Detecting cliques in CONGEST networks

Artur Czumaj[1]
Department of Computer Science and Centre for Discrete Mathematics and its Applications (DIMAP), University of Warwick, UK
A.Czumaj@warwick.ac.uk

Christian Konrad[2]
Department of Computer Science, University of Bristol, UK
christian.konrad@bristol.ac.uk

Abstract

The problem of detecting network structures plays a central role in distributed computing. One of the fundamental problems studied in this area is to determine whether for a given graph H, the input network contains a subgraph isomorphic to H or not. We investigate this problem for H being a clique K_ℓ in the classical distributed CONGEST model, where the communication topology is the same as the topology of the underlying network, and with limited communication bandwidth on the links.

Our first and main result is a lower bound, showing that detecting K_ℓ requires $\Omega(\sqrt{n}/\mathfrak{b})$ communication rounds, for every $4 \le \ell \le \sqrt{n}$, and $\Omega(n/(\ell\mathfrak{b}))$ rounds for every $\ell \ge \sqrt{n}$, where $\mathfrak{b}$ is the bandwidth of the communication links. This result is obtained by using a reduction to the set disjointness problem in the framework of two-party communication complexity. We complement our lower bound with a two-party communication protocol for listing all cliques in the input graph, which up to constant factors communicates the same number of bits as our lower bound for K_4 detection. This demonstrates that our lower bound cannot be improved using the two-party communication framework.

2012 ACM Subject Classification Theory of Computation → Design and Analysis of Algorithms → Distributed Algorithms; Networks → Network Algorithms

Keywords and phrases Lower bounds; CONGEST; subgraph detection; two-party communication.

Digital Object Identifier 10.4230/LIPIcs.DISC.2018.16

1 Introduction

We study the problem of detecting network structures in a distributed environment, which is a fundamental problem in modern computing. Our focus is on the *subgraph detection problem*, in which for a given graph H, one wants to determine whether the network graph G contains a subgraph isomorphic to H or not. We investigate this problem for H being a clique K_ℓ for $\ell \ge 4$.

The nowadays classical distributed *CONGEST model* (see, e.g., [18]) is a variant of the classical LOCAL model of distributed computation (where in each round network nodes can send through all incident links messages of unrestricted size) with limited communication bandwidth. The distributed system is represented as a network (undirected graph)

[1] Research partially supported by the Centre for Discrete Mathematics and its Applications (DIMAP), by EPSRC award EP/D063191/1, by EPSRC award EP/N011163/1, and by an IBM Faculty Award.
[2] Most of work on this paper has been carried out while the author was at the University of Warwick, where he was supported by the Centre for Discrete Mathematics and its Applications (DIMAP) and by EPSRC award EP/N011163/1.

$G = (V, E)$ with $n = |V|$ nodes, where network nodes execute distributed algorithms in synchronous rounds, and the nodes collaborate to solve a graph problem with input G. Each node is assumed to have a unique identifier from $\{0, \ldots, \text{poly}(n)\}$. In any single round, all nodes can:

(i) perform an unlimited amount of local computation,
(ii) send a possibly different $\mathfrak{b}$-bit message to each of their neighbors, and
(iii) receive all messages sent to them.

We measure the *complexity* of an algorithms by the number of synchronous rounds required.

In accordance with the standard terminology in the literature, we assume $\mathfrak{b} = \mathcal{O}(\log n)$; we note though that our analysis generalizes to other settings of $\mathfrak{b}$ in a straightforward manner. (We note that in our lower bound for detecting K_4 and K_ℓ in Section 2, to ensure full generality of presentation, we will make the analysis parametrized by the message size $\mathfrak{b}$, in which case we will refer to such model of distributed computation as $\mathsf{CONGEST}_\mathfrak{b}$, the $\mathsf{CONGEST}$ model with messages of size $\mathfrak{b}$.)

Our goal is, for a given network $G = (V, E)$ and $\ell \geq 4$, to solve the *subgraph detection problem* for a clique K_ℓ, that is, to design an algorithm in the $\mathsf{CONGEST}$ model such that

(i) if G contains a copy of K_ℓ, then with probability $\geq \frac{2}{3}$ at least one node outputs 1, and
(ii) if G does not contain a copy of K_ℓ, then with probability $\geq \frac{2}{3}$ no node outputs 1.

The subgraph detection problem is a local problem: it can be solved efficiently solely on the basis of local information. In particular, in the $\mathsf{CONGEST}$ model, the problem of finding K_ℓ in a graph can be trivially solved in $\mathcal{O}(n)$ rounds, or in fact, in $\mathcal{O}(\max_{u \in V} \deg_G(u))$ rounds, where $\deg_G(u)$ denotes the degree of node u in G. Indeed, if each node sends its entire neighborhood to all its neighbors, then afterwards, each node will be aware of all its neighbors and of their neighbors. Therefore, in particular, each node will be able to detect all cliques it belongs to. Since for each node u, the task of sending its entire neighborhood to all its neighbors can be performed in $\mathcal{O}(\deg_G(u))$ rounds in the $\mathsf{CONGEST}$ model, the total number of rounds for the entire network is $\mathcal{O}(\max_{u \in V} \deg_G(u)) = \mathcal{O}(n)$ rounds. In view of this simple observation, the main challenge in the clique K_ℓ detection problem is whether this task can be performed in a *sublinear number of rounds.*

1.1 Our results

In this paper, we give the first non-trivial lower bound for the complexity of detecting a clique K_ℓ in the $\mathsf{CONGEST}_\mathfrak{b}$ model, for $\ell \geq 4$. In Theorem 5, we prove that every algorithm in the $\mathsf{CONGEST}_\mathfrak{b}$ model that with probability at least $\frac{2}{3}$ detects K_ℓ, for $\ell \geq 4$ and $\ell = \mathcal{O}(\sqrt{n})$, requires $\Omega(\sqrt{n}/\mathfrak{b})$ rounds. Further, if $\ell = \omega(\sqrt{n})$, then $\Omega(n/(\ell\,\mathfrak{b}))$ rounds are required. We are not aware of any other non-trivial (super-constant) lower bound for this problem in the $\mathsf{CONGEST}_\mathfrak{b}$ model.

We complement our lower bound with a two-party communication protocol for listing all cliques in the input graph (see Theorem 10), which up to constant factors communicates the same number of bits as our lower bound for K_4 detection. This demonstrates that our lower bound is essentially tight in this framework, and cannot be improved using the two-party communication approach.

1.2 Techniques: Framework of two-party communication complexity

Our main results, the lower bound of clique detection in Theorem 5 and the upper bound in Theorem 10, rely on the *two-party communication complexity* framework and the use of a tight lower bound for the set disjointness problem in this framework.

We consider the classical two-party communication complexity setting (cf. [16]) in which two players, Alice and Bob, each have some private input X and Y. The players' goal is to compute a joint function $\mathfrak{f}(X, Y)$, and the complexity measure used is the number of bits Alice and Bob must exchange to compute $\mathfrak{f}(X, Y)$. In the two-party communication problem of *set disjointness*, Alice's input is $X \in \{0, 1\}^n$ and Bob holds $Y \in \{0, 1\}^n$, and their goal is to compute $\mathrm{DISJ}_n(X, Y) := \bigvee_{i=1}^{n} X_i \wedge Y_i$. In a seminal work, Kalyanasundaram and Schnitger [14] showed that in any randomized communication protocol, the players must exchange $\Omega(n)$ bits to solve the set disjointness problem with constant success probability.

▶ **Theorem 1** ([14]). *The randomized two-party communication complexity of set disjointness is $\Omega(n)$. That is, for any constant $p > 0$, any randomized two-party communication protocol that computes $\mathrm{DISJ}_n(X, Y)$ with probability at least p, has two-party communication complexity $\Omega(n)$.*

Our main result, the lower bound for detecting K_ℓ in the CONGEST model, relies on a reduction from the two-party communication problem of set disjointness. The two-party communication framework, and, in particular, the two-party set disjointness problem, have been frequently used in the past to construct lower bounds for the CONGEST model, see, e.g., [4, 7, 9, 11, 15]. A typical approach relies on a construction of a special graph $G = (V, E)$ with some fixed edges and some edges depending on the input of Alice and Bob. One partitions the nodes of G into two disjoint sets V_A and V_B. Let $\mathcal{C}$ be the (V_A, V_B)-cut, that is, the set of edges in G with one endpoint in V_A and one endpoint in V_B. Let E_A be the edge set of $G[V_A]$ (subset of E on vertex set V_A) and E_B be the edge set of $G[V_B]$. We consider a scenario where Alice's input is represented by the subgraph $G_A = (V, E_A \cup \mathcal{C}) \subseteq G$ and Bob's input is represented by $G_B = (V, E_B \cup \mathcal{C}) \subseteq G$. (We denote this way of distributing the vertex and edge sets as the *vertex partition model*.) In order to learn any information about the structure of $G[A] \setminus \mathcal{C}$ and $G[B] \setminus \mathcal{C}$, and hence about the input of the other player, Alice and Bob must communicate through the edges of the cut $\mathcal{C}$. Therefore, in order to obtain a lower bound for a problem in the $\mathsf{CONGEST}_\mathfrak{b}$ model, one wants to construct G to ensure that it has some property (in our case, contains a copy of K_ℓ) if and only if the corresponding instance of set disjointness is such that $\mathrm{DISJ}_n(X, Y) = 1$, and in order to determine the required property, one has to communicate a large part of (essentially the entire graph) $G[A]$ through $\mathcal{C}$. With this approach, if the cut $\mathcal{C}$ has size $|\mathcal{C}|$, and the private inputs of Alice and Bob (edges in $G[A] \setminus \mathcal{C}$ or $G[B] \setminus \mathcal{C}$) are of size $\mathfrak{s}$, one can apply Theorem 1 to argue that the round complexity of any distributed algorithm in the $\mathsf{CONGEST}_\mathfrak{b}$ model for a given problem is $\Omega(\frac{\mathfrak{s}}{|\mathcal{C}| \cdot \mathfrak{b}})$. The central challenge is to ensure that for the encoded set disjointness instance of size $\mathfrak{s}$ and the cut of size $|\mathcal{C}|$, the ratio $\frac{\mathfrak{s}}{|\mathcal{C}|}$ is as large as possible.

For example, Drucker et al. [7] incorporated a similar approach to obtain a lower bound for the subgraph detection problem in a *broadcast* variant of the $\mathsf{CONGEST}_\mathfrak{b}$ model (in fact, even for a (stronger) broadcast variant of the CONGESTED CLIQUE model), where nodes are required to send the same message through all their incident edges. The lower bound construction requires sending $\Omega(n^2)$ bits through the cut of size $\mathcal{O}(n^2)$, but the fact that in the broadcast variant of the $\mathsf{CONGEST}_\mathfrak{b}$ model every node is required to send the same message via all incident edges, at most $\mathcal{O}(n\,\mathfrak{b})$ bits can be transmitted through the cut, yielding a lower bound of $\Omega(\frac{n}{\mathfrak{b}})$. (In particular, for the *broadcast* variant of the $\mathsf{CONGEST}_\mathfrak{b}$ model, Drucker et al. [7, Theorem 15] proved that detecting a clique K_ℓ, $\ell \geq 4$, requires $\Omega(\frac{n}{\mathfrak{b}})$ rounds.) Note however that in the (non-broadcast) $\mathsf{CONGEST}_\mathfrak{b}$ model, this construction does not give any not-trivial bound, since $\frac{\mathfrak{s}}{|\mathcal{C}|} = \mathcal{O}(1)$.

Our main building block for our lower bound is the construction of $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-*lower-bound graphs* (see Section 3.1 for the precise definition) that can be used to encode a set disjointness instance of size $\mathfrak{s} = \Omega(n^2)$ such that the cut is of size $|\mathcal{C}| = \mathcal{O}(n^{3/2})$. By incorporating these bounds in the framework described above, this construction leads to the first non-trivial lower bound of $\Omega(\frac{\sqrt{n}}{\mathfrak{b}})$ for the subgraph detection problem in the $\mathsf{CONGEST}_{\mathfrak{b}}$ model for the clique K_4. This construction can also be extended to detect larger cliques, yielding the lower bound of $\Omega(\frac{n}{(\ell+\sqrt{n})\,\mathfrak{b}})$ for detecting any K_ℓ with $\ell \geq 4$.

Since these are the first superconstant lower bounds for detecting a clique in the CONGEST model and since the best upper bound for these problems is still $\mathcal{O}(n)$, the next goal is to understand to what extent these bounds could be improved and whether the existing approach could be used for that task. Do we need $\Omega(\frac{\sqrt{n}}{\mathfrak{b}})$ communication rounds to detect any clique K_ℓ (with $\ell \geq 4$, $\ell = \mathcal{O}(\sqrt{n})$) in the $\mathsf{CONGEST}_{\mathfrak{b}}$ model, or maybe we need as many as a linear number of rounds? While we do not know the answer to this question, and in fact, this question is the main open problem left by this paper, we can prove that any better lower bound would require a significantly different approach, going beyond the two-party communication framework in the vertex partition model.

Indeed, let us consider the vertex partition model in the two-party communication framework, as defined above. The input consists of an undirected $G = (V, E)$ with an arbitrary vertex partition $V = V_A \,\dot\cup\, V_B$. We consider a scenario where Alice is given the subgraph $G_A = (V, E_A \cup \mathcal{C}) \subseteq G$ and Bob is given $G_B = (V, E_B \cup \mathcal{C}) \subseteq G$, where $\mathcal{C}$ is the (V_A, V_B)-cut in G. The arguments in our construction of lower-bound graphs in Theorem 9 imply that for some inputs, any two-party communication protocol in the vertex partition model for the problem of listing all cliques in a given graph with n nodes requires communication of $\Omega(\sqrt{n}\,|\mathcal{C}|)$ bits between Alice and Bob. We will prove in Section 4 (Theorem 10) that this lower bound is asymptotically tight in the two-party communication framework in the vertex partition model. We show that there is a two-party communication protocol in the vertex partition model for listing *all cliques* that uses $\mathcal{O}(\sqrt{n}\,|\mathcal{C}|)$ communication rounds, where $\mathcal{C}$ is the set of shared edges between Alice and Bob. This shows that we cannot obtain stronger lower bounds for the K_ℓ-detection problem, for $\ell = \mathcal{O}(\sqrt{n})$, in the CONGEST model using the two-party communication framework in the vertex partition model.

1.3 Related works

As a fundamental primitive, subgraph detection and listing in the CONGEST model has been recently receiving attention from multiple authors, focusing mainly on randomized complexity. However, despite major efforts, for the CONGEST model, relatively little is known about the complexity of the subgraph detection problem.

Prior to our work, no non-trivial results about the complexity of clique K_ℓ ($\ell \geq 4$) detection in the CONGEST model have been known. While there is a trivial lower bound of a constant number of rounds, and as we mentioned earlier, one can easily solve the problem in $\mathcal{O}(n)$ rounds in the CONGEST model, no sublinear upper bounds nor superconstant lower bounds have been known.

In a recent breakthrough in this area, Izumi and Le Gall [12] raised some hopes that maybe these problems could be solved in a sublinear number of rounds in the CONGEST model. They considered the subgraph detection problem for the smallest interesting subgraph H, the triangle K_3, and presented a very clever algorithm that detects a triangle in $\widetilde{\mathcal{O}}(n^{2/3})$ rounds. Further, they also showed that the related problem of finding all triangles (triangle listing) can be solved in $\widetilde{\mathcal{O}}(n^{3/4})$ rounds. Very recently, these results were improved by Chang et al. [5], who showed that both triangle detection and enumeration can be solved in $\widetilde{\mathcal{O}}(\sqrt{n})$ rounds in the CONGEST model. There is no non-trivial lower bound for the triangle detection problem, though it is known (cf. [12, 17]) that the more complex triangle listing problem requires $\Omega(n^{1/3}/\log n)$ rounds, even in the CONGESTED CLIQUE model. It can also be shown that the problem of listing all triangles such that each node v learns all triangles that it is part of significantly harder than the general triangle listing problem and requires $\Omega(n/\log n)$ rounds [12, Proposition 4.4]. While rather disappointingly, we do not know how to extend any of these upper bounds to other cliques K_ℓ with $\ell \geq 4$, the previously mentioned works for triangle detection raise hope that detecting cliques K_ℓ could potentially be solved in a sublinear number of rounds. In fact, even for K_3, we do not even know whether detecting a triangle K_3 can be solved in a polylogarithmic or even a constant number of rounds in the CONGEST model (the lower bound of $\Omega(n^{1/3}/\log n)$ rounds in the CONGESTED CLIQUE model (cf. [12, 17]) holds only for a more complex problem of detecting *all triangles*).

Even et al. [8] noted that the problem of detecting trees is significantly simpler and designed a randomized color-coding algorithm that detects any constant-size *tree* on ℓ nodes in $\mathcal{O}(\ell^\ell)$ rounds.

As for lower bounds for the subgraph detection problem in the CONGEST model, until very recently, the only hardness results known in the literature have been for cycles. For any fixed $\ell \geq 4$, there is a polynomial lower bound for detecting the ℓ-cycle C_ℓ in the CONGEST model [7], where it has been shown that detecting C_ℓ requires $\Omega(\mathrm{ex}(n, C_\ell)/\log n)$ rounds, where $\mathrm{ex}(n, C_\ell)$ is the Turán number for cycles, that is, the largest possible number of edges in a C_ℓ-free graph over n vertices. In particular, for odd-length cycles (of length 5 or more), the lower bound of [7] is $\Omega(n/\log n)$, and it is $\Omega(\sqrt{n}/\log n)$ for $\ell = 4$. Very recently, Korhonen and Rybicki [15] improved the lower bound for all even-length cycles to $\Omega(\sqrt{n}/\log n)$. Further, Gonen and Oshman [11] extended these lower bounds for C_ℓ-freeness to some related classes of graphs, though still with some cyclic underlying structure. (As mentioned above, we note that Drucker et al. [7] presented lower bounds for other graphs, but this was in a *broadcast* variant of the CONGESTED CLIQUE model, where nodes are required to send the same message on all their edges. In particular, for the broadcast variant of the CONGESTED CLIQUE model, Drucker et al. [7] proved that detecting a clique K_ℓ, $\ell \geq 4$, requires $\Omega(n/\log n)$ rounds.)

The only lower bound for the subgraph detection problem for H significantly other than cycles, is a very recent work of Fischer et al. [9], who demonstrated that the subgraph detection problem is hard even for some subgraphs H of constant size. In particular, for any constant $\ell \geq 2$, there is a graph H with a constant number of vertices and edges such that the problem of finding H in a network of size n requires time $\Omega(n^{2-\frac{1}{\ell}}/\mathfrak{b})$ in the CONGEST model, where $\mathfrak{b}$ is the bandwidth of each communication links.

There has also been some recent research for the *deterministic* subgraph detection problem in the CONGEST model. For example, Drucker et al. [7] designed an $\mathcal{O}(\sqrt{n})$ round algorithm for C_4 detection, and Even et al. [8] and Korhonen and Rybicki [15] obtained path and tree detection algorithms requiring only a constant number of rounds. Korhonen and Rybicki [15] considered also deterministic subgraph detection (for paths, cycles, trees, pseudotrees, and on d-degenerate graphs) in the weaker broadcast CONGEST model, where nodes send the same message to all neighbors in each communication round. In the CONGESTED CLIQUE model, deterministic subgraph detection algorithms were given by Dolev et al. [6] and Censor-Hillel et al. [3].

We summarize earlier results together with our new results in Table 1.

Paper	Time bound	Problem	Model
[8]	$\mathcal{O}(\ell^\ell)$	Detecting a tree on ℓ nodes	CONGEST
folklore	$\mathcal{O}(n)$	Detecting K_ℓ, $\ell \geq 3$	CONGEST
[5]	$\widetilde{\mathcal{O}}(\sqrt{n})$	Detecting triangle K_3	CONGEST
[5]	$\widetilde{\mathcal{O}}(\sqrt{n})$	Triangle listing	CONGEST
[9]	$\Omega(n^{2-\frac{1}{\ell}}/\log n)$	Detecting some H of size $\mathcal{O}(\ell)$	CONGEST
[7]	$\Omega(n/\log n)$	Detecting C_ℓ, $\ell \geq 5$, ℓ odd	CONGEST
[7, 15]	$\Omega(\sqrt{n}/\log n)$	Detecting C_ℓ, $\ell \geq 4$, ℓ even	CONGEST
[12, 17]	$\Omega(n^{1/3}/\text{poly-log}(n))$	Triangle listing	CONGESTED CLIQUE
[7]	$\Omega(n/\log n)$	Detecting K_ℓ for $\ell \geq 4$	br. CONGESTED CLIQUE
Thm. 4	$\Omega(\sqrt{n}/\log n)$	Detecting K_4	CONGEST
Thm. 5	$\Omega(\sqrt{n}/(\ell \log n))$	Detecting K_ℓ for $\ell \geq 4$	CONGEST

Table 1 Prior (randomized) results for the problem of detecting a given subgraph H, or for listing all copies of H, in the CONGEST model (less relevant results (upper bounds) for the CONGESTED CLIQUE model are omitted; note that lower bounds for CONGESTED CLIQUE hold also for CONGEST and lower bounds for broadcast CONGESTED CLIQUE, abbreviated by br. CONGESTED CLIQUE in the table, do not imply any bounds for CONGEST).

1.3.1 Property testing of H-freeness

Since there have been so few positive results for the original subgraph detection problem, recently there have been some advances in a relaxation of this problem, a closely related (and significantly simpler) problem of *testing subgraphs freeness* in the *framework of property testing for distributed computations* (see, e.g., [1, 8]). In the property testing setting, an algorithm has to decide, with probability at least $\frac{2}{3}$, if the input graph is (a) H-free (i.e., does not contain a subgraph isomorphic to H) or (b) ε-far from being H-free (that is, the goal is to distinguish whether the input graph G is H-free or one needs to modify more than $\varepsilon|E(G)|$ edges of G to obtain a graph that is H-free); in the intermediate case, the algorithm can perform arbitrarily (see e.g., [3, 8] for more details). Property testing of H-freeness in the CONGEST model has received a lot of attention lately (see, e.g., [1, 2, 8, 9, 10]). In particular, it has been shown [8] that testing H-freeness can be done in $\mathcal{O}(1/\varepsilon)$ round in the CONGEST model for any constant-size graph H containing an edge (x, y) such that any cycle in H contains at least one of x, y. This implies testing in $\mathcal{O}(1/\varepsilon)$ rounds of any cycle C_k, and of any subgraph H on five (or less) vertices except K_5. Further, for any $\ell \geq 5$, K_ℓ-freeness can be tested in $\mathcal{O}((\varepsilon \cdot |E(G)|)^{\frac{1}{2}-\frac{1}{\ell-2}}/\varepsilon)$ rounds [8]. For trees, Even et al. [8] show that testing if the input graph is T-free for a tree T on ℓ vertices can be done in $\mathcal{O}(\ell^{1+\ell^2}/\varepsilon^\ell)$ rounds the CONGEST model.

