Threshold Graphs for Hardness of Approximation

Reina Itakura
Mentor: Prof. Karthik C. S.

DIMACS, 7.16.25

Maximum Inner Product

The Problem (decision ver.):

- Input: Points $P \subseteq \{0,1\}^d$, $d = O(\log n)$
- Output: $\exists a, b \in P$ s.t. $\langle a, b \rangle \geq s$?

$$\langle a, b \rangle = \sum_{i=1}^{d} a_i \cdot b_i$$

- Trivial Solution: $O(n^2 d)$
- Solution for $d = O(\log n)$: $O(n^2)$ [AWY15]
- Applications to machine learning.

Exact Solving

Definition (Strong Exponential Time Hypothesis (SETH))

For every $\epsilon > 0$, $\exists k$ s.t. k-SAT cannot be solved in time $O(2^{(1-\epsilon)n})$ even by a randomized algorithm.

Assuming SETH,

- MaxIP cannot be solved in time $O(n^{2-\epsilon}\text{poly}(d))$ for $d = \omega(\log(n))$

Approximate Solving

Assuming SETH, there IS an algebraic proof of hardness of approximation of MaxIP using Distributed PCP.

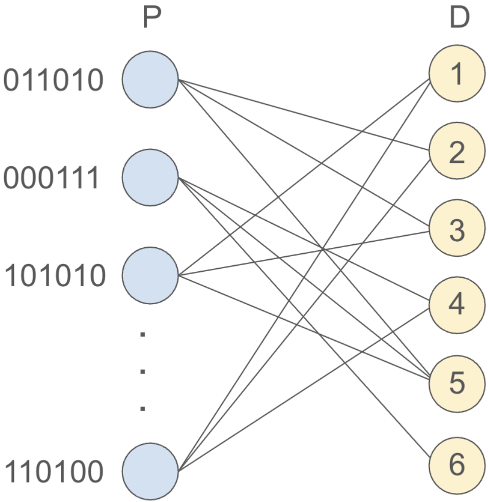
- Assuming SETH, 100-approximate MaxIP is not possible in time $O(n^{2-\epsilon}\text{poly}(d))$ $\forall \epsilon > 0$. $d = n^{o(1)}$ [ARW17]
- Assuming SETH, $\forall \delta > 0$, $\exists \epsilon > 0$ s.t. $(1+\epsilon)$-approximate MaxIP requires $O(n^{2-\delta})$ time for $d = O(\log n)$.[Rub18]

Combinatorial Proof

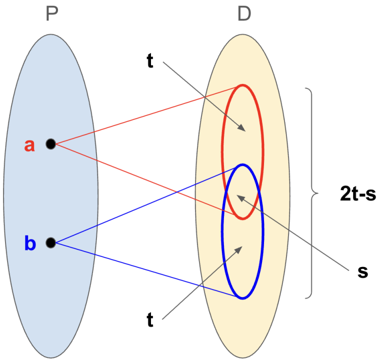
Motivation:

1. We want a combinatorial proof.
2. Want to prove for lower dimensions: $d = O(\log n)$

Threshold Graph Reduction



Threshold Graph Reduction



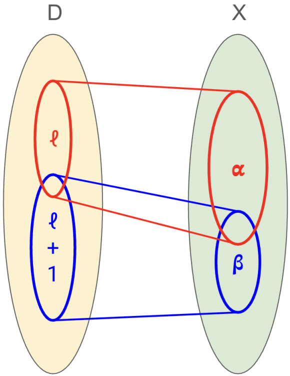
Notice:

- $\langle a, b \rangle \geq s$ implies
$$\mathcal{N}(a, b) \leq 2t - s.$$
- $\langle a, b \rangle < s$ implies
$$\mathcal{N}(a, b) \geq 2t - s + 1.$$

Threshold Graph

Magical Threshold Graph: $T = (D \sqcup X, E)$, parameters (ℓ, α, β).

1. Every ℓ nodes in D must share $\geq \alpha$ common neighbors in X.
2. Every $\ell + 1$ nodes in D share $\leq \beta$ common neighbors in X.



