Graph Reconstruction in the Congested Clique

Pedro Montealegre* Sebastian Perez-Salazar† Ivan Rapaport†
Ioan Todinca‡

June 13, 2017

Abstract

The congested clique model is a message-passing model of distributed computation where the underlying communication network is the complete graph of n nodes. In this paper we consider the situation where the joint input to the nodes is an n-node labeled graph G, i.e., the local input received by each node is the indicator function of its neighborhood in G. Nodes execute an algorithm, communicating with each other in synchronous rounds and their goal is to compute some function that depends on G. In every round, each of the n nodes may send up to $n-1$ different b-bit messages through each of its $n-1$ communication links. We denote by R the number of rounds of the algorithm. The product Rb, that is, the total number of bits received by a node through one link, is the *cost* of the algorithm.

The most difficult problem we could attempt to solve is the *reconstruction problem*, where nodes are asked to recover all the edges of the input graph G. Formally, given a class of graphs $\mathcal{G}$, the problem is defined as follows: if $G \notin \mathcal{G}$, then every node must reject; on the other hand, if $G \in \mathcal{G}$, then every node must end up, after the R rounds, knowing all the edges of G. It is not difficult to see that the cost Rb of any algorithm that solves this problem (even with public coins) is at least $\Omega(\log |\mathcal{G}_n|/n)$, where $\mathcal{G}_n$ is the subclass of all n-node labeled graphs in $\mathcal{G}$. In this paper we prove that previous bound is tight and that it is possible to achieve it with only $R = 2$ rounds. More precisely, we exhibit (i) a one-round algorithm that achieves this bound for hereditary graph classes; and (ii) a two-round algorithm that achieves this bound for arbitrary graph classes. Later, we show that the bound $\Omega(\log |\mathcal{G}_n|/n)$ cannot be achieved in one-round for arbitrary graph classes, and we give tight algorithms for that case.

From (i) we recover all known results concerning the reconstruction of graph classes in one round and bandwidth $\mathcal{O}(\log n)$: forests, planar graphs, cographs, etc. But we also get new one-round algorithms for other hereditary graph classes such as unit disc graphs, interval graphs, etc. From (ii), we can conclude that any problem restricted to a class of graphs of size $2^{\mathcal{O}(n \log n)}$ can be solved in the congested clique model in two rounds, with bandwidth $\mathcal{O}(\log n)$. Moreover, our general two-round algorithm is valid for any set of labeled graphs, not only for graph classes (which are sets of labeled graphs closed under isomorphims).

*Facultad de Ingeniería y Ciencias, Univ. Adolfo Ibáñez, Santiago, Chile, pedro.montealegre@uai.cl
†DIM-CMM (UMI 2807 CNRS), Univ. de Chile, Santiago, Chile, {sperez,rapaport}@dim.uchile.cl
‡Univ. Orléans, INSA Centre Val de Loire, LIFO EA 4022, Orléans, France, ioan.todinca@univ-orleans.fr

1 Introduction

The *congested clique* model –a message-passing model of distributed computation where the underlying communication network is the complete graph [20]– is receiving increasingly more attention [4, 6, 7, 8, 11, 12, 13, 15, 18, 22]. There are deep connections between the congested clique model and popular distributed systems such as the k-machine model [17] or MapReduce [14]. Moreover, with the emergence of large-scale networks, this model has started to be used in other areas such as distributed convex learning [1].

The congested clique model is defined as follows. There are n nodes which are given distinct identities (IDs), that we assume for simplicity to be numbers between 1 and n. In this paper we consider the situation where the joint input to the nodes *is a graph* G. More precisely, each node v receives as input an n-bit boolean vector $x_v \in \{0,1\}^n$, which is the indicator function of its neighborhood in G. Note that the input graph G is an arbitrary n-node graph, *a subgraph of the communication network* K_n.

Nodes execute an algorithm, communicating with each other in synchronous rounds and their goal is to compute some function f that depends on G. In every round, each of the n nodes may send up to $n-1$ different b-bit messages through each of its $n-1$ communication links. When an algorithm stops *every node must know* $f(G)$. We call $f(G)$ the *output* of the distributed algorithm. The parameter b is known as the *bandwidth* of the algorithm. We denote by R the *number of rounds*. The product Rb represents the total number of bits received by a node through one link, and we call it the *cost* of the algorithm.

An algorithm may be deterministic or randomized. We distinguish two sub-cases of randomized algorithms: the private-coin setting, where each node flips its own coin; and the public-coin setting, where the coin is shared between all nodes. An ε-error algorithm $\mathcal{A}$ that computes a function f is a randomized algorithm such that, for every input graph G, $\Pr(\mathcal{A} \text{ outputs } f(G)) \geq 1-\varepsilon$. In the case where $\varepsilon \to 0$ as $n \to \infty$, we say that $\mathcal{A}$ computes f with high probability (whp).

Function f defines the problem to be solved. A $0-1$ function corresponds to a decision problem (such as connectivity [13]). For other, more general types of problems, f should be defined, in fact, as a relation. This happens, for instance, when we want to construct a minimum spanning tree [12], a 3-ruling set [15], all-pairs shortest-paths [6], etc.

The most difficult problem we could attempt to solve is the *reconstruction problem*, where nodes are asked to reconstruct the input graph G. In fact, if at the end of the algorithm every node v has full knowledge of G, then it could answer any question concerning G. (This holds because in the congested clique model nodes have unbounded computational power and the only cost is related to communication).

In centralized, classical graph algorithms, a widely used approach to cope with NP-hardness is to restrict the class of graphs where the input G belongs. Consider, for instance, the coloring problem, where the goal is to determine the minimum number of colors that we can assign to the vertices of G such that no two vertices sharing the

same edge have the same color [10]. It is known that, if the input is restricted to the class $\mathcal{G}$ of interval graphs, the coloring problem is polynomial [10]. Nevertheless, if we restrict it to planar graphs, the problem remains NP-complete [10]. We are going to use the same approach here, in the congested clique model. But, as we are going to explain later, surprisingly, the complexity of the reconstruction problem *will only depend on the cardinality* of the subclass of n-node graphs in $\mathcal{G}$.

Formally, for any fixed set of graphs $\mathcal{G}$ we are going to introduce two problems. The first one, the *strong recognition problem* $\mathcal{G}$-Strong-Rec, is the following.

$\mathcal{G}$-Strong-Rec

Input: An arbitrary graph G

Output: $\begin{cases} \text{all the edges of } G & \text{if } G \in \mathcal{G}; \\ \text{reject} & \text{otherwise.} \end{cases}$

Recall that the output is computed by *every node* of the network. In other words, every node of an algorithm that solves $\mathcal{G}$-Strong-Rec must end up knowing whether G belongs to $\mathcal{G}$; and, in the positive cases, every node also finishes knowing all the edges of G. Note that, in principle, $\mathcal{G}$ could be defined as the set of all graphs.

We also define a *weak recognition problem* $\mathcal{G}$-Weak-Rec. This is a promise problem, where the input graph G is promised to belong to $\mathcal{G}$. In other words, for graphs that do not belong to $\mathcal{G}$, the behavior of an algorithm that solves $\mathcal{G}$-Weak-Rec does not matter.