2 Lower bound results: Detecting a clique requires $\widetilde{\Omega}(\sqrt{n})$ rounds

In this section we prove our hardness results showing that any algorithm in the $\text{CONGEST}_{\mathfrak{b}}$ model that detects a K_ℓ with probability at least $\frac{2}{3}$ requires $\Omega(\sqrt{n}/\mathfrak{b})$ rounds, for every $\ell = \mathcal{O}(\sqrt{n})$ and $\ell \geq 4$, and requires $\Omega(\frac{n}{\ell \mathfrak{b}})$ rounds if $\ell = \omega(\sqrt{n})$ (Theorems 4 and 5); or in short, $\Omega(\frac{n}{(\ell+\sqrt{n})\,\mathfrak{b}})$ rounds, for every $\ell \geq 4$. Our lower bound for the complexity of detecting K_ℓ in the CONGEST model relies on a reduction to the two-party communication

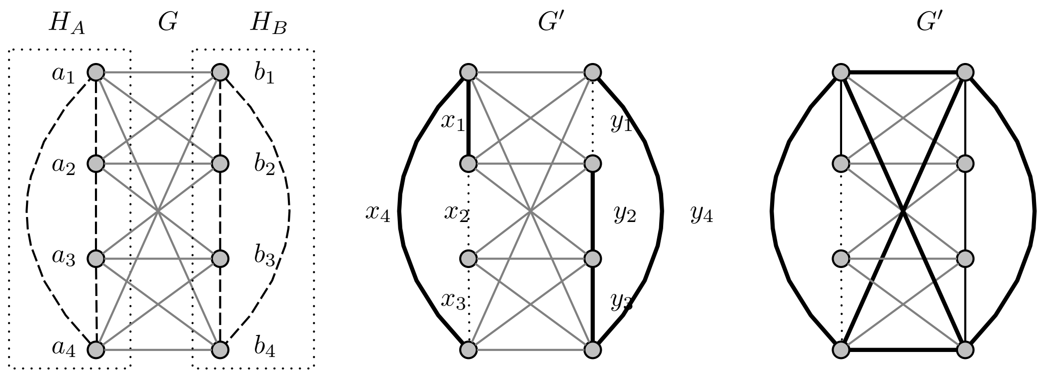

Figure 1 Left: Example of a $(4, 12)$-lower-bound graph $G = (A, B, E)$. The dotted edges are the edges of the associated graphs H_A and H_B (observe that H_A and H_B form cycles of length 4, which are bipartite). For $1 \le i \le 4$, let $\mathcal{E}_i$ be the edge set of subgraph $G[\{a_i, a_{(i \bmod 4)+1}, b_i, b_{(i \bmod 4)+1}\}]$. Observe that $E = \bigcup_{i \le 4} \mathcal{E}_i$, and, for every i, $G[\mathcal{E}_i]$ is isomorphic to $K_{2,2}$. Observe further that for $i \ne j$, $G[A(\mathcal{E}_i) \cup B(\mathcal{E}_j)]$ is not isomorphic to $K_{2,2}$. **Center:** Graph G' as in the proof of Theorem 3 obtained from the set disjointness instance with $X = (1, 0, 0, 1)$ and $Y = (0, 1, 1, 1)$. Graph G' contains a K_4 if and only if the set disjointness instance evaluates to 1. **Right:** The highlighted edges form a K_4.

complexity lower bound for the set disjointness problem (cf. Theorem 1 in Section 1.2), which we implement with the help of lower-bound graphs (cf. Section 2.1).

2.1 Lower-bound graphs

Our reduction to the two-party communication complexity lower bound for the set disjointness problem relies on a notion of a *lower-bound graph* (cf. Figure 1).

▶ **Definition 2.** Let $G = (A, B, E)$ be a bipartite graph with $|A| = |B| = n$ and let k, m be integers. Then G is called a (k, m)*-lower-bound graph* if:

1. $|E| \le m$.
2. The edge set E is the union of (not necessarily disjoint) sets $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ such that, for every i, $1 \le i \le k$, the edge-induced subgraph $G[\mathcal{E}_i]$ *is isomorphic to* $K_{2,2}$.
3. For every i, j, $1 \le i, j \le k$, $i \ne j$, the vertex-induced subgraph $G[A(\mathcal{E}_i) \cup B(\mathcal{E}_j)]$ is *not* isomorphic to $K_{2,2}$ (For a set of edges $E' \subseteq E$ we denote the set of incident A-vertices by $A(E')$. The set $B(E')$ is defined similarly.).
4. Define two *graphs associated with* G, $H_A = (A, E_A)$ and $H_B = (B, E_B)$. H_A is the graph on vertex set A, where $a_1, a_2 \in A$ are adjacent if and only if there exists an index i with $A(\mathcal{E}_i) = \{a_1, a_2\}$. Similarly, H_B is the graph on vertex set B, where $b_1, b_2 \in B$ are adjacent if and only if there exists an index j with $B(\mathcal{E}_j) = \{b_1, b_2\}$. Then, we require that H_A and H_B are *bipartite*.

2.2 Using lower-bound graphs and set disjointness to prove the hardness of clique detection

With the notion of lower-bound graphs at hand, we can formalize our reduction to the two-party communication complexity lower bound for set disjointness to obtain the following central theorem.

▶ **Theorem 3.** *Let G be a (k,m)-lower-bound graph. Then, detecting a K_4 in the CONGEST$_{\mathfrak{b}}$ model with probability at least $\frac{2}{3}$ requires $\Omega\left(\frac{k}{m\mathfrak{b}}\right)$ rounds.*

Proof. Let $\mathcal{A}$ be an algorithm in the CONGEST$_{\mathfrak{b}}$ model for K_4 detection, that is, such that with probability at least $\frac{2}{3}$, if G contains a K_4 then at least one node outputs 1 and if G contains no copy of K_4 then no node outputs 1. We will show that $\mathcal{A}$ can be used to solve the two-party set disjointness problem for instances of size k.

Consider a set disjointness instance (X,Y) of size k. Let $G=(A,B,E)$ be a (k,m)-lower-bound graph, let $\mathcal{E}_1,\mathcal{E}_2,\ldots,\mathcal{E}_k$ be the edge partition as in Item 2 of Definition 2, and let $H_A=(A,E_A)$ and $H_B=(B,E_B)$ be the graphs associated with G (Item 4 in Definition 2). Alice constructs the set $E'_A \subseteq E_A$ such that for every i with $X_i=1$, the edge between $A(\mathcal{E}_i)$ is included in E'_A. Similarly, Bob constructs the set $E'_B \subseteq E_B$ such that for every i with $Y_i=1$, the edge between $B(\mathcal{E}_i)$ is included in E'_B.

We first show that the graph $G' := G \cup (E'_A \cup E'_B)$ contains a K_4 if and only if $\mathrm{DISJ}_n(X,Y)=1$. Indeed, since by Item 4 of Definition 2, the graphs H_A and H_B are bipartite (and thus the subgraphs $G'[A]$ and $G'[B]$ are bipartite too), any copy of K_4 in G' must consist of two vertices from A and two vertices from B. Let a_1,a_2 be any pair of distinct vertices in A and b_1,b_2 be any pair of distinct vertices in B. Observe that if there is no $\mathcal{E}_i$ such that $\{a_1,a_2\}=A(\mathcal{E}_i)$ or there is no $\mathcal{E}_i$ such that $\{b_1,b_2\}=B(\mathcal{E}_i)$ then it is impossible for the nodes a_1,a_2,b_1,b_2 to form a K_4, since this would imply that either $a_1a_2 \notin E'_A$ or $b_1b_2 \notin E'_B$. Assume therefore that $\{a_1,a_2\}=A(\mathcal{E}_i)$ and $\{b_1,b_2\}=B(\mathcal{E}_j)$, for some i,j. Next, suppose that $i \neq j$. Then $G[\{a_1,a_2,b_1,b_2\}]$ is not isomorphic to $K_{2,2}$, by Item 3 of Definition 2. Hence, assume that $i=j$. Then $G[\{a_1,a_2,b_1,b_2\}]$ forms a $K_{2,2}$ if and only if $X_i=Y_i=1$, which in turn implies $\mathrm{DISJ}_n(X,Y)=1$.

