Solving the INDUCED SUBGRAPH problem in the randomized multiparty simultaneous messages model [*]

J. Kari[1], M. Matamala[2], I. Rapaport[2], and V. Salo[2]

[1] Department of Mathematics and Statistics, University of Turku, Finland
[2] DIM-CMM (UMI 2807 CNRS), Universidad de Chile

Abstract. We study the message size complexity of recognizing, under the broadcast congested clique model, whether a fixed graph H appears in a given graph G as a minor, as a subgraph or as an induced subgraph. The n nodes of the input graph G are the players, and each player only knows the identities of its immediate neighbors. We are mostly interested in the one-round, simultaneous setup where each player sends a message of size $\mathcal{O}(\log n)$ to a referee that should be able then to determine whether H appears in G. We consider randomized protocols where the players have access to a common random sequence. We completely characterize which graphs H admit such a protocol. For the particular case where H is the path of 4 nodes, we present a new notion called twin ordering, which may be of independent interest.

1 Introduction

Yao, in his seminal paper of 1979 [27], not only introduced the two-party communication model but also the much more restricted two-party *simultaneous messages* communication model (SM). The SM model is defined as follows. Alice and Bob wish to evaluate together a function $f : X \times Y \to \{0, 1\}$. Alice receives her input x, Bob receives his input y. Both Alice and Bob send simultaneously a message to a referee, who sees none of the input. The referee then announces the function value $f(x, y)$. Of course, the goal of the game is to minimize the size of the messages. Many results have been obtained in this model and, in particular, clear separations have been proved between the deterministic and the randomized settings [5, 9, 19].

The extension of the SM model to many players is direct and it is defined as follows. There are n players. These n players wish to evaluate together a function $f : X_1 \times \ldots \times X_n \to \{0, 1\}$. Each player receives an input $x_i \in X_i$. The n players send simultaneously a message to the referee who uses these messages in order to compute the boolean function $f(x_1, \ldots, x_n)$. We call this model the *multiparty simultaneous messages* communication model (MSM).

The already defined *number-in-hand* multiparty communication model is more general than the MSM model because, in the number-in-hand model, many rounds are allowed and different communication modes can be considered [12, 14, 17, 21, 26]. In fact, the MSM model corresponds to the one-round, synchronous, shared-whiteboard number-in-hand model.

The *broadcast congested clique model* is exactly the number-in-hand model but where the joint input, instead of being $(x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$, *is a graph* [13, 16]. This input graph is distributed among the nodes, which are the parties of the communication game. More precisely, in the broadcast congested clique model, the joint input to the n nodes is an undirected n-node graph G, with node v receiving the list of its neighbors in G. Each node broadcasts, in each round, a b-bit message (written on a whiteboard, which is visible to every node).

In this paper we are interested in the simultaneous messages (one-round) broadcast congested clique model SM-BCAST. We assume that the ID of each node is a unique number between 1 and n and that the only information each node has, besides n and its own ID, is the list of IDs of its neighbors in G. These nodes need to send, simultaneously, a b-bit message to the referee allowing him to answer, typically, questions of the form "Does the input graph G belong to the graph class $\mathcal{C}$?".

If there is no restriction on the message size then there is a trivial simultaneous protocol that allows the referee to reconstruct any graph: given an input graph G (with an arbitrary assignment of IDs to each of the n nodes), every node sends the binary vector $x \in \{0,1\}^n$ corresponding to the indicator function of its neighborhood. Clearly, this information determines G completely.

If we restrict the message size then reconstructing G becomes much more difficult. Despite this, in [6] it was proved that if (an upper bound on) the *degeneracy* of G is known in advance, then it is possible to reconstruct G with a one-round protocol of $\mathcal{O}(\log n)$ message size. More precisely,

Proposition 1 (Lemma 2 of [6]). *Let m be a positive number. Then, it is possible to decide deterministically, in the* SM-BCAST *model, the class of m-degenerate graphs using messages of size $\mathcal{O}(m^2 \log n)$. Moreover, if the degeneracy of G is (upper bounded by) m then G can be completely reconstructed by the referee.*

The degeneracy m of a graph is defined as follows: G is m-*degenerate* if one can remove from G a node r of degree at most m, and then proceed recursively on the resulting graph $G' = G - r$, until obtaining an empty graph; the *degeneracy* of G is the smallest m such that G is m-degenerate. For instance, the degeneracy of trees is 1, and the degeneracy of planar graphs is at most 5. Many other important graph classes have bounded degeneracy and this is the reason why previous result is surprising.

In [1, 2, 15] the authors introduced a beautiful and powerful technique for *graph sketching*. This technique works both for streaming models and for the SM-BCAST model. It allows the referee to decide whether the input graph is

connected when each node sends one message of size $\mathcal{O}(\log^3 n)$. The protocol for generating the messages is randomized.

Some negative results for the SM-BCAST model have also been obtained. In [7] the authors prove that it is impossible to decide whether the input graph G has diameter at most 3 or whether G has a triangle unless the messages sent by the nodes are all of size $\Omega(n)$, even if randomness is allowed. Deciding whether the input graph G contains a cycle requires at least one node to write a message of length at least $\lceil \log d \rceil - 1$, where d is the maximum degree of G [4].

It should be pointed out that negative results in the general broadcast congested clique model yield negative results in the SM-BCAST model. In fact, if one can prove that any solution for some problem in the broadcast congested clique model allowing messages of size at most b needs at least r rounds, then one can conclude that any solution of the same problem in the SM-BCAST model needs messages of size at least $\Omega(rb)$.

Let H be a fixed graph. The question we address in this paper is the following: "Does H appear in the input graph G?" In graph theory, the word "appear" has at least three interpretations: H may appear as a *minor* of G, as a *subgraph* of G or as an *induced subgraph* of G.

1.1 Minors

An interesting application of Proposition 1 is related to the problem of detecting the presence of particular minors in the input graph G. The study of graph classes defined by graph minors is one of the most important branches of graph theory, culminating in the Robertson–Seymour theorem [22], also known as the Graph Minor Theorem, which states that every minor-closed family of graphs is defined by a *finite* set of forbidden minors. Many classes of graphs are minor-closed, and have known characterizations in terms of minors. For example, the famous theorem of Kuratowski states that planar graphs are exactly those not containing K_5 or $K_{3,3}$ as minors.