$\mathcal{G}$-Weak-Rec

Input: $G \in \mathcal{G}$

Output: all the edges of G

For any positive integer n we define $\mathcal{G}_n$ as the set of n-node graphs in $\mathcal{G}$. There is an obvious lower bound for Rb, even for the weak reconstruction problem $\mathcal{G}$-Weak-Rec and even in the public-coin setting. In fact, $Rb = \Omega(\log |\mathcal{G}_n|/n)$. This can be easily seen if we note that, in the randomized case, there must be at least one outcome of the coin tosses for which the correct algorithm reconstructs the input graph in at least $(1-\varepsilon)$ of the cases. Therefore, $n + (n-1)Rb = \Omega((1-\varepsilon)\log|\mathcal{G}_n|) = \Omega(\log|\mathcal{G}_n|)$. The value $(n-1)Rb + n$ corresponds to the total number of bits received by any node v of the network: $(n-1)Rb$ bits are received from the other nodes and n bits are known by v at the beginning of the algorithm (this is the indicator function of its neighborhood). This implies that $Rb = \Omega(\log|\mathcal{G}_n|/n)$. In this paper we are going to prove that this bound is tight even with $R = 1$ (if $\mathcal{G}$ is an hereditary class of graphs) and $R = 2$ (in the general case).

We point out that our reconstruction algorithms may be applied not only to G itself but also to some subgraph of G. For instance, consider the situation where we generate

a new graph H by performing (locally) a random sampling on the edges of G. Since H typically belongs to a smaller class of graphs (whp), reconstructing H may result in an efficient strategy to infer some properties of G [26].

1.1 Our Results

We start this paper by studying a very natural family of graph classes known as *hereditary*. A class $\mathcal{G}$ is hereditary if, for every graph $G \in \mathcal{G}$, every induced subgraph of G also belongs to $\mathcal{G}$. Many graph classes are hereditary: forests, planar graphs, bipartite graphs, k-colorable graphs, bounded tree-width graphs, d-degenerate graphs, etc. [5]. Moreover, any intersection class of graphs –such as interval graphs, chordal graphs, unit disc graphs, etc.– is also hereditary [5].

In Section 3 we give, for every hereditary class of graphs $\mathcal{G}$, a one-round private-coin randomized algorithm that solves $\mathcal{G}$-Strong-Rec with bandwidth

$$\mathcal{O}(\max_{k\in[n]} \log |\mathcal{G}_k|/k + \log n).$$

We emphasize that our algorithm runs in one-round, and therefore it runs in the *broadcast congested clique*, a restricted version of the congested clique model where, in every round, the $n-1$ messages sent by a node must be the same. (This equivalence will be explained in Section 2). We also remark that for many hereditary graph classes, including all classes listed above, our algorithm is tight. Moreover, our result implies that $\mathcal{G}$-Strong-Rec can be solved in one-round with bandwidth $\mathcal{O}(\log n)$ when $\mathcal{G}$ is the class of forests, planar graphs, interval graphs, unit-circle graphs, or any other hereditary graph class $\mathcal{G}$ such that $|\mathcal{G}_n| = 2^{\mathcal{O}(n \log n)}$.

In Section 4 we give a very general result, showing that two rounds are sufficient to solve $\mathcal{G}$-Strong-Rec in the congested clique model, for any set of graphs $\mathcal{G}$. More precisely, we provide a two-round deterministic algorithm that solves $\mathcal{G}$-Weak-Rec and a two-round private-coin randomized algorithm that solves $\mathcal{G}$-Strong-Rec whp. We also give a three-round deterministic algorithm solving $\mathcal{G}$-Strong-Rec. All algorithms run using bandwidth $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$, so they are asymptotically optimal when $|\mathcal{G}_n| = 2^{\Omega(n \log n)}$.

Our result implies, in particular, that $\mathcal{G}$-Strong-Rec can be solved in two rounds with bandwidth $\mathcal{O}(\log n)$, when $\mathcal{G}$ is any set of graphs of size $2^{\mathcal{O}(n \log n)}$. The only property of the set of graphs $\mathcal{G}$ used by our algorithm is the cardinality of $\mathcal{G}_n$. Our algorithm does not require $\mathcal{G}$ to be closed under isomorphisms.

In Section 5 we revisit the one-round case. We show that our general algorithm can be adapted to run in one round (i.e., in the broadcast congested clique model) by allowing a larger bandwidth, and then we show that this is tight. More precisely, we show that, for every set of graphs $\mathcal{G}$, there is a one-round deterministic algorithm that solves $\mathcal{G}$-Weak-Rec, and a one-round private-coin algorithm that solves $\mathcal{G}$-Strong-Rec whp, both of them using bandwidth $\mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$.

Then we show that there are classes of graphs $\mathcal{G}$ satisfying that $|\mathcal{G}_n| \leq 2^{\mathcal{O}(n)}$ such that every algorithm (deterministic or randomized) that solves $\mathcal{G}$-Weak-Rec in the

broadcast congested clique model has cost $Rb = \Omega(\sqrt{\log |\mathcal{G}_n|})$. Therefore, with respect to the bandwidth, our general one-round algorithms for solving $\mathcal{G}$-Weak-Rec and $\mathcal{G}$-Strong-Rec are tight (up to a logarithmic factor).

Our one-round algorithm that solves $\mathcal{G}$-Strong-Rec uses private coins. Is it possible to achieve the same deterministically? Our last result gives a negative answer to this question. Consider, for a set of graphs $\mathcal{G}$, the *recognition problem* $\mathcal{G}$-Recognition, which consists in deciding whether the input graph G belongs to $\mathcal{G}$. We show that there exists a set of graphs $\mathcal{S}$, satisfying $|\mathcal{S}_n| \leq 2^n$, such that any one-round deterministic algorithm that solves $\mathcal{S}$-Recognition requires bandwidth $\Omega(n) = \Omega(\log |\mathcal{S}_n|)$. Clearly, the same lower-bound is valid for any deterministic algorithm that solves $\mathcal{S}$-Strong-Rec. This is far from our bandwidth $\mathcal{O}(\sqrt{n \log n}) = \mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$.

1.2 Related Work

All known results concerning the reconstruction of graphs have been obtained in the context of hereditary graph classes. For instance, let $\mathcal{G}$ be the class of *cograph*, that is, the class of graphs that do not contain the 4-node path as an induced subgraph. This class is obviously hereditary. In [16], the authors presented a one-round public-coin algorithm that solves $\mathcal{G}$-Strong-Rec with bandwidth $\mathcal{O}(\log n)$. Note that $|\mathcal{G}_n| = \Theta(2^{n \log n})$. Therefore, the result we get in this paper is stronger, because our one-round algorithm needs the same bandwidth but uses private coins.

In [3, 21] it is shown that, if $\mathcal{G}$ is the class of *d-degenerate* graphs, then there is a one-round deterministic algorithm that solves $\mathcal{G}$-Strong-Rec with bandwidth $\mathcal{O}(d \log n) = \mathcal{O}(\log n)$. A graph G is d-degenerate if one can remove from G a vertex r of degree at most d, and then proceed recursively on the resulting graph $G' = G - r$, until obtaining the empty graph. Note that planar graphs (or more generally, bounded genus graphs), bounded tree-width graphs, graphs without a fixed graph H as a minor, are all d-degenerate, for some constant $d > 0$. Since the class of d-degenerate graphs is hereditary and satisfies $|\mathcal{G}_n| = \Theta(2^{n \log n})$, it follows, from this paper, the existence of a one-round private-coin randomized algorithm that solves $\mathcal{G}$-Strong-Rec with bandwidth $\mathcal{O}(\log n)$. However, the result of [3] for this particular class is stronger, since their algorithm is deterministic.

