Journal of Combinatorics
Volume 8, Number 2, 323–340, 2017

On the geodetic rank of a graph

Mamadou Moustapha Kanté, Rudini M. Sampaio,
Vinícius F. dos Santos, and Jayme L. Szwarcfiter

A graph convexity is a finite graph G, together with a family of subsets $\mathcal{C}$ of its vertices, such that $\emptyset, V(G) \in \mathcal{C}$, and $\mathcal{C}$ is closed under intersections. The members of $\mathcal{C}$ are called convex sets. The graph convexity is geodetic when its convex sets are closed under shortest paths. For a subset $S \subseteq V(G)$, the smallest convex set containing S, denoted by $H(S)$, is the hull of S. On the other hand, S is convexly independent when $v \notin H(S \setminus \{v\})$, for any $v \in S$. The rank of G is the cardinality of its largest convexly independent set. In this paper, we consider complexity aspects of the determination of the rank in the geodetic convexity. Among the results, we prove that it is NP-hard to approximate the geodetic rank of bipartite graphs by a factor of $n^{1-\varepsilon}$, for every $\varepsilon > 0$. On the other hand, we describe polynomial time algorithms for finding the rank of P_4-sparse graphs and split graphs. Also, by applying monadic second-order logic we obtain further complexity results, including a linear time algorithm for determining the rank of a distance-hereditary graph. Some of the results obtained are extended to other graph convexities.

AMS 2000 subject classifications: Primary 05C85, 68R10; secondary 68Q25.
Keywords and phrases: Bipartite graphs, geodetic convexity, inapproximability, monophonic convexity, NP-completeness, P_3-convexity, P_3^*-convexity, rank.

1. Introduction

Convexity spaces have been considered in different branches of mathematics [32]. The study of convexities applied to graphs has started more recently, about 50 years ago [22, 29, 31]. Abstract convexity parameters, when considered on graph convexities [23], give rise to interesting graph parameters. In particular, complexity aspects related to the computation of these parameters are the main goal of various recent papers.

The computation of convexity parameters for a graph depends on the particular convexity being considered. The most common graph convexities so far considered are those whose convex sets are based on some family of paths $\mathcal{P}$ of the graph, known as path convexities, where a set is convex if it is closed for paths of $\mathcal{P}$. In this context, the convexity which has been most considered is the geodetic convexity, which is precisely the graph convexity whose convex sets are closed under shortest paths, that is, $\mathcal{P}$ is the collection of all shortest paths of the graph. See, for instance, [3, 7, 13, 24]. There are some other types of path convexities, as the monophonic convexity [8, 19, 21], P_3-convexity [9, 10], P_3^*-convexity [2], m^3-convexity [20], triangle-path convexity [12, 11]. They are defined by letting the convex sets be closed under induced paths, paths of order 3, induced paths of order 3, induced paths of length at least three, and paths whose only possible chords are those at distance two in the path, respectively.

In the present paper, we are mainly concerned with the computation of the rank of a graph. Complexity aspects of the rank of a graph convexity have been considered in [30] on the P_3-convexity and on the monophonic convexity where the problem is shown to be NP-complete. As a corollary of a result of [30], it is also NP-complete for bipartite graphs on the P_3^*-convexity. In the rest of this paper we are mainly interested in the geodetic rank. We show that finding such a parameter leads to an NP-hard problem. Moreover, it is NP-hard even to find an approximation for the rank of a bipartite graph, within a factor of $n^{1-\varepsilon}$, for every $\varepsilon > 0$. In contrast, we describe polynomial time algorithms for the geodetic rank of P_4-sparse graphs, with consequences for the rank on the P_3^*-convexity, and of split graphs. In addition, by employing monadic second-order logic, we describe results concerning polynomial time determination of the rank of graphs having fixed clique-width, in some graph convexities. As a consequence, we show that both the geodetic rank and the monophonic rank of a distance-hereditary graph can be computed in linear time.

2. Preliminaries

For graph theoretical concepts and terminology, see the book by Bondy and Murty [6].

Let G be a simple finite graph, with vertex set $V(G)$ and $\mathcal{C}$ a family of subsets of $V(G)$. The pair $(G, \mathcal{C})$ is a *graph convexity*, when $\emptyset \in \mathcal{C}$, $V(G) \in \mathcal{C}$ and, if $S_1, S_2 \in \mathcal{C}$, then $S_1 \cap S_2 \in \mathcal{C}$. The subsets $C \in \mathcal{C}$ are called *convex sets*. The *convex hull* of a subset $S \subseteq V(G)$ with respect of a graph G and a convexity $\mathcal{C}$, denoted by $H_{\mathcal{C},G}(S)$, is the smallest convex set which contains S. When the convexity and graph being considered are clear from the context, we will omit the subscript. If $H(S) = V(G)$, we say that S is a *hull set*.

Let G be a graph and $\mathcal{P}$ be a family of paths in G. Given a subset $S \subseteq V(G)$, let $I(S)$ be the set of vertices belonging to the paths of $\mathcal{P}$ between two vertices of S. In this paper, we are mainly concerned with $\mathcal{P}$ being all shortest paths (corresponding to the geodetic convexity or g-convexity for short), and also induced paths of length 2 (corresponding to the P_3^*-convexity), respectively.

There are many different graph parameters. In general, their computation strongly depends on the particular convexity under consideration. One of the parameters, so far most studied, is the hull number. For a graph convexity $(G, \mathcal{C})$, the *hull number* of G is the least cardinality hull set of G. The computation of this parameter has been considered in some different papers, as [1, 17, 18, 23, 27].

Another important convexity parameter is the rank of a graph [32]. For some graph convexity $(G, \mathcal{C})$, we say that a subset $S \subseteq V(G)$ is *convexly independent* when $v \notin H(S \setminus \{v\})$, for each $v \in S$. The *rank* of G, denoted by $rk_{\mathcal{C}}(G)$, is the cardinality of its largest convexly independent set. In the following we denote respectively by $rk_g(G)$, $rk_P(G)$, $rk_{P_3}(G)$ and $rk_{P_3^*}(G)$ the rank of G with respect to the following convexities: geodetic convexity, monophonic convexity, paths of order 3 and induced paths of order 3.

3. Computational complexity of the geodetic rank

In this section, we prove that the geodetic rank is NP-Complete even in bipartite graphs of diameter 3. In fact, we prove a stronger statement.

Theorem 1. *For every $\varepsilon > 0$, approximating the geodetic rank of a bipartite graph by a factor $n^{1-\varepsilon}$ is NP-hard.*

In order to prove this theorem, we need to present some basic results on approximation algorithms and reductions, following the terminology of Ausiello et al. [4] and Crescenzi [16].

Given an optimization problem P, let $opt_P(I)$ denote the optimal solution value for some instance I of P and, for a solution S of I, let $val_P(I, S)$ denote the associated value. Given an instance I of P and a solution S of I, the *performance ratio* $\mathcal{R}_P(I, S)$ is defined by

$$\mathcal{R}_P(I,S) \ = \ \max\left\{\frac{opt_P(I)}{val_P(I,S)}, \frac{val_P(I,S)}{opt_P(I)}\right\}.$$

Given a constant $r \geq 1$, an *r-approximation algorithm* for P is an algorithm that, applied to any instance I of P, runs in time polynomial in the size of I and produces a solution S such that $\mathcal{R}(I, S) \leq r$. If such an algorithm exists for a constant r, then we say that P belongs to APX.

We say that a problem P is *r-inapproximable in polynomial time* if there is no r-approximation polynomial time algorithm for P. Given a function $r(n)$, we say that a problem P is *$O(r(n))$-inapproximable in polynomial time* if there is a function $r'(n) = O(r(n))$ such that the problem P is $r'(n)$-inapproximable in polynomial time.

A reduction from P_1 to P_2 consists of a pair (f, g) of polynomial-time computable functions such that, for any instance I of P_1, (a) $f(I)$ is an instance of P_2, and (b) $g(I, S)$ is a feasible solution of I, for any feasible solution S of $f(I)$.

A *continuous reduction* from P_1 to P_2 is a 3-tuple (f, g, γ), where (f, g) is a reduction from P_1 to P_2 and $\gamma \geq 1$ is a constant, such that, if $\mathcal{R}_{P_2}(f(I), S) \leq r$ $(r \geq 1)$, then $\mathcal{R}_{P_1}(I, g(I, S)) \leq \gamma r$ for each instance I of P_1 and for every feasible solution S of $f(I)$. From this definition, if there is a polynomial time r-approximation algorithm for P_2 for some $r \geq 1$, then there is a polynomial time γr-approximation algorithm for P_1. Consequently, if P_1 is $O(r(n))$-inapproximable in polynomial time, then P_2 is also $O(r(n))$-inapproximable in polynomial time, where $r(n)$ is a function such that $\lim_{n\to\infty} r(n) = \infty$.

Proof of Theorem 1. We obtain a continuous reduction from the Set Packing Problem. Given a family $\mathcal{S} = \{S_1, \ldots, S_m\}$ of finite sets, the objective of the Set Packing Problem is to determine the size of a maximum set packing of $\mathcal{S}$, which is a family of mutually disjoint sets of $\mathcal{S}$. Given an instance $\mathcal{S} = \{S_1, \ldots, S_m\}$ of the Set Packing Problem, it is easy (polynomial) to check if $opt(\mathcal{S}) < 4$ (just check all subfamilies with 4 sets $S_{k_1}, S_{k_2}, S_{k_3}, S_{k_4}$). We will then consider instances $\mathcal{S}$ such that $opt(\mathcal{S}) \geq 4$.

Given an instance $\mathcal{S} = \{S_1, \ldots, S_m\}$ of Set Packing, we construct a bipartite graph $f(\mathcal{S}) = G$, which will be an instance of the geodetic rank problem, as follows. Let $\{a_1, \ldots, a_n\} = S_1 \cup \ldots \cup S_m$. Create two vertices z_1 and z_2. For every $i \in [n]$, create vertices a'_i and a''_i, and create the edges $a'_i z_1$, $a''_i z_1$, $a'_i z_2$ and $a''_i z_2$. For every $k \in [m]$, create a vertex S'_k. Create a vertex W, edges $z_1 W, z_1 W$ and, for every $k \in [m]$, create the edge $S'_k W$. If $a_i \in S_k$, create the edges $a'_i S'_k$ and $a''_i S'_k$. Let $A' = \{a'_1, \ldots, a'_n\}$, $A'' = \{a''_1, \ldots, a''_n\}$ and $\mathcal{S}' = \{S'_1, \ldots, S'_m\}$. This constructed graph G is clearly bipartite, with bipartitions $A' \cup A'' \cup \{W\}$ and $\mathcal{S}' \cup \{z_1, z_2\}$.

Given a set packing $\{S_{k_1}, S_{k_2}, \ldots\}$ of $\mathcal{S}$, we can obtain the convexly independent set $\{S'_{k_1}, S'_{k_2}, \ldots\}$ of G (notice that all minimum paths between two vertices in this set passes through the vertex W).

On the other way, given a convexly independent set C of $V(G)$, let $g(\mathcal{S}, C) = \{S_k : S'_k \in C\}$. We claim that $g(\mathcal{S}, C)$ is a set packing of $\mathcal{S}$.

In fact, we prove a stronger statement. We claim that a subset $C \subseteq V(G)$ with $|C| \geq 4$ is a convexly independent set if and only if $C \subseteq \mathcal{S}'$ and $g(\mathcal{S}, C)$ is a set packing of $\mathcal{S}$.

See Figure 1:

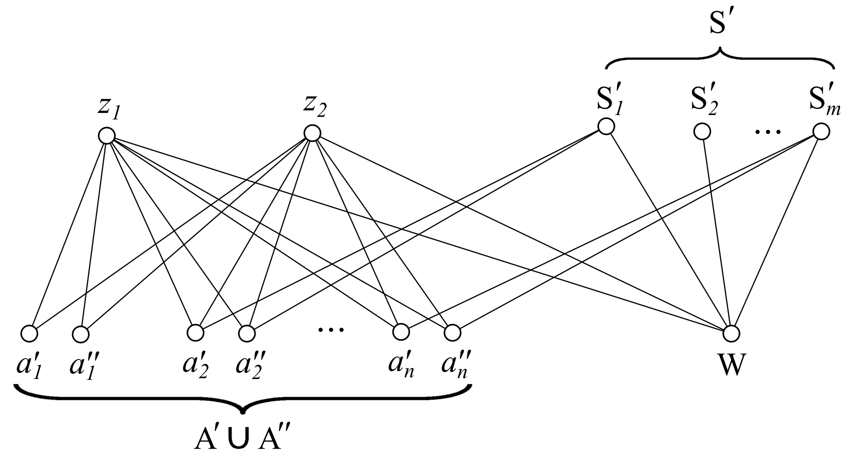

Figure 1.

In order to prove this claim, suppose that C is a convexly independent set with at least 4 vertices. Clearly C does not contain 2 vertices u, v such that $hull(\{u, v\}) = V(G)$. Then C does not contain both z_1 and z_2, and C does not contain two vertices of $A' \cup A'' \cup \{W\}$. Also notice that C does not contain two vertices S'_k and S'_ℓ such that $S_k \cap S_\ell \neq \emptyset$, since $hull(\{S'_k, S'_\ell\}) = V(G)$.

If $z_1 \in C$, then, since $|C| \geq 4$ and C has at most one vertex of $A' \cup A'' \cup \{W\}$, C contains a vertex $S'_k \in \mathcal{S}'$, a contradiction since $hull(z_1, S'_k) = V(G)$. Then $z_1, z_2 \notin C$. Since $|C| \geq 4$, then C contains at least two vertices $S'_k, S'_\ell \in \mathcal{S}'$ (clearly $S_k \cap S_\ell = \emptyset$). Therefore, $W \notin C$, since $W \in hull(\{S'_k, S'_\ell\})$. Assume that C contains a vertex $a'_i \in A'$. Then a'_i cannot be adjacent to both S'_k and S'_ℓ. Consider that a'_i is not adjacent to S'_k. Then there are minimum paths from a'_i to S'_k passing through z_1 and z_2 and, consequently, $hull(\{a'_i, S'_k\}) = V(G)$, a contradiction. Thus $C \cap A' = \emptyset$ and analogously $C \cap A'' = \emptyset$. Therefore $C \subseteq \mathcal{S}'$ and, if $S'_k, S'_\ell \in C$, then $S_k \cap S_\ell = \emptyset$ (in other words, $g(\mathcal{S}, C)$ is a set packing of $\mathcal{S}$).

For the converse, assume that $C \subseteq \mathcal{S}'$ and $g(\mathcal{S}, C)$ is a set packing of $\mathcal{S}$. Therefore, since $S_k \cap S_\ell = \emptyset$ for every $S'_k, S'_\ell \in C$, there is no minimum path between two vertices of C passing through a vertex of $A' \cup A''$. Consequently

$hull(C') \subseteq C' \cup \{W\}$ for every $C' \subseteq C$. This implies that C is convexly independent.

With this, we conclude directly that $(f, g, 1)$ is a continuous reduction from the set packing problem to the geodetic rank problem. It is known that, unless P=NP, there can be no polynomial time algorithm that approximates the maximum clique to within a factor better than $n^{1-\varepsilon}$, for any $\varepsilon > 0$ [33]. Since Set-Packing is as hard to approximate as the Maximum Clique Problem [5], we are done. □

4. P_4-sparse graphs

In 2011, Campos et al. [9] obtained linear time algorithms for many P_3-convexity parameters on $(q, q-4)$-graphs, for every fixed q, which are the graphs such that every set with at most q vertices induces at most $q-4$ P_4's. Cographs and P_4-sparse graphs are respectively the $(4, 0)$-graphs and the $(5, 1)$-graphs. In this paper, we obtain linear time algorithms for the geodetic rank on P_4-sparse graphs.

Given graphs G_1 and G_2, the disjoint union $G_1 \cup G_2$ is the graph obtained from the union of the vertex sets and the edge sets, and the join $G_1 + G_2$ is the graph obtained from $G_1 \cup G_2$ including all edges between G_1 and G_2.

It is known that every cograph G is a vertex, or a disjoint union or a join of two cographs. The next lemma is valid for the geodetic convexity and the P_3^*-convexity.

Lemma 2 (Union). *Let G_1 and G_2 be two graphs. Then $rk(G_1 \cup G_2) = rk(G_1) + rk(G_2)$ for any path convexity.*

Lemma 3. *Let G_1 and G_2 be two graphs and let $G = G_1 + G_2$. If G_1 and G_2 are complete graphs, then $rk_g(G) = rk_{P_3^*}(G) = n$, where n is the number of vertices of G. If G_1 and G_2 are not complete graphs, then $rk_g(G) = rk_{P_3^*}(G) = \omega(G)$, where $\omega(G)$ is the size of the largest clique of G. If G_1 is not complete and G_2 is complete, then $rk_g(G) = rk_{P_3^*}(G) = \max\{rk_{P_3^*}(G_1), \omega(G)\}$.*

Proof. If G is complete, it is clear that $V(G)$ is convexly independent and then the rank is $n = |V(G)|$ for both convexities. Assume that G_1 and G_2 are not complete and let S be a convexly independent set of G with at least 3 vertices. If S contains two non-adjacent vertices of G_1, then these two vertices form a hull set, a contradiction. The same for G_2. Then every convexly independent set with at least 3 vertices is a clique of G and then $rk_g(G) = rk_{P_3^*}(G) = \omega(G)$.

Now assume that G_1 is not complete and G_2 is complete. Let S be a convexly independent set of G. If S contains a vertex of G_2, then S cannot contain two non-adjacent vertices of G_1 and, consequently, S is a clique. If S does not contain a vertex of G_2, then S is also a convexly independent set of G_1. Therefore, $rk_g(G) = rk_{P_3^*}(G) = \max\{rk_{P_3^*}(G_1), \omega(G)\}$. □

It is known [26] that every P_4-sparse graph G is a vertex, or is the union or the join of two P_4-sparse graphs, or is a spider (R, K, S) such that either $R = \emptyset$ or $G[R]$ is P_4-sparse. A *spider* (R, K, S) is a graph $G = (R \cup K \cup S,\ E)$ such that $K = \{k_1, \ldots, k_p\}$ and $S = \{s_1, \ldots, s_p\}$, for $p \geq 2$, induce a clique and a stable set, respectively; either s_i is adjacent to k_j if and only if $i = j$ (a thin spider), or s_i is adjacent to k_j if and only if $i \neq j$ (a thick spider); and every vertex of R is adjacent to each vertex of K and non-adjacent to each vertex of S. Moreover, this decomposition of P_4-sparse graphs can be obtained in linear time [26].

Notice that, if G is a spider (R, K, S), then the only shortest paths that are not edges and induced P_3's in G are between two vertices of S (and only if G is a thin spider). Therefore the geodetic convexity and the P_3^*-convexity coincide in G if G is not a thin spider and the computation of the parameters in one convexity will imply the computation in the other convexity.

Lemma 4. *If G is a spider (R, K, S) and $R \neq \emptyset$, then $rk_g(G) = rk_{P_3^*}(G) = rk_{P_3^*}(G[R]) + k$, where $k = |K| = |S|$.*

Proof. At first, notice that the set C' formed by S union a largest P_3^*-convexly independent set of $G[R]$ is a convexly independent set with $|C'| = rk_{P_3^*}(G[R]) + k \geq k + 1$ vertices in both convexities. Suppose, by contradiction, that there exists a set C which is a convexly independent set of G in the geodetic convexity with more than $|C'|$ vertices. If C contains two adjacent vertices $k_i \in K$ and $s_j \in S$, then C cannot contain any vertex of $R \cup K \setminus \{k_i\}$ and, consequently $|C| \leq k + 1$, a contradiction. Then C does not contain two adjacent vertices $k_i \in K$ and $s_j \in S$. This implies that $|C \setminus R| \leq k$. Consequently $|C \cap R| > rk_{P_3^*}(G[R])$ and $C \cap R$ is not P_3^*-convexly independent in $G[R]$. If C contains a vertex of K, then C cannot contain two non-adjacent vertices of R and, consequently, $C \cap R$ is a clique, which is also a P_3^*-convexly independent set, a contradiction. If C does not contain a vertex of K, then $C \cap R$ is also a P_3^*-convexly independent set of $G[R]$, a contradiction. Then C' is a largest convexly independent set of G. □

Lemma 5. *If G is a spider (R, K, S) and $R = \emptyset$, then $rk_g(G) = k$, where $k = |K| = |S|$.*

Proof. Let C be a convexly independent set of G in the geodetic convexity. It is clear that if $C \subseteq K$ or $K \subseteq S$ then $|C| \leq k$, and hence we can assume that $C \cap K \neq \emptyset$ and $C \cap S \neq \emptyset$. If C contains two adjacent vertices $k_i \in K$ and $s_j \in S$, then C cannot contain any other vertex of $K \cup S \setminus \{k_i, s_j\}$ when G is a thin spider and, consequently $|C| = 2$. And if G is a thick spider, then $k_i \notin C$ and therefore $|C| \leq k$.

In the case G is a thick spider, if C contains two non adjacent vertices, which are s_i and k_i, then any other vertex from $K \setminus k_i$ is forbidden, and since k_i is adjacent to all other vertices from $S \setminus s_i$ we can conclude that $|C| \leq k$. Now in the case G is a thin spider and $|C| > 2$, since every vertex of $C \cap K$ excludes a vertex of S, and vice-versa, we can also conclude that $|C| \leq k$.

Finally, because K and S are convexly independent sets, we can conclude that $rk_g(G) = k$. □

Since the P_3^*-convexity coincide with the g-convexity in thick spiders, we have the following corollary.

Corollary 6. *If G is a thick spider (R, K, S) and $R = \emptyset$, then $rk_{P_3^*}(G) = k$, where $k = |K| = |S|$.*

Finally, if G is a thin spider, we can compute the parameters of the P_3^*-convexity with the following lemma.

Lemma 7. *If G is a thin spider (R, K, S) and $R = \emptyset$, then $rk_{P_3^*}(G) = k+1$, where $k = |K| = |S|$.*

Proof. At first, notice that the set C' formed by S and one vertex of K is P_3^*-convexly independent with $k+1$ vertices. Suppose, by contradiction, that there exists a P_3^*-convexly independent set C of G with more than $|C'|$ vertices. Then $|C| \geq k+2$ and, by the pigeon-hole principle, C contains two adjacent vertices $k_i \in K$ and $s_i \in S$. Moreover, C contains a vertex $k_j \neq k_i$ of K, a contradiction, since there exists a shortest path from s_i to k_j passing through k_i. □

With this, we obtain a linear time algorithm for the geodetic rank on P_4-sparse graphs.

5. Split graphs

In this section, we describe a polynomial time algorithm for determining the geodetic rank of a split graph. First, we present a necessary condition for a set of vertices of a graph to be convexly independent, in general graphs.

The following notation is employed. For a graph G and $v \in V(G)$, $N_G(v)$ and $N_G[v]$ denote the open and closed neighborhoods of v, respectively. For $S \subseteq V(G)$, let $N_G[S] = \cup_{v \in S} N_G[v]$, $G[S]$ denote the subgraph of G induced by the vertices of S, and $H_G(S)$ denote the hull of S. We let $N_G^2[v] = N_G[N_G[v]]$. For vertices $v_i, v_j \in V(G)$, let $d_G(v_i, v_j)$ to denote the *distance* between v_i and v_j in G, that is, the length of a shortest $v_i - v_j$ path. When the context is clear, we may also drop the subscripts in these notations, and simply write $N[v]$, $d(v_i, v_j)$, and so on. Finally, denote by $\omega(G)$ and $\alpha(G)$ the sizes of the maximum clique and independent set of G, respectively.

Let $\{S_1, \ldots, S_k\}$ be a family of subsets $S_i \subseteq V(G)$ of vertices of a graph G. Then $S_1, \ldots, S_k$ is *distance-regular* when each pair of subsets S_i, S_j satisfies

$$d(v_i, v_j) = d(v'_i, v'_j),$$

for all $v_i, v'_i \in S_i$ and $v_j, v'_j \in S_j$.

It follows from the definition that the subsets $\{S_1, \ldots, S_k\}$ must be vertex disjoint.

Theorem 8. *Let $S \subseteq V(G)$ be a convexly independent set of vertices of G. Then $G[S]$ is a family of (vertex disjoint) distance-regular cliques of G.*

Proof. Let $\{S_1, \ldots, S_k\}$ be the set of connected components of $G[S]$. Suppose some S_i is not a clique. Then S_i contains an induced P_3. Let v_1, v_2, v_3 be such a path. It follows that v_2 belongs to a shortest $v_1 - v_3$ path. Consequently, $v_2 \in H(S \setminus \{v_2\})$, contradicting S to be convexly independent.

We now know that the connected components $\{S_1, \ldots, S_k\}$ of $G[S]$ are all cliques. Next, again by contrary, suppose $G[S]$ is not distance-regular. Then, for some pair of subsets S_i, S_j there are vertices $v_i, v'_i \in S_i$ and $v_j, v'_j \in S_j$, such that $d(v_i, v_j) \neq d(v'_i, v'_j)$. Without loss of generality, let $d(v_i, v_j) < d(v'_i, v'_j)$. Examine the alternatives for the relative values of $d(v_i, v'_j)$. Because $d(v_j, v'_j) = 1$, it follows $d(v_i, v'_j) \leq d(v_i, v_j) + 1$. If $d(v_i, v'_j) = d(v_i, v_j) + 1$ then v_j belongs to a shortest $v_i - v'_j$ path, implying $H(S \setminus \{v_j\}) = H(S)$, contradicting S to be convexly independent. Then $d(v_i, v'_j) = d(v_i, v_j)$. In this situation, since $d(v'_i, v'_j) > d(v_i, v_j)$ and $d(v_i, v'_i) = 1$, it follows that v_i belongs to a shortest $v'_i - v'_j$ path, again contradicting S to be convexly independent. Therefore $G[S]$ is indeed a family of distance-regular cliques of G. □

Corollary 9. *Let S be a convexly independent set of a bipartite graph G with $|S| \geq 3$. Then S is an independent set of G.*

Proof. Since G is bipartite and by the above theorem, it follows that $G[S]$ consists of an independent set of vertices and edges. However, suppose $G[S]$ contains an edge v_1v_2. Take any third vertex v_3 of $G[S]$, $v_3 \neq v_1, v_2$. Then the parities of $d(v_1, v_3)$ and $d(v_2, v_3)$ are distinct, meaning that $d(v_1, v_3) \neq d(v_2, v_3)$. Then $G[S]$ is not distance-regular, contradicting S to be convexly independent. Therefore S contains solely an independent set of vertices of G. □

Next, we turn to split graphs. The following notations are employed. Let G be a connected split graph, $V(G) = I \cup C$, where I is an independent set and C a maximal clique of G, and $I, C \neq \emptyset$. Let B denote the bipartite graph obtained from G, by removing all internal edges of C. For some $v \in I$, we denote by $E_v(B)$ the set of edges incident to v in B, C_v the set $C \setminus N(v)$ and we let I_v be the set $\{v' \in I \mid d(v, v') = 2 \text{ and } N(v') \cap C_v \neq \emptyset\}$. Let $B_v := B[N_B^2[v]] - E_v(B)$. Clearly, B and B_v are bipartite graphs, and v is an isolated vertex of B_v.

For $v \in I$ and $w \in N(v)$, we let $\simeq_v^w$ be the binary relation on I_v such that for every $v', v'' \in I_v$ we have $v' \simeq_v^w v''$ if $N(v') \cap N(v) = N(v'') \cap N(v)$ and $v', v'' \in N_G^2[v] \cap N_G^2[w]$. Clearly, the binary relation $\simeq_v^w$ is an equivalence relation, and we denote by $[v']_{\simeq_v^w}$ the equivalence class of v'. We let $S_{vw}^{[v']_{\simeq_v^w}}$ be the set $\{v, w\} \cup [v']_{\simeq_v^w} \cup \{v'' \in I \setminus (I_v \cup \{v\}) \mid N(v'') \subseteq N(v')\}$. It is worth noticing that $S_{vw}^{[v']_{\simeq_v^w}}$ is an independent set of B_v.

The next theorem describes exactly the convexly independent sets of a split graph.

Theorem 10. *Let G be a connected split graph, and let $S \subseteq V(G)$, $|S| \geq 3$. The following affirmatives are equivalent:*

1. *S is convexly independent.*
2. *S is an independent set of B or an independent set of B_v containing v, for some $v \in I$. Moreover, if S is of the latter type, then two vertices in $(S \cap I) \setminus \{v\}$ cannot have a common neighbor in C_v, and if $|S \cap C| > 1$, then S cannot contain a vertex in I_v, and if $S \cap C = \{w\}$, then $S \subseteq S_{vw}^{[v']_{\simeq_v^w}}$ for some $v' \in I_v$.*

Proof. ((1) $\Longrightarrow$ (2)). From Theorem 8 we know that $G[S]$ is a family of disjoint distance-regular cliques. If $G[S]$ consists of exactly one clique then $S \subseteq C$ or $S \subseteq N[v]$, for some $v \in I$. In the former alternative S is an independent set of B, while in the latter an independent set of B_v containing v. Examine the situations where $G[S]$ contains more than one clique. If all cliques of $G[S]$ are singletons then S is an independent set of G, hence also of B.

If $G[S]$ is not connected and all the connected components are singletons then $S \subseteq I$ and then is an independent set of B. Now, exactly one of the connected components of $G[S]$, say C_1, is not a singleton and is included in $N_G[v]$ for some v and all the other connected components are singleton vertices in I. Let $S_I := S \cap I$ and $S_C := S \cap C$. If S_C is not included in the neighbors of a vertex from S_I, then S is an independent set of B, otherwise, $C_1 = S_C \cup \{v\}$. Now, since I is an independent set no vertex in $S_I \setminus \{v\}$ can be adjacent to a vertex from S_C because $G[S]$ is distance-regular (by Theorem 8). Hence, S is an independent set of B_v containing v. Now, if two vertices in $S_I \setminus \{v\}$ have a common neighbor in C_v, then S_C would be included in $H_g(S_I)$ because if $w \in C_v$ is a common neighbor of v_1 and v_2, both in $S_I \setminus \{v\}$, it would be generated by $\{v_1, v_2\}$, and then every vertex in $N(v)$ is in a shortest path between v and w. If $|S_C| \geq 2$, then no vertex in $S_I \setminus \{v\}$ can have a neighbor in C_v. Indeed, let w_1 and w_2 be two neighbors of v in S_C and let $v' \in S_I \setminus \{v\}$ have a neighbor z in C_v. Then, z is in a shortest path between w_1 and v', *i.e.*, $w_2 \in H_g(S \setminus w_2)$, contradicting S being a convexly independent set. Suppose finally that $S_C = \{w\}$ and let $v' \in S_I \setminus \{v\}$ with a neighbor z in C_v and let $w' \in N_G(v) \cap N_G(S_I \setminus \{v\})$ with $w' \neq w$. Notice that w' must exist because for each $v'' \in S_I \setminus \{v\}$, $d(w, v'') = 2$ and hence $d(v, v'') = 2$. If $w' \notin N(v')$, then $w \in H_g(S_I)$ because w' would be generated by S_I, then z would be generated by w' and v' and finally w would be generated by v and z. Therefore, for all $v'' \in S_I \setminus \{v\}$ we have $N_G[v''] \cap N_G(v) \subseteq N_G(v') \cap N_G(v)$ and if v'' have a neighbor in C_v then $v' \simeq_v^w v''$. Hence, $S \subseteq S_{vw}^{[v']_{\simeq_v^w}}$.

((2) $\implies$ (1)). First, suppose that S is an independent set of B. If $S \subseteq I$ then S is an independent set of G, hence convexly independent. If $S \subseteq C$ then $G[S]$ is a single clique, clearly also convexly independent. In the remaining alternatives, the two sets $S_C := S \cap C$ and $S_I := S \cap I \neq$ are both non empty. Because S is an independent set, $N_G(S_I) \subseteq C$ and C is a clique we can conclude that $H_g(C' \cup S_I) = C' \cup S_I \cup N_G(S_I)$ for every $C' \subset S_C$. Consequently, S is convexly independent.

Examine the alternative where S is an independent set of B_v containing v, for some $v \in I$ and S is not an independent set of B. Then the two sets $S_C := S \cap C$ and $S_I := S \cap I$ are non empty. If $S_I = \{v\}$ then S is a clique of G, *i.e.*, convexly independent. Otherwise $S_I \setminus \{v\} \neq \emptyset$. First notice that because the vertices in S_I are simplicial vertices, none can be in a shortest path and hence no vertex $v' \in S_I$ is in $H_g(S \setminus \{v'\})$. Assume first that $|S_C| \geq 2$. Now, because $|S_C| \geq 2$, by hypothesis $S_I \cap I_v = \emptyset$. Hence, $H_g(C' \cup (S_I \setminus \{v\})) = C' \cup S_I \cup N_G(S_I \setminus \{v\})$ for each $C' \subset S_C$.

Therefore, S is convexly independent. Suppose finally that $S_C := \{w\}$, *i.e.*, $S \subseteq S_{vw}^{[v']_{\simeq_v^w}}$ for some $v' \in I_v$. Since two vertices v_1 and v_2 in $S_I \cap [v']_{\simeq_v^w}$ cannot share a common vertex in C_v and C is a clique, then $C_v \cap H_g(S) = \emptyset$. And because $w \notin N_G(v')$ and $N_G(S_I \setminus \{v\}) = N_G(v')$, we can conclude that $w \notin H_g(S \setminus \{w\})$, *i.e.*, S is convexly independent. □

From Theorem 10 we can characterize the maximum convexly independent sets in split graphs. Let G be a split graph. For each $v \in I$ and each $w \in N_G[v]$ we let $\alpha_{vw} := \max\{|S_{vw}^{[v']_{\simeq_v^w}}| \mid \forall v_1, v_2 \in [v']_{\simeq_v^w},\ (N(v_1) \cap N(v_2)) \cap C_v = \emptyset\}$.

Corollary 11. *Let G be a split graph, then*

$$rk(G) = max_{v \in I} \left\{ \alpha(B_v[V(B_v) \setminus I_v]), \alpha(B), \max_{w \in N(v)} \{\alpha_{vw}\} \right\}.$$

Proof. By Theorem 10 we did not take into account in the above equality only the sets $S_{vw}^{[v']_{\simeq_v^w}}$ such that there exist $v_1, v_2 \in [v']_{\simeq_v^w}$ and $(N(v_1) \cap N(v_2)) \cap C_v \neq \emptyset$. However, such sets are not convexly independent and for each such set $S_{vw}^{[v']_{\simeq_v^w}}$, the set $S_{vw}^{[v']_{\simeq_v^w}} \setminus \{w\}$ is convexly independent and is included in I. Therefore, $\max\{|S_{vw}^{[v']_{\simeq_v^w}}| \mid \exists v_1, v_2 \in [v']_{\simeq_v^w},\ (N(v_1) \cap N(v_2)) \cap C_v \neq \emptyset\} \leq \alpha(B)$. This concludes the proof. □

The algorithm for efficiently computing the rank of a split graph follows directly from Corollary 11. The maximum independent set of a bipartite graph having n vertices can be found by computing a matching [28], requiring $O(n^{2.5})$ time. For each $v \in I$ and each $w \in N[v]$ the equivalence classes of $\simeq_v^w$ can be found in $O(n^2)$ time and the sets $S_{vw}^{[v']_{\simeq_v^w}}$ be constructed also in $O(n^2)$ time. Corollary 11 implies that the rank of G can be determined by $O(n)$ computations of a maximum independent sets of bipartite graphs, and the computations of $O(n^2)$ sets $S_{vw}^{[v']_{\simeq_v^w}}$. Then the overall complexity of the algorithm is $O(n^4)$.

6. Monadic second-order logic

Monadic second-order logic is a logical language which has proved its importance in complexity theory the last three decades due to its algorithmic relations with graph complexity measures such as *tree-width* (see for instance the book [14]). In this section we consider the definability of the rank function in monadic second-order logic, and its consequences in graphs of bounded *clique-width*. We refer to [14] for more information since we recall only the necessary definitions for the understanding. We now review *monadic second-order formulas* on graphs. We will use lower case variables $x, y, z, \dots$ (resp. upper case variables $X, Y, Z, \dots$) to denote vertices (resp. subsets of vertices) of graphs. The *atomic formulas* are $x = y$, $x \in X$ and $edg(x, y)$ where edg denotes the adjacency relation in graphs. The set MS_1 of *monadic second-order formulas* is the set of formulas formed from atomic formulas with Boolean connectives $\wedge$, $\vee$, $\neg$, $\Longrightarrow$, $\Longleftrightarrow$, *element quantifications* $\exists x$ and $\forall x$, and *set quantifications* $\exists X$ and $\forall X$. An occurrence of a variable which is not under the scope of a quantifier is called a *free variable*. We often write $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$ to express that the formula φ has $x_1, \dots, x_m, Y_1, \dots, Y_q$ as free variables and $G \models \varphi(a_1, \dots, a_m, Z_1, \dots, Z_q)$ to say that the formula $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$ holds in G when substituting $(a_1, \dots, a_m) \in V(G)^m$ to element variables $(x_1, \dots, x_m)$ and $(Z_1, \dots, Z_q) \in (2^{V(G)})^q$ to set variables $(Y_1, \dots, Y_q)$ in the formula φ. The following is an example of a formula expressing that two vertices x and y are connected by an induced path of length 2.

$$\neg edg(x, y) \wedge \exists z(edg(x, z) \wedge edg(z, y)).$$

If $\varphi(x_1, \dots, x_m, Y_1, \dots, Y_q)$ is a formula in MS_1, we let opt_φ, with $opt \in \{\min, \max\}$, be the problem that consists in, given a graph G, to finding a tuple $(Z_1, \dots, Z_q)$ of $(2^{V(G)})^q$ such that

$$\sum_{1 \le i \le q} |Z_i| = opt \left\{ \sum_{1 \le i \le q} |W_i| \mid G \models \varphi(a_1, \dots, a_m, W_1, \dots, W_q) \right\}.$$

It is well-known that many optimization graph problems, *e.g.*, minimum feedback-vertex set, maximum clique, …, correspond to opt_φ for some MS_1 formula φ.

Clique-width is a graph complexity measure introduced by Courcelle and Olariu in [15] and has many algorithmic applications. Roughly, a graph has *clique-width* at most k if it can be obtained from the empty graph by successively applying the following operations (1) the disjoint union of two graphs, (2) add all the edges between vertices labeled i and vertices labeled j, $i \neq j$, $i, j \in \{1, \dots, k\}$, (3) relabel the vertices labeled i into j, $i, j \in \{1, \dots, k\}$ (4) creation of a graph with a single vertex labeled $i \in \{1, \dots, k\}$. The obtained expression is called a *clique-width expression*. Examples of graph classes with bounded clique-width are distance-hereditary graphs, graphs of bounded tree-width, $(q, q-4)$-graphs (for instance P_4-sparse graphs), distance-hereditary graphs, P_4-tidy graphs, … We refer to the book [14], which summarizes the most important results in this area, for the exact definition of clique-width and for more information on it.

Theorem 12 ([14, Theorem 6.56, Page 499])**.** *Let k be a fixed constant. For every MS_1 formula $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$, opt_φ, for $opt \in \{\min, \max\}$, can be solved in linear time in any graph of clique-width at most k, provided the clique-width expression is given. If the clique-width expression is not given, one can compute one using at most 2^{k+1} labels in cubic time.*

Proposition 13. *Let $\mathcal{C}$ be a graph convexity and let $B_\mathcal{C}$ be the betweenness relation associated with $\mathcal{C}$. Let $\mathscr{C}$ be a graph class. If there is a monadic second-order formula $\varphi_{B,\mathscr{C}}(x, z, y)$ expressing that z is between x and y for vertices x, y, z of a graph $G \in \mathscr{C}$, then there is a formula $ind_{\mathcal{C},\mathscr{C}}(X)$ expressing that X is convexly independent for every subset X of the vertex set of a graph $G \in \mathscr{C}$.*

Proof. The first part of the proof is already given in [27], but we include it for completeness. We write $X \subseteq Y$ and $X \subsetneq Y$ as shortcuts for the formulas $\forall x(x \in X \implies x \in Y)$ and $X \subseteq Y \ \wedge\ \exists y(y \in Y \ \wedge\ y \notin X)$ respectively. The following formula $Closed(X)$, with free set variable X, says that X is in $\mathcal{C}$.

$$\forall x, y(x \in X \ \wedge\ y \in X \implies \neg\left(\exists z(\varphi_{B,\mathscr{C}}(x, z, y) \wedge \ \not z \in X)\right).$$

The validity of the formula $Closed(X)$ is trivial since it states that for each $x, y \in X$, all $z \in V(G)$ such that $B_\mathcal{C}(x, z, y)$ holds must be in X, which is exactly the definition of convex sets. The following formula, $ConvexHull(X, Y)$, with free set variables X and Y, says that Y is the convex hull of X

$$Closed(Y) \wedge X \subseteq Y \wedge \forall Z(X \subseteq Z \wedge Z \subsetneq Y \implies \neg Closed(Z)).$$

The formula $ConvexHull(X, Y)$ states that Y is closed, contains X and for any $X \subseteq Z \subseteq Y$ we have that Z is not a convex set, which is the definition of a convex hull set. Therefore, $ConvexHull(X, Y)$ states that Y is the convex hull of X.

We can now write the formula $ind_{\mathcal{C},\mathscr{C}}(X)$ stating that X is convexly independent, the validity of which is easy to check. (We write $X \setminus \{z\}$ below because one can write a formula $\phi(Z)$ stating that $Z = X \setminus \{z\}$.)

$$\forall z(\forall Y(z \in X \wedge ConvexHull(X \setminus \{z\}, Y) \implies \neg z \in Y)). \qquad \square$$

As corollaries we get the following.

Proposition 14. *Let k be a fixed positive integer. For every graph G of clique-width at most k one can compute $rk_{P_3^*}(G)$, $rk_{P_3}(G)$ and $rk_P(G)$ in cubic time.*

Proof. The following formula $\varphi_{P_3}(x,z,y) := edg(x,z) \wedge edg(z,y)$ expresses that z is a common neighbor of x and y. The formula $\varphi_{P_3^*}(x,z,y) := \neg edg(x,y) \wedge edg(x,z) \wedge edg(z,y)$ expresses that z is in a chordless path of length 2 between x and y. A formula $\varphi_P(x,z,y)$ stating that z is in a chordless path between x and y is given in [27]. From Proposition 13 and Theorem 12 one can conclude. □

Proposition 15. *For every distance-hereditary graph G one can compute $rk_g(G)$ and $rk_P(G)$ in linear time.*

Proof. From the definition of distance-hereditary graphs we know that every chordless path is a shortest path. Hence, z is in a shortest path between x and y if and only if z is in a chordless path between x and y and the geodetic and monophonic convexities coincide. Therefore there is a formula $\varphi_{dh}(x,z,y)$ stating that z is in a shortest path between x and y for all vertices x, z, y of a distance-hereditary graph (see [27]). Since distance-hereditary graphs have clique-width at most 3 and a clique-width expression can be constructed in linear time [25], one can conclude with Proposition 13 and Theorem 12. □

Acknowledgments

This research was supported by FAPERJ, Capes (Project STIC-AmSud 2012), CNPq (Project Universal/Bolsa de Produtividade), by Funcap (Project PRONEM 2011) and by CNRS (GDRI Web Science). We would like to thank Manoel Campêlo, UFC, Fortaleza, Brazil, for his valuable comments.

References

[1] J. Araujo, V. Campos, F. Giroire, N. Nisse, L. Sampaio, R. Soares. On the hull number of some graph classes. *Theoretical Computer Science* **475** (2013), 1–12. MR3018479

[2] R. Araújo, R. M. Sampaio, V. F. dos Santos; J. L. Szwarcfiter. The convexity of induced paths of order three and applications: complexity aspects. *Submitted*, 2014.

[3] D. Artigas, S. Dantas, M. C. Dourado, J. L. Szwarcfiter, Y. Yamaguchi. On the contour of graphs. *Discrete Applied Mathematics* **161** (2013), 1356–1362. MR3043087

[4] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, M. Protasi. *Combinatorial Optimization Problems and Their Approximability Properties*, Springer, Berlin, 1999. MR1734026

[5] G. Ausiello, A. D'Atri, M. Protasi. Structure preserving reductions among convex optimization problems. *J. Comput. System Sci.* **21** (1980), 136–153. MR0589808

[6] J. A. Bondy, U. S. R. Murty. *Graph Theory*, Springer, 2008. MR2368647

[7] F. Buckley, F. Harary. *Distance in Graphs.* Addison-Wesley, 1990. MR1045632

[8] J. Caceres, C. Hernando, M. Mora, I. M. Pelayo, M. L. Puertas, C. Seara. On monophonic sets in graphs. In *20th British Combinatorial Conference*, Durham, England, 2005.

[9] V. Campos, R. Sampaio, A. Silva, J. L. Szwarcfiter. Graphs with few P4's under the convexity of paths of order three. *Discrete Applied Mathematics*, to appear. http://dx.doi.org/10.1016/j.dam.2014.05.005. MR3354816

[10] C. C. Centeno, M. C. Dourado, L. D. Penso, D. Rautenbach, J. L. Szwarcfiter. Irreversible conversion on graphs. *Theoretical Computer Science* **412**, 3693–3700. MR2839698

[11] M. Changat, J. Mathew. On triangle path convexity in graphs. *Discrete Mathematics* **206** (1999), 91–95. MR1665388

[12] M. Changat, H. M. Mulder, G. Sierksma. Convexities related to path properties in graphs. *Discrete Mathematics* **290** (2005), 117–131. MR2123384

[13] V. Chepoi. Separation of two convex sets in convexity structures. *J. Geometry* **50** (1994), 30–51. MR1280625

[14] B. Courcelle, J. Engelfriet. Graph Structure and Monadic Second-Order Logic – A Language-Theoretic Approach. *Encyclopedia of mathematics and its applications 138*, Cambridge University Press 2012, ISBN 978-0-521-89833-1, pp. I–XIV, 1–728. MR2962260

[15] B. Courcelle, S. Olariu. Upper bounds to the clique width of graphs. *Discrete Applied Mathematics* **101** (2000), 77–114. MR1743732

[16] P. Crescenzi. A short guide to approximation preserving reductions. *Proc. 12th Annual IEEE Conference on Computational Complexity*, IEEE Computer Society, (1997) 262–273. MR1758143

[17] M. C. Dourado, J. G. Gimbