

# Brief Announcement: Does Preprocessing Help under Congestion?

Klaus-Tycho Foerster\*    Janne H. Korhonen†    Joel Rybicki†    Stefan Schmid\*

\*Faculty of Computer Science, University of Vienna, Austria

†Institute of Science and Technology Austria

## Abstract

This paper investigates the power of preprocessing in the CONGEST model. Schmid and Suomela (ACM HotSDN 2013) introduced the SUPPORTED CONGEST model to study the application of distributed algorithms in Software-Defined Networks (SDNs). In this paper, we show that a large class of lower bounds in the CONGEST model still hold in the SUPPORTED model, highlighting the robustness of these bounds. This also raises the question how much does preprocessing help in the CONGEST model.

## 1 Introduction

Common models of distributed computation typically consider scenarios where the computation always starts from *scratch*, i.e., in an *unknown* communication topology. However, in many practical scenarios, the communication topology does not change as frequently as the problem input. For example, the distributed algorithm may always be run in networks whose topology is known in advance, but the input instance may vary. In such cases, it is natural to *support* distributed algorithms by allowing preprocessing of the underlying network topology [11].

With this in mind, Schmid and Suomela [11] proposed two SUPPORTED models of distributed computation to enhance distributed algorithms with the power of preprocessing: the SUPPORTED LOCAL and SUPPORTED CONGEST. Subsequently, Korhonen and Rybicki considered subgraph detection problems [8] in the SUPPORTED CONGEST model, whereas Foerster et al. [6] investigated the power of the SUPPORTED LOCAL model. In this paper, we focus on SUPPORTED CONGEST.

**Contribution.** We observe that many lower bounds in the standard CONGEST model still hold under such preprocessing. Given that intuitively preprocessing seems to be very powerful, this may come as a surprise. This raises the question of how much preprocessing actually helps in the CONGEST model. Indeed, it may be either that the power of preprocessing is very limited or that the current lower bounds for non-supported CONGEST are not tight. In the light of this, we propose the following challenge: *is there a separation between supported and non-supported models?* If the answer is no, then there may be a way to easily simulate preprocessing, and thus, simplify algorithm design in the CONGEST model. In the converse case, preprocessing may offer a practical way to accelerate current distributed algorithms.

**Model.** In the SUPPORTED CONGEST model, the communication topology is an undirected graph  $H = (V, E)$  and each node has a unique identifier of size  $O(\log n)$  bits. The logical state is given by an undirected subgraph, i.e., the *input* graph  $G \subseteq H$ , which inherits the identifiers in  $H$ . The computation proceeds in two steps: First, in the preprocessing phase, the nodes may compute

any function on  $f(H)$  and store the result locally. In the second phase, the nodes are tasked to solve a problem instance on the input graph  $G$  in the CONGEST model. To this end, the edges of  $H$  may be used for communication and additionally the local outputs of the preprocessing. Note that the congested clique [9] model is a special case of the SUPPORTED model: the support  $H$  is simply a clique. In addition, one may also restrict the communication of the SUPPORTED model to just the input graph  $G$  after preprocessing; this model is called the *passive* SUPPORTED model.

## 2 Lower Bounds for the SUPPORTED CONGEST Model

We show that CONGEST lower bounds obtained using the now standard *family of lower bound graphs* construction [2] easily translate to the SUPPORTED model. To this end, we adapt here the proof of Abboud et al. [2]. Using existing constructions for families of lower bound graphs then immediately give the lower bounds shown in Table 1. We note that this technique does not directly cover the lower bounds of Das Sarma et al. [10], though we believe they can be similarly translated.

**Two-party communication complexity.** Let  $f: \{0, 1\}^{2k} \rightarrow \{0, 1\}$  be a Boolean function. In the two-party communication game on  $f$ , there are two players who receive a private  $k$ -bit string  $x_0$  and  $x_1$  as input, and the task is to have at least one of the players compute  $f(x) = f(x_0, x_1)$ . The *deterministic communication complexity*  $\text{CC}(f)$  of a function  $f$  is the maximum number of bits the two players need to exchange in the worst case (over all deterministic protocols and input strings) in order to compute  $f(x_0, x_1)$ . Similarly, the *randomised communication complexity*  $\text{RCC}(f)$  is the worst-case complexity of protocols, which compute  $f$  with probability at least  $2/3$  on all inputs.

