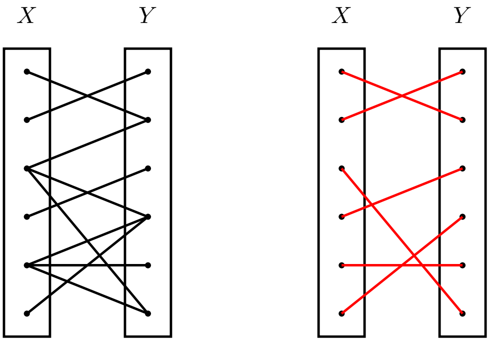
# Perfect Matchings in Bipartite Graphs

Suppose $k = n$. If the associated bipartite graph has a *perfect matching* (PM), i.e., a set of pairwise disjoint edges that span all the vertices then Problem 1 admits an affirmative solution.

The converse also holds: If $G_{\mathcal{A}}$ has no PM then the star array does **not** have this property:

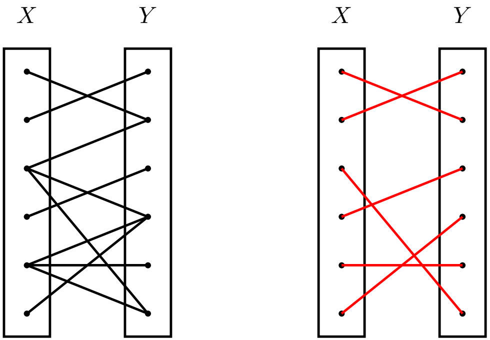
Hall's theorem: $G(X, Y)$ has PM iff $\forall S \subseteq X, |N(S)| \geq |S|$. Here, $N(S)$ is the set of neighbors of $S$.

# Hall's Theorem: An illustration




Each $S \subseteq X$ satisfies $|N(S)| \geq |S|$. So, $G$ has a perfect matching.

It states that $G(X, Y)$ has a perfect matching iff $\forall S \subseteq X, |N(S)| \geq |S|$.




Each $S \subseteq X$ satisfies $|N(S)| \geq |S|$. So, $G$ has a perfect matching.

It states that $G(X,Y)$ has a perfect matching iff $\forall S \subseteq X, |N(S)| \geq |S|$.
What is an analogous result in the case when $|X| = k$ and $|Y| = n$?

Definition

*$G = G(X, Y)$ is said to have the* **Normalized Matching Property (NMP)** *if*

$$\frac{|N(S)|}{|Y|} \geq \frac{|S|}{|X|}$$

*for all $S \subseteq X$.*

# The Normalized Matching Property in Bipartite graphs

Definition

*$G = G(X, Y)$ is said to have the* **Normalized Matching Property (NMP)** *if*

$$\frac{|N(S)|}{|Y|} \geq \frac{|S|}{|X|}$$

*for all $S \subseteq X$.*

In particular, if $|X| = |Y|$, then this is the familiar Hall's condition for the existence of PM in $G$.

# The Normalized Matching Property in Bipartite graphs

Definition

*$G = G(X, Y)$ is said to have the* **Normalized Matching Property (NMP)** *if*

$$\frac{|N(S)|}{|Y|} \geq \frac{|S|}{|X|}$$

*for all $S \subseteq X$.*

In particular, if $|X| = |Y|$, then this is the familiar Hall's condition for the existence of PM in $G$.

**Notation:** For $A \subseteq X, B \subseteq Y$, $G(A, B)$ denotes the subgraph induced by the vertices in $A \cup B$. $e(A, B) := |\{(A \times B) \cap E(G)\}|$.

# Equivalent Criteria

NMP in bipartite graphs is rather well-understood due to the following

**Theorem**

*(Kleitman '74) The following are equivalent:*

- *(NMP)* $G$ *with* $|X| = k, |Y| = n$ *has NMP.*
- *(LYM) For any independent set* $I$ *in* $G$, $\frac{|I \cap X|}{k} + \frac{|I \cap Y|}{n} \leq 1$.
- *(REG) There exists* $w : E \to \mathbb{N} \cup \{0\}$ *such that* $\sum_{\substack{e \ni x \\ e \in E}} w(e)$ *(resp.* $\sum_{\substack{e \ni y \\ e \in E}} w(e)$*) is equal for all* $x \in X$ *(resp. for all* $y \in Y$*).*

# Equivalent Criteria

NMP in bipartite graphs is rather well-understood due to the following

**Theorem**

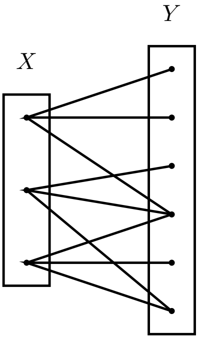
*(Kleitman '74) The following are equivalent:*

- *(NMP)* $G$ *with* $|X| = k, |Y| = n$ *has NMP.*
- *(LYM) For any independent set* $I$ *in* $G$, $\frac{|I \cap X|}{k} + \frac{|I \cap Y|}{n} \leq 1$.
- *(REG) There exists* $w : E \to \mathbb{N} \cup \{0\}$ *such that* $\sum_{\substack{e \ni x \\ e \in E}} w(e)$ *(resp.* $\sum_{\substack{e \ni y \\ e \in E}} w(e)$*) is equal for all* $x \in X$ *(resp. for all* $y \in Y$*).*

By condition REG, it follows that the first problem reduces to whether or not the corresponding graph has NMP.

# Structural Characterization: An illustration

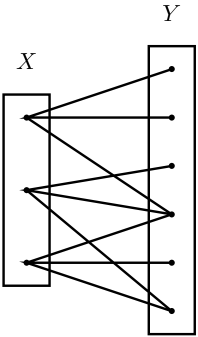
In the first example, the graph $G_{\mathcal{A}}$ is



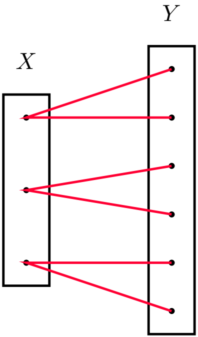

Every $S \subseteq X$ satisfies $|N(S)| \geq 2|S|$

# Structural Characterization: An illustration

In the first example, the graph $G_{\mathcal{A}}$ is




Every $S \subseteq X$ satisfies $|N(S)| \geq 2|S|$

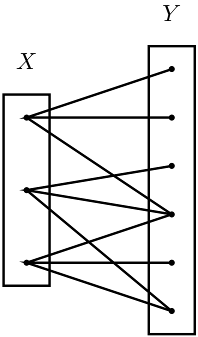



$G$ is spanned by an $X$ 2-thrill

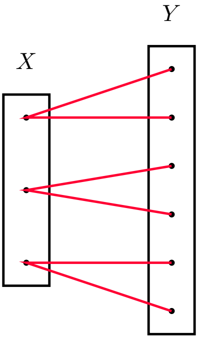
True in general, i.e. when $\frac{n}{k} = q \in \mathbb{N}$.