Let H be a fixed graph. We say that H is a minor of G if H can be extracted from G by deleting edges, deleting nodes and contracting edges. We say that G is H-minor free if G does not have H as a minor. H-minor free graphs have bounded degeneracy [18, 24, 25]. This fact, together with Proposition 1, allows us to conclude that, in the SM-BCAST model, it is possible to decide deterministically whether H is a minor of G using messages of size $\mathcal{O}(\log n)$. Moreover, if G is H-minor free then G can be completely reconstructed by the referee. This implies that for *every* minor-closed class $\mathcal{C}$, there must be an $\mathcal{O}(\log n)$ message size deterministic protocol that decides class $\mathcal{C}$ (and that even reconstructs the input graphs belonging to the class). Unfortunately, for many minor-closed classes we have not discovered the corresponding finite set of forbidden minors yet, and therefore, we can only conclude the *existence* of such protocol (as it occurs with the existence of polynomial time algorithms for recognizing minor-closed classes in the sequential, classical setting).

1.2 Subgraphs

We say that G *contains* H if H is a (not necessarily induced) subgraph of G. The problem H-SUBGRAPH consists in deciding whether H is a subgraph of G. Proposition 1 can also be used for tackling H-SUBGRAPH. In fact, Drucker, Kuhn and Oshman [13] made the following remark: the degeneracy of graphs which do not contain H as a subgraph (H-subgraph free graphs) can be upper bounded in terms of the Turán number $ex(n, H)$, the maximal number of edges of an n-node graph which does not contain a subgraph isomorphic to H. More precisely, they showed that the degeneracy of H-subgraph free graphs with n nodes is at most $4ex(n, H)/n$. This gives the following result.

Proposition 2 ([13]). *Let H be a fixed graph. Then, the problem H-SUBGRAPH can be solved in the* SM-BCAST *model with a $\mathcal{O}(ex(n, H)^2 \log n/n^2)$ message size deterministic protocol.* ⋆

The previous proposition gives some interesting upper bounds. For instance, if H is a tree or a forest then $ex(n, H) = \Theta(n)$ [3]. Therefore, in this case, H-SUBGRAPH can be solved with messages of size $\mathcal{O}(\log n)$. It is also known that $ex(n, C_{2\ell}) = \Theta(n^{1+1/\ell})$, where $C_{2\ell}$ is the even length cycle of length 2ℓ [8]. In other words, $C_{2\ell}$-SUBGRAPH can be solved with messages of size $\mathcal{O}(n^{2/\ell} \log n)$.

The authors in [13] obtained interesting lower bounds which also depend on the Turán number. For instance, consider the ℓ-node cycle C_ℓ. They show that if $\ell \geq 4$, then any protocol that solves C_ℓ-SUBGRAPH needs at least $\Omega(ex(n, C_\ell)/(nb))$ rounds, where b is the message size each node can broadcast in each round. This yields a lower bound of $\Omega(ex(n, C_\ell)/n)$ message size for the SM-BCAST model. Considering that $ex(n, C_\ell) = \Theta(n^2)$ if $\ell > 3$ is odd and that $ex(n, C_\ell) = \Theta(n^{1+2/\ell})$ if ℓ is even we conclude the following: any randomized protocol which solves C_ℓ-SUBGRAPH in the SM-BCAST model uses messages of size at least $\Omega(n)$ if ℓ is odd and $\Omega(n^{2/\ell})$ if ℓ is even.

In the case of triangles C_3 they obtain $\Omega(n/(e^{\mathcal{O}(\sqrt{\log n})}b))$ rounds as a lower bound for the deterministic case. This yields a lower bound of $\Omega(n/e^{\mathcal{O}(\sqrt{\log n})})$ for any deterministic protocol that solves C_3-SUBGRAPH in the SM-BCAST model. In Corollary 5, we state a similar result more generally: if H contains a cycle, then messages of polynomial size are needed in the problem H-SUBGRAPH.

1.3 Induced subgraphs

An *induced subgraph* of a graph $G = (V, E)$ is a graph $G' = (V', E')$ with $V' \subseteq V$ and such that $vw \in E'$ if and only if $vw \in E$. In other words, the edges of the induced graph G' are all those whose endpoints are both in V'. A class of graphs

---

⋆ In [13] the authors say that the message size is $\mathcal{O}(ex(n, H) \log n/n)$. But this bound is an optimistic interpretation of the upper bound of Proposition 1, because instead of considering m^2 they consider m. The conclusions they obtain do not depend on this issue.

$\mathcal{G}$ is said to be *hereditary* if every induced subgraph of every member of $\mathcal{G}$ is also in $\mathcal{G}$.

A graph G is H-free if H is not an induced subgraph of G. It is easy to show that a graph class $\mathcal{G}$ is hereditary if and only if $\mathcal{G}$ is defined by a (finite or infinite) set $\mathcal{H}$ of forbidden graphs. More precisely, $\mathcal{G}$ is hereditary if and only if for some $\mathcal{H}$, $\mathcal{G} = \{G \mid G \text{ is } H\text{-free, for every } H \in \mathcal{H}\}$.

There is no analog of the Graph Minor Theorem for induced subgraphs, and many classes of graphs have an infinite minimal set of forbidden induced subgraphs. For instance, the class of bipartite graphs is hereditary and the (minimal) set of forbidden induced subgraphs is the set of odd cycles. There are, however, many interesting classes of graphs defined by finite families of forbidden induced subgraphs. For example, the class P_4-free, the class of graphs without an induced copy of the 4-node path P_4, is the class of *cographs*. It arises often in algorithmic graph theory, and also plays a major role in this article.

Problem H-INDUCED SUBGRAPH consists in deciding whether H is an induced subgraph of G. This problem has not yet been addressed in the congested clique model (with the exception of H being a clique, because K_k is an induced subgraph of G if and only if it is a subgraph of G). This work intends to initiate this research line.

1.4 Notation

In this work a "graph" is always a "simple undirected labeled graph". In particular, the nodes of the n-node input graph $G = (V, E)$ are labeled by their IDs. The open neighborhood of a node $v \in V$ is denoted by $N_G(v)$ and corresponds to the set of nodes which are adjacent to v. The closed neigborhood is $N_G[v] = N_G(v) \cup \{v\}$.

