

Reachability and Shortest Paths in the Broadcast CONGEST Model

Shiri Chechik

Tel Aviv University, Israel

shiri.chechik@gmail.com

Doron Mukhtar

Tel Aviv University, Israel

doron.muk@gmail.com

Abstract

In this paper we study the time complexity of the single-source reachability problem and the single-source shortest path problem for directed unweighted graphs in the Broadcast CONGEST model. We focus on the case where the diameter D of the underlying network is constant.

We show that for the case where $D = 1$ there is, quite surprisingly, a very simple algorithm that solves the reachability problem in $1(!)$ round. In contrast, for networks with $D = 2$, we show that any distributed algorithm (possibly randomized) for this problem requires $\Omega(\sqrt{n}/\log n)$ rounds. Our results therefore completely resolve (up to a small polylog factor) the complexity of the single-source reachability problem for a wide range of diameters.

Furthermore, we show that when $D = 1$, it is even possible to get an almost 3-approximation for the all-pairs shortest path problem (for directed unweighted graphs) in just 2 rounds. We also prove a stronger lower bound of $\Omega(\sqrt{n})$ for the single-source shortest path problem for unweighted directed graphs that holds even when the diameter of the underlying network is 2. As far as we know this is the first lower bound that achieves $\Omega(\sqrt{n})$ for this problem.

2012 ACM Subject Classification Theory of computation → Distributed algorithms

Keywords and phrases Distributed algorithms, Broadcast CONGEST, distance estimation, small diameter

Digital Object Identifier 10.4230/LIPIcs.DISC.2019.11

1 Introduction

Reachability and shortest path are two of the most fundamental problems in graph algorithms. In this paper, we study the single-source reachability (SSR) problem and the single-source shortest path (SSSP) problem in the Broadcast CONGEST model of distributed computing.

The CONGEST model [17] is one of the most studied message-passing models in the field of distributed computing. In this model, a synchronized n -vertex communication network is modeled by an undirected graph N whose vertices correspond to the processors in this network and whose edges correspond to the communication links between them. Each vertex has a unique $O(\log n)$ -bit identifier initially known only to itself and its neighbors in N . The vertices communicate in discrete rounds, where in each round each vertex receives the messages that were previously sent to it, performs some unbounded local computation and then sends messages of $O(\log n)$ bits to all or some of its neighbors. The vertices work together on some common task (such as computing distances in the network) and the complexity is measured by the number of communication rounds needed to complete this task. The Broadcast CONGEST model is a more restrictive variant of the CONGEST model where every vertex has to send (broadcast) the same message to all of its neighbors in each round.

In this paper we focus on directed and unweighted graphs. In the SSR problem, we are asked to identify all the vertices in a given graph G for which there is a directed path from some designated vertex s called the source. In the SSSP problem, we are further asked to

11:2 Reachability and Shortest Paths in the Broadcast CONGEST Model

compute for each such vertex its distance (the number of edges in a shortest path) from the source s . In the CONGEST model as well as in other similar message-passing models, we assume that the communication network N is identical to the underlying graph of G (where G is the input graph for the SSR\SSSP problem). We also assume that the communication between the vertices is bi-directional (regardless of directions of the edges in G). Initially, each vertex in the network knows whether it is the source or not, and it also knows its set of incoming and outgoing edges in G . In the distributed SSR problem, each vertex has to determine whether it is reachable from the source or not, and in the distributed SSSP problem, each vertex has to determine its distance from the source.

Related Work

Distance computation problems (such as the SSSP problem) have been widely studied in many models of distributed computing. It is not hard to see that in many synchronous message-passing models, problems such as SSR and SSSP require $\Omega(D)$ rounds (where D is the diameter of the underlying network). While this lower bound can be easily matched when messages of unbounded size are allowed, the situation for models that require the messages to be of bounded size is far more involved.

In the CONGEST model, it is possible to solve the directed single-source reachability problem in $\tilde{O}(\sqrt{n}D^{1/4} + D)$ rounds with high probability [10]. Many variants of the SSSP problem (directed\undirected, exact\approximate etc.) were studied over the years (see, e.g., [16, 11, 2, 7, 9, 8]). In particular, for directed and weighted graphs, there is a randomized algorithm that solves the SSSP problem in $\tilde{O}(\sqrt{nD})$ rounds [8]. We note that many of the above mentioned algorithms (such as [10, 8]) actually work in the more restrictive Broadcast CONGEST model. Regarding lower bounds, Das Sarma et al. [6] showed that in the CONGEST model the time complexity of any (possibly randomized) algorithm for the directed single-source reachability problem is $\Omega(\sqrt{n}/\log n)$. However, this lower bound was shown only for graphs of underlying diameter $\Omega(n^\delta)$ for some $0 < \delta < 1/2$. For smaller diameters, similar but weaker lower bounds were shown e.g. $\Omega(\sqrt{n}/\log n)$ for graphs of underlying diameter $\Theta(\log n)$. The smallest constant diameter for which a non-trivial lower bound is known is 3 where it was shown to require $\Omega((n/\log n)^{1/4})$ rounds.

For the related all-pairs shortest path (APSP) problem, many algorithms with near-optimal complexities for the approximate version of this problem and for the case of unweighted graphs were developed over the years (e.g., [12, 14, 15, 16]). Recently, many algorithms with improved complexities for the case of weighted graphs were devised [7, 13, 1] culminating with the $\tilde{O}(n)$ -time randomized algorithm of [3].

Our Results

In this paper we study the time complexity of the SSR problem and the SSSP problem (for directed unweighted graphs) in the Broadcast CONGEST model for networks of constant diameter. Specifically, we show that even for networks of diameter 2, any distributed algorithm (possibly randomized) for the SSR problem requires $\Omega(\sqrt{n}/\log n)$ rounds. In contrast, we show that quite surprisingly for networks of diameter 1, this problem (or even the more general all-pairs reachability problem) can be solved deterministically in 1 round. Moreover, we show that for networks of diameter 1 one can compute in 2 rounds a $(3, 2)$ -approximation for the APSP problem (for directed unweighted graphs), where by (α, β) -approximation we mean α multiplicative approximation and β additive approximation.

The algorithm for the approximate APSP problem (resp. for the all-pairs reachability problem) allows each vertex to compute a $(3, 2)$ -approximation for the distance between every pair of vertices in the graph (resp. determine reachability for every such pair). We note that if one can compute a $(2 - \epsilon)$ -approximation for the APSP problem (for some $1 \geq \epsilon > 0$) such that there is some vertex v that knows the computed estimation for every pair of vertices, then this vertex can recover the whole graph. This means that v must receive in this case $\Theta(n^2)$ bits of information from its neighbors (simply because there are $\Theta(2^{n^2})$ possible graphs on these vertices), but in each round, v can get at most $O(n \log n)$ bits from its neighbors and so $\Omega(n / \log n)$ rounds are required for solving this problem.

Our results show a large gap between networks of diameter 1 and 2. As upper bounds of $\tilde{O}(\sqrt{n})$ are already known for the SSR problem when the underlying network has constant or poly-logarithmic diameter (e.g., [10, 8]), we completely resolve (up to poly-log factors) the SSR problem when the diameter of the underlying network is constant or even poly-logarithmic. Our algorithms are very simple (we see this as a plus and not a minus). In addition, we show a stronger lower bound of $\Omega(\sqrt{n})$ for the SSSP problem for unweighted directed graphs in the Broadcast CONGEST model that holds even when the diameter of the underlying network is 2. As far as we know this is the first lower bound that achieves $\Omega(\sqrt{n})$ for this problem.

Further Related Work

A closely related model to the CONGEST when the underlying communication network has diameter 1 is the Congested Clique model. The Congested Clique model is a synchronous message-passing model in which the underlying communication network is the complete graph on n vertices but the graph G on which the solution needs to be obtained can be an arbitrary graph on n vertices (that is, each vertex initially knows its neighbors in G and can exchange messages of size $O(\log n)$ with any vertex in the graph even if they are not adjacent in G).

Censor-Hillel et al. [5] adapted parallel matrix multiplication algorithms to this model. Using these algorithms, they obtained better algorithms for subgraph detection and distance computation. In particular, they showed a $\tilde{O}(n^{1/3})$ -round algorithm for solving the APSP problem for weighted directed graphs, and even more efficient algorithms for unweighted undirected graphs or distance approximation. Recently, it was shown [4] that the SSSP problem for weighted undirected graphs can be solved in $\tilde{O}(n^{1/6})$ rounds.

We note that for problems such as SSSP or APSP (for weighted graphs) the Congested Clique model is actually a special case of the CONGEST model when the diameter of the underlying network is 1. To see this, note that one can always transform the input graph G into a complete graph by adding edges of very large weight. Therefore, either one can show a constant upper bound for the weighted SSSP problem in the Congested Clique model which will be quite a breakthrough or our upper bound shows a separation between the SSR problem and the SSSP problem for directed weighted graphs of underlying diameter 1 (and even between the all-pairs reachability problem and the SSSP problem).

2 Preliminaries

In the following, we assume that all directed graphs are simple (i.e., they do not contain self-loops or multiple edges, but they may contain anti-parallel edges). For a graph H , we respectively denote by $V(H)$ and $E(H)$ its vertex set and edge set. The out-degree and in-degree of a vertex v in a directed graph H are denoted by $d_{out}(v)$ and $d_{in}(v)$, respectively.

For a directed graph H and a vertex v in H , we denote by $N_{out}(v)$ its set of outgoing neighbors, and by $N_{in}(v)$ its set of ingoing neighbors. Given a directed graph $H = (V, E)$ and a set $A \subseteq V$, we denote by A^c the set $V \setminus A$. The underlying diameter of a directed graph H is defined to be the diameter of its underlying graph. For a graph H and two vertices v and u in $V(H)$, we denote by $d(v, u, H)$ the distance from v to u in H . All logarithms in this paper are of base 2.

The rest of paper is organized as follows. In section 3, we show an algorithm that solves the all-pairs reachability problem in one round for networks of diameter 1. In section 4, we show that in two rounds one can compute an approximation for the APSP problem (also for networks of diameter 1). In section 5, we prove lower bounds for computing reachability and distances in networks of diameter 2.

3 All-Pairs Reachability for Networks of Diameter 1

In this section we show that when the diameter of the underlying network is 1, the directed single-source reachability problem can be solved in $O(1)$ rounds in the Broadcast CONGEST model. In fact, we show that it can be solved in a single round. Furthermore, our algorithm can solve the much more general problem of all-pairs reachability (again in a single round). The algorithm is extremely simple. Every vertex simply sends its in-degree and out-degree to all its neighbors in the underlying network, and then, by using this information only, each vertex can determine (by a simple computation) which vertex is reachable from which. This requires messages of at most $2\lceil \log_2 n \rceil$ bits (moreover, if there are no anti-parallel edges then, as the underlying diameter is 1, the in-degree plus out-degree of every vertex is exactly $n - 1$ and therefore it is enough to send only the in-degree and so $\lceil \log_2 n \rceil$ bits are enough).

The next lemma shows that when the underlying diameter of some directed graph H is 1, we can determine if $E(H) \cap (A \times A^c) = \emptyset$ by using the in and out degrees of the vertices in A , for every subset of vertices $A \subseteq V(H)$.

► **Lemma 3.1.** *For every directed graph $H = (V, E)$ with underlying diameter 1 and every set $A \subseteq V$, we have $\sum_{v \in A} (d_{in}(v) - d_{out}(v)) = |A^c \times A|$ if and only if $E \cap (A \times A^c) = \emptyset$.*

Proof. Let $H = (V, E)$ be a directed graph with underlying diameter 1 and let A be some subset of V . We have $\sum_{v \in A} d_{out}(v) = |E \cap (A \times V)| = |E \cap (A \times A)| + |E \cap (A \times A^c)|$ and similarly $\sum_{v \in A} d_{in}(v) = |E \cap (V \times A)| = |E \cap (A \times A)| + |E \cap (A^c \times A)|$. It follows that

$$\sum_{v \in A} (d_{in}(v) - d_{out}(v)) = |E \cap (A^c \times A)| - |E \cap (A \times A^c)| \quad (1)$$

Now, for showing the first direction, assume that $\sum_{v \in A} (d_{in}(v) - d_{out}(v)) = |A^c \times A|$. By equation (1), we have $|A^c \times A| = |E \cap (A^c \times A)| - |E \cap (A \times A^c)|$. As $|E \cap (A^c \times A)| \leq |A^c \times A|$, we get that $|E \cap (A \times A^c)| \leq 0$ and so $E \cap (A \times A^c) = \emptyset$.

