

Distributed Subgraph Finding: Progress and Challenges

Keren Censor-Hillel  

Department of Computer Science, Technion, Haifa, Israel

Abstract

This is a survey of the exciting recent progress made in understanding the complexity of distributed subgraph finding problems. It overviews the results and techniques for assorted variants of subgraph finding problems in various models of distributed computing, and states intriguing open questions.

2012 ACM Subject Classification Networks → Network algorithms; Theory of computation → Distributed algorithms

Keywords and phrases distributed algorithms, subgraph finding, limited bandwidth

Digital Object Identifier 10.4230/LIPIcs.ICALP.2021.3

Category Invited Talk

Funding This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 755839.

Acknowledgements The author thanks Francois Le Gall for clarifications and Dean Leitersdorf for useful feedback on a preliminary version of this survey.

1 Introduction

Distributed subgraph finding is motivated by considering a user in a network whose connections are unknown to its users. Typical examples could be friends or followers in social networks, or computing devices in large networks. A prime question is what is the local structure of the network, e.g., do two connected users share common connections, or are they part of a slightly larger cycle or clique? While some social networks provide the user with information about common connections they have with their connections, the general fundamental question that arises is that of finding small subgraphs in such distributed settings. This type of questions also emerges when processing huge graphs by distributed systems. For example, Hirvonen, Rybicki, Schmid, and Suomela [33] study distributed algorithms for finding large cuts in triangle-free graphs, and Pettie and Su [46] study coloring in triangle-free graphs.

This article surveys the state-of-the-art in distributed subgraph finding. We focus on synchronous settings, in which the main complexity measure is the number of communication rounds that is required in order to find a subgraph H . We will elaborate on the computational models and on the possible interpretations of what it means to *find* a subgraph, but let us start with a warm-up.

Warm-up: locality. The first simple observation is that in a synchronous network with no additional restrictions, the round complexity of finding a subgraph H by the devices in the network is $\Theta(k)$, where k is the diameter of H . The reason for this is that in a single round of communication, all nodes of the network can learn the neighbors of their neighbors,¹ and by induction, within t rounds each node can learn the topology within its t -neighborhood.

¹ This assumes that all nodes begin with knowledge of their neighbors. Removing this assumption incurs another round of communication to this argument.



© Keren Censor-Hillel;

licensed under Creative Commons License CC-BY 4.0

48th International Colloquium on Automata, Languages, and Programming (ICALP 2021).

Editors: Nikhil Bansal, Emanuela Merelli, and James Worrell; Article No. 3; pp. 3:1–3:14

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



To be more precise, what we get is that within $O(k)$ rounds, each node is able to *list* all the instances of H to which it belongs. This is the most powerful variant of subgraph finding.

The clear drawback of the above example is that it sweeps the under the carpet the complexity of sending potentially very large messages in a single round. Therefore, while the above classic LOCAL model [42] is suitable for studying many other tasks, it misrepresents the complexity of subgraph finding. We will thus focus mainly on two distributed settings that capture limited bandwidth, and in which the complexity of some subgraph finding problems is related. Additional settings will be discussed towards the end of the survey.

Computational models. The standard model that imposes restrictions on the bandwidth over the above setting is the CONGEST model [45]. This is a synchronous model, in which each of n nodes can send a message to each of its neighbors in every round, where the size of messages is bounded by $O(\log n)$ bits. This choice of this bound on the size of messages allows sending a node identifier within a single message. Typically, the graph that is processed is the graph that underlies the communication network.

Another important model is the CONGESTED CLIQUE model [43], which we will abbreviate as the CLIQUE model in this document, in which the n nodes are part of a fully connected network, with the same message bound of $O(\log n)$ bits as in the CONGEST model. Here, the input graph is an arbitrary graph G on n nodes, which is typically assigned through a bijection to the nodes of the CLIQUE, such that each node receives as input the edges in G that are adjacent to its assigned node.

Subgraph finding. The extreme case mentioned above, where each node in the system lists all instances of H to which it belongs is referred to as *membership listing* (or sometimes *local listing* or *local enumeration*). In settings with bounded bandwidth, this is typically a difficult variant that is known to require many rounds of computation for many subgraphs (see, e.g., discussion about triangle finding in Section 2). A weaker variant is that of *listing* or *enumeration*, in which it is required that every instance of the subgraph H is output by some node, but not necessarily a node which belongs to that instance.

Apart from these two listing variants, it is in many cases important to merely *detect* whether the input contains a copy of H or not. The standard formalization of distributed detection is that if there is no copy of H then all nodes output **false**, and otherwise at least one node outputs **true**. The reason for not demanding a unanimous output is to avoid incurring an overhead just for propagating the information of existence of H , for example if a graph contains a single copy of H and has many nodes that are very far from that instance.

Similarly to the case of listing, the detection problem also has a membership variant, which requires each node to output **true** or **false** based on whether it is a part of a copy of H or not.

Outline. Section 2 overviews the case in which H is a triangle. It covers both the CLIQUE and the CONGEST models. Sections 3 and 4 address larger cliques and larger cycles in both models, respectively. Finally, Section 5 briefly mentions additional subgraphs and additional settings.

2 Triangle Finding

A naïve simulation of the warm-up algorithm for membership listing in the CONGEST or CLIQUE models, in which all neighbors are sent to each other neighbor one by one, gives a trivial $O(\Delta)$ -round algorithm for triangle membership listing, where Δ is the maximum degree in the graph. However, it is possible to do better, as we overview in this section.

2.1 Triangle Finding in the CLIQUE Model

We begin with the CLIQUE model.

Triangle listing in the CLIQUE model. The first non-trivial algorithm for triangle finding is due to Dolev, Lenzen, and Peled [19]. This is a deterministic triangle listing algorithm for the CLIQUE model, which has a complexity of $O(n^{1/3}/\log n)$ rounds. The simplicity of this algorithm turned out to be a huge advantage for later additional results, as we will see. The algorithm works as follows: The vertices of the graph are partitioned into $n^{1/3}$ subsets $S_1, \dots, S_{n^{1/3}}$, each of $n^{2/3}$ nodes. Each of the n nodes receives a different tuple of three of these subsets. A node that receives $S_{i_1}, S_{i_2}, S_{i_3}$ for indices $1 \leq i_1, i_2, i_3 \leq n^{1/3}$ (that are not necessarily different) collects all edges with one endpoint in one of the three subsets and one endpoint in another, that is, this node collects all edges in $E(S_{i_1}, S_{i_2}) \cup E(S_{i_1}, S_{i_3}) \cup E(S_{i_2}, S_{i_3})$, and reports all triangles that it finds. It is straightforward to see that all triangles are listed by this algorithm since the number of 3-tuples of subsets is n and so each is handled by some node.