# Structural Characterization: An illustration

In the first example, the graph $G_{\mathcal{A}}$ is




Every $S \subseteq X$ satisfies $|N(S)| \geq 2|S|$




$G$ is spanned by an $X$ 2-thrill

True in general, i.e. when $\frac{n}{k} = q \in \mathbb{N}$.But what about when $n/k \ni \mathbb{N}$? Is there an appropriate generalisation? We shall return to this later.

# A structural characterization for NMP when $k \mid n$

A copy of $K_{1,q}$ in $G$ is called an $X$ $q$-fan when the star is 'pointed' at a vertex in $X$.

# A structural characterization for NMP when $k \mid n$

A copy of $K_{1,q}$ in $G$ is called an $X$ $q$-fan when the star is 'pointed' at a vertex in $X$.

A bunch of vertex disjoint $X$ $q$-fans together is called an $X$ $q$-thrill. In particular, an $X$ $1$-thrill is simply a matching.

# A structural characterization for NMP when $k \mid n$

A copy of $K_{1,q}$ in $G$ is called an $X$ $q$-fan when the star is 'pointed' at a vertex in $X$.

A bunch of vertex disjoint $X$ $q$-fans together is called an $X$ $q$-thrill. In particular, an $X$ 1-thrill is simply a matching.

**Lemma**

*If $n = qk$ for some $q \in \mathbb{N}$. Then $G$ has NMP iff $G$ is spanned by an $X$ $q$-thrill.*

# A structural characterization for NMP when $k \mid n$

A copy of $K_{1,q}$ in $G$ is called an $X$ $q$-fan when the star is 'pointed' at a vertex in $X$.

A bunch of vertex disjoint $X$ $q$-fans together is called an $X$ $q$-thrill. In particular, an $X$ 1-thrill is simply a matching.

**Lemma**

*If $n = qk$ for some $q \in \mathbb{N}$. Then $G$ has NMP iff $G$ is spanned by an $X$ $q$-thrill.*

Proof: Clone $q$ copies of each vertex of $X$ and apply Hall's theorem.

# A structural characterization for NMP when $k \mid n$

A copy of $K_{1,q}$ in $G$ is called an $X$ $q$-fan when the star is 'pointed' at a vertex in $X$.

A bunch of vertex disjoint $X$ $q$-fans together is called an $X$ $q$-thrill. In particular, an $X$ $1$-thrill is simply a matching.

### Lemma

*If $n = qk$ for some $q \in \mathbb{N}$. Then $G$ has NMP iff $G$ is spanned by an $X$ $q$-thrill.*

Proof: Clone $q$ copies of each vertex of $X$ and apply Hall's theorem.

The second problem also reduces to determining if the corresponding graph has NMP.

# Complexity of checking for NMP

Checking if a given $G(X, Y)$ with $|X| = k, |Y| = n$ can be done in $\mathrm{Poly}(n, k)$:

- Clone each $x \in X$ into $x_1, \ldots, x_n$,
- Clone each $y \in Y$ into $y_1, \ldots, y_k$,
- Check if the resulting graph has a PM. To determine if a graph $G(V, E)$ has a PM can be done in $O(|E|\sqrt{|V|})$.

# Complexity of checking for NMP

Checking if a given $G(X, Y)$ with $|X| = k, |Y| = n$ can be done in $\text{Poly}(n, k)$:

- Clone each $x \in X$ into $x_1, \ldots, x_n$,
- Clone each $y \in Y$ into $y_1, \ldots, y_k$,
- Check if the resulting graph has a PM. To determine if a graph $G(V, E)$ has a PM can be done in $O(|E|\sqrt{|V|})$.

For Problem 2, how do we check if NMP holds in that associated graph?

# Complexity of checking for NMP

Checking if a given $G(X, Y)$ with $|X| = k, |Y| = n$ can be done in $\mathrm{Poly}(n, k)$:

- Clone each $x \in X$ into $x_1, \ldots, x_n$,
- Clone each $y \in Y$ into $y_1, \ldots, y_k$,
- Check if the resulting graph has a PM. To determine if a graph $G(V, E)$ has a PM can be done in $O(|E|\sqrt{|V|})$.

For Problem 2, how do we check if NMP holds in that associated graph?

Is it even true?! If not, how true is it?

# NMP in posets

- NMP in ranked posets is a very important property and a crucial hypothesis in several conjectures.

# NMP in posets

- NMP in ranked posets is a very important property and a crucial hypothesis in several conjectures.
- Many interesting posets (Boolean lattice, poset of flats in finite projective space etc) are all NMP posets though their corresponding graphs are relatively very sparse.

# NMP in posets

- NMP in ranked posets is a very important property and a crucial hypothesis in several conjectures.
- Many interesting posets (Boolean lattice, poset of flats in finite projective space etc) are all NMP posets though their corresponding graphs are relatively very sparse.

**Question**: How dense must a ‘typical’ bipartite graph be for it to have NMP?

# Enter Randomness: $\mathbb{G}(k, n, p)$

$\mathbb{G}(n, p)$: The Erdős-Rényi random graph: Each pair $(u, v)$ is an edge *independently* with probability $p$.

# Enter Randomness: $\mathbb{G}(k, n, p)$

$\mathbb{G}(n, p)$: The Erdős-Rényi random graph: Each pair $(u, v)$ is an edge *independently* with probability $p$.

$\mathbb{G}(k, n, p)$: Random bipartite graph with the vertex partition $(X, Y)$ with $|X| = k, |Y| = n$: Each $(x, y) \in X \times Y$ is an edge *independently* with probability $p$.

# Enter Randomness: $\mathbb{G}(k, n, p)$

$\mathbb{G}(n, p)$: The Erdős-Rényi random graph: Each pair $(u, v)$ is an edge *independently* with probability $p$.

$\mathbb{G}(k, n, p)$: Random bipartite graph with the vertex partition $(X, Y)$ with $|X| = k, |Y| = n$: Each $(x, y) \in X \times Y$ is an edge *independently* with probability $p$.

A **Graph Property** is a subset of all graphs closed under graph isomorphism. A graph property $\mathcal{P}$ is *monotone* if the collection is closed w.r.t. taking supergraphs, i.e., if $G \in \mathcal{P}$ and $G \subset H$ then $H \in \mathcal{P}$.

