



Maximizing the expected number of components in an online search of a graph

Malgorzata Sulkowska, Fabrício Siqueira Benevides

► To cite this version:

Malgorzata Sulkowska, Fabrício Siqueira Benevides. Maximizing the expected number of components in an online search of a graph. *Discrete Mathematics*, 2022, 345 (1), pp.112668. 10.1016/j.disc.2021.112668 . hal-03462292

HAL Id: hal-03462292

<https://hal.science/hal-03462292>

Submitted on 2 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MAXIMIZING THE EXPECTED NUMBER OF COMPONENTS IN AN ONLINE SEARCH OF A GRAPH

FABRÍCIO SIQUEIRA BENEVIDES AND MAŁGORZATA SULKOWSKA

ABSTRACT. The following optimal stopping problem is considered. The vertices of a graph G are revealed one by one, in a random order, to a selector. He aims to stop this process at a time t that maximizes the expected number of connected components in the graph \tilde{G}_t , induced by the currently revealed vertices. The selector knows G in advance, but different versions of the game are considered depending on the information that he gets about \tilde{G}_t . We show that when G has N vertices and maximum degree of order $o(\sqrt{N})$, then the number of components of \tilde{G}_t is concentrated around its mean, which implies that playing the optimal strategy the selector does not benefit much by receiving more information about \tilde{G}_t . Results of similar nature were previously obtained by M. Lasoń for the case where G is a k -tree (for constant k). We also consider the particular cases where G is a square, triangular or hexagonal lattice, showing that an optimal selector gains cN components and we compute c with an error less than 0.005 in each case.

1. INTRODUCTION

Let $G = (V, E)$ be a graph on N vertices. Let \mathcal{S} be the set of all permutations of V . We consider the following online stopping problem. Select uniformly at random a permutation $\sigma \in \mathcal{S}$, say $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$. The vertices of G emerge, one by one, following the order given by σ . For $t \in \{1, 2, \dots, N\}$, let $\tilde{G}_t(\sigma)$, or simply \tilde{G}_t , be the graph induced by $\{\sigma_1, \sigma_2, \dots, \sigma_t\}$. We think about t as time and at each time step a player, who knows what the graph G is, must take a decision based on some information that he receives about \tilde{G}_t : either he continues the process and reveals the next vertex or he stops the game and gains as payoff the number of connected components of \tilde{G}_t . In case he decides to reveal another vertex, he is not allowed to go back to the previous step of the game. If the player never takes the decision to stop, the game finishes with \tilde{G}_N selected and the payoff is equal to the number of components of G .

The presented question may be treated as one of many generalizations of the celebrated *secretary problem* that attracted the attention of mathematical society in the early 1960's (consult [15] and [5]). In the secretary problem the player observes elements of a linear order emerging one by one in some random permutation. At a given time step he can see the order induced by the elements that have already appeared. His task is to stop the search maximizing the probability that the element that has just appeared is the maximal one in the whole order. Series of papers in which the linear order has been replaced by a partial order followed the work of

Date: October 8, 2021.

1991 Mathematics Subject Classification. Primary: 60G40, Secondary: 60K35.

Key words and phrases. optimal stopping, secretary problem, 2-dimensional lattice, simple graph.

Stadje [18]. Optimal strategies for a particular posets as well as universal algorithms for the whole families of posets have been presented in [16], [7], [17], [12] and [6]. Kubicki and Morayne were the first ones to investigate the optimal stopping problem on a directed graph choosing a directed path as the underlying structure, [13]. The link between the directed path case and the classical secretary problem was given by the authors in [1]. Universal algorithms for graphs were formulated by Goddard et al. in [8] and by Sulkowska in [19]. This paper continues the study of optimal stopping algorithms for graphs. However, the approach to the subject is slightly different since now the aim is to maximize the expected number of components at the stopping moment instead of the probability that the last vertex belongs to some previously defined set. In turn, the study of components is another classical topic in the area of random graphs. The first paper that puts optimal stopping for graphs in the setup of counting components is [14] by Lason.

One can introduce various versions of the presented stopping game, depending on exactly what information the player receives about \tilde{G}_t . In every version, the player knows G in advance and his task is to find the strategy that maximizes the expected payoff. The following three versions may serve as examples.

1. Blind game. At time t the selector knows only the number of vertices that have already appeared (i.e., t). He has no other information about the revealed structure. In fact, he gains no information during the game.
2. Partial information game. The selector can see an unlabeled graph isomorphic to \tilde{G}_t . In particular, he knows how many edges or components are there at time t , but he does not know exactly which vertices of G have been selected. This is a classical setup for many optimal stopping problems considered in the past (comparable to the setup in the secretary problem).
3. Full information game. The selector knows $\{\sigma_1, \sigma_2, \dots, \sigma_t\}$, and since he knows G , he knows \tilde{G}_t . Thus he gets all information that is available at time t .

In [14], Lason considers the above three versions for the case when G is a k -tree, for some constant k . Surprisingly, the maximum expected payoff for a selector that plays optimally with full information is very close to the one for an optimal selector playing the blind game. In this article, we prove that a similar statement holds for any graph G with N vertices and maximum degree bounded from above by $o(\sqrt{N})$. Throughout the rest of the paper we are going to refer only to blind and full information games, as the expected payoff for the partial information game is between those of the other two.

We also study the cases where G is a square, a triangular or a hexagonal lattice and we provide tight estimates for the expected payoff in those three cases. The study of lattices is motivated by the relation (explained below) between our problem and the well studied site percolation problem on 2-dimensional lattices. Another motivation is the fact that the results from [14] are stated for k -trees, which are maximal (with respect to the inclusion of edges) k -degenerate graphs and at the same time maximal graphs with treewidth equal to k . In contrast, 2-dimensional lattices are also k -degenerate (a square lattice and a hexagonal lattice are 2-degenerate, while a triangular lattice is 3-degenerate) but all have unbounded treewidth. It turns out that the maximum expected payoff for 2-dimensional lattices is smaller than the one for k -trees in a non-negligible way.

2. FORMAL MODEL AND NOTATION

Our graph $G = (V, E)$ with $V = \{v_1, v_2, \dots, v_N\}$ is always finite, simple and undirected. The degree of $v \in V$ will be denoted by $\deg(v)$. An induced subgraph $G' = (W, E \cap W^2)$, where $W \subseteq V$, is called a *component* of G if it is a maximal connected induced subgraph of G . We define a probability space $(\mathcal{S}, \mathcal{P}, \mathbb{P})$, where \mathcal{P} is the set of all subsets of \mathcal{S} and the probability measure is defined by $\mathbb{P}[\{\sigma\}] = 1/N!$ for any permutation $\sigma \in \mathcal{S}$. A *stopping time* is a function $\tau : \mathcal{S} \rightarrow \{1, 2, \dots, N\}$ such that its value on a permutation σ , say $t = \tau(\sigma)$, depends only on the information the selector gathered up to time t , which is information the selector gets about \tilde{G}_t (the graph induced by $\{\sigma_1, \sigma_2, \dots, \sigma_t\}$). E.g., in the blind game the selector gains no information during the game thus τ needs to be the same for all $\sigma \in \mathcal{S}$, it is simply a constant function depending only on G . Formally, let $\mathcal{P}_1 \subseteq \mathcal{P}_2 \subseteq \dots \subseteq \mathcal{P}_N \subseteq \mathcal{P}$ be a sequence of σ -algebras (such a sequence is called a *filtration*). A random variable $\tau : \mathcal{S} \rightarrow \{1, 2, \dots, N\}$ is a stopping time with respect to a filtration $(\mathcal{P}_t)_{t=1}^N$ if $\tau^{-1}(t) \in \mathcal{P}_t$ for all $t \leq N$. In the blind game we define $\mathcal{P}_t = \{\emptyset, \mathcal{S}\}$ for every t . In the full information game, for each $t \in \{1, 2, \dots, N\}$, the sets A in \mathcal{P}_t are those with the following property: if $\sigma = (\sigma_1, \dots, \sigma_N) \in A$, then for every $\pi = (\pi_1, \dots, \pi_N) \in \mathcal{S}$ such that $\{\sigma_1, \dots, \sigma_t\} = \{\pi_1, \dots, \pi_t\}$ we have $\pi \in A$.

A stopping algorithm is any algorithm that produces a stopping time. Let \mathcal{T} denote the set of all stopping times (note that the definition of \mathcal{T} depends on whether we are playing the blind or the full information version, but we omit this in the notation). Let \tilde{C}_t be the number of components of \tilde{G}_t . We say that τ^* is optimal if

$$\tau^* = \operatorname{argmax}_{\tau \in \mathcal{T}} \mathbb{E}[\tilde{C}_\tau],$$

that is, it maximizes the expected number of components at the time it stops. Here, $\mathbb{E}[\tilde{C}_\tau]$ is understood as $\frac{1}{N!} \sum_{\sigma \in \mathcal{S}} \tilde{C}_{\tau(\sigma)}(\sigma)$.

This is the classical optimal stopping setup, but in our proofs it is going to be more convenient to work with a different (probability) model. Assume the graph G is given and let $p \in [0, 1]$. Each vertex of G is declared *open* with probability p and *closed* with probability $1 - p$, independently of the other vertices. By G_p we denote the graph induced by the set of open vertices. Let C_p be the number of connected components of G_p . When G is an infinite lattice, the problem of deciding for what values of p there exists (with high probability) an infinite connected component in G_p is known as *site percolation*. Due to its huge number of applications this problem was overly studied by mathematicians as well as physicists (see, for example, the book [4]). Both, theoretical arguments and computer simulations were used in order to investigate percolation phenomenon, especially in the context of phase transitions. However, we have not found articles providing good estimates for C_p for general values of p (especially when p is far from the percolation threshold). For some general estimates consult [20, 9].

Intuitively, for N sufficiently large and $t \in \{1, 2, \dots, N\}$, letting $p = t/N$, one should expect that \tilde{C}_t and C_p behave similarly. For the sake of completeness, we will prove that this is indeed the case for graphs with maximum degree bounded by $o(\sqrt{N})$. To show that C_p is concentrated around its mean we use classical concentration result from [11] known as Azuma's inequality or McDiarmid's inequality. In order to compare C_p and \tilde{C}_t , or in general G_p and \tilde{G}_t , we consider a coupling of those random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = [0, 1]^N$, \mathcal{F} is the

family of all Borel sets of Ω , and \mathbb{P} is the uniform distribution (i.e., with probability density function equal to 1 everywhere). Given $\omega = (\omega_1, \dots, \omega_N) \in \Omega$, we interpret ω_i as the arrival time of the vertex v_i of G . Note that each ω induces (almost surely) a permutation σ such that $\sigma = (v_{i_1}, v_{i_2}, \dots, v_{i_N})$ for $\omega_{i_1} < \omega_{i_2} < \dots < \omega_{i_N}$. If this is the case, we write $\omega \rightsquigarrow \sigma$. Note that, in this way, the distribution that ω induces on the set of permutations \mathcal{S} is the uniform distribution.

For $\omega \in \Omega$, by $\tilde{G}_t(\omega)$ we understand the graph induced by $\{\sigma_1, \sigma_2, \dots, \sigma_t\}$, where σ is induced by ω . Note that, for any fixed $t \in \{1, \dots, N\}$, we have that $\tilde{G}_t(\omega)$ has the same distribution as $\tilde{G}_t(\sigma)$ where σ is drawn directly from $(\mathcal{S}, \mathcal{P}, \mathbb{P})$. Similarly, given $p \in [0, 1]$, we can define $G_p(\omega)$ as the graph induced by the vertices v_i for which $\omega_i \leq p$ and such graph has the same distribution as G_p defined in the previous paragraph. Note that, with this notation,

$$\mathbb{E}[\tilde{C}_\tau] = \int_{\Omega} \tilde{C}_{\tau(\omega)}(\omega) d\omega,$$

where $\tau(\omega)$ is naturally understood as $\tau(\sigma)$, where σ is the permutation induced by ω .

By $\tilde{V}_t(\omega)$, $\tilde{E}_t(\omega)$ and $\tilde{C}_t(\omega)$ we denote, respectively, the number of vertices, edges and components of the random graph $\tilde{G}_t(\omega)$. Whenever the context is clear we write \tilde{V}_t , \tilde{E}_t and \tilde{C}_t for short. Similarly, we denote by V_p , E_p and C_p the analogous random variables with respect to G_p . Note that \tilde{V}_t is a constant equal to t , while V_p follows the binomial distribution with parameters N and p .

For $S \subseteq \mathcal{S}$ let $\Omega_S = \bigcup_{\sigma \in S} \{\omega : \omega \rightsquigarrow \sigma\}$. In particular when $S = \{\sigma\}$, we simply use Ω_σ . Note that $\tilde{E}_t(\omega)$ is a random variable that is constant on Ω_σ (and the same holds for $\tilde{C}_t(\omega)$). This value will be interchangeably denoted by $\tilde{E}_t(\sigma)$ (and similarly $\tilde{C}_t(\sigma)$ for $\tilde{C}_t(\omega)$).

3. BLIND VERSUS FULL INFORMATION

This whole section is devoted to proving quite a surprising result. We show that in many situations the maximum expected payoff while playing in the full information mode is very close to the maximum expected payoff while playing in the blind mode. Precisely speaking, we are going to prove the following theorem.

Theorem 1. *Let G be a graph on N vertices. Let τ^f be the optimal algorithm while playing in a full information mode and let τ^b be the optimal algorithm while playing in a blind mode. For every $\varepsilon \in (0, 1)$ there exists N_ε such that if $N \geq N_\varepsilon$ and the maximum degree of G is bounded by $D_{\varepsilon, N} = \frac{\varepsilon^2}{32} \sqrt{N}$, then*

$$\mathbb{E}[\tilde{C}_{\tau^b}] \leq \mathbb{E}[\tilde{C}_{\tau^f}] \leq \mathbb{E}[\tilde{C}_{\tau^b}] + \varepsilon N.$$

Remark. Note that the stated bounds are useful when $\mathbb{E}[\tilde{C}_{\tau^b}]$ is of order N . Determining precisely for which families of graphs this condition holds is left as a future work. In the next section we give natural examples of graphs that fulfill this requirement.

Before we prove the main theorem, we state several technical lemmas that will be helpful later on. The first one is a concentration result (known as McDiarmid's inequality) which is a version of Azuma's inequality tailored for combinatorial applications (see [10] and [11]).

Lemma 1. *Let Z_1, Z_2, \dots, Z_M be independent random variables, with Z_j taking values in a set Λ_j . Assume that a function $g : \Lambda_1 \times \Lambda_2 \times \dots \times \Lambda_M \rightarrow \mathbb{R}$ satisfies, for some constants b_j , where $j \in \{1, 2, \dots, M\}$, the following Lipschitz condition:*

(L) *if two vectors $\mathbf{z}, \mathbf{z}' \in \Lambda_1 \times \Lambda_2 \times \dots \times \Lambda_M$ differ only in j^{th} coordinate, then $|g(\mathbf{z}) - g(\mathbf{z}')| \leq b_j$.*

Then the random variable $X = g(Z_1, Z_2, \dots, Z_M)$ satisfies, for any $t \geq 0$,

$$\mathbb{P}[X \geq \mathbb{E}[X] + t] \leq \exp \left\{ \frac{-2t^2}{\sum_{j=1}^M b_j^2} \right\},$$

$$\mathbb{P}[X \leq \mathbb{E}[X] - t] \leq \exp \left\{ \frac{-2t^2}{\sum_{j=1}^M b_j^2} \right\}.$$

The above lemma generalizes many known concentration bounds for sums of independent random variables. For example, compare it to Hoeffding's inequality [10].

