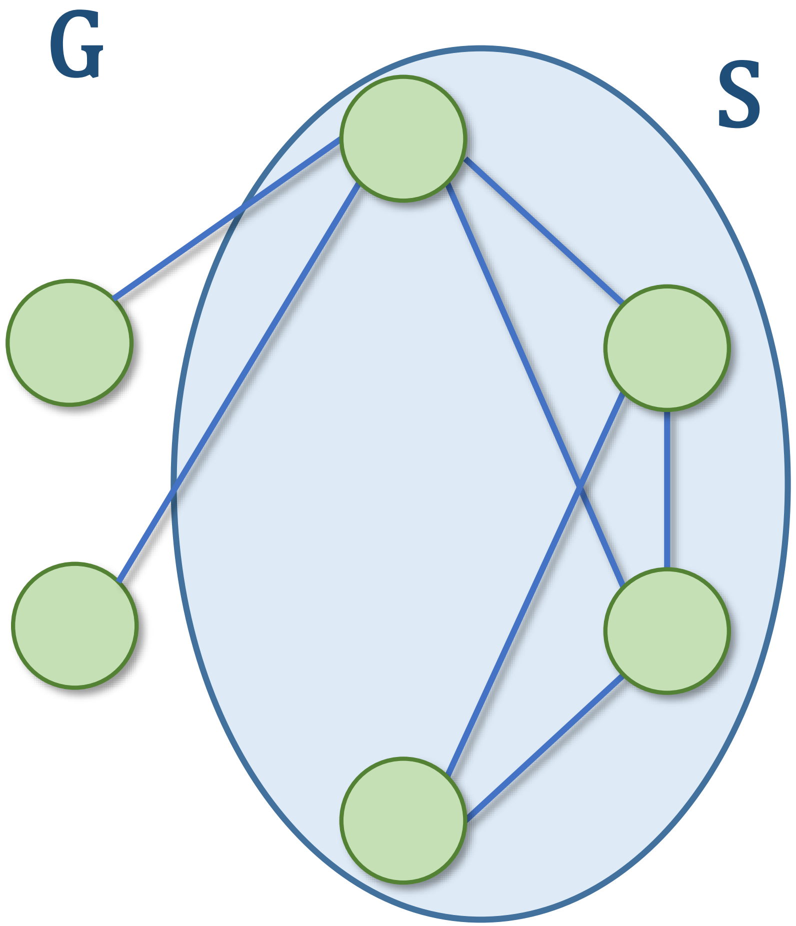
# ALMOST-POLYNOMIAL RATIO ETH-HARDNESS OF APPROXIMATING DENSEST K-SUBGRAPH

Pasin Manurangsi
UC Berkeley

# DENSEST K-SUBGRAPH (DKS)




## INPUT

- An undirected graph $G = (V, E)$
- A positive integer $k$

## GOAL

- Find a subset of vertices $S \subseteq V$ of size $k$ that maximizes the number of edges within $S$

# DENSEST K-SUBGRAPH: WHAT IS KNOWN?

## HARDNESS RESULTS

| | | |
|---|---|---|
| *[Feige02]* | Random 3SAT Hypothesis | Some constant |
| *[Khot04]* | $NP \not\subseteq \cap_{\varepsilon>0} BPTIME(2^{n^{\varepsilon}})$ | Some constant |
| *[Raghavendra-Steurer10]* | Small Set Expansion Hypothesis | Every constant |
| *[Alon-Arora-Manokaran-Moshkovitz-Weinstein11]* | Planted Clique Hypothesis | $2^{\Omega(\log^{2/3} n)}$ |
| *[Braverman-Ko-Rubinstein-Weinstein17]* | ETH | Some constant |
| **Conjecture** *[BCCFV10, AAMMW11,…]* | NP-hard | $n^c$ for some $c > 0$ |

## APPROXIMATION ALGORITHMS

(Note: $n = |V|$)

| | |
|---|---|
| *[Kortsarz-Peleg93]* | $O(n^{0.385\ldots})$ |
| *[Feige-Kortsarz-Peleg01]* | $O(n^{0.316\ldots})$ |
| *[Bhaskara-Charikar-Chlamtac-Feige-Vijayaraghavan10]* | $O(n^{0.25+\varepsilon})$ |

