A new lower bound on the independence number of graphs

Eric Angel, Romain Campigotto, Christian Laforest

A New Lower Bound on the Independence Number of Graphs[☆]

Eric Angel[a], Romain Campigotto[b], Christian Laforest[c,*]

[a]*Laboratoire IBISC, EA 4526 – Université d'Évry-Val d'Essonne, IBGBI, 23 boulevard de France, 91 037 Évry Cedex, France*
[b]*LAMSADE, CNRS UMR 7243 – Université Paris-Dauphine, place du Maréchal de Lattre de Tassigny, 75 775 Paris Cedex 16, France*
[c]*LIMOS, CNRS UMR 6158 – Université Blaise Pascal, Clermont-Ferrand, Campus des Cézeaux, 24 avenue des Landais, 63 173 Aubière Cedex, France*

Abstract

We propose a new lower bound on the independence number of a graph. We show that our bound compares favorably to recent ones (e.g. [12]). We obtain our bound by using the *Bhatia-Davis* inequality applied with analytical results (minimum, maximum, expectation and variance) of an algorithm for the VERTEX COVER problem.

Keywords: independence number, graphs, variance, expectation, vertex cover

1. Introduction

The *independence number* of a graph is one of the most fundamental and well-studied graph parameters [14]. In view of its computational hardness [11], various bounds have been proposed [5, 7, 8, 15, 16]. Obtaining such bounds is an important and hard challenge. The first one was given independently by *Y. Caro* and *V. Wei* [6, 17]. The most recent is due to *J. Harant* [12].

In this paper, we propose a new lower bound for this parameter. We prove that our bound is often strictly better than the ones recently proposed in [13] and [12]. To obtain it, we use the *Bhatia-Davis* inequality (see [4]) applied with the parameters capturing the behavior of an algorithm to compute the well-known VERTEX COVER problem. Sect. 2 contains these preliminary analytical results. In Sect. 3, we prove our lower bound and compare it.

Definitions, Notations, Problems

Graphs $G = (V, E)$ considered throughout this paper are undirected, finite, simple and unweighted. We denote by n the number of vertices ($n = |V|$). For

[☆]Work partially supported by the French Agency for Research under the DEFIS project ANR-09-EMER-010 *TODO*.

*Corresponding author: Christian Laforest, LIMOS, Université Blaise Pascal, France. Email: christian.laforest@isima.fr

any vertex $u \in V$, we denote by $N(u)$ the set of *neighbors* of u (i.e. the set of vertices sharing an edge with u), $d(u) = |N(u)|$ the *degree* of vertex u and $\Delta = \max\{d(u) \mid u \in V\}$ the *maximum degree* of G.

The Vertex Cover Problem. The VERTEX COVER problem is a well-known classical **NP**-complete optimization graph problem [10]. A *cover* C of G is a subset of vertices such that every edge contains (or *is covered by*) at least one vertex of C, that is $C \subseteq V$ and $\forall e = uv \in E$, one has $u \in C$ or $v \in C$ (or both). The VERTEX COVER problem is to find a cover of minimum size.

The Independence Number of a Graph. A set of vertices $I \subseteq V$ in a graph G is called *independent* if the graph induced by vertices of I in G contains no edge. The *independence number* α of G is the largest cardinality $|I|$ among all independent sets I of G.

2. Preliminary Results

We focus in this section on the VERTEX COVER problem. More precisely, we analyze a simple algorithm, that we call LL: for any graph G, we express the size of the best and the worst solutions it can produce; we also express the expectation and the variance of the sizes. These results will be used in Sect. 3 to express our main result, namely our lower bound on independence number of any graph (thanks to the *Bhatia-Davis* inequality).

Labeling of Nodes, Left and Right Neighbors. In a *labeled graph*, denoted by $G = (V, L, E)$, the n vertices of G are labeled by a given function L (called *labeling*) such that each vertex $u \in V$ has a unique *label* $L(u) \in \{1, \dots, n\}$ (L is a bijection between V and $\{1, \dots, n\}$). We denote by $\mathbb{L}(G)$ the set of all possible labelings L for a graph $G = (V, E)$. Given a labeled graph $G = (V, L, E)$ and a vertex $u \in V$, v is called a *right neighbor* (resp. *left neighbor*) of u if $v \in N(u)$ and if v has a label larger (resp. smaller) than u.

The algorithm LL can be described as follows on any labeled graph $G = (V, L, E)$. For each vertex $u \in V$, u is added to the cover if it has at least one *right neighbor*.

It is easy to see that LL constructs a cover for any graph G, given any labeling L of G. Of course, the solution (and size) it returns strongly depends on L.

Property 1 ([2]). *For any graph G:*

1. *There exists $L^* \in \mathbb{L}(G)$ such that LL returns an optimal solution on $G = (V, L^*, E)$.*
2. *LL cannot construct a cover of size greater than $n - c$ (where is c the number of connected components of G).*
3. *With equiprobability on the labelings of G with n vertices, we have:*
   (a) *The probability that a vertex u of G is in the solution is $1 - \frac{1}{d(u)+1}$.*

(b) *The expected size of the solution produced by* LL *on* G *is*

$$\mathbb{E}\left[\mathsf{LL}(G)\right] \quad = \quad n - \sum_{u \in V} \frac{1}{d(u)+1} \ . \tag{1}$$

All these properties are easy and are proved in [2]. To be complete, we recall the ideas. Let L^* labeling first the vertices of any optimal vertex cover C^* and after all the other vertices. Clearly LL applied on L^* returns C^*. For the second property, suppose that $c = 1$. In any labeling, the last labeled vertex will not be chosen, producing a solution of size at most $n-1$ (generalization to any c is trivial).

The probability $1 - \frac{1}{d(u)+1}$ is the complement of the result shown in the famous proof of *Y. Caro* and *V. Wei* for the *Turan's theorem* (see [1]), since the complement of any independent set is a vertex cover (and vice-versa). The expectation $\mathbb{E}\left[\mathsf{LL}(G)\right]$ is then easily obtained from the previous probability and the linearity of the expectation.

We complete these results by the *variance* of LL in what follows.

Theorem 1. *Let* $G = (V, E)$ *be any graph. By considering the* $n!$ *labelings of* $\mathbb{L}(G)$ *with equiprobability assumption, the variance of* LL *is*

$$\begin{aligned} \mathbb{V}\left[\mathsf{LL}(G)\right] \quad = \quad & \sum_{u \in V} \frac{d(u)}{(d(u)+1)^2} - 2 \sum_{uv \in E} \frac{1}{(d(u)+1)(d(v)+1)} \\ & + \ 2 \sum_{uv \notin E} \frac{d_{uv}}{(d(u)+1)(d(v)+1)(2+d(u)+d(v)-d_{uv})} \ , \end{aligned} \tag{2}$$

with $d_{uv} = |\{w \in V \mid uw \in E \wedge wv \in E\}|$.

Proof. We can express $\mathbb{V}\left[\mathsf{LL}(G)\right]$ by a sum of covariances (for more details, see e.g. Chap. 9 of [9]). Let C be a cover returned by algorithm LL. For each vertex $u \in V$, we associate a random variable $\mathbb{1}_u$ such that $\mathbb{1}_u = 1$ if $u \in C$ and 0 otherwise. Let $X = \sum_{u \in V} \mathbb{1}_u$ be the random variable giving the size of C. We have

$$\mathbb{V}\left[X\right] \quad = \quad \sum_{u \in V} Cov(\mathbb{1}_u, \mathbb{1}_u) + 2 \sum_{\substack{\{u,v\} \subseteq V \\ u<v}} Cov(\mathbb{1}_u, \mathbb{1}_v) \ , \tag{3}$$

