How Long It Takes for an Ordinary Node with an Ordinary ID to Output?

Laurent Feuilloley*

Institut de Recherche en Informatique Fondamentale (IRIF),
CNRS and University Paris Diderot,
France

Abstract. In the context of distributed synchronous computing, processors perform in rounds, and the time complexity of a distributed algorithm is classically defined as the number of rounds before all computing nodes have output. Hence, this complexity measure captures the running time of the slowest node(s). In this paper, we are interested in the running time of the ordinary nodes, to be compared with the running time of the slowest nodes. The *node-averaged* time-complexity of a distributed algorithm on a given instance is defined as the average, taken over every node of the instance, of the number of rounds before that node output. We compare the node-averaged time-complexity with the classical one in the standard LOCAL model for distributed network computing. We show that there can be an exponential gap between the node-averaged time-complexity and the classical time-complexity, as witnessed by, e.g., leader election. Our first main result is a positive one, stating that, in fact, the two time-complexities behave the same for a large class of problems on very sparse graphs. In particular, we show that, for LCL problems on cycles, the node-averaged time complexity is of the same order of magnitude as the "slowest node" time-complexity. In addition, in the LOCAL model, the time-complexity is computed as a worst case over all possible identity assignments to the nodes of the network. In this paper, we also investigate the *ID-averaged* time-complexity, when the number of rounds is averaged over all possible identity assignments of size $O(\log n)$. Our second main result is that the ID-averaged time-complexity is essentially the same as the expected time-complexity of *randomized* algorithms (where the expectation is taken over all possible random bits used by the nodes, and the number of rounds is measured for the worst-case identity assignment). Finally, we study the node-averaged ID-averaged time-complexity. We show that 3-colouring the n-node ring requires $\Theta(\log^* n)$ rounds if the number of rounds is averaged over the nodes, or if the number of rounds is averaged over the identity assignments. In contrast, we show that 3-colouring the ring requires only $O(1)$ rounds if the number of rounds is averaged over the nodes, and over the identity assignments.

* The author received additional support from ANR project DESCARTES, and Inria project GANG.

1 Introduction

The LOCAL model [22] is a standard model of distributed network computing. In this model, the network is abstracted as a graph, and the nodes perform in rounds to solve some task. At each round, each node can send messages to its neighbours in the graph, receive messages and perform some computation. The complexity of an algorithm solving some task is measured by the number of rounds before the task is completed, which usually depends on the size of the network, that is, its number of nodes.

A classic assumption in the LOCAL model is that the nodes know the size of the network *a priori*. As a consequence, in many algorithms, each node can compute from the start how many rounds are needed to solve the task, and stops after that number of rounds. There have been efforts to remove such *a priori* knowledge of the parameters of the graph (*e.g.* the arboricity [3] and the maximum degree [19]). Quite recently a general technique, called *pruning algorithms*, has been developed to remove the assumption that nodes know the size n of the network [15]. In other words, [15] provides a method to transform a non-uniform algorithm into a uniform algorithm. The basic idea is to guess the number of nodes and to apply a non-uniform algorithm with this guess. The output can be incorrect, as the algorithm is only certified to be correct when it is given the actual number of nodes in the graph. The technique consists in virtually removing from the graph the nodes that have correct outputs, and to repeat the previous procedure with a new guess that is twice as large as the previous guess. Eventually all nodes have an output after a certain number of iterations, and the solution that is computed is correct. Note that with the resulting uniform algorithm some nodes can output very quickly, and some others can output much later. So far, only the classic measure of complexity, *i.e.* the time before all nodes stop and output, has been studied, even for such algorithms. In other words, only the behaviour of the *slowest* node has been considered. In this paper, we introduce a new measure of complexity, which is the *average* measure, in opposition to the usual measure which is a *worst-case* measure. More precisely, we define the running time of a node as the number of rounds before it outputs, and consider the average of the running times of the nodes. We argue that, when studying the locality of problems and of algorithms, it is worth to also consider this measure. Indeed it describes the typical local behaviour of the algorithm, that is, the behaviour of an *ordinary* node.

In some contexts partial solutions are useful. For example, consider the scenario in which two tasks are to be performed one after the other. In such case, it may happen that, on some part of the graph a partial solution for the first task is computed quickly. We can take advantage of this to start the second task in that part of the network, while the other nodes are still working on the first task. Note that knowing if the first task is finished can be impossible locally, and one has to design the second algorithms such it start at different rounds on different nodes. Consider a second scenario in which a global operator has to take a decision based on the outcome of a local algorithm. In that case, a partial solution may also be sufficient. For example the operator can detect that the network is in a bad state, and start immediately a recovery procedure without waiting for all nodes to finish. Such situations are a motivation for the study of graph property testing, where a centralized algorithm probes the network on a sublinear number of nodes and take a decision based on this partial knowledge. We refer to the survey on graph property testing [13] for more examples of applications. When such partial solutions are useful, one would like to design algorithm that stop as soon as possible, and the average of the running time of the nodes is then a measure one would like to minimize.

Another classical assumption in the LOCAL model is that the nodes are given distinct identifiers. These identifiers (or IDs for short), are distinct binary strings on $O(\log n)$ bits, that is, distinct integers from a polynomially large space. The usual way to measure the complexity of an algorithm is again to consider its the worst-case behaviour, that is, the performance of the algorithm on the worst ID assignment. We argue that the average performances over all ID assignments, is also worth considering. Indeed many lower bounds are based on the fact that, as the identifiers can be viewed as set by an adversary, they do not really

help to break symmetry. For example, on a path, one may consider the identifier assignment $1, 2, ..., n$, and argue that if nodes only consider the relative ordering of the identifiers in their neighbourhoods, then many nodes have the same view, and thus they cannot break symmetry. It is interesting to study if such specific constructions are required, or if one can design lower bounds that are robust against arbitrary ID assignment. We cannot expect that IDs are always set in a perfect way for the task we consider, but it may seem excessive to consider that they are set in an adversarial way, which naturally leads to the question of random assignments. We study the complexity of algorithms on random ID assignment, as the average over all possible ID assignment of the running time of the slowest node. Finally, the typical behaviour of an algorithm can arguably be the expected running time of an ordinary node on a random ID assignment. That is, the standard complexity but averaged on both nodes and ID assignments.

For the sake of concreteness, here is an example of the type of questions tackled in this paper. Consider the classic task of 3-colouring a ring of n nodes. It is known that this task requires $\Omega(\log^* n)$ rounds [16]. This bound also holds for randomized algorithms [20]. The question tackled in this paper are of the following form: is it the case that a node typically outputs after a constant number of rounds, or is the $\Omega(\log^* n)$ lower bound robust to this spatial averaging? And what about the complexity of the problem on a random ID assignment?

