International Journal of Parallel, Emergent and Distributed Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gpaa20

Distributed maximal independent set computation driven by finite-state dynamics

Eric Goles, Laura Leal, Pedro Montealegre, Ivan Rapaport & Martín Ríos-Wilson

To cite this article: Eric Goles, Laura Leal, Pedro Montealegre, Ivan Rapaport & Martín Ríos-Wilson (2023) Distributed maximal independent set computation driven by finite-state dynamics, International Journal of Parallel, Emergent and Distributed Systems, 38:1, 85-97, DOI: 10.1080/17445760.2022.2153248

To link to this article: https://doi.org/10.1080/17445760.2022.2153248

Published online: 06 Dec 2022.

Submit your article to this journal

Article views: 53

View related articles

View Crossmark data

CrossMark

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gpaa20

INTERNATIONAL JOURNAL OF PARALLEL, EMERGENT AND DISTRIBUTED SYSTEMS
2023, VOL. 38, NO. 1, 85–97
https://doi.org/10.1080/17445760.2022.2153248

Taylor & Francis
Taylor & Francis Group

Check for updates

Distributed maximal independent set computation driven by finite-state dynamics

Eric Goles[a], Laura Leal[b], Pedro Montealegre[a], Ivan Rapaport[c] and Martín Ríos-Wilson[a]

[a]Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Penalolen, Chile; [b]Departamento de Ingeniería Matemática, Universidad de Chile, Santiago de Chile, Chile; [c]DIM-CMM (UMI 2807 CNRS), Universidad de Chile, Santiago de Chile, Chile

ABSTRACT

A Maximal Independent Set (MIS) is an inclusion maximal set of pairwise non-adjacent vertices. The computation of an MIS is one of the core problems in distributed computing. In this article, we introduce and analyze a finite-state distributed randomized algorithm for computing a Maximal Independent Set (MIS) on arbitrary undirected graphs. Our algorithm is self-stabilizing (reaches a correct output on any initial configuration) and can be implemented on systems with very scarce conditions. We analyze the convergence time of the proposed algorithm, showing that in many cases the algorithm converges in logarithmic time with high probability.

ARTICLE HISTORY
Received 15 November 2022
Accepted 25 November 2022

KEYWORDS
Maximal independent set; local algorithms; distributed algorithms; randomized algorithms

1. Introduction

The Maximal Independent Set problem (MIS) is one of the main problems in distributed computing. In its simplest version, it consists in finding an inclusion-maximal set of pairwise non-adjacent vertices on an undirected graph. The MIS corresponds to a specific case of a wide problem in distributed graph algorithms, known as symmetry breaking. When a distributed algorithm is executed, the nodes of a distributed system are assumed to be in the same state, but in the successive time-steps the nodes are expected to play different roles, hence breaking the symmetry.

The MIS has a trivial solution in the classical sequential setting, where a greedy algorithm sequentially picks an arbitrary vertex, includes it in the maximal independent set, and removes that vertex together with all its neighbors. In the 80's, Karp and Wigderson [1] mentioned that the MIS is an interesting problem in non-centralized computation. Soon after that, Luby [2] and Alon, Babai, and Itai [3] presented simple distributed randomized algorithms solving MIS in $\mathcal{O}(\log n)$ time. Since then, this problem has been studied extensively in the distributed setting. In the LOCAL model, the fastest deterministic MIS algorithms for general graphs run in $\mathcal{O}(\log^5 n)$ [4], and $\mathcal{O}(\Delta + \log^* n)$ time [5]. Ghaffari [6] also obtained a $\mathcal{O}(\log \Delta) + 2^{\mathcal{O}(\sqrt{\log\log n})}$ time randomized algorithm on general graphs, and a $\mathcal{O}(\log a + \sqrt{\log n})$ time randomized algorithm for graphs of arboricity a. With respect to lower bounds, Linial [7] proved that computing an MIS on an n-cycle requires time $\Omega(\log^* n)$. Moreover, Kuhn et al. [8] showed a $\Omega(\sqrt{\log n})$ lower-bound on the round complexity on general graphs.

Another branch of research regarding the MIS problem in the distributed setting consists in considering models with limited resources. One example is the *beeping model* [9], where the nodes are limited to an extremely harsh system of communication. On each round, a node can either broadcast

CONTACT Pedro Montealegre p.montealegre@uai.cl Diagonal las Torres 2640, Peñalolén, Santiago, Chile

© 2022 Informa UK Limited, trading as Taylor & Francis Group

a signal (a beep) or hear whether a neighbor emitted a beeping signal, but noes it is not capable of distinguishing the number nor the sources of the beeping signals it receives. In this model, Afek et al. [9] showed that an MIS can be computed in time $\mathcal{O}(\log^3 n)$ when the nodes have to know the size of the graph n and have poly$(\log n)$ sized memory. The *stone-age model* is another relevant model with limited resources, where the memory of each node is limited to a constant not depending on the size of the graph. In this model, Emek and Wattenhofer [10] give an MIS algorithm with a running time of $\mathcal{O}(\log^2 n)$. Interestingly, this algorithm requires that the nodes start in a particular initial state in order to be capable of performing correct computation.

An algorithm is called *self-stabilizing* [11] if it can reach a correct output starting from any initial state. The motivation for this kind of algorithm is the capacity of distributed systems to self-repair a faulty configuration when one of the parties crashes. For instance, consider a maximal independent set, where one of the nodes crashes. It is possible that the rest of the nodes do not form a maximal independent set in the remaining graph (for instance, consider an MIS in a complete graph where the unique marked node crashes). In that context, a self-stabilizing MIS algorithm should be able to reach an MIS for any initial state configuration of the nodes.

In a keynote talk of SIROCCO 2022 [12], George Giakkoupis presented two extremely simple randomized algorithms for MIS. His algorithms have two interesting properties: they are self-stabilizing and they require only two or three states. Despite this simplicity, the algorithm has hardly been studied before, and its convergence time is not been settled yet. In this article, we propose a variant of the algorithm of [12] and study its convergence time both numerically and analytically.

Notation: For a positive integer k, we denote by $[k]$ the set $\{1,\dots,k\}$. Also, for a set S, we denote by $x \in_U S$ the process of taking an element of S uniformly at random. On inputs of size n, we say that an event occurs with high probability if it occurs with probability greater or equal than $1-1/n$.