Let H be a graph. The number of nodes of H is denoted by $|H|$. Its complement $\overline{H}$ is the graph with the same set of nodes $V(H)$ but such that $e \in E(\overline{H}) \iff e \notin E(H)$. We write $H_1 \cong H_2$ when the two graphs are isomorphic. Let v be a node of H. We denote by $H - v$ the graph with $|H| - 1$ nodes where, besides removing v, we also remove all the edges incident to v. Similarly, we denote by $H - e$ the graph obtained by removing the edge e from H.

The path of k nodes is denoted by P_k, the cycle of k nodes is denoted by C_k, the clique of k nodes is denoted by K_k. The disjoint union of H_1 and H_2 is denoted by $H_1 + H_2$. The disjoint union of t isomorphic graphs is denoted by tH (where each of the t graphs is isomorphic to H).

A *deterministic protocol* $\mathcal{P}$ in the SM-BCAST model describes the mechanisms of the nodes (for generating the messages) and the mechanism of the referee (for retrieving the final result) that correctly computes the output on all inputs. An ϵ-*error randomized protocol* $\mathcal{P}$ for some problem is a protocol in which every node and the referee are allowed to use a *public* sequence of random bits, and for every input the referee outputs the correct answer with probability at least $1 - \epsilon$. The *cost* of a protocol $\mathcal{P}$, denoted $C(\mathcal{P})$, is the length of the longest message sent to the referee. The *deterministic message size complexity*, denoted $C(f)$, is the minimum cost of any deterministic protocol computing f.

Analogously, we denote as $C_\epsilon(f)$ the message size complexity for ϵ-error (public) randomized protocols.

1.5 Our results

We study the message size complexity of the problem of determining whether a fixed graph H "appears" in a given graph G, mostly under the one-round SM-BCAST model. In particular, we are interested in finding out which graphs H admit (deterministic or randomized) solutions with message size that is logarithmic in n, the number of nodes of the input graph G. Note that a $\log(n)$-size message allows one to identify a node in G, so each node can broadcast the identities of a bounded number of nodes.

As already discussed in Section 1.1, for any graph H, a logarithmic message size is enough to determine – even deterministically – if H is a minor of an arbitrary input graph G.

By Section 1.2, the same is true for the problem of determining whether H is a subgraph of a given G when H is a forest. In other words, if H is a forest, then H-SUBGRAPH can be decided by a deterministic protocol with simultaneous messages of logarithmic size. On the other hand, in Section 2 we prove that if H is not a forest then any protocol (even randomized) requires polynomial size messages. These results are summarized in Corollary 5.

Our results of Section 3 concern the appearance of H as an induced subgraph in G (with $|V(H)| \geq 3$, because otherwise the problem is trivial). Corollary 6 (together with Comment 1) states that polynomial message size is required to solve H-INDUCED SUBGRAPH – even with a randomized protocol – for all H except for $H \in \{P_1 + P_2, P_3, P_4\}$. These are exactly the graphs of order at least three that both themselves and their complements are without cycles. We then provide a randomized protocol with logarithmic message size for the case $H = P_3$ (equivalently $H = P_1 + P_2$) in Proposition 7. Note that P_3-INDUCED SUBGRAPH is equivalent to asking if a graph G is a disjoint union of cliques.

Our most involved result is the one of Section 4, where we provide a randomized protocol with logarithmic message size for problem P_4-INDUCED SUBGRAPH (Proposition 10). For doing this we give a characterization of P_4-free graphs (or cographs) based on the notion of twin ordering. This characterization of cographs is, to the best of our knowledge, a new one.

We are not aware of deterministic one-round solutions for P_3- and P_4-INDUCED SUBGRAPH problems, so these remain open. However, the problems can be solved with logarithmic message size in two rounds (Proposition 8) and in $2(h-1)$ rounds (Proposition 11), respectively, where h bounds the cograph level to be checked.

Every connected cograph has diameter 2. Proposition 10 tells us that cographs can be recognized, in the SM-BCAST model, with a randomized $\mathcal{O}(\log n)$ message size protocol with $1/n^c$ error. It is interesting to point out that, from the paper of Holzer and Pinsker [16], one can conclude that for deciding whether a graph has diameter 2, the size of the messages must be $\Omega(n)$, even if randomness is allowed.

2 Lower bounds for detecting subgraphs and induced subgraphs

As mentioned in Section 1.2, it follows from [13] that any randomized ϵ-error protocol that solves the problem C_ℓ-SUBGRAPH in the SM-BCAST model uses messages of size $\Omega(n)$ for $\ell > 3$ odd and $\Omega(n^{2/\ell})$ for ℓ even

The following two propositions generalize these results from cycles C_ℓ to arbitrary graphs H that contain a cycle. Our proofs work also in the case $H = C_3$ of a triangle, and the same proofs provide the lower bounds also for the H-INDUCED SUBGRAPH problem.

The proofs are reductions from the INDEX problem. Consider the INDEX function in the two players SM model: the first player, say Alice, has as input an N-bit boolean vector x and the second player, Bob, has an integer $q \in [1, N]$. Then INDEX$(x, q) = x_q$, the q'th coordinate of Alice's vector. We will use the fact that for any $\epsilon < 1/2$, any public coin randomized protocol for INDEX requires $\Omega(N)$ bits (see, e.g., [19] for a proof).

Let H and G be two disjoint graphs, let ab be an edge of H and let r, t be two nodes of G. We denote by $G_{rt} \oplus H_{ab}$ the graph obtained from G and $H - ab$ by identifying nodes r and a, and nodes t and b. Then, the set of nodes of $G_{rt} \oplus H_{ab}$ is $V(G) \cup V(H) \setminus \{a, b\}$, where we still denote by r (resp. t), the new node obtained under the identification of r and a (resp. t and b). We call $\tilde{G}$ the subgraph of $G_{rt} \oplus H_{ab}$ induced by the set of nodes $V(G)$; we have $\tilde{G} \cong G$. We call $\tilde{H}$ the subgraph of $G_{rt} \oplus H_{ab}$ induced by the set $U := (V(H) \cup \{r, t\}) \setminus \{a, b\}$. We notice that $\tilde{H} \cong H$ if and only if rt is an edge of G.