We will also need a concentration inequality for a binomial random variable, $X \sim \text{Bin}(N, p)$, for an additive error of order \sqrt{N}/ε . In this range, it will be enough (and more convenient for us) to simply use Chebyshev's inequality.

Lemma 2. *Let X be a random variable following the binomial distribution with parameters N and $p \in [0, 1]$. Then, for every $\varepsilon > 0$,*

$$\mathbb{P} \left[|X - \mathbb{E}[X]| \geq 4\sqrt{N}/\varepsilon \right] \leq \frac{\varepsilon^2}{64}.$$

Proof. By Chebyshev's inequality we get

$$\mathbb{P} \left[|X - \mathbb{E}[X]| \geq 4\sqrt{N}/\varepsilon \right] \leq \frac{\varepsilon^2 \text{Var}[X]}{16N} = \frac{\varepsilon^2 Np(1-p)}{16N} \leq \frac{\varepsilon^2}{64}.$$

□

In the next lemma we show a trivial bound for the *mean absolute deviation* of a binomial random variable.

Lemma 3. *Let X be a random variable following the binomial distribution with parameters N and $p \in [0, 1]$. Then*

$$\mathbb{E}[|X - \mathbb{E}[X]|] \leq \frac{1}{2}\sqrt{N}.$$

Proof. For any random variable T we have $\text{Var}[T] = \mathbb{E}[T^2] - (\mathbb{E}[T])^2 \geq 0$ thus $(\mathbb{E}[T])^2 \leq \mathbb{E}[T^2]$. Setting $T = |X - \mathbb{E}[X]|$ we get

$$(\mathbb{E}[|X - \mathbb{E}[X]|])^2 \leq \mathbb{E}[(X - \mathbb{E}[X])^2] = \text{Var}[X],$$

which gives

$$\mathbb{E}[|X - \mathbb{E}[X]|] \leq \sqrt{\text{Var}[X]} = \sqrt{Np(1-p)} \leq \frac{1}{2}\sqrt{N}.$$

□

Remark. There are sharper results than Lemma 3. One can show that for such X , we have $\mathbb{E}[|X - \mathbb{E}[X]|] = \sqrt{2/\pi} \sqrt{\text{Var}[X]} + O(N^{-1/2})$ (consult [3] or [2]). But this improvement of only a constant factor does not yield any improvement to our main theorem.

The next few lemmas compare the number of components in the two models that we use in this article.

Lemma 4. *Let $t \in \{1, 2, \dots, N\}$. Let G be a graph on N vertices with the maximum degree at most D and $\tilde{C}_t, C_{t/N}$ be built from the probability space Ω (as in Section 2). Then for every $\varepsilon > 0$*

$$\mathbb{P} \left[\tilde{C}_t > C_{t/N} + \frac{4D\sqrt{N}}{\varepsilon} \right] \leq \frac{\varepsilon^2}{64}.$$

Proof. Note that for each $\omega \in \Omega$ we have

$$(1) \quad \tilde{C}_t \leq C_{t/N} + D |V_{t/N} - \tilde{V}_t| = C_{t/N} + D |V_{t/N} - t|.$$

Indeed, the graphs \tilde{G}_t and $G_{t/N}$ arose according to the same permutation, induced by ω . Then, they differ by $|V_{t/N} - \tilde{V}_t|$ vertices and each additional vertex may increase the number of components by at most one and decrease the number of components by at most $D - 1$. Recall that $V_{t/N}$ follows the binomial distribution with parameters N and t/N thus $\mathbb{E}[V_{t/N}] = t$ and, by Lemma 2, we know that

$$\mathbb{P} \left[|V_{t/N} - t| < 4\sqrt{N}/\varepsilon \right] \geq 1 - \frac{\varepsilon^2}{64}.$$

Therefore,

$$\mathbb{P} \left[\tilde{C}_t \leq C_{t/N} + 4D\sqrt{N}/\varepsilon \right] \geq 1 - \frac{\varepsilon^2}{64}.$$

□

Lemma 5. *Let $t \in \{1, 2, \dots, N\}$. Let G be a graph on N vertices with the maximum degree bounded by D and $\tilde{C}_t, C_{t/N}$ be built from the probability space Ω (as in Section 2). Then*

$$\mathbb{E}[C_{t/N}] \leq \mathbb{E}[\tilde{C}_t] + \frac{1}{2}D\sqrt{N}.$$

Proof. Analogously to (1), for all $\omega \in \Omega$, we write

$$C_{t/N} \leq \tilde{C}_t + D |V_{t/N} - t|.$$

Since $V_{t/N}$ is binomial with expected value t , taking expectation (in Ω) on both sides, the conclusion follows from Lemma 3. □

The next lemma shows that in the blind game we do not loose much if we look at G_p for $p = t/N$ instead of \tilde{G}_t . Also, for a good bound, it is enough to consider a finite number of values of p : $1/N, 2/N, \dots, N/N$.

Lemma 6. *Let G be a graph on N vertices with the maximum degree bounded by D and let τ^b be an optimal algorithm while playing in a blind mode. Then*

$$(2) \quad \max_{t \in \{1, \dots, N\}} \mathbb{E}[C_{t/N}] \leq \mathbb{E}[\tilde{C}_{\tau^b}] + \frac{1}{2}D\sqrt{N}.$$

Proof. In a blind mode the selector does not gain any new information during the game. He can actually decide in advance when to stop. Of course, the only reasonable strategy is to stop at time t maximizing the expected number of components. Therefore,

$$\mathbb{E}[\tilde{C}_{\tau^b}] = \max_{t \in \{1, \dots, N\}} \mathbb{E}[\tilde{C}_t].$$

Now, let t^b be such that $\mathbb{E}[C_{t^b/N}] = \max_{t \in \{1, \dots, N\}} \mathbb{E}[C_{t/N}]$. It follows that:

$$\mathbb{E}[\tilde{C}_{\tau^b}] = \max_{t \in \{1, \dots, N\}} \mathbb{E}[\tilde{C}_t] \geq \mathbb{E}[\tilde{C}_{t^b}] \geq \mathbb{E}[C_{t^b/N}] - \frac{1}{2}D\sqrt{N},$$

where the last inequality follows from Lemma 5. Now, equation (2) follows directly. \square

In our final lemma, we show that C_p , for constant p , is concentrated around its mean for certain G . In particular, the condition in the lemma is satisfied when the maximum degree of G is $o(\sqrt{N})$.

Lemma 7. *Let $p \in [0, 1]$. For every $\varepsilon \in (0, 1)$ there exists N_ε such that if G is a graph with $N \geq N_\varepsilon$ vertices and $\sum_{j=1}^N \deg(v_j)^2 \leq \delta_\varepsilon N^2$, where $\delta_\varepsilon = \frac{\varepsilon^2/64}{\ln(64/\varepsilon^2)}$, then C_p satisfies*

$$\mathbb{P}[C_p \geq \mathbb{E}[C_p] + (\varepsilon/8)N] \leq \varepsilon^2/64.$$

Proof. Let $p \in [0, 1]$. For $i \in \{1, 2, \dots, N\}$ and $\omega \in \Omega$, define Z_i as follows

$$Z_i = \begin{cases} 1 & \text{if } \omega_i \leq p, \\ 0 & \text{if } \omega_i > p. \end{cases}$$

Recall that ω_i represents the arrival time of v_i . Thus Z_i simply indicates whether v_i belongs to G_p or not. Put $g(Z_1, Z_2, \dots, Z_N) = C_p$. Note that for two vectors $\mathbf{z}, \mathbf{z}' \in \{0, 1\}^N$ that differ only in j^{th} coordinate we have $|g(\mathbf{z}) - g(\mathbf{z}')| \leq \deg(v_j)$ unless v_j is isolated in G . Indeed, whenever we add a single vertex v_j to the graph, the number of components can increase by at most one or decrease by at most $\deg(v_j) - 1$. If $\deg(v_j) = 0$ the number of components always increases by one when v_j appears. For $j \in \{1, 2, \dots, N\}$ define

$$b_j = \begin{cases} \deg(v_j) & \text{if } \deg(v_j) > 0, \\ 1 & \text{if } \deg(v_j) = 0. \end{cases}$$

We have

$$\sum_{j=1}^N b_j^2 \leq N + \sum_{j=1}^N \deg(v_j)^2,$$

where N represents the upper bound for the sum of squared ones while summing over vertices isolated in G . By Lemma 1, setting $t = (\varepsilon/8)N$, for sufficiently large N we obtain

$$\begin{aligned} \mathbb{P}[C_p \geq \mathbb{E}[C_p] + (\varepsilon/8)N] &\leq \exp \left\{ \frac{-2(\varepsilon/8)^2 N^2}{N + \sum_{j=1}^N \deg(v_j)^2} \right\} \leq \exp \left\{ \frac{-2\varepsilon^2 N^2}{64(N + \delta_\varepsilon N^2)} \right\} \\ &\leq \exp \left\{ \frac{-2\varepsilon^2 N^2}{64 \cdot 2\delta_\varepsilon N^2} \right\} = \exp \left\{ \frac{-\varepsilon^2}{64 \cdot \delta_\varepsilon} \right\} = \varepsilon^2/64. \end{aligned} \quad \square$$

Now we can use the concentration of C_p to prove the main theorem of this section. Hereby we partially follow the lines of the proof of Lemma 3.5 from [14].

Proof of Theorem 1. The lower bound is trivial since any stopping algorithm for the blind game is also a stopping algorithm for the full information game.

Now, let $T = \{0, \lfloor \frac{\varepsilon}{8}N \rfloor, \lfloor \frac{2\varepsilon}{8}N \rfloor, \dots, \lfloor \frac{\lfloor 8/\varepsilon \rfloor \varepsilon}{8}N \rfloor\} = \{t_0, t_1, \dots, t_{\lfloor 8/\varepsilon \rfloor}\}$. Put $t_{\lfloor 8/\varepsilon \rfloor+1} = N+1$. For $i \in \{0, 1, \dots, \lfloor 8/\varepsilon \rfloor\}$ partition \mathcal{S} into $\bigcup S_{t_i}$, where S_{t_i} is the set of permutations σ for which $\tau^f(\sigma) \in [t_i, t_{i+1})$. We can write

$$\mathbb{E}[\tilde{C}_{\tau^f}] = \frac{1}{N!} \sum_{\sigma \in \mathcal{S}} \tilde{C}_{\tau^f(\sigma)} = \frac{1}{N!} \sum_{t \in T} \sum_{\sigma \in S_t} \tilde{C}_{\tau^f(\sigma)},$$

or, equivalently,

$$\mathbb{E}[\tilde{C}_{\tau^f}] = \int_{\Omega} \tilde{C}_{\tau^f(\omega)} d\omega = \sum_{t \in T} \left(\int_{\Omega_{S_t}} \tilde{C}_{\tau^f(\omega)} d\omega \right).$$

Note that for $t \in T$ and $\omega \in \Omega_{S_t}$, letting $\omega \rightsquigarrow \sigma$, implies $\sigma \in S_t$ and therefore $\tilde{C}_{\tau^f(\omega)} = \tilde{C}_{\tau^f(\sigma)} \leq \tilde{C}_t + \frac{\varepsilon}{8}N+1$, as each new vertex adds at most one component. By Lemma 4, for any given t , we know that $\mathbb{P} \left[\tilde{C}_t > C_{t/N} + 4D_{\varepsilon,N} \sqrt{N}/\varepsilon \right] \leq \frac{\varepsilon^2}{64}$ (recall that in the statement of Theorem 1 we define $D_{\varepsilon,N} = \varepsilon^2 \sqrt{N}/32$). As $4D_{\varepsilon,N} \sqrt{N}/\varepsilon = (\varepsilon/8)N$, this means that

$$\tilde{C}_t \leq C_{t/N} + (\varepsilon/8)N$$

except for at most a $\varepsilon^2/64$ fraction of the whole Ω . One can also easily check that the assumptions of Lemma 7 are satisfied; for δ_ε from Lemma 7 we get $\sum_{i=1}^N \deg(v_i)^2 \leq \sum_{i=1}^N D_{\varepsilon,N}^2 \leq (\varepsilon^4/32^2)N^2 \leq \delta_\varepsilon N^2$. Thus for N_ε from Lemma 7, for any given t , we get that for $N \geq N_\varepsilon$ at most $\frac{\varepsilon^2}{64}$ fraction of all ω 's do not satisfy the inequality

$$C_{t/N} \leq \mathbb{E}[C_{t/N}] + (\varepsilon/8)N.$$

Putting it together we get that for $N \geq N_\varepsilon$

$$\tilde{C}_t \leq \mathbb{E}[C_{t/N}] + (\varepsilon/4)N$$

except for at most $\frac{\varepsilon^2}{32}$ fraction of the whole Ω , where we can use that $\tilde{C}_t \leq N$. Thus, for each $t \in T$ we get:

$$\begin{aligned} \int_{\Omega_{S_t}} \tilde{C}_{\tau^f(\omega)} d\omega &\leq \int_{\Omega_{S_t}} \left(\tilde{C}_t + \frac{\varepsilon}{8}N + 1 \right) d\omega \\ &\leq \int_{\Omega_{S_t}} \left(\mathbb{E}[C_{t/N}] + \frac{3\varepsilon}{8}N + 1 \right) d\omega + \frac{\varepsilon^2}{32}N. \end{aligned}$$

Now, summing the above inequality over $t \in T$, for $N \geq N_\varepsilon$, and since $\varepsilon < 1$, we get

$$\begin{aligned}
\mathbb{E}[\tilde{C}_{\tau^f}] &\leq \sum_{t \in T} \left(\int_{\Omega_{S_t}} \left(\mathbb{E}[C_{t/N}] + \frac{3\varepsilon}{8}N + 1 \right) d\omega \right) + \left(\left\lfloor \frac{8}{\varepsilon} \right\rfloor + 1 \right) \frac{\varepsilon^2}{32}N \\
&\leq \sum_{t \in T} \left(\int_{\Omega_{S_t}} \left(\mathbb{E}[C_{t/N}] + \frac{3\varepsilon}{8}N + 1 \right) d\omega \right) + \frac{4\varepsilon}{8}N \\
&\leq \sum_{t \in T} \left(\int_{\Omega_{S_t}} \left(\max_{t \in T} \mathbb{E}[C_{t/N}] + \frac{3\varepsilon}{8}N + 1 \right) d\omega \right) + \frac{4\varepsilon}{8}N \\
&= \left(\max_{t \in T} \mathbb{E}[C_{t/N}] + \frac{3\varepsilon}{8}N + 1 \right) \sum_{t \in T} \left(\int_{\Omega_{S_t}} 1 d\omega \right) + \frac{4\varepsilon}{8}N \\
&= \max_{t \in T} \mathbb{E}[C_{t/N}] + 1 + \frac{7\varepsilon}{8}N \leq \max_{t \in \{1, \dots, N\}} \mathbb{E}[C_{t/N}] + 1 + \frac{7\varepsilon}{8}N \\
&\leq \mathbb{E}[\tilde{C}_{\tau^b}] + \frac{1}{2}D_{\varepsilon, N}\sqrt{N} + 1 + \frac{7\varepsilon}{8}N \\
&\leq \mathbb{E}[\tilde{C}_{\tau^b}] + \frac{\varepsilon}{8}N + \frac{7\varepsilon}{8}N = \mathbb{E}[\tilde{C}_{\tau^b}] + \varepsilon N.
\end{aligned}$$

For the previous to last inequality refer to Lemma 6. The last inequality follows from the fact that $\frac{1}{2}D_{\varepsilon, N}\sqrt{N} + 1 < (\varepsilon/8)N$ for $\varepsilon < 1$. \square

4. PLAYING ON LATTICES

In this section we study a particular family of graphs, namely 2-dimensional lattices: square, triangular and hexagonal one (see Figure 1). Their degrees are bounded by a constant thus all satisfy the assumptions of Theorem 1. This means that playing a full information game we can not gain significantly more than playing blind. Even though we refer to the three particular lattices here, the reader will notice that the proofs of theorems from this section may be adapted for more general subgraphs of the infinite lattices. The next two lemmas together with Theorem 1 justify that giving the upper and lower bounds just for the value $\mathbb{E}[C_p]$ we get the upper and lower bounds for the gain of either blind or full information game.

Lemma 8. *Let G be a graph on N vertices with maximum degree at most D . Let τ^b be the optimal stopping time while playing in a blind mode. Let $g : [0, 1] \rightarrow \mathbb{R}$ be a function satisfying $\mathbb{E}[C_p] \leq Ng(p)$ and attaining its maximum at p_{\max} . Then*

$$\mathbb{E}[\tilde{C}_{\tau^b}] \leq Ng(p_{\max}) + \frac{1}{2}D\sqrt{N}.$$

Proof. As in (1), for all $\omega \in \Omega$ we write

$$\tilde{C}_{\tau^b} \leq C_{\tau^b/N} + D|V_{\tau^b/N} - \tau^b|.$$

Since we are considering a blind game, the value of τ^b is constant. Thus, taking expectation on both sides of the above inequality and using Lemma 3 we get

$$\begin{aligned}
\mathbb{E}[\tilde{C}_{\tau^b}] &\leq \mathbb{E}[C_{\tau^b/N}] + \frac{1}{2}D\sqrt{N} \leq \max_{t \in \{1, \dots, N\}} \mathbb{E}[C_{t/N}] + \frac{1}{2}D\sqrt{N} \\
&\leq \sup_{p \in [0, 1]} \mathbb{E}[C_p] + \frac{1}{2}D\sqrt{N} \leq Ng(p_{\max}) + \frac{1}{2}D\sqrt{N}.
\end{aligned}$$

□

Lemma 9. *Let G be a graph on N vertices with the maximal degree bounded by D . Let τ^b be the optimal stopping time while playing in a blind mode. Let $f : [0, 1] \rightarrow \mathbb{R}$ be a function continuous on $[0, 1]$, differentiable on $(0, 1)$ such that $Nf(p) \leq \mathbb{E}[C_p]$ and attaining its unique maximum at p_{\max} . Let also $|f'(p)| < b$ for some constant b and every $p \in (p_{\max} - 1/N, p_{\max} + 1/N) \cap (0, 1)$. Then*

$$\mathbb{E}[\tilde{C}_{\tau^b}] \geq Nf(p_{\max}) - \frac{1}{2}D\sqrt{N} - b.$$

Proof. By Lemma 6

$$\mathbb{E}[\tilde{C}_{\tau^b}] \geq \max_{t \in \{1, \dots, N\}} \mathbb{E}[C_{t/N}] - \frac{1}{2}D\sqrt{N} \geq N \cdot \max_{t \in \{1, \dots, N\}} f(t/N) - \frac{1}{2}D\sqrt{N}.$$

Let $t^* = \min_{t \in \{1, \dots, N\}} \{t : t/N \geq p_{\max}\}$. By the mean value theorem there exists $c \in (p_{\max} - 1/N, p_{\max} + 1/N) \cap (0, 1)$ such that

$$f(p_{\max}) - f(t^*/N) = |f'(c)| \cdot |t^*/N - p_{\max}| \leq b/N.$$

Therefore,

$$\mathbb{E}[\tilde{C}_{\tau^b}] \geq Nf(t^*/N) - \frac{1}{2}D\sqrt{N} \geq Nf(p_{\max}) - \frac{1}{2}D\sqrt{N} - b.$$

□

Thus throughout this section we focus on investigating the expected number of components only in the random graph G_p . Equivalently, we study parameters of a random graph obtained in a process of site percolation. Site percolation on lattices has been widely and deeply studied. Nevertheless, to the best of our knowledge, the exact values of the expected number of components in a random graph that evolved in a site percolation process are still not known for 2-dimensional lattices. Some precise upper and lower bounds were given only for p being close to the critical probability (consult [4]). The three theorems of this section give quite tight upper and lower bounds for the maximal value of $\mathbb{E}[C_p]$ in cases when G is a square, a triangular or a hexagonal lattice.

From this point, the graphs G that we consider are always planar. While C_p is a graph parameter that does not depend on a particular embedding of G , we will study it indirectly via Euler's formula, expressing the number of components in terms of the number of vertices, edges and faces of a particular embedding of G in the plane. That is, we study plane graphs.

Whenever we mention a square, triangular or hexagonal lattice we mean their particular embedding as in Figure 1. For our proofs the exact shape of the outer face is not relevant, but only the fact the number of vertices and faces that are not typical is $o(N)$ (in particular, $o(N)$ of the vertices of G belong to its outer face or have neighbors in the outer face). Other important fact about our embeddings are: in the square lattice (respectively, triangular/hexagonal) all the inner vertices (vertices that do not belong to the outer face) are of degree four (resp., six/three) and all the inner faces are squares (resp. triangles/hexagons).

In this section F_p stands for the number of faces in the plane graph G_p . By $F_p^{(k)}$ we denote the number of k -faces in the same plane graph, i.e., faces having exactly k edges in the boundary (taking into consideration that if an edge belongs to the boundary of only one face, then it is counted with multiplicity 2 for such face).

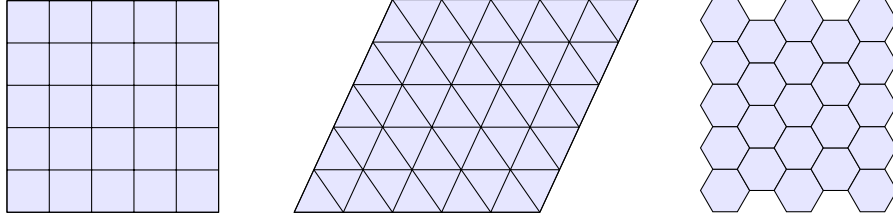
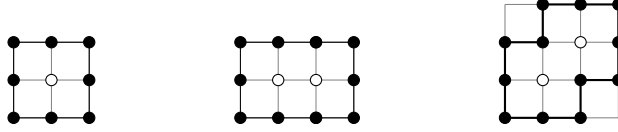


FIGURE 1. Square, triangular and hexagonal lattice.


 FIGURE 2. Examples of faces in G_p surrounding 1 or 2 (white) vertices from $V \setminus V_p$.

