New Bounds for the CLIQUE-GAP Problem using Graph Decomposition Theory

Vladimir Braverman, Zaoxing Liu, Tejasvam Singh, N. V. Vinodchandran, Lin F. Yang

Johns Hopkins University, University of Nebraska-Lincoln

zaoxing@jhu.edu

Aug. 24, 2015

Overview

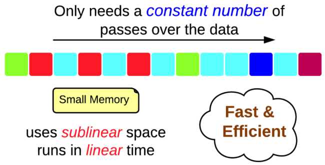
1. Introduction
   - CLIQUE-GAP Problem
   - Prior Results
   - Our Results
2. Upper Bound
   - One-pass Streaming Algorithm
3. Lower Bound
   - Reduction from multi-party set disjointness problem

CLIQUE-GAP Problem Definition

Insert-only Stream

A graph stream $G = (V, E)$ is an sequence of m edges, where $|V| = n$ and $|E| = m$.

A typical streaming algorithm is:



CLIQUE-GAP(r, s):

Given a graph stream G, integer r and s with $0 \leq s \leq r$, output "1" if G has a r-clique or "0" if G has no $(s+1)$-clique. The output can be either 0 or 1 if the size of the max-clique $w(G)$ is in $[s+1, r]$.

Related Problems

- Max-Cut in Data Stream. [Ahn and Guha 2009, Zelke 2011]
- Triangle Counting in Data Stream. [Buriol et al. 2006, Pavan et al. 2013, Cormode and Jowhari 2014]
- Max-Clique Problem. [Feige 2004, Khot and Ponnuswami 2006]

Prior Results

Halldórsson, Sun, Szegedy, and Wang (*ICALP 2012*) investigated the space complexity of the CLIQUE-GAP(r, s):

- they give matching upper and lower bounds for CLIQUE-GAP(r, s) for any r and $s = c\log(n)$, for some constant c.
- for smaller values of s, the bounds are unknown.

In our paper, we answered the above open problem: for $s = \tilde{O}(\log(n))$ and for any $r > s$, we prove that the space complexity of CLIQUE-GAP problem is $\tilde{\Theta}(\frac{ms^2}{r^2})$.

Our Results

- Upper Bound: we give a one-pass streaming algorithm that solves CLIQUE-GAP(r, s) using $\tilde{O}(ms^2/r^2)$ space.
- Lower Bound: we give a lower bound of $\tilde{\Omega}(ms^2/r^2)$ on the space complexity of CLIQUE-GAP(r, s) when $s = O(\log n)$, by showing a new connection between graph decomposition theory (Chung, Erdös, and Spencer '83, and Chung '81) and the multi-party set disjointness problem in communication complexity

Our Results

In addition,

- we extend our results to a lower bound theorem for the general promise problem GAP($\mathcal{P}, \mathcal{Q}$), which distinguishes between any two graph properties $\mathcal{P}$ and $\mathcal{Q}$ satisfying some restrictions (will clarity later).
- Our results for the CLIQUE-GAP problem can be extended to distinguish between graphs with at least T triangles and triangle-free graphs.
- We also give a new lower bound for the space complexity of CLIQUE-GAP($r, 2$) in the incidence model.

Pseudocode

Input A graph stream G with m edges; positive integers r and s.

Output "1" if a clique of order r is detected in G; "0" if G is $(s+1)$-clique free.

Init Set $p = 40(s+1)/r$.
Set memory buffer M empty.
Compute n pairwise independent bits $\{Q_v | \text{for all } v \in V\}$ using $O(\log n)$ space such that for each $v \in V$, $Pr[Q_v = 1] = p$.

While not the end of the stream:
Read an edge $e = (a, b)$.
Insert e into M if $Q_a = 1$ and $Q_b = 1$.
If there is an $(s+1)$-clique in M, **then** output "1".

Output "0"

The space complexity is $\tilde{O}(ms^2/r^2)$.