A cycle in $G_{rt} \oplus H_{ab}$ is called a *crossing* cycle if it contains nodes from $V(G) \setminus \{r, t\}$ and from $V(H) \setminus \{a, b\}$. Then, the length of a crossing cycle is at least the distance in $H - ab$ between a and b, which we denote by k_H, plus the distance in $G - rt$ between r and t, which we denote by k_G.

Let $\mathcal{P}$ be a protocol for a graph problem. We denote by $\mathcal{P}(|G|, v, N_G(v))$ the message generated in the protocol by node v having neighborhood $N_G(v)$ in a graph with $|G|$ nodes.

We first consider the case that H contains a cycle of odd length.

Proposition 3. *Let H be a non-bipartite graph. Any randomized ϵ-error protocol that solves H-INDUCED SUBGRAPH or H–SUBGRAPH uses messages of size $\Omega(n)$.*

Proof. Let $N \in \mathbb{N}$ be even. Consider the following instance of the INDEX problem. Alice receives the indicator vector of a set $X \subseteq V_L \times V_R$ where V_L and V_R are two disjoint sets of cardinality $|V_L| = |V_R| = \frac{N}{2}$. Bob receives a couple (p, q). The question the referee needs to answer is whether $(p, q) \in X$. We already know that any ϵ-error randomized protocol that solves this problem needs Alice to send $\Omega(N^2)$ bits.

Suppose that there exists a randomized ϵ-error protocol $\mathcal{P}$ that solves H-INDUCED SUBGRAPH or H-SUBGRAPH using messages of size $c(n)$. We are going to use $\mathcal{P}$ to solve INDEX.

Consider the N-node graph $G = (V_L \cup V_R, E)$ with $E = X$. Let a, b be two nodes of H such that the edge ab lies in a shortest odd cycle (it must exist, H is non-bipartite), and let k be the length of this cycle. Let $i \in V_L$ and $j \in V_R$. Then, any odd-length crossing cycle of $G_{ij} \oplus H_{ab}$ has length at least $k + 2$: paths in G between i and j have odd length, and the shortest even-length path in $H - ab$ between a and b has length $k - 1$. Hence, any cycle of length k in $G_{ij} \oplus H_{ab}$ is either included in G or in $\tilde{H}$. But as k is odd and G is bipartite, any cycle of length k belongs to $\tilde{H}$. In conclusion, H is an (induced) subgraph of $G_{ij} \oplus H_{ab}$ if and only if ij is an edge of G: if H is a subgraph of $G_{ij} \oplus H_{ab}$, then ij is an edge of G as otherwise $\tilde{H}$ has fewer cycles of length k than H and, conversely, if ij is an edge of G then $\tilde{H} \cong H$ is an induced subgraph of $G_{ij} \oplus H_{ab}$.

Alice can take advantage of the previous fact in order to generate a message from her input X. She generates N messages, one for each node in G.

For each $i \in V_L$ she generates the message (M_a^i, M^i), where

 − $M_a^i = \mathcal{P}(|G| + |H| - 2, i, N_G(i) \cup N_H(a) \setminus \{b\})$ is the message node i would send in the graph $G_{ij'} \oplus H_{ab}$ with $j' \in V_R$ arbitrary.
 − $M^i = \mathcal{P}(|G| + |H| - 2, i, N_G(i))$ is the message node i would send in the graph $G_{i'j'} \oplus H_{ab}$ with $i' \in V_L \setminus \{i\}$, $j' \in V_R$, both arbitrary.

For each $j \in V_R$ she generates the message (M_b^j, M^j), where

 − $M_b^j = \mathcal{P}(|G| + |H| - 2, j, N_G(j) \cup N_H(b) \setminus \{a\})$ is the message node j would send in the graph $G_{i'j} \oplus H_{ab}$ with $i' \in V_L$ arbitrary.
 − $M^j = \mathcal{P}(|G| + |H| - 2, j, N_G(j))$ is the message node j would send in the graph $G_{i'j'} \oplus H_{ab}$ with $i' \in V_L$, $j' \in V_R \setminus \{j\}$, both arbitrary.

Suppose that Bob sends (p, q) to the referee. How can the referee decide whether $(p, q) \in X$? He simply simulates protocol $\mathcal{P}$ considering for node p the message M_a^p, for node q the message M_b^q and for every other node r the message M^r (recall that H is fixed, known by the referee).

The size of the message sent by Alice is $\mathcal{O}(Nc(N + |H| - 2))$. Therefore, since the randomized complexity of INDEX is $\Omega(N^2)$ and $|H|$ is constant, we conclude that $c(N) = \Omega(N)$. □

Another reduction in the same style provides a lower bound in the case of bipartite H containing cycles.

Proposition 4. *Let H be a bipartite graph containing a cycle. Any randomized ϵ-error protocol that solves H-INDUCED SUBGRAPH or H–SUBGRAPH uses messages of size $\Omega(n^{2/k})$ where k is the (even) length of the shortest cycle in H.*

Proof. Let $N = ex(n, C_k)$ for some n, and let $G = (V, E)$ be a graph with n nodes and N edges which does not contain a subgraph isomorphic to C_k. Recall that $ex(n, C_k) = \Theta(n^{1+2/k})$ for even k.

Consider the following instance of the INDEX problem. Alice receives a vector $X \in \{0, 1\}^N$ and Bob receives a natural number $p \in [1, N]$. The question the

referee needs to answer is whether $X_p = 1$. We already know that any ϵ-error randomized protocol that solves this problem needs Alice to send $\Omega(N)$ bits.

Suppose that there exists a randomized ϵ-error protocol $\mathcal{P}$ that solves either H-INDUCED SUBGRAPH or H-SUBGRAPH using messages of size $c(n)$. We are going to use $\mathcal{P}$ to solve INDEX.

Let $e_1, e_2, \ldots, e_N$ be an enumeration of the edges of G, and consider the subgraph $G' = (V, E')$ of G with $e_i \in E' \iff X_i = 1$. Let a, b be two nodes of H such that the edge ab lies in a shortest cycle (that is, on a cycle of length k).