For the second direction, assume that $E \cap (A \times A^c) = \emptyset$. Since in addition H has underlying diameter 1, every vertex in A^c must have an outgoing edge to every vertex in A and so $E \cap (A^c \times A) = A^c \times A$. It follows, by equation (1), that $\sum_{v \in A} (d_{in}(v) - d_{out}(v)) = |A^c \times A|$. ◀

The next lemma shows that when the underlying diameter of some directed graph H is 1, the in and out degrees of all the vertices in H are enough to determine which vertices are reachable from any given vertex in H .

► **Lemma 3.2.** *For every directed graph H with underlying diameter 1, every ordering (v_1, \dots, v_n) of its vertices such that $d_{out}(v_1) \leq \dots \leq d_{out}(v_n)$ and every $i \in \{1, \dots, n\}$, there exists an index $k \in \{i, \dots, n\}$ such that the set of reachable vertices from v_i in H is equal to $\{v_1, \dots, v_k\}$. Moreover, k is the minimal index in $\{i, \dots, n\}$ for which $(n - k)k = \sum_{j=1}^k (d_{in}(v_j) - d_{out}(v_j))$.*

Proof. Let $H = (V, E)$ be a directed graph with underlying diameter 1, let (v_1, \dots, v_n) be an ordering of its vertices such that $d_{out}(v_1) \leq \dots \leq d_{out}(v_n)$ and let $i \in \{1, \dots, n\}$. Let A be the set of all the reachable vertices from v_i in H , and note that we must have $E \cap (A \times A^c) = \emptyset$ (as otherwise v_i can reach a vertex from A^c which is of course contradiction to the definition of A and A^c).

Let k be the highest index in $\{i, \dots, n\}$ for which $v_k \in A$ (such an index must exist as $v_i \in A$). Clearly, we have $A \subseteq \{v_1, \dots, v_k\}$. We claim that we must also have $\{v_1, \dots, v_k\} \subseteq A$. Since $v_k \in A$ and $E \cap (A \times A^c) = \emptyset$, the set A must contain at least $d_{out}(v_k) + 1$ vertices (the vertex v_k and its $d_{out}(v_k)$ outgoing neighbors). It also follows that every vertex in A^c must have out-degree at least $d_{out}(v_k) + 1$. To see this, note that every vertex in A^c must have an outgoing edge to every vertex in A (as $E \cap (A \times A^c) = \emptyset$ and the underlying diameter of H is 1). Therefore, it must be that $v_j \in A$ for all $j \in \{1, \dots, k\}$ as $d_{out}(v_j) \leq d_{out}(v_k)$ for every such j . We conclude that $A = \{v_1, \dots, v_k\}$.

Now, as $E \cap (A \times A^c) = \emptyset$ we get from Lemma 3.1 that $(n - k)k = \sum_{j=1}^k (d_{in}(v_j) - d_{out}(v_j))$. We are left to show that k is the minimal index in $\{i, \dots, n\}$ with this property. Assume towards a contradiction that there exists $m \in \{i, \dots, n\}$ such that $m < k$ and $\sum_{j=1}^m (d_{in}(v_j) - d_{out}(v_j)) = (n - m)m$. Let $B = \{v_1, \dots, v_m\}$. By Lemma 3.1 we get that $E \cap (B \times B^c) = \emptyset$, and, in particular, that v_k is not reachable from v_i (as $v_i \in B$ and $v_k \notin B$) which is a contradiction. ◀

Lemma 3.2 can be easily turned into an algorithm that solves the all-pairs reachability problem in one round (when the diameter of the underlying network is 1) as follows. Each vertex v in the graph starts by broadcasting the values of $d_{in}(v)$ and $d_{out}(v)$. After receiving the messages, v sorts the vertices in non-decreasing order of their out-degree. Let (v_1, \dots, v_n) be that ordering. It then finds for every $i \in \{1, \dots, n\}$ the minimal index $k_i \in \{i, \dots, n\}$ such that $(n - k_i)k_i = \sum_{j=1}^{k_i} (d_{in}(v_j) - d_{out}(v_j))$ and deduces by Lemma 3.2 that the set of reachable vertices from v_i is $\{v_1, \dots, v_{k_i}\}$. We conclude the following:

► **Corollary 3.3.** *In the Broadcast CONGEST model, there is a deterministic algorithm that solves the all-pairs reachability problem in one round when the diameter of the underlying network is 1.*

We also note that the time complexity of the internal computation of each vertex is $O(n^2)$.

4 APSP Approximation for Networks of Diameter 1

In the previous section, we showed that it is possible to solve the all-pairs reachability problem in one round for networks of diameter 1. Here we show that it is actually possible to compute an approximation to the distance between all pairs of vertices in two rounds.

Let $G = (V, E)$ be a directed graph on n vertices and underlying diameter 1. For every non-negative integer $i < n$, we let $A(i)$ be the set of all the vertices $u \in V$ whose out-degree is greater than i and that have some in-going neighbor whose out-degree is at most i , that is, $A(i) = \{u \mid (d_{out}(u) > i) \text{ and } (\exists w \in V \text{ s.t. } (w, u) \in E \text{ and } d_{out}(w) \leq i)\}$. We also set $M(i)$ to be \perp if $A(i) = \emptyset$ and $\max\{d_{out}(v) \mid v \in A(i)\}$ otherwise.

11:6 Reachability and Shortest Paths in the Broadcast CONGEST Model

For each $i \in \{d_{out}(v) \mid v \in V\}$, we set $f_0[i] = i$, and then for each $k \in \{1, \dots, n\}$, we further set $f_k[i]$ to be \perp if $f_{k-1}[i] = \perp$ and to $M(f_{k-1}[i])$ otherwise. We first prove some basic properties.

▷ **Claim 4.1.** For every $v \in V$ and $k \in \{1, \dots, n\}$ such that $f_k[d_{out}(v)] \neq \perp$, we have (i) $f_i[d_{out}(v)] \neq \perp$ for every $i \in \{0, \dots, k\}$, and (ii) $f_i[d_{out}(v)] < f_{i+1}[d_{out}(v)]$ for every $i \in \{0, \dots, k-1\}$.

Proof. The first property follows directly from the definition of the sequence. For the second property, note that if $f_{i+1}[d_{out}(v)] \neq \perp$ holds for some $i \in \{0, \dots, k-1\}$, then $f_{i+1}[d_{out}(v)]$ must be equal to the maximum out-degree of the vertices in $A(f_i[d_{out}(v)])$. As, by definition, $A(f_i[d_{out}(v)])$ contains only vertices whose out-degree is greater than $f_i[d_{out}(v)]$, it must be that $f_{i+1}[d_{out}(v)] > f_i[d_{out}(v)]$. \triangleleft

Note that, in particular, this claim implies that $f_n[d_{out}(v)] = \perp$ for every $v \in V$. As otherwise, we would get that $0 \leq f_0[d_{out}(v)] < f_1[d_{out}(v)] < \dots < f_n[d_{out}(v)]$, and so that $f_n[d_{out}(v)] \geq n$ which is impossible as the maximum possible out-degree is $n-1$.

▷ **Claim 4.2.** For every $k \in \{0, \dots, n-1\}$ and every two vertices x and y in V such that y is reachable from x , if $f_k[d_{out}(x)] \neq \perp$ and $d_{out}(y) > f_k[d_{out}(x)]$ then $f_{k+1}[d_{out}(x)] \neq \perp$.

Proof. Let x and y be two vertices in V such that y is reachable from x , and let $k \in \{0, \dots, n-1\}$ be such that $f_k[d_{out}(x)] \neq \perp$ and $d_{out}(y) > f_k[d_{out}(x)]$. First, note that by definition and Claim 4.1, we must have $d_{out}(x) = f_0[d_{out}(x)] \leq f_k[d_{out}(x)]$. This fact together with the assumption that $d_{out}(y) > f_k[d_{out}(x)]$ implies that G must contain an edge (u, v) such that $d_{out}(u) \leq f_k[d_{out}(x)]$ and $d_{out}(v) > f_k[d_{out}(x)]$. To see this, note that there must be a path π from x to y (as y is reachable from x) and $d_{out}(x) \leq f_k[d_{out}(x)] < d_{out}(y)$. This is possible only if π contains an edge (u, v) such that $f_k[d_{out}(x)] \geq d_{out}(u)$ and $f_k[d_{out}(x)] < d_{out}(v)$. It follows that $M(f_k[d_{out}(x)]) \neq \perp$ and so $f_{k+1}[d_{out}(x)] \neq \perp$. \triangleleft

In the next claims, we show how the defined sequences can be used to estimate the distances between the vertices in the graph.

▷ **Claim 4.3.** For every two vertices x and y in V such that y is reachable from x , there exists an index $k \in \{0, \dots, n-1\}$ such that $f_k[d_{out}(x)] \neq \perp$ and $d_{out}(y) \leq f_k[d_{out}(x)]$. Moreover, if k is the minimal index with this property, then the distance from x to y is at least k .

Proof. Let x and y be two vertices in V such that y is reachable from x , and let π be some shortest path from x to y . Let k' be the maximal index in $\{0, \dots, n\}$ for which $f_{k'}[d_{out}(x)] \neq \perp$ (such an index must exist as $f_0[d_{out}(x)] \neq \perp$ always holds). Note that $k' \leq n-1$ as we must have $f_n[d_{out}(v)] = \perp$. It follows that $d_{out}(y) \leq f_{k'}[d_{out}(x)]$ (as if $d_{out}(y) > f_{k'}[d_{out}(x)]$ then, by Claim 4.2, we must have $f_{k'+1}[d_{out}(x)] \neq \perp$ which is a contradiction to the maximality of k') and so the required index exists.

Now, let k be the minimal index in $\{0, \dots, n-1\}$ with the required property. We claim that π contains at least $k+1$ different vertices. Indeed, if $k=0$ then π clearly contains at least 1 vertex (or more if $x \neq y$). This leaves us with the case in which $k > 0$. Note that for every $t \in \{0, \dots, k-1\}$, we have $d_{out}(x) = f_0[d_{out}(x)] \leq f_t[d_{out}(x)]$ and $d_{out}(y) > f_t[d_{out}(x)]$. This means that π contains an edge (x'_t, x_t) such that $d_{out}(x'_t) \leq f_t[d_{out}(x)]$ and $d_{out}(x_t) > f_t[d_{out}(x)]$, for every $t \in \{0, \dots, k-1\}$. It follows that for every such t , there exists a vertex x_t in π such that $f_{t+1}[d_{out}(x)] \geq d_{out}(x_t) > f_t[d_{out}(x)]$. Note that $d_{out}(x) = f_0[d_{out}(x)] < d_{out}(x_0) < \dots < d_{out}(x_{k-1})$ which implies that π contains at least $k+1$ different vertices. \triangleleft

▷ **Claim 4.4.** For every two vertices x and y in V , if $d_{out}(x) \geq d_{out}(y)$ then G contains a path from x to y of length at most 2.

Proof. Let x and y be two vertices in V such that $d_{out}(x) \geq d_{out}(y)$ and assume towards a contradiction that the claim does not hold. We must have $x \neq y$ and $N_{in}(y) \cap (\{x\} \cup N_{out}(x)) = \emptyset$ as otherwise the distance from x to y would be at most 2. Since the underlying diameter of G is 1, we get that $\{x\} \cup N_{out}(x) \subseteq N_{out}(y)$, and so that $d_{out}(y) = |N_{out}(y)| \geq |\{x\} \cup N_{out}(x)| > |N_{out}(x)| = d_{out}(x)$ which is a contradiction. \triangleleft

▷ **Claim 4.5.** For every $k \in \{0, \dots, n-1\}$ and every two vertices x and y in V , if $f_k[d_{out}(x)] \neq \perp$ and $f_k[d_{out}(x)] \geq d_{out}(y)$, then G contains a path from x to y of length at most $3k + 2$.

Proof. Let $k \in \{0, \dots, n-1\}$ and $x, y \in V$ be such that $f_k[d_{out}(x)] \neq \perp$ and $f_k[d_{out}(x)] \geq d_{out}(y)$. Let $x_0 = x$ and for each $i \in \{1, \dots, k\}$ let x_i and x'_i be two vertices such that $(x'_i, x_i) \in E$ and $d_{out}(x_i) = f_i[d_{out}(x)]$ and $d_{out}(x'_i) \leq f_{i-1}[d_{out}(x)]$ (note that such vertices must exist as $f_i[d_{out}(x)] \neq \perp$ holds for every such i , by Claim 4.1).

We first prove that for each $i \in \{0, \dots, k\}$ there exists in G a path from x_0 to x_i of length at most $3i$. We do this by induction on i . The base case ($i = 0$) is trivial. Assume that the claim holds for some $i \in \{0, \dots, k-1\}$ and prove it for $i+1$. We have $d_{out}(x_i) = f_i[d_{out}(x)] \geq d_{out}(x'_{i+1})$ and so, by Claim 4.4, we get that there is a path in G from x_i to x'_{i+1} of length at most 2. As in addition $(x'_{i+1}, x_{i+1}) \in E$, it must be that there is a path of length at most 3 from x_i to x_{i+1} which together with the induction hypothesis gives a path of length at most $3(i+1)$ from x_0 to x_{i+1} .

Note that we also have $d_{out}(x_k) = f_k[d_{out}(x)] \geq d_{out}(y)$ and so, by Claim 4.4, there is a path of length at most 2 from x_k to y . We conclude that there is a path of length at most $3k+2$ from x_0 to y . \triangleleft

Now, we put everything together.

▷ **Claim 4.6.** For every two vertices x and y in V , the following holds:

1. If y is not reachable from x , then there is no index $k \in \{0, \dots, n-1\}$ such that $f_k[d_{out}(x)] \neq \perp$ and $d_{out}(y) \leq f_k[d_{out}(x)]$.
2. If y is reachable from x , then such an index exists, and moreover $d(x, y, G) \leq 3k+2 \leq 3d(x, y, G) + 2$ where k is the minimal index in $\{0, \dots, n-1\}$ for which this property holds.

Proof. The first part follows from Claim 4.5 as the existence of such an index would imply the existence of a path from x to y . For the second part, note that by Claim 4.3 such an index exists. Let k be the minimal index with that property. We have $d(x, y, G) \geq k$ (by Claim 4.3) and $d(x, y, G) \leq 3k+2$ (by Claim 4.5), that is, we have $d(x, y, G) \leq 3k+2 \leq 3d(x, y, G) + 2$. \triangleleft

The above claim can be easily turned into an algorithm. Each vertex starts by broadcasting its out-degree to all the vertices in the network. In the next round, each vertex finds the maximal out-degree of its outgoing neighbors and broadcasts this value. By using this information, each vertex can compute for every $u \in V$ and $i \in \{0, \dots, n\}$ the value of $f_i[d_{out}(u)]$, and then it can compute an estimation for the distance between every pair of vertices x and y by finding the required index as in Claim 4.6.

5 Lower Bounds for Networks of Diameter 2

In this section we prove lower bounds for the single-source reachability problem and the closely related single-source shortest path problem for unweighted directed graphs in the Broadcast CONGEST model that hold even when the underlying network has diameter 2.

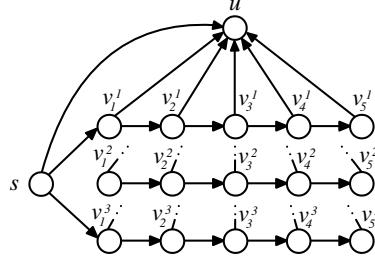


Figure 1 An illustration of the graph $G(k, q, \sigma)$ for $k = 3$, $q = 5$ and $\sigma = 101$.

5.1 The Single-Source Reachability Problem

We start this section by describing a family (parameterized by two positive integers k and q) of directed graphs with underlying diameter at most 2 which we denote by $F_{k,q}$. This family will be used later on to prove the required lower bound.

The Family $F_{k,q}$. For two positive integers $k, q \in \mathbb{Z}$ and a k -bit string $\sigma \in \{0, 1\}^k$, we define the directed graph $G(k, q, \sigma)$ to be the graph that consists of:

- k vertex-disjoint directed paths P_1, \dots, P_k with q vertices each (that is, $P_i = (V_i, E_i)$ where $V_i = \{v_1^i, \dots, v_q^i\}$ and $E_i = \{(v_j^i, v_{j+1}^i) \mid j \in \{1, \dots, q-1\}\}$ for every $i \in \{1, \dots, k\}$).
- A source vertex s that has an outgoing edge to v_1^i (the first vertex of P_i) if the i -th bit of σ is 1, for every $i \in \{1, \dots, k\}$.
- A sink vertex u to which s and every vertex in P_1, \dots, P_k has an outgoing edge.

In other words, the vertex set of the graph $G(k, q, \sigma)$ is $V_1 \cup \dots \cup V_k \cup \{s, u\}$ and its edge set is $E_1 \cup \dots \cup E_k \cup \{(s, v_1^i) \mid i \in \{1, \dots, k\}\} \cup \{(x, u) \mid x \in \{s\} \cup V_1 \cup \dots \cup V_k\}$ (see Figure 1 for an illustration). For two positive integers k and q , we define the family $F_{k,q}$ to be the set $\{G(k, q, \sigma) \mid \sigma \in \{0, 1\}^k\}$.