The round complexity of the algorithm follows by proving that each node needs to send and receive $O(n^{4/3})$ edges in total, which are to and from locations that are known to all nodes (we will discuss this knowledge property later), since the partition to subsets is hardcoded and so it is known to all nodes. **Sending:** Take a node v and assume that it is in the subset S_i . There can be at most $n^{2/3}$ edges between v and nodes in S_j and these edges need to be sent to all nodes that have S_i and S_j in their 3-tuple. Since there are $n^{1/3}$ such 3-tuples, these $n^{2/3}$ edges need to be sent to $n^{1/3}$ nodes. Repeating this for all $n^{1/3}$ possibilities for j gives a total of $n^{2/3+1/3+1/3} = n^{4/3}$ edges that v has to send. **Receiving:** Each node needs to learn 3 subsets of edges, each containing at most $n^{2/3} \cdot n^{2/3} = n^{4/3}$ edges. To conclude the complexity analysis, one can use the simple claim that [19] proves, which states that a routing task in which each node needs to send and receive n messages in a known pattern can be done in 2 rounds. This means that the $O(n^{4/3})$ sent and received messages per node are divided by n , yielding a complexity of $O(n^{1/3})$ rounds. Noticing that the partition and routing are fixed, one can refrain from sending actual edge identifiers and replace them with a bit mask, which saves a logarithmic factor and results in a complexity of $O(n^{1/3}/\log n)$ rounds.

This complexity turns out to be optimal. The first lower bound for this task was of $\Omega(n^{1/3}/\log^3 n)$ rounds given by Pandurangan, Robinson and Scquizzato [44], and it was followed by a tight lower bound of $\Omega(n^{1/3}/\log n)$ rounds given by Izumi and Le Gall [36]. The lower bound is proven using an information-theoretic argument, which bounds by $\Omega(n^{4/3})$ the entropy of the transcript that a certain node sees on a random graph in which each edge appears independently at random with probability $1/2$. As the entropy of the transcript is a lower bound on its length, and since each node can receive at most $O(n \log n)$ bits per round, the lower bound on the round complexity follows. This approach also allowed [36] to obtain a lower bound of $\Omega(n/\log n)$ rounds for *local* listing of triangles, in which each node needs to output a list of all the triangles which include it.

In addition to the aforementioned algorithm, [19] also gives a triangle-listing algorithm whose complexity improves upon the above in the case that the graph contains *many* triangles. Specifically, they provide a randomized algorithm that completes in $O(n^{1/3}/(t^{2/3} + 1) + 1)$ rounds in expectation, where t is the number of triangles in the graph, and in $O(\min\{n^{1/3} \log^{2/3} n/(t^{2/3} + 1), n^{1/3}\})$ rounds with high probability.

Triangle detection in the CLIQUE model. The approach of [19] inherently lists all triangles. A natural question is whether the decision problem of triangle detection is easier than listing, for which the answer turns out to be affirmative. Censor-Hillel, Kaski, Korhonen, Lenzen, Paz, and Suomela [12] show how to perform matrix multiplication over a ring in the CLIQUE model within $O(n^{0.158})$ rounds. More precisely, the complexity is $O(n^{1-2/\omega})$ rounds, where ω is the matrix multiplication exponent, currently known to be bounded by 2.3728596 due to Alman and Vassilevska-Williams [2]. Ring matrix multiplication directly carries over to triangle detection with the same complexity. The main approach of [12] is to simulate matrix multiplication algorithms for parallel settings in the CLIQUE model. This includes the so-called parallel 3D matrix multiplication, as well as bilinear Strassen-like algorithms.

Elaborating upon the matrix multiplication algorithms is beyond the scope of this survey. Yet, we must mention a crucial component that underlies these CLIQUE algorithms, as well as numerous additional CLIQUE algorithms for various tasks, which is the ingenious routing technique of Lenzen [41]. In a nutshell, this technique provides a way to route in $O(1)$ rounds any set of messages in which each node needs to send and receive at most n messages.

The above complexity of $O(n^{0.158})$ rounds for triangle detection in the CLIQUE model is the current state-of-the-art. For reasons that will be explained shortly, unlike the $\Omega(n^{1/3})$ lower bound for triangle listing in this model, there is no known lower bound for triangle detection. We thus have the following major open problem in distributed triangle finding.

► **Open Problem 2.1.** *What is the complexity of triangle detection in the CLIQUE model? In particular, is there an algorithm that is faster than $O(n^{0.158})$ rounds? Can any lower bound be proven?*

The second part of the above question which asks for a lower bound for triangle detection in the CLIQUE model is considered hard: Very roughly speaking, Drucker, Kuhn, and Oshman [20] show that the CLIQUE model is strong enough to simulate certain circuits, which implies that for a wide range of problems, any super-constant lower bound in the CLIQUE model would result in a major breakthrough in circuit complexity. It is noteworthy that the problem of triangle listing does not fall into this category of problems due to its large outputs, which explains why this does not contradict the aforementioned lower bound of $\Omega(n^{1/3})$ rounds for triangle listing.

For the broadcast version of this model, a.k.a. the BROADCAST CONGESTED CLIQUE model, in which the messages sent by a node in a certain round must all be identical, an $\Omega(n/e^{O(\sqrt{\log n})} \log n)$ -round lower bound on deterministic triangle detection is given by [20], through a reduction from 3-party number-on-forehead set disjointness.

Triangle listing in sparse graphs in the CLIQUE model. The first algorithms for triangle finding in sparse graphs were given by Dolev, Lenzen, and Peled [19]. One algorithm has a round complexity of $O(\Delta^2/n + 1)$ and another has a round complexity of $O(A^2/n + \log_{2+n/A^2} n)$, where A is the arboricity of the graph.² The latter implies a round complexity of $O(m^2/n^3)$, in terms of the number of edges, m .

