



# Detecting cliques in CONGEST networks

Artur Czumaj<sup>1</sup> · Christian Konrad<sup>2</sup>

Received: 29 December 2018 / Accepted: 9 December 2019 / Published online: 21 December 2019  
© The Author(s) 2019

## 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  $K_\ell$  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  $K_\ell$  requires  $\Omega(\sqrt{n}/b)$  communication rounds, for every  $4 \leq \ell \leq \sqrt{n}$ , and  $\Omega(n/(\ell b))$  rounds for every  $\ell \geq \sqrt{n}$ , where  $b$  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.

**Keywords** CONGEST model · Subgraph detection · Lower bounds · Two-party communication · Set-disjointness

## 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  $K_\ell$  for  $\ell \geq 4$ .

The nowadays classical distributed *CONGEST model* (see, e.g., [21]) 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)  $G = (V, E)$  with  $n = |V|$  nodes, where network nodes execute distributed algorithms in synchronous rounds, and the nodes collaborate to solve a graph problem with input  $G$ . Further, every node has a unique identifier from  $\{0, \dots, \text{poly}(n)\}$ . In any single round, all nodes can:

- (i) perform an unlimited amount of local computation,
- (ii) send a possibly different  $b$ -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  $b = \mathcal{O}(\log n)$ ; we note though that our analysis generalizes to other settings of  $b$  in a straightforward manner. (We note that in our lower bound for detecting  $K_4$  and  $K_\ell$  in Sect. 2, to ensure full generality of presentation, we will make the analysis parameterized by the message size  $b$ , in which case we will refer to such model of distributed computation as  $\text{CONGEST}_b$ , the CONGEST model with messages of size  $b$ .)

---

Research partially supported by the Centre for Discrete Mathematics and its Applications (DIMAP), by EPSRC Award EP/D063191/1, and by EPSRC award EP/N011163/1. Most of work on this paper was carried out while C.K. was at the University of Warwick.

---

✉ Christian Konrad  
christian.konrad@bristol.ac.uk

Artur Czumaj  
A.Czumaj@warwick.ac.uk

<sup>1</sup> Department of Computer Science and Centre for Discrete Mathematics and its Applications (DIMAP), University of Warwick, Coventry, UK

<sup>2</sup> Department of Computer Science, University of Bristol, Bristol, UK

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

- (i) if  $G$  contains a copy of  $K_\ell$ , then with high probability<sup>1</sup> at least one node outputs 1, and
- (ii) if  $G$  does not contain any copy of  $K_\ell$ , then with high probability no node outputs 1.

Since standard success probability amplification techniques cannot easily be applied for the subgraph detection problem in the CONGEST model, our problem definition requires algorithms to succeed with high probability. The lower bounds given in this paper however also apply to algorithms that succeed with only constant probability (e.g.,  $\frac{2}{3}$ ).

The subgraph detection problem is a local problem: it can be solved efficiently solely on the basis of local information. In particular, in the CONGEST model, the problem of finding  $K_\ell$  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 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  $K_\ell$  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  $K_\ell$  in the  $\text{CONGEST}_b$  model, for  $\ell \geq 4$ . In Theorem 4, we prove that every algorithm in the  $\text{CONGEST}_b$  model that with probability at least  $\frac{2}{3}$  detects  $K_\ell$ , for  $\ell \geq 4$  and  $\ell = \mathcal{O}(\sqrt{n})$ , requires  $\Omega\left(\frac{\sqrt{n}}{b}\right)$  rounds. Further, if  $\ell = \omega(\sqrt{n})$ , then  $\Omega\left(\frac{n}{\ell b}\right)$  rounds are required. We are not aware of any other non-trivial (super-constant) lower bound for this problem in the  $\text{CONGEST}_b$  model.

We complement our lower bound with a two-party communication protocol for listing all cliques in the input graph (see Theorem 6), 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

<sup>1</sup> We say that an event occurs *with high probability* (in short w.h.p.) if the probability of it happening is at least  $1 - \frac{1}{n}$ .