The simulation of $\mathcal{A}$ on G' is executed as follows. Suppose that $\mathcal{A}$ runs in r rounds. Alice simulates vertices A and Bob simulates vertices B. In round i, Alice sends all messages from A with destinations in B to Bob, and Bob sends all messages from B with destinations in A to Alice. Since the cut between A and B is of size m, Alice and Bob exchange messages with overall $m\mathfrak{b}$ bits per round. Thus, overall they communicate $rm\mathfrak{b}$ bits. Since the algorithm allows them to solve set disjointness, by Theorem 1, we have $rm\mathfrak{b}=\Omega(k)$. Thus, $\mathcal{A}$ requires $\Omega(\frac{k}{m\mathfrak{b}})$ rounds. ◀

In Theorem 9 in Section 3, we prove the existence of a $(\Omega(n^2),\mathcal{O}(n^{3/2}))$-lower-bound graph. By combining Theorem 9 with Theorem 3, we obtain the following main result.

▶ **Theorem 4.** *Every algorithm in the CONGEST$_{\mathfrak{b}}$ model that detects a K_4 with probability at least $\frac{2}{3}$ requires $\Omega(\sqrt{n}/\mathfrak{b})$ rounds.*

2.3 Detection of K_ℓ for $\ell \geq 5$

The lower bound construction given in Theorem 3 can be extended to the task of detecting K_ℓ, for $\ell \geq 5$ (see also Figure 2). To this end, we add a clique on $\ell-4$ new nodes to graph G' (from the proof of Theorem 3) and connect each of these nodes to every vertex in $A \cup B$. Observe that this increases the cut between A and B by $n(\ell-4)$ edges. For $\ell=\mathcal{O}(\sqrt{n})$, there are only $\mathcal{O}(n^{3/2})$ additional edges, which implies that the same lower bound as for K_4 holds. If $\ell=\omega(\sqrt{n})$, then the number of additional edges is significant, since the size of the cut increases by more than a constant factor. In this case, the round complexity is $\Omega(\frac{n^2}{n(\ell-4)\,\mathfrak{b}})=\Omega(\frac{n}{\ell\,\mathfrak{b}})$. Similarly as before, the encoded set disjointness instance evaluates to 1 if and only if G' contains a clique of size ℓ. We thus conclude with the following theorem.

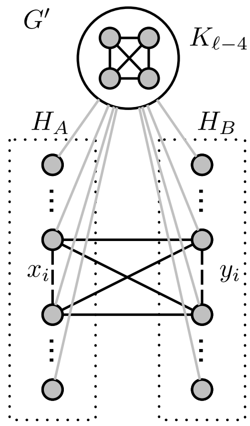

Figure 2 Extension of our lower bound for K_4 detection to K_ℓ detection, for $\ell \geq 5$. We add a clique $K_{\ell-4}$ on $\ell - 4$ new vertices to the graph G' and connect every vertex of the clique to every other vertex of G'. Then the resulting graph contains a clique on ℓ vertices if and only if the encoded set disjointness instance evaluates to 1, i.e., $x_i = y_i = 1$, for some i.

▶ **Theorem 5.** *Every algorithm in the CONGEST$_\mathfrak{b}$ model that detects K_ℓ, for $\ell \geq 4$ and $\ell = \mathcal{O}(\sqrt{n})$, with probability at least $\frac{2}{3}$ requires $\Omega(\sqrt{n}/\mathfrak{b})$ rounds. If $\ell = \omega(\sqrt{n})$, then $\Omega(n/(\ell\,\mathfrak{b}))$ rounds are required.*

3 Lower-bound graph construction

In this section, we prove the existence of a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower-bound graph (see Definition 2), which is our main technical tool. We will show in Theorem 9 that Algorithm 1 below constructs a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower-bound graph with high probability (observe that a non-zero probability already suffices to prove the existence of such a graph).

3.1 Construction of a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower-bound graph

We proceed as follows. We start our construction with a bipartite random graph $G = (A, B, E)$ with $|A| = |B| = n$, where every potential edge ab between $a \in A$ and $b \in B$ is included with probability $p = \frac{1}{\sqrt{n}}$. Observe that for any $a_1, a_2 \in A$ ($a_1 \neq a_2$) and $b_1, b_2 \in B$ ($b_1 \neq b_2$), the probability that $G[\{a_1, a_2, b_1, b_2\}]$ is isomorphic to a $K_{2,2}$ is p^4. We therefore expect G to contain $\binom{n}{2}^2 p^4$ copies of $K_{2,2}$, and we prove in Lemma 6 below that, with high probability, the actual number of copies of $K_{2,2}$ does not deviate significantly from its expectation. Let $\mathcal{K}$ denote the set of copies of $K_{2,2}$ in G.

In the peeling phase, we greedily compute a subset $\mathcal{H} \subseteq \mathcal{K}$ such that at the end, the graph induced by the edges of $\mathcal{H}$ is a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower bound graph. When inserting a set $K = \{a_1, a_2, b_1, b_2\} \in \mathcal{K}$ into $\mathcal{H}$, we make sure that the following three properties are fulfilled:

1. We ensure that later on we will never add a $K' = \{a'_1, a'_2, b'_1, b'_2\}$ such that either $\{a_1, a_2, b'_1, b'_2\}$ or $\{a'_1, a'_2, b_1, b_2\}$ form a $K_{2,2}$. To this end, when inserting K into $\mathcal{H}$, for every $K' \in \mathcal{K}$ that contains the same pair of A-vertices (or B-vertices), we add its pair of B vertices (resp. pair of A vertices) to set F_B (resp. F_A), indicating that this is

Algorithm 1. Construction of a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower-bound graph:

Input: Integer n, let $p = \frac{1}{\sqrt{n}}$.

1. **Random Graph:**
   Let $G = (A, B, E)$ with $|A| = |B| = n$ be the bipartite random graph where
   for every $a \in A, b \in B$ the edge ab is included in E with probability p.
   Let $\mathcal{K}$ be the family of sets $\{a_1, a_2, b_1, b_2\}$ with $a_1, a_2 \in A$, $a_1 \neq a_2$, $b_1, b_2 \in B$, $b_1 \neq b_2$
   and $G[\{a_1, a_2, b_1, b_2\}]$ isomorphic to $K_{2,2}$.
   For $S \subseteq A \cup B$, let $\mathcal{K}(S) := \{K \in \mathcal{K} : S \subseteq K\}$.
2. **Peeling Process:**
   Let $A' \subseteq A$ and $B' \subseteq B$ be a uniform random sample of A and B, respectively,
   where every vertex is included with probability $\frac{1}{2}$.
   $\mathcal{H} \leftarrow \{\}$, $F_A \leftarrow \{\}$, $F_B \leftarrow \{\}$.
   for every $K = \{a_1, a_2, b_1, b_2\} \in \mathcal{K}$ **do**
   if $|\mathcal{K}(\{a_1, a_2\})| \leq 6$ and $|\mathcal{K}(\{b_1, b_2\})| \leq 6$ and $|\{a_1, a_2\} \cap A'| = |\{b_1, b_2\} \cap B'| = 1$ and
   $\{a_1, a_2\} \notin F_A$ and $\{b_1, b_2\} \notin F_B$ **then**
   $\mathcal{H} \leftarrow \mathcal{H} \cup K$.
   For every $\{a_1, a_2, b_3, b_4\} \in \mathcal{K}(\{a_1, a_2\})$, add $\{b_3, b_4\}$ to F_B.
   For every $\{a_3, a_4, b_1, b_2\} \in \mathcal{K}(\{b_1, b_2\})$, add $\{a_3, a_4\}$ to F_A.
   end if
   end for
3. **Lower Bound Graph H:**
   For $K = \{a_1, a_2, b_1, b_2\} \in \mathcal{H}$, let E_K be the edge set $\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}$.
   return $H := (A, B, \bigcup_{K \in \mathcal{H}} E_K)$.

a forbidden pair. Then, when inserting an element of $\mathcal{K}$ into $\mathcal{H}$, we make sure that its pairs of A and B vertices are not forbidden.