# Threshold for graph property

$p_0 = p_0(n)$ is a threshold for a property $\mathcal{P}$ if $\forall p(n)$,

$$\Pr[\mathbb{G}(n,p) \text{ has } \mathcal{P}] \to \begin{cases} 0, & \text{if } p/p_0 \to 0 \\ 1, & \text{if } p/p_0 \to \infty \end{cases}$$

# Threshold for graph property

$p_0 = p_0(n)$ is a threshold for a property $\mathcal{P}$ if $\forall p(n)$,

$$\Pr[\mathbb{G}(n,p) \text{ has } \mathcal{P}] \to \begin{cases} 0, & \text{if } p/p_0 \to 0 \\ 1, & \text{if } p/p_0 \to \infty \end{cases}$$

Theorem (Bollobás, Thomason, 85)

*Every monotone graph property has a threshold.*

# Threshold for Perfect matchings in $\mathbb{G}(n, n, p)$

**Theorem (Erdős-Rényi, 66')**

*For $\varepsilon > 0$, and $n \gg 0$,*

- *If $p < \frac{(1-\varepsilon)\log n}{n}$, then whp $\mathbb{G}(n, n, p)$ does not have PM.*
- *If $p > \frac{(1+\varepsilon)\log n}{n}$, then $\mathbb{G}(n, n, p)$ has PM. whp.*

Here *whp* (with high probability) means

$$\mathbb{P}(\mathbb{G}(n, n, p) \text{ has PM}) \to 1 \text{ as } n \to \infty.$$

$\frac{\log n}{n}$ is a *sharp* threshold for the existence of perfect matchings in $\mathbb{G}(n, n, p)$.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.
- Same argument for $\mathbb{G}(k, n, p)$ to give: If $p < \frac{(1-\varepsilon)\log n}{k}$, $\mathbb{G}(k, n, p)$ does not have NMP *whp*.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.
- Same argument for $\mathbb{G}(k, n, p)$ to give: If $p < \frac{(1-\varepsilon)\log n}{k}$, $\mathbb{G}(k, n, p)$ does not have NMP *whp*.

**A Heuristic:**

- Clone each vertex of $X$ $n/k$ times to get a new graph $\mathbb{G}'(X', Y)$.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.
- Same argument for $\mathbb{G}(k, n, p)$ to give: If $p < \frac{(1-\varepsilon)\log n}{k}$, $\mathbb{G}(k, n, p)$ does not have NMP *whp*.

**A Heuristic:**

- Clone each vertex of $X$ $n/k$ times to get a new graph $\mathbb{G}'(X', Y)$.
- By Kleitman's theorem, $G$ has NMP if and only if the new graph has a perfect matching.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.
- Same argument for $\mathbb{G}(k,n,p)$ to give: If $p < \frac{(1-\varepsilon)\log n}{k}$, $\mathbb{G}(k,n,p)$ does not have NMP *whp*.

**A Heuristic:**

- Clone each vertex of $X$ $n/k$ times to get a new graph $\mathbb{G}'(X', Y)$.
- By Kleitman's theorem, $G$ has NMP if and only if the new graph has a perfect matching.
- If $\mathbb{G}' \sim \mathbb{G}(n,n,p')$ (?!!) we need $p' \gtrsim \frac{\log n}{n}$.

# Threshold for NMP?

Suppose $p < \frac{(1-\varepsilon)\log n}{n}$. Let $N =$ the number of isolated vertices in $Y$.

- $\mathbb{E}(N) = n(1-p)^n$ and by standard concentration bounds (Chernoff), $N > 0$ *whp*.
- Same argument for $\mathbb{G}(k,n,p)$ to give: If $p < \frac{(1-\varepsilon)\log n}{k}$, $\mathbb{G}(k,n,p)$ does not have NMP *whp*.

**A Heuristic:**

- Clone each vertex of $X$ $n/k$ times to get a new graph $\mathbb{G}'(X',Y)$.
- By Kleitman's theorem, $G$ has NMP if and only if the new graph has a perfect matching.
- If $\mathbb{G}' \sim \mathbb{G}(n,n,p')$ (?!!) we need $p' \gtrsim \frac{\log n}{n}$.
- Each vertex of $X$ = union of $n/k$ vertices of $X'$, so threshold for NMP is $\frac{n}{k} \cdot \frac{\log n}{n} = \frac{\log n}{k}$.

# Our Results: A sharp threshold for NMP

**Theorem**

*Suppose* $\varepsilon > 0, k \gg_{\varepsilon} 0$, *and* $k \leq n < \exp(k)$. *Then*

- *If* $p < \frac{(1-\varepsilon)\log n}{k}$, *then whp* $\mathbb{G}(k, n, p)$ *does not have NMP.*

# Our Results: A sharp threshold for NMP

**Theorem**

*Suppose* $\varepsilon > 0, k \gg_\varepsilon 0$, *and* $k \leq n < \exp(k)$. *Then*

- *If* $p < \frac{(1-\varepsilon)\log n}{k}$, *then whp* $\mathbb{G}(k,n,p)$ *does not have NMP.*
- *If* $p > \frac{(1+\varepsilon)\log n}{k}$, *then* $\mathbb{G}(k,n,p)$ *has NMP whp.*

# Our Results: A sharp threshold for NMP

**Theorem**

*Suppose* $\varepsilon > 0, k \gg_\varepsilon 0$, *and* $k \leq n < \exp(k)$. *Then*

- *If* $p < \frac{(1-\varepsilon)\log n}{k}$, *then whp* $\mathbb{G}(k,n,p)$ *does not have NMP.*
- *If* $p > \frac{(1+\varepsilon)\log n}{k}$, *then* $\mathbb{G}(k,n,p)$ *has NMP whp.*

$\frac{\log n}{k}$ is a sharp threshold for NMP in $\mathbb{G}(k,n,p)$.

# Pseudorandom Graphs: Brief Introduction

What does it mean to say that a graph behaves 'random-like'?

# Pseudorandom Graphs: Brief Introduction

What does it mean to say that a graph behaves 'random-like'?

Theorem (Erdős, Goldberg, Pach & Spencer '88)

*Let $p = p(n) \leq 0.99$. Then asymptotically almost surely, in the binomial random graph $\mathbb{G}(n, p)$, for any two subsets $X, Y \subseteq V(G)$,*

$$|e(X, Y) - p|X||Y|| \leq O(\sqrt{pn|X||Y|}).$$

# Pseudorandom Graphs: Brief Introduction

What does it mean to say that a graph behaves 'random-like'?