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 4 and the upper bound in Theorem 6, 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. [19]) in which two players, Alice and Bob, each have some private input  $X$  and  $Y$ . The players' goal is to compute a function  $f(X, Y)$ , and the complexity measure used is the number of bits Alice and Bob exchange to compute  $f(X, Y)$ . 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

$$\text{DISJ}_n(X, Y) := \bigvee_{i=1}^n X_i \wedge Y_i .$$

In a seminal work, Kalyanasundaram and Schnitger [17] 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** [17] *The randomized two-party communication complexity of set disjointness is  $\Omega(n)$ . That is, for any constant  $p > \frac{1}{2}$ , any randomized two-party communication protocol that computes  $\text{DISJ}_n(X, Y)$  with probability at least  $p$ , has two-party communication complexity  $\Omega(n)$ .*

Our main result, the lower bound for detecting  $K_\ell$  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., [5, 9, 12, 14, 18]. 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 *static vertex partition model*. A non-static vertex partition model was considered for example in [22] and will be

discussed further below. From now on, we refer to the static vertex partition model simply by 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  $\text{CONGEST}_b$  model, one wants to construct  $G$  to ensure that

- it has some property (in our case, contains a copy of  $K_\ell$ ) if and only if the corresponding instance of set disjointness is such that  $\text{DISJ}_n(X, Y) = 0$ , 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  $\varepsilon$ , one can apply Theorem 1 to argue that the round complexity of any distributed algorithm in the  $\text{CONGEST}_b$  model for a given problem is  $\Omega(\frac{\varepsilon}{|\mathcal{C}| \cdot b})$ . The central challenge is to ensure that for the encoded set disjointness instance of size  $\varepsilon$  and the cut of size  $|\mathcal{C}|$ , the ratio  $\frac{\varepsilon}{|\mathcal{C}|}$  is as large as possible.

For example, Drucker et al. [9] incorporated a similar approach to obtain a lower bound for the subgraph detection problem in a *broadcast* variant of the  $\text{CONGEST}_b$  model (in fact, even for a (stronger) broadcast variant of the  $\text{CONGESTED CLIQUE}$  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  $\text{CONGEST}_b$  model every node is required to send the same message via all incident edges, at most  $\mathcal{O}(n \cdot b)$  bits can be transmitted through the cut, yielding a lower bound of  $\Omega(\frac{n}{b})$ . (In particular, for the *broadcast* variant of the  $\text{CONGEST}_b$  model, Drucker et al. [9, Theorem 15] proved that detecting a clique  $K_\ell$ ,  $\ell \geq 4$ , requires  $\Omega(\frac{n}{b})$  rounds.) Note however that in the (non-broadcast)  $\text{CONGEST}_b$  model, this construction does not give any non-trivial bound, since  $\frac{\varepsilon}{|\mathcal{C}|} = \mathcal{O}(1)$ .

The main building block for our lower bound is the construction of  $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graphs (see Sect. 3.1 for the precise definition) that can be used to encode a set disjointness instance of size  $\varepsilon = \Omega(n^2)$  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  $\Omega(\frac{\sqrt{n}}{b})$  for the subgraph detection problem in the  $\text{CONGEST}_b$  model for the clique  $K_4$ . This construction can also be extended to detect larger cliques, yielding the lower bound of  $\Omega(\frac{n}{(\ell + \sqrt{n}) \cdot b})$  for detecting any  $K_\ell$  with  $\ell \geq 4$ .

Since these are the first superconstant lower bounds for detecting a clique (with  $\ell \geq 4$ ) in the  $\text{CONGEST}$  model and

since only very recently we have seen that  $K_3, K_4, K_5$  can be detected in  $o(n)$  rounds [6,7,10,15], 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  $\Omega(\frac{\sqrt{n}}{b})$  communication rounds to detect any clique  $K_\ell$  (with  $\ell \geq 4, \ell = \mathcal{O}(\sqrt{n})$ ) in the  $\text{CONGEST}_b$  model, or maybe we need substantially more 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 5 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 Sect. 4 (Theorem 6) 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 communicates  $\mathcal{O}(\sqrt{n} |\mathcal{C}|)$  bits, where  $\mathcal{C}$  is the set of shared edges between Alice and Bob. This shows that we cannot obtain stronger lower bounds for the  $K_\ell$ -detection problem, for  $\ell = \mathcal{O}(\sqrt{n})$ , in the  $\text{CONGEST}$  model using the two-party communication framework in the vertex partition model.

In [22], a non-static version of the two-party vertex partition model was considered for proving lower bounds in the  $\text{CONGEST}$  model for problems such as minimum spanning tree. In the non-static version, the partitioning of the vertex set between Alice and Bob evolves as the algorithm progresses. Our two-party communication protocol shows that our lower bound in the static vertex partitioning model is optimal up to constant factors. While we do not believe that stronger lower bounds for the clique detection problem can be proved in a non-static vertex partition model, the existence of our two-party communication protocol does not rule out this possibility.