Another example of reconstruction with one-round algorithms can be found in [8]. There, the authors consider the class of graphs defined by one forbidden subgraph H. They show that such classes can be reconstructed deterministically with cost $Rb = \mathcal{O}((ex(n, H) \log n)/n)$, where $ex(n, H)$ is the *Turán number*, defined as be the maximum number of edges in an n-node graph not containing an isomorphic copy of H as a subgraph. For example, if C_4 is the cycle of length 4, then $ex(n, C_4) = \mathcal{O}(n^{3/2})$. This implies that, if we define $\mathcal{G}$ as the class of graphs not containing C_4 as a subgraph, then there is a one-round deterministic algorithm that solves $\mathcal{G}$-Strong-Rec with bandwidth $\mathcal{O}(\sqrt{n} \log n)$.

2 Preliminaries

2.1 Some Graph Terminology

Two graphs G and H are *isomorphic* if there exists a bijection $\varphi : V(G) \to V(H)$ such that any pair of vertices u, v are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in H. A *class of graphs* $\mathcal{G}$ is a set of graphs which is *closed under isomorphisms*, i.e., if G belongs to $\mathcal{G}$ and H is isomorphic to G, then H also belongs to $\mathcal{G}$. For a class of graphs $\mathcal{G}$ and $n > 0$, we call $\mathcal{G}_n$ the subclass of n-node graphs in $\mathcal{G}$.

For a graph $G = (V, E)$ and $U \subseteq V$ we denote $G[U]$ the subgraph of G induced by U. More precisely, the vertex set of $G[U]$ is U and the edge set consists of all of the edges in E that have both endpoints in U. A class of graphs $\mathcal{G}$ is *hereditary* if it is closed under taking induced subgraphs, i.e., for every $G = (V, E) \in \mathcal{G}$ and every $U \subseteq V$, the induced subgraph $G[U] \in \mathcal{G}$.

For a graph $G = (\{v_1, \ldots, v_n\}, E)$, we call $A(G)$ its *adjacency matrix*, i.e., the 0-1 square matrix of dimension n where $[A(G)]_{ij} = 1$ if and only if v_i is adjacent to v_j. Let M be a square matrix of dimension n, and let $i \in [n] = \{1, \ldots, n\}$. We call M_i the i-th row of M. Let N be another square matrix of dimension n. We denote by $d_r(M, N)$ the *row-distance* between M and N, that is, the number of rows that are different between M and N. In other words, $d_r(M, N) = \{i \in [n] : M_i \neq N_i\}$. For $k > 0$ and $G = (V, E)$, let us call $B(G, k)$ the set of all graphs $H = (V, E')$ such that $d_v(A(G), A(H)) = k$.

2.2 One-Round Algorithms in the Congested Clique

The *broadcast congested clique* is a restricted version of the congested clique model where each node is forced, in each round, to send the same message through its $n - 1$ communication links. But, if we consider one-round algorithms, the two models are the same. In fact, suppose that there is a one-round algorithm $\mathcal{A}$ (deterministic or randomized) in the congested clique with bandwidth b. We can transform it into an algorithm $\mathcal{B}$ in the broadcast version with bandwidth $b + 1$ as follows. We fix a vertex, say the one with ID 1, and every node j broadcasts the message it would send to node 1 on algorithm $\mathcal{A}$, plus one bit indicating whether node j and node 1 are adjacent in G. After this communication round of $\mathcal{B}$, every node knows the messages node 1 would have received after the communication round of algorithm $\mathcal{A}$. Moreover, every node knows the neighborhood of node 1. The result follows from the fact that, with this information, node 1 knows the output. Obviously, as we will see in this paper, when multi-round algorithms are considered, the broadcast congested clique model is much less powerful than the congested clique model.

2.3 Fingerprints

The following technique, that we call *fingerprints*, is based on a result known as the Schwartz Zippel Lemma, used in verification of polynomial identities [25]. Let n be a positive integer and p be a prime number. In the following, we denote by $\mathbb{F}_p$ the finite field of size p (we refer to the book of Lidl and Niederreiter [19] for further details and

definitions involving finite fields). A polynomial $P \in \mathbb{F}_p[X]$ of *degree* d is an expression of the form $P(x) = \sum_{i=0}^{d} a_i x^i$, where $a_i \in \mathbb{F}_p$ and $a_i \neq 0$ for each $0 \leq i \leq d$. We denote by $\mathbb{F}_p[X]$ the polynomial ring on $\mathbb{F}_p$. An element $b \in \mathbb{F}_p$ is called a *root* of a polynomial $P \in \mathbb{F}_p[X]$ if $P(b) = 0$.

Let n be a positive integer, p and q be two prime numbers such that $q < n < p$. For each $a \in \mathbb{F}_q^n$ and $t \in \mathbb{F}_p$, consider the polynomial $FP(a, \cdot) \in \mathbb{F}_p[X]$ defined as

$$FP(a,t) = \sum_{i \in [n]} a_i t^{i-1}.$$

For $t \in \mathbb{F}_p$, we call $FP(a,t)$ the *fingerprint* of a and t. Note in the last expression that the coordinates of a are interpreted as elements of $\mathbb{F}_p$. The following lemma is direct. Since the proof is very short we include it here.

Lemma 1 *[19] Let n be a positive integer, p and q be two prime numbers such that $q < n < p$. Let $a, b \in (\mathbb{F}_q)^n$ such that $a \neq b$. Then, $|\{t \in \mathbb{F}_p : P(a,t) = P(b,t)\}| \leq n$.*

Proof Note that $P(a,t) = P(b,t)$ implies that $P(a-b,t) = P(a,t) - P(b,t) = 0$. Since $P(a-b,t)$ is a polynomial of degree at most n in $\mathbb{F}_p[X]$, it has at most n roots in $\mathbb{F}_p$. Therefore $|\{t \in \mathbb{F}_p : P(a,t) = P(b,t)\}| \leq n$. □

We extend the definition of fingerprints to matrices. Let M be a square matrix of dimension n and coordinates in $\mathbb{F}_q$, and let T be an element of $(\mathbb{F}_q)^n$. We call $FP(M,T) \in (\mathbb{F}_p)^n$ the *fingerprint of M and T*, defined as $FP(M,T) = (FP(M_1,T_1), \ldots, FP(M_n,T_n))$, where M_i is the i-th row of M, for each $i \in [n]$. Moreover, for a graph of size n, and $T \in (\mathbb{F}_p)^n$ we call $FP(G,T)$ the fingerprint of $A(G)$ and T.

3 Reconstructing Hereditary Graph Classes in One Round

In this section we start giving the positive result. Later we explain the consequence of this result on well-known hereditary graph classes.

Theorem 1 *Let $\mathcal{G}$ be an hereditary class of graphs. There exists a one-round private-coin algorithm that solves $\mathcal{G}$-*Strong-Rec *whp and bandwidth $\mathcal{O}(\max_{k \in [n]}(\log(|\mathcal{G}_k|)/k) + \log n)$.*

Proof In the algorithm, nodes use a prime number p, whose value will be chosen later. The algorithm consists in: (1) Each node i picks t_i in $\mathbb{F}_p$ uniformly at random (using private coins), and computes $FP(x_i, t_i)$. (2) Each node communicates t_i and $FP(x_i, t_i)$. (3) Every node constructs $T = (t_1, \ldots t_n)$ and $FP(G,T) = (FP(x_1,t_1), \ldots, F(x_n,t_n))$ from the messages sent in the communication round. Finally: (4) Every node looks in $\mathcal{G}_n$ for a graph H such that $FP(H,T) = FP(G,T)$. If such graph H exists, the algorithm outputs H, otherwise it *rejects*. The description of the algorithm is given in Algorithm 1.

Now we aim to show that, if $H \in \mathcal{G}_n$ satisfies $FP(H,T) = FP(G,T)$, then $G = H$ whp. Let T in $(\mathbb{Z}_p)^n$, picked uniformly at random. Then,

$$Pr(\exists H \in \mathcal{G}_n \text{ s.t. } H \neq G \text{ and } FP(G,T) = FP(H,T)) \\ \leq \sum_{k \in [n]} Pr(\exists H \in \mathcal{G}_n \cap B(G,k) \text{ s.t. } FP(G,T) = FP(H,T)).$$

Suppose that $H \neq G$ and let $k > 0$ such that H belongs to $|B(G,k) \cap \mathcal{G}_n|$. Then, from Lemma 1, we deduce that $Pr(FP(G,T) = FP(H,T)) \leq \left(\frac{n}{p}\right)^k$. It follows that

$$Pr(\exists H \in \mathcal{G}_n \text{ s.t. } H \neq G \text{ and } FP(G,T) = FP(H,T)) \leq \left(\frac{n}{p}\right)^k \cdot |\mathcal{G}_n \cap B(G,k)|.$$

We now claim that $|\mathcal{G}_n \cap B(G,k)| \leq \binom{n}{k} |\mathcal{G}_k|$. Indeed, we can interpret a graph H in $B(G,k)$ as a graph built by picking k vertices $\{v_1, \dots v_k\}$ of $\mathcal{G}$ and then adding or removing edges between those vertices. Since we are looking for graphs in $|\mathcal{G}_n \cap B(G,k)|$, and $\mathcal{G}$ is hereditary, the graph induced by $\{v_1, \dots, v_k\}$ must belong to $\mathcal{G}_k$. Therefore, $|\mathcal{G}_n \cap B(G,k)| \leq \binom{n}{k} |\mathcal{G}_k|$. This claim implies that

$$Pr(\exists H \in \mathcal{G}_n \text{ s.t. } H \neq G \text{ and } FP(G,T) = FP(H,T)) \leq \sum_{k \in [n]} \left(\frac{n^2 \cdot e \cdot (|\mathcal{G}_k|)^{1/k}}{p}\right)^k.$$

Let $f : \mathbb{N} \to \mathbb{R}$ be defined as $f(n) = n \cdot \max_{k \in [n]} \frac{\log |\mathcal{G}_k|}{k}$. Note that this function is increasing, satisfies $f(n)/n \leq f(n+1)/(n+1)$, and $\log |\mathcal{G}_n| \leq f(n)$. Therefore,

$$Pr(\exists H \in \mathcal{G}_n \text{ s.t. } H \neq G \text{ and } FP(G,T) = FP(H,T)) \leq \sum_{k \in [n]} \left(\frac{n^2 \cdot 2^{(f(n)/n)}}{p}\right)^k.$$

We now fix p as the smallest prime number greater than $n^3 \cdot e \cdot 2^{(f(n)/n)}$, and we deduce that

$$Pr(\exists H \in \mathcal{G}_n \text{ s.t. } H \neq G \text{ and } FP(G,T) = FP(H,T)) \leq \frac{1}{n}.$$

Then, with probability at least $1 - 1/n$, either $G = H$ or $F(H,T) \neq F(G,T)$, for every $H \in \mathcal{G}_n$. Hence, the algorithm solves $\mathcal{G}$-Strong-Rec whp.

Note that the bandwidth required by node i in the algorithm equals the number of bits required to represent the pair $(t_i, F(x_i, t_i))$, which are two integers in $[p]$. Therefore, the bandwidth of the algorithm is

$$2\lceil \log p \rceil = \mathcal{O}(f(n)/n + \log n) = \mathcal{O}(\max_{k \in [n]} (\log(|\mathcal{G}_k|)/k) + \log n).$$

□

Algorithm 1: $\mathcal{G}$-WEAK-REC when $\mathcal{G}$ is hereditary. Algorithm executed by node i

Compute p, the smallest prime greater than $n^3 \cdot e \cdot 2^{f(n)/n}$, where $f(n) = n \cdot \max_{k\in[n]} \frac{\log|\mathcal{G}_k|}{k}$;
Pick $T_i \in \mathbb{F}_p$ uniformly at random using private coins ;
Compute $FP(x_i, T_i)$;
Communicate $FP(x_i, T_i)$ and T_i ;
Receive $T = (T_1, \ldots, T_n)$ and $FP(G, T)$;
Look for $H \in \mathcal{G}_n$ such that $FP(H, T) = FP(G, T)$;
If H exists and is unique, output H. Otherwise, reject.

Corollary 1 *Let $\mathcal{G}$ be an hereditary class of graphs, and f be an increasing function such that $|\mathcal{G}_n| = 2^{\theta(nf(n))}$. Then, our private-coin algorithm solves $\mathcal{G}$-STRONG-REC whp, in one-round, with bandwidth $\Theta(\log|\mathcal{G}_n|/n + \log n)$. This matches the lower bound on the cost Rb (which must be satisfied even in the public coin setting).*

Proof We simply note the existence of constants $c_1, c_2 > 0$ such that:

$$\max_{k\in[n]}(\log(|\mathcal{G}_k|)/k) \leq c_2 \cdot \max_{k\in[n]} f(k) \leq c_2 \cdot f(n) \leq (c_2/c_1) \cdot (\log(|\mathcal{G}_n|)/n).$$

Therefore, the algorithm of Theorem 1 uses bandwidth $\mathcal{O}(\log(|\mathcal{G}_n|)/n)$. □

In [24], Scheinerman and Zito showed that hereditary graph classes have a very specific growing rate. They showed ([24], Theorem 1) that, for any hereditary class of graphs $\mathcal{G}$, one of the following behaviors must hold:

- $|\mathcal{G}_n|$ is *constant*; meaning that $|\mathcal{G}_n| \leq 2$ for all n sufficiently large.
- $|\mathcal{G}_n|$ is *polynomial*, meaning that $|\mathcal{G}_n| = n^{\Theta(1)}$.
- $|\mathcal{G}_n|$ is *exponential* , meaning that $|\mathcal{G}_n| = 2^{\Theta(n)}$.
- $|\mathcal{G}_n|$ is *factorial* , meaning that $|\mathcal{G}_n| = 2^{\Theta(n\log n)}$.
- $|\mathcal{G}_n|$ is *super-factorial* , meaning that $|\mathcal{G}_n| = 2^{\omega(n\log n)}$.

Corollary 1 implies that our algorithm is tight for any factorial hereditary class of graphs. For example, the class of forests, planar graphs, interval graphs, unit disc graphs, circle graphs, etc., are factorial. Therefore, the bandwidth required to reconstruct them in one-round is $\Theta(\log n)$. Moreover, constant, polynomial and exponential hereditary classes can be also reconstructed with bandwidth $\mathcal{O}(\log n)$.