The dynamics. Let $G=(V,E)$ be a simple finite undirected graph and $k \geq 2$ an integer. A *configuration* is a function that assigns to each node a *state* in $[k]$. Formally, a configuration is given by a function $x: V \to \{0\} \cup [k]$. For each node $u \in V$ we denote x_u the state of u on configuration x. The nodes in a state different than zero are called *marked nodes*, while the nodes in state 0 are called *unmarked*.

We say that a configuration *represents an independent set of* G when no edge has both endpoints in a state different than 0. Formally,

$$\forall \{u,v\} \in E, \quad x_u \cdot x_v = 0$$

Additionally, we say that a configuration *represents a maximal independent set of* G when it represents an independent set and, in addition, no edge has both endpoints in state 0. Formally,

$$\forall u \in V, \quad x_u \neq 0 \vee (\exists v \in N(u), x_u \neq 0)\,.$$

Let us consider the following stochastic dynamic over graph configurations. Given a configuration x, the next configuration x' is computed synchronously according to a rule that we denote MIS-Dynamics: Synchronously, the new state x'_u of $u \in V$ is computed as follows:

(i) If $x_u = 0$ and $\forall\, v \in N(u), x_v = 0$, then $x'_u \in_U [k]$.
(ii) If $x_u \neq 0$ and $\exists v \in N(u), x_u = x_v$ and $\forall\, v \in N(u), x_v \leq x_u$, then $x'_u \in_U [k]$.
(iii) If $x_u \neq 0$ and $\exists v \in N(u), x_v > x_u$, then $x'_u = 0$.
(iv) $x'_u = x_u$ otherwise.

In the special case of two-state configurations $x: V \to \{0,1\}$, we consider the following stochastic dynamics, that we call 2-MIS-Dynamics: Synchronously, we compute for each $u \in V$ the new state x'_u as follows:

(i) If $x_u = 0$ and $\forall\, v \in N(u), x_v = 0$, then $x_u = 1$.
(ii) If $x_u \neq 0$ and $\exists v \in N(u), x_v \neq 0$, then $x_u \in_U \{0, 1\}$.
(iii) $x'_u = x_u$ otherwise.

Given a configuration x (also called initial configuration), the *trajectory* of x, denoted $\{x^t\}_{t\geq 0}$, is the random variable representing the evolution of the MIS-Dynamics where x^t is obtained from x^{t-1} for each $t > 1$, and $x^0 = x$.

We say that a configuration is a *fixed point* for the MIS-Dynamics if $x' = x$ with probability 1. Observe that a configuration x is a fixed point of the MIS-Dynamics if and only if x represents a maximal independent set. Indeed, let x be a fixed point of the MIS-Dynamics. We say that a node u is *stabilized* on x if one of the following conditions is satisfied:

(1) $x_u \neq 0$ and every $v \in N(u)$ satisfies $x_v = 0$,
(2) $x_u = 0$ and there exists a stabilized neighbor of u such that $x_u \neq 0$.

1.1. Our contribution

We study the convergence time of the MIS-Dynamics and 2-MIS-Dynamics both empirically and analytically.

We first simulate the dynamics on different graph classes, namely: Complete graphs, stars, graphs of Erdös-Renyi, random trees, and graphs of bounded degeneracy. On the one hand, we observe that the 2-MIS-Dynamics converges to a MIS in logarithmic time in expectation for all classes. We also observe that the convergence time tends to increase with the density of the input graph. In fact, for complete graphs, we the convergence time is $\Omega(\log^2(n))$ with non-negligible probability. On the other hand, we observe that the MIS-Dynamics converges to an MIS in logarithmic time, both in expectation and with high probability. Moreover, the computation time decreases as the number of states augments.

Then, we analytically study the convergence time of the MIS-Dynamics. First, we show that with high probability, the dynamic converges in time $\mathcal{O}(\alpha \log n)$, where α is the size of a maximum independent set of the input graph. Then, we extend our analysis to the 2-MIS-Dynamics, showing that with high probability it converges in time $\mathcal{O}(\alpha \log^2 n)$. Finally, we show that restricted to the class of d-degenerate graphs, the 2-MIS-Dynamics converges in time $\mathcal{O}(\log n)$ with high probability.

1.2. Structure of the article

We begin giving some background and preliminaries in Section 2. In Section 3, we report the results of our computational simulations. Then, in Section 4 we give bounds for the convergence time of the MIS-Dynamics and 2-MIS-Dynamics on arbitrary graphs. In Section 5, we study the 2-MIS-Dynamics on graphs of bounded degeneracy. We finish with a discussion in Section 6.

2. Preliminaries

In this article, all graphs are simple, finite, and undirected. Given a node v of a graph $G = (V, E)$, the *neighborhood of* v, denoted by $N(v)$, is the set of vertices adjacent to v. Formally $N(v) = \{u \in V : \{u, v\} \in E\}$. The degree of a node u, denoted $d(u)$ is the cardinality of $N(v)$. Given a set of nodes $U \subseteq V$, the *subgraph of G induced by U*, denoted $G[U]$, is the graph defined by the vertex set U, and all the edges in E with both endpoints in U. A graph is called *connected* if there is a path between every pair of vertices. A *connected component* of a graph G is an inclusion maximal connected set of vertices. A connected graph without cycles is called a *tree*. A graph where every node has degree d is called d-regular.

A set of nodes S is called an *independent set* if the graph induced by it has no edges. An inclusion-maximal independent set is simply called *maximal independent set*. The cardinality of an independent set of maximal cardinality is denoted $\alpha(G)$, and is called the *independence number* of G. The problem of computing $\alpha(G)$ is NP-Hard. The range of its value goes from 1 for the complete graph, to $n-1$ for the case of the star graph. There are a number of combinatorial lower bounds for $\alpha(G)$ with respect to some graph parameters. In this article, we make use of the following simple result.

Proposition 2.1 ([13]): *Let G be an n-node graph of maximum degree Δ. Then, $\alpha \geq n/\Delta$.*

A graph property is called an *hereditary property* if it is closed under taking induced subgraphs. An example of a hereditary property is bounded degeneracy. A graph G is called *d-degenerate* if every subgraph of G (including G itself) contains a vertex of degree at most d. Alternatively, a graph has degeneracy d if it can be decomposed successively removing vertices of degree at most d. In the following lemma, we now show that in a d-degenerate graph, for most vertices their degree is bounded by $4d-2$.

Lemma 2.2: *A connected graph of degeneracy d contains at least n/2 nodes of degree at most $4d-2$.*

Proof: First, observe that a graph of degeneracy d has at most dn edges. Let us call U the set of nodes of a degree greater or equal than $4d-1$. Then, for the hand-shaking lemma:

$$2dn = \sum_{v \in V} d(v) \geq |U|(4d-1) + (n - |U|).$$

The previous bound implies that $|U| \leq \frac{(2d-1)n}{4d-2} \leq n/2$. We deduce that the cardinality of the set vertices of degree at most $4d-2$ is at least $n/2$. ■

The graphs of Erdös-Rényi-Gilbert (in the following graphs of Erdös-Rényi for simplicity) are a randomized model of graphs where a graph is constructed by connecting labeled nodes, where each edge is included in the graph with probability p, independently from every other edge. The following result states that with a high probability the independence number of a graph of Erdös-Rényi is bounded by the logarithm of the number of nodes.

Proposition 2.3 ([14]): *For each $p \in (0, 1)$ and sufficiently large n, the independence number of a graph of Erdös-Rényi with parameters n and p is $\mathcal{O}(\log n)$ with high probability.*

3. Experimental results

In this section, we report the empirical analysis of the 2-MIS-Dynamics and the MIS-Dynamics.

In Figure 1 we show the results of a study of the 2-MIS-Dynamics and the MIS-Dynamics of k states with $k \in \{2, 3, 4, 9\}$ over complete graphs. We observe that on all cases the average convergence time has a logarithmic grow with respect to the size of the graph. We also observe that the average convergence time of the MIS-Dynamics decreases as the number of states k grows. The 2-MIS-Dynamics has a better convergence time than the three-state MIS-Dynamics for small values of n, but for complete graphs on more than 16 nodes the 2-MIS-Dynamics has a larger average convergence time than all the MIS-Dynamics for every $k \geq 2$.

We then extend the analysis to other graph classes. In Figure 2 we report our results for complete, star, random trees (Barabasi-Albert graphs with $m = 1$) and 4-regular graphs. In all cases, the results are roughly the same than those we obtained for the complete graphs. That is to say, the average convergence time $\mathcal{O}(\log n)$ and it decreases as we increase the number of states in the dynamics. We complement our analysis with a study of the worst-case convergence time (i.e. the maximum convergence time over all initial configurations). In Figure 3 we report our results, which were obtained in the same way as Figure 2, except that we take the maximum over all observed convergence times for each graph size.

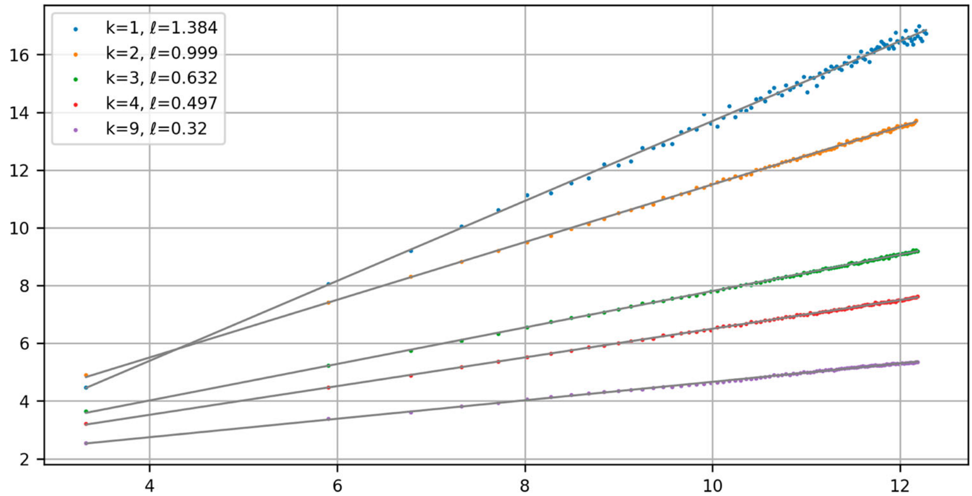

Figure 1. Plot of the average convergence time for the 2-MIS-Dynamics and MIS-Dynamics. On the x-axis we have the number of nodes in $\log_2$ scale, while in the y-axis are given the number of iterations. The different dynamics are represented with different colors. The line labeled $k = 1$ corresponds to the 2-MIS-Dynamics while the other lines represent the MIS-Dynamics with $k \in \{2, 3, 4, 9\}$ states. Each line represents the average convergence time of 5000 initial configurations picked uniformly at random. We also give the slope ℓ of the least squares regression line corresponding to the points.

Interestingly, in the worst-case analysis on complete graphs, we observe that the convergence time of the 2-MIS-Dynamics behaves significantly different than on the rest of the classes, and also with respect to the average case. In fact, our results suggest that in the worst case the convergence time of the 2-MIS-Dynamics is $\mathcal{O}(\log^2 n)$. This fact is indeed verified in the next section.

Finally, we explore how the density of the input graph influences the average and worst-case convergence-time of the 2-MIS-Dynamics and MIS-Dynamics. To do so, we fixed a number of nodes to $n = 500$, and simulate the dynamics in two families of graphs. First, on Erdös-Rényi graphs for different probabilities p. Second, in d-degenerate graphs for different values of d. Our results are reported in Figure 4. We observe that for the 2-MIS-Dynamics and the MIS-Dynamics, the maximum average convergence times are found on graphs with a mean degree around $n/2$. With respect to the worst-case convergence-time, we observe a similar behavior for the MIS-Dynamics, while for the 2-MIS-Dynamics tends to increase with the density.

4. A bound on the convergence time of the MIS-Dynamics on arbitrary graphs

In this section, we show that for every initial configuration, the MIS-Dynamics converges to a fixed point (hence a configuration that represents a maximal independent set) in $\mathcal{O}(\alpha \cdot \log n)$ time-steps on average. For simplicity, our analysis focuses on the case where $k = 2$, as it can be trivially generalized to the case where $k > 2$.

For a configuration x, we define the following *energy functional*:

$$S(x) = \sum_{\{u,v\} \in E} \delta(x_u, x_v), \quad \text{where } \delta(a,b) = \begin{cases} 1 & \text{if } a = b \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let us fix an initial configuration x^0 and a time-step t. We denote by S^t the random variable that equals $S(x^t)$. A configuration x^t such that $S^t = 0$ is called a *zero-energy configuration*. Observe that all configurations that represent an independent set are zero-energy configurations (but the converse is not true). We call E^t_m the random variable representing the set of edges of G having on time-step t with

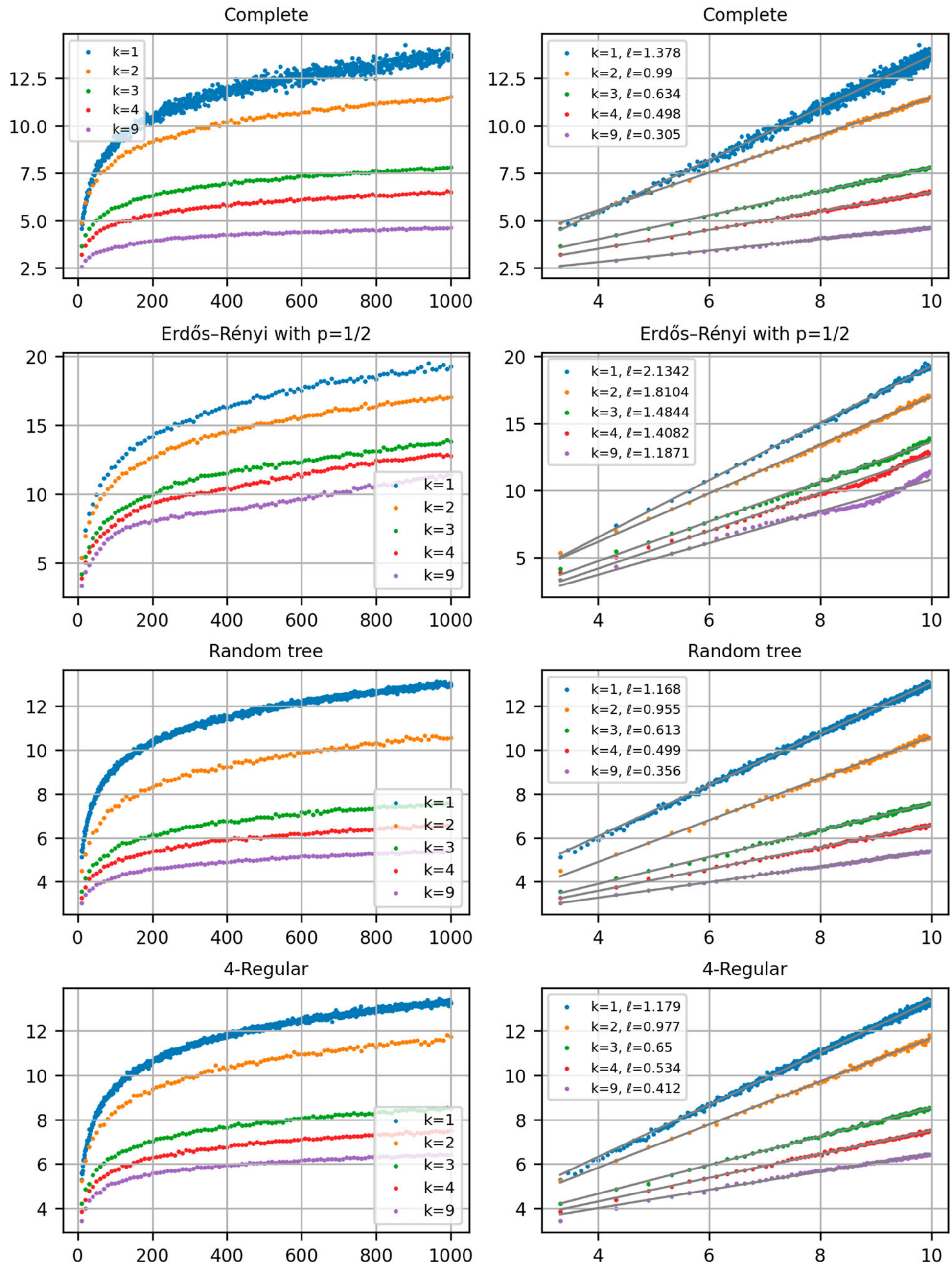

Figure 2. Plots of the average convergence time for the 2-MIS-Dynamics and MIS-Dynamics for different graph classes. On the x-axis we have the number of nodes, while on the y-axis are given the number of iterations. In the left column, we have the linear scale, while in the right column we give the x-axis in $\log_2$ scale. The different dynamics are represented in different colors. The line labeled $k = 1$ corresponds to the 2-MIS-Dynamics while the other lines represent the MIS-Dynamics with $k \in \{2, 3, 4, 9\}$ states. Each line represents the average convergence time of 1000 initial configurations picked uniformly at random. We also give the slope ℓ of the least squares regression line corresponding to the points.

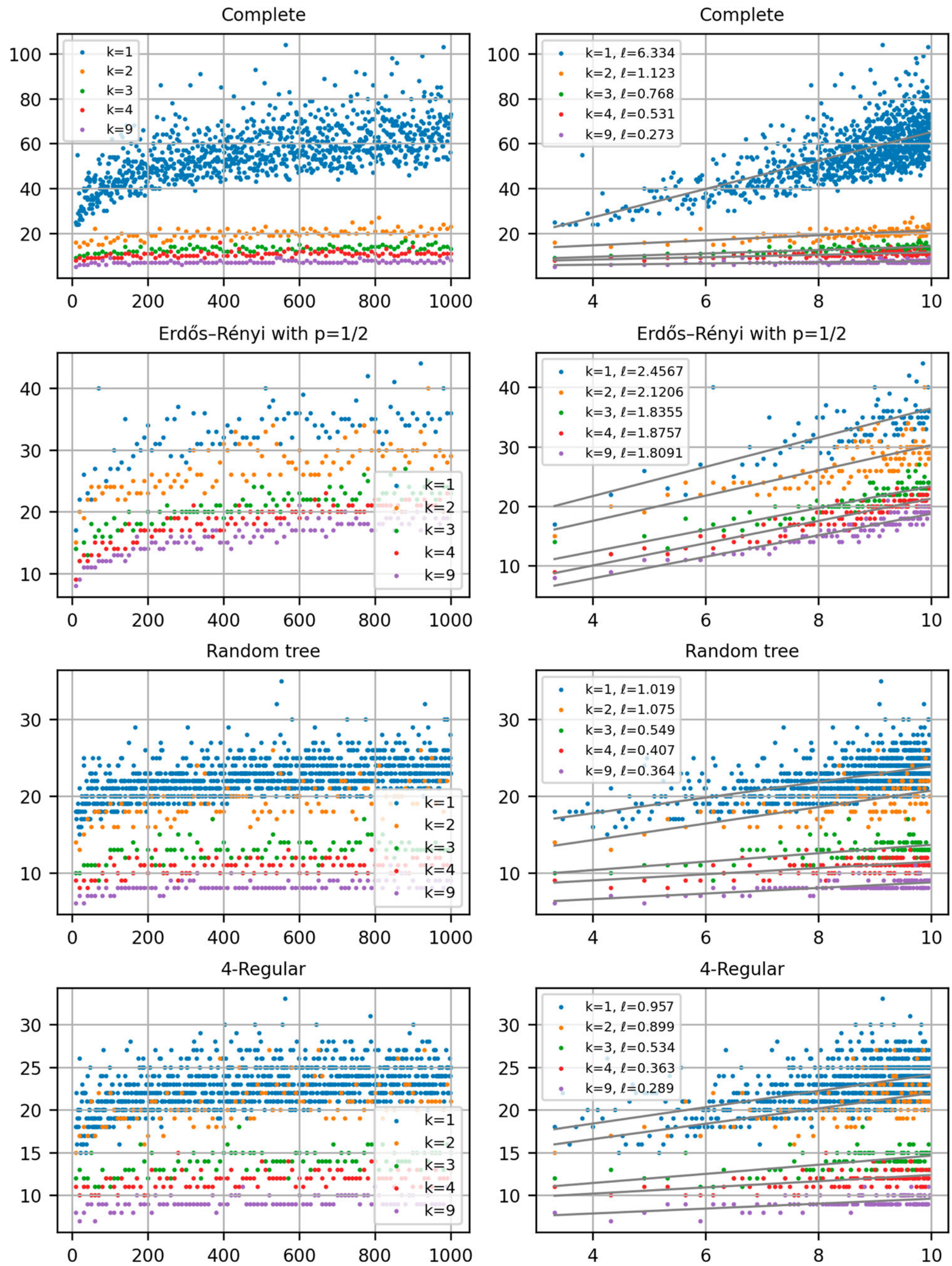

Figure 3. Plots of the worst-case convergence time for the 2-MIS-Dynamics and MIS-Dynamics for different graph classes. On the *x*-axis we have the number of nodes, while on the *y*-axis are given the number of iterations. In the left column, we have the linear scale, while in the right column we give the *x*-axis in $\log_2$ scale. The different dynamics are represented in different colors. The line labeled $k = 1$ corresponds to the 2-MIS-Dynamics while the other lines represent the MIS-Dynamics with $k \in \{2, 3, 4, 9\}$ states. Each line represents the average convergence time of 1000 initial configurations picked uniformly at random. We also give the slope ℓ of the least squares regression line corresponding to the points.

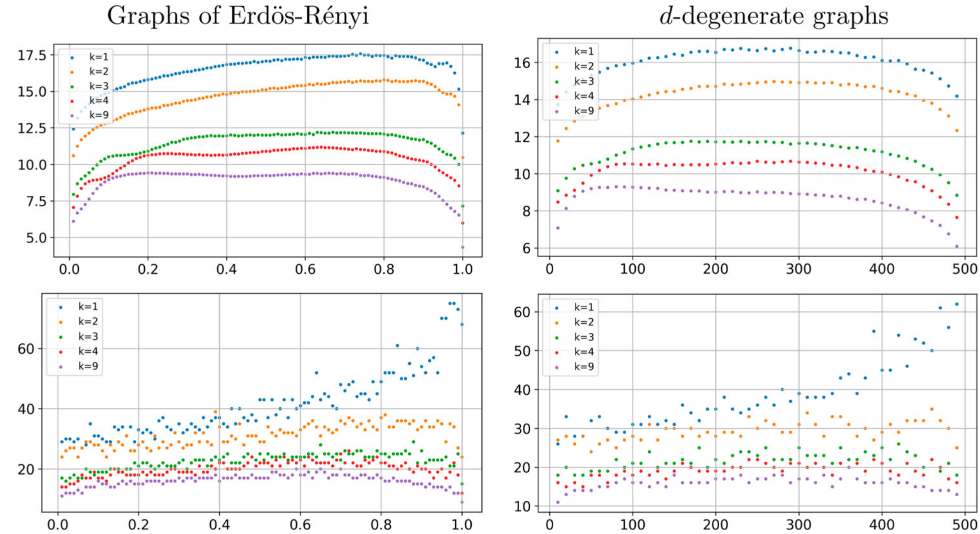

Figure 4. Plots of the influence of the edge density on convergence time for the 2-MIS-Dynamics and MIS-Dynamics. On the four plots we represent the convergence time of the 2-MIS-Dynamics ($k = 1$) and the MIS-Dynamics on $k \in \{2, 3, 4, 9\}$ states on graphs with $n = 500$ nodes. In the left column we represent the behavior of the convergence-times on Erdös-Renyi graphs, where the x-axis represents the probability p of adding the edge between two vertices. In the right column we represent the behavior of the convergence-times on d-degenerate graphs. In that case, the x-axis represents the different values of the degeneracy d. In the top row the y-axis represent the average convergence times of the corresponding dynamics, while the bottom row we show the worst-case convergence times. Each point is computed taking 10,000 random initial conditions, each one on a different graph picked uniformly at random.

both endpoints marked with the same value. Formally,

$$E_m^t = \{\{u, v\} \in E : x_u^t = x_v^t \neq 0\}$$

Observe that $S^t = |E_m^t|$. We also denote by A^t and B^t the random variables representing the following sets of edges:

$$A^t = \{u \in V : x_u^{t-1} = 0 \text{ and } x_u^t \neq 0\}$$
$$B^t = \{u \in V : x_u^{t-1} \neq 0 \text{ and } x_u^t \neq 0\}$$

We say that a time-step t such that $A^t \neq \emptyset$ is a *marking time step*.

Lemma 4.1: *Let $\{x^t\}_{t\geq 0}$ be a trajectory and let t_0 be a marking time-step. Let C be a connected component of the graph induced by A^{t_0}. Then C contains a node that is never unmarked. Formally, there is $u \in C$ satisfying $x_u^t \neq 0$ for every $t \geq t_0$.*

Proof: Let us suppose by contradiction that there is a time step $t_1 > t_0$ where every node in C has visited state 0 at least once on some configuration of $\{x^{t_0}, \ldots, x^{t_1}\}$. From all possible choices of t_1, we pick the minimum one. Let v be a node such that $x_v^t \neq 0$ for every $t_0 \leq t < t_1$ and $x_v^{t_1} = 0$. Then, the only possibility is that $x_v^{t_1-1} = 1$ and that v has a neighbor $u \in N(v)$ such that $x_u^{t_1-1} = 2$. Observe that $x_u^{t_1} \neq 0$ (it impossible that a node in state 2 switches to 0 in the next time-step). Then, by definition of T, there must exist a time-step $t_0 < t < t_1 - 1$ such that $x_u^t = 0$. However, we are assuming that v is in a state different than 0 on all the configurations of that interval. Therefore, it is impossible that u switches to a state different than 0 on a time-step between t and $t_1 - 1$. This contradicts the choice of u. We deduce that at least one node of C is never unmarked. ∎

Lemma 4.2: *The trajectory of every configuration has at most α marking time steps.*

Proof: Now let us call T the set of time-steps t such that $A^t \neq \emptyset$. Given a time-step $t \in T$, we know by Lemma 4.1 that there is a node $v(t) \in V$ that is never unmarked after time-step t. We have that the set $\{v(t)\}_{t \in T}$ forms an independent set of G. Indeed, let us pick two different time-steps $t_1, t_2 \in T$ such that $t_1 < t_2$. By definition of A^{t_2} we have that $x_{v(t_2)}^{t_2-1} = 0$ and $x_{v(t_2)}^{t_2} \neq 0$. Since $t_1 < t_2$ we have that $x_{v(t_1)}^{t_2-1} \neq 0$. We deduce that $v(t_1)$ and $v(t_2)$ cannot be adjacent. We conclude that $|T| \leq \alpha(G)$. ■

Lemma 4.3: *Let $t > 0$ be a non-marking time-step. Then, $\mathbb{E}(S^t) \leq S^{t-1}/2$.*

Proof: First, we observe that, since $A^t = \emptyset$, every edge that contributes to S^t also contributes to S^{t-1}. Formally, $E_m^t \subseteq E_m^{t-1}$. Indeed, let $e = \{u, v\}$ be an edge contained in E_m^t. Clearly u and v belong to B^t, as A^t is empty. Then, $x_u^{t-1} \neq 0$ and $x_v^{t-1} \neq 0$. Moreover, $x_u^{t-1} = x_v^{t-1}$, as otherwise at least one of the endpoints would be 0 on time-step t. Therefore e is also contained in E_m^{t-1}. Now observe that for every $e \in E_m^{t-1}$, the probability that $e \in E_m^t$ is fewer or equal than the probability that both endpoints of e remain marked and choose the same state, which is 1/2. We deduce that $\mathbb{E}(S^t) \leq S^{t-1}/2$. ■

Lemma 4.4: *Let t be a zero energy time-step. Then at least one of the following holds:*

- *$t + 1$ is marking,*
- *$t + 2$ is marking,*
- *x^{t+1} is a fixed point*

Proof: Let t be a time-step satisfying that $S^t = 0$, and let us assume that $A^{t+1} = \emptyset$. We show first that x^{t+1} must represent an independent set. Observe that $A^{t+1} = \emptyset$ implies $S^{t+1} = 0$. Then, in x^t or x^{t+1} no edge has both endpoints in state different than 0. Suppose that there exist an edge $\{u, v\}$ such that $x_u^{t+1} = 1$ and $x_v^{t+2} = 2$. Then necessarily $x_u^t = x_v^t = 0$, which contradicts the assumption of $A^{t+1} = \emptyset$. We deduce x^{t+1} represents an independent set. Now suppose that x^{t+1} does not represent a maximal independent set, that is to say, there is a node w such that $x_v^{t+1} = 0$ for all $v \in N(w) \cup \{w\}$. Then necessarily $x_w^{t+2} = 1$, implying that $A^{t+2} \neq \emptyset$. ■

Now we are ready to prove the main result of this section.

Theorem 4.5: *For every initial configuration, the MIS-Dynamics converges to a configuration representing a maximal independent set in $\mathcal{O}(\alpha \cdot \log n)$ time-steps with high probability.*

Proof: Let x^0 be an arbitrary initial configuration. From Lemma 4.3 we know that, with high probability, in at most $\mathcal{O}(\log n)$ time-steps the trajectory visits a zero-energy configuration or a marking time-step. Let $t > 0$ be a marking time-step or a time-step where the trajectory visits a zero-energy configuration. Combined with Lemmas 4.4 and 4.1 we know that either x^{t+1} is a fixed point (so we are done), or either t, $t + 1$ or $t + 2$ is marking. In the later cases, we repeat the analysis by taking x^t, x^{t+1} or x^{t+2} as the initial configuration. By Lemma 4.2 we know that the number of repetitions is bounded by α. We deduce that with high probability the dynamic converges to a configuration representing a maximal independent set in $\mathcal{O}(\alpha \cdot \log n)$ time-steps. ■

4.1. The 2-MIS-Dynamics on arbitrary graphs

We now adapt our result to the 2-MIS-Dynamics. The difference is found in the result given on Lemma 4.1. In fact, that lemma does not hold for the 2-MIS-Dynamics. Indeed, on every marking time-step t, there exists a non-zero probability that all nodes of A^t become unmarked. Nevertheless, the next lemma shows that at least one marked node is stabilized with a constant probability.

Lemma 4.6: *Let $\{x^t\}_{t\geq 0}$ be a trajectory and let t_0 be a marking time-step. Let C be a connected component of the graph induced by A^{t_0}. Let $\mathcal{E}$ the event where at least one node of C is never unmarked. Formally,*

$$\mathcal{E} : \exists\, u \in C, \quad \forall\, t \geq 0,\ x_u^t \neq 0.$$

Then, there exists an absolute constant $c \in (0,1)$ satisfying that $Pr(\mathcal{E}) \geq c$.

Proof: Let m be the number of edges in $G[C]$. If $m = 0$ then C consists on an isolated marked node, which is stabilized with probability 1. If $0 < m \leq 144$ we have that $Pr(\mathcal{E}) \geq 2^{-144}$. Indeed, when $m \leq 144$ we have that C contains at most 144 nodes. A lower-bound on $Pr(\mathcal{E})$ is the event where on one time-step all except one fixed node of C becomes 0. The probability of such an event is lower-bounded by 2^{-144}. In the following, we assume that $m > 144$.

Let us denote $e_1, \dots, e_m$ the edges of $G[C]$. For each $i \in [m]$ we denote by e_i^t the random variable that equals 1 if edge e_i is marked on all time-steps in $\{t_0, \dots, t_0 + t\}$. We also denote $m^t = \sum_{i\in[m]} e_i^t$. In words, m^t is the random variable representing the edges that have both endpoints marked on all time-steps in $\{t_0, \dots, t_0 + t\}$. Observe that for each $i, j \in [m]$, $\mathbb{E}(e_i^t) = 2^{-2t}$ and by linearity of the expectation, $\mathbb{E}(m^t) = 2^{-2t}m$.

We aim to bound the probability that the actual value of m^t has a large gap with respect to its expectation. By Chebyshev's inequality we have that

$$Pr(|m^t - \mathbb{E}(m^t)| > a) \leq \frac{\mathrm{Var}(m^t)}{a^2}.$$

By Bienyamé's identity, we know that

$$\mathrm{Var}(m^t) = \sum_{i\in[m]} \mathrm{Var}(e_i^t) + \sum_{i\in[m]} \sum_{j\in[m]\setminus\{i\}} \mathrm{Cov}(e_i^t, e_j^t)$$

where

$$\mathrm{Var}(e_i^t) = \mathbb{E}((e_i^t)^2) - (\mathbb{E}(e_i^t))^2 = 2^{-2t} - 2^{-4t}.$$

and

$$\mathrm{Cov}(e_i^t, e_j^t) = \mathbb{E}(e_i^t e_j^t) - \mathbb{E}(e_i^t)\mathbb{E}(e_i^t) = \begin{cases} 2^{-3t} - 2^{-4t} & \text{if } e_i \cap e_j \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

Then, if we denote by Δ the maximum degree of $G[C]$,

$$\mathrm{Var}(m^t) \leq m \cdot (2^{-2t} - 2^{-4t}) + m \cdot 2\Delta \cdot (2^{-3t} - 2^{-4t}).$$

Observe that $\Delta^2 \leq |C|\Delta \leq 2m$. Therefore,

$$\mathrm{Var}(m^t) \leq m \cdot (2^{-2t} - 2^{-4t}) + 2^{3/2} m^{3/2} \cdot (2^{-3t} - 2^{-4t}).$$

Now let us pick $k \leq m$ a variable to be fixed later. If we choose

$$\tau = \left\lfloor \frac{\log(m) - \log(k)}{2} \right\rfloor,$$

we have that $\mathbb{E}(m^\tau) \in [k, 2k]$ and

$$\begin{aligned} \mathrm{Var}(m^\tau) &\leq m \cdot \left(\frac{k}{m} - \left(\frac{k}{m}\right)^2\right) + 2^{3/2} m^{3/2} \cdot \left(\left(\frac{k}{m}\right)^{3/2} - \left(\frac{k}{m}\right)^2\right) \\ &= k + k\sqrt{8k} - \left(\frac{k^2}{m} - \frac{\sqrt{8}k^2}{\sqrt{m}}\right) \leq k + k\sqrt{8k} \end{aligned}$$

Then,

$$Pr(|m^\tau - \mathbb{E}(m^\tau)| > k/2) \leq \frac{4}{k} + \frac{4\sqrt{8}}{\sqrt{k}}.$$

If we pick $k = 144$ we obtain that $Pr(|m^\tau - \mathbb{E}(m^\tau)| > 72) < 0.98$. In other words, with probability greater than 0.02 we have that $m^\tau \in [\mathbb{E}(m^\tau) - 72, \mathbb{E}(m^\tau) + 72] \subseteq [72, 360]$. Conditioned to that event, we have that with probability greater or equal than 2^{-360} a fixed node u is stabilized on time-step $\tau + 1$.

We conclude that $Pr(\mathcal{E}) \geq c = 0.02 \cdot 2^{-360}$. ■

Using the previous lemma, we can show that, with high probability, there are $\mathcal{O}(\alpha \cdot \log n)$ marking time-steps on a 2-MIS-Dynamics.

Lemma 4.7: *For every initial configuration, with high probability, there are $\mathcal{O}(\alpha \cdot \log n)$ marking time-steps on the 2-MIS-Dynamics.*

Proof: From Lemma 4.6, we know that there exists a constant c such that at least one marked node is stabilized with a probability greater than c. Then, on $\mathcal{O}(\log n)$ marking time-steps at least one node is stabilized with high probability. We deduce that the trajectory of every initial configuration visits $\mathcal{O}(\alpha \cdot \log n)$ marking time-steps with high probability. ■

Observe that we can show results analogous to Lemmas 4.3 and 4.4 using exactly the same proofs. We deduce the main result of this subsection.

Theorem 4.8: *For every initial configuration, the 2-MIS-Dynamics converges to a configuration representing a maximal independent set in $\mathcal{O}(\alpha \cdot \log^2 n)$ time-steps with high probability.*

5. The 2-MIS-Dynamics on graphs of bounded degeneracy

In this section, we show that the 2-MIS-Dynamics converges to a configuration that represents a maximal independent set in time $\mathcal{O}(\log n)$ on average.

Let G be an arbitrary graph. We denote by $G_{\leq d}$ the sub-graph of G induced by the nodes of degree at most d and by α_d the size of a maximum independent set of $G_{\leq d}$.

Lemma 5.1: *Let G be a graph, let d be a positive integer, and let x be a configuration of G with no stabilized vertices. Then, on average, $\Omega(\frac{\alpha_d}{4^{d+1}})$ nodes are stabilized after two time-steps.*

Proof: For each $u \in V$ and $t > 0$, let us call $P(u, t)$ the probability that u is stabilized on time-step t. We claim that $P(u, 2) \geq 2^{-(2d_u+1)}$. Indeed, if $x_u = 1$ then $P(u, 1) \geq 2^{-(d_u+1)}$, hence $P(u, 2) \geq 2^{-(2d_u+1)}$. If $x_u = 0$ then, on time step $t = 1$ the probability that u and all its neighbors are unmarked is at least 2^{-d_u}. Hence $P(u, 2) \geq 2^{-(2d_u+1)}$.

Now let U be a maximum independent set of $G_{\leq d}$, and let W be the random variable representing the subset of U that is stabilized on time-step $t = 2$. Then, for each $u \in U, Pr(u \in W) \geq 2^{-(2d+1)}$. Hence

$$\mathbb{E}(|W|) \geq \frac{|U|}{2^{2d+1}} \geq \frac{\alpha_d}{4^{d+1}}.$$ ■

Theorem 5.2: *For every initial configuration over a d-degenerate graph, the 2-MIS-Dynamics converges to a configuration representing a maximal independent set in $\mathcal{O}(\log n)$ time-steps with high probability.*

Proof: Let G be an arbitrary n-node graph of degeneracy d, and let x be an arbitrary configuration of G. Without loss of generality, we assume that G has no stabilized vertices. Otherwise, we pick the set of nodes U that are not stable on x, and continue the reasoning with the subgraph of G induced by U. Observe that $G[U]$ is a d-degenerate graph as the property is hereditary.

Let W be the set of nodes in G of degree at most $4d-2$. From Lemma 2.2 we know that $|W| \geq n/2$. Now let us call α_W the size of a maximum independent set of the graph induced by W. From Proposition 2.1 we have that

$$\alpha_W \geq \frac{|W|}{4d-2} \geq \frac{n}{2(4d-2)}.$$

Then, from Lemma 5.1 we know that after two time-steps, the expected number of stabilized nodes is:

$$\frac{\alpha_W}{4^{4d-1}} \geq \frac{n}{2(4d-2)4^{4d-1}}.$$

Let us denote $c(d) = 2(4d-2)4^{4d-1}$. The previous bound implies that after

$$T \geq \frac{2}{\log(c(d)) - \log(c(d)-1)} \cdot \log(n)$$

time-steps, the expected number of non-stabilized nodes is at most $1/n$. By the Markov inequality, we deduce that on time-step T the probability that all nodes are stabilized is at least $1 - 1/n$. We conclude that, with high probability, on $\mathcal{O}(\log n)$ time-steps all nodes are stabilized. ■

6. Discussion

We have presented a very simple dynamics that converges to configurations that represent a maximal independent set of the input graph on any initial configuration with probability 1. Our experimental results suggest that in average the convergence time of our dynamics is $\mathcal{O}(\log n)$. Moreover, the dynamics on three or more states converges in $\mathcal{O}(\log n)$ steps with high probability, while in the case of two states, the convergence time is $\mathcal{O}(\log^2 n)$ with high probability.

The results given in Theorems 4.5 and 4.8 confirm the observations on graphs with constant independence number (such as complete graphs). Theorem 5.2 also confirms the observations of the 2-MIS-Dynamics on graphs of bounded degeneracy. Finally, Proposition 2.3 together with Theorems 4.5 and 4.8 implies poly-logarithmic convergence time for graphs of Erdös-Rényi in average, as well as with high probability.

We remark that the convergence time of our dynamics is not settled for general graphs. The existence of graph classes with large convergence time (on average or with high probability) is a possibility that we do not completely explore in this article. In this sense, we conjecture that our results can be improved in order to show a poly-logarithmic convergence time for every graph class.

From a more general perspective, we believe that there is an interesting research line related to the definition of simple finite-state dynamics that can work as strategies to efficiently compute to other graph structures, different that maximal independent sets, such as maximal matchings, minimal dominating sets, etc.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was possible with the support of the supercomputing infrastructure of the NLHPC [grant number ECM-02] and the support of Centro de Modelamiento Matemático (CMM) , FB210005 BASAL funds for centers of excellence from ANID-Chile . E.G. acknowledges ANID-FONDECYT 1200006. L.L. acknowledges the Chilean government through the National Agency for Research and Development (ANID) [Agencia Nacional de Investigación y Desarrollo]/Scholarship Program/BECA DOCTORADO NACIONAL/2019-21191440. P.M. acknowledges ANID-PAI 77170068. I.R. acknowledges ANID-FONDECYT 1220142. M.R.-W. acknowledges ANID-FONDECYT Postdoctorado 3220205.

References

[1] Karp RM, Wigderson A. A fast parallel algorithm for the maximal independent set problem. J ACM. 1985;32(4):762–773.
[2] Luby M. A simple parallel algorithm for the maximal independent set problem. In: Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing, Providence Rhode Island USA, May 6 - 8; 1985. p. 1–10.
[3] Alon N, Babai L, Itai A. A fast and simple randomized parallel algorithm for the maximal independent set problem. J Algorithms. 1986;7(4):567–583.
[4] Ghaffari M, Grunau C, Rozhoň V. Improved deterministic network decomposition. In: Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA). SIAM; 2021. p. 2904–2923.
[5] Barenboim L, Elkin M, Kuhn F. Distributed $(\delta+1)$-coloring in linear (in δ) time. SIAM J Comput. 2014;43(1):72–95.
[6] Ghaffari M. An improved distributed algorithm for maximal independent set. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete algorithms. SIAM; 2016. p. 270–277.
[7] Linial N. Locality in distributed graph algorithms. SIAM J Comput. 1992;21(1):193–201.
[8] Kuhn F, Moscibroda T, Wattenhofer R. What cannot be computed locally! In: Proceedings of the Twenty-Third Annual ACM Symposium on Principles of Distributed Computing; St. John's Newfoundland, Canada July 25 - 28, 2004. p. 300–309.
[9] Afek Y, Alon N, Bar-Joseph Z, et al. Beeping a maximal independent set. Distrib Comput. 2013;26(4):195–208.
[10] Emek Y, Wattenhofer R. Stone age distributed computing. In: Proceedings of the 2013 ACM Symposium on Principles of Distributed Computing; Montréal Québec, Canada, July 22 - 24, 2013. p. 137–146.
[11] Hedetniemi SM, Hedetniemi ST, Jacobs DP, et al. Self-stabilizing algorithms for minimal dominating sets and maximal independent sets. Comput Math Appl. 2003;46(5–6):805–811.
[12] Giakkoupis G. Simple efficient distributed processes on graphs. In: 2022 Keynote Talk of the 29th International Colloquium of Structural Information and Communication Complexity; 2022 Jun 28. Paderborn, Germany: SIROCCO; 2022.
[13] Caro Y. New results on the independence number. Tel-Aviv University; 1979. (Technical report).
[14] Bollobás B, Erdös P. Cliques in random graphs. In: Mathematical Proceedings of the Cambridge Philosophical Society; Vol. 80. Cambridge University Press; 1976. p. 419–427.