### 1.3 Related works

As a fundamental primitive in networks analysis, subgraph detection and listing in the  $\text{CONGEST}$  model has been recently receiving attention from multiple authors, focusing mainly on randomized complexity. However, despite major

efforts, until very recently relatively little has been known about the complexity of the subgraph detection problem.

For a very long time we did not know whether one can detect any  $K_\ell$  in a sublinear number of rounds in the CONGEST model. In a recent breakthrough in this area, Izumi and Le Gall [15] considered the subgraph detection problem for the smallest interesting subgraph  $H$ , the triangle  $K_3$ , and showed that one can detect a triangle in  $\tilde{O}(n^{2/3})$  rounds in the CONGEST model. Further, they also showed that the related problem of finding all triangles (triangle listing) can be solved in  $\tilde{O}(n^{3/4})$  rounds. Very recently, these results were improved by Chang et al. [7] and then by Chang and Saranurak [6], who showed that both triangle detection and listing can be solved in  $\tilde{O}(n^{1/2})$  and  $\tilde{O}(n^{1/3})$  rounds, respectively.

Regarding lower bounds for  $K_3$ , it is known that randomized single round algorithms for triangle detection require messages of size  $\Omega(\Delta)$  [12], and deterministic ones require messages of size  $\Omega(\Delta \log n)$  [1]. No non-trivial lower bound on the number of rounds for the triangle detection problem is known in the CONGEST<sub>b</sub> model, for any  $b \geq 2$ , though it is known (cf. [15,20]) that the more complex triangle listing problem requires  $\Omega(n^{1/3}/\log n)$  rounds in both the CONGEST and the CONGESTED CLIQUE models. 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 [15, Proposition 4.4].

Before our paper has been made available, no sublinear rounds CONGEST algorithms for detecting or listing cliques  $K_\ell$  have been known for any  $\ell \geq 4$ . While there is a trivial lower bound of a constant number of rounds and 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. However, very recently, building on the ideas from Chang et al. [7], Eden et al. [10] presented the first sublinear rounds algorithms for the next two smallest cliques,  $K_4$  and  $K_5$ . They gave randomized algorithms that detect and list copies of  $K_4$  and  $K_5$  in  $\mathcal{O}(n^{5/6+o(1)})$  and  $\mathcal{O}(n^{21/22+o(1)})$  rounds, respectively.

While rather disappointingly, we do not know how to extend any of these upper bounds to other cliques  $K_\ell$  with  $\ell \geq 6$ , the previously mentioned works for triangle detection raise hope that detecting cliques  $K_\ell$  could potentially be solved in a sublinear number of rounds for all  $\ell \geq 3$ . Furthermore, 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 [15,20] holds only for a more complex problem of detecting *all triangles*).

Even et al. [11] noted that the problem of detecting trees is significantly simpler and designed a randomized color-

coding algorithm that detects any constant-size *tree* on  $\ell$  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  $\ell$ -cycle  $C_\ell$  in the CONGEST model [9], where it has been shown that detecting  $C_\ell$  requires  $\Omega(\text{ex}(n, C_\ell)/\log n)$  rounds, where  $\text{ex}(n, C_\ell)$  is the Turán number for cycles, that is, the largest possible number of edges in a  $C_\ell$ -free graph over  $n$  vertices. In particular, for odd-length cycles (of length 5 or more), the lower bound of [9] is  $\Omega(n/\log n)$ , and it is  $\Omega(\sqrt{n}/\log n)$  for  $\ell = 4$ . Very recently, Korhonen and Rybicki [18] improved the lower bound for all even-length cycles to  $\Omega(\sqrt{n}/\log n)$ . Further, Gonen and Oshman [14] extended these lower bounds for  $C_\ell$ -freeness to some related classes of graphs, though still with some cyclic underlying structure. (As mentioned above, we note that Drucker et al. [9] 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. [9] proved that detecting a clique  $K_\ell$ ,  $\ell \geq 4$ , 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. [12], 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  $\Omega(n^{2-\frac{1}{\ell}}/b)$  in the CONGEST model, where  $b$  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. [9] designed an  $\mathcal{O}(\sqrt{n})$  round algorithm for  $C_4$  detection, and Even et al. [11] and Korhonen and Rybicki [18] obtained path and tree detection algorithms requiring only a constant number of rounds. Korhonen and Rybicki [18] 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. [8] and Censor-Hillel et al. [4].

We summarize earlier and new results in Table 1.

### 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

**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 do not imply any bounds for CONGEST)

| Paper     | Time bound                               | Problem  | Model                      |
|-----------|--|--|----------------------------|
| [11]      | $\mathcal{O}(\ell^\ell)$                 | Detecting a tree on $\ell$ nodes               | CONGEST                    |
| folklore  | $\mathcal{O}(n)$                         | Detecting $K_\ell, \ell \geq 3$                | CONGEST                    |
| [15]      | $\tilde{\mathcal{O}}(n^{2/3})$           | Detecting triangle ( $K_3$ )                   | CONGEST                    |
| [15]      | $\tilde{\mathcal{O}}(n^{3/4})$           | Triangle listing ( $K_3$ )                     | CONGEST                    |
| [7]       | $\tilde{\mathcal{O}}(n^{1/2})$           | Triangle detection and listing ( $K_3$ )       | CONGEST                    |
| [6]       | $\tilde{\mathcal{O}}(n^{1/3})$           | Triangle detection and listing ( $K_3$ )       | CONGEST                    |
| [10]      | $\mathcal{O}(n^{5/6+o(1)})$              | $K_4$ detection and listing                    | CONGEST                    |
| [10]      | $\mathcal{O}(n^{21/22+o(1)})$            | $K_5$ detection and listing                    | CONGEST                    |
| [12]      | $\Omega(n^{2-\frac{1}{\ell}} / \log n)$  | Detecting some $H$ of size $\mathcal{O}(\ell)$ | CONGEST                    |
| [9]       | $\Omega(n / \log n)$                     | Detecting $C_\ell, \ell \geq 5, \ell$ odd      | CONGEST                    |
| [9,18]    | $\Omega(\sqrt{n} / \log n)$              | Detecting $C_\ell, \ell \geq 4, \ell$ even     | CONGEST                    |
| [15,20]   | $\Omega(n^{1/3} / \text{poly-log}(n))$   | Triangle listing ( $K_3$ )                     | CONGESTED CLIQUE           |
| [9]       | $\Omega(n / \log n)$                     | Detecting $K_\ell$ for $\ell \geq 4$           | Broadcast CONGESTED CLIQUE |
| Theorem 4 | $\Omega(n / ((\sqrt{n} + \ell) \log n))$ | Detecting $K_\ell$ for $\ell \geq 4$           | CONGEST                    |

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., [2,11]). 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)  $\varepsilon$ -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., [4,11] for more details). Property testing of  $H$ -freeness in the CONGEST model has received a lot of attention lately (see, e.g., [2,3,11–13]). In particular, it has been shown [11] 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  $\ell \geq 5$ ,  $K_\ell$ -freeness can be tested in  $\mathcal{O}((\varepsilon \cdot |E(G)|)^{\frac{1}{2}-\frac{1}{\ell-2}} / \varepsilon)$  rounds [11]. For trees, testing if the input graph is  $T$ -free for a tree  $T$  on  $\ell$  vertices can be done in  $\mathcal{O}(\ell^{1+\ell^2} / \varepsilon^\ell)$  rounds in the CONGEST model [11].

### 1.4 Outline

We begin in Sect. 2.1 with a definition of lower-bound graphs and then, in Sects. 2.2, 2.3, we show how to combine lower-bound graphs and the lower bound for set disjointness to prove the hardness of clique detection. A construction of  $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graphs is given in Sect. 3.

Section 4 provides our upper bound, a two-party communication protocol in the vertex partition model for listing all cliques. Section 5 gives some final conclusions.