Main Theorem for Lower Bound

For any $0 < \delta < 1/2$ there exists a global constant $c > 0$ such that for any $0 < s < r$, $M > 0$, there exists graph families $\mathcal{G}_1$ and $\mathcal{G}_2$ that satisfy the following:

- for all graph $G_1 \in \mathcal{G}_1$, $|E(G_1)| = m \geq M$, G_1 has a r-clique;
- for all graph $G_2 \in \mathcal{G}_2$; $|E(G_2)| = m \geq M$, G_2 has no $(s+1)$-clique;
- any randomized one-pass streaming algorithm $\mathcal{A}$ that distinguishes whether $G \in \mathcal{G}_1$ or $G \in \mathcal{G}_2$ with probability at least $1 - \delta$ uses at least $cm/(r^2 \log_s^2 r)$ memory bits.

For $s = O(\log n)$ our lower bound matches, up to polylogarithmic factors, the upper bound.

Theorem (The Communication Complexity of Multi-party Set-disjointness Problem)

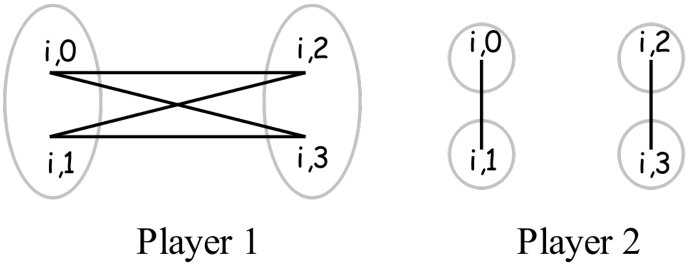
Denote the multi-party set disjointness problem as DISJ_k^n*. Any randomized one-way communication protocol that solves* DISJ_k^n *correctly with probability* $> 3/4$ *requires* $\Omega(n/k)$ *bits of communication.*

We show a reduction from DISJ_k^n to CLIQUE-GAP problem using the idea of graph decomposition.

Examples

Example (CLIQUE-GAP$(4, 2)$)

The reduction is from $\mathrm{DISJ}_2^{n/4}$ to CLIQUE-GAP$(4, 2)$



Reduction from $\mathrm{DISJ}_2^{n/4}$ to CLIQUE-GAP$(4, 2)$

For any instance of $\mathrm{DISJ}_2^{n/4}$, where Player 1 holds a set $S_1 \subset [n/4]$ and Player 2 holds a set $S_2 \subset [n/4]$, construct an instance G with n vertices of CLIQUE-GAP$(4, 2)$ as follows.

- Denote the set of n vertices by $\{v_{i,j} | i = 1, 2, 3, \ldots, n/4, j = 0, 1, 2, 3\}$. Naturally this vertex set can be partitioned into n/4 groups, each of size 4 (denoting as $V_i \equiv \{v_{i,0}, v_{i,1}, v_{i,2}, v_{i,3}\}$ for $i = 1, 2, 3, \ldots, n/4$).
- Partition $V_i = V_{i,0} \cup V_{i,1}$, where $V_{i,0} = \{v_{i,0}, v_{i,1}\}$ and $V_{i,1} = \{v_{i,2}, v_{i,3}\}$. Further partition $V_{i,0} = V_{i,0,0} \cup V_{i,0,1}$ and $V_{i,1} = V_{i,1,0} \cup V_{i,1,1}$, where $V_{i,0,0} = \{v_{i,0}\}$, $V_{i,0,1} = \{v_{i,1}\}$, $V_{i,1,0} = \{v_{i,2}\}$ and $V_{i,1,1} = \{v_{i,3}\}$.