**Theorem (Erdős, Goldberg, Pach & Spencer '88)**

*Let $p = p(n) \leq 0.99$. Then asymptotically almost surely, in the binomial random graph $\mathbb{G}(n, p)$, for any two subsets $X, Y \subseteq V(G)$,*

$$|e(X, Y) - p|X||Y|| \leq O(\sqrt{pn|X||Y|}).$$

Here is one way to capture 'random-like' behavior. Write $p = \frac{e(G)}{\binom{n}{2}}$.

- (CUT SIZES) If $U, W$ are subsets of $V(G_n)$, then

$$e(U, W) \approx p|U||W|.$$

# Pseudorandom graphs: An introduction

## Definition

(Thomason) A graph $G$ on vertex set $V$ is $(p, \beta)$-jumbled if, for all vertex subsets $X, Y \subseteq V(G)$,

$$|e(X,Y) - p|X||Y|| \leq \beta\sqrt{|X||Y|}$$

# Pseudorandom graphs: An introduction

**Definition**

(Thomason) A graph $G$ on vertex set $V$ is $(p, \beta)$-jumbled if, for all vertex subsets $X, Y \subseteq V(G)$,

$$|e(X,Y) - p|X||Y|| \leq \beta\sqrt{|X||Y|}$$

In the context of bipartite graphs:

**Definition (Following Thomason '89)**

Suppose $0 < p < 1$ and $0 \leq \varepsilon < 1$. A bipartite graph $G(X,Y)$ with $|X| = k \leq n = |Y|$ is called T-pseudorandom with parameters $(p, \varepsilon)$ if

- For each $x \in X$, $d(x) \geq pn$,
- For $x \neq x', x, x' \in X$, $|N(x) \cap N(x')| \leq p^2 n(1+\varepsilon)$. Any two distinct vertices of $X$ have at most $p^2 n(1+\varepsilon)$ common neighbours.

# Main theorem of Thomason

These graphs are rightfully called pseudorandom because

**Theorem**

*Let $G(X, Y)$ be a bipartite graph with $|X| = k \leq n = |Y|$, which is T-pseudorandom with parameters $(p, \varepsilon)$. Then for every subset $A \subseteq X$ with $1/p \leq |A|$ and every subset $B \subseteq Y$,*

$$|e(A, B) - p|A||B|| \leq \sqrt{pn|A||B|(1 + \varepsilon p|A|)}.$$

# Examples of T-pseudorandom graphs

Let $X$ be the points in projective $d$-space over $\mathbb{F}_q$, $Y$ be the 'hyperplanes', then the corresponding incidence bipartite graph has vertex parts of sizes $|X| = |Y| = n := 1 + q + \cdots + q^{d-1}$, and is T-pseudorandom with parameters

$$p = n^{-1/2}(1 + o(1)), \quad \varepsilon = 0.$$

# Examples of T-pseudorandom graphs

Let $X$ be the points in projective $d$-space over $\mathbb{F}_q$, $Y$ be the 'hyperplanes', then the corresponding incidence bipartite graph has vertex parts of sizes $|X| = |Y| = n := 1 + q + \cdots + q^{d-1}$, and is T-pseudorandom with parameters

$$p = n^{-1/2}(1 + o(1)), \quad \varepsilon = 0.$$

The point-block incidence graphs for symmetric designs are also T-pseudorandom.

# A robust model for pseudorandomness

T-pseudorandomness is algorithmically easily verifiable as it is combinatorial in definition. It also has a certain sense of robustness:

# A robust model for pseudorandomness

T-pseudorandomness is algorithmically easily verifiable as it is combinatorial in definition. It also has a certain sense of robustness:

### Lemma (N.B., D. Kush, 2019+)

*Let $0 < \varepsilon < \frac{1}{2}$ and suppose $G(X, Y)$ is a T-pseudorandom bipartite graph with parameters $(p_0, \varepsilon_0)$ with $|X| = k \leq |Y| = n$, and suppose $p_0 \geq \frac{1}{\sqrt{k}}$. Then, for any integer $\varepsilon^3 n/2 \leq D \leq \varepsilon^3 n$, there exist subsets $C_X \subseteq X$ and $C_Y \subseteq Y$ such that*

- *$|C_Y| = D$ and $|C_X| \leq \eta k$, where $\eta = \exp(-\frac{C}{\varepsilon})$ for some fixed constant $C$,*
- *$G(X \setminus C_X, Y \setminus C_Y)$ is T-pseudorandom with parameters $(p_1, \varepsilon_1)$ where $p_1 = p_0(1 - \varepsilon)$ and $\varepsilon_1 \leq 5(\varepsilon_0 + 3\varepsilon)$.*

Allows for efficient randomized algorithmic constructions of several T-pseudorandom bipartite graphs.

# ‘Almost’ NMP in Bipartite Graphs: NMP-Approximability

Informally: If one can remove a small proportion of vertices from both parts s.t. the resulting graph has NMP, then it is ‘NMP-approximable’.

# 'Almost' NMP in Bipartite Graphs: NMP-Approximability

Informally: If one can remove a small proportion of vertices from both parts s.t. the resulting graph has NMP, then it is 'NMP-approximable'.
Formally,

**Definition (NMP-Approximability)**

Suppose $\varepsilon > 0$. For functions $f, g : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(x), g(x) \to 0$ as $x \to 0$, a bipartite graph $G(X, Y)$ is said to be $(f, g, \varepsilon)$-NMP approximable if there are subsets $\mathcal{X} \subseteq X$ and $\mathcal{Y} \subseteq Y$ such that:

- $\frac{|\mathcal{X}|}{|X|} \leq f(\varepsilon)$, $\frac{|\mathcal{Y}|}{|Y|} \leq g(\varepsilon)$
- $G(X \setminus \mathcal{X}, Y \setminus \mathcal{Y})$ has NMP.

# NMP-Approximability in T-pseudorandom graphs

Henceforth $|X| = k, |Y| = n$, and $k \leq n$.

# NMP-Approximability in T-pseudorandom graphs

Henceforth $|X| = k, |Y| = n$, and $k \leq n$.