where $u < v$ means that we count only once $Cov(\mathbb{1}_u, \mathbb{1}_v)$ in the sum;
$Cov(\mathbb{1}_u, \mathbb{1}_v) = \mathbb{E}\left[\mathbb{1}_u \mathbb{1}_v\right] - \mathbb{E}\left[\mathbb{1}_u\right] \cdot \mathbb{E}\left[\mathbb{1}_v\right]$ is the *covariance* of random variables $\mathbb{1}_u$ and $\mathbb{1}_v$. With Property 1, we get

$$\mathbb{E}\left[\mathbb{1}_u\right] \quad = \quad 1 - \frac{1}{d(u)+1} \ . \tag{4}$$

Now, by using (4), we can give closed formulas for each term of (3).

First, as we have $Cov(\mathbb{1}_u, \mathbb{1}_u) = \mathbb{E}\left[\mathbb{1}_u^2\right] - \mathbb{E}\left[\mathbb{1}_u\right]^2 = \mathbb{E}\left[\mathbb{1}_u\right] - \mathbb{E}\left[\mathbb{1}_u\right]^2$, we obtain

$$Cov(\mathbb{1}_u, \mathbb{1}_u) = 1 - \frac{1}{d(u)+1} - \left(1 - \frac{1}{d(u)+1}\right)^2 = \frac{d(u)}{(d(u)+1)^2} \quad . \tag{5}$$

Now, for $Cov(\mathbb{1}_u, \mathbb{1}_v)$, we distinguish the case where $uv \in E$ and the one where $uv \notin E$.

When $uv \in E$, at least u or v is in the solution. So $\mathbb{1}_u\mathbb{1}_v = 1$ if the "second" vertex is also in the solution. If $L(u) < L(v)$ (resp. $L(v) < L(u)$) then u (resp. v) is in the solution and v (resp. u) is *not* in the solution with probability $\frac{1}{d(v)+1}$ (resp. $\frac{1}{d(u)+1}$): see Property 1. Hence, both vertices u and v are in the solution (and $\mathbb{1}_u\mathbb{1}_v = 1$) with probability $1 - (\frac{1}{d(u)+1} + \frac{1}{d(v)+1})$ ($L(u) < L(v)$ and $L(v) < L(u)$ are two disjoints cases). As $\mathbb{1}_u\mathbb{1}_v = 0$ or 1, we get

$$\mathbb{E}\left[\mathbb{1}_u\mathbb{1}_v\right] = 1 - \frac{1}{d(u)+1} - \frac{1}{d(v)+1} \quad .$$

Thus, if $uv \in E$,

$$\begin{aligned} Cov(\mathbb{1}_u, \mathbb{1}_v) &= 1 - \frac{1}{d(u)+1} - \frac{1}{d(v)+1} - \left(1 - \frac{1}{d(u)+1}\right)\left(1 - \frac{1}{d(v)+1}\right) \\ &= \frac{-1}{(d(u)+1)(d(v)+1)} \quad . \end{aligned} \tag{6}$$

We treat now the case where $uv \notin E$. We prove later that

$$\mathbb{E}\left[\mathbb{1}_u\mathbb{1}_v\right] = 1 - \frac{(d(u)+d(v)-d_{uv}+1)(d(u)+d(v)+2)}{(d(u)+1)(d(v)+1)(d(u)+d(v)-d_{uv}+2)} \quad . \tag{7}$$

Before making the proof of (7), suppose first that the result is correct. In this case, when $uv \notin E$, we have

$$\begin{aligned} Cov(\mathbb{1}_u, \mathbb{1}_v) &= 1 - \frac{(d(u)+d(v)-d_{uv}+1)(d(u)+d(v)+2)}{(d(u)+1)(d(v)+1)(d(u)+d(v)-d_{uv}+2)} \\ &\quad - \left(1 - \frac{1}{d(u)+1}\right)\left(1 - \frac{1}{d(v)+1}\right) \\ &= \frac{d_{uv}}{(d(u)+1)(d(v)+1)(d(u)+d(v)-d_{uv}+2)} \quad . \end{aligned} \tag{8}$$

Therefore, by replacing the expressions of covariances of (5), (6) and (8) in the formula (3), we obtain (2).

Hence, to complete the proof of Theorem 1, we just need now to prove (7), with the hypothesis that $uv \notin E$.

By definition here, $\mathbb{E}\left[\mathbb{1}_u\mathbb{1}_v\right] = \mathbb{P}\left[u \in C \wedge v \in C\right]$ (where C designates a solution produced by LL). By decomposing the sub-cases of the complementary probability, this expectation is also equal to

$$1 - \left(\mathbb{P}\left[u \notin C\right] + \mathbb{P}\left[v \notin C\right] - \mathbb{P}\left[u \notin C \wedge v \notin C\right]\right) \quad .$$

We already saw that $\mathbb{P}[u \notin C] = \frac{1}{d(u)+1}$ and similarly $\mathbb{P}[v \notin C] = \frac{1}{d(v)+1}$.

Let us compute now $\mathbb{P}[u \notin C \wedge v \notin C]$. We consider the case where $L(v) < L(u)$ (the other is symmetric). To get the result, we just need to consider the relative numberings of u, v and their $d(u)+d(v)-d_{uv}$ neighbors (the labels of the others vertices play no role here on the fact that u or v be in C). As $L(v) < L(u)$, u must have the highest of these $d(u)+d(v)-d_{uv}+2$ labelings (otherwise, u or v is in C). Choose $d(v)+1$ of them, fix v at the first chosen position (i.e. it gets the highest label), and assign the remaining $d(v)$ neighbors of v arbitrarily to the remaining $d(v)$ chosen positions. Finally, assign the $d(u)-d_{uv}$ vertices (which are neighbors of u but not of v) arbitrarily in the remaining $d(u)-d_{uv}$ positions. This gives

$$\binom{d(u)+d(v)-d_{uv}+1}{d(v)+1} \cdot d(v)! \cdot (d(u)-d_{uv})! \;=\; \frac{(d(u)+d(v)-d_{uv}+1)!}{d(v)+1}$$

cases. Replacing $d(u)$ by $d(v)$ gives the symmetric result when $L(v) > L(u)$. Hence, dividing by the $(d(u)+d(v)-d_{uv}+2)!$ possible (restricted) labelings, we obtain

$$\begin{aligned}
\mathbb{P}[u \notin C \wedge v \notin C] &= \frac{1}{(d(u)+d(v)-d_{uv}+2)}\left(\frac{1}{d(u)+1}+\frac{1}{d(v)+1}\right)\\
&= \frac{d(u)+d(v)+2}{(d(u)+d(v)-d_{uv}+2)(d(u)+1)(d(v)+1)}\ .
\end{aligned}$$

Combining all these results, we get $\mathbb{E}[\mathbb{1}_u \mathbb{1}_v]$ that is

$$1-\frac{1}{d(u)+1}-\frac{1}{d(v)+1}+\frac{d(u)+d(v)+2}{(d(u)+d(v)-d_{uv}+2)(d(u)+1)(d(v)+1)}\ .$$

Easy simplifications lead to (7) and ends the proof of Theorem 1. □

3. A New Lower Bound for the Independence Number

We can now use the result of *R. Bhatia* and *C. Davis* [4] linking the variance ϑ, the expected value μ, the minimum m and the maximum M of any random variable (with bounded distribution) by the formula $\vartheta \leq (M-\mu)(\mu-m)$. Transforming this formula we get $m \leq \mu - \frac{\vartheta}{M-\mu}$. With the help of Property 1, we replace ϑ by $\mathbb{V}[\mathsf{LL}(G)]$, μ by $\mathbb{E}[\mathsf{LL}(G)]$, m by OPT (the size of the optimal vertex cover of G) and M by $n-c$ to get

$$OPT \quad \leq \quad \mathbb{E}[\mathsf{LL}(G)] - \frac{\mathbb{V}[\mathsf{LL}(G)]}{n-c-\mathbb{E}[\mathsf{LL}(G)]}\ .$$

Now, as algorithm LL constructs a cover C of $G=(V,E)$, it is well known (and not difficult to prove) that $I = V - C$ is an independent set of G. Hence,

$\alpha = n - OPT$. Using this with the previous inequality, we obtain our lower bound on α:

$$\alpha \quad \geq \quad n - \left(\mathbb{E}\left[\mathsf{LL}(G)\right] - \frac{\mathbb{V}\left[\mathsf{LL}(G)\right]}{n - c - \mathbb{E}\left[\mathsf{LL}(G)\right]} \right) \ . \tag{9}$$

Note that this bound can be calculated for any graph, except when $n - c = \mathbb{E}[\mathsf{LL}(G)]$. In this case, this means that the average size of covers constructed by LL is the worst one, of size $n-c$. As LL can construct the optimal one, this means that $OPT = \mathbb{E}[\mathsf{LL}(G)] = n - c$ (otherwise, we would have $\mathbb{E}[\mathsf{LL}(G)] < n - c$). As $\alpha = n - OPT$, and as $OPT = n - c$ here, our lower bound becomes:

$$\text{if } \mathbb{E}\left[\mathsf{LL}(G)\right] \ = \ n - c \qquad \text{then} \qquad \alpha \ \geq \ n - (n - c) = c \ . \tag{10}$$

With this additional case (10), our bound can be computed in polynomial time for any graph G (c is the number of connected components of G and $\mathbb{E}[\mathsf{LL}(G)]$ can be computed with the formula of Property 1 and $\mathbb{V}[\mathsf{LL}(G)]$ with the formula of Theorem 1).

In the following of this section, we compare our lower bound to the recent one of [12] by *J. Harant* and to a previous one [13] proposed by *J. Harant* and *I. Schiermeyer*. We prove analytically in Sect. 3.1 that on at least one infinite family of graphs (the wheel graphs with even number of vertices), the difference between our bound and the one of [12] tends to infinity when the number of vertices grows, meaning that on this family our bound gets better when n grows. In Sect. 3.2, we compare the bounds by calculating them numerically on hundreds of graphs. We show that our bound is better on random graphs and random bipartite graphs.

3.1. Comparison of Lower Bounds on Independence Number on Wheel Graphs

In this section, we compare our lower bound (9) to the one of Harant in [12] in a family of graphs, namely the wheels. The *wheel graph* $W(n)$ has n vertices and is composed of a cycle of $n - 1$ vertices and an additional "central" vertex directly connected to all the $n - 1$ other ones. We prove in this section that when n (restricted here to even n for simplifications) tends to infinity, our lower bound minus the Harant's lower bounds tends to infinity, proving that our lower bound is greater and thus better on these graphs.

To do that, we need the bound of *J. Harant*. Here is the exact formulation of his theorem as it can be founded in [12].

Theorem 2 (Theorem 2 in [12]). *Let $G = (V, E)$ be a finite, simple, connected and non-complete graph on $n \geq 3$ vertices, where $|E| \geq n$. Moreover, let $\alpha \leq \frac{n}{2}$ be the independence number, Δ be the maximum degree of G, n_j be the number of vertices of degree j in G and*

$$\begin{aligned} x(j) \ = \quad & \frac{j(j+1)}{j(j+1)-2} \left(\left(\frac{2}{j+1} - (\Delta - j) \right) n_\Delta + \cdots + \left(\frac{2}{j+1} - 1 \right) n_{j+1} \right. \\ & + \left. \frac{2n_j}{j+1} + \cdots + \frac{2n_1}{2} - 2 \right) \end{aligned}$$

for $j \in \{2, \ldots, \Delta\}$*. Then, there is a unique* $j_0 \in \{2, \ldots, \Delta\}$ *such that* $0 \leq x(j_0) < n_\Delta + \cdots + n_{j_0}$ *and*

$$\alpha \quad \geq \quad \frac{x(j_0) + n_{j_0+1} + 2n_{j_0+2} + \cdots + (\Delta - j_0)n_\Delta}{2} + 1 \ .$$

To obtain the following results, we used Maple©software since these are tedious elementary formal calculations that are done easily and safely by a computer.