2. We make sure that the insertion of K will not prevent too many other sets K' from being inserted into $\mathcal{H}$. To this end, we guarantee that there are at most six other sets in $\mathcal{K}$ that share the same pair of A vertices and at most six other sets that share the same pair of B vertices. We prove in Lemma 7 that most $K \in \mathcal{K}$ fulfill this property.
3. It is required that the graphs G_A and G_B as defined in Item 4 of Definition 2 are bipartite. We therefore partition the sets A and B randomly into subsets A' and $A \setminus A'$, and B' and $B \setminus B'$, and only add K to $\mathcal{H}$ if exactly one of its A vertices is in A' and one of its B vertices is in B'.

In the last step of the algorithm, we assemble graph H as the union of the edges contained in the copies of $K_{2,2}$ in $\mathcal{H}$.

3.2 Analysis of Algorithm 1

Our analysis relies on some basic properties of the structure of subgraphs of random graphs (for a more complete treatment of related problems, see, e.g., [13, Chapter 3]). We prove three high probability claims about the construction in Algorithm 1: that the random graph G contains many copies of $K_{2,2}$ (Lemma 6), that only a small fraction of pairs of A vertices are contained in more than six copies of $K_{2,2}$ (Lemma 7), and finally that the resulting graph H contains $\Omega(n^2)$ copies of $K_{2,2}$ (Lemma 8). With these three claims at hand, we will complete the analysis to prove in Theorem 9 that with high probability, the output of Algorithm 1 is a $(\Omega(n^2), \mathcal{O}(n^{3/2}))$-lower-bound graph.

We begin with a proof that in Algorithm 1, the random graph G contains many copies of $K_{2,2}$.

▶ **Lemma 6.** *Suppose that $p \geq \frac{1}{n}$. Then there is a constant C such that*

$$\mathbb{P}\left[|\mathcal{K}| \leq \frac{9}{10}\binom{n}{2}^2 p^4\right] \leq C \cdot \frac{1}{n^2 p} \ .$$

Proof. We will compute the expectation and the variance of $|\mathcal{K}|$ and then use Chebyshev's inequality to bound the probability that $|\mathcal{K}|$ deviates substantially from its expectation.

Let $\mathcal{X}$ be the family of all sets $\{a_1, a_2, b_1, b_2\}$ with $a_1, a_2 \in A$, $a_1 \neq a_2$, $b_1, b_2 \in B$, $b_1 \neq b_2$, and for $X \in \mathcal{X}$ let $\chi(X)$ be the indicator variable of the event "$G[X]$ is isomorphic to $K_{2,2}$". Then:

$$\mathbb{E}|\mathcal{K}| = \sum_{X \in \mathcal{X}} \mathbb{P}\left[\chi(X) = 1\right] = |\mathcal{X}| p^4 = \binom{n}{2}^2 p^4 \ ,$$

since $K_{2,2}$ contains 4 edges. To bound the variance $\mathbb{V}|\mathcal{K}|$, we use the identity $\mathbb{V}|\mathcal{K}| = \mathbb{E}|\mathcal{K}|^2 - (\mathbb{E}|\mathcal{K}|)^2$:

$$\mathbb{E}|\mathcal{K}|^2 = \mathbb{E}\left(\sum_{X \in \mathcal{X}} \chi(X)\right)^2 = \mathbb{E} \sum_{X,Y \in \mathcal{X}} \chi(X) \cdot \chi(Y) = \sum_{X,Y \in \mathcal{X}} \mathbb{E}(\chi(X) \cdot \chi(Y)) \ .$$

We distinguish the following cases:

- $|X \cap Y| = 0$. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^8$. Observe that there are $t_0 = \binom{n}{2}^2 \binom{n-2}{2}^2$ such pairs.
- $|X \cap Y| = 1$. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^8$. There are $t_1 = 4\binom{n}{2}^2 \binom{n-2}{2}\binom{n-2}{1}$ such pairs.
- $|X \cap Y| = 2$ and the intersection consists of either two A-vertices or two B-vertices. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^8$ and there are $t_{2,1} = 2 \cdot \binom{n}{2}^2 \binom{n-2}{2}$ such pairs.
- $|X \cap Y| = 2$ and the intersection consists of one A-vertex and one B-vertex. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^7$ and there are $t_{2,2} = 4 \cdot \binom{n}{2}^2 \cdot (n-2)^2$ such pairs.
- $|X \cap Y| = 3$. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^6$. There are $t_3 = 4 \cdot \binom{n}{2}^2 \cdot (n-2)$ such pairs.
- $|X \cap Y| = 4$. Then, $\mathbb{E}(\chi(X) \cdot \chi(Y)) = p^4$. There are $t_4 = \binom{n}{2}^2$ such pairs.

A quick sanity check shows that $t_0 + t_1 + t_{21} + t_{22} + t_3 + t_4 = \binom{n}{2}^4$. We thus obtain:

$$\begin{aligned}\mathbb{V}|\mathcal{K}| = \mathbb{E}|\mathcal{K}|^2 - (\mathbb{E}|\mathcal{K}|)^2 &= p^8(t_0 + t_1 + t_{2,1}) + p^7 t_{2,2} + p^6 t_3 + p^4 t_4 - \binom{n}{2}^4 p^8 \\ &\leq p^7 t_{2,2} + p^6 t_3 + p^4 t_4 = \mathcal{O}(p^7 n^6) \ ,\end{aligned}$$

where the last equality holds for every $p \geq \frac{1}{n}$. We apply Chebyshev's inequality and obtain:

$$\mathbb{P}\left[\Big||\mathcal{K}| - \mathbb{E}|\mathcal{K}|\Big| \geq \frac{1}{10}\mathbb{E}|\mathcal{K}|\right] \leq \frac{100\mathbb{V}|\mathcal{K}|}{(\mathbb{E}|\mathcal{K}|)^2} = C \cdot \frac{1}{n^2 p} \ ,$$

for some constant C. ◀

Next, we prove that only a small fraction of pairs of A vertices are contained in more than six copies of $K_{2,2}$.

▶ **Lemma 7.** *Let $p = \frac{1}{\sqrt{n}}$. For every constant $\delta > 0$, with high probability, there are at most $(1+\delta)n^2/10$ pairs of distinct vertices $a_1, a_2 \in A$ with $|\mathcal{K}(\{a_1, a_2\})| > 6$.*

Proof. Let $a_1, a_2 \in A$, $a_1 \neq a_2$ be arbitrary vertices. Let $B(\{a_1, a_2\}) \subseteq B$ be the set of vertices b such that $a_1 b, a_2 b \in E$. Observe that $|\mathcal{K}(\{a_1, a_2\})| = \binom{|B(\{a_1,a_2\})|}{2}$. By linearity of expectation, $\mathbb{E}|B(\{a_1, a_2\})| = np^2 = 1$.