For any $e_i = (r, t)$, $1 \leq i \leq N$, consider the graph $G'_{rt} \oplus H_{ab}$. Then any crossing cycle of $G'_{rt} \oplus H_{ab}$ has length at least $k + 1$: $k_G \geq 2$ and $k_H = k - 1$. By definition, G' has no cycle of length k. Hence, any cycle of length k in $G'_{rt} \oplus H_{ab}$ must appear in the subgraph $\tilde{H}$ of $G'_{rt} \oplus H_{ab}$. Therefore, if H is a subgraph of $G'_{rt} \oplus H_{ab}$ then rt is an edge of G' as otherwise $\tilde{H}$ has fewer cycles of length k than H. And of course, conversely, presence of edge rt in G' means that $\tilde{H} \cong H$ is an induced subgraph of $G'_{rt} \oplus H_{ab}$.

Alice can take advantage of the previous fact in order to generate a message from her input X. She generates n messages, one for each node in G'. For each $i \in V$ she generates the message (M^i, M^i_a, M^i_b), where

- $M^i_a = \mathcal{P}(n + |H| - 2, i, N_{G'}(i) \cup (N_H(a) \setminus \{b\}))$ is the message node i would send in the graph $G'_{ij} \oplus H_{ab}$ with $j \in V \setminus \{i\}$ arbitrary.

- $M^i_b = \mathcal{P}(n + |H| - 2, i, N_{G'}(i) \cup (N_H(b) \setminus \{a\}))$ is the message node i would send in the graph $G'_{ij} \oplus H_{ba}$ with $j \in V \setminus \{i\}$ arbitrary.

- $M^i = \mathcal{P}(n + |H| - 2, i, N_{G'}(i))$ is the message node i would send in the graph $G'_{i'j'} \oplus H_{ab}$ with $i', j' \in V \setminus \{i\}$, both arbitrary.

Suppose that Bob sends p to the referee. How can the referee decide whether $X_p = 1$? If $e_p = (y, z)$ he simply simulates the protocol $\mathcal{P}$ considering for node y the message M^y_a, for node z the message M^z_b and for every other node i the message M^i (recall that H is fixed, known by the referee).

The size of the message sent by Alice is $\mathcal{O}(nc(n + |H| - 2))$. Therefore, since the randomized complexity of INDEX is $\Omega(N)$ we conclude that $c(n + |H| - 2)$ is $\Omega(N/n)$. We have $N = ex(n, C_k) = \Omega(n^{1 + 2/k})$, which proves the claim. $\quad\square$

Combining Propositions 3 and 4 with the observations of Section 1.2, we obtain the following:

Corollary 5 *If H is a forest, then the problem H-SUBGRAPH can be decided by a deterministic protocol with simultaneous messages of logarithmic size. If H contains a cycle, then a randomized protocol with simultaneous messages for H-SUBGRAPH requires messages of polynomial size.*

3 The problem H-INDUCED SUBGRAPH

Lemma 1. *Let H be a fixed graph. The problems H-INDUCED SUBGRAPH and $\overline{H}$-INDUCED SUBGRAPH are equivalent. More precisely, there exists a protocol*

with message size b for solving H-INDUCED SUBGRAPH if and only if there exists a protocol with message size b for solving $\overline{H}$-INDUCED SUBGRAPH.

Proof. Let H be a fixed graph. Suppose that we have a protocol for solving $\overline{H}$-INDUCED SUBGRAPH. We can use this protocol for solving H-INDUCED SUB-GRAPH as follows. Let G be the input graph. Note that H is an induced subgraph of G if and only if $\overline{H}$ is an induced subgraph of $\overline{G}$. Therefore, every node v can consider the nodes that are not its neighbors and apply the protocol for detecting $\overline{H}$ with this new, complementary neighborhood. Of course, if there is enough information for reconstructing $\overline{G}$ (when the answer is positive) then there is enough information for reconstructing G. $\qquad\square$

Corollary 6 *Let H be an arbitrary graph with at least 3 nodes. If $H \notin \{P_1 + P_2, P_3, 2P_2, C_4, P_4\}$ then any randomized ϵ-error protocol that solves H-INDUCED SUBGRAPH uses messages of size $\Omega(n)$.*

Proof. This follows directly from Proposition 3 and Lemma 1 because, the graphs listed above, are the only graphs with at least 3 nodes which are bipartite both themselves and their complements. $\qquad\square$

Comment 1 *Notice that $\overline{P_1 + P_2} = P_3$, $\overline{2P_2} = C_4$ and $\overline{P_4} = P_4$. Therefore, in order to understand completely problem H-INDUCED SUBGRAPH, the only problems we need to study are P_3-INDUCED SUBGRAPH, C_4-INDUCED SUBGRAPH and P_4-INDUCED SUBGRAPH. The case $H = C_4$ in Proposition 4 directly provides an $\Omega(n^{1/2})$ lower bound on the message size for any randomized ϵ-error protocol that solves C_4-INDUCED SUBGRAPH. Therefore, the only two cases for which we do not know the message size complexity yet are $H = P_3$ and $H = P_4$.*

3.1 The problem P_3-INDUCED SUBGRAPH

Notice that a graph is P_3-free if and only if it is the disjoint union of cliques. There is a classical randomized "fingerprint" technique for testing whether two vectors are equal. We are going to use this technique for solving P_3-INDUCED SUBGRAPH. It works as follows. Let $n^{c+3} < p \le 2n^{c+3}$ be a prime number. A value $t \in \mathbb{Z}_p$ is chosen uniformly at random using $\mathcal{O}(\log(n))$ public random bits. Given an n-bits vector $a = (a_1, \ldots, a_n)$, consider the polynomial $P_a = a_1 + a_2 X + a_3 X^2 + \ldots a_n X^{n-1}$ in $\mathbb{Z}_p[X]$ and let $FP(a, t) = P_a(t)$. The value $FP(a, t)$ is sometimes called the "fingerprint" of vector a. Clearly two equal vectors have equal fingerprints, and, more important, for any two different vectors a and b, the probability that $FP(a, t) = FP(b, t)$ is at most $1/n^{c+2}$ (because the polynomial $P_a - P_b$ has at most n roots and t was chosen uniformly at random, thus the probability that t is a root of $P_a - P_b$ is at most $1/n^{c+2}$, see [20]).