Theorem (N.B., D. Kush, 2019+)

*Suppose* $0 \leq \varepsilon < 1$*, and let* $\omega : \mathbb{N} \to \mathbb{R}^+$ *be a non-negative valued function that satisfies* $\omega(k) \to \infty$ *as* $k \to \infty$*. There exists an integer* $k_0 = k_0(\varepsilon, \omega)$ *such that the following holds.*

*Suppose* $p \geq \frac{\omega(k)}{k}$ *and suppose* $G = G(X, Y)$ *is T-pseudorandom with parameters* $(p, \varepsilon)$*. Then* $G$ *is* $(f, g, \varepsilon)$*-NMP-approximable with*

$$f(x), g(x) = O\left(x^{1/4} \log(1/x)\right).$$

*Moreover, the deletion sets can be determined in polynomial time.*

# Something about the proofs: Threshold for NMP

Start with the LYM characterization for NMP: Let $p > \frac{(1+\varepsilon)\log n}{k}$ and $\mathbb{G} = \mathbb{G}(k, n, p)$ **not** have NMP.

# Something about the proofs: Threshold for NMP

Start with the LYM characterization for NMP: Let $p > \frac{(1+\varepsilon)\log n}{k}$ and $\mathbb{G} = \mathbb{G}(k, n, p)$ **not** have NMP.

Kleitman's theorem $\Rightarrow$ there exists $I = I_X \cup I_Y$ in $\mathbb{G}$ with $|I_X| = \ell$ and $|I_Y| \geq \left\lceil n\left(1 - \frac{\ell}{k}\right)\right\rceil$ for some $\ell > 0$.

# Something about the proofs: Threshold for NMP

Start with the LYM characterization for NMP: Let $p > \frac{(1+\varepsilon)\log n}{k}$ and $\mathbb{G} = \mathbb{G}(k,n,p)$ **not** have NMP.

Kleitman's theorem $\Rightarrow$ there exists $I = I_X \cup I_Y$ in $\mathbb{G}$ with $|I_X| = \ell$ and $|I_Y| \geq \left\lceil n\left(1 - \frac{\ell}{k}\right)\right\rceil$ for some $\ell > 0$. From the union bound,

$$\mathbb{P}(\mathbb{G} \text{ does not have NMP}) \leq \sum_{\ell=1}^{k} P_\ell$$

where for $1 \leq \ell \leq k$,

$$\begin{aligned} P_\ell &= \binom{k}{\ell}\binom{n}{\left\lceil n\left(1-\frac{\ell}{k}\right)\right\rceil}(1-p)^{\ell\left\lceil n\left(1-\frac{\ell}{k}\right)\right\rceil} \text{ for } \ell < k \\ P_k &= n\cdot(1-p)^k \leq \frac{1}{n^\varepsilon}. \end{aligned}$$

# The LYM approach

After some calculations (!) one can show $\sum_\ell P_\ell = o(1)$ if $n \gg k$ or if $p > \frac{10 \log n}{k}$.

To get the sharp threshold we need other ideas, more 'structure'.

# The proof of Erdős-Rényi for PM

Recall the Erdős-Rényi theorem: Sharp threshold for PM is $\frac{\log n}{n}$.
Suppose $p > \frac{(1+\varepsilon)\log n}{n}$ and $\mathbb{G}(n, n, p)$ does not admit PM, then there exists $S \subseteq X$ s.t. $|N(S)| < |S|$.

# The proof of Erdős-Rényi for PM

Recall the Erdős-Rényi theorem: Sharp threshold for PM is $\frac{\log n}{n}$. Suppose $p > \frac{(1+\varepsilon)\log n}{n}$ and $\mathbb{G}(n, n, p)$ does not admit PM, then there exists $S \subseteq X$ s.t. $|N(S)| < |S|$.

Let $S$ be a *minimal* such set. Then one has

- $|S| \leq \frac{n}{2}$
- $|N(S)| = |S| - 1$
- Every vertex in $N(S)$ is adjacent to at least $2$ vertices of $S$.

# The proof of Erdős-Rényi for PM

Recall the Erdős-Rényi theorem: Sharp threshold for PM is $\frac{\log n}{n}$.
Suppose $p > \frac{(1+\varepsilon)\log n}{n}$ and $\mathbb{G}(n, n, p)$ does not admit PM, then there exists $S \subseteq X$ s.t. $|N(S)| < |S|$.

Let $S$ be a *minimal* such set. Then one has

- $|S| \leq \frac{n}{2}$
- $|N(S)| = |S| - 1$
- Every vertex in $N(S)$ is adjacent to at least $2$ vertices of $S$.

A Union bound gives the following bound on the 'error' probability:

$$\sum_{|S|=1}^{n/2} \binom{n}{|S|} \binom{n}{|S|-1} (1-p)^{|S| \cdot (n-|S|+1)} \left( \binom{|S|}{2} \cdot p^2 \right)^{|S|-1}$$

# The proof of Erdős-Rényi for PM

Recall the Erdős-Rényi theorem: Sharp threshold for PM is $\frac{\log n}{n}$.
Suppose $p > \frac{(1+\varepsilon)\log n}{n}$ and $\mathbb{G}(n, n, p)$ does not admit PM, then there exists $S \subseteq X$ s.t. $|N(S)| < |S|$.

Let $S$ be a *minimal* such set. Then one has

- $|S| \le \frac{n}{2}$
- $|N(S)| = |S| - 1$
- Every vertex in $N(S)$ is adjacent to at least $2$ vertices of $S$.

A Union bound gives the following bound on the 'error' probability:

$$\sum_{|S|=1}^{n/2} \binom{n}{|S|} \binom{n}{|S|-1} (1-p)^{|S|\cdot(n-|S|+1)} \left( \binom{|S|}{2} \cdot p^2 \right)^{|S|-1}$$

The large amount of 'structure' revealed by considering the *minimal* violating set was critical!

# Completing the Proof: Extra 'structure'

### Fact

*If* $G(X,Y)$ *has NMP, then* $G(Y,X)$ *also has NMP, i.e., for any* $T \subseteq Y, |N_X(T)| \geq \frac{k}{n}|T|$.