**Definition 1.** Let  $f_n: \{0, 1\}^{2k(n)} \rightarrow \{0, 1\}$  and  $C: \mathbb{N} \rightarrow \mathbb{N}$  be functions and  $\Pi$  a graph predicate. Suppose that there exists a constant  $n_0$  such that for all  $n > n_0$  and  $x_0, x_1 \in \{0, 1\}^{k(n)}$  there exists a (weighted) graph  $G(n, x_0, x_1)$  satisfying the following properties:

1.  $G(n, x_0, x_1)$  satisfies  $\Pi$  if and only if  $f_n(x_0, x_1) = 1$ ,
2.  $G(n, x_0, x_1) = (V_0 \cup V_1, E_0 \cup E_1 \cup S)$ , where
  - $V_0$  and  $V_1$  are disjoint and  $|V_0 \cup V_1| = n$ ,
  - $E_i \subseteq V_i \times V_i$  for  $i \in \{0, 1\}$ ,
  - $S \subseteq V_0 \times V_1$  is a cut and has size at least  $C(n)$ , and
  - the (weighted) subgraph  $G_i = (V_i, E_i)$  only depends on  $i$ ,  $n$  and  $x_i$ , i.e.,  $G_i = G_i(n, x_i)$ .

If  $\mathcal{G}(n) = \{G(n, x) : x \in \{0, 1\}^{2k(n)}\}$ , then  $\mathcal{F} = (\mathcal{G}(n))_{n > n_0}$  is a *family of lower bound graphs*.

**Theorem 2.** Let  $\mathcal{F}$  be a family of lower bound graphs. Any algorithm deciding  $\Pi$  on a graph family  $\mathcal{H}$  containing  $\bigcup \mathcal{G}(n)$  for all  $n > n_0$  in the *passive* or *active* SUPPORTED CONGEST model with bandwidth  $b(n)$  needs

$$\Omega\left(\frac{\text{CC}(f_n)}{C(n)b(n)}\right) \quad \text{and} \quad \Omega\left(\frac{\text{RCC}(f_n)}{C(n)b(n)}\right)$$

*deterministic and randomised rounds, respectively.*

*Proof.* Suppose  $\mathcal{A}$  is an algorithm that decides  $\Pi$  on the graph family  $\mathcal{H}$  in  $T(n)$  communication rounds. We now construct a two-player protocol  $\pi$  that computes  $f_n(x_0, x_1)$  by simulating  $\mathcal{A}$ . Let  $x_0, x_1 \in \{0, 1\}^{k(n)}$  be the input and  $G = G(n, x_0, x_1)$  and  $H = \bigcup \mathcal{G}(n)$ .

Given its input  $x_i$ , player  $i$  can locally construct the subgraph  $G_i(n, x_i) \subseteq G(n, x_0, x_1)$ . Note that given  $G_i(x_i)$ , the support graph  $H$  does not reveal any information about  $E_{1-i}$  or  $x_{1-i}$  to player  $i$ , since for any  $y \in \{0, 1\}^{k(n)}$  we have  $G_{1-i}(y) \subseteq H[V_{1-i}]$ . Simulating any messages sent between vertices of  $G_i(n, x)$  can be done without any communication with player  $1-i$ . Any messages

from  $V_i$  to  $V_{1-i}$  must go across the cut  $S$  and are communicated by player  $i$  to player  $1 - i$ . As in each round each player communicates at most  $b(n)$  bits over any edge in  $S$ , the total amount of bits communicated during the course of  $T$  rounds is at most  $2b(n)|S(x)|T(n) \geq 2b(n)C(n)T(n)$ , which must be at least  $\text{CC}(f_n)$  for deterministic algorithms and  $\text{RCC}(f_n)$  for randomised algorithms. Thus, the claim follows by observing that  $\frac{\text{CC}(f_n)}{2b(n)C(n)} \leq T(n)$  and  $\frac{\text{RCC}(f_n)}{2b(n)C(n)} \leq T(n)$ .  $\square$

Lower bound	Problem
$\Omega(n^{1/2}/\log n)$	4-cycle [4], $2k$ -cycle [8], Girth $((2 - \epsilon)$ -apx.) [7]
$\Omega(n/\log n)$	$(2k + 1)$ -cycle [4], APSP, Diameter $((3/2 - \epsilon)$ -apx.) [7]
$\Omega(n/(\log n)^2)$	Diameter on sparse graphs [1]
$\Omega(n/(\log n)^3)$	Diameter and radius $((3/2 - \epsilon)$ -apx.), eccentricities $((5/3 - \epsilon)$ -apx.), all on sparse graphs [1]
$\Omega(n^{2-1/k}/(k \log n))$	Subgraph detection (for any $k$ ) [5]
$\Omega(n^2/(\log n)^2)$	Min. vertex cover, max. independent set, chrom. number $((4/3 - \epsilon)$ -apx.), weighted 8-cycle [3]
$\Omega(n^2)$	Identical subgraphs (det. only) [3]

Table 1: Lower bounds that transfer from the CONGEST to the SUPPORTED CONGEST model.

### 3 Towards New Algorithmic Opportunities?

We saw in the last section that many lower bounds from the CONGEST model translate directly to the SUPPORTED CONGEST model, even though intuitively, the SUPPORTED model may seem significantly more powerful. This raises the question if the SUPPORTED model is actually a stronger model in a meaningful sense or if the prior lower bounds were so strong that they easily transferred.