Calculation of the Bound of Harant $H(2p)$ *on* $W(2p)$*.* We can note that Theorem 2 can be applied here since all the hypothesis are satisfied. To compute the bound of Harant on $W(n)$, that we note $H(n)$, we need to detail parameters used in the theorem. To simplify the calculations, we restrict ourselves to the case where n in even, and we write $n = 2p$. In $W(2p)$, $\Delta = n-1 = 2p-1$, there are $n-1$ vertices of degree 3 and one vertex of degree $n-1$. Thus, $n_3 = 2p-1$, $n_\Delta = n_{2p-1} = 1$ and all the other n_i are 0.

The next step is to find the index j_0. We claim that $j_0 = p+2$. To prove that, we just have to show that $0 \leq x(j_0) < n_\Delta + \cdots + n_{j_0}$. As $j_0 > 3$ (when p is large enough), $n_\Delta + \cdots + n_{j_0} = n_\Delta = 1$. We get $x(j_0) = \frac{1}{2}\frac{(7+p)(p+2)}{p^2+5p+4}$. Clearly, $x(j_0) \geq 0$. We have to prove that $x(j_0) < 1$. This is true iff $(7+p)(p+2) - 2p^2 - 10p - 8 < 0$. After simplifications, this is equivalent to $-p^2 - p + 6 < 0$. When p is large enough, this is clearly true.

Hence, the (unique) index j_0 of Theorem 2 is $p+2$. We can now write the formula of the lower bound $H(2p)$ of Harant on $W(2p)$:

$$H(2p) \quad = \quad \frac{(x(j_0) + 2p - 1 - p - 2)}{2} + 1 \ = \ \frac{1}{4}\,\frac{7p + 6 + 9p^2 + 2p^3}{p^2 + 5p + 4} \ .$$

Calculation of our Bound $U(2p)$. We do the same kind of work to get the lower bound given by our method in (9). To do that, we need to calculate the expectation $\mathbb{E}\left[\mathsf{LL}(W(2P))\right]$, noted here $Exp(2p)$ to simplify, and the variance on $W(2p)$, noted here $Var(2p)$. We use here again the fact that $\Delta = n-1 = 2p-1$, and that there is one vertex of degree $n-1$ and $n-1$ vertices of degree 3. After calculations with (1) and Theorem 1 and reorganizations and several simplifications (using Maple©again), we get

$$Exp(2p) = \frac{3p}{2} + \frac{1}{4} - \frac{1}{2p} \quad \text{and } Var(2p) = \frac{(2p)^2}{112} + \frac{2p}{24} + \frac{3}{2(2p)} - \frac{1}{(2p)^2} - \frac{185}{336} \ .$$

Using (9), we can now express our lower bound, noted $U(2p)$:

$$U(2p) \quad = \quad 2p - \left(Exp(2p) - \frac{Var(2p)}{n - 1 - Exp(2p)}\right) \ = \ \frac{2}{21}\,\frac{(6p-11)p}{p-2} \ .$$

Comparison of the Two Bounds $H(2p)$ *and* $U(2p)$*.* After simplifications, we obtain $U(2p) - H(2p) = \frac{1}{84}\frac{6p^4+47p^3-17p^2-184p+252}{(p-2)(p^2+5p+4)}$.

From that, we get $\lim_{p\to+\infty}(U(2p) - H(2p)) = +\infty$. This proves that there is an infinite number of graphs for which our bound is better than Harant's one and that the difference increases.

3.2. Numerical Comparisons of Lower Bounds on Independence Number on Random Graphs

In Sect. 3.1, we proved by analytical arguments that our lower bound is strictly better than the one of [12] and that the difference tends to infinity for at least a particular family of graphs. On others graphs, like cycles, it can be shown that the lower bound of Harant is better than ours.

As studying, even restricted class of graphs, is a long and hard analytical task, we decided to compare the bounds numerically on several graphs.

Bounds Compared. We compared our bound with the ones of [12] (see Theorem 2) and [13] that has an analytical expression similar to Theorem 2. We do not include it here but we can note that the hypothesis given by the authors to apply it is only that G must be connected and non-complete. No other hypothesis is required. At the opposite, Theorem 2 is valid only if the independence number α is less than or equal to $\frac{n}{2}$. We must keep this point in mind when we make calculations on random graphs, since we cannot be sure that this constraint is satisfied a priori.

We programed these three bounds using Maple©. We calculated the integer values of these bounds (for example, the bound of Theorem 2 calculated is $\left\lceil \frac{x(j_0)+n_{j_0+1}+2n_{j_0+2}+\cdots+(\Delta-j_0)n_\Delta}{2}+1 \right\rceil$).

Results Obtained. We calculated the bounds on random graphs produced by Maple©(library `RandomGraphs`) using the option `connected` to be sure to work on connected graphs. As we cannot be sure that $\alpha \leq \frac{n}{2}$, we computed Theorem 2: if the index j_0 exists then we calculated the final value; otherwise, we cannot use the formula. Note that even if we obtain a result, we are not sure that it is valid (because maybe $\alpha > \frac{n}{2}$ and thus we are not in the validity conditions).

For each experiment, we randomly generated an integer n in a range R_n (typically $[5, 250]$) and a probability p in a range R_p (e.g. $[0.05, 0.95]$). Then we generated a connected random graph G with n vertices and probability p for each possible edge[1]. We computed the three bounds on G and compared them.

We evaluated more than 1500, with different ranges of vertices (but mostly $[5, 250]$) and different ranges of probabilities R_p ($[0.05, 0.95]$ but also "dense" graphs $[0.5, 0.95]$ and "sparse" graphs $[0.05, 0.15]$). On these 1500 calculations, the bound of Theorem 2 "failed" 3 times (no parameter j_0 found). In 561 cases, our bound is the same than the best of [12] and [13]. In 935 cases, our bound is better than the best of the two others by one unit and in 4 cases, our bound is one unit worst than the best of the two others.

Thus, in 99.73% cases, our bound is at least as good as the others and in 62.33%, it is strictly better.

We describe now other numerical comparisons on random *bipartite* graphs. For that, we generated randomly two integers a and b in a range $[5, 125]$ and a

[1] Maple© generates a random tree and then generates random edges with probability p.

probability p in a given range. Then we generated random bipartite connected graphs with one part V_1 of the bi-partition having a vertices and the second one V_2 having b vertices (hence, the whole graphs contains between 10 and 250 vertices); edges (between V_1 and V_2 only) are randomly generated with probability p. Here we only compared our bound with the one of [13] (in too many cases, Theorem 2 could not be computed, probably because of the α parameter). We did 200 comparisons here. In 100% of the cases, our bound is at least as good as the other. In 89.5%, our bound is strictly better. The maximum of difference between the two bounds is 63 units (on one experiment). In average, our bound in better by 6.14 units.

We also made computations on wheel graphs analytically studied in Sect. 3.1. They confirm our analysis (even on wheels with an odd number of vertices).

4. Conclusion

In a first part of this paper, we gave an analytical formula expressing the variance of an algorithm (called LL) for the VERTEX COVER problem. As the others parameters (min., max. and average sizes) were already been calculated in [2], we refine the knowledge of the behavior of LL. We can note that all these bounds are exact and can be polynomially calculated on any graph G. It is an important result in itself since we showed in [2] that LL is adapted to treat huge graphs. We also experimented LL in [3] and shown that it is able to construct a solution on graphs of up to 100.10^9 vertices and edges in around 7 hours on a laptop. For such huge graphs, the precise knowledge of the statistical behavior of the algorithm is important.

In a second part, we used these results to propose a new lower bound on the independence number of a graph. We compared our bound to the recent ones proposed in [13] and [12]. We showed that for an infinite number of graphs, our bounds is strictly better. We also compared it numerically on hundreds of random graphs.

From these results, we can conclude that: our bound can always be calculated, for any graph, without restrictions (this is not the case of the ones of [13] and [12]); for many graphs, our bounds is at least as good as the two others and often strictly better.

A future work is to improve our lower bound by applying the *Bhatia-Davis* inequality with other algorithms for which we can express (exactly or approximatively) min., max., average sizes and variance.

References

[1] N. Alon and J. H. Spencer. *The Probabilistic Method, second edition.* John Wiley & Sons, Inc., New York, 2008.

[2] E. Angel, R. Campigotto, and C. Laforest. Analysis and Comparison of Three Algorithms for the Vertex Cover Problem on Large Graphs with Low Memory Capacities. *Algorithmic Operations Research*, 6(1):56–67, 2011.

[3] E. Angel, R. Campigotto, and C. Laforest. Implementation and Comparison of Heuristics for the Vertex Cover Problem on Huge Graphs. In *11th International Symposium on Experimental Algorithms (SEA)*, volume LNCS 7276, pages 39–50, 2012.

[4] R. Bhatia and C. Davis. A Better Bound on the Variance. *American Mathematical Monthly (Math. Assoc. of America)*, 107(4):353–357, 2000.

[5] A. Bossecker and D. Rautenbach. Interpolating between Bounds on the Independence Number. *Discrete Mathematics*, 310(17–18):2398–2403, 2010.

[6] Y. Caro. New Results on the Independence Number. Technical report, Tel-Haviv University, 1979.

[7] D. A. Cartwright, M. A. Cueto, and E. A. Tobis. The Maximum Independent Sets of de Bruijn Graphs of Diameter 3. *The Electronic Journal of Combinatorics*, 18(1), 2011.

[8] F. R. K. Chung. The Average Distance and the Independence Number. *Journal of Graph Theory*, (2):229–235, 1988.

[9] W. Feller. *An Introduction to Probability Theory and its Applications, Vol. 1, third edition*. John Wiley & Sons, Inc., 1968.

[10] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, 1979.

[11] J. Håstad. Clique is Hard to Approximate Within $n^{1-\epsilon}$. In *37th Annual Symposium on Foundations of Computer Science*, pages 627–636, 1996.

[12] J. Harant. A Lower Bound on Independence in Terms of Degrees. *Discrete Applied Mathematics*, 159:966–970, 2011.

[13] J. Harant and I. Schiermeyer. A Lower Bound on Independence Number of a Graph in Terms of Degrees. *Discussiones Mathematicae Graph Theory*, 26:431–437, 2006.

[14] V. Lozin and D. de Werra. Foreword: Special Issue on Stability in Graphs and Related Topics. *Discrete Applied Mathematics*, 132:1–2, 2003.

[15] D. Mubayi and J. Williford. On the Independence Number of the Erdős-Rényi and Projective Norm Graphs and a Related Hypergraph. *Journal of Graph Theory*, 56(2):113–127, 2007.

[16] S. M. Selkow. A Probabilistic Lower Bound on the Independence Number of Graphs. *Discrete Mathematics*, 132(1-3):363–365, 1994.

[17] V. K. Wei. A Lower Bound on the Stability Number of a Simple Graph. Technical Report 81-11217-9, Bell Laboratories, Murray Hill, NJ, 1981.