Proposition 7. *For any constant $c > 0$, P_3-INDUCED SUBGRAPH can be solved with a randomized $\mathcal{O}(\log n)$ message size protocol with $1/n^c$ error.*

Proof. Let $x_i \in \{0,1\}^n$ be the input vector of node i, i.e., the characteristic function of its closed neighborhood $N[i] = N(i) \cup \{i\}$. A protocol for P_3-Induced Subgraph consists in each node sending two numbers: its degree d_i and its fingerprint $m_i = FP(x_i, t)$. Let $l(m)$ be the number of nodes that send the same fingerprint m. The referee concludes that the input graph G is the disjoint union of cliques (and therefore P_3 is not an induced subgraph of G) if and only if all the nodes with the same fingerprint m have degree $l(m) - 1$. It is not difficult to realize that the previous protocol fails if and only if there are at least two nodes i, j with different neighborhoods such that $FP(m_i, t) = FP(m_j, t)$. For each fixed pair of nodes this probability is at most $1/n^{c+2}$, so altogether the probability of a wrong answer is at most $1/n^c$. $\qquad \square$

3.2   A deterministic protocol for P_3-Induced Subgraph

Recall that when more than one round is allowed the messages, instead of being sent to a referee, are written on a shared whiteboard.

Proposition 8. *There exists a $\mathcal{O}(\log n)$ message size deterministic two-round protocol for solving P_3-Induced Subgraph.*

Proof. Let G be the input graph. Our protocol does the following. In the first round each node v writes on the whiteboard its own ID together with the minimum ID of its closed neighborhood $M_v = \min\{\text{ID}(u) \mid u \in N_G[v]\}$. In the second round each node v writes only one bit. It writes the bit 1 if and only if for all $u \in N_G[v]$ $M_u = M_v$ and for all $u \notin N_G[v]$ $M_u \neq M_v$.

Obviously, every node writes a 1 in the second round if and only if G is a disjoint union of cliques. If G is indeed a disjoint union of cliques then, with the information written on the whiteboard, it is possible to reconstruct it. $\qquad \square$

Open Problem 1 *Is it possible to solve deterministically, in the* SM-BCAST *model, the problem P_3-Induced Subgraph using messages of size $\mathcal{O}(\log n)$?*

4   The problem P_4-Induced Subgraph

Let G_1 and G_2 be two disjoint graphs. The join operation $G_1 \star G_2$ consists in connecting all the nodes of G_1 with all the nodes of G_2. Formally, it is defined as follows: $G_1 \star G_2 = \overline{(\overline{G_1} + \overline{G_2})}$. The class of *cographs* is defined recursively. First, an isolated node K_1 is a cograph. Second, $G \neq K_1$ is a cograph if and only if G is the join or the union of two disjoint cographs [11, 23].

In this paper we provide a new characterization of cographs based on a new notion we introduce here that we call *twin ordering*. Two nodes u and v of a graph G are called twins if $N_G(u) \setminus \{v\} = N_G(v) \setminus \{u\}$.

A twin ordering of an n-node graph is an ordering $v_1, \ldots, v_n$ such that for each $j \geq 2$, the vertex v_j has a twin in the graph induced by $\{v_1, \ldots, v_j\}$.

Proposition 9. *For a graph G the following are equivalent.*

1. *G is a cograph.*
2. *Every non trivial induced subgraph of G has a pair of twins.*
3. *G is P_4-free.*
4. *G has a twin ordering.*

Proof. The equivalence between the first three characterizations was proved in [11] and [23]. It is clear that the second implies the fourth. Moreover, not only there exists a twin ordering, but one can find it by repeatedly picking an arbitrary node having a twin and removing such node. This follows from the assumption that every non trivial induced subgraph has a pair of twins.

We prove that if G has a twin ordering, then it is P_4-free. Take any subset of nodes $U = \{v_t, v_l, v_k, v_j\}$, with $t < l < k < j$. For the sake of contradiction, let us assume that the graph induced by U is P_4. Among the choices for U, pick one with j as small as possible. From hypothesis, there is a $i < j$ such that v_i and v_j are twins in the graph induced by $\{v_1, \ldots, v_j\}$. Since P_4 has no pairs of twins we get that $v_i \notin U$. But this is a contradiction with the choice of U because the graph induced by $\{v_t, v_l, v_k, v_i\}$ is P_4 and $\max\{t, l, k, i\} < j$. $\qquad\square$

Let $G = (V, E)$ be an n-node graph with $V \subseteq \mathbb{N}$. The *canonical ordering* of G is the twin ordering of G defined as follows. Instead of picking an arbitrary node having a twin we select the lexicographically first pair of twins. Then we choose, among these two nodes, the smaller one. The process continues, starting by removing v_n, until no further twins appear. So, a canonical ordering of an arbitrary graph is of the form $v_k, \ldots, v_n$ and $k = 1$ if and only if G is a cograph.

Let p be a prime and let $\phi = (\phi_w)_{w \in V}$ be a linearly independent family of polynomials in $\mathbb{Z}_p[X]$. Let $q = (q_w)_{w \in V}$ and $\bar{q} = (\bar{q}_w)_{w \in V}$ be defined by $q_w = \sum_{w' \in N_G(w)} \phi_{w'}$ and $\bar{q}_w = q_w + \phi_w$, for each $w \in V$. We also define $\alpha_{u,v} = \phi_u - \phi_v$, $\beta_{u,v} = q_u - q_v$ and $\gamma_{u,v} = \bar{q}_u - \bar{q}_v$, for each $u, v \in V$. The *derivated polynomials* of family ϕ are the following polynomials: $(\alpha_{u,v})_{u,v \in V}, (\beta_{u,v})_{u,v \in V}, (\gamma_{u,v})_{u,v \in V}$.