# Completing the Proof: Extra 'structure'

## Fact

*If $G(X,Y)$ has NMP, then $G(Y,X)$ also has NMP, i.e., for any $T \subseteq Y, |N_X(T)| \geq \frac{k}{n}|T|$.*

## Lemma

*Suppose $G(X,Y)$ with $|X| = k \leq n = |Y|$ does **not** have NMP. Then either there exists*

- *$S \subset X$ that violates NMP for $G(X,Y)$ with $|S| \leq \frac{k}{2}$, or*
- *$T \subset Y$ that violates NMP for $G(Y,X)$ with $|T| < \frac{n}{2} + \frac{n}{k}$.*

# Completing the Proof: Extra 'structure'

**Fact**

*If $G(X,Y)$ has NMP, then $G(Y,X)$ also has NMP, i.e., for any $T \subseteq Y, |N_X(T)| \geq \frac{k}{n}|T|$.*

**Lemma**

*Suppose $G(X,Y)$ with $|X| = k \leq n = |Y|$ does **not** have NMP. Then either there exists*

- *$S \subset X$ that violates NMP for $G(X,Y)$ with $|S| \leq \frac{k}{2}$, or*
- *$T \subset Y$ that violates NMP for $G(Y,X)$ with $|T| < \frac{n}{2} + \frac{n}{k}$.*

**Fact**

*Suppose $p > \frac{(1+\varepsilon)\log n}{k}$. For any fixed $r \in \mathbb{N}$, $d(x) \geq r$ for all $x \in X$ and $d(y) \geq r$ for all $y \in Y$ whp.*

# Proof of NMP-approximability

$G(X,Y)$ is T-pseudorandom with parameters $(p,\varepsilon)$ with $p \geq \frac{\omega(k)}{k}$.

Suppose $\frac{n}{k} = \frac{L}{\ell}$ with $\gcd(\ell, L) = 1$ and $\ell, L = O(1)$.

# Proof of NMP-approximability

$G(X, Y)$ is T-pseudorandom with parameters $(p, \varepsilon)$ with $p \geq \frac{\omega(k)}{k}$.

Suppose $\frac{n}{k} = \frac{L}{\ell}$ with $\gcd(\ell, L) = 1$ and $\ell, L = O(1)$.

Main difficulties:

- Unlike in the case when $k \mid n$ there is no canonical structure that certifies NMP.
- When $n/k \pmod 1$ is 'large' then a cloning argument fails spectacularly.

# Proof of NMP-approximability

$G(X,Y)$ is T-pseudorandom with parameters $(p,\varepsilon)$ with $p \geq \frac{\omega(k)}{k}$.

Suppose $\frac{n}{k} = \frac{L}{\ell}$ with $\gcd(\ell, L) = 1$ and $\ell, L = O(1)$.

Main difficulties:

- Unlike in the case when $k \mid n$ there is no canonical structure that certifies NMP.
- When $n/k \pmod 1$ is 'large' then a cloning argument fails spectacularly.

New Idea: A decomposition type theorem, i.e., want a spanning subgraph of $G$ which certifies NMP (Especially when $n/k \ni \mathbb{N}$).

# New structure: Euclidean trees $T_{\ell,L}$

Suppose $\ell < L$. Consider the Euclidean algorithm on the pair $(\ell, L)$ as follows.

# New structure: Euclidean trees $T_{\ell,L}$

Suppose $\ell < L$. Consider the Euclidean algorithm on the pair $(\ell, L)$ as follows.

$$L = q_m \ell + r_{m-1}, \qquad 0 < r_{m-1} < r_m = \ell,$$

# New structure: Euclidean trees $T_{\ell,L}$

Suppose $\ell < L$. Consider the Euclidean algorithm on the pair $(\ell, L)$ as follows.

$$
\begin{aligned}
L &= q_m \ell + r_{m-1}, && 0 < r_{m-1} < r_m = \ell, \\
\ell &= q_{m-1} r_{m-1} + r_{m-2}, && 0 < r_{m-2} < r_{m-1}, \\
\cdots &= \cdots && \\
r_3 &= q_2 r_2 + r_1, && 0 < r_1 < r_2, \\
r_2 &= q_1 r_1, && r_1 = 1.
\end{aligned}
$$

# New structure: Euclidean trees $T_{\ell,L}$

Suppose $\ell < L$. Consider the Euclidean algorithm on the pair $(\ell, L)$ as follows.

$$
\begin{aligned}
L &= q_m \ell + r_{m-1}, && 0 < r_{m-1} < r_m = \ell, \\
\ell &= q_{m-1} r_{m-1} + r_{m-2}, && 0 < r_{m-2} < r_{m-1}, \\
\cdots &= \cdots && \\
r_3 &= q_2 r_2 + r_1, && 0 < r_1 < r_2, \\
r_2 &= q_1 r_1, && r_1 = 1.
\end{aligned}
$$

If we set $r_{m+1} = L, r_m = \ell, r_0 = 0$, then we may write

$$r_{i+1} = q_i r_i + r_{i-1} \text{ for } 1 \leq i \leq m.$$

# New structure: Euclidean trees $T_{\ell,L}$

Suppose $\ell < L$. Consider the Euclidean algorithm on the pair $(\ell, L)$ as follows.

$$
\begin{aligned}
L &= q_m \ell + r_{m-1}, && 0 < r_{m-1} < r_m = \ell, \\
\ell &= q_{m-1} r_{m-1} + r_{m-2}, && 0 < r_{m-2} < r_{m-1}, \\
\cdots &= \cdots \\
r_3 &= q_2 r_2 + r_1, && 0 < r_1 < r_2, \\
r_2 &= q_1 r_1, && r_1 = 1.
\end{aligned}
$$

If we set $r_{m+1} = L, r_m = \ell, r_0 = 0$, then we may write

$$r_{i+1} = q_i r_i + r_{i-1} \text{ for } 1 \leq i \leq m.$$

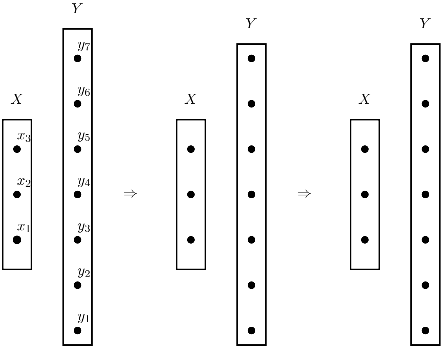
Construct a family of trees called Euclidean trees in $m$ steps: In step $i$ (if even) add an $X$ $q_i$-thrill from the 'first' left $r_i$ vertices $\{x_1, \ldots, x_{r_i}\}$ *into* $\{y_{r_{(i-1)}+1}, \ldots, y_{r_{(i+1)}}\}$.