# DENSEST K-SUBGRAPH: OUR RESULTS

## HARDNESS RESULTS

| | | |
|---|---|---|
| *[Feige02]* | Random 3SAT Hypothesis | Some constant |
| *[Khot04]* | $NP \not\subseteq \cap_{\varepsilon>0} BPTIME(2^{n^{\varepsilon}})$ | Some constant |
| *[Raghavendra-Steurer10]* | Small Set Expansion Hypothesis | Every constant |
| *[Alon-Arora-Manokaran-Moshkovitz-Weinstein11]* | Planted Clique Hypothesis | $2^{\Omega(\log^{2/3} n)}$ |
| *[Braverman-Ko-Rubinstein-Weinstein17]* | ETH | Some constant |
| **Conjecture** *[BCCFV10, AAMMW11,…]* | NP-hard | $n^c$ for some $c > 0$ |

## OUR RESULTS

| | |
|---|---|
| ETH | $n^{1/\text{polyloglog}\, n}$ |
| Gap-ETH | $n^{o(1)}$ |

# DENSEST K-SUBGRAPH: OUR RESULTS

Holds even for *"perfect completeness"*

**THEOREM**
Assuming ETH, no polynomial time algorithm can, on every input graph $G$ that contains $k$-clique, find a $n^{-1/\text{polyloglog}\, n}$-dense $k$-subgraph.

Note: $\text{Den}(S) = |E(S)|/\binom{k}{2}$

This regime is easy! *[Feige-Seltser97]*

- There is an $(1+\varepsilon)$-approximation algorithm with running time $n^{O(\log n/\varepsilon)}$
- There is an $n^{\varepsilon}$-approximation algorithm with running time $n^{O(1/\varepsilon)}$

**OUR RESULTS**

| | |
|---|---|
| ETH | $n^{1/\text{polyloglog}\, n}$ |
| Gap-ETH | $n^{o(1)}$ |

**THEOREM** Assuming Gap-ETH, no $n^{\tilde{O}(1/\varepsilon^{1/3})}$-time algorithm can, on every $G$ that contains $k$-clique, find a $n^{-\varepsilon}$-dense $k$-subgraph.

# THE HYPOTHESES: ETH & GAP-ETH

## EXPONENTIAL TIME HYPOTHESIS (ETH)

*[Impagliazzo-Paturi99]*

No $2^{o(n)}$-time algorithm can solve 3SAT.

($n$ is the number of variables.)

## GAP EXPONENTIAL TIME HYPOTHESIS (GAP-ETH)

*[Dinur16, M-Raghavendra16]*

No $2^{o(n)}$-time algorithm can distinguish between

(i) a satisfiable 3CNF formula, and,

(ii) a 3CNF formula which is not even 0.99-satisfiable.

# ETH/GAP-ETH & HARDNESS OF APPROXIMATION

## THE BIRTHDAY REPETITION FRAMEWORK

*[Aaronson-Impagliazzo-Moshkovitz14]*

**3SAT INSTANCE $\Phi$**
- **Size:** $n$
- **YES:** $\mathrm{val}(\Phi) = 1$
- **NO:** $\mathrm{val}(\Phi) < 1$
- **Lower Bound:** $2^{\Omega(n)}$

→ PCP Thm *[Dinur07]* →

**3SAT INSTANCE $\Phi'$**
- **Size:** $n' = n \,\mathrm{polylog}\, n$
- **YES:** $\mathrm{val}(\Phi') = 1$
- **NO:** $\mathrm{val}(\Phi') < 0.99$
- **Lower Bound:** $2^{\Omega(n'/\mathrm{polylog}\, n')}$

→

**DKS INSTANCE $(G, k)$**
- **Size:** $N = 2^{n'/\mathrm{polylog}\, n'} = 2^{o(n)}$
- **YES:** k-Clique
- **NO:** Every k-subgraph has density $< N^{-1/\mathrm{polyloglog}\, N}$
- **Lower Bound:** $N^{\omega(1)}$

Not necessary, assuming Gap-ETH

Used in other works before and since:
- **Densest $k$-Subgraph** *[BKRW17]*
- **$\varepsilon$-best $\varepsilon$-Nash** *[Braverman-Ko-Weinstein15]*
- **Community Detection** *[Rubinstein17]*, etc.

# THE REDUCTION

Parameter: $q$ $(1 \leq q \leq n)$

Representative Case:
- $q = n / \text{polylog}\, n$
- $N = \exp(n/\text{polylog}\, n)$
- Gap: $\exp(n/\text{polylog}\, n) = N^{1/\text{polyloglog}\, N}$

**3SAT $\Phi$**

**Size:** $n$

**YES:** $\text{val}(\Phi) = 1$

**NO:** $\text{val}(\Phi) < 0.99$

**DKS INSTANCE** $(G = (V, E), k)$

**Size:** $N = \binom{n}{q} 2^q$

**YES:** $k$-Clique

**NO:** Every $k$-subgraph has density $< \exp(-\Omega(q^4/n^3))$

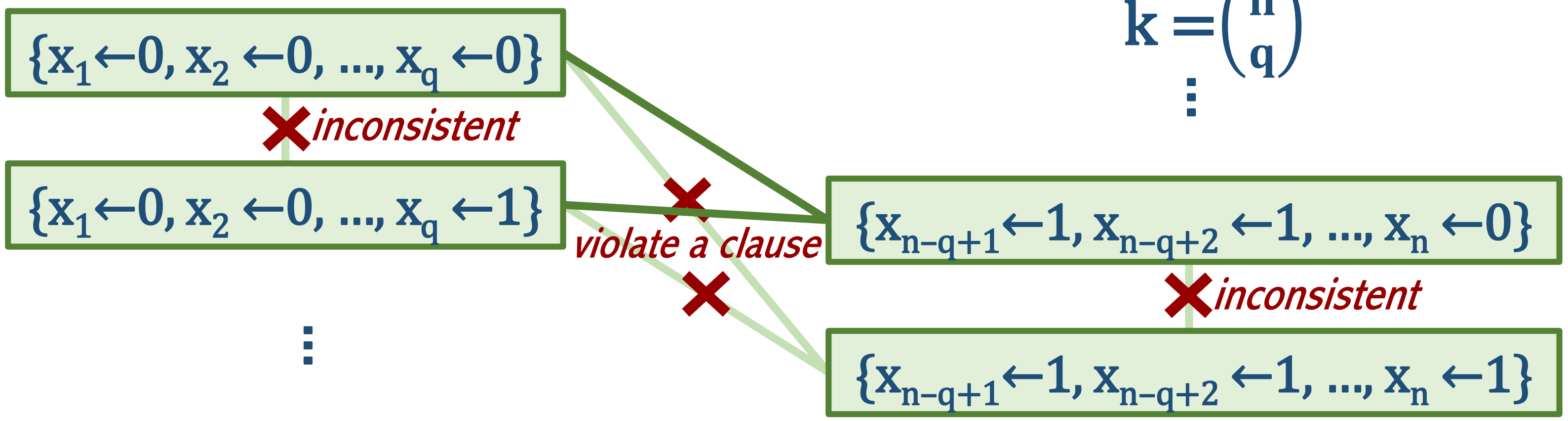
**Just pick all partial assignments to a satisfying assignment!**

$(x_1 \vee \sim x_3 \vee x_5)$
$\wedge (x_1 \vee x_2 \vee \sim x_n)$
$\wedge \ldots$

$V = \{$all partial assignments to $q$ variables$\}$

$E = \{(u, v)$ that are *consistent* and *do not violate any clauses*$\}$

$k = \binom{n}{q}$



# SOUNDNESS PROOF

## Goal

$\text{val}(\Phi) < 0.99 \Rightarrow$ every k-subgraph of G has density $< \exp(-\Omega(q^4/n^3))$

## A Less Ambitious Goal

$\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $k \cdot \exp(-\Omega(q^4/n^3))$

# SOUNDNESS PROOF

## Goal

$\text{val}(\Phi) < 0.99 \Rightarrow$ every k-subgraph of G has density $< \exp(-\Omega(q^4/n^3))$

## A Less Ambitious Goal

$\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $k \cdot \exp(-\Omega(q^2/n))$

## An Even Less Ambitious Goal

G does not contain a biclique of size larger than k

# SOUNDNESS PROOF (CONT.)

**Goal:** G does not contain a biclique of size larger than $\binom{n}{q}$

Immediate from the observations!

L

$\{x_1 \leftarrow 0, x_2 \leftarrow 0, \ldots, x_q \leftarrow 0\}$

$\{x_1 \leftarrow 0, x_2 \leftarrow 1, \ldots, x_q \leftarrow 0\}$

⋮

R

$\{x_{n-q+1} \leftarrow 1, x_{n-q+2} \leftarrow 1, \ldots, x_n \leftarrow 0\}$

$\{x_{n-q+1} \leftarrow 1, x_{n-q+2} \leftarrow 1, \ldots, x_n \leftarrow 1\}$

⋮

"Flattening"

| $A(L) = \{x_1 \leftarrow 0, x_2 \leftarrow 0, x_2 \leftarrow 1, \ldots\}$ | $A(R) = \{\ldots, x_n \leftarrow 0, x_n \leftarrow 1\}$ |
|---|---|
| $x_i \leftarrow 0, x_i \leftarrow 1$ | - |
| $x_i \leftarrow 0$ | $x_i \leftarrow 0$ |
| $x_i \leftarrow 0$ | - |
| - | - |