Our results Our first result is that averaging on the nodes can have a dramatic effect on the time complexity of solving a task in the LOCAL model. Indeed, for leader election on cycles, there is an exponential gap between the node-averaged complexity and the classic complexity. That is the slowest node outputs after a number of rounds that is exponentially larger than the time complexity of an ordinary node. This contrasts with our next result, for very sparse graphs. We say that a graph is q-sparse, if every ball of radius r has at most $q.r$ nodes. For q-sparse graphs, we show that, for many classic tasks, the two measures are of the same order of magnitude. More precisely for a class of tasks that generalizes the class of *locally checkable labellings* (LCL for short) [21], we show the following lemma, that we call *local average lemma*. For a given algorithm, either no node has a running time much larger than the average in its neighbourhood, or there exists an algorithm that is strictly better, that is an algorithm that has smaller running time for every node in every graph. As a consequence when proving lower bounds for these problems, one can use the fact that there is no *peak* in the distribution of the running times of the nodes. Then, to show that the average running time is large, it suffices to show that there is a large enough number of nodes that are far enough one from the other and that have large running time . This local average lemma can be used to show, for example, that for LCL problems on cycles, the landscape of complexities for an ordinary node and for the slowest node is the same, that is, for every problem the complexity is either $\Theta(1)$, $\Theta(\log^* n)$ or $\Theta(n)$.

We then move on to averaging on the identifier assignments. That is, we consider the expected behaviour of deterministic algorithms on random ID assignments. This topic happens to be related with the expected complexity of randomized algorithms. We show that even though these two models have specific properties, namely the independence of the random strings for the randomized algorithms, and the uniqueness of the identifiers for random ID assignment, the complexities are essentially the same. It follows that the results known for randomized algorithms can be translated to average over the identifiers.

Finally we prove that averaging on both nodes and IDs, can have an important effect on the complexity. We take the example of 3-colouring an n-node cycle. From the previous results of the paper, and from the literature, we know that this task has complexity $\Omega(\log^* n)$ for both the average on the nodes and the average on the identifiers. Quite surprisingly, when averaging on both the nodes and the ID assignment, the complexity becomes constant. In other words, deterministic and randomized complexity of ordinary nodes are clearly separated. Such separation contrast with the situation when considering the classic measure,

as randomized constant-time algorithms for LCL, can be derandomized to get constant-time deterministic algorithms [21].

Related works The LOCAL model was defined in [16], and a standard book on the topic is [22]. The problem of leader election, studied in section 3, is a classic problem in distributed computing [2, 18].

Deterministic algorithms stopping after different number of rounds on different nodes have been studied in contexts where the parameters of graphs, such as the degree or the number of vertices, are unknown. Such algorithms are called uniform algorithm, because it is the same algorithms that is run on every graph, independently of the parameters. A work that is particularly relevant to us is [15]. In this paper the authors prove that for a wide class of problems, one can remove the assumption that the nodes know the size n of the network. This is done by applying a general method to transform a non-uniform algorithm into a uniform algorithm, without increasing of the asymptotic running time. In this framework, called *pruning algorithms*, some nodes may stop very early and some may run for much longer time. Such algorithms justify the study of the behaviour of an ordinary node and not only of the behaviour of the slowest node.

The local average lemma of section 4 applies to problems that are local in the sense that the nodes can check in constant time if a given solution is correct. This is an extension of the well-studied notion of locally checkable labelling (or LCL for short) [21], which is similar but requires in addition that the size of the inputs and of the outputs are bounded. Also the set of correct labellings usually studied, *e.g.* in distributed decision [9], including in LCL, do not depend on the identifiers of the graph, a restriction that is not needed here.

Randomized algorithms that turn out to be equivalent to algorithms working on random ID assignment form a well-studied subject, going back to the 80s with algorithms for maximal independent sets [1, 17]. Recently, improvements on classic problems have been obtained [12, 14] along with an exponential separation between randomized and deterministic complexity [6] (see also [4]). In [12], the author, by advocating the study of the so-called *local complexity* for a randomized algorithms, conveys the same message as the current paper: the behaviour of a typical node is worth considering, even if some nodes of the graph have much worst behaviour.

In this paper, we consider two relaxations of the measure of complexity, from worst-case to average, on the nodes and on the IDs. An aspect that we do not consider is the structure of the graph. We refer to [11] and references therein, for the topic of local algorithms on random graphs.

Finally, part of the results of this paper appeared in a brief announcement at PODC 2015 by the current author [8].

2 Model and definitions

The graph considered in this paper are simple connected graphs, and throughout the text n will denote the number of nodes in the graph. The distance between two nodes is the number of edges on a shortest path between these nodes, that is, the hop-distance. The k-neighbourhood of a node v in a graph G, is the graph induced by the nodes at distance at most k from v. Every node is given a distinct identifier on $O(\log n)$ bits, or equivalently an integer from a polynomially large range.

The algorithms studied in this paper can be defined in two ways. In both definitions, the nodes are synchronized and work in rounds, and for both the computational power of the nodes is unbounded. In the first definition, at each round, every node can exchange messages with its neighbours, and perform some computation. There is no bound on the size of the messages. A given node chooses an output after some number of rounds, and different nodes can stop at different rounds. After the output, a node can continue to transmit messages and perform computations, but it cannot change its output. In other words, the nodes

do not go to a sleep mode once they have output, but the output is irrevocable. In the second definition, each node starts with the information of its 0-neighbourhood, and increases the size of this view at each round. That is, after k rounds, it knows its k-neighbourhood, that is it knows the structure of the graph in this neighbourhood, along with the identifiers and the inputs of the nodes. At some round, it chooses an output and stops. These two definitions are equivalent. On one hand, if we start from the first definition, we can assume that each round every node sends to its neighbours all the information it has about the graph (remember that the message size is unbounded)[1]. Then after k rounds, a node has gathered the information about its k-neighbourhood. On the other hand, given a k-neighbourhood, a node can simulate the run of the other nodes, and compute the messages that it would receive if the nodes were using a message-passing algorithm.

The running time of a node is the number of rounds before it outputs. With the second definition, the running time of the algorithm can be described in a more combinatorial way: it is the minimum k such that the node can choose an output with a view of radius k. Given a graph G, an identifier assignment I (from the set of legal ID assignments that we denote $\mathcal{ID}$), an algorithm A, and a node v, we denote by $r_{G,I,A}(v)$ the running time of node v in this context. When the context is clear, we simply use $r(v)$. We now define the different measures of complexity used in this paper. Given a graph G, and an algorithm A, we call *complexity of the slowest node complexity* or *classical complexity*, and *complexity of an ordinary node* or *node-averaged complexity* respectively, the following quantities:

$$\max_{I\in\mathcal{ID}} \max_{v\in G} r_{G,I,A}(v) \qquad \text{and} \qquad \max_{I\in\mathcal{ID}} \frac{1}{n}\sum_{v\in G} r_{G,I,A}(v)$$

In the second part of this paper, we consider the average on the identifier assignments and the average on both the identifiers and the nodes, that is, the following measures:

$$\frac{1}{|\mathcal{ID}|}\sum_{I\in\mathcal{ID}} \left(\max_{v\in G} r_{G,I,A}(v)\right) \qquad \text{and} \qquad \frac{1}{|\mathcal{ID}|}\sum_{I\in\mathcal{ID}} \left(\frac{1}{n}\sum_{v\in G} r_{G,I,A}(v)\right)$$

The tasks or problems that we want to solve in a distributed manner, are formalized with the notion of *language*. A language $\mathcal{L}$ is a set of configurations of the form (G, I, x, y), where G is a graph, I an identifier assignment, and x and y are functions from the nodes of the graph to a set of labels. We are interested in constructing these languages, which means that given a graph G, an ID assignment I and inputs given by the function x, we want to compute locally a function y such that (G, I, x, y) is in the language $\mathcal{L}$. The languages considered are such that for every (G, I, x), there exists a legal output y. Note that usually, the identifier assignment is not part of the language [9, 10, 21], but our results hold for this more general version.

In section 3, we use the most general option regarding the knowledge of n by the nodes, we assume such knowledge for lower bounds, whereas for upper bounds we do not require it. For section 4, we assume that nodes do not have the knowledge of n. For the randomized part we assume this knowledge for the sake of simplicity, and refer the reader to subsection 4.4 of [15] for a technique to remove such assumption for randomized algorithm.

Throughout the paper, the expression *with high probability* means with probability at least $1 - 1/n$.

[1] There is a subtlety here, which is that after k rounds in the message-passing algorithm a node cannot know the edges that are between nodes at distance exactly k from it. For the sake of simplicity, we consider the proper k-neighbourhoods, as it does not affect the asymptotic of the algorithms.

3 Exponential gap for a global language

The complexity of an ordinary node is bounded by the complexity of the slowest node by definition. In this section, we show that the gap between these two quantities can be exponential.

Theorem 1 *The gap between the averaged-node complexity and the classical complexity can be exponential.*

We illustrate this phenomenon on the classic problem of leader election. The language of leader election is the set of graphs, with no inputs and binary outputs, such that exactly one node has label 1, and the others have label 0. It does not depend on the ID assignment. It is folklore that this problem has classic complexity $\Theta(n)$.

Proposition 1 (Folklore). *Leader election on an n-node ring requires $\Theta(n)$ rounds (for the slowest node).*

We prove this statement for completeness. The complexity of leader election in various models is discussed in [2, 18].

Proof. Let A be an algorithm for leader election, which has access to the size of the graph. Suppose that the slowest node complexity of A is $c(n) \in o(n)$. Let n_0 be a large enough constant such that $2c(n_0)+1 < n_0/2$. Consider a ring R_1 of length n_0. After running the algorithm A on R_1, a node v_1 is elected to be the leader. This node v_1 outputs 1, after at most $c(n_0)$ steps. That is, v_1 outputs based on a view that contains at most $2c(n_0) + 1$ nodes. Because of the definition of n_0, this view contains less than $n_0/2$ nodes. Let I_1 be the set of identifiers in this view. Now consider another ring R_2 of length n_0, whose set of identifiers does not contain any of the IDs of I_1. Again, a node v_2 is designated as the leader, and its view contains less than $n_0/2$ nodes. Now consider the ring made by concatenating the two views, and adding dummy nodes with fresh identifiers, to make sure that the ring has size n_0. Because the identifiers are all distinct, this is a proper instance. Then, as v_1 and v_2 have the same view as in R_1 and R_2 respectively, with the same graph size n_0, they output the same as in R_1 and R_2. That is, they both output 1, and thus produce a configuration that is not in the language, which a contradiction. □

Proposition 2. *The complexity of an ordinary node for leader election on an n-node ring is $O(\log n)$.*

Proof. Consider the following algorithm. Each node increases its radius until one of the two following situation occurs. First, if it detects an ID that is larger than its own, then it outputs 0. Second, if it can see the whole ring and that no ID is larger than its own, then it outputs 1. It is easy to see that this algorithm is correct as only the node with maximum ID can output 1. Note that this algorithm is order-invariant in the sense of [21], *i.e.* the algorithm does not take into account the identifiers themselves, but only their relative ordering in its view. In particular, the algorithm does not require the knowledge of n. We show that the averaged-node complexity of this algorithm is logarithmic in n.

Let us first make an observation. Consider the nodes with the k largest identifiers, and mark them. The nodes that are not marked form k paths, some of them possibly empty. A key property is that the behaviour of the algorithm on one path is independent of the other paths. More precisely, we claim that on a given path the algorithm will have the same behaviour whatever the sizes and the identifier distributions of the other paths are. Fix a path, and a node v, in this path. By definition, v has an identifier that is smaller than the ones of the two marked nodes at the endpoints of the path. Therefore, it stops either before, or just when reaching one of the marked nodes, and it outputs 0. As a consequence it will never get to know the rest of the cycle. This simple observation implies that we can study the behaviour of the algorithm on each path separately.

Let p be integer, and let us consider a path of length p with two additional marked nodes at each endpoint. In order to study the behaviour of the algorithm on this path, it is sufficient to consider all the relative ordering of identifiers on this path, because it is an order-invariant algorithm. Marked nodes can be replaced by nodes with IDs larger than every ID in the path. Let $a(p)$ be the maximum over all these identifier assignments of the sum of the running time of the nodes. We claim that this function obeys the following recursion formula:

$$a(p) = \max_{1 \leq k \leq \lceil p/2 \rceil} \{k + a(k-1) + a(p-k)\}.$$

Consider the node v with the largest identifier in the path. It must reach one of the endpoints to stop. Then if we mark this node, the behaviour of the algorithm on the two subpath is independent of the context, and the maximum sums of running times in each path is $a(p_1)$ and $a(p_2)$ for the first subpath of length p_1 and the second of length p_2 respectively. Then the only parameter is the distance k from v to the closest endpoint. Given such an integer k, $a(p)$ is then equal to $k + a(k-1) + a(p-k)$. One can then check by induction that this maximum is always met for the value $k = \lceil p/2 \rceil$. Then an alternative formula is:

$$a(p) = \left\lceil \frac{p}{2} \right\rceil + a\left(\left\lceil \frac{p}{2} \right\rceil\right) + a\left(\left\lceil \frac{p}{2} \right\rceil - 1\right)$$

The sequence $a(n)$, defined by the induction formula above, along with initial values $a(0) = 0$ and $a(1) = 1$, is known to be in $\theta(n \log n)$. For references and more information about this sequence, see [25]. Consequently, the sum of the running times of the nodes is equal to the sum of the running time of the leader, which is $n/2$, and of $a(n-1)$. This is because, we can mark the node that has the largest ID, and consider the rest of graph as a path. Thereafter, the complexity of an ordinary node is logarithmic in n. □

Note that analysis of the same flavour already exist in the literature, see for example [24] p.125. Theorem 1 follows from propositions 1 and 2.

4 Local average lemma and application

This section is devoted to proving that, for local languages on very sparse graphs, the complexity of an ordinary node is basically the same as the one of the slowest node. This proof is based on a *local average lemma*. Given a graph and an algorithm, let us define informally a *peak*, as a node with high running time, whose neighbours at some distance have much smaller running times in average. The lemma states that, for local languages, and for algorithm that are somehow optimal, there is no such peak.

In order to give an intuition of this lemma, let us use the example of the 3-colouring a cycle. Consider an algorithm for the problem, and three adjacent nodes u, v and w, in this order, in a cycle. We claim that if $r(v) > \max(r(u), r(w)) + 1$, then the algorithm can be speeded up. Indeed after $\max(r(u), r(w)) + 1$ steps, v can simulate the computation of u and w, deduce the colours they output, and output a non-conflicting colour. As a consequence if one wants to prove a lower bound on the average of the running times, one can assume that, for every node, at least one of its neighbours has a similar running time, namely at least its running time minus one.

In this section the algorithm do not have the knowledge of n. In order to state the lemma we need to introduce a few notions.

Class LCL^* We consider a large class of distributed problems that we call LCL^*, which includes the well-known class of LCL problems [21], and the more general class LD [10]. A language $\mathcal{L}$ is in LCL^*, if there exists a constant-time *verification algorithm*. That is, an algorithm $\mathcal{V}$ performing in a constant number of

rounds, with binary output, *accept* or *reject*, such that for every configuration (G, I, x, y), $\mathcal{V}$ accepts at every node, if and only the graph is in the language $\mathcal{L}$. The running time of $\mathcal{V}$ is called the *verification* radius. No bound on the size of the inputs and output is necessary, and the language can depend on the identifiers.

q-sparse graphs A graph is q-sparse if any ball of radius r contains at most qr nodes. For example a cycle is 3-sparse.

Minimal algorithms The lemma has the following shape: given a node v whose running time is r, the nodes of its neighbourhood have running times whose average is roughly r. This type of statement cannot hold in general as we could artificially increase the radius of a node by modifying the algorithm. But as we are interested in lower bounds for the node-averaged complexity, we can consider algorithms that are in some sense minimal. More precisely, let A and A' be two distributed algorithms for some language $\mathcal{L}$. We say that A is smaller than A', if on every graph, every ID assignment and inputs, and on every node, the running time of A is at most the running time of A'. For lower bounds on the node-averaged complexity, it is sufficient to study algorithms that are minimal for this ordering. Indeed, if an algorithm that is not minimal has low complexity, then there exists a minimal algorithms that has at most this complexity.

Lemma 1 (Local average lemma). *Let $\mathcal{L}$ be a language in LCL^* with verification radius t, and A be a minimal algorithm for $\mathcal{L}$. There exists two positive constants α and β, such that on any q-sparse graph, ID assignment, inputs, and node v, the average of the running times of the nodes at distance at most $r(v)/2$ from v, is at least $\alpha.r(v) - \beta$.*

Let us denote by $B(v, k, G, I, x)$ the subgraph of G, with identifiers I, and inputs x, induced by the nodes at distance at most k from a node v. Likewise, given two integers $k_1 < k_2$, let $S(v, k_1, k_2, G, I, x)$ be the induced graphs with IDs and inputs, induced by the set of nodes whose distance to v is at least k_1 and at most k_2. When the context is unambiguous, we omit the information G, I and x.

Let $\mathcal{L}$ be a language of LCL^*. There exists a verification algorithm $\mathcal{V}$, such that a configuration (G, I, x, y) is in the language $\mathcal{L}$ if and only if $\mathcal{V}$ accepts at all node. Let t be the verification radius of $\mathcal{V}$. Let $\mathcal{L}$, A, G, I, v and x be respectively, a language, a minimal algorithm, a graph and an ID assignment, a node and an input assignment as in the lemma.

In order to prove the lemma, we will first prove the following claim.

Claim. For every integer k:

$$r(v) \leq 2k + 2t + \max_{u \in S(v,k,k+2t)} r(u)$$

Proof. Suppose the inequality does not hold for some k. Let us use the following notations:

$$M = \max_{u \in S(v,k,k+2t,G,I,x)} r(u) \qquad \text{and} \qquad B = B(v, k + 2t + M, G, I, x).$$

As in the example of 3-colouring at the beginning of this section, we define a new algorithm A', designed to be smaller than A. On a node w of a graph G', with ID assignment I', and inputs x', the behaviour of A' differs from the one of A only if the following conditions are fulfilled:

(1) The running time of w, $r_{G',I',A}(w)$, is at least $2k + 2t + M$;
(2) The node w is at distance at most k from a node whose neighbourhood at distance $k + 2t + M$ is exactly $B(v, k + 2t + M, G, I, x)$.

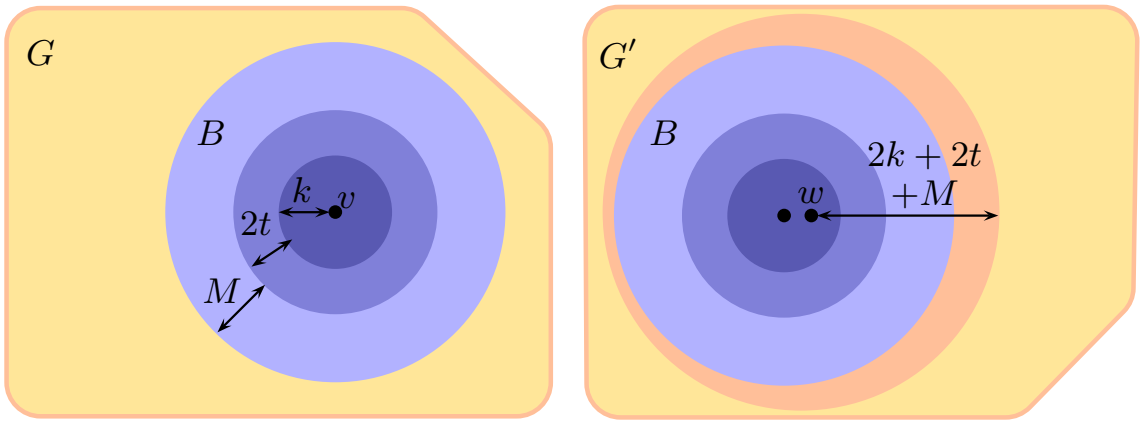

Fig. 1. This figure illustrates the definition of the algorithm A' in the proof of lemma 1. On the left is the original graph G with node v, along with the ball B around it. The behaviour algorithm A' differs from the algorithm A only if it is in the situation of the node w on the right: it has running time at least $2k+2t+M$, and it is at distance at most k from a node whose $(k+2t+M)$-neighbourhood is exactly B.

See figure 1. When the two conditions are fulfilled, let w_G be the node of G, whose position in B, ID, and input, are the same as the ones of w in G'. In that case, the algorithm A' stops at round $2k + 2t + M$, and outputs the same label as A does on w_G, in (G, I, x).

The algorithm A' is correct on G by construction, as it has exactly the same outputs as A. We prove that A' is correct on any graph. Consider the behaviour of the verification algorithm $\mathcal{V}$, on a node z of a graph G', with IDs I' after the run of A'. This node may reject, only if it can detect a difference between the outputs of A and A'. That is, only if A' has an output that is different from the output of A, on a node y of the t-neighbourhood of z. Because of the condition (2) in the definition of A', y is at distance at most k from a node v' whose neighbourhood at distance $k + 2t + M$ is exactly B. Then z is at distance at most $t + k$ from this node v'. It is then sufficient to show that the algorithm A' has correct outputs on the ball that is centred in v' and has radius $k + 2t$, as the whole view of $\mathcal{V}$ on z is contained in this ball.

First note that, the nodes at distance at most k from v' in G', have the same output with A', as the nodes at distance at most k from v in G with A. This is by definition A'. Second, remember that $M = \max_{u \in S(v,k,k+2t,G,I)} r(u)$. As a consequence, in G, the nodes of $S(v, k, k + 2t, G, I)$ stop before they see nodes outside of B. The same holds in G': as the $(k + 2t + M)$-neighbourhood of v' is also B, the nodes of $S(v', k, k + 2t, G', I')$ have the same behaviour as is in the previous case, that is, they stop before they discover that they are not in G. Therefore, the nodes of $S(v', k, k + 2t, G', I')$ output as if they were in $S(v, k, k + 2t, G, I)$.[2] Then A' has the same outputs as A, and as A is correct, A' is correct.

The algorithm A' is strictly smaller that A, indeed no running time has been increased, and the running time of v in G has been reduced to $2k + 2t + M$. Consequently A is not minimal, which is a contradiction. □

The final step of the proof lemma 1, requires some computations, and is given in appendix A. We give an intuition of this proof, considering a simplified version of the inequality of claim 4. Suppose we have the following inequality: $r(v) \leq \max_{S_k} r(u)$ where S_k is the set of nodes at distance exactly k. The quantity $\max_{S_k} r(u)$ is upper bounded by $\sum_{S_k} r(u)$. Then, summing both terms of the inequality, for k ranging

[2] Remember that the nodes do not have the knowledge of the size of the network, thus they have exactly the same information in G and G'.

from 1 to $r(v)$, one gets $r(v)^2 \leq \sum_S r(u)$, where S is the ball of radius $r(v)$, without v. Now because of q-sparsity, there are at most $qr(v)$ nodes in S, and then $\sum_S r(u) \leq qa_S r(v)$, where a_S is the average running time in S. Then $r(v) \leq q.a_S$.

4.1 Applications

Thanks to the lemma, establishing a lower bound for node-averaged complexity of languages in LCL^* for very sparse graphs, boils down to show a simpler fact. It is sufficient to prove that a constant fraction of the nodes are close enough to nodes with running times similar to the running time of the slowest node. We illustrate this type of proof with LCL problems on cycles. It is known that for such problems, the slowest node complexity can only take three forms: $O(1)$, $\Theta(\log^* n)$ or $\Theta(n)$. See for example [5] for a recent presentation of this classification.[3] We prove that the landscape is the same for ordinary nodes.

Theorem 2 *For* LCL *on cycles, the node-averaged complexity has the same asymptotic classification as the slowest node complexity.*

Proof. Remember that the slowest node complexity is an upper bound on the average node complexity. Thereafter, it is sufficient to only prove the two lower bounds: $\Omega(\log^* n)$ and $\Omega(n)$.

Let us first focus on the case $\Theta(n)$. In this case, there exists a constant γ, such that on every cycle on n nodes, for large enough n, at least one node v has a running time at least γn. As we consider a lower bound, we can assume that the algorithm is minimal. As lemma 1 applies, the average complexity in the $(\gamma n/2)$-neighbourhood of v is at least $\alpha\gamma n - \beta$, where α and β are constants. Thereafter, the sum of the running time, in the $(\gamma n/2)$-neighbourhood of v is bounded from below by $\alpha\gamma^2 n^2 - \beta\gamma n$ (or n^2). Hence the average complexity for the whole cycle is in $\Omega(n)$.

Now let us now consider the case of classical complexity $\Theta(\log^* n)$. Consider any minimal algorithm A for the language $\mathcal{L}$ we consider. Again, let γ be a constant, such that the slowest node complexity is at least $\gamma \log^* n$, for large enough n. There exists a ring R_1 on n nodes, such that a node v_1 has running time $r_1 \geq \gamma \log^*(n)$. Then let H_1 be the graph that is composed of only the r_1-neighbourhood of v_1, and let I_1 be the set of identifiers of this segment. Now consider another ring R_2 on n nodes, with no identifiers from I_1, such that there exists a node v_2 with running time $r_2 \geq \gamma \log^*(n)$. Let H_2 be H_1 concatenated with the r_1-neighbourhood of v_1. Note that because no identifier from I_1 is present in R_2, H_2 has distinct identifiers. This operation can be repeated, until H_k has more than $n/2$ nodes. Let H be H_k, completed in an arbitrary way to get a full cycle of size n with distinct identifiers.

Note that as we performed the operation at most a linear number of times, the fact of removing some identifiers at each step is harmless as the identifier space is supposed to be polynomially large. Also note that the $\Theta(\log^* n)$ lower bound for the classical complexity is not affected by the constraints we add on the identifier space. This is because the lower bounds proofs do not rely on the particular shape of this space, which can even be assumed to be $\{1...n\}$[16].

We claim that on this cycle H with this ID assignment, the average node complexity is $\delta \log^*(n)$ for some constant δ. Indeed the nodes v_i, for i ranging from 1 to k, have the same neighbourhood as in R_i respectively, thus have running times r_i respectively. Then using lemma 1 we get for every i the nodes at distance at most $r_i/2$ from v_i have running time at least $\alpha r_i - \beta$. And by construction there is a constant fraction of the nodes of H that are in this case. As for every i, $r_i \geq \gamma \log^*(n)$, the node-averaged complexity is in $\Omega(\log^* n)$. □

[3] Even if not stated explicitly in [5], this classification also holds in the context where no knowledge of n is assumed. This is because the $\Theta(\log^* n)$ bound relies on the construction of a maximal independent set, and that MIS is a problem for which the construction of [15] works.