Reduction from $\mathrm{DISJ}_2^{n/4}$ to CLIQUE-GAP$(4, 2)$

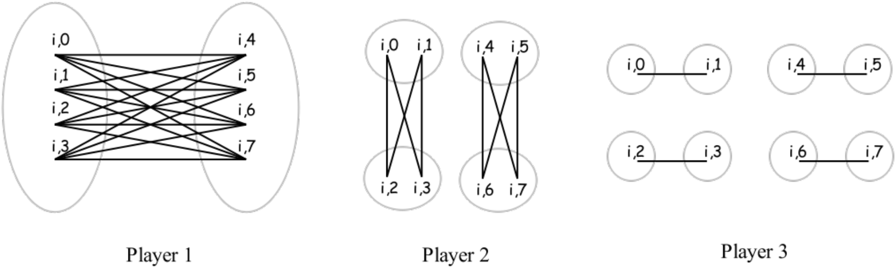
Player 1 places all edges of the complete bipartite graphs between $V_{i,0}$ and $V_{i,1}$ if $i \in S_1$. Player 2 places all edges between $V_{i,0,0}$ and $V_{i,0,1}$ and edges between $V_{i,1,0}$, $V_{i,1,1}$ if $i \in S_2$.

- If $S_1 \cap S_2 = \{i\}$, then there is a clique on vertex set V_i (which is of size 4).
- If $S_1 \cap S_2 = \emptyset$, since both Player 1 and Player 2 have only bipartite graph edges on disjoint vertex sets, the output graph is triangle free.

Examples

Example (CLIQUE-GAP(8, 3))

The reduction is from $\mathrm{DISJ}_3^{n/8}$ to CLIQUE-GAP(8, 3)



Finally we can generalize to a reduction from $\mathrm{DISJ}_{\lceil \log_s r \rceil}^{n/r}$ to CLIQUE-GAP(r, s).

Sketch of the Proof

1. Since any one-way communication protocol that solves $\mathrm{DISJ}^{n/r}_{\lceil \log_s r \rceil}$ need in total $\Omega(n/r \log_s r)$ communication cost, a one-pass streaming algorithm requires $\Omega(n/r \log_s r)/(\log_s r - 1)) = \Omega(n/r \log_s^2 r)$.
2. From our construction, for each hard instance we know $m = \Omega(r^2 \times n/r) = \Omega(nr)$. Hence any one-pass streaming algorithm that solves CLIQUE-GAP(r, s) requires $\Omega(m/r^2 \log_s^2 r)$ space.

A General Problem GAP($\mathcal{P}$, $\mathcal{Q}$)

GAP($\mathcal{P}$, $\mathcal{Q}$)

Let $\mathcal{P}$ and $\mathcal{Q}$ be two sets of graphs (graph properties) such that $\mathcal{P} \cap \mathcal{Q} = \emptyset$. Given an input graph G, an algorithm for GAP($\mathcal{P}$, $\mathcal{Q}$) should output "1" if $G \in \mathcal{P}$ and '0' if $G \in \mathcal{Q}$. For $G \notin \mathcal{P} \cup \mathcal{Q}$, the algorithm can output "1" or "0".

Lower Bound for General Problem GAP($\mathcal{P}, \mathcal{Q}$)

Let $\mathcal{P}, \mathcal{Q}$ be two graph properties such that

- $\mathcal{P} \cap \mathcal{Q} = \emptyset$;
- If $G'' \in \mathcal{P}$ and G'' is a subgraph of G', then $G' \in \mathcal{P}$;
- If $G', G'' \in \mathcal{Q}$ and $V(G') \cap V(G'') = \emptyset$, then $\tilde{G} = (V(G') \cup V(G''), E(G') \cup E(G'')) \in \mathcal{Q}$;

Let G_0 be an arbitrary graph in $\mathcal{P}$. Given any graph G with m edges and n vertices, if a one-pass streaming algorithm $\mathcal{A}$ solves $GAP(\mathcal{P}, \mathcal{Q})$ correctly with probability at least $3/4$, then $\mathcal{A}$ requires $\Omega(\frac{n}{|V(G_0)|} \frac{1}{\alpha_*^2(G_0, \mathcal{Q})})$ space in the worst case.

Thank You!