Let u and v be twins. We associate to $G - v$ the polynomials $(\phi'_w)_{w \in V \setminus \{v\}}$ by $\phi'_w = \phi_w$, when $w \neq u$, and $\phi'_u = \phi_u + \phi_v$. By using this construction, starting with $\phi_u(x) = x^u$, and following the canonical ordering $v_k, \ldots v_n$, we obtain polynomials ϕ_u^i, for each $k \leq i \leq n$ and each u in the graph $G - \{v_n, \ldots, v_{i+1}\}$. We call these polynomials the *basic* polynomials of G. The *canonical family of polynomials* of G is the union of basic polynomials and their derivated. This family has at most $n \times 3n^2 = 3n^3$ polynomials.

We say that a vector $m = ((a_w, b_w))_{w \in V} \in (\mathbb{Z}_p)^{2n}$ is *valid for G at $t \in \mathbb{Z}_p$* if there is a linearly independent family of polynomials $(\phi_w)_{w \in V}$ in $\mathbb{Z}_p[X]$ such that $a_w = \phi_w(t)$ and $b_w = q_w(t)$, for each $w \in V$.

Lemma 2. *Let $m = ((a_w, b_w))_{w \in V} \in (\mathbb{Z}_p)^{2n}$ be* valid *for G at t. Let u, v be twins in G such that $a_u \neq a_v$. Then, the following vector $m' = ((a'_w, b'_w))_{w \in V \setminus \{v\}} \in (\mathbb{Z}_p)^{2n-2}$ is valid for $G - v$ at t. For each $w \in V \setminus \{u, v\}$: $a'_w = a_w$ and $b'_w = b_w$. For $w = u$: $a'_u = a_u + a_v$ and $b'_u = b_u - a_v \delta_{uv}$, where $\delta_{uv} \in \{0, 1\}$ and $\delta_{uv} = 1$ if and only if $a_u + b_u = a_v + b_v$.*

Proof. Let $(\phi_w)_{w \in V}$ be a linearly independent family of polynomials associated to m. Since u and v are twins and $a_u \neq a_v$, we have that $a_u + b_u = a_v + b_v$ if and only if u and v are adjacent. Hence, $\delta_{uv} = 1$ if and only if u and v are adjacent.

Let $(\phi'_w)_{w \in V \setminus \{v\}}$ be given by $\phi'_w = \phi_w$ for each $w \neq u$, and $\phi'_u = \phi_u + \phi_v$. Clearly, this family is linearly independent.

For $w \neq u$ we have that $a'_w = a_w = \phi_w(t) = \phi'_w(t)$. Moreover, $b'_w = b_w$ and $b_w = q_w(t)$. Since u and v are twins either both are in $N_G(w)$ or none. In both cases we have that $b'_w = q'_w(t)$. By definition, $a'_u = a_u + a_v = \phi_u(t) + \phi_v(t) = \phi'_u(t)$. Also, by definition, $b'_u = b_u - \delta_{uv} a_v$. We know that $b_u = q_w(t) = \delta_{uv}\phi_v(t) + q'_w(t)$. Hence, $b'_u = q'_w(t)$. □

Proposition 10. *For any constant $c > 0$, P_4-Induced Subgraph can be solved with a randomized $\mathcal{O}(\log n)$ message size protocol with $1/n^c$ error.*

Proof. Let $G = (V, E)$ be an n-node graph. Let p be prime with $3n^{c+4} \leq p \leq 6n^{c+4}$. The protocol applied to G is the following. Each node sends to the referee the message m_w such that $m = (m_w)_{w \in V}$ is valid for G at t, where t is picked uniformly at random in $\mathbb{Z}_p$. Each node computes such a message by defining $\phi_w(x) = x^w$.

On input $m \in (\mathbb{Z}_p)^{2n}$ the referee iterates at most $n - 1$ times trying to build the canonical ordering $\{v_1, \dots, v_n\}$. In iteration i he starts with a graph G^i and a vector $m^i \in (\mathbb{Z}_p)^{2(n-i+1)}$ (with $G^1 = G$ and $m^1 = m$). He determines if there is a pair of nodes u and v in G^i such that $a^i_u \neq a^i_v$ and either $b^i_u = b^i_v$ or $a^i_u + b^i_u = a^i_v + b^i_v$ He selects, among all these, the lexicographically first pair of nodes. If no such pair exists, then he rejects. Otherwise, he sets $G^{i+1} = G^i - v$, setting $v_{n-i+1} = v$ (w.l.o.g, we assume that $v < u$). Then he computes m^{i+1} from m^i according to Lemma 2. If the referee reaches iteration $n - 1$ he accepts.

What is the probability that the referee does not construct the canonical ordering of G? This could happen only when the chosen t is a zero of at least one member of the canonical family of G. As this family has at most $3n^3$ polynomials, this occurs with probability at most $3n^3(n/p) \leq 1/n^c$. □

4.1 A deterministic protocol for P_4-Induced Subgraph

The definition of cographs by closure operations comes with the following natural hierarchy, which we call the *bottom-up hierarchy*, and which will be needed in the proof of Proposition 11. First, $\Sigma_0 = \Pi_0 = \{K_1\}$. Second, for $i \geq 0$, Σ_{i+1} is the set of disjoint unions of graphs in Π_i and Π_{i+1} is the set of joins of graphs in Σ_i. A graph G is a cograph if and only if $G \in \Sigma_i$ for some i. Notice that Σ_1 corresponds exactly to the class of disjoint unions of isolated nodes $K_1 + \dots + K_1$. On the other hand, Π_1 corresponds to the class of all cliques. More precisely, $\Pi_1 = \{K_n\}_{n>0}$, where K_n is the n-node clique.

Notice that $G \in \Sigma_i \iff \overline{G} \in \Pi_i$. We can prove this by induction. This is obviously true for $i = 1$. Assume now that $G = G_1 + G_2 \in \Sigma_i$ for some $i > 1$. The result follows from the induction hypothesis because $\overline{G} = \overline{G_1 + G_2} = \overline{G_1} \star \overline{G_2}$.

Σ_2 is exactly the class of P_3-free graphs because, as we saw previously, P_3-free graphs are exactly the disjoint unions of cliques. Since $\overline{P_3} = P_1 + P_2$ and

considering the previous observation, we conclude that Π_2 is the class of $(P_1 + P_2)$-free graphs.