Super-factorial hereditary classes of graphs might be more troublesome. Indeed, in [2] it is shown that there exist super-factorial hereditary classes $\mathcal{G}$ such that the succession $\log|\mathcal{G}_n|$ might oscillate, roughly, between $cn\log n$ and $n^{1+c'}$, for two constants $c, c' > 0$. For these classes, the upper bound given by our algorithm does not match the lower

bound $\Omega(\log|\mathcal{G}_n|/n)$. We remark, however, that there are also super-factorial classes of graphs where our algorithm is non-trivial and tight. For example, if $\mathcal{G}$ is the class of chordal-bipartite graphs, we have that $|\mathcal{G}_n| = 2^{\Theta(n\log^2 n)}$. Therefore, they can be reconstructed in one-round with bandwidth $\Theta(\log^2 n)$.

4 Reconstructing Arbitrary Graph Classes in Two Rounds

In this section we show that there exists a two-round private-coin algorithm in the congested clique model that solves $\mathcal{G}$-Strong-Rec whp and bandwidth $\mathcal{O}(\log|\mathcal{G}_n|/n + \log n)$. Our algorithm is based, roughly, on the same ideas used to reconstruct hereditary classes of graphs. But the problem we encounter is the following: while in the case of hereditary classes of graphs, we had for every graph G and $k > 0$, a bound on the number of graphs contained in $B(G,k)\cap\mathcal{G}_n$, this is not the case in an arbitrary family of graphs $\mathcal{G}$. Therefore, fingerprints alone are not able to differentiate graphs. To cope with this obstacle, we use Error Correcting Codes.

4.1 Error Correcting Codes

Consider the following technique, introduced by Reed and Solomon [23], originally used to produce safe communication in a noisy channel. (This technique has also been used in randomized protocols for multiparty communication complexity [9]).

Definition 1 *Let $0 \le k \le n$, and let q be the smallest prime number greater that $n+k$. An* error correcting code with parameters (n,k) *is a mapping $C : \{0,1\}^n \to (\mathbb{F}_q)^{n+k}$, satisfying:*

1) *For every $x \in \{0,1\}^n$ and $i \in [n]$, $C(x)_i = x_i$.*

2) *For each $x, y \in \{0,1\}^n$, $x \neq y$ implies $|\{i \in [n+k] : C(x)_i \neq C(y)_i\}| \ge k$.*

For sake of completeness, we give the construction of an error correcting code with parameters (n,k). For $x \in \{0,1\}^n$, let P_x be the unique polynomial in $\mathbb{F}_q[X]$ satisfying $P_x(i) = x_i$ for each $i \in [n]$. The function C is then defined as $C(x) = (P_x(1), \ldots, P_x(n+k))$. This function satisfies both property (1) from the definition of P_x, and property (2) because two different polynomials of degree n can be equal in at most $n-1$ different values.

We now adapt the definition of error correcting codes to graphs.

Definition 2 *For a graph G, we call $C(G)$ the square matrix of dimension $n+k$ with elements in $\mathbb{F}_q$ defined as follows.*

- *For each $i \in [n]$, the i-th row of $C(G)$ is $C(A(G)_i) \in (\mathbb{F}_q)^{n+k}$ (recall that $A(G)_i$ is the i-th row of the adjacency matrix of G).*

- *For each $i \in [k]$, the $(n+i)$-th row of $C(G)$ is the vector $(C(x_1)_{n+i}, \ldots, C(x_n)_{n+i}, \vec{0}) \in (\mathbb{F}_q)^{n+k}$, where $\vec{0}$ is the zero-vector of $\mathbb{F}_q^d$, and $C(x)_j \in \mathbb{F}_q$ is the j-th element of $C(x)$.*

We can represent $C(x)$ as a pair $(x, \tilde{x})$, where $\tilde{x}$ belongs to $(\mathbb{F}_q)^k$. Similarly, for a graph G, we can represent $C(G)$ as the matrix:

$$C(G) = \begin{bmatrix} A(G) & A\tilde{(G)} \\ A\tilde{(G)}^T & 0 \end{bmatrix}.$$

where $A\tilde{(G)}$ is the matrix with rows $C(A(G)_i)_{n+1}, \ldots, C(A(G)_i)_{n+k}$, $i \in [n]$. Note that $C(G)$ is symmetric.

Remark 1 *Note that $d_r(C(G), C(H)) > k$, for every two different n-node graphs H and G. Indeed, if $G \neq H$, there exists $i \in [n]$ such that $A(G)_i$ is different than $A(H)_i$. Then, by definition of C, $|\{j \in [n+k] : C(A(G))_{i,j} \neq C(A(H))_{i,j}\}| > k$. This means that $d_r(C(G), C(H)) > k$, because $C(G)$ and $C(H)$ are symmetric matrices.*

4.2 Optimal Reconstruction of Arbitrary Graph Classes in Two Rounds

Lemma 2 *Let $\mathcal{G}$ be a set of graphs, C the error correcting code with parameters (n, k), and let p be the smallest prime number greater than $(n+k) \cdot |\mathcal{G}_n|^{2/k}$. Then, there exists $T \in (\mathbb{F}_p)^{n+k}$ depending only on $\mathcal{G}$, satisfying $FP(C(G), T) \neq FP(C(H), T)$ for all different $G, H \in \mathcal{G}_n$.*

Proof From the remark at the end of the last subsection, we know that $d_r(C(G), C(H)) > k$, for every two different n-node graphs H and G. Then, if we pick $T \in (\mathbb{F}_p)^{n+k}$ uniformly at random we have from Lemma 1:

$$Pr(FP(C(G), T) = FP(C(H), T)) < \left(\frac{n+k}{p}\right)^k.$$

Then, by the union bound

$$Pr(\exists G, H \in \mathcal{G}_n \text{ s.t. } G \neq H \text{ and } FP(C(G), T) = FP(C(H), T)) < \left(\frac{n+k}{p}\right)^k \cdot |\mathcal{G}_n|^2 \leq 1.$$

The last inequality follows from the choice of p. Therefore, there must exist a $T \in (\mathbb{F}_p)^{n+k}$ such that $FP(C(G), T) \neq FP(C(H), T)$, for all different $G, H \in \mathcal{G}_n$. □

Theorem 2 *Let $\mathcal{G}$ be a set of graphs. The following holds:*

1) There exists a two-round deterministic algorithm in the congested clique model that solves $\mathcal{G}$-Weak-Rec with bandwidth $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$.

2) There exists a three-round deterministic algorithm in the congested clique model that solves $\mathcal{G}$-STRONG-REC *with bandwidth* $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$.

3) There exists a two-round private-coin algorithm in the congested clique model that solves $\mathcal{G}$-STRONG-REC *with bandwidth* $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$ *whp.*

Proof The first algorithm we are going to explain here, Algorithm 2, is deterministic and solves $\mathcal{G}$-WEAK-REC with bandwidth $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$. The algorithms for (2) and (3) are slight modifications of Algorithm 2 and will also be explained in this proof.