Lemma 10. *Let H be a connected, planar graph with v vertices, e edges, with $e \geq 2$, and f faces. Let also $f^{(3)} = f^{(5)} = f^{(7)} = 0$, where $f^{(s)}$ denote the number of s -faces in H . Then*

$$f \leq (e + 2f^{(4)} + f^{(6)})/4.$$

Proof. Since H has at least two edges, we have

$$\begin{aligned} 2e &= \sum_{k \geq 3} k f^{(k)} \geq 4f^{(4)} + 6f^{(6)} + \sum_{k \geq 8} k f^{(k)} \\ &\geq 4f^{(4)} + 6f^{(6)} + 8(f - f^{(4)} - f^{(6)}) = 8f - 4f^{(4)} - 2f^{(6)}. \end{aligned}$$

□

Theorem 2. *Let G be a square lattice on $n \times n = N$ vertices. Then*

$$\lim_{N \rightarrow \infty} \frac{\sup_{p \in (0,1)} \mathbb{E}[C_p]}{N} \in (0.12953, 0.13268).$$

Proof. Let us start with a lower bound for $\sup_{p \in (0,1)} \mathbb{E}[C_p]$. The graph G_p is planar for any $p \in (0, 1)$. Therefore, by Euler's formula we write

$$C_p = V_p - E_p + F_p - 1.$$

We call an inner face *empty* if it does not surround any vertex from $V \setminus V_p$. Note that since G is a square lattice, the number of empty faces in G_p equals $F_p^{(4)}$ or $F_p^{(4)} - 1$. The number of inner faces in G_p that surround exactly k vertices from $V \setminus V_p$ will be denoted by $F_p^{[k]}$. For example, all *non-empty* faces (up to rotations) surrounding exactly one or two vertices are shown in Figure 2. We have

$$C_p \geq V_p - E_p + F_p^{(4)} + F_p^{[1]} + F_p^{[2]} - 2.$$

Now,

$$\begin{aligned}\mathbb{E}[V_p] &= Np, & \mathbb{E}[E_p] &= (2N - 2n)p^2 \leq 2Np^2, \\ \mathbb{E}[F_p^{(4)}] &= (n - 1)^2 p^4 + o(N) = Np^4 + o(N), \\ \mathbb{E}[F_p^{[1]}] &= Np^8(1 - p) + o(N) \quad \text{and} \\ \mathbb{E}[F_p^{[2]}] &= 2Np^{10}(1 - p)^2 + 2Np^{12}(1 - p)^2 + o(N).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[C_p] &\geq N(p - 2p^2 + p^4 + p^8(1 - p) + 2p^{10}(1 - p)^2 + 2p^{12}(1 - p)^2) + o(N) \\ &= N(p - 2p^2 + p^4 + p^8 - p^9 + 2p^{10} - 4p^{11} + 2p^{12} - 4p^{13} + 2p^{14}) + o(N) \\ &= Nf(p) + o(N).\end{aligned}$$

The function $f(p)$ attains its unique maximum on $[0, 1]$ at p_{\max} approximately 0.27 and $f(p_{\max}) > 0.12953$.

Now we calculate an upper bound for $\sup_{p \in (0,1)} \mathbb{E}[C_p]$. Graph G_p has C_p components, name them H_1, H_2, \dots, H_{C_p} . For $i \in \{1, 2, \dots, C_p\}$ let e_i denote the number of edges in H_i and f_i the number of faces in H_i if H_i is considered as a standalone graph. Let J_p be the set of indices $j \in \{1, 2, \dots, C_p\}$ such that H_j has at least two edges. Note that $|J_p| = C_p - \bar{V}_p - \bar{E}_p$ where \bar{V}_p stands for the number of isolated vertices in G_p and \bar{E}_p for the number of isolated edges. By Lemma 10, we know that $f_j \leq (e_j + 2f_j^{(4)} + f_j^{(6)})/4$ for each $j \in J_p$. Therefore

$$\sum_{j \in J_p} f_j \leq \frac{1}{4} \sum_{j \in J_p} e_j + \frac{1}{2} \sum_{j \in J_p} f_j^{(4)} + \frac{1}{4} \sum_{j \in J_p} f_j^{(6)}.$$

We have $\sum_{j \in J_p} e_j = E_p - \bar{E}_p$ and $\sum_{j \in J_p} f_j = F_p + |J_p| - 1$; indeed, the outer face of G_p was counted once in every f_j . Denote $F_p^{(4*)} = \sum_{j \in J_p} f_j^{(4)}$ and $F_p^{(6*)} = \sum_{j \in J_p} f_j^{(6)}$. We have

$$F_p + |J_p| - 1 \leq \frac{1}{4}(E_p - \bar{E}_p) + \frac{1}{2}F_p^{(4*)} + \frac{1}{4}F_p^{(6*)}$$

and since $|J_p| = C_p - \bar{V}_p - \bar{E}_p$ we obtain

$$F_p \leq \frac{1}{4}E_p + \frac{3}{4}\bar{E}_p + \bar{V}_p + \frac{1}{2}F_p^{(4*)} + \frac{1}{4}F_p^{(6*)} - C_p + 1.$$

Using it together with the Euler's formula, $C_p = V_p - E_p + F_p - 1$, we get

$$(3) \quad C_p \leq \frac{1}{2} \left(V_p - \frac{3}{4}E_p + \frac{3}{4}\bar{E}_p + \bar{V}_p + \frac{1}{2}F_p^{(4*)} + \frac{1}{4}F_p^{(6*)} \right).$$

The number of isolated vertices and the number of isolated edges in G_p satisfy $\mathbb{E}[\bar{V}_p] = Np(1 - p)^4 + o(N)$ and $\mathbb{E}[\bar{E}_p] = 2Np^2(1 - p)^6 + o(N)$. Consider $F_p^{(4*)} = \sum_{j \in J_p} f_j^{(4)}$. Note that, unless $|J_p| = 1$ (in which case $F_p^{(4*)} = F_p^{(4)}$ or $F_p^{(4*)} = F_p^{(4)} + 1$), the value of $F_p^{(4*)}$ is equal to the number of 4-faces in G_p plus the number of graphs H_j whose outer face is a 4-face. In Figure 3 we consider all cases, up to rotation, where the outer face of a component is a 4-face. Therefore, recalling that $\mathbb{E}[F_p^{(4)}] = Np^4 + o(N)$,

$$\mathbb{E}[F_p^{(4*)}] = Np^4 + Np^4(1 - p)^8 + 4Np^3(1 - p)^7 + 2Np^3(1 - p)^8 + o(N).$$

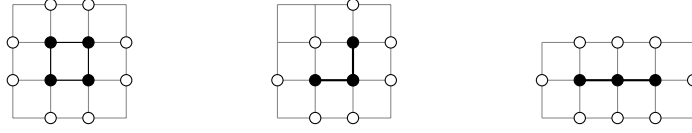


FIGURE 3. Examples of H_j 's having the outer face being a 4-face. Black vertices belong to the face, white ones must be on $V \setminus V_p$ and non-marked vertices are irrelevant.

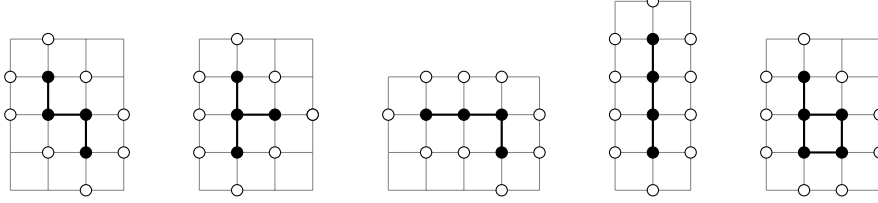


FIGURE 4. Examples of H_j 's having the outer face being a 6-face. Black vertices belong to the face, white ones must be on $V \setminus V_p$ and non-marked vertices are irrelevant.