# Illustrative example: $T_{3,7}$

$$\begin{aligned} 7 &= 2 \cdot 3 + 1 \\ 3 &= 3 \cdot 1. \end{aligned}$$

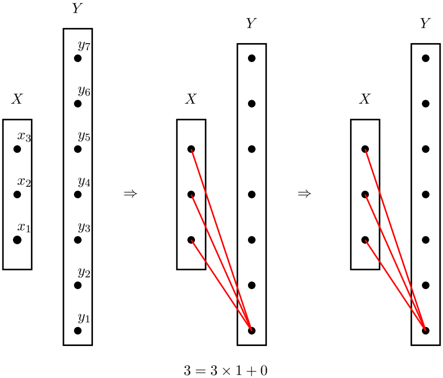
# Illustrative example: $T_{3,7}$

$$\begin{aligned} 7 &= 2 \cdot 3 + 1 \\ 3 &= 3 \cdot 1. \end{aligned}$$



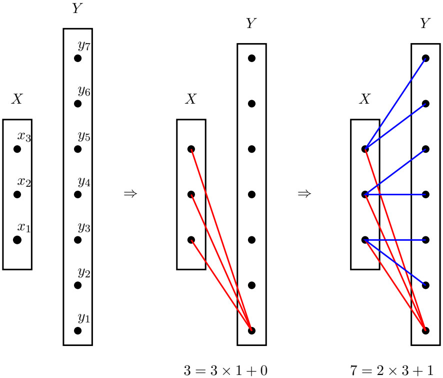
# Illustrative example: $T_{3,7}$

$$\begin{aligned} 7 &= 2 \cdot 3 + 1 \\ 3 &= 3 \cdot 1. \end{aligned}$$



# Illustrative example: $T_{3,7}$

$$\begin{aligned} 7 &= 2 \cdot 3 + 1 \\ 3 &= 3 \cdot 1. \end{aligned}$$

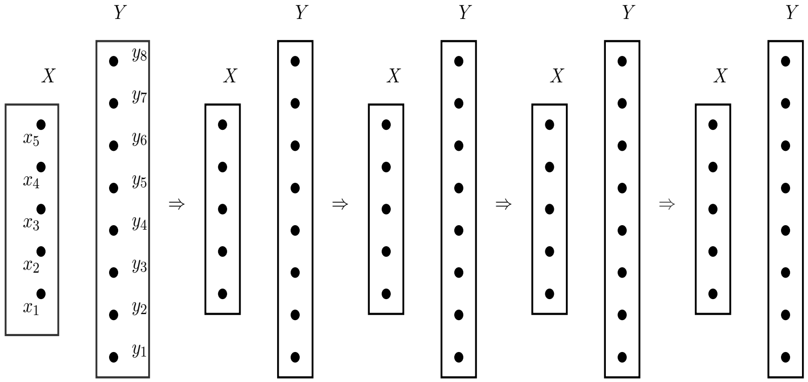


# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.

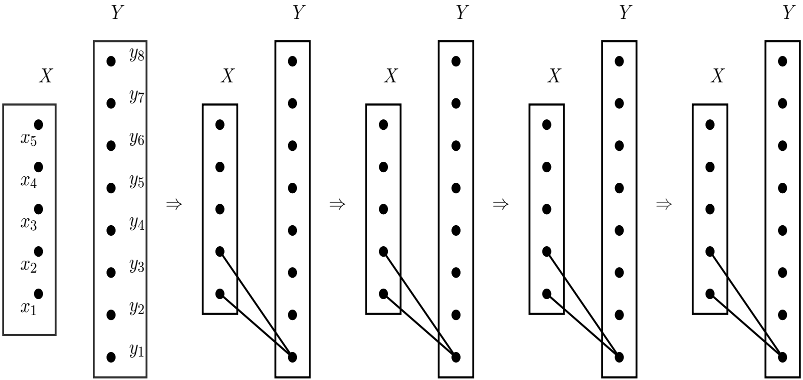
# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.



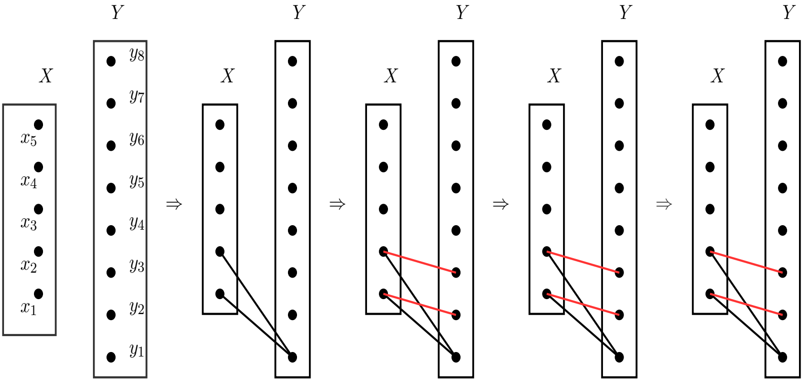
# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.



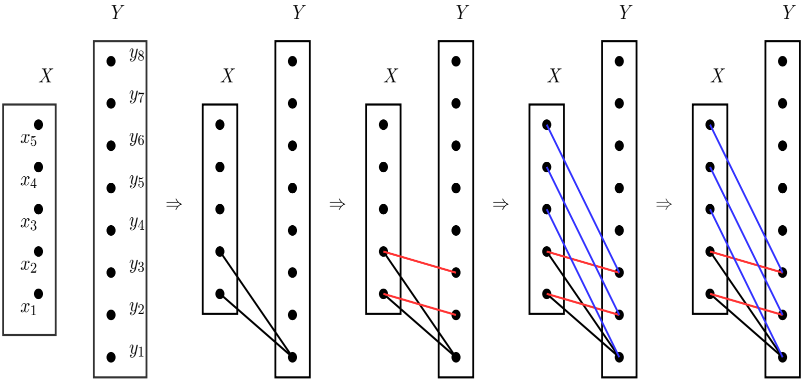
# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.



# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.



# Another Example: $T_{5,8}$

The Euclidean algorithm gives $m = 4$, $(r_2, r_3, r_4, r_5) = (2, 3, 5, 8)$, $(q_1, q_2, q_3, q_4) = (2, 1, 1, 1)$.

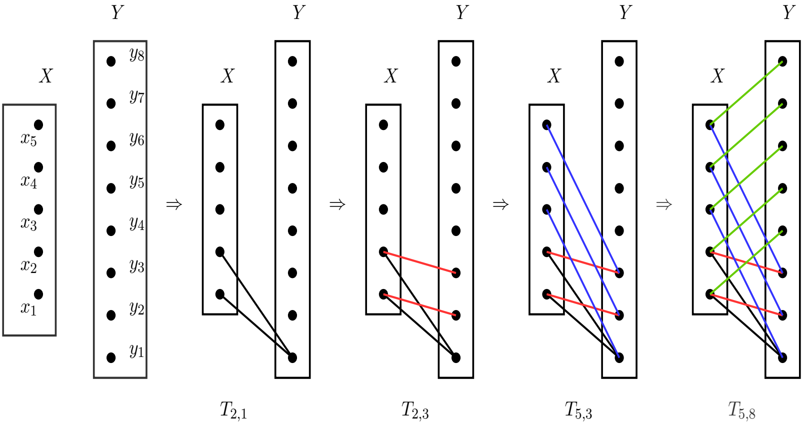



Figure: The Euclidean $(5, 8)$-tree process. $T_{5,8}$ evolves as $T_{2,1} \Rightarrow T_{2,3} \Rightarrow T_{5,3} \Rightarrow T_{5,8}$.