**First separation results.** Prior work on the SUPPORTED LOCAL model [6] already pointed out that computing an upper bound on the network size separates the LOCAL ( $\Omega(D)$  rounds) and the SUPPORTED LOCAL model (0 rounds). Analogous results hold if the support graph is promised to have certain (monotone) properties that apply to all its subgraphs, e.g., being  $k$ -colorable. These results directly carry over to the SUPPORTED CONGEST model, providing a 0 vs  $\Omega(D)$  round separation, even in an identifier-independent setting. On the other hand, in the SUPPORTED LOCAL model, all problems can be solved trivially in diameter time, but can the SUPPORTED CONGEST go further? Observe that the naïve problem of collecting all identifiers does not provide an  $\Omega(n^2)$  separation in the SUPPORTED CONGEST model, as the problem may only depend on the input graph, which may omit nodes present in the support graph. Notwithstanding, we can alter the problem s.t. each node has to e.g. output two sets  $I_0, I_1$  of identifiers with  $|I_0| = |I_1|$ , where  $I_1$  contains a superset of all identifiers in  $G$  and none of the identifiers in  $I_0$  appear in  $G$ .

**Open questions and possibilities.** While the separation results for restricted graph classes can be directly used to accelerate many specialized algorithms (e.g., coloring when the support graph has a small chromatic number), we leave it as an open question how the SUPPORTED model can be leveraged outside the case of collecting identifiers and providing upper bounds on the graph size, even though the latter is sometimes needed as an input for some algorithms. We believe that exciting possibilities arise, no matter the outcome to this open question. For example, if the CONGEST model could simulate the SUPPORTED model with negligible overhead beyond the previously mentioned exceptions, the SUPPORTED model could greatly simplify algorithm

design by incorporating preprocessing. On the other hand, even a strong separation could lead to significantly faster algorithms in neighboring research areas, e.g. for Software Defined Networks [11].

## References

- [1] Amir Abboud, Keren Censor-Hillel, and Seri Khoury. [Near-linear lower bounds for distributed distance computations, even in sparse networks](#). In *Proc. 30th International Symposium on Distributed Computing (DISC 2016)*, pages 29–42. 2016.
- [2] Amir Abboud, Keren Censor-Hillel, Seri Khoury, and Ami Paz. [Smaller cuts, higher lower bounds](#), 2019. arxiv:1901.01630 [cs.DC].
- [3] Keren Censor-Hillel, Seri Khoury, and Ami Paz. [Quadratic and near-quadratic lower bounds for the CONGEST model](#). In *Proc. 31st International Symposium on Distributed Computing (DISC 2017)*. 2017.
- [4] Andrew Drucker, Fabian Kuhn, and Rotem Oshman. [On the power of the congested clique model](#). In *Proc. Symposium on Principles of Distributed Computing (PODC 2014)*, pages 367–376. 2014.
- [5] Orr Fischer, Tzlil Gonen, Fabian Kuhn, and Rotem Oshman. [Possibilities and impossibilities for distributed subgraph detection](#). In *Proc. 30th Symposium on Parallelism in Algorithms and Architectures (SPAA 2018)*, pages 153–162. 2018.
- [6] Klaus-Tycho Foerster, Juho Hirvonen, Stefan Schmid, and Jukka Suomela. [On the power of preprocessing in decentralized network optimization](#). In *Proc. 39th IEEE International Conference on Computer Communications (INFOCOM 2019)*. 2019.
- [7] Silvio Frischknecht, Stephan Holzer, and Roger Wattenhofer. [Networks cannot compute their diameter in sublinear time](#). In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms (SODA 2012)*, pages 1150–1162. 2012.
- [8] Janne H. Korhonen and Joel Rybicki. [Deterministic subgraph detection in broadcast CONGEST](#). In *Proc. 21st International Conference on Principles of Distributed Systems (OPODIS 2017)*. 2017.
- [9] Zvi Lotker, Elan Pavlov, Boaz Patt-Shamir, and David Peleg. [MST construction in  \$O\(\log \log n\)\$  communication rounds](#). In *Proc. 15th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2003)*, pages 94–100. 2003.
- [10] Atish Das Sarma, Stephan Holzer, Liah Kor, Amos Korman, Danupon Nanongkai, Gopal Pandurangan, David Peleg, and Roger Wattenhofer. [Distributed verification and hardness of distributed approximation](#). *SIAM Journal on Computing*, 41 (5): pages 1235–1265, 2012.
- [11] Stefan Schmid and Jukka Suomela. [Exploiting locality in distributed SDN control](#). In *Proc. 2nd ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN 2013)*, pages 121–126. 2013.