Pandurangan, Robinson, and Squizzato [44] give a randomized triangle listing algorithm that completes within $\tilde{O}(m/n^{5/3} + 1)$ rounds, w.h.p. At the heart of the algorithm lies the partitioning approach of [19], but more is needed in order to exploit sparsity. In [44], the

² The arboricity of a graph is the minimal number of forests that contain all of its edges. While bounded by the maximum degree Δ , the arboricity can in some cases be much smaller, e.g., a star has a linear maximum degree but its arboricity is 1.

partition is random, and the routing of edges to the nodes to which they are assigned is done in a randomized load balanced manner. This approach is what handles load balancing of the information that needs to be routed in the system in a way which is sensitive to the sparsity of the graph, rather than optimizing only for the worst case. Moreover, the aforementioned $\tilde{\Omega}(n^{1/3})$ lower bound of [44] for triangle listing in general graphs follows from a $\tilde{\Omega}(m/n^{5/3})$ lower bound for graphs with m edges.³

Censor-Hillel, Leitersdorf, and Turner [14] obtained an algorithm for sparse graphs, with a complexity of $O(m/n^{5/3} + 1)$ which is similar to, but slightly improves upon, the complexity of the algorithm of [44], by polylogarithmic factors. The algorithm of [14] is deterministic and is based on an algorithm for sparse matrix multiplication, which completes in $O((\rho_S \rho_T n)^{1/3} + 1)$ rounds, where S and T are the input matrices and ρ_A is the average number of non-zero elements per row of a matrix A (i.e., it is the number of non-zero elements of A , divided by n).⁴ While the aforementioned semi-ring matrix multiplication algorithm of [12] assigns the n^3 element-wise multiplication to the nodes in an optimal manner, it applies to the worst case. The general approach used by [14] for sparse matrix multiplication is to assign the element-wise multiplications to nodes in a way which is optimal given the input sparsity. The way this is done is by load balancing the number of non-zero elements that each node needs to send and receive in order for all element-wise multiplications to be computed.

Notes on matrix multiplication in the CLIQUE model. The first algorithm for matrix multiplication in this model is due to Drucker, Kuhn, and Oshman [20], who showed a randomized complexity of $O(n^{\omega-2}) \approx O(n^{0.373})$ rounds with high probability, for semirings. A deterministic $O(n^{1/3})$ -round algorithm for matrix multiplication over a semiring is given in [12].

Censor-Hillel, Dory, Korhonen, and Leitersdorf [8] provide two additional sparse matrix multiplication algorithms, used for distance computations. Compared to [14], these algorithms also benefit from sparsity of the output matrix $P = ST$. More concretely, their first algorithm completes in $O((\rho_S \rho_T \rho_P)^{1/3} / n^{2/3} + 1)$ rounds. This matches the complexity algorithm of [14] for a general P , but improves upon it for sparse output matrices (recall that the multiplication of two sparse matrices need not be sparse in general). The second algorithm of [8] pays an additive $\log n$ over the first one, i.e., has a round complexity of $O((\rho_S \rho_T \rho)^{1/3} / n^{2/3} + \log n)$, but here ρ_P is replaced by ρ , which stands for the number of elements per row that are needed. That is, there is no need to compute any element in the output matrix P that is not among the ρ smallest ones in its row (given an appropriate definition of the total order on matrix elements). In other words, the complexity of this algorithm does not depend on the sparsity of $P = ST$, but on some sparsity parameter that is given as input.

Additional algebraic algorithms in the CLIQUE model with important applications were given by Le Gall [27]. These include fast algorithms for rectangular matrix multiplications, as well as for multiple instances of matrix multiplication.

As explained earlier, we do not expect a lower bound for matrix multiplication in the CLIQUE model. However, it is shown in [12] that a near-linear number of rounds is required in the BROADCAST CONGESTED CLIQUE model.

Finally, we mention that matrix multiplication based algorithms also give results for *counting* for some small subgraphs [12, 14].

³ In fact, the results of [44] hold for the k -machine model (see Klauck, Nanongkai, Pandurangan, and Robinson [39]), which is similar to the CLIQUE model, but has k nodes rather than n . For a general k , the complexity of the algorithm is $\tilde{O}(m/n^{5/3} + n/k^{4/3})$, and the lower bound is $\tilde{\Omega}(m/k^{5/3})$.

⁴ The sparse matrix multiplication algorithm of [14] can also be converted to the k -machine model, in which it has a complexity of $O(n^{4/3}(\rho_S \rho_T)^{1/3} / k^{5/3} + 1)$ rounds.

2.2 Triangle Finding in the CONGEST Model

We now move to triangle finding problems in the CONGEST model.

The first breakthroughs in this model are the first non-trivial (sublinear) algorithms due to Izumi and Le Gall [36]. These are randomized algorithms for triangle listing and detection in $O(n^{3/4} \log n)$ and $O(n^{2/3} \log^{2/3} n)$ rounds w.h.p., respectively. The approach of these algorithms is to split the task of finding triangles into two parts: one which looks for triangles that are ϵ -heavy and another which looks for other triangles, where an ϵ -heavy triangle is one in which at least one of its edges appears in at least n^ϵ triangles. Very roughly speaking, heavy triangles can be detected within $O(n^{1-\epsilon})$ rounds by randomly sampling which edges to send, and can be listed in $O(n^{1-\epsilon/2})$ rounds by randomly hashing the edges sent to different neighbors. A more involved argument shows that non-heavy triangles can be listed in $O(n^{1-\epsilon} + n^{(1+\epsilon)/2} \log n)$ rounds, w.h.p. Carefully plugging in the right values of $n^\epsilon = \tilde{O}(n^{1/3})$ and $n^\epsilon = \tilde{O}(n^{1/2})$ then gives the round complexities for triangle detection and listing.

The breakthrough that came after [36] is due to Chang, Pettie, and Zhang [15], who showed triangle listing (and thus also detection) in the CONGEST model in $\tilde{O}(n^{1/2})$ rounds, w.h.p. In a nutshell, the main methodology of this algorithm has two elements: One element is an algorithm for decomposing the graph into well-connected components which could behave somewhat similarly to the CLIQUE model. The second element is to have the nodes of each component search for triangles that have an edge in the component.