Our next goal is to show that any distributed algorithm that solves the single-source reachability problem for all the graphs in $F_{k,q}$ requires a significant number of rounds. We start with the following lemma:

► **Lemma 5.1.** *Let k and q be two positive integers. Let $G \in F_{k,q}$ and let φ be some legal assignment of identifiers to its vertices. Let A be some deterministic distributed algorithm (in the Broadcast CONGEST model) that solves the single-source reachability problem on the instance (G, φ, s) using at most t rounds (for some non-negative integer $t < q$). For each $i \in \{1, \dots, k\}$, the output of the vertex v_q^i by the end of the last round is just function of the initial input of the vertices v_{q-t}^i, \dots, v_q^i and the sequence of messages that v_q^i received from u .*

Proof. Let $i \in \{1, \dots, k\}$. We will show by induction on $0 \leq j \leq t$ that by the end of the j -th round the state of each vertex $v \in \{v_{q-t+j}^i, \dots, v_q^i\}$ is just a function of the initial input of the vertices in its ball of radius j (in the underlying graph of P_i) and the sequence of messages that it received from u up to this round.

The base case ($j = 0$) clearly holds as the state of each vertex in $\{v_{q-t}^i, \dots, v_q^i\}$ by the end of round 0 can depend only on its initial input. Assume now that the claim holds for some $0 \leq j < t$ and prove it for $j + 1$. Let $r \in \{q - t + j + 1, \dots, q\}$. The state of v_r^i by the end of the $(j + 1)$ -th round is a function of its state at the end of the previous round and the messages that it received from its neighbors in the underlying network (which are u , v_{r-1}^i and possibly v_{r+1}^i).

The messages that v_r^i has received from its neighbors in V_i are, by the induction hypothesis, functions of the inputs of the vertices in the balls of radius j (in P_i) around these neighbors and the sequence of messages that they received from u (up to round j). As u broadcasts the same message to all the vertices in each round, we get that these messages are just a function of the inputs of the vertices in the ball of radius $j + 1$ around v_r^i (in P_i) and the sequence of messages that v_r^i received from u (up to round j). As the previous state of v_r^i is, by the induction hypothesis, also a function of the initial inputs of the vertices in its ball of radius j (in P_i) and the sequence of message that it received from u , the claim follows. ◀

► **Lemma 5.2.** *Let k and q be two positive integers and let φ be some legal assignment of identifiers to $V_1 \cup \dots \cup V_k \cup \{s, u\}$. For every deterministic algorithm A (in the Broadcast CONGEST model), if A solves the single-source reachability problem on all the instances in $\{(G, \varphi, s) \mid G \in F_{k,q}\}$ and uses messages of size at most B bits (for some $B \geq 1$), then A requires at least $\min\{q - 1, k/(2B)\}$ rounds.*

Proof. Let A be some deterministic algorithm that satisfies the requirements of the lemma and let t be its running time. We can assume that $t \leq q - 2$ as otherwise there is nothing to show.

For each $G \in F_{k,q}$, we let $\text{out}(G)$ be the sequence $(\text{out}(v_q^1, G), \dots, \text{out}(v_q^k, G))$ where $\text{out}(v_q^i, G)$ is the output of v_q^i when A is invoked on (G, φ, s) , for every $i \in \{1, \dots, k\}$. Lemma 5.1 implies that for each $G \in F_{k,q}$ the value of $\text{out}(G)$ is just a function of the initial inputs in (G, φ, s) of the vertices $\bigcup_{i=1}^k \{v_q^{i-t}, \dots, v_q^i\}$ and the sequence of messages that u had broadcast. Since we assumed that $t \leq q - 2$, the initial inputs of these vertices is the same in all $\{(G, \varphi, s) \mid G \in F_{k,q}\}$, and so we can have $\text{out}(G) \neq \text{out}(G')$ for two graphs G and G' in $F_{k,q}$ only if the sequence of messages that u had broadcast in the corresponding invocations was different.

In each round, u may send a message that contains at most B bits, that is, a message with 0 bits, or with 1 bit and so on. Therefore, there are $1+2+\dots+2^B \leq 2^{2B}$ different messages that u may send in each round. It follows that there are at most 2^{2Bt} possible sequences and so we get that $|\{\text{out}(G) \mid G \in F_{k,q}\}| \leq 2^{2Bt}$. Note also that $|\{\text{out}(G) \mid G \in F_{k,q}\}| = 2^k$ as for each $G \in F_{k,q}$ the output should be different. We conclude that $2^k \leq 2^{2Bt}$ and so $t \geq k/2B$. ◀

► **Corollary 5.3.** *In the Broadcast CONGEST model, there is no deterministic algorithm that solves the single-source reachability problem in $o(\sqrt{n/\log n})$ rounds even when the diameter of the underlying network is always 2.*

Proof. Assume towards a contradiction that there exists a deterministic algorithm A that solves the above problem in $T(n) = o(\sqrt{n/\log n})$ rounds. As A works in the CONGEST model, there must be some constant $c \geq 1$ such that the number of bits in any message that the algorithm may send (when it is invoked on inputs of size $n > 1$) is at most $c \cdot \log n$.

Since $T(n) = o(\sqrt{n/\log n})$, there must be some integer $n_0 \geq 16$ for which $T(n) \leq \frac{1}{10c} \sqrt{n/\log n}$ holds for every $n > n_0$. Choose an integer $m > n_0$ such that both $k = \sqrt{m \log m}$ and $q = \sqrt{m/\log m}$ are positive integers. Lemma 5.2 implies that there must be some $G \in F_{k,q}$ and some assignment of identifiers φ to $V(G)$ such that invoking the algorithm on (G, φ, s) requires at least $\min\{\frac{1}{2}q, \frac{1}{4c} \frac{k}{\log m}\} = \min\{\frac{1}{2} \sqrt{\frac{m}{\log m}}, \frac{1}{4c} \sqrt{\frac{m}{\log m}}\} = \frac{1}{4c} \sqrt{\frac{m}{\log m}}$ rounds. But, we also have $|V(G)| > m > n_0$ and so the algorithm must take at most $\frac{1}{5c} \sqrt{m/\log m}$ rounds on (G, φ, s) , a contradiction. ◀

In the next lemma, we show that the same lower bound holds for distributed randomized algorithms as well.