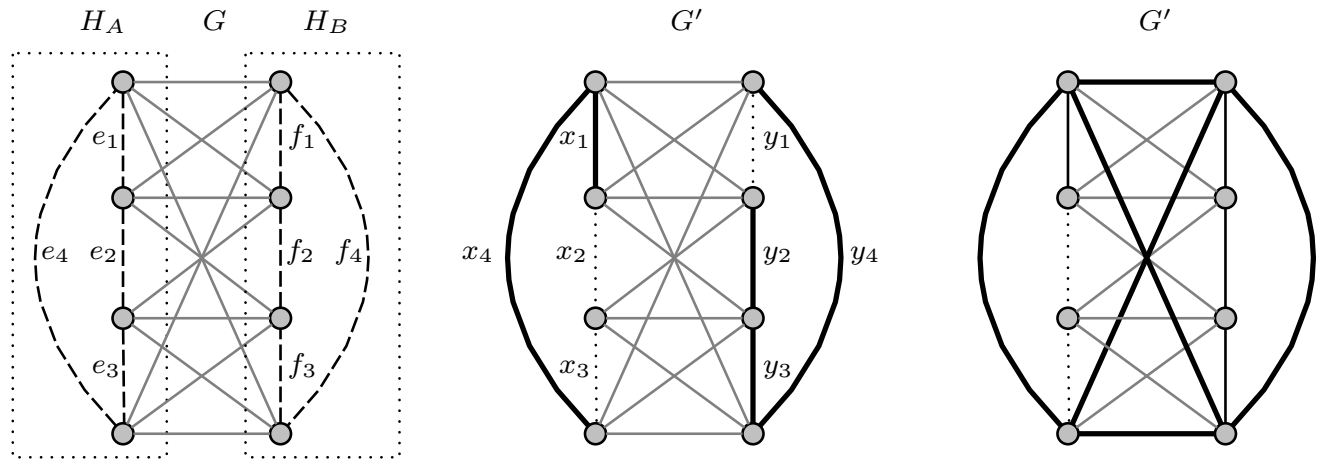
## 2 Lower bound (clique detection needs $\tilde{\Omega}(\sqrt{n})$ rounds)

In this section we prove our hardness results showing that any algorithm in the  $\text{CONGEST}_b$  model that detects a  $K_\ell$  with probability at least  $\frac{2}{3}$  requires  $\Omega(\sqrt{n}/b)$  rounds, for every  $\ell = \mathcal{O}(\sqrt{n})$  and  $\ell \geq 4$ , and requires  $\Omega(\frac{n}{\ell b})$  rounds if  $\ell = \omega(\sqrt{n})$  (Theorems 3 and 4); or in short,  $\Omega(\frac{n}{(\ell+\sqrt{n})b})$  rounds, for every  $\ell \geq 4$ . Our lower bound for the complexity of detecting  $K_\ell$  in the CONGEST model relies on a reduction to the two-party communication complexity lower bound for the set disjointness problem (cf. Theorem 1 in Sect. 1.2), which we implement with the help of lower-bound graphs (cf. Sect. 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. Fig. 1).

**Definition 1** 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  $|E| \leq m$  and there exist bipartite graphs  $H_A = (A, E_A)$  and  $H_B = (B, E_B)$  with  $E_A =$



**Fig. 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  $1 \leq i \leq 4$ , observe that the graph  $G \cup \{e_i, f_i\}$  contains one  $K_4$  consisting of the edges  $e_i, f_i$  and the edges of the subgraph of  $G$  induced by the vertices incident to  $e_i$  and  $f_i$  (which forms a  $K_{2,2}$ ). For every  $\{e_1, \dots, e_k\}, E_B = \{f_1, \dots, f_k\}$ , and  $|E_A| = |E_B| = k$  on vertex sets  $A$  and  $B$ , respectively, so that:

1. The graph  $G \cup \{e_i, f_i\}$  contains a  $K_4$ , for every  $1 \leq i \leq k$ , and
2. the graph  $G \cup \{e_i, f_j\}$  does not contain a  $K_4$ , for every  $1 \leq i, j \leq k$  with  $i \neq j$ .

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

**Proof** Let  $\mathcal{A}$  be an algorithm in the  $CONGEST_b$  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, and let  $H_A = (A, E_A)$  and  $H_B = (B, E_B)$  with  $E_A = \{e_1, \dots, e_k\}$  and  $E_B = \{f_1, \dots, f_k\}$  be the associated graphs to  $G$  as in Definition 1. Alice constructs the set  $E'_A \subseteq E_A$  such that for every  $i$  with  $X_i = 1$ , the edge  $e_i$  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  $f_i$  is included in  $E'_B$ .

$i \neq j$ , the graph  $G \cup \{e_i, f_j\}$  does not contain a  $K_4$ . *Center* Graph  $G'$  as in the proof of Theorem 2 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 0. *Right* The highlighted edges form a  $K_4$

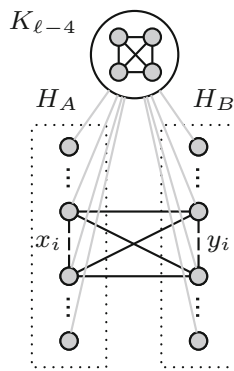
We first argue that the graph  $G' := G \cup (E'_A \cup E'_B)$  contains a  $K_4$  if and only if  $DISJ_n(X, Y) = 0$ . Indeed, since by Definition 1, the graphs  $H_A$  and  $H_B$  are bipartite (and thus the subgraphs  $G'[A]$  and  $G'[B]$  are bipartite too), any copy of  $K_4$  in  $G'$  must consist of two vertices from  $A$  and two vertices from  $B$ .