In more detail, [15] showed how to compute the following expander decomposition in $O(n^{1/2})$ rounds. This decomposition is a partition of the set of edges of the graph into three sets. In the first set E_m , each connected component has minimum degree $n^{1/2}$ and conductance $\Omega(1/\text{polylog } n)$. The graph induced by the edges in the second set, E_s , is such that its arboricity is at most $n^{1/2}$. The third set of remaining edges, E_r , is at most a constant fraction of the total number of edges, and the triangle listing algorithm which we discuss in what follows recurses over this remaining set of edges. The algorithm for the decomposition itself is beyond the scope of this survey, and here we only mention that it is rather far from being merely a distributed implementation of its centralized counterpart.⁵

The triangle listing algorithm over the two first sets above then works as follows. Since the arboricity in E_s is bounded by $n^{1/2}$, triangles with at least one edge in this set are listed in a straightforward manner, using an orientation that is deduced by the decomposition algorithm. Then, triangles with an edge in any high-conductance component (i.e., in E_m) are listed by having the nodes of each component mimic a variant of the triangle listing algorithm in the CLIQUE model of Dolev, Lenzen, and Peleg [19]. This mimicking has three aspects. First, it replaces the deterministic partition of [19] with a randomized partition, in order to have a good probability for a good balance of information within the component. Second, the fact that a component has large conductance indeed implies that it has a small mixing time, but this alone is insufficient for efficiently exchanging large amounts of information within the component. To this end, the algorithm makes use of the random-walk based routing techniques of Ghaffari, Kuhn, and Su [28] and Ghaffari and Li [29], whose complexities improve as the mixing time decreases. Finally, even with these ingredients, some nodes of a component may have too many edges that touch them that are not inside the component, preventing them from efficiently using their inner-component edges for routing this large amount of information. Thus, some additional edges are added to the set E_r of remaining edges that the decomposition left for recursing over.

⁵ The actual decomposition result is more general, with a parameter δ which is tuned here to be $\delta = 1/2$ in order to optimize the running time of the triangle listing algorithm that uses it.

The decomposition approach was refined by Chang and Saranurak [16], allowing the complexity of triangle listing to drop to $\tilde{O}(n^{1/3})$ rounds, w.h.p. Since the $\tilde{\Omega}(n^{1/3})$ lower bound for triangle listing in the CLIQUE model directly applies also in CONGEST, this complexity is near-optimal (up to polylog n factors). At a high level, the decomposition obtained in [16] improves upon the one in [15] in that it does not have E_s at all (the bounded arboricity part) and in the complexity of obtaining it. In addition, a somewhat modified version of the routing algorithms of [28, 29] is introduced, for the triangle listing usage of the decomposition.

Chang and Saranurak [17] then show deterministic algorithms for expander decomposition and for expander routing. These yield round complexities of $O(n^{0.58})$ and $n^{2/3} + o(1)$ for deterministic triangle detection and triangle listing, respectively.

An additional algorithm for deterministic triangle listing is given by Huang, Pettie, Zhang, and Zhang [34]. The complexity of this algorithm, given in terms of the maximum degree Δ , is $O(\Delta/\log n + \log \log \Delta)$ rounds, w.h.p. For $\Delta = \tilde{O}(n^{1/3})$, this is faster compared with the algorithm of [17], while the latter is faster for the larger range of Δ .

Since the complexities of both [17] and [34] do not reach yet the lower bound of $\tilde{\Omega}(n^{1/3})$ rounds, we note the following open question.

► **Open Problem 2.2.** *What is the complexity of deterministic triangle listing in the CONGEST model? Does randomization help for this problem?*

The above are triangle listing algorithms so they clearly also solve the detection variant. However, as opposed to triangle listing, for triangle detection we currently do not have good lower bounds. What we do know is the following. Abboud, Censor-Hillel, Khoury, and Lenzen [1] showed that triangle detection cannot be solved in the CONGEST model within a single round by a deterministic algorithm. Specifically, they showed that any single-round algorithm requires a bandwidth of $\Delta \log n$ bits. Fischer, Gonen, Kuhn, and Oshman [24] showed that this also holds for randomized algorithms, by showing that randomized single-round algorithms require a bandwidth of Δ bits. Both works also addressed the round complexity of 1-bit bandwidth algorithms, with a lower bound of $\Omega(\log^* n)$ rounds given in [1], which was improved to $\Omega(\log n)$ rounds in [24].⁶

We will later (Section 4) describe some lower bounds for finding other subgraphs in the CONGEST model which are based on reductions from 2-party communication complexity problems, and there we will see why these techniques do not give any meaningful lower bound for triangles. Moreover, in the spirit of the aforementioned argument of Drucker, Kuhn, and Oshman [20], Eden, Fiat, Fischer, Kuhn, and Oshman [21] showed that a lower bound of $\omega(\log n)$ for triangle detection in CONGEST would imply major breakthroughs in circuit complexity. This leaves us with another curious gap in our knowledge of triangle finding in this model.

► **Open Problem 2.3.** *What is the complexity of triangle detection in the CONGEST model (randomized and deterministic)?*

3 Finding Larger Cliques

The CLIQUE model. The deterministic triangle listing algorithm of [19] for the CLIQUE model can be easily generalized to list larger cliques. To find cliques of size p for some integer $p \geq 3$, the set of nodes is partitioned into $n^{1/p}$ sets. Each node is assigned a p -tuple of sets

⁶ The function $\log^* n$ counts the number of times the logarithm function needs to be applied starting from n until the value drops to at most 1.

and learns all edges between any two of its sets. A similar argument to that of triangles shows that indeed all p -cliques are listed by this algorithm, and that its round complexity is $O(n^{1-2/p}/\log n)$. In fact, it is easy to see that this algorithm lists all instances of any subgraph H of p nodes.

This complexity is optimal, due to a lower bound of $\tilde{\Omega}(n^{1-2/p})$ rounds by Fischer, Gonen, Kuhn, and Oshman [24], which generalizes the aforementioned lower bound for triangle listing of [44] and [36], using additional machinery. In the BROADCAST CONGESTED CLIQUE model, Drucker, Kuhn, and Oshman [20] show that $\Omega(n/\log n)$ rounds are needed for p -cliques, even for the stronger detection variant, for almost all values of $p \geq 4$ (as long as $p \leq (1 - \epsilon)n$ for some constant $\epsilon > 0$).

For sparse graphs, Censor-Hillel, Le Gall, and Leitersdorf [11] show that listing can be complete within $\tilde{O}(m/n^{1+2/p} + 1)$ rounds, for $p \geq 3$. This follows from their CONGEST approach which is discussed below, and is tight up to polylogarithmic factors using the lower bound technique of [24, 36, 44].