1) Let p be the first prime greater than $2n \cdot |\mathcal{G}_n|^{2/n}$ (then $p \leq 4n \cdot |\mathcal{G}_n|^{2/n}$), and let q be the smallest prime number greater than $2n$. In the algorithm, node i first computes $C(x_i)$, where C is the error correcting code with parameters (n, n). Then, for each $j \in [n]$ node i communicates $C(x_i)_{j+n}$ to node j. This communication round requires bandwidth $\lceil \log q \rceil = \mathcal{O}(\log n)$. After the first communication round, node i knows $C(x_i)$ and $(C(x_1)_{i+n}, \ldots, C(x_n)_{i+n})$, i.e., it knows rows i and $i+n$ of matrix $C(G)$. Each node computes a vector $T \in (\mathbb{F}_p)^{2n}$ such that $FP(C(G), T) \neq FP(C(H), T)$, for all different $G, H \in \mathcal{G}_n$ (each node computes the same T). The existence of T is given by Lemma 2. Then, node i communicates (broadcasts) $P(C(G)_i, T_i)$ and $P(C(G)_{i+n}, T_{i+n})$. This communication round requires bandwidth $2\lceil \log p \rceil = \mathcal{O}((\log |\mathcal{G}_n|)/n + \log n)$. After the second communication round, each node knows $P(C(G), T)$. Then, they locally compute the unique $H \in \mathcal{G}_n$ such that $P(C(H), T) = P(C(G), T)$. Since G belongs to $\mathcal{G}_n$, then necessarily $G = H$.

2) Suppose now that we are solving $\mathcal{G}$-STRONG-REC. In this case G does not necesarily belong to $\mathcal{G}_n$. After receiving the fingerprints of $C(G)$, nodes look for a graph H in $\mathcal{G}_n$ that satisfies $F(C(G), T) = F(C(H), T)$ (line 9 in Algorithm 2). If such a graph exists, we call it a *candidate*. Otherwise, every node decides that G is not in $\mathcal{G}_n$, so they *reject*. Note that, if the candidate exists, then it is unique, since $P(C(H_1), T) \neq P(C(H_2), T)$ for all different H_1, H_2 in $\mathcal{G}_n$. So, if the candidate H exists, each node i checks whether the neighborhood of vertex i on G and H are equal, and announces the answer in the third round (communicating one bit). If every node announces affirmatively, then they output $G = H$. Otherwise, it means that G is not in $\mathcal{G}_n$, so every node *rejects*.

3) We now show that, if we allow the algorithm to be randomized, then we can spare the third round. In fact, nodes only need to run Algorithm 3 after the first round of Algorithm 2. Let us explain this now. Let $p' \in [n^2, 2n^2]$ be a prime number. In the second round, node i picks $S_i \in \mathbb{F}_p$, and it communicates, together with $FP(C(G)_i, T_i)$ and $FP(C(G)_{i+n}, T_{i+n})$, also S_i. After the second round of communication, if a candidate $H \in \mathcal{G}_n$ exists, each node computes $S = (S_1, \ldots, S_n)$, $FP(G, S) = (FP(x_1, S_1), \ldots, F(x_n, S_n)$. If $F(G, S) = F(H, S)$, then nodes deduce that $G = H$. Otherwise, they deduce that $G \notin \mathcal{G}_n$ and *rejects*. Note that if G belongs to $\mathcal{G}_n$, then the algorithm always give the correct answer. Otherwise, it rejects whp. Indeed, if $G \notin \mathcal{G}_n$, then $H \neq G$, and from Lemma 1, $Pr(FP(G, T) = FP(H, T)) \leq 1/n$. □

Note that our private-coin algorithm for $\mathcal{G}$-STRONG-REC has one-sided error. In fact, if the input graph belongs to $\mathcal{G}$, then our algorithm reconstructs it with probability 1.

Algorithm 2: $\mathcal{G}$-Weak-Rec. Algorithm executed by node i

Compute $C(x_i)$, where C is the error-correcting-code with parameters (n, n);
Communicate the element $n + j$ of $C(x_i)$ to player j ;
Receive $C(x_1)_{n+i}, \ldots, C(x_n)_{n+i}$;
Call $C(x_{i+n}) = (C(x_1)_{n+i}, \ldots, C(x_n)_{n+i}, \vec{0})$, where $\vec{0}$ is the zero vector of $(\mathbb{F}_p)^n$;
Compute p as the smallest prime greater than $2n \cdot |\mathcal{G}_n|^{2/n}$;
Compute T, the vector in $\mathbb{F}_p^{2n}$, given by Lemma 2 ;
Compute and communicate (broadcast) $FP(C(x_i), T_i)$ and $FP(C(x_{n+i}), T_{n+i})$;
Receive $FP(C(G), T)$;
Look for $H \in \mathcal{G}_n$ such that $FP(C(H), T) = FP(C(G), T)$;
Output H.

Algorithm 3: Checking a candidate H. Algorithm executed by node i

Compute p', the smallest prime number such that $p' > n^2$;
Pick $T_i \in \mathbb{F}_p$ uniformly at random using private coins ;
Compute $FP(x_i, T_i)$;
Communicate $FP(x_i, T_i)$ and T_i ;
Receive $T = (T_1, \ldots, T_n)$ and $FP(G, T)$;
Output H if $FP(H, T) = FP(G, T)$, otherwise reject.

On the other hand, if G is not contained in $\mathcal{G}$, then our algorithm fails to discard the candidate with probability at most $1/n$.

5 Revisiting the One Round Case

In this section we revisit the one-round case (and therefore the broadcast congested clique model). But instead of studying hereditary graph classes we study arbitrary graph classes, and we show that for this general case we need a larger bandwith. Our results are tight, not only in terms of the bandwidth, but also in the necessity of using randomization.

Theorem 3 *Let $\mathcal{G}$ be a set of graphs. The following holds:*

1) *There exists a one-round deterministic algorithm in the congested clique model that solves $\mathcal{G}$-Weak-Rec with bandwidth $\mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$.*

2) *There exists a two-round deterministic algorithm in the broadcast congested clique model that solves $\mathcal{G}$-Strong-Rec with cost $\mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$.*

3) *There exists a one-round private-coin algorithm in the congested clique model that solves $\mathcal{G}$-Strong-Rec with bandwidth $\mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$ whp.*

Proof The algorithm in this case is very similar to the one we provided in the proof of Theorem 2. Let k be a parameter whose value will be chosen at the end of the proof, and let C be the error-correcting-code with parameters (n, k). Let p be the smallest prime number greater than $2n \cdot |\mathcal{G}|^{2/k}$. Let $T \in (\mathbb{F}_p)^{n+k}$ be the vector given by Lemma 2, corresponding to $\mathcal{G}$.

In the algorithm, every node i computes $C(x_i)$, and communicates $FP(C(x_i), T_i)$ together with $C(x_i)_{n+1}, \ldots, C(x_i)_{n+k} \in (\mathbb{F}_q)^k$, where q is the smallest prime greater than $k + n$. Note that the communication round requires bandwidth

$$\mathcal{O}(\log p + k \cdot \log(n + k)) = \mathcal{O}(\log |\mathcal{G}_n|/k + (k + 1) \cdot \log n).$$

After the communication round, every node knows $FP(C(x_i), T_i)$, for all $i \in [n]$, and also knows the matrix $A(\widetilde{G})$. Therefore, every node can compute $F(C(x_i), T_i)$, for all $i \in \{n + 1, \ldots, n + k\}$, and, moreover, compute $F(C(G), T)$.

From the construction of T, there is at most one graph $H \in \mathcal{G}_n$ such that $F(C(G), T) = F(C(H), T)$. Therefore, if G belongs to $\mathcal{G}$, every node can reconstruct it. On the other hand, if we are solving $\mathcal{G}$-Strong-Rec, then we proceed as in the algorithm of Theorem 2, either testing whether $H = G$ in one more round, or sending a fingerprint of G to check with high probability if a candidate $H \in \mathcal{G}_n$ such that $F(C(G), T) = F(C(H), T)$ is indeed equal to G. This verification requires to send $\mathcal{O}(\log n)$ more bits, which fits in the asymptotic bound of the bandwidth.