# Proof Outline of NMP-Approximability

**Lemma**

*Euclidean Trees have NMP.*

# Proof Outline of NMP-Approximability

**Lemma**

*Euclidean Trees have NMP.*

Write $\frac{n}{k} = \frac{L}{\ell}$ with $(\ell, L) = 1$.

Partition $X = X_1 \sqcup \cdots \sqcup X_\ell$ and $Y = Y_1 \sqcup \cdots \sqcup Y_L$. Replicate the Euclidean $(\ell, L)$-process, with the vertices $x_i, y_j$ replaced by the blocks $X_i, Y_j$. The following lemma is key:

# Proof Outline of NMP-Approximability

**Lemma**

*Euclidean Trees have NMP.*

Write $\frac{n}{k} = \frac{L}{\ell}$ with $(\ell, L) = 1$.

Partition $X = X_1 \sqcup \cdots \sqcup X_\ell$ and $Y = Y_1 \sqcup \cdots \sqcup Y_L$. Replicate the Euclidean $(\ell, L)$-process, with the vertices $x_i, y_j$ replaced by the blocks $X_i, Y_j$. The following lemma is key:

**Lemma (Informal)**

*Suppose $q \in \mathbb{N}$ and $U \subseteq X$ and $V \subseteq Y$ both are large enough subsets such that $|V| = q|U|$. Then there exist 'small' subsets $A \subseteq U, B \subseteq V$ such that $G(U \setminus A, V \setminus B)$ is spanned by an $X$ $q$-thrill.*

# Proof of NMP-Approximability: Our main structural theorem

**Theorem**

*Suppose $G(X, Y)$ is T-pseudorandom with parameters $(p, \varepsilon)$ with $p \geq \frac{\omega(k)}{k}$, and suppose $k \gg 0$. Suppose $\frac{n}{k} = \frac{L}{\ell}$ with $(\ell, L) = 1$ and $\ell, L = O(1)$. Then there exist sets $\mathcal{X} \subset X, \mathcal{Y} \subset Y$ s.t. $|\mathcal{X}| \leq O(\varepsilon)k, |\mathcal{Y}| \leq O(\varepsilon)n$ s.t. $G(X \setminus \mathcal{X}, Y \setminus \mathcal{Y})$ is spanned by vertex disjoint copies of $T_{\ell,L}$.*

# Proof of NMP-Approximability: Our main structural theorem

**Theorem**

*Suppose $G(X, Y)$ is T-pseudorandom with parameters $(p, \varepsilon)$ with $p \geq \frac{\omega(k)}{k}$, and suppose $k \gg 0$. Suppose $\frac{n}{k} = \frac{L}{\ell}$ with $(\ell, L) = 1$ and $\ell, L = O(1)$. Then there exist sets $\mathcal{X} \subset X, \mathcal{Y} \subset Y$ s.t. $|\mathcal{X}| \leq O(\varepsilon)k, |\mathcal{Y}| \leq O(\varepsilon)n$ s.t. $G(X \setminus \mathcal{X}, Y \setminus \mathcal{Y})$ is spanned by vertex disjoint copies of $T_{\ell,L}$.*

In general tweak $(n, k)$ to a 'close-enough' $(n', k')$ such that $\frac{n'}{k'} = \frac{L}{\ell}$ with $\ell, L = O_\varepsilon(1)$.

# An approx. version of problem 2

The proof of NMP-approximability works in more general pseudorandom paradigms as well.

# An approx. version of problem 2

The proof of NMP-approximability works in more general pseudorandom paradigms as well.

## Theorem (N.Balachandran, D. K., 2019)

*Suppose* $X, Y \subset \mathbb{F}_q, |Y| = 10|X| \geq q/100$. *Then for any multiplicative subgroup* $H \subset \mathbb{F}_q^*$ *of size at least* $q^{1/2+\varepsilon}$, *one can delete at most* $O(q^{1-\varepsilon})$ *elements from both* $X, Y$ *s.t. in the remaining sets, problem 2 has an affirmative answer for* $H$ *as well (in place of quadratic residues).*

# An approx. version of problem 2

The proof of NMP-approximability works in more general pseudorandom paradigms as well.

**Theorem (N.Balachandran, D. K., 2019)**

*Suppose $X, Y \subset \mathbb{F}_q, |Y| = 10|X| \geq q/100$. Then for any multiplicative subgroup $H \subset \mathbb{F}_q^*$ of size at least $q^{1/2+\varepsilon}$, one can delete at most $O(q^{1-\varepsilon})$ elements from both $X, Y$ s.t. in the remaining sets, problem 2 has an affirmative answer for $H$ as well (in place of quadratic residues).*

**Fact**

*The corresponding bipartite graph $\Gamma_q(H)$ is $(q, |H|, \sqrt{q})$-pseudorandom.*

# THANK YOU!

# Induction Step Outline

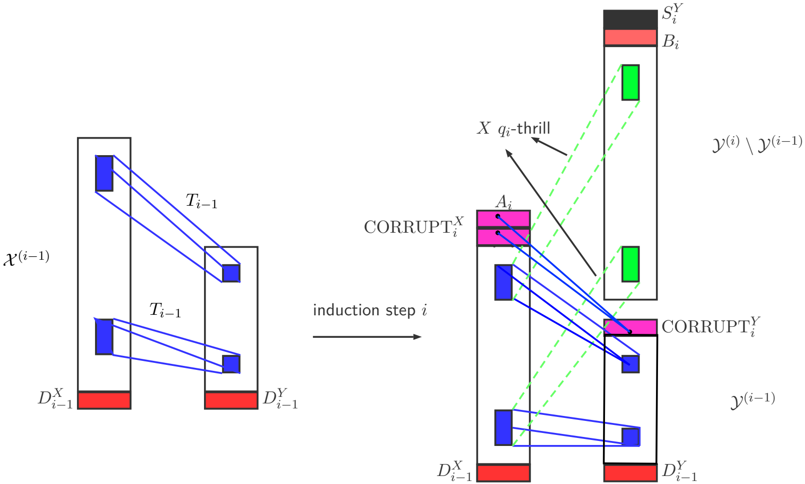



Figure: Induction step