As opposed to the listing variant, for p -clique detection the aforementioned hardness of obtaining lower bounds in the CLIQUE model by [20] kicks in, and we do not know any non-constant lower bound (or any larger than 1 lower bound, for that matter), leaving the complexity of p -clique detection open for $p > 4$.

► **Open Problem 3.1.** *What is the complexity of p -clique detection in the CLIQUE model (randomized and deterministic), for $p \geq 4$?*

The CONGEST model. Algorithms for finding larger cliques in the CONGEST model, as in the case of triangles, are also based on the conductance decomposition algorithms of Chang, Pettie, and Zhang [15] and of Chang and Saranurak [16]. The first sublinear algorithms for larger cliques in CONGEST are due to Eden, Fiat, Fischer, Kuhn, and Oshman [21]. They show that 4-cliques can be listed in $\tilde{O}(n^{5/6+o(1)})$ rounds and that 5-cliques can be listed in $\tilde{O}(n^{21/22+o(1)})$ rounds, w.h.p. These listing algorithms are more involved compared with the triangle listing algorithm, due to the need to handle, for example, a 4-clique with one edge within a certain component and another edge within a different component, which imposes a complex challenge that becomes even worse as p grows.

Following [21], the work of Censor-Hillel, Le Gall, and Leitersdorf [11] provided algorithms for listing p -cliques in $\tilde{O}(n^{p/(p+2)})$ rounds, w.h.p., for all $p \geq 4$ except for $p = 5$. For $p = 5$, this work gave an algorithm which completes in $\tilde{O}(n^{3/4+o(1)})$ rounds, w.h.p. The main approach of these algorithms is iterating over the decomposition in a way that balances the minimum degree and the arboricity thresholds. Within each cluster, sparsity-aware listing helps speeding up the computation. While these algorithms improved upon the state-of-the-art for $p = 4, 5$ and were the first sublinear algorithms for $p \geq 6$, they fell short of the [24] lower bound of $\tilde{\Omega}(n^{1-2/p})$ rounds. The question of whether clique listing in the CONGEST model is as easy as, or harder than, its CLIQUE counterpart, was recently answered by Censor-Hillel, Chang, Le Gall, and Leitersdorf [7], using the newer conductance decomposition of Chang and Saranurak [16] and additional mechanisms for optimal sparsity-aware listing and for efficient transmission of edges in the clusters.

For $p = 4$, the tight listing of [7] also implies that 4-clique detection is not easier than its listing counterpart in the CONGEST model. This is due to a lower bound of $\tilde{\Omega}(n^{1/2})$ for 4-clique detection in CONGEST, by Czumaj and Konrad [18]. The lower bound of [18] is more general, and says that detecting p -cliques in the CONGEST model requires $\tilde{\Omega}(n^{1/2}/p)$ rounds for $p \leq n^{1/2}$, and $\tilde{\Omega}(n/p)$ rounds for $p \geq n^{1/2}$. Thus, for values of p larger than 4, there is still a gap between detection and listing.

► **Open Problem 3.2.** *What is the complexity of p -clique detection in the CONGEST model (randomized and deterministic), for $p > 4$?*

The lower bound of [18] uses a reduction from 2-party set disjointness. The same work also shows that listing can be done sufficiently fast by the two parties, which shows that if the detection problems are harder, a different lower bound technique must be used.

4 Finding Larger Cycles

The CLIQUE model. Since the algorithm of [19] easily lists any subgraph of p nodes, we have that, in particular, p -cycles can be listed in the CLIQUE model within $O(n^{1-2/p})$ rounds (deterministically).

The detection variant was then addressed by Censor-Hillel, Kaski, Korhonen, Lenzen, Paz, and Suomela [12], who show a deterministic constant-round algorithm for detecting 4-cycles. In addition, they show an $O(n^{0.158})$ -round algorithm for detecting p -cycles, for any constant p . More accurately, the complexity is $2^{O(p)}n^{0.158}$ rounds. This is a randomized algorithm that relies on the color coding technique of Alon, Yuster, and Zwick [3], in order to avoid too much congestion that would be caused by collecting all possible candidates for cycle nodes. More recently, Censor-Hillel, Fischer, Gonen, Le Gall, Leitersdorf, and Oshman [10] showed that $2p$ -cycles can be detected in $O(1)$ rounds for any p , using a technique which also allows fast girth approximation. This is, notably, a deterministic algorithm. For odd values of p , it is still not known whether there is a constant-round p -cycle detection algorithm.

► **Open Problem 4.1.** *What is the complexity of p -cycle detection in the CLIQUE model (randomized and deterministic), for odd values of p ?*

The CONGEST model. For 4-cycle detection, Drucker, Kuhn, and Oshman [20] provide a tight complexity of $\tilde{\Theta}(n^{1/2})$, by providing both an algorithm and a lower bound. They also show that for any odd value of p , detecting p -cycles requires $\tilde{\Omega}(n)$ rounds. The upper bound is obtained by setting a threshold of $T = 2n^{1/2}$, and defining heavy nodes as nodes with at least T neighbors. First, all non-heavy nodes send all their neighbors to all their neighbors, in T rounds. This detects 4-cycles which have at least 2 non-neighboring non-heavy nodes. Then, a heavy node v with more than T heavy neighbors reports that there exists a 4-cycle. Finally, a heavy node v with at most T heavy neighbors sends its heavy neighbors to all of its neighbors, again in T rounds. The reason for which too many heavy neighbors imply a 4-cycle is because this means that the total number of neighbors of heavy neighbors is at least T^2 , which is more than $2n$, and so there must be a node other than v which is connected to two of the neighbors of v , and hence we have a 4-cycle. Otherwise, if v sends its heavy neighbors to all of its neighbors, then any 4-cycle with two neighboring nodes that are heavy is detected by one of its nodes (by a simple case analysis).