The optimal value of k, that is, the one which minimizes the bandwidth, is such that $k = \mathcal{O}\left(\sqrt{\frac{\log |\mathcal{G}_n|}{\log n}}\right)$. Threfore, the bandwidth is $\mathcal{O}(\sqrt{\log |\mathcal{G}_n| \log n} + \log n)$. □

5.1 Tightness of our Algorithms

In this subsection we show that our algorithms for solving $\mathcal{G}$-Weak-Rec and $\mathcal{G}$-Strong-Rec are tight, from two different perspectives. First, from the point of view of the bandwidth, we show that there are classes of graphs $\mathcal{G}$ satisfying $|\mathcal{G}_n| \leq 2^{\mathcal{O}(n)}$ such that every algorithm (deterministic or randomized) solving $\mathcal{G}$-Weak-Rec in the broadcast congested clique model has cost $Rb = \Omega(\sqrt{\log |\mathcal{G}_n|})$. This lower bound matches the upper *one-round* bound given in Theorem 3 (up to logarithmic factors).

Then, we show that, when restricted to one-round algorithms, the use of randomization is necessary in order to have non-trivial general algorithms solving $\mathcal{G}$-Strong-Rec. Indeed, we prove that there exists a set of graphs $\mathcal{G}$ satisfying $|\mathcal{G}_n| \leq 2^n$ such that, every one-round deterministic algorithm that solves $\mathcal{G}$-Strong-Rec, requires bandwidth $\Omega(n)$.

Theorem 4 *There exists a class of graphs $\mathcal{G}$ satisfying $|\mathcal{G}_n| \leq 2^{\mathcal{O}(n)}$ such that, any ϵ-error public-coin algorithm in the broadcast congested clique model that solves $\mathcal{G}$-Weak-Rec, has cost $Rb = \Omega(\sqrt{n}) = \Omega(\sqrt{\log |\mathcal{G}_n|})$.*

Proof Let $\mathcal{G}^+$ be the class of graphs defined as follows: G belongs to $\mathcal{G}_n^+$ if and only if G is the disjoint union of a graph H of $\lceil\sqrt{n}\rceil$ nodes and $n - |H|$ isolated nodes. Note that $|\mathcal{G}_n^+| = \binom{n}{\lceil\sqrt{n}\rceil} \cdot 2^{\binom{\lceil\sqrt{n}\rceil}{2}} \leq 2^{\mathcal{O}(n)}$. Indeed, there are $2^{\binom{\lceil\sqrt{n}\rceil}{2}} = 2^{\mathcal{O}(n)}$ labeled graphs of size $\lceil\sqrt{n}\rceil$, and at most $\binom{n}{\lceil\sqrt{n}\rceil} = 2^{\mathcal{O}(\sqrt{n}\log n)}$ different labelings of a graph of $\sqrt{n}$ nodes using n labels (so $\mathcal{G}^+$ is closed under isomorphisms).

Let $\mathcal{A}$ be an ϵ-error public-coin algorithm solving $\mathcal{G}^+$-Weak-Rec in $R(n)$ rounds and bandwidth $b(n)$, on input graphs of size n.

Consider now the following algorithm $\mathcal{B}$ that solves $\mathcal{U}$-Weak-Rec, where $\mathcal{U}$ is the set of all graphs: on input graph G of size n, each node $i \in [n]$ supposes that it is contained in a graph G^+ formed by G plus $n^2 - n$ isolated vertices with identifiers $(n+1), \ldots, n^2$. Note that G^+ belongs to $\mathcal{G}^+$. Then, node i simulates $\mathcal{A}$ as follows: at each round, node $i \in [n]$ produces the message of node i in G^+ according to $\mathcal{A}$. Note that the messages produced by nodes labeled $(n+1), \ldots, n^2$ do not depend on G, so they can be produced by any node of G. Since $\mathcal{A}$ solves $\mathcal{G}^+$-Weak-Rec, at the end of the algorithm every node knows all the edges of G^+, so they reconstruct G ignoring vertices labeled $(n+1), \ldots, n^2$.

We deduce that algorithm $\mathcal{B}$ solves $\mathcal{U}$-Weak-Rec. Note that the cost of $\mathcal{B}$ is $n^2R(n)b(n)$ on input graphs of size n. We deduce that $n^2R(n)b(n) = \Omega(n)$, i.e., the cost of $\mathcal{A}$ is $\Omega(\sqrt{n})$. □

We say that an algorithm *recognizes* $\mathcal{G}$ if the algorithm decides whether an input graph G belongs to $\mathcal{G}$. We call $\mathcal{G}$-Recognition the problem of recognizing $\mathcal{G}$.

Theorem 5 *There exists a set of graphs $\mathcal{G}$ satisfying $|\mathcal{G}_n| \leq 2^n$ such that, and any one-round deterministic algorithm in the congested clique model that solves $\mathcal{G}$-Recognition, requires bandwidth $\Omega(n)$.*

Proof We prove this theorem by a counting argument. Our goal is to show that there are more *small* sets of graphs than one-round deterministic algorithms capable to recognize them.

We first count the number of sets of graphs (not necessarily closed under taking isomorphism) containing 2^n different graphs of size n. We call the family of these sets $\mathcal{C}$. There are $2^{\binom{n}{2}}$ possible graphs of size n, so $\binom{2^{\binom{n}{2}}}{2^n}$ possible choices for graphs in $\mathcal{C}$. We deduce that there exists $c_1 > 0$ such that $|\mathcal{C}| \geq 2^{c_1 \cdot n^2 \cdot 2^n}$.

On the other hand, we count the number of one-round deterministic algorithms that recognize a set of graphs in $\mathcal{C}$ with bandwidth at most β. A one-round deterministic algorithm is composed of two parts: the algorithm before the communication round, and the algorithm after the communication. The first part of an algorithm is defined by the messages that a node sends on each input. The input of a node is its neighborhood represented by a Boolean vector of size n, and an integer representing its label. Therefore, the first part of an algorithm is defined by the messages corresponding to all the $n2^n$ possible inputs. Since the bandwidth is β, we obtain that there are $2^{n\beta 2^n}$ possible choices for the first part of an algorithm.

The second part of an algorithm is defined by a function $f_{\mathcal{G}} : (\{0,1\}^b)^n \to \{0,1\}$, such that if $m = (m_1, \ldots, m_n)$ are the messages sent by the nodes in the communication round, then $f(m) = 1$ if and only if m was produced from an input graph belonging to $\mathcal{G}$. The crucial observation is that this implies that f can output 1 in at most 2^n inputs. Therefore, the number of possible second parts of an algorithm is $\sum_{i \in [2^n]} \binom{2^{n\beta}}{i} \leq (1 + 2^{n\beta})^{2^n} \leq 2^{c_2 \cdot n\beta 2^n}$, where $c_2 > 0$ is a constant.

We deduce that the number of one-round deterministic algorithms with bandwidth β that are capable to recognize a set of graphs in $\mathcal{C}$ is at most $2^{c_3 n\beta 2^n}$, with $c_3 > 0$. Since we are considering only deterministic algorithms, two different sets must be recognized by two different algorithms. This implies that $2^{c_3 n\beta 2^n}$ must be greater than $2^{c_1 n^2 2^n}$, so $\beta = \Omega(n)$.