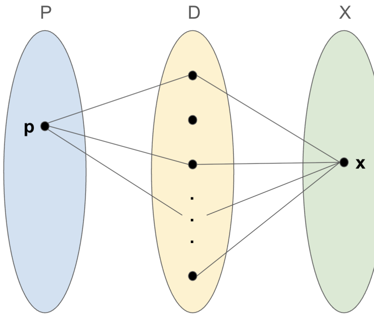
Threshold Graph Composition

Input:

1. MaxIP instance $R = (P \sqcup D, E_R)$
2. $(2t - s, \alpha, \alpha/2)$ threshold graph $T = (D \sqcup X, E_T)$

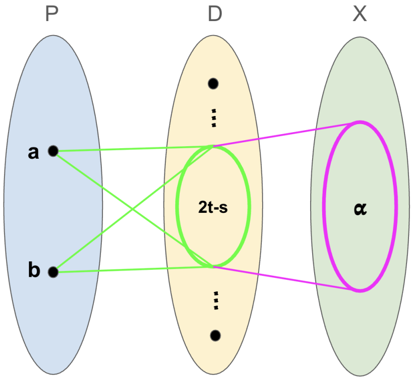
Output: 2-gap MaxIP instance $U = (P \sqcup X, E)$.

- Connect $p \in P$ and $x \in X$ if $\mathcal{N}(p) \subseteq \mathcal{N}(x)$.



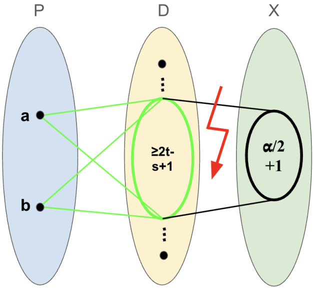
Completeness

1. Suppose $\exists \langle a, b \rangle = s$.
2. a and b have $2t - s$ **total** neighbors.
3. $2t - s$ nodes in D share $\geq \alpha$ **common** neighbors in X.
4. These α nodes are adjacent to both a and b in resulting instance.



Soundness

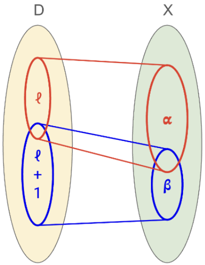
1. Suppose $\forall \langle a, b \rangle < s$.
2. Any a and b have $2t - s + 1$ **total** neighbors.
3. BWOC suppose a and b share $> \alpha/2$ nodes in final instance.
4. Each of $\alpha/2$ nodes must be adjacent to dimension nodes of a and b.
5. $\implies \geq 2t - s + 1$ nodes in D are adjacent to $> \alpha/2$ nodes. contradiction.



Issue: Don't Have!

Magical Threshold Graph: $T = (D \sqcup X, E)$, parameters (ℓ, α, β).

1. Every ℓ nodes in D must share $\geq \alpha$ common neighbors in X.
2. Every $\ell + 1$ nodes in D share $\leq \beta$ common neighbors in X.



- Don't know how to construct.
- Want to find a workaround.

1-center

The Problem:

- Input: Set of n points P.
- Output: Find point in P that minimizes max distance to other points in P.

Applications to machine learning, computational biology, etc.

Exact 1-center

Assuming the hitting set conjecture (HSC), no algorithm in time $O(n^{2-\epsilon})$ can solve discrete 1-center problem in Edit/Ulam/ℓ_p metric, $d = \omega(\log n)$[Abb+23]

Approximation

Input: Set of n points P.

Output: Decide between

- **(Yes)** $\exists p \in P$ s.t. $\max_{x \in P}\{\text{dist}(p, x)\} \leq \alpha$.
- **(No)** $\forall p \in P$, $\max_{x \in P}\{\text{dist}(p, x)\} > (1 + \epsilon)\alpha$.

However,

- Distributed PCP fails.
- Combinatorial proof (HSC) goes through with same threshold graph composition. (don't have threshold graph)

Acknowledgements

I would like to thank my mentor Karthik C. S., his graduate student Mursalin, my groupmates Gary and Mayank, and the DIMACS REU for supporting our research.

I would also like to thank the NSF for funding my research through the grant CCF-2447342.

References

[Abb+23] Amir Abboud et al. *On Complexity of 1-Center in Various Metrics*. 2023. arXiv: 2112.03222 [cs.CC]. URL: https://arxiv.org/abs/2112.03222.

[ARW17] Amir Abboud, Aviad Rubinstein, and Ryan Williams. *Distributed PCP Theorems for Hardness of Approximation in P*. 2017. arXiv: 1706.06407 [cs.CC]. URL: https://arxiv.org/abs/1706.06407.

[AWY15] Amir Abboud, Ryan Williams, and Huacheng Yu. "More applications of the polynomial method to algorithm design". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '15. San Diego, California: Society for Industrial and Applied Mathematics, 2015, pp. 218–230.

[Rub18] Aviad Rubinstein. "Hardness of approximate nearest neighbor search". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 1260–1268. ISBN: 9781450355599. DOI: 10.1145/3188745.3188916. URL: https://doi.org/10.1145/3188745.3188916.