The matching lower bound for 4-cycles is a good point of reference for understanding lower bounds for the CONGEST model that rely on reductions from 2-party communication complexity. It relies on the fact that there is a 4-cycle-free graph G on n nodes with $\Theta(n^{3/2})$ edges, due to Erdős [22], whose work implies that this is the Turán Number of 4-cycles [48]. The setting is as follows. Each of two players, Alice and Bob, has an input string of k bits, $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$, respectively. By communicating with each other, the players need to compute the set disjointness function $Disj(x, y)$, whose value is 1 if and only if the input strings represent disjoint subsets of the set $\{1, \dots, k\}$, that is, if and only if there is no index $1 \leq i \leq k$ such that $x_i = y_i = 1$. The 2-party set disjointness problem is known

to require exchanging $\Omega(k)$ bits, even by randomized protocols, due to Kalyanasundaram and Schmitger [38], Razborov [47], and Bar-Yossef, Jayram, Kumar, and Sivakumar [4]. The reduction to distributed detection of 4-cycles is as follows. Each of the two players takes a subgraph of the 4-cycle free graph G according to their respective input. That is, the edges of G are mapped to the set $\{1, \dots, k\}$, with a value of n for which $k = \Theta(n^{3/2})$. Each player imagines a subgraph of G that has an edge for any index in which their input is 1. The players then imagine that their two subgraphs are connect by a perfect matching: each node in Alice's subgraph is connected to the respective node in Bob's subgraph (the nodes of both subgraphs are the nodes of G). Now, it is easy to see that the combined graph contains a 4-cycle if and only if the input strings x, y are disjoint, as the only 4-cycles that can occur are those that have two edges from the perfect matching, and two edges that represent the same index in the bit strings of the players. Thus, if Alice and Bob can simulate a distributed algorithm for 4-cycle detection, they solve set disjointness. They simulate a given algorithm as follows. Any message that the algorithm sends between two nodes that are in the same subgraph (either that of Alice or that of Bob) can be internally simulated by the respective player. The only messages that need to be explicitly sent between the two players are those that are sent along the edges of the perfect matching. There are n such edges, and so simulating a round of the detection algorithms costs $O(n \log(n))$ bits of communication between the two players. Since $\Omega(k) = \Omega(n^{3/2})$ is a lower bound on the total number of bits that need to be exchanged for solving set disjointness, we get that the distributed 4-cycle detection algorithm must consist of at least $\Omega(k/n \log(n)) = \Omega(n^{3/2}/n \log(n)) = \tilde{\Omega}(n^{1/2})$ rounds.

A remark on triangle detection. While this approach for reductions which imply CONGEST lower bounds is heavily used in the literature, it is doomed to fail for triangle detection. The reason is that any triangle in a graph has at least one player which knows at least two of its nodes and thus all of its edges, which nullifies any attempt for a lower bound constructions, regardless of the 2-party problem or the graph construction.

Korhonen and Rybicki [40] then showed that the $\tilde{\Omega}(n^{1/2})$ lower bound for 4-cycles can be extended and applies to p -cycles for all even values of p . Notably, in the spirit of the results of [20], it is shown in Censor-Hillel, Fischer, Gonen, Oshman, Le Gall, and Leitersdorf [10] that going above this lower bound for $p = 6$ would imply new lower bounds in circuit complexity, which are considered hard to obtain. The reason that this barrier for lower bounds works in the CONGEST model is because it is shown that it is sufficient to consider high-conductance clusters, and that these can simulate circuits in a similar manner to the CLIQUE model.

On the upper bound side, Korhonen and Rybicki [40] showed that p -cycle detection for any constant p can be done in a linear in n number of rounds. For odd values of p , this is optimal due to the above lower bound. They also showed that this can be done faster for degenerate graphs. Fischer, Gonen, Kuhn, and Oshman [24] showed how to detect $2p$ -cycles within $O(n^{1-1/p(p-1)})$ rounds, for $p \geq 2$. This was subsequently improved by Eden, Fiat, Fischer, Kuhn, and Oshman [21], who showed how to detect $2p$ -cycles in $\tilde{O}_p(n^{1-2/(p^2-p+2)})$ rounds for odd $p \geq 3$, and in $\tilde{O}_p(n^{1-2/(p^2-2p+4)})$ rounds for even $p \geq 4$. They also show that as opposed to the case of 4-cliques, the listing variant of 4-cycles is harder than its detection counterpart, requiring $\tilde{\Omega}(n)$ rounds. For $p = 3, 4, 5$, Censor-Hillel, Fischer, Gonen, Oshman, Le Gall, and Leitersdorf [10] then showed improved algorithms (randomized) for detecting $2p$ -cycles, which completes within $\tilde{O}(n^{1-1/p})$ rounds. Still, the upper bounds here are above the respective $\tilde{\Omega}(n^{1/2})$ lower bound.

► **Open Problem 4.2.** *What is the complexity of p -cycle detection in the CONGEST model (randomized and deterministic), for even value of $p \geq 6$?*

5 Additional Variants of Distributed Subgraph Finding

Finding Other Subgraphs. The literature has also been studying additional subgraphs, apart from cliques and cycles.

Drucker, Kuhn, and Oshman [20] study the complexity of finding trees and complete bipartite subgraphs in the BROADCAST CONGESTED CLIQUE model. Korhonen and Rybicki [40] show that trees can be detected in $O(1)$ rounds in the BROADCAST CONGEST model.

On the lower bound front, Gonen and Oshman [31] showed lower bounds for finding subgraphs that are created from smaller ones using certain allowed operations. Fischer, Gonen, Kuhn, and Oshman [24] showed that for any $p \geq 4$, there exists a subgraph H of size p such that H detection requires $\tilde{\Omega}(n^{2-\Theta(1/p)})$ rounds. Eden, Fiat, Fischer, Kuhn, and Oshman [21] showed that this complexity is roughly the right one, by providing an upper bound of $\tilde{O}(n^{2-2/(3p+1)} + o(1))$ rounds. In particular, their result shows that there does not exist a constant-sized subgraph for which detection requires a truly quadratic number of rounds.

Subgraph Freeness Testing. The variant of testing for H -freeness has also been studied in the distributed setting, initially introduced by Brakerski and Patt-Shamir [6].

The definition of distributed testing for H -freeness requires that if there is no instance of H in the graph then all the nodes output **true**, but instead of requiring at least one node to output **false** in case there is an instance of H as would be an analog to the detection problem, testing only requires at least one node to output **false** in case the graph is far from being H -free. Here, being ϵ -far from having a property is the same as its standard definition by Goldreich, Goldwasser, and Ron [30], and means that no matter which ϵ -fraction of the graph is changed (removed, in the case of H -freeness), the resulting graph must still have a copy of H . For triangles and other small subgraphs, this typically means that there are many instances of H in a graph that is far from being H -free.

For testing triangle-freeness, Censor-Hillel, Fischer, Schwartzman, and Vasudev [9] showed a $O(1/\epsilon^2)$ -round algorithm. This was later improved by Even, Fischer, Fraigniaud, Gonen, Levi, Medina, Montealegre, Olivetti, Oshman, Rapaport, and Todinca [23] and Fraigniaud and Olivetti [25] to a complexity of $O(1/\epsilon)$ rounds. These, along with the work of Fraigniaud, Rapaport, Salo, and Todinca [26] also show fast testing for larger cliques, cycles, and additional subgraphs.

Subgraph Finding in Dynamic Networks. Subgraph finding has also been studied from the perspective of dynamic networks. Bonne and Censor-Hillel [5] characterize the bandwidth that is required for various clique finding problems by algorithms that work in a dynamic setting and must produce the correct answer immediately at the end of the round in which a topology change occurs.

Subgraph finding in a harsh dynamic setting in which the number of topology changes per round is unlimited, was studied by Censor-Hillel, Kolobov, and Schwartzman [13], who showed upper and lower bounds for small subgraphs.

Subgraph Finding in Quantum Networks. Izumi, Le Gall, and Magniez [37] have shown that in a quantum CONGEST model, triangle detection can be solved within $\tilde{O}(n^{1/4})$ rounds, using a distributed version developed by Izumi and Le Gall in [35] of a *Grover Search* [32]. We note that the listing variant for triangles cannot be improved using quantum tools, since it is obtained through information theoretic arguments, which apply to this setting as well.

References

- 1 Amir Abboud, Keren Censor-Hillel, Seri Khoury, and Christoph Lenzen. Fooling views: a new lower bound technique for distributed computations under congestion. *Distributed Computing*, 33:545–559, 2020. doi:10.1007/s00446-020-00373-4.
- 2 Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA)*, 2021.
- 3 Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *J. ACM*, 42(4):844–856, 1995. doi:10.1145/210332.210337.
- 4 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. doi:10.1016/j.jcss.2003.11.006.
- 5 Matthias Bonne and Keren Censor-Hillel. Distributed detection of cliques in dynamic networks. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming, (ICALP)*, pages 132:1–132:15, 2019. doi:10.4230/LIPIcs.ICALP.2019.132.
- 6 Zvika Brakerski and Boaz Patt-Shamir. Distributed discovery of large near-cliques. *Distributed Comput.*, 24(2):79–89, 2011. doi:10.1007/s00446-011-0132-x.
- 7 Keren Censor-Hillel, Yi-Jun Chang, François Le Gall, and Dean Leitersdorf. Tight distributed listing of cliques. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2021.
- 8 Keren Censor-Hillel, Michal Dory, Janne H. Korhonen, and Dean Leitersdorf. Fast approximate shortest paths in the congested clique. *Distributed Computing*, 2020. doi:10.1007/s00446-020-00380-5.
- 9 Keren Censor-Hillel, Eldar Fischer, Gregory Schwartzman, and Yadu Vasudev. Fast distributed algorithms for testing graph properties. *Distributed Computing*, 32(1):41–57, 2019. doi:10.1007/s00446-018-0324-8.
- 10 Keren Censor-Hillel, Orr Fischer, Tzlil Gonen, François Le Gall, Dean Leitersdorf, and Rotem Oshman. Fast distributed algorithms for girth, cycles and small subgraphs. In *Proceedings of the 34th International Symposium on Distributed Computing (DISC)*, pages 33:1–33:17, 2020. doi:10.4230/LIPIcs.DISC.2020.33.
- 11 Keren Censor-Hillel, François Le Gall, and Dean Leitersdorf. On distributed listing of cliques. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 474–482, 2020. doi:10.1145/3382734.3405742.
- 12 Keren Censor-Hillel, Petteri Kaski, Janne H. Korhonen, Christoph Lenzen, Ami Paz, and Jukka Suomela. Algebraic methods in the congested clique. *Distributed Computing*, 32(6):461–478, 2019. doi:10.1007/s00446-016-0270-2.
- 13 Keren Censor-Hillel, Victor I. Kolobov, and Gregory Schwartzman. Finding subgraphs in highly dynamic networks. *CoRR*, abs/2009.08208, 2020, To appear in SPAA. arXiv:2009.08208.
- 14 Keren Censor-Hillel, Dean Leitersdorf, and Elia Turner. Sparse matrix multiplication and triangle listing in the congested clique model. *Theor. Comput. Sci.*, 809:45–60, 2020. doi:10.1016/j.tcs.2019.11.006.
- 15 Yi-Jun Chang, Seth Pettie, and Hengjie Zhang. Distributed triangle detection via expander decomposition. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 821–840, 2019. doi:10.5555/3310435.3310486.
- 16 Yi-Jun Chang and Thatchaphol Saranurak. Improved distributed expander decomposition and nearly optimal triangle enumeration. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 66–73, 2019. doi:10.1145/3293611.3331618.
- 17 Yi-Jun Chang and Thatchaphol Saranurak. Deterministic distributed expander decomposition and routing with applications in distributed derandomization. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 377–388, 2020. doi:10.1109/FOCS46700.2020.00043.

- 18 Artur Czumaj and Christian Konrad. Detecting cliques in CONGEST networks. In *Proceedings of the 32nd International Symposium on Distributed Computing (DISC)*, pages 16:1–16:15, 2018. doi:10.4230/LIPIcs.DISC.2018.16.
- 19 Danny Dolev, Christoph Lenzen, and Shir Peled. "tri, tri again": Finding triangles and small subgraphs in a distributed setting - (extended abstract). In *Proceedings of the 26th International Symposium on Distributed Computing (DISC)*, pages 195–209, 2012. doi:10.1007/978-3-642-33651-5_14.
- 20 Andrew Drucker, Fabian Kuhn, and Rotem Oshman. On the power of the congested clique model. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 367–376, 2014. doi:10.1145/2611462.2611493.
- 21 Talya Eden, Nimrod Fiat, Orr Fischer, Fabian Kuhn, and Rotem Oshman. Sublinear-time distributed algorithms for detecting small cliques and even cycles. In *Proceedings of the 33rd International Symposium on Distributed Computing (DISC)*, pages 15:1–15:16, 2019. doi:10.4230/LIPIcs.DISC.2019.15.
- 22 Paul Erdős. On sequences of integers no one of which divides the product of two others and on some related problems. *Inst. Math. Mech. Univ. Tomsk*, 2:74–82, 1938.
- 23 Guy Even, Orr Fischer, Pierre Fraigniaud, Tzlil Gonen, Reut Levi, Moti Medina, Pedro Montealegre, Dennis Olivetti, Rotem Oshman, Ivan Rapaport, and Ioan Todinca. Three notes on distributed property testing. In *Proceedings of the 31st International Symposium on Distributed Computing (DISC)*, pages 15:1–15:30, 2017. doi:10.4230/LIPIcs.DISC.2017.15.
- 24 Orr Fischer, Tzlil Gonen, Fabian Kuhn, and Rotem Oshman. Possibilities and impossibilities for distributed subgraph detection. In *Proceedings of the 30th Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 153–162, 2018. doi:10.1145/3210377.3210401.
- 25 Pierre Fraigniaud and Dennis Olivetti. Distributed detection of cycles. In Christian Scheideler and Mohammad Taghi Hajiaghayi, editors, *Proceedings of the 29th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA 2017, Washington DC, USA, July 24-26, 2017*, pages 153–162. ACM, 2017. doi:10.1145/3087556.3087571.
- 26 Pierre Fraigniaud, Ivan Rapaport, Ville Salo, and Ioan Todinca. Distributed testing of excluded subgraphs. In Cyril Gavoille and David Ilcinkas, editors, *Distributed Computing - 30th International Symposium, DISC 2016, Paris, France, September 27-29, 2016. Proceedings*, volume 9888 of *Lecture Notes in Computer Science*, pages 342–356. Springer, 2016. doi:10.1007/978-3-662-53426-7_25.
- 27 François Le Gall. Further algebraic algorithms in the congested clique model and applications to graph-theoretic problems. In *Proceedings of the 30th International Symposium on Distributed Computing (DISC)*, pages 57–70, 2016. doi:10.1007/978-3-662-53426-7_5.
- 28 Mohsen Ghaffari, Fabian Kuhn, and Hsin-Hao Su. Distributed MST and routing in almost mixing time. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 131–140, 2017. doi:10.1145/3087801.3087827.
- 29 Mohsen Ghaffari and Jason Li. New distributed algorithms in almost mixing time via transformations from parallel algorithms. In *Proceedings of the 32nd International Symposium on Distributed Computing (DISC)*, pages 31:1–31:16, 2018. doi:10.4230/LIPIcs.DISC.2018.31.
- 30 Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998. doi:10.1145/285055.285060.
- 31 Tzlil Gonen and Rotem Oshman. Lower bounds for subgraph detection in the CONGEST model. In *Proceedings of the 21st International Conference on Principles of Distributed Systems (OPODIS)*, pages 6:1–6:16, 2017. doi:10.4230/LIPIcs.OPODIS.2017.6.
- 32 Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing (STOC 1996)*, pages 212–219, 1996. doi:10.1145/237814.237866.

- 33 Juho Hirvonen, Joel Rybicki, Stefan Schmid, and Jukka Suomela. Large cuts with local algorithms on triangle-free graphs. *Electron. J. Comb.*, 24(4):P4.21, 2017. URL: <http://www.combinatorics.org/ojs/index.php/eljc/article/view/v24i4p21>.
- 34 Dawei Huang, Seth Pettie, Yixiang Zhang, and Zhijun Zhang. The communication complexity of set intersection and multiple equality testing. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1715–1732, 2020. doi:10.1137/1.9781611975994.105.
- 35 Taisuke Izumi and François Le Gall. Quantum distributed algorithm for the all-pairs shortest path problem in the CONGEST-CLIQUE model. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC 2019)*, pages 84–93, 2019. doi:10.1145/3293611.3331628.
- 36 Taisuke Izumi and François Le Gall. Triangle finding and listing in CONGEST networks. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 381–389, 2017. doi:10.1145/3087801.3087811.
- 37 Taisuke Izumi, François Le Gall, and Frédéric Magniez. Quantum distributed algorithm for triangle finding in the CONGEST model. In *Proceedings of the 37th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 23:1–23:13, 2020. doi:10.4230/LIPIcs.STACS.2019.49.
- 38 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discret. Math.*, 5(4):545–557, 1992. doi:10.1137/0405044.
- 39 Hartmut Klauck, Danupon Nanongkai, Gopal Pandurangan, and Peter Robinson. Distributed computation of large-scale graph problems. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 391–410, 2015. doi:10.1137/1.9781611973730.28.
- 40 Janne H. Korhonen and Joel Rybicki. Deterministic subgraph detection in broadcast CONGEST. In James Aspnes, Alysson Bessani, Pascal Felber, and João Leitão, editors, *21st International Conference on Principles of Distributed Systems, OPODIS 2017, Lisbon, Portugal, December 18-20, 2017*, volume 95 of *LIPIcs*, pages 4:1–4:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.OPODIS.2017.4.
- 41 Christoph Lenzen. Optimal deterministic routing and sorting on the congested clique. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 42–50, 2013. doi:10.1145/2484239.2501983.
- 42 Nathan Linial. Locality in distributed graph algorithms. *SIAM J. Comput.*, 21(1):193–201, 1992. doi:10.1137/0221015.
- 43 Zvi Lotker, Boaz Patt-Shamir, Elan Pavlov, and David Peleg. Minimum-weight spanning tree construction in $O(\log \log n)$ communication rounds. *SIAM J. Comput.*, 35(1):120–131, 2005. doi:10.1137/S0097539704441848.
- 44 Gopal Pandurangan, Peter Robinson, and Michele Scquizzato. On the distributed complexity of large-scale graph computations. In *Proceedings of the 30th on Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 405–414, 2018.
- 45 David Peleg. *Distributed Computing: A Locality-Sensitive Approach*. Society for Industrial and Applied Mathematics, USA, 2000.
- 46 Seth Pettie and Hsin-Hao Su. Fast distributed coloring algorithms for triangle-free graphs. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II*, volume 7966 of *Lecture Notes in Computer Science*, pages 681–693. Springer, 2013. doi:10.1007/978-3-642-39212-2_59.
- 47 Alexander A. Razborov. On the distributional complexity of disjointness. In Mike Paterson, editor, *Automata, Languages and Programming, 17th International Colloquium, ICALP90, Warwick University, England, UK, July 16-20, 1990, Proceedings*, volume 443 of *Lecture Notes in Computer Science*, pages 249–253. Springer, 1990. doi:10.1007/BFb0032036.
- 48 Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48:436–452, 1941.