Finally, we construct $\mathcal{G}$ by picking, for each n, one set of graphs contained in $\mathcal{C}$ that can not be recognized by any algorithm of bandwidth $o(n)$. □

Remark 2 *Note that for any set of graphs $\mathcal{G}$, problem $\mathcal{G}$-Strong-Rec is at least as hard as $\mathcal{G}$-Recognition. We conclude that there exists a set of graphs $\mathcal{G}$ satisfying $|\mathcal{G}_n| \leq 2^n$ such that, any one-round deterministic algorithm that solves $\mathcal{G}$-Strong-Rec, requires bandwidth $\Omega(n)$. Note that, since in this case $|\mathcal{G}_n| \leq 2^n$, from Theorem 3 we know that $\mathcal{G}$-Strong-Rec can be solved using a one-round private-coin algorithm with bandwidth $\mathcal{O}(\sqrt{n \log n})$ whp.*

6 Discussion

In this paper we showed that all graph classes can be optimally reconstructed in two rounds in the congested clique model. But our algorithm is randomized, it uses private-coins. A natural question is the following: is it possible to achieve the same deterministically? In other words, given an arbitrary graph class $\mathcal{G}$, is it always possible to solve $\mathcal{G}$-Strong-Rec with a two-round deterministic algorithm with bandwidth $\mathcal{O}(\log |\mathcal{G}_n|/n + \log n)$? (Note that this is true for the weak version of the reconstruction problem $\mathcal{G}$-Weak-Rec).

We also restricted the reconstruction problem to one-round algorithms. We showed that, if $\mathcal{G}$ is an hereditary graph class such as forests, planar graphs, interval graphs, unit disc graphs, chordal bipartite graphs, bounded tree-widh graphs, d-degenerate graphs, etc., then $\mathcal{G}$-Strong-Rec can be solved, whp, with a one-round private-coin algorithm that uses bandwidth $\mathcal{O}(\log |\mathcal{G}_n|/n)$. Can we extend this result to every hereditary class of graphs?

A related problem is the recognition problem, where we simply want to decide whether the input graph belongs to the class $\mathcal{G}$. It seems that sometimes we can not solve the recognition problem without solving the reconstruction problem. This seems to be true in the case of trees and, more generally, in the case of d-degenerate graphs. But this is not always the case. Sometimes, solving the recognition problem requires a much smaller bandwidth. For example, consider the class of split graphs. A split

graph is a graph where the vertices can be partitioned into a clique and an independent set (these two sets are connected arbitrarily). The class of split graphs contains $2^{\Omega(n^2)}$ graphs of size n, so it cannot be reconstructed with cost $o(n)$. However, split graphs can be characterized solely by their degree sequences (see [5]), so they can be recognized by a one-round deterministic algorithm, where each node sends its degree ($\mathcal{O}(\log n)$ bits). It is an interesting challlenge to understand the cases where we can solve the recognition problem *without* solving the reconstruction problem.

References

[1] Yossi Arjevani and Ohad Shamir. Communication complexity of distributed convex learning and optimization. In *Adv. Neural Inf. Process. Syst.*, pages 1756–1764, 2015.

[2] József Balogh, Béla Bollobás, and David Weinreich. The penultimate rate of growth for graph properties. *European Journal of Combinatorics*, 22(3):277 – 289, 2001.

[3] Florent Becker, Martín Matamala, Nicolas Nisse, Ivan Rapaport, Karol Suchan, and Ioan Todinca. Adding a referee to an interconnection network: What can(not) be computed in one round. In *25th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 508–514, 2011.

[4] Andrew Berns, James Hegeman, and Sriram Pemmaraju. Super-fast distributed algorithms for metric facility location. *Automata, Languages, and Programming*, pages 428–439, 2012.

[5] Andreas Brandstädt, Van Bang Le, and Jeremy P Spinrad. *Graph classes: a survey*. SIAM, 1999.

[6] Keren Censor-Hillel, Petteri Kaski, Janne H. Korhonen, Christoph Lenzen, Ami Paz, and Jukka Suomela. Algebraic methods in the congested clique. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 143–152, 2015.

[7] Danny Dolev, Christoph Lenzen, and Shir Peled. "Tri, tri again": Finding triangles and small subgraphs in a distributed setting - (ext. abstract). In *26th International Symposium on Distributed Computing (DISC)*, pages 195–209, 2012.

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

[9] Orr Fischer, Rotem Oshman, and Uri Zwick. Public vs. private randomness in simultaneous multi-party communication complexity. In *Int. Colloquium on Structural Information and Communication Complexity (SIROCCO)*, pages 60–74. Springer, 2016.

[10] Michael R Garey and David S Johnson. *Computers and intractability*, volume 29. wh freeman New York, 2002.

[11] Mohsen Ghaffari. An improved distributed algorithm for maximal independent set. In *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 270–277. Society for Industrial and Applied Mathematics, 2016.

[12] Mohsen Ghaffari and Merav Parter. Mst in log-star rounds of congested clique. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 19–28. ACM, 2016.

[13] James W Hegeman, Gopal Pandurangan, Sriram V Pemmaraju, Vivek B Sardeshmukh, and Michele Scquizzato. Toward optimal bounds in the congested clique: Graph connectivity and mst. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 91–100. ACM, 2015.

[14] James W. Hegeman and Siriam V. Pemmaraju. Lessons from the congested clique applied to MapReduce. In *Int. Colloquium on Structural Information and Communication Complexity (SIROCCO)*, volume 8576 of *LNCS*, pages 149–164, 2014.

[15] James W. Hegeman, Sriram V. Pemmaraju, and Vivek Sardeshmukh. Near-constant-time distributed algorithms on a congested clique. In *28th International Symposium on Distributed Computing (DISC)*, pages 514–530, 2014.

[16] Jarkko Kari, Martin Matamala, Ivan Rapaport, and Ville Salo. Solving the induced subgraph problem in the randomized multiparty simultaneous messages model. In *21st Int. Colloquium, on Structural Information and Communication Complexity (SIROCCO)*, pages 370–384, 2015.

[17] Hartmut Klauck, Danupon Nanongkai, Gopal Pandurangan, and Peter Robinson. Distributed computation of large-scale graph problems. In *26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 391–410. SIAM, 2015.

[18] Christoph Lenzen. Optimal deterministic routing and sorting on the congested clique. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 42–50, 2013.

[19] Rudolf Lidl and Harald Niederreiter. *Introduction to finite fields and their applications*. Cambridge university press, 1994.

[20] Z. Lotker, B. Patt-Shamir, E. Pavlov, and D. Peleg. Minimum-weight spanning tree construction in O (log log n) communication rounds. *SIAM J. Comput.*, 35(1):120–131, 2005.

[21] Pedro Montealegre and Ioan Todinca. Brief announcement: Deterministic graph connectivity in the broadcast congested clique. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 245–247. ACM, 2016.

[22] Boaz Patt-Shamir and Marat Teplitsky. The round complexity of distributed sorting. In *ACM Symposium on Principles of Distributed Computing (PODC)*, pages 249–256, 2011.

[23] Irving S Reed and Gustave Solomon. Polynomial codes over certain finite fields. *J. Soc. Ind. Appl. Math.*, 8(2):300–304, 1960.

[24] Edward R Scheinerman and Jennifer Zito. On the size of hereditary classes of graphs. *Journal of Combinatorial Theory, Series B*, 61(1):16–39, 1994.

[25] Jacob T Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. of the ACM*, 27(4):701–717, 1980.

[26] Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. on Computing*, 40(4):981–1025, 2011.