► **Lemma 5.4.** *Let k and q be two positive integers and let φ be some legal assignment of identifiers to $V_1 \cup \dots \cup V_k \cup \{s, u\}$. For every randomized algorithm A (in the Broadcast CONGEST model), if A correctly solves the SSR problem on each instance in $\{(G, \varphi, s) \mid G \in F_{k,q}\}$ with probability $> 1/2$ and uses messages of size at most B bits (for some $B \geq 1$), then A requires at least $\min\{q - 1, (k - 1)/(2B)\}$ rounds.*

Proof. Clearly, it is sufficient to show that this lower bound holds in a model that generates a public random string first, announces it to every vertex in the graph and then every vertex proceeds deterministically as usual. Let A be a randomized algorithm that works in the above model and solves the SSR problem on every instance in $F = \{(G, \varphi, s) \mid G \in F_{k,q}\}$ with probability $> 1/2$. We can assume that its running time t is at most $q - 2$. As in the proof of Lemma 5.2, we can show that, for every fixed random string r , the algorithm (given that string) can succeed on at most 2^{2Bt} of the instances in F .

For each graph G in $F_{k,q}$, let R_G be the event that the algorithm fails on (G, φ, s) . Note that we must have $\sum_{G \in F_{k,q}} P(R_G) \geq |F_{k,q}| - 2^{2Bt}$. By assumption, we must also have $0.5|F_{k,q}| > \sum_{G \in F_{k,q}} P(R_G)$ and so $0.5|F_{k,q}| > |F_{k,q}| - 2^{2Bt}$ which implies that $t > (k - 1)/2B$. ◀

5.2 The Single-Source Shortest Path Problem

The result of the previous section already gives a lower bound of $\Omega(\sqrt{n/\log n})$ for the directed SSSP problem (or even for the approximate version of it) as, by definition, any algorithm that solves this problem must also solve the SSR problem.

In this section we show a slightly stronger lower bound of $\Omega(\sqrt{n})$ for this problem which holds even when the diameter of the underlying network is 2 and even when all the vertices in the input graph are guaranteed to be reachable from the given source. As in the previous section, we start by describing a family J_k of unweighted directed graphs with underlying diameter 2 which will be used to prove the lower bound.

The Family J_k . For a positive integer k and a sequence σ of k numbers from $\{1, \dots, k\}$, we define the directed graph $G(k, \sigma)$ to be the graph that consists of:

- k vertex-disjoint directed paths P_1, \dots, P_k where each $P_i = (V_i, E_i)$ contains $\sigma(i) + k$ vertices. For each path P_i , we denote by u^i its first vertex and by v_1^i, \dots, v_k^i its last k vertices.
- A source vertex s that has an outgoing edge to the first vertex of every path in $\{P_1, \dots, P_k\}$.
- A sink vertex u to which s and every vertex in P_1, \dots, P_k has an outgoing edge.

In other words, the vertex set of the graph $G(k, \sigma)$ is $V_1 \cup \dots \cup V_k \cup \{s, u\}$ and its edge set is $E_1 \cup \dots \cup E_k \cup \{(s, u^i) \mid i \in \{1, \dots, k\}\} \cup \{(x, u) \mid x \in \{s\} \cup V_1 \cup \dots \cup V_k\}$ (see Figure 2 for an illustration). For a positive integer k , we define the family J_k to be the set $\{G(k, \sigma) \mid \sigma \in \{1, \dots, k\}^k\}$.

We say that a collection of assignments $\{\varphi_G \mid G \in J_k\}$ is a consistent set of assignments for the family J_k , if φ_G is a legal assignment of identifiers to $V(G)$ and $\varphi_G(x) = \varphi_{G'}(x)$, for every $G, G' \in J_k$ and $x \in \{u\} \cup (\bigcup_{i=1}^k \{v_1^i, \dots, v_k^i\})$.

► **Lemma 5.5.** *Let $k > 1$ be some integer and let $\{\varphi_G \mid G \in J_k\}$ be some consistent set of assignments for J_k . For every deterministic algorithm A (in the Broadcast CONGEST model), if A solves the SSSP problem on all the instances in $\{(G, \varphi_G, s) \mid G \in J_k\}$, then A requires $\Omega(k)$ rounds.*

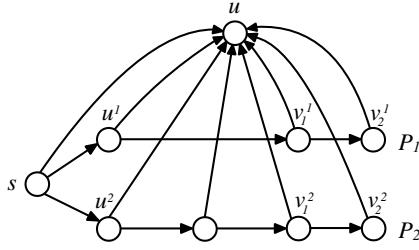


Figure 2 An illustration of the graph $G(k, \sigma)$ for $k = 2$ and $\sigma = (1, 2)$.

Proof. Let A be some deterministic algorithm that satisfies the requirements of the lemma and let t be its running time. We can assume that $t \leq k - 2$ as otherwise there is nothing to show. As A works in the CONGEST model, there must be some constant $c \geq 1$ such that the number of bits in any message that the algorithm may send on any input of size $n > 1$ is at most $c \cdot \log n$.

Consider invoking the algorithm A on the instance (G, φ_G, s) for some $G \in J_k$. In each round, u may send one message containing at most $c \cdot \log(|V(G)|) \leq 4c \cdot \log(k)$ bits to all the vertices in the graph. Given that, it is easy to see (by a proof similar to Lemma 5.1) that the output of every v_k^i is just a function of its initial input, the initial input of at most $t \leq k - 2$ vertices that precede it in the path P_i , and the sequence of messages that u had broadcast.

For each $G \in J_k$, we let $\text{out}(G)$ be the sequence $(\text{out}(v_k^1, G), \dots, \text{out}(v_k^k, G))$ where $\text{out}(v_k^i, G)$ is the output of v_k^i when A is invoked on (G, φ_G, s) , for every $i \in \{1, \dots, k\}$. By the observation above, for each $i \in \{1, \dots, k\}$ the vertices in P_i whose initial input may affect the output of v_k^i are just v_2^i, \dots, v_k^i . Since the initial input of each of these vertices is the same in each of the instances in $\{(G, \varphi_G, s) \mid G \in J_k\}$, we get that $\text{out}(G) \neq \text{out}(G')$ can hold for some graphs G and G' in J_k only if the sequence of messages that u had broadcast in the corresponding invocations was different.

Straightforward calculations show that the number of such sequences is at most $(2^{8c \cdot \log(k)})^t = k^{8c \cdot t}$, and so $|\{\text{out}(G) \mid G \in J_k\}| \leq k^{8c \cdot t}$. Since we assumed that the algorithm is correct, we must have $\text{out}(G) \neq \text{out}(G')$ for every two different graphs G and G' in J_k (as, by construction, we cannot have $d(s, v_k^i, G) = d(s, v_k^i, G')$ for every $i \in \{1, \dots, k\}$), and so $|\{\text{out}(G) \mid G \in J_k\}| = |J_k| = k^k$. We conclude that $k^k \leq k^{8c \cdot t}$ and so $t \geq k/(8c)$. \blacktriangleleft

► Corollary 5.6. *In the Broadcast CONGEST model, there is no deterministic algorithm that solves the SSSP problem in $o(\sqrt{n})$ rounds even when the diameter of the underlying network is always 2.*

Proof. Assume towards a contradiction that there exists a deterministic algorithm A that solves the above problem in $T(n) = o(\sqrt{n})$ rounds. As before, we can assume that there is a constant $c \geq 1$ such that the number of bits in any message that A may send on any input of size $n > 1$ is at most $c \cdot \log n$.

Since $T(n) = o(\sqrt{n})$, there must be some positive integer n_0 for which $T(n) \leq \frac{1}{18c} \sqrt{n}$ holds for every $n > n_0$. Let $k > 1$ be an integer such that $k^2 > n_0$. The proof of Lemma 5.5 implies that there must be some $G \in J_k$ and some assignment of identifiers φ to $V(G)$ such that invoking the algorithm on (G, φ, s) requires at least $k/(8c)$ rounds. But, $|V(G)| > k^2 > n_0$ and so the algorithm must take at most $\frac{1}{18c} \sqrt{|V(G)|} \leq \frac{1}{9c} k$ rounds on (G, φ, s) , a contradiction. \blacktriangleleft

By a proof similar to that of the previous section, it is possible to show that the same lower bound holds for randomized distributed algorithms as well.

References

- 1 Udit Agarwal, Vijaya Ramachandran, Valerie King, and Matteo Pontecorvi. A Deterministic Distributed Algorithm for Exact Weighted All-Pairs Shortest Paths in $\tilde{O}(n^{3/2})$ Rounds. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing (PODC '18)*, pages 199–205, 2018.
- 2 Ruben Becker, Andreas Karrenbauer, Sebastian Krininger, and Christoph Lenzen. Near-Optimal Approximate Shortest Paths and Transshipment in Distributed and Streaming Models. In *Proceedings of the 31st International Symposium on Distributed Computing (DISC '17)*, pages 7:1–7:16, 2017.
- 3 Aaron Bernstein and Danupon Nanongkai. Distributed Exact Weighted All-pairs Shortest Paths in Near-linear Time. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC '19)*, pages 334–342, 2019.
- 4 Keren Censor-Hillel, Michal Dory, Janne H. Korhonen, and Dean Leitersdorf. Fast Approximate Shortest Paths in the Congested Clique. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC '19)*, pages 74–83, 2019.
- 5 Keren Censor-Hillel, Petteri Kaski, Janne H. Korhonen, Christoph Lenzen, Ami Paz, and Jukka Suomela. Algebraic Methods in the Congested Clique. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC '15)*, pages 143–152, 2015.
- 6 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. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing (STOC '11)*, pages 363–372, 2011.
- 7 Michael Elkin. Distributed Exact Shortest Paths in Sublinear Time. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC '17)*, pages 757–770, 2017.
- 8 Sebastian Forster and Danupon Nanongkai. A Faster Distributed Single-Source Shortest Paths Algorithm. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS '18)*, pages 686–697, 2018.
- 9 Mohsen Ghaffari and Jason Li. Improved Distributed Algorithms for Exact Shortest Paths. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC'18)*, pages 431–444, 2018.
- 10 Mohsen Ghaffari and Rajan Udwani. Brief Announcement: Distributed Single-Source Reachability. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC '15)*, pages 163–165, 2015.
- 11 Monika Henzinger, Sebastian Krininger, and Danupon Nanongkai. A Deterministic Almost-tight Distributed Algorithm for Approximating Single-source Shortest Paths. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing (STOC '16)*, pages 489–498, 2016.
- 12 Stephan Holzer and Roger Wattenhofer. Optimal Distributed All Pairs Shortest Paths and Applications. In *Proceedings of the 2012 ACM Symposium on Principles of Distributed Computing (PODC '12)*, pages 355–364, 2012.
- 13 Chien-Chung Huang, Danupon Nanongkai, and Thatchaphol Saranurak. Distributed Exact Weighted All-Pairs Shortest Paths in $\tilde{O}(n^{5/4})$ Rounds. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS '17)*, pages 168–179, 2017.
- 14 Christoph Lenzen and Boaz Patt-Shamir. Fast Partial Distance Estimation and Applications. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC '15)*, pages 153–162, 2015.
- 15 Christoph Lenzen and David Peleg. Efficient Distributed Source Detection with Limited Bandwidth. In *Proceedings of the 2013 ACM Symposium on Principles of Distributed Computing (PODC '13)*, pages 375–382, 2013.

- 16 Danupon Nanongkai. Distributed Approximation Algorithms for Weighted Shortest Paths. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing (STOC '14)*, pages 565–573, 2014.
- 17 David Peleg. *Distributed Computing: A Locality-sensitive Approach*. Society for Industrial and Applied Mathematics, 2000.