**Observation 1**

$L \subseteq \binom{A(L)}{q}$ and

$R \subseteq \binom{A(R)}{q}$

**Observation 2** $|A(L)| + |A(R)| \leq 2n$

# SOUNDNESS PROOF: THE GOOD, THE BAD AND THE UGLY

**Goal:** $\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $\exp(-\Omega(q^2/n))\binom{n}{q}$

**L**

$\{x_1 \leftarrow 0, x_2 \leftarrow 0, \ldots, x_q \leftarrow 0\}$

$\{x_1 \leftarrow 0, x_2 \leftarrow 1, \ldots, x_q \leftarrow 0\}$

$\vdots$

**R**

$\{x_{n-q+1} \leftarrow 1, x_{n-q+2} \leftarrow 1, \ldots, x_n \leftarrow 0\}$

$\{x_{n-q+1} \leftarrow 1, x_{n-q+2} \leftarrow 1, \ldots, x_n \leftarrow 1\}$

$\vdots$

"Flattening"

| $A(L) = \{x_1 \leftarrow 0, x_2 \leftarrow 0, x_2 \leftarrow 1, \ldots\}$ | $A(R) = \{\ldots, x_n \leftarrow 0, x_n \leftarrow 1\}$ | |
|---|---|---|
| $x_i \leftarrow 0, x_i \leftarrow 1$ | - | "bad" variable |
| $x_i \leftarrow 0$ | $x_i \leftarrow 0$ | "good" variable |
| $x_i \leftarrow 0$<br>- | -<br>- | "ugly" variable |

**Observation 1**

$L \subseteq \binom{A(L)}{q}$ **and**

$R \subseteq \binom{A(R)}{q}$

**Observation 2** $|A(L)| + |A(R)| \leq 2n$

# SOUNDNESS PROOF: THE UGLY

**Goal:** $\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $\exp(-\Omega(q^2/n))\binom{n}{q}$

$|A(L)| + |A(R)| \leq n$

**Extreme Case 1** every variable is ugly

**Observation 1** $L \subseteq \binom{A(L)}{q}$ and $R \subseteq \binom{A(R)}{q}$

**Observation 2** $|A(L)| + |A(R)| \leq 2n$

| A(L) | A(R) | |
|---|---|---|
| $x_i \leftarrow 0$, $x_i \leftarrow 1$ | - | "bad" variable |
| $x_i \leftarrow 0$ | $x_i \leftarrow 0$ | "good" variable |
| $x_i \leftarrow 0$<br>- | -<br>- | "ugly" variable |

# SOUNDNESS PROOF: THE BAD

**Goal:** $\mathrm{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $\exp(-\Omega(q^2/n))\binom{n}{q}$

**Extreme Case 2** every variable is bad

Not valid vertices! →

**Claim**
Most elements of $\binom{A(L)}{q}$ contains both $x_i \leftarrow 0$ and $x_i \leftarrow 1$ for some $i$!

**Proof of Claim**

**"BIRTHDAY PARADOX"**
In expectation, a random element of $\binom{A(L)}{q}$ contains $\Omega(q^2/n)$ such $i$'s!

→ Only $\exp(-\Omega(q^2/n))$ fraction of $\binom{A(L)}{q}$ are actual vertices

**Observation 1** $L \subseteq \binom{A(L)}{q}$ and $R \subseteq \binom{A(R)}{q}$

**Observation 2** $|A(L)| + |A(R)| \leq 2n$

| A(L) | A(R) | |
|---|---|---|
| $x_i \leftarrow 0, x_i \leftarrow 1$ | - | "bad" variable |
| $x_i \leftarrow 0$ | $x_i \leftarrow 0$ | "good" variable |

# SOUNDNESS PROOF: THE GOOD

**Goal:** $\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size larger than $\exp(-\Omega(q^2/n))\binom{n}{q}$

## Extreme Case 3 every variable is good

"good" variable

| A(L) | A(R) |
|---|---|
| $x_i \leftarrow 0$ | $x_i \leftarrow 0$ |

⇓

$A(L) = A(R)$ is an actual assignment!

⇓

$0.01m$ clauses unsatisfied

Suppose a clause $(x_i \vee \sim x_j \vee x_k)$ is unsatisfied.