Now, consider $F_p^{(6*)} = \sum_{j \in J_p} f_j^{(6)}$. Note that $f_j^{(6)} \neq 0$ if, and only if, $f_j^{(6)} = 1$ and the outer face of H_j is a 6-face (see examples in Figure 4). We get

$$\mathbb{E}[F_p^{(6*)}] = 8Np^4(1-p)^8 + 8Np^4(1-p)^9 + 2Np^4(1-p)^{10} + 8Np^5(1-p)^9 + o(N).$$

Finally,

$$\begin{aligned} \mathbb{E}[C_p] &\leq \frac{1}{2}N \left(p - \frac{3}{2}p^2 + \frac{3}{2}p^2(1-p)^6 + p(1-p)^4 + \frac{1}{2}p^4 + \frac{1}{2}p^4(1-p)^8 + 2p^3(1-p)^7 \right. \\ &\quad \left. + p^3(1-p)^8 + 2p^4(1-p)^8 + 2p^4(1-p)^9 + \frac{1}{2}p^4(1-p)^{10} + 2p^5(1-p)^9 \right) + o(N) \\ &= N \left(p - 2p^2 + p^4 + \frac{43}{2}p^6 - \frac{165}{2}p^7 + \frac{535}{4}p^8 - 112p^9 + \frac{81}{2}p^{10} \right. \\ &\quad \left. + \frac{17}{2}p^{11} - \frac{29}{2}p^{12} + \frac{11}{2}p^{13} - \frac{3}{4}p^{14} \right) + o(N) = Ng(p) + o(N). \end{aligned}$$

The function $g(p)$ attains its maximum on $[0, 1]$ at p'_{\max} approximately 0.29 and $g(p'_{\max}) < 0.13268$. \square

Remark. Graph G , being a square lattice on $n \times n = N$ vertices, is 2-degenerate. It can be easily transformed into a maximal 2-degenerate graph G' by adding $2n - 3 = \Theta(\sqrt{N})$ edges (note that a 2-degenerate graph on n^2 vertices has at most $2n^2 - 3$ edges while the lattice has exactly $2n^2 - 2n$ edges). Note that Theorem 2 still holds for G' since these additional edges may decrease the number of gained components by at most $o(N)$. In [14] it was proved that $\left(\frac{k^k}{(k+1)^{k+1}} + o(1) \right) N$ constitutes the upper bound for the expected number of components while playing full information game on a maximal k -degenerate graph on N vertices. Examples of graphs that attain this maximum are k -trees. Note that G' serves as an example of a maximal 2-degenerate graph which does not attain this upper bound. Indeed $0.13268 < 2^2/3^3 \approx 0.148$. One difference between G' and a k -tree is its unbounded treewidth.

However, we do not know how (and whether) this parameter really influences the gain of the game.

Theorem 3. *Let G be a triangular lattice on N vertices (drawn as in Figure 1). Then*

$$\lim_{N \rightarrow \infty} \frac{\sup_{p \in (0,1)} \mathbb{E}[C_p]}{N} \in (0.09629, 0.10107).$$

Proof. For the lower bound for $\sup_{p \in (0,1)} \mathbb{E}[C_p]$ we write

$$C_p \geq V_p - E_p + F_p^{(3)} + F_p^{(6)} - 1.$$

(Note that $F_p^{(4)}$ and $F_p^{(5)}$ are either 0 or 1, being 1 only if they count the outer face of G_p , so we do not lose much by ignoring those terms). For G being a triangular lattice we have $\mathbb{E}[V_p] = Np$, $\mathbb{E}[E_p] = 3Np^2 + o(N)$, $\mathbb{E}[F_p^{(3)}] = 2Np^3 + o(N)$ and $\mathbb{E}[F_p^{(6)}] = Np^6(1-p) + o(N)$ (where the $(1-p)$ in the last expression comes from the fact that the vertex of G that belongs to the interior of the hexagon cannot be selection to G_p). Thus

$$\mathbb{E}[C_p] \geq N(p - 3p^2 + 2p^3 + p^6(1-p)) + o(N) = Nf(p) + o(N).$$

The function $f(p)$ attains its unique maximum on $[0, 1]$ at p_{\max} approximately 0.21 and $f(p_{\max}) > 0.09629$.

For the upper bound we follow the lines of the proof of Theorem 2 just this time for the number of edges in H_j we use the following inequality

$$2e_j \geq 3f_j^{(3)} + 4f_j^{(4)} + 5f_j^{(5)} + 6(f_j - f_j^{(3)} - f_j^{(4)} - f_j^{(5)}).$$

Hence

$$\sum_{j \in J_p} f_j \leq \frac{1}{3} \sum_{j \in J_p} e_j + \frac{1}{2} \sum_{j \in J_p} f_j^{(3)} + \frac{1}{3} \sum_{j \in J_p} f_j^{(4)} + \frac{1}{6} \sum_{j \in J_p} f_j^{(5)}.$$

Again denote $F_p^{(3*)} = \sum_{j \in J_p} f_j^{(3)}$, $F_p^{(4*)} = \sum_{j \in J_p} f_j^{(4)}$ and $F_p^{(5*)} = \sum_{j \in J_p} f_j^{(5)}$. By Euler's formula applied to G_p , we have

$$C_p \leq \frac{1}{2} \left(V_p - \frac{2}{3} E_p + \frac{2}{3} \bar{E}_p + \bar{V}_p + \frac{1}{2} F_p^{(3*)} + \frac{1}{3} F_p^{(4*)} + \frac{1}{6} F_p^{(5*)} \right).$$

The fact that

$$\begin{aligned} \mathbb{E}[\bar{E}_p] &= 3Np^2(1-p)^8 + o(N), & \mathbb{E}[\bar{V}_p] &= Np(1-p)^6 + o(N), \\ \mathbb{E}[F_p^{(3*)}] &= 2Np^3 + 2Np^3(1-p)^9 + o(N), \\ \mathbb{E}[F_p^{(4*)}] &= 9Np^3(1-p)^{10} + 3Np^4(1-p)^{10} + o(N) \quad \text{and} \\ \mathbb{E}[F_p^{(5*)}] &= 12Np^4(1-p)^{11} + 6Np^5(1-p)^{11} + o(N) \end{aligned}$$

implies

$$\begin{aligned} \mathbb{E}[C_p] &\leq \frac{1}{2} N \left(p - 2p^2 + 2p^2(1-p)^8 + p(1-p)^6 + p^3 + p^3(1-p)^9 \right. \\ &\quad \left. + 3p^3(1-p)^{10} + p^4(1-p)^{10} + 2p^4(1-p)^{11} + p^5(1-p)^{11} \right) + o(N) \\ &= Ng(p) + o(N). \end{aligned}$$

The function $g(p)$ attains its maximum on $[0, 1]$ at p'_{\max} approximately 0.24 and $g(p'_{\max}) < 0.10107$. \square

Remark. Graph G , which is a triangular lattice on N vertices, is 3-degenerate. Again, it can be easily transformed into a maximal 3-degenerate graph G' by adding $\Theta(\sqrt{N})$ edges. Such G' serves as another example of a maximal k -degenerate graph which does not attain the upper bound from [14]. Indeed $0.10107 < 3^3/4^4 \approx 0.1055$.

Theorem 4. *Let G be a hexagonal lattice on N vertices (drawn as in Figure 1). Then*

$$\lim_{N \rightarrow \infty} \frac{\sup_{p \in (0,1)} \mathbb{E}[C_p]}{N} \in (0.16738, 0.17144).$$

Proof. For G being a hexagonal lattice we have $\mathbb{E}[V_p] = Np$, $\mathbb{E}[E_p] = (3/2)Np^2 + o(N)$, $\mathbb{E}[F_p^{(6)}] = (N/2)p^6 + o(N)$ and $\mathbb{E}[F_p^{(12)}] = Np^{12}(1-p) + o(N)$. This time for the lower bound we write

$$C_p \geq V_p - E_p + F_p^{(6)} + F_p^{(12)} - 1$$

thus

$$\mathbb{E}[C_p] \geq N \left(p - \frac{3}{2}p^2 + \frac{1}{2}p^6 + p^{12}(1-p) \right) + o(N) = Nf(p) + o(N).$$

The function $f(p)$ attains its unique maximum on $[0, 1]$ at p_{\max} approximately 0.34 and $f(p_{\max}) > 0.16738$.

For the upper bound we use again the inequality (3). Since now $\mathbb{E}[F_p^{(4*)}] = 3Np^3(1-p)^5 + o(N)$ and $\mathbb{E}[F_p^{(6*)}] = (N/2)p^6 + (N/2)p^6(1-p)^6 + Np^4(1-p)^6 + 6Np^4(1-p)^6 + o(N)$ we get

$$\begin{aligned} \mathbb{E}[C_p] &\leq \frac{1}{2}N \left(p - \frac{9}{8}p^2 + \frac{9}{8}p^2(1-p)^4 + p(1-p)^3 + \frac{3}{2}p^3(1-p)^5 \right. \\ &\quad \left. + \frac{1}{8}p^6 + \frac{1}{8}p^6(1-p)^6 + \frac{1}{4}p^4(1-p)^6 + \frac{3}{2}p^4(1-p)^6 \right) + o(N) \\ &= Ng(p) + o(N). \end{aligned}$$

The function $g(p)$ attains its maximum on $[0, 1]$ at p'_{\max} approximately 0.36 and $g(p'_{\max}) < 0.17144$. \square

The following corollary summarizes the results for lattices.

Corollary 5. *Let τ^b be the optimal stopping time while playing a blind game and τ^f be the optimal stopping time while playing a full information game on a lattice with N vertices. Let $c^b = (1/N)\mathbb{E}[\tilde{C}_{\tau^b}]$ and $c^f = (1/N)\mathbb{E}[\tilde{C}_{\tau^f}]$. By Theorems 2, 3 and 4 together with Lemmas 8 and 9, for sufficiently large N , we have:*

<i>Lattice</i>	<i>Lower bound for c^b</i>	<i>Upper bound for c^b</i>	<i>gap</i>
<i>square</i>	0.12953	0.13268	0.00315
<i>triangular</i>	0.09629	0.10107	0.00478
<i>hexagonal</i>	0.16738	0.17144	0.00406

Furthermore, for every $\varepsilon \in (0, 1)$ and for sufficiently large N , by Theorem 1 we have:

$$c^b \leq c^f \leq c^b + \varepsilon.$$

5. FINAL COMMENTS AND QUESTIONS

Corollary 5 presents tight bounds for the expected number of components that an optimal blind strategy gains for each lattice and shows that with full information the gain is almost the same. It does not find the exact optimal stopping time, but one can easily verify that if the player (blindly) stops at time $\tau = \lfloor p_{\max} N \rfloor$, where p_{\max} is a value of p that maximizes one of the functions $f(p)$ used in the proofs for the lower bound in Theorems 2, 3 or 4, then, in the respective lattice, the value of $(1/N)\mathbb{E}[\tilde{C}_\tau]$ belongs to the interval given in Corollary 5.

Interpreting the proofs of Theorems 2, 3 and 4, one concludes that at the moment when the expected number of components is maximized, most faces are very small. However, if we take into account only the expected number of isolated vertices and isolates edges while counting components, we would get much worse lower bound for the expected number of components (say, of order $0.103N$ in the square lattice). This indicates that the number of small faces is indeed relevant. On the other hand, we tried to include slightly larger faces in our proof, what yielded a longer case analysis, but ended up with no significant improvements.

A natural open question is whether one can relax the condition about the maximum degree in Theorem 1, e.g., to graphs with the maximum degree of order $o(N)$. It is also of interest to determine for which graphs the optimal blind strategy returns the linear expected number of components. This would precisely define the family of graphs for which the bounds from Theorem 1 are useful.

Acknowledgements. This research was partially supported by Polish National Science Center - grant MINIATURA 3, DEC-2019/03/X/ST6/00657, by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - 88881.569474/2020-01, by CNPQ (Proc.437841/2018-9, Proc.314374/2018-3).

REFERENCES

- [1] Fabrício Siqueira Benevides and Małgorzata Sulkowska. Percolation and best-choice problem for powers of paths. *J. Applied Probability*, 54(2):343–362, 2017.
- [2] Daniel Berend and Aryeh Kontorovich. A sharp estimate of the binomial mean absolute deviation with applications. *Stat. Probabil. Lett.*, 83(4):1254–1259, 2013.
- [3] Colin R. Blyth. Expected absolute error of the usual estimator of the binomial parameter. *Am. Stat.*, 34(3):155–157, 1980.
- [4] Béla Bollobás and Oliver Riordan. *Percolation*. Cambridge University Press, 2006.
- [5] Thomas S. Ferguson. Who solved the secretary problem? *Statist. Sci.*, 4(3):282–289, 08 1989.
- [6] Ragnar Freij and Johan Wästlund. Partially ordered secretaries. *Electron. Commun. Probab.*, 15:504–507, 2010.
- [7] Bryn Garrod and Robert Morris. The secretary problem on an unknown poset. *Random Struct. Algor.*, 43(4):429–451, 2013.
- [8] Wayne Goddard, Ewa Kubicka, and Grzegorz Kubicki. An efficient algorithm for stopping on a sink in a directed graph. *Oper. Res. Lett.*, 41(3):238–240, 2013.
- [9] Geoffrey Grimmett. On the number of clusters in the percolation model. *J. London Math. Soc.*, s2-13(2):346–350, 1976.
- [10] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, 58(301):13–30, 1963.
- [11] Svante Janson and Andrzej Ruciński. The infamous upper tail. *Random Struct. Algor.*, 20(3):317–342, 2002.
- [12] Jakub Kozik. Dynamic threshold strategy for universal best choice problem. *Proceedings of 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms*, pages 439–452, 2010.

- [13] Grzegorz Kubicki and Michał Morayne. Graph-theoretic generalization of the secretary problem: The directed path case. *SIAM J. Discrete Math.*, 19(3):622–632, 2005.
- [14] Michał Lasoń. Optimal stopping for many connected components in a graph. *Random Struct. Algor.*, 59(2):267–287, 2021.
- [15] Dennis V. Lindley. Dynamic programming and decision theory. *Appl. Stat. - J. Roy. St. C*, 10(1):39–51, 1961.
- [16] Michał Morayne. Partial-order analogue of the secretary problem the binary tree case. *Discret. Math.*, 184(1-3):165–181, 1998.
- [17] John Preater. The best-choice problem for partially ordered objects. *Oper. Res. Lett.*, 25(4):187–190, 1999.
- [18] Wolfgang Stadje. Efficient stopping of a random series of partially ordered points. *Multiple Criteria Decision Making Theory and Application. Lecture Notes in Economics and Mathematical Systems*, 177:430–447, 1980.
- [19] Małgorzata Sulkowska. The best choice problem for upward directed graphs. *Discret. Optim.*, 9(3):200–204, 2012.
- [20] M.F. Sykes and John W. Essam. Exact critical percolation probabilities for site and bond problems in two dimensions. *J. Math. Phys.*, 5(8):1117–1127, 1964.

DEPARTAMENTO DE MATEMÁTICA, CAMPUS DO PICI, UNIVERSIDADE FEDERAL DO CEARÁ, BRAZIL.

Email address: `fabricio@mat.ufc.br`

WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY, DEPARTMENT OF FUNDAMENTALS OF COMPUTER SCIENCE, POLAND; UNIVERSITÉ CÔTE D'AZUR, CNRS, INRIA, I3S, FRANCE.

Email address: `malgorzata.sulkowska@pwr.edu.pl`