Suppose first that  $G'$  contains a  $K_4$  and let  $a_1, a_2 \in A$  and  $b_1, b_2 \in B$  be the vertices incident to this  $K_4$ . Since  $a_1$  and  $a_2$  are connected, this implies that  $a_1, a_2$  are the endpoints of an edge from  $E_A$ . Let  $e_i \in E_A$  be this edge. Furthermore, since  $b_1$  and  $b_2$  are connected,  $b_1, b_2$  are necessarily the endpoints of an edge from  $E_B$ . Let  $f_j \in E_B$  be this edge. Since  $G$  is a lower-bound graph, by Definition 1 we obtain that  $i = j$ . Hence, since Alice and Bob included  $e_i$  and  $f_j = f_i$  in  $G'$ , we have  $X_i = Y_i = 1$  and thus  $DISJ_n(X, Y) = 0$ .

Next, suppose that  $G'$  does not contain a  $K_4$ . Then, for every  $1 \leq i \leq k$ , Alice and Bob have not both included the edges  $e_i$  and  $f_i$  (since otherwise there would be a  $K_4$ ). This implies that for every  $1 \leq i \leq k$ ,  $X_i \wedge Y_i = 0$  holds and thus  $DISJ_n(X, Y) = 1$ .

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 at most  $m$ , Alice and Bob exchange messages with overall at most  $mb$  bits per round. Thus, overall they communicate at most  $rm$  bits. Since the algorithm allows them to solve set disjointness, by Theorem 1, we have  $rm = \Omega(k)$ . Thus,  $\mathcal{A}$  requires  $\Omega\left(\frac{k}{mb}\right)$  rounds.  $\square$

**Fig. 2** Extension of our lower bound for  $K_4$  detection to  $K_\ell$  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  $\ell$  vertices if and only if the encoded set disjointness instance evaluates to 0, i.e.,  $x_i = y_i = 1$ , for some  $i$



In Theorem 5 in Sect. 3, we prove the existence of a  $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graph. By combining Theorem 5 with Theorem 2, we obtain the following main result.

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

### 2.3 Detection of $K_\ell$ for $\ell \geq 5$

The lower bound construction given in Theorem 2 can be extended to the task of detecting  $K_\ell$ , for  $\ell \geq 5$  (see also Fig. 2). To this end, we add a clique on  $\ell - 4$  new nodes to graph  $G'$  (from the proof of Theorem 2) 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  $\Omega(\frac{n^2}{n(\ell-4)b}) = \Omega(\frac{n}{\ell b})$ . Similarly as before, the encoded set disjointness instance evaluates to 0 if and only if  $G'$  contains a clique of size  $\ell$ . We thus conclude with the following theorem.

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

## 3 Lower-bound graph construction

In this section, we construct our main technical tool and prove the existence of a  $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graph, see Definition 1. We will show in Theorem 5 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 $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graphs

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 1 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  $G$ .

**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  $\{a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2\}$ .

**return**  $H := (A, B, \bigcup_{K \in \mathcal{H}} E_K)$ .

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 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  $K_{2,2}$  later on. 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 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 2 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 1 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., [16, 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 1), that only a small fraction of pairs of  $A$  vertices are contained in more than six copies of  $K_{2,2}$  (Lemma 2), and finally that the resulting graph  $H$  contains  $\Omega(n^2)$  copies of  $K_{2,2}$  (Lemma 3). With these three claims at hand, we will complete the analysis to prove in Theorem 5 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 1** *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:

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

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

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

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  $t_0 + t_1 + t_{2,1} + t_{2,2} + t_3 + t_4 = \binom{n}{2}^4$ . We thus obtain:

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

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

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

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 2** *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  $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots \cup \mathcal{X}_{n-1}$ , 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  $\mu = \frac{n}{10}$ ):

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

for any constant  $\delta$ . 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  $\mathcal{K}(\{a_1, a_2\}) > \binom{4}{2} = 6$ .  $\square$

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

**Lemma 3** *With high probability, the number of copies of  $K_{2,2}$  in  $H$  is  $|\mathcal{H}| = \Omega(n^2)$ .*

**Proof** By Lemma 1, 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 2, with high probability,  $|\mathcal{K}'| \geq |\mathcal{K}| - 2 \cdot (1 + \delta)n^2/10$ , for any small constant  $\delta$ .