**Claim:** $x_i, x_j$ cannot appear together in any vertex in L

L: $\{\ldots, x_i \leftarrow 0, x_j \leftarrow 1, \ldots\}$ ⋮

✗ *violate clause*

R: $\{\ldots, x_k \leftarrow 0, \ldots\}$ ⋮

$(x_i, x_j)$ is a ***"prohibited pair"*** that can't appear together in any vertex of L

→ There are at least $0.01m$ prohibited pairs

→ **"BIRTHDAY PARADOX"** Only $\exp(-\Omega(q^2/n))$ fraction of $\binom{A(L)}{q}$ can be vertices of L

# SOUNDNESS PROOF: FROM BICLIQUE TO DENSE SUBGRAPHS

**What we have proved:**

$\text{val}(\Phi) < 0.99 \Rightarrow$ G does not contain a biclique of size $\binom{n}{q} \exp(-\Omega(l^2/n))$

?

**What we actually want:**

$\text{val}(\Phi) < 0.99 \Rightarrow$ every $\binom{n}{q}$-subgraph of G has density $< \exp(-\Omega(q^4/n^3))$

# WHY BICLIQUE? Aviad Rubinstein told me to do so…

Different from clique: $t$-clique free graph can be $(1\text{-}1/(t-1))$-dense!

**THEOREM** *[Kővári-Sós-Turán54]*

Every $t$-biclique free graph on $N$ vertices is $O(N^{-1/t})$-dense.

Need $t \leq \log N$ to even get constant bound

Our bound is not (nearly) enough…

**THEOREM** *[Alon02]*

Every graph on $N$ vertices that contains $< \varepsilon^{t^2} N^{2t}$ copies of $t$-biclique is $O(\varepsilon)$-dense.

Only need to prove that G does not contain many bicliques.

# SOUNDNESS PROOF: BOUNDING NUMBER OF BICLIQUES

**What we actually showed:**

For each flattening (A(L), A(R)),
L is a subset of a small subset of $\binom{A(L)}{q}$ and
R is a subset of a small subset of $\binom{A(R)}{q}$

## Bounding number of $t$-bicliques

Fix (A(L), A(R)). Only a small number of choices for each vertex in L. Same for R.

→ Sum over all choices of (A(L), A(R))

→ With appropriate t, Alon's Theorem gives the desired bound for density of $\binom{n}{q}$-subgraph!

# A DIFFERENT PERSPECTIVE: PARAMETERIZED INAPPROX. OF BICLIQUE

**THEOREM** *[Kővári-Sós-Turán54]* Every $t$-biclique free graph on $n$ vertices is $O(n^{-1/t})$-dense.

**THEOREM** *[Alon02]* Every graph on $N$ vertices that contains $< \varepsilon^{2t^2} N^{2t}$ copies of $t$-biclique is $O(\varepsilon)$-dense.

→ Subsample each vertex w.p. $p = 0.1/(\varepsilon^t N)$ →

$t$-biclique free graph on $O(1/\varepsilon^t)$ vertices

→ Apply KST Theorem →

The graph must be $O(\varepsilon)$-dense

## OUR REDUCTION

Parameterized Hardness of Approximating Biclique!

**3SAT INSTANCE $\Phi$**

**Size:** $n$

**YES:** $\text{val}(\Phi') = 1$

**NO:** $\text{val}(\Phi') < 0.99$

→

**DKS INSTANCE $(G, k)$**

**Size:** $N = 2^{o(n)}$

**YES:** $k$-Clique

**NO:** $G$ contains $\ll k^{2t}$ copies of $t$-biclique for $t \ll k$

→ Subsample each vertex w.p. $p$ →

**DKS INSTANCE $(G', pk)$**

**Size:** $N' = 2^{o(n)}$

**YES:** $(pk)$-Clique

**NO:** $G$ is $t$-biclique free (where $t \ll pk$)

# CONCLUSION

- Sub-exponential time reduction from gap 3SAT to DENSEST K-SUBGRAPH
- Assuming ETH, gives $n^{1/\text{polyloglog}\, n}$-ratio inapproximability for DKS
- Assuming Gap-ETH, gives $n^{o(1)}$-ratio inapproximability for DKS
- Assuming Gap-ETH, gives parameterized inapproximability for biclique

# OPEN QUESTIONS

- NP-hardness of approximation of DENSEST K-SUBGRAPH?
- Polynomial ratio hardness of approximation of DENSEST K-SUBGRAPH (assuming ETH/Gap-ETH)?

# THANK YOU! QUESTIONS?