Let G be a cograph. In [10] the authors define the *height* of G as the minimum i such that $G \in \Sigma_i \cup \Pi_i$. We do not know if there is a one-round deterministic protocol. However, any fixed level of the bottom-up hierarchy has a deterministic protocol with a bounded number of rounds:

Proposition 11. *Let $h > 0$ be a fixed positive integer. Then, there exists a $2(h-1)$-round protocol for the classes Σ_h and Π_h with messages of size $\log n$.*

Proof. We prove the existence/correctness/complexity of such protocols by induction on h. For $h = 2$ we use the two-round protocol of Proposition 8.

For the general case, first note that if the distance between two nodes is finite and strictly larger than 2, then P_4 is an induced subgraph of G. Let us now describe the protocol for Σ_h when $h > 2$ (the one for Π_h is symmetric). In the first round of the protocol, every node v writes on the whiteboard the minimum ID of the nodes in its closed neighborhood $N_G[v] = N_G(v) \cup \{v\}$. In the second round, every node v writes the minimum ID among the IDs written by the nodes in its closed neighborhood.

If the graph G is indeed P_4-free, then after these two rounds, every node knows the partition $G = G_1 + G_2 + \cdots + G_k$ of G into its connected components. In the third round, every node includes in its message the verification that this partition is correct: if some node in G_i is connected to a node in G_j with $i \neq j$ then, its message will state this fact. If this happens, then protocol concludes that G is not in Σ_h: in this case G is not even a cograph, because it contains an induced path of length 3, that is, a copy of P_4.

Assuming G is a cograph, every node knows its partition into connected components after the second round. Thus, in the third round, we can start performing the protocol for Π_{h-1} separately in each of the connected components G_i. If some G_i is not in Π_{h-1}, then $G \notin \Sigma_h$. If all of these graphs are in Π_{h-1}, the recursively called protocols reconstruct the graphs G_i, and our protocol for Σ_h reconstructs G as their disjoint union. $\square$

Open Problem 2 *Is it possible to solve deterministically, in the SM-BCAST model, the problem P_4-INDUCED SUBGRAPH using messages of size $\mathcal{O}(\log n)$?*

References

1. K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *Proc. of the 23rd Annual ACM-SIAM Symp. on Discrete Algorithms (SODA 2012)*, pages 459–467, 2012.
2. K. J. Ahn, S. Guha, and A. McGregor. Graph sketches: Sparsification, spanners, and subgraphs. In *Proc. of PODS 2012*, pages 5–14, 2012.
3. M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi. The exact solution of the Erdos-T. Sós conjecture for (large) trees, in preparation.
4. H. Arfaoui, P. Fraigniaud, D. Ilcinkas, and F. Mathieu. Distributedly testing cycle-freeness. In *WG 2014*, volume 8747 of *LNCS*, pages 15–28, 2014.

5. L. Babai and P. G. Kimmel. Randomized simultaneous messages: Solution of a problem of Yao in communication complexity. In *Proc. of the 12th Annual IEEE Conference on Computational Complexity*, pages 239–246, 1997.

6. F. Becker, M. Matamala, N. Nisse, I. Rapaport, K. Suchan, and I. Todinca. Adding a referee to an interconnection network: What can (not) be computed in one round. In *Proc. of IPDPS 2011*, pages 508–514, 2011.

7. F. Becker, I. Rapaport, P. Montealegre, and I. Todinca. The simultaneous number-in-hand communication model for networks: Private coins, public coins and determinism. In *Proc. of SIROCCO 2014*, pages 83–95, 2014.

8. J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *Journal of Combinatorial Theory, Series B*, 16(2):97–105, 1974.

9. A. Chakrabarti, Y. Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. of FOCS 2001*, pages 270–278, 2001.

10. M. Chudnovsky, A. Scott, and P. Seymour. Excluding pairs of graphs. *Journal of Combinatorial Theory, Series B*, 106:15–29, 2014.

11. D. G. Corneil, H. Lerchs, and L. S. Burlingham. Complement reducible graphs. *Discrete Applied Mathematics*, 3(3):163–174, 1981.

12. A. Drucker, F. Kuhn, and R. Oshman. The communication complexity of distributed task allocation. In *Proc. of PODC 2012*, pages 67–76, 2012.

13. A. Drucker, F. Kuhn, and R. Oshman. On the power of the congested clique model. In *Proc. of PODC 2014*, pages 367–376, 2014.

14. A. Gronemeier. Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness. In *Proc. of STACS 2009*, pages 505–516, 2009.

15. S. Guha, A. McGregor, and D. Tench. Vertex and hyperedge connectivity in dynamic graph streams. In *Proc. of PODS 2015*, pages 241–247, 2015.

16. S. Holzer and N. Pinsker. Approximation of distances and shortest paths in the broadcast congest clique. *CoRR*, abs/1412.3445, 2014.

17. T. S. Jaymar. Hellinger strikes back: A note on the multi-party information complexity of AND. In *Random–Approx*, volume 5687 of *LNCS*, pages 562–573, 2009.

18. A. V. Kostochka. Lower bound of the hadwiger number of graphs by their average degree. *Combinatorica*, 4(4):307–316, 1984.

19. I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

20. E. Kushilevitz. Communication complexity. *Adv. in Computers*, 44:331–360, 1997.

21. J. M. Phillips, E. Verbin, and Q. Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *Proc. of the 23rd Annual ACM-SIAM Symp. on Discrete Algorithms (SODA 2012)*, pages 486–501, 2012.

22. N. Robertson and P. D. Seymour. Graph minors XX: Wagner's conjecture. *Journal of Combinatorial Theory, Series B*, 92(2):325–357, 2004.

23. D. Seinsche. On a property of the class of n-colorable graphs. *Journal of Combinatorial Theory, Series B*, 16(2):191–193, 1974.

24. A. Thomason. An extremal function for contractions of graphs. In *Mathematical Proc. of the Cambridge Philosophical Society*, volume 95, pages 261–265, 1984.

25. A. Thomason. The extremal function for complete minors. *Journal of Combinatorial Theory, Series B*, 81(2):318–338, 2001.

26. D. P. Woodruff and Q. Zhang. When distributed computation is communication expensive. In *DISC 2013*, volume 8205 of *LNCS*, pages 16–30, 2013.

27. A. Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. of STOC 1979*, pages 209–213, 1979.