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:

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

for  $\delta < \frac{1}{8}$ .  $\square$

With Lemmas 1–3 at hand, we are now ready to complete the analysis and show that the graph  $H$  fulfills Definition 1 of a lower bound graph.

**Theorem 5** *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  $n \in \mathbb{N}$ , there exists a  $(\Omega(n^2), \mathcal{O}(n^{3/2}))$ -lower-bound graph.*

**Proof** We need to check that the output graph  $H$  of Algorithm 1 with  $p = \frac{1}{\sqrt{n}}$  fulfills Definition 1. First, observe that graph  $G$  has  $\mathcal{O}(n^2 p) = \mathcal{O}(n^{3/2})$  edges with high probability (by a Chernoff bound), and hence  $H$  also has  $\mathcal{O}(n^{3/2})$  edges.

We now show that graphs  $H_A = (A, E_A)$  and  $H_B = (B, E_B)$  with  $E_A = \{e_1, \dots, e_k\}$  and  $E_B = \{f_1, \dots, f_k\}$  as in Definition 1 exist, where  $k = |\mathcal{H}|$ . To this end, let  $\mathcal{H} = \{K_1, K_2, \dots, K_k\}$  and for every  $K_i = \{a_1, a_2, b_1, b_2\}$ , let  $e_i = a_1 a_2$  and  $f_i = b_1 b_2$ . Observe that  $H_A$  and  $H_B$  are bipartite, since by construction every  $e_i$  connects a vertex from  $A'$  to a vertex from  $A \setminus A'$ , and every  $f_i$  connects a vertex from  $B'$  to a vertex from  $B \setminus B'$ .

Next, we show that the graphs  $H_A$  and  $H_B$  fulfill the two items of Definition 1. To this end, first observe that for every  $1 \leq i \leq k$  the graph  $G \cup \{e_i, f_i\}$  with  $e_i = a_1 a_2$  and  $f_i = b_1 b_2$  contains a  $K_4$ : Since  $K_i = \{a_1, a_2, b_1, b_2\}$ , the subgraph  $G[\{a_1, a_2, b_1, b_2\}]$  is isomorphic to  $K_{2,2}$  which in turn implies that  $G[\{a_1, a_2, b_1, b_2\}] \cup \{e_i, f_i\}$  is isomorphic to a  $K_4$ .

Next, for the sake of a contradiction, assume that there exists a  $1 \leq i, j \leq k$  with  $i < j$  (the case  $i > j$  is similar and omitted) so that the graph  $G \cup \{e_i, f_j\}$  contains a  $K_4$ . Then, by construction of Algorithm 1, when  $K_i$  was inserted into  $\mathcal{H}$ , the edge  $f_j$  was declared to be forbidden and inserted in  $F_B$ . It is thus impossible that  $K_j$  was inserted into  $\mathcal{H}$  at a later stage.

Last, by Lemma 3 we have  $k = |\mathcal{H}| = \Omega(n^2)$  which completes the proof of this theorem.  $\square$

## 4 Two-party communication protocol for listing 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  $K_\ell$ -detection problem, for  $\ell = \mathcal{O}(\sqrt{n})$ , in the CONGEST model (cf. Theorem 4) 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

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

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 6** *There is a two-party communication protocol in the vertex partition model for listing all cliques (of all sizes) that communicates  $\mathcal{O}(\sqrt{n} |\mathcal{C}|)$  bits, 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  $K_\ell$ , for  $\ell \geq 4$ , in the classical distributed CONGEST model. We show that detecting  $K_\ell$  requires  $\Omega\left(\frac{n}{(\ell + \sqrt{n}) \mathfrak{b}}\right)$  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 true complexity of  $K_\ell$  detection in the CONGEST model is  $\tilde{\Theta}(\sqrt{n})$ , for  $\ell = \mathcal{O}(\sqrt{n})$ , or one needs substantially more rounds. 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  $\tilde{\mathcal{O}}(\sqrt{n})$ -communication rounds algorithm for detecting a triangle [7] raises some hopes that maybe also  $K_4$  could be detected in  $\tilde{\mathcal{O}}(\sqrt{n})$  rounds.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abboud, A., Censor-Hillel, K., Khoury, S., Lenzen, C.: Fooling views: a new lower bound technique for distributed computations under congestion. [arXiv:1711.01623](https://arxiv.org/abs/1711.01623) (2017)
2. Brakerski, Z., Patt-Shamir, B.: Distributed discovery of large near-cliques. *Distrib. Comput.* **24**(2), 79–89 (2011)
3. Censor-Hillel, K., Fischer, E., Schwartzman, G., Vasudev, Y.: Fast distributed algorithms for testing graph properties. In: Proceedings of the 30th International Symposium on Distributed Computing (DISC), pp. 43–56 (2016)
4. Censor-Hillel, K., Kaski, P., Korhonen, J.H., Lenzen, C., Paz, A., Suomela, J.: Algebraic methods in the congested clique. In: Proceedings of the 35th Annual ACM Symposium on Principles of Distributed Computing (PODC), pp. 143–152 (2015)
5. Censor-Hillel, K., Khoury, S., Paz, A.: Quadratic and near-quadratic lower bounds for the CONGEST model. In: Proceedings of the 31st International Symposium on Distributed Computing (DISC), pp. 10:1–10:16 (2017)

6. Chang, Y., Saranurak, T.: Improved distributed expander decomposition and nearly optimal triangle enumeration. In: Proceedings of the 39th Annual ACM Symposium on Principles of Distributed Computing (PODC), pp. 66–73 (2019)
7. Chang, Y.J., Pettie, S., Zhang, H.: Distributed triangle detection via expander decomposition. In: Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 821–840 (2019)
8. Dolev, D., Lenzen, C., Peled, S.: “Tri, tri again”: finding triangles and small subgraphs in a distributed setting. In: Proceedings of the 26th International Symposium on Distributed Computing (DISC), pp. 195–209 (2012)
9. Drucker, A., Kuhn, F., Oshman, R.: On the power of the congested clique model. In: Proceedings of the 33rd Annual ACM Symposium on Principles of Distributed Computing (PODC), pp. 367–376 (2014)
10. Eden, T., Fiat, N., Fischer, O., Kuhn, F., Oshman, R.: Sublinear-time distributed algorithms for detecting small cliques and even cycles. In: Proceedings of the 33rd International Symposium on Distributed Computing (DISC), pp. 15:1–15:16 (2019)
11. Even, G., Fischer, O., Fraigniaud, P., Gonen, T., Levi, R., Medina, M., Montealegre, P., Olivetti, D., Oshman, R., Rapaport, I., Todinca, I.: Three notes on distributed property testing. In: Proceedings of the 31st International Symposium on Distributed Computing (DISC), pp. 15:1–15:30 (2017)
12. Fischer, O., Gonen, T., Kuhn, F., Oshman, R.: Possibilities and impossibilities for distributed subgraph detection. In: Proceedings of the 30th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), pp. 153–162 (2018)
13. Fraigniaud, P., Olivetti, D.: Distributed detection of cycles. In: Proceedings of the 29th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), pp. 153–162 (2017)
14. Gonen, T., Oshman, R.: Lower bounds for subgraph detection in the CONGEST model. In: Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS), pp. 6:1–6:16 (2017)
15. Izumi, T., Gall, F.L.: Triangle finding and listing in CONGEST networks. In: Proceedings of the 37th Annual ACM Symposium on Principles of Distributed Computing (PODC), pp. 381–389 (2017)
16. Janson, S., Łuczak, T., Ruciński, A.: Random Graphs. Wiley, Hoboken (2011)
17. Kalyanasundaram, B., Schnitger, G.: The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.* **5**(4), 545–557 (1992)
18. Korhonen, J.H., Rybicki, J.: Deterministic subgraph detection in broadcast CONGEST. In: Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS), pp. 4:1–4:16 (2017)
19. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press, Cambridge (1997)
20. Pandurangan, G., Robinson, P., Scquizzato, M.: Tight bounds for distributed graph computations. [arXiv:1602.08481](https://arxiv.org/abs/1602.08481) (2016)
21. Peleg, D.: Distributed Computing: A Locality-Sensitive Approach. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia (2000)
22. Sarma, A.D., Holzer, S., Kor, L., Korman, A., Nanongkai, D., Pandurangan, G., Peleg, D., Wattenhofer, R.: Distributed verification and hardness of distributed approximation. *SIAM J. Comput.* **41**(5), 1235–1265 (2012)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.