Let $\mathcal{X}$ be the family of all sets of vertices $\{a_1, a_2\} \subseteq A$ with $a_1 \neq a_2$. Partition now $\mathcal{X}$ into disjoint subsets such that $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_{n-1}$, where $|\mathcal{X}_i| = n/2$ and, for every $1 \leq i \leq n-1$, all elements of $\mathcal{X}_i$ are pairwise disjoint (such a partitioning corresponds to partitioning the complete graph K_n into $n-1$ perfect matchings). For a pair of vertices $P \in \mathcal{X}$, let $\chi(P)$ be the indicator variable of the event "$|B(P)| \geq 5$". Recall that $\mathbb{E}|B(P)| = np^2 = 1$ (since $p = 1/\sqrt{n}$). Hence, by Markov's inequality, we have $\mathbb{P}[\chi(P) = 1] \leq \frac{1}{5}$.

For every $1 \leq i \leq n-1$ we have $\mathbb{E}\sum_{P \in \mathcal{X}_i} \chi(P) \leq \frac{1}{5}\frac{n}{2} = \frac{n}{10}$. Observe further that for every $P, Q \in \mathcal{X}_i$, $P \neq Q$, the random variables $B(P)$ and $B(Q)$ are independent. Thus, by a Chernoff bound (for $\mu = \frac{n}{10}$):

$$\mathbb{P}\left[|\sum_{S \in \mathcal{X}_i} \chi(S) - \mu| \geq \delta\mu\right] \leq 2\exp\left(-\mu\delta^2/3\right) = e^{-\Theta(n)} \ ,$$

for any constant δ. Thus, applying the union bound for every $1 \leq i \leq n-1$, with high probability, at most $(1+\delta)\frac{n}{10} \cdot (n-1) \leq (1+\delta)n^2/10$ pairs of vertices are both connected to at least 5 vertices of B. Hence, at most $(1+\delta)n^2/10$ pairs of vertices $\{a_1, a_2\}$ are such that $\mathcal{K}(\{a_1, a_2\}) > \binom{4}{2} = 6$. ◀

In the next lemma, we show that our resulting graph H contains $\Omega(n^2)$ copies of $K_{2,2}$.

▶ **Lemma 8.** *With high probability, the number of copies of $K_{2,2}$ in H is $|\mathcal{H}| = \Omega(n^2)$.*

Proof. By Lemma 6, we have $|\mathcal{K}| \geq \frac{9}{40}(n-1)^2$ with high probability. Let $\mathcal{K}' \subseteq \mathcal{K}$ be the subset of sets $\{a_1, a_2, b_1, b_2\}$ with $\mathcal{K}(\{a_1, a_2\}) \leq 6$ and $\mathcal{K}(\{b_1, b_2\}) \leq 6$. By Lemma 7, with high probability, $|\mathcal{K}'| \geq |\mathcal{K}| - 2 \cdot (1+\delta)n^2/10$, for any small constant δ.

Let $\mathcal{K}'' \subseteq \mathcal{K}'$ be the subset of sets $\{a_1, a_2, b_1, b_2\}$ with $|\{a_1, a_2\} \cap A'| = |\{b_1, b_2\} \cap B'| = 1$. Observe that every set $X \in \mathcal{K}'$ is included in $\mathcal{K}''$ with probability $\frac{1}{4}$. Thus, by a Chernoff bound, $|\mathcal{K}''| \geq |\mathcal{K}'|/8$ with high probability.

We argue next that the insertion of any set $K \in \mathcal{K}'$ can block at most $2 \cdot 6^2 = 72$ other sets of $\mathcal{K}'$ from being inserted into $\mathcal{H}$. Consider thus a set $K = \{a_1, a_2, b_1, b_2\} \in \mathcal{K}'$ that is added to $\mathcal{H}$. This inserts at most six pairs $\{a_3, a_4\}$ into F_A and six pairs $\{b_3, b_4\}$ into F_B, since $\mathcal{K}(\{a_1, a_2\}) \leq 6$ and $\mathcal{K}(\{b_1, b_2\}) \leq 6$. Since each pair in F_A or in F_B can block at most another six sets of $\mathcal{K}'$, overall at most $2 \cdot 6^2 = 72$ sets of $\mathcal{K}'$ can be blocked by the insertion of K into $\mathcal{H}$.

Hence:

$$|\mathcal{H}| \geq \frac{|\mathcal{K}''|}{72} \geq \frac{|\mathcal{K}'|}{8 \cdot 72} \geq \frac{(|\mathcal{K}| - 2 \cdot (1+\delta)n^2/10)}{8 \cdot 72} \geq \frac{(\frac{9}{40}(n-1)^2 - (1+\delta)n^2/5)}{8 \cdot 72} = \Omega(n^2) \ ,$$

for $\delta < \frac{1}{8}$. ◀

With Lemmas 6–8 at hand, we are now ready to complete the analysis and show that the graph H fulfills Definition 2 of a lower bound graph.

▶ **Theorem 9.** *With high probability, the output of Algorithm 1 is a* $(\Omega(n^2), \mathcal{O}(n^{3/2}))$*-lower-bound graph. In particular, for every natural n, there exists a* $(\Omega(n^2), \mathcal{O}(n^{3/2}))$*-lower-bound graph.*

Proof. We need to check that all items of Definition 2 are fulfilled with $p = \frac{1}{\sqrt{n}}$. Concerning Item 1, observe that graph G has $\mathcal{O}(n^2p) = \mathcal{O}(n^{3/2})$ edges with high probability (by a Chernoff bound).

For each $K \in \mathcal{H}$, let E_K denote the edge set added to graph H as in Step 3 of the algorithm. Item 2 holds, since $E(H) = \bigcup_{K\in\mathcal{H}} E_K$, and $H[E_K]$ is isomorphic to $K_{2,2}$, for every K, and by Lemma 8.

Concerning Item 3, observe that when $K = \{a_1, a_2, b_1, b_2\}$ is inserted into $\mathcal{H}$, then every $\{a_1, a_2, b_3, b_4\}$ such that $G[\{a_1, a_2, b_3, b_4\}]$ is isomorphic to $K_{2,2}$ will not be inserted at a later stage, since $\{b_3, b_4\}$ is inserted into F_B. For the same reason, every $\{a_3, a_4, b_1, b_2\}$ such that $G[\{a_3, a_4, b_1, b_2\}]$ is isomorphic to $K_{2,2}$ will not be inserted into $\mathcal{H}$. This proves Item 3.

Concerning Item 4, observe that for every $\{a_1, a_2, b_1, b_2\}$ that is included in $\mathcal{H}$, we have $|\{a_1, a_2\}\cap A'| = |\{b_1, b_2\}\cap B'| = 1$. Hence, H_A and H_B as defined in Item 4 are bipartite. ◀

4 Two-party communication protocol for listing all cliques

We consider a two-party communication protocol in the vertex partition model for listing all cliques (of all sizes) in a given graph. The input consists of an undirected graph $G = (V, E)$ with an arbitrary vertex partition $V = V_A \,\dot\cup\, V_B$. Let $\mathcal{C}$ be the (V_A, V_B)-cut, E_A be the edge set of $G[V_A]$, and E_B be the edge set of $G[V_B]$. We consider a scenario where Alice is given the subgraph $G_A = (V, E_A \cup \mathcal{C}) \subseteq G$ and Bob is given $G_B = (V, E_B \cup \mathcal{C}) \subseteq G$. The objective is for Alice and Bob to detect all cliques (of all sizes) of G and to minimize the number of bits communicated.

We show that in such framework, there is a two-party communication protocol for listing all cliques (of all sizes) that uses $\mathcal{O}(\sqrt{n}\,|\mathcal{C}|)$ bits of communication, where $\mathcal{C}$ are the edges shared by Alice and Bob. This shows that we cannot improve our lower bounds for the K_ℓ-detection problem, for $\ell = \mathcal{O}(\sqrt{n})$, in the CONGEST model (cf. Theorem 5) using the two-party communication framework in the vertex partition model.

Observe that without any communication between the two players, Alice can detect every clique that contains at most one vertex of V_B, and, similarly, Bob can detect every clique that contains at most one vertex of V_A (in particular, listing all triangles does not require any communication). Our task is hence to detect every clique consisting of at least two V_A vertices and at least two V_B vertices. We consider two cases:

1. Suppose that $|\mathcal{C}| \geq n^{3/2}$. Then Alice sends all edges E_A to Bob by encoding all entries in the adjacency matrix of $G[V_A]$, which requires at most $n^2 \leq \sqrt{n}|\mathcal{C}|$ bits. Since Bob then knows the entire graph G, he can detect all cliques.
2. Suppose that $|\mathcal{C}| < n^{3/2}$. For any vertex $v \in V$, let d_v be the number of edges of $\mathcal{C}$ incident to v, let $V_{\leq\sqrt{n}} \subseteq \{v \in V_A : d_v \leq \sqrt{n}\}$, and let $V_{>\sqrt{n}} = V_A \setminus V_{\leq\sqrt{n}}$. We first show how to detect every clique that contains at least one vertex of $V_{\leq\sqrt{n}}$. Then, we show how to detect every clique that does not contain any vertex of $V_{\leq\sqrt{n}}$.
   a. For every $v \in V_{\leq\sqrt{n}}$, Bob sends the induced subgraph $G_B[\Gamma_G(v) \cap V_B]$ (its adjacency matrix) to Alice (observe that Bob knows the set $V_{\leq\sqrt{n}}$ without communication). This requires at most $\sqrt{n}\,|\mathcal{C}|$ bits, since

$$\sum_{v\in V_{\leq\sqrt{n}}} d_v^2 \leq \sqrt{n} \sum_{v\in V_{\leq\sqrt{n}}} d_v \leq \sqrt{n}\,|\mathcal{C}| \enspace .$$

Alice can thus detect any clique that contains at least one vertex of $V_{\leq\sqrt{n}}$.

b. Observe that $|V_{>\sqrt{n}}| \leq \frac{|\mathcal{C}|}{\sqrt{n}}$. Alice sends the entire subgraph $G_A[V_{>\sqrt{n}}]$ (again, its adjacency matrix) to Bob. This requires at most $\sqrt{n}\,|\mathcal{C}|$ bits, since

$$|V_{>\sqrt{n}}|^2 \leq \left(\frac{|\mathcal{C}|}{\sqrt{n}}\right)^2 \leq |\mathcal{C}| \cdot \frac{|\mathcal{C}|}{n} \leq \sqrt{n}|\mathcal{C}| \ ,$$

using the assumption $|\mathcal{C}| \leq n^{3/2}$. Bob can thus detect every clique that does not contain any vertex of $V_{\leq\sqrt{n}}$.

We thus obtain the following theorem:

▶ **Theorem 10.** *There is a two-party communication protocol in the vertex partition model for listing all cliques (of all sizes) that uses $\mathcal{O}(\sqrt{n}\,|\mathcal{C}|)$ communication rounds, where $\mathcal{C}$ is the set of shared edges between Alice and Bob.*

5 Conclusions

In this paper, we give the first non-trivial lower bound for the problem of detecting a clique K_ℓ, for $\ell \geq 4$, in the classical distributed CONGEST model. We show that detecting K_ℓ requires $\Omega(\frac{n}{(\ell+\sqrt{n})\,\mathfrak{b}})$ communication rounds, for every $\ell \geq 4$, where $\mathfrak{b}$ is the bandwidth of the communication links. Our lower bound is complemented by a matching upper bound obtained by a two-party communication protocol in the vertex partition model for listing all cliques of all sizes. This demonstrates that our lower bound cannot be improved using the two-party communication framework.

We leave as a great open question whether the complexity of clique detection in the CONGEST model is sublinear, or one needs $\widetilde{\Theta}(n)$ communication rounds to detect even a copy of K_4. Since the two-party communication approach used in our lower bound cannot be improved further, we do not have any intuition whether the lower bound is tight, or could be improved significantly. On the other hand, the very recent $\widetilde{\mathcal{O}}(\sqrt{n})$-communication rounds algorithm for detecting a triangle [5] raises some hopes that maybe also K_4 could be detected in a sublinear number of rounds.

References

1. Zvika Brakerski and Boaz Patt-Shamir. Distributed discovery of large near-cliques. *Distributed Computing*, 24(2):79–89, 2011.
2. Keren Censor-Hillel, Eldar Fischer, Gregory Schwartzman, and Yadu Vasudev. Fast distributed algorithms for testing graph properties. In *Proceedings of the 30th International Symposium on Distributed Computing (DISC)*, pages 43–56, 2016.
3. Keren Censor-Hillel, Petteri Kaski, Janne H. Korhonen, Christoph Lenzen, Ami Paz, and Jukka Suomela. Algebraic methods in the congested clique. In *Proceedings of the 35th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 143–152, 2015.
4. Keren Censor-Hillel, Seri Khoury, and Ami Paz. Quadratic and near-quadratic lower bounds for the CONGEST model. In *Proceedings of the 31st International Symposium on Distributed Computing (DISC)*, pages 10:1–10:16, 2017.
5. Yi-Jun Chang, Seth Pettie, and Hengjie Zhang. Distributed triangle detection via expander decomposition. *CoRR*, abs/1807.06624, 2018. `arXiv:1807.06624`.

6 Danny Dolev, Christoph Lenzen, and Shir Peled. “Tri, tri again”: Finding triangles and small subgraphs in a distributed setting. In *Proceedings of the 26th International Symposium on Distributed Computing (DISC)*, pages 195–209, 2012.

7 Andrew Drucker, Fabian Kuhn, and Rotem Oshman. On the power of the congested clique model. In *Proceedings of the 33rd Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 367–376, 2014.

8 Guy Even, Orr Fischer, Pierre Fraigniaud, Tzlil Gonen, Reut Levi, Moti Medina, Pedro Montealegre, Dennis Olivetti, Rotem Oshman, Ivan Rapaport, and Ioan Todinca. Three notes on distributed property testing. In *Proceedings of the 31st International Symposium on Distributed Computing (DISC)*, pages 15:1–15:30, 2017.

9 Orr Fischer, Tzlil Gonen, Fabian Kuhn, and Rotem Oshman. Possibilities and impossibilities for distributed subgraph detection. In *Proceedings of the 30th on Symposium on Parallelism in Algorithms and Architectures*, (SPAA), pages 153–162, New York, NY, USA, 2018. ACM.

10 Pierre Fraigniaud and Dennis Olivetti. Distributed detection of cycles. In *Proceedings of the 29th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 153–162, 2017.

11 Tzlil Gonen and Rotem Oshman. Lower bounds for subgraph detection in the CONGEST model. In *Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS)*, pages 6:1–6:16, 2017.

12 Taisuke Izumi and François Le Gall. Triangle finding and listing in CONGEST networks. In *Proceedings of the 37th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 381–389, 2017.

13 Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. John Wiley & Sons, 2011.

14 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

15 Janne H. Korhonen and Joel Rybicki. Deterministic subgraph detection in broadcast CONGEST. In *Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS)*, pages 4:1–4:16, 2017.

16 Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

17 Gopal Pandurangan, Peter Robinson, and Michele Scquizzato. Tight bounds for distributed graph computations. *CoRR*, abs/1602.08481, 2016.

18 David Peleg. *Distributed Computing: A Locality-Sensitive Approach*. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia, PA, 2000.