This "extract and glue" technique works on other class of graphs, and similar bounds can thus be achieved. Nevertheless it is not true that, for any LCL problem and any graph class, the classic complexity is the same as the node-averaged complexity, as the following proposition shows.

Proposition 3. *There exists a graph class $\mathcal{C}$ for which 3-colouring has slowest node complexity $\Omega(\log^* n)$ but node-averaged complexity in $O(1)$.*

Proof. Consider first the following construction. Start with a path of even length k, and index the nodes along the path from v_1 to v_k. Create three new nodes and link them to the node v_k. Now for the nodes v_i with $1 < i < k$, if the index i is even, then add a node v'_i and the edge (v_i, v'_i). We call this construction a *short leg*. If the index i is odd, add two nodes v'_i and v''_i, and two edges (v_i, v'_i) and (v'_i, v''_i). This is a *long leg*. For both construction, the node v_i is called the *basis* of the leg. Let us call such a graph an even-odd caterpillar. Now the graphs of $\mathcal{C}$ are the ones that can be constructed the following way: take an even-odd caterpillar based on a path of length k, and a cycle of length $\alpha \log^* k$ (where α is a large enough constant), and add an edge between an arbitrary node of the cycle and v_1. See figure 2.

Every algorithm must colour the $\alpha \log^* k$ cycle, and as the size of the graph is linear in k, the identifiers space is polynomial in k. Then Linial's lower bound applies on the cycle, and the slowest node complexity is $\Omega(\log^* k)$.

Let us now show that there exists an algorithm with constant node-averaged complexity for 3-colouring in this graph class. Every node first gathers its 3-hop neighbourhood. From this view it can deduce its position in the graph, and its behaviour for the following steps. More precisely, for every node v:

- if all the (direct) neighbours of v have degree two, then it is a node of the cycle, then it runs the Cole-Vishkin procedure for 3-colouring a cycle [7]. It does not take into account the rest of the graph;
- if it is the basis of a short leg, or the middle of a long leg, then it takes colour 1;
- if it is the basis of a long leg, or has degree 1, then it takes colour 2;
- if it has degree four, then it is v_k and it takes colour 1;
- if it has degree two and both its neighbours have degree three, then it is v_1, and it waits until both its neighbours have output, and it outputs a non-conflicting colour.

See figure 2.

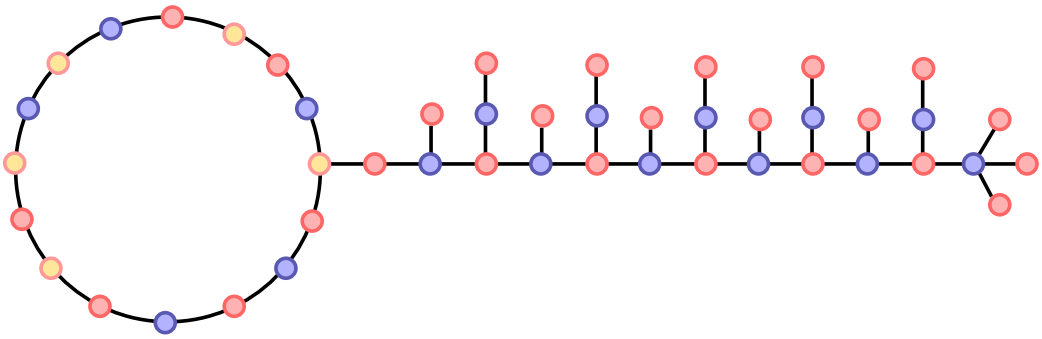

Fig. 2. The figure illustrates proof of proposition 3. It takes $O(\log^* n)$ rounds to 3-colour the cycle on the left, but it take constant time to colour the even-odd caterpillar on the right, as a 2-colouring is hard-coded in the structure of the graph. In this picture, colour 1 is blue, colour 2 is red, and colour 3 is yellow.

This algorithm uses at most $\log^* n$ rounds on the cycle and v_1, and constant time in the even-odd caterpillar. As the cyclic part has negligible size, the average node complexity is constant. □

5 Random ID assignments and randomized algorithms

We move on to the second topic of this paper, where the randomized aspects are considered. The standard definition of the complexity in the LOCAL model not only considers the slowest node, but also the worst-case distribution of the identifiers. In this section we investigate the impact of replacing this measure by the running time of the slowest node, on a random ID assignments. In other words, given a graph, we consider the average of the slowest-node running time over all possible ID assignments.

The main result is the equivalence between such measure, and the complexity of randomized algorithms. Here, the complexity of a randomized algorithm is the expectancy of the number of rounds before every node finishes. Note that the two concepts have similar flavour, but are distinct On one hand, the random inputs of a randomized algorithm are independent, while in a random ID assignment, the identifiers are not independent. On the other hand, the IDs are distinct, while the random inputs can be equal. On a high level, the equivalence is similar to Yao's principle [26], that relates the performance of a randomized algorithm on a worst-case instance, and the complexity of a deterministic algorithm on a random instance. Also note that in the literature, the usual complexity of randomized algorithms is not the one we consider, but the time needed to output a correct solution with high probability. That is, Monte-Carlo algorithms are considered instead of Las Vegas algorithms. We discuss briefly this point at the end of the section.

For the following theorem, randomized algorithms are given random strings of size $O(\log n)$, and not infinite such strings. This hypothesis is not excessive as most algorithm use a small amount of randomness. For example the celebrated MIS algorithm of [17] for bounded degree graphs, can be described as using random strings of size bounded by $O(\log n)$.

Theorem 3 *Given a problem, the expected slowest-node complexity of randomized algorithms, is equal to the expected deterministic slowest-node complexity on identifier assignment taken uniformly at random.*

Proof. It is part of the folklore that randomized algorithms do not need identifiers: they can generate such IDs with high probability by taking a integer in a cubic range uniformly at random. Here a probability of success equal to $1 - 1/n$ would be slightly too weak, so we make it $1 - 1/n^2$.

Lemma 2. *If n numbers are taken independently uniformly at random between 1 and n^4, these numbers are pairwise distinct with probability $1 - 1/n^2$.*

Proof. The probability of two fixed numbers being equal is $1/n^4$. Then by union bound, the probability that a pair of numbers have the same value is bounded by the number of such pairs $n(n-1)/2$ multiplied by the former probability. Then the probability of collision is bounded by $1/n^2$, thus with probability $1 - 1/n^2$ the numbers are pairwise distinct. □

Remember that a randomized algorithm can be formalized as a deterministic algorithm having an auxiliary input, this input being a large enough random number. We consider an algorithm A with an auxiliary input that can either be the ID or the random bits, and show that with high probability the behaviour is the same.

As stated in proposition 2, taking independently and uniformly at random n numbers from $[n^4]$ provides a list of distinct numbers with probability $1-1/n^2$. Also when this sampling succeeds, that is when the numbers are distinct, the outcome is uniform among all distinct identifiers assignments, because the identifiers are taken independently uniformly at random.

Let D be a deterministic algorithm, and let c be its average slowest-node complexity on identifier assignments taken uniformly at random. Let R be a randomized algorithm, that first picks random numbers in $[n^4]$, and then runs D, until D stops or until the node basically sees whole graph, and in the last case it outputs a colour such that the colouring is correct. The algorithm R has probability at least $(1 - 1/n^2)$ to stop with D that has expected runtime c, and probability at most $1/n^2$ to stop after at most n rounds. Then the expected runtime is upper bounded by $(1 - 1/n^2)c + 1/n^2.n$ which asymptotically is c. Conversely, suppose that a randomized algorithm has expected complexity c. We claim that using the same algorithm using the identifier as random strings provides a deterministic algorithm with average complexity c. Suppose it is not the case. Then, the randomized algorithm must have complexity c when the numbers are distinct, and c' when they are non-distinct. The expected runtime is $(1 - 1/n^2)c + 1/n^2c'$, which is asymptotically c as c' can be assumed to be at most n, which is a contradiction.

Thus theorem 3 holds. □

A similar result can be obtained for the more classic context of Monte-Carlo algorithm. That is, when one considers the time before the nodes have stopped and output a proper solution with high probability, then the complexity of randomized algorithms and of deterministic algorithm on random identifiers are the same.

A related topic is to minimize the amount of randomness used by randomized algorithms. The amount of random bits necessary to perform a computation is usually not considered as a resource to be minimized in the LOCAL model. Whereas it is considered in centralized computing, see [23] for a precise example. Here, it is possible to do a small step in that direction, if we consider algorithms and languages that are local. In this case, it is not necessary to have all IDs of the graph that are different one from the other. In a local algorithm, the nodes see only a small neighbourhood of the graph, and thus only such neighbourhoods need to have distinct IDs. This is one of the ingredient of recent breakthroughs in the field, such as the speed-up theorem from [6] (see theorem 6 in the paper).

Let s be the maximum number of nodes that a node can see when it runs the local algorithm at hand. Then the following holds:

Proposition 4. *Taking uniformly at random numbers from $\left[n^2s^2\right]$ is sufficient to have locally distinct identifiers with high probability.*

Proof. Consider a ball of size s. The probability that two nodes of this ball have the same identifier is upper bounded by $s^2/(n^2s^2) = 1/n^2$. Then by union bound on all the centres of balls, one gets a probability of collision of $1/n$. □

5.1 Node-averaged randomized complexity

After considering an average on the nodes, and on the identifiers assignment separately, we consider both averages together. That is we consider the behaviour of an ordinary node on an ordinary ID assignment. In the light of the previous subsection, this is equivalent to consider node-averaged complexity of randomized algorithms. This new measure can be unexpectedly low, as we illustrate on the example 3-colouring.

Theorem 2 implies that the node-averaged complexity of 3-colouring of a cycle is $\Theta(\log^* n)$. It is also known that the randomized complexity is $\Theta(\log^* n)$, if one considers Monte-Carlo algorithms with probability of success greater than one half [20]. Then the expected running time is also in $\Theta(\log^* n)$. This contrasts with the following result.

Proposition 5. *For 3-colouring on a ring, the expected complexity of an ordinary node is constant.*

Proof. The algorithm we consider, consists in repeating a simple procedure. At each round every node that has not yet an output, take a colour at random among the colours that are still available. That is, it takes a colour that as not yet been output by a neighbour. Note that this is always possible, as the nodes have degree two, and choose among three colours. After the sampling, if there is no conflict, then the node outputs the colour. If there is a conflict, then the colour is forgotten, and the node continue to the next round. If the node outputs a colour, we say that it *succeeds*, otherwise it *fails*.

Given an arbitrary partial colouring obtained after some rounds, the probability that a fixed node succeeds is lower bounded by $\alpha = 5/12$. This number is obtained by case analysis. It corresponds to the case where, the current node has both neighbours without outputs, but both nodes at distance two with outputs, and these outputs are different. Let $\beta = 1 - \alpha$. Also, let V_k be the number of nodes that have not yet output after round k, with $V_0 = n$. The following holds by linearity of the expectation.

$$\mathbb{E}(|V_k| \mid |V_{k-1}|) = \sum_{v \in V_{k-1}} \mathbb{P}(v \text{ does } not \text{ stop at round } k) \leq \beta |V_{k-1}|$$

We can apply the previous inequality repeatedly, and get: $\mathbb{E}(V_k) \leq \beta^k n$. The number of nodes that stop at round k is precisely $V_k - V_{k-1}$, then the sum of the running times is:

$$\sum_k k(V_k - V_{k-1}) \leq \sum_k kV_k.$$

The expected sum of the running time is then upper bounded by $\sum_k k\beta^k n$. Then the node-averaged expected sum is $\sum_k k\beta^k$. As $\beta < 1$, $\sum_k k\beta^k$ is a constant, thus the expected complexity of an ordinary node in a random ID assignment is constant. □

Note that having a constant complexity when looking at a more local measure, is not particular to this example. For example in [12], the author designs an algorithm for maximal independent set that terminates after $O(\log deg(v) + \log(1/\epsilon))$ rounds, with probability at least $1 - \epsilon$, where $deg(v)$ is the degree of node v.

6 Conclusion and open questions

This paper introduces the notions of node-averaged and ID-averaged complexities. We think these measures are meaningful when analysing algorithm that do not have the knowledge of the size of the network, or in contexts where partial solutions are useful. Also, very local complexities, as the one of subsection 5.1 and the one advocated in [12], are natural measures that one would like to understand better. Our results illustrate that these complexities can have interesting behaviours. The natural next step is to generalize these results, to more problems, and larger graph classes.

7 Acknowledgements

I would like to thank Juho Hirvonen, Tuomo Lempiinen and Jukka Suomela for fruitful discussions, and Pierre Fraigniaud for both discussions, and help for the writing. I thank the reviewers for helpful comments, and Mohsen Ghaffari for pointing out that randomized node-averaged complexity could be considered.

Bibliography

[1] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of algorithms*, 7(4):567–583, 1986.

[2] Hagit Attiya and Jennifer Welch. *Distributed Computing: Fundamentals, Simulations, and Advanced Topics*. Wiley, 2004.

[3] Leonid Barenboim and Michael Elkin. Sublogarithmic distributed MIS algorithm for sparse graphs using nash-williams decomposition. *Distributed Computing*, 22(5-6):363–379, 2010. doi:10.1007/s00446-009-0088-2.

[4] Sebastian Brandt, Orr Fischer, Juho Hirvonen, Barbara Keller, Tuomo Lempiäinen, Joel Rybicki, Jukka Suomela, and Jara Uitto. A lower bound for the distributed lovász local lemma. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 479–488, 2016. doi:10.1145/2897518.2897570.

[5] Sebastian Brandt, Juho Hirvonen, Janne H. Korhonen, Tuomo Lempiäinen, Patric R. J. Östergård, Christopher Purcell, Joel Rybicki, Jukka Suomela, and Przemyslaw Uznanski. LCL problems on grids. *CoRR*, abs/1702.05456, 2017. arxiv:1702.05456.

[6] Yi-Jun Chang, Tsvi Kopelowitz, and Seth Pettie. An exponential separation between randomized and deterministic complexity in the LOCAL model. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 615–624, 2016. doi:10.1109/FOCS.2016.72.

[7] Richard Cole and Uzi Vishkin. Deterministic coin tossing with applications to optimal parallel list ranking. *Information and Control*, 70(1):32–53, 1986. doi:10.1016/S0019-9958(86)80023-7.

[8] Laurent Feuilloley. Brief announcement: Average complexity for the LOCAL model. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, PODC 2015, Donostia-San Sebastián, Spain, July 21 - 23, 2015*, pages 335–337, 2015. doi:10.1145/2767386.2767446.

[9] Laurent Feuilloley and Pierre Fraigniaud. Survey of distributed decision. *Bulletin of the EATCS*, 119, 2016. EATCS:The Distributed Computing Column by Stefan Schmid.

[10] Pierre Fraigniaud, Amos Korman, and David Peleg. Towards a complexity theory for local distributed computing. *J. ACM*, 60(5):35, 2013. doi:10.1145/2499228.

[11] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 369–376, 2014. doi:10.1145/2554797.2554831.

[12] Mohsen Ghaffari. An improved distributed algorithm for maximal independent set. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 270–277, 2016. doi:10.1137/1.9781611974331.ch20.

[13] Oded Goldreich. Introduction to testing graph properties. In *Property Testing - Current Research and Surveys*, pages 105–141, 2010. doi:10.1007/978-3-642-16367-8_7.

[14] David G. Harris, Johannes Schneider, and Hsin-Hao Su. Distributed $(\Delta+1)$-coloring in sublogarithmic rounds. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 465–478, 2016. doi:10.1145/2897518.2897533.

[15] Amos Korman, Jean-Sébastien Sereni, and Laurent Viennot. Toward more localized local algorithms: removing assumptions concerning global knowledge. *Distributed Computing*, 26(5-6):289–308, 2013. doi:10.1007/s00446-012-0174-8.

[16] Nathan Linial. Locality in distributed graph algorithms. *SIAM J. Comput.*, 21(1):193–201, 1992.

[17] Michael Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM J. Comput.*, 15(4):1036–1053, 1986. doi:10.1137/0215074.

[18] Nancy A. Lynch. *Distributed Algorithms*. Morgan Kaufmann, 1996. ISBN 1-55860-348-4.

[19] Topi Musto. Knowledge of degree bounds in local algorithms. Master's thesis, University of Helsinki, 2011.

[20] Moni Naor. A lower bound on probabilistic algorithms for distributive ring coloring. *SIAM J. Discrete Math.*, 4(3):409–412, 1991. doi:10.1137/0404036.

[21] Moni Naor and Larry J. Stockmeyer. What can be computed locally? *SIAM J. Comput.*, 24(6):1259–1277, 1995. doi:10.1137/S0097539793254571.

[22] David Peleg. *Distributed Computing: A Locality-Sensitive Approach*. SIAM, 2000.

[23] Seth Pettie and Vijaya Ramachandran. Minimizing randomness in minimum spanning tree, parallel connectivity, and set maxima algorithms. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA.*, pages 713–722, 2002. acm:545381.545477.

[24] Nicola Santoro. *Design and analysis of distributed algorithms*, volume 56. John Wiley & Sons, 2006.

[25] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. A000788.

[26] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pages 222–227, 1977. doi:10.1109/SFCS.1977.24.

A Proof of lemma 1

Lemma 3 (Local average lemma). *Let $\mathcal{L}$ be a language in LCL^* with verification radius t, and A be a minimal algorithm for $\mathcal{L}$. There exists two positive constants α and β, such that on any q-sparse graph, ID assignment, inputs, and node v, the average of the running times of the nodes at distance at most $r(v)/2$ from v, is at least $\alpha.r(v) - \beta$.*

Proof. From claim 4, for every k,

$$r(v) - 2k - 2t \leq \max_{u \in S(v,k,k+2t)} r(u).$$

The following inequality follows:

$$r(v) - 2k - 2t \leq \sum_{u \in S(v,k,k+2t)} r(u).$$

Let us sum the inequality above, for k ranging from 1 to $r(v)/2 - 2t$. We assume without loss of generality that t and $r(v)$ are positive. The left-hand term is then:

$$\sum_{k=1}^{r(v)/2-2t} (r(v) - 2k - 2t) = \frac{r(v)^2}{4} - tr(v) + \frac{r(v)}{2} - 2t \geq \frac{r(v)^2}{4} - 3tr(v).$$

The right-hand term is:

$$\sum_{k=1}^{r(v)/2-2t} \sum_{u \in S(v,k,k+2t)} r(u) \leq (2t+1) \times \sum_{u \in S(v,1,r(v)/2)} r(u).$$

Because of q-sparsity, the number of nodes in $S(v, 1, r(v)/2)$ is bounded by $(r(v)/2).q$, thus

$$(2t+1) \times \sum_{u \in S(v,1,r(v)/2)} r(u) \leq (2t+1) \frac{qr(v)}{2|S(v,1,r(v)/2)|} \sum_{u \in S(v,1,r(v)/2)} r(u).$$

Putting the pieces together, and simplifying the terms, we get:

$$\frac{r(v)^2}{4} - 3tr(v) \leq 2tqr(v) \frac{1}{|S(v,1,r(v)/2)|} \sum_{u \in S(v,1,r(v)/2)} r(u).$$

Dividing by $r(v)$, and defining $\alpha = \frac{1}{8tq}$ and $\beta = \frac{3t}{2q}$, we get:

$$\alpha.r(v) - \beta \leq \frac{1}{|S(v,1,r(v)/2)|} \sum_{u \in S(v,1,r(v)/2)} r_u$$

Which is the desired formula, as the right-hand term is the node-averaged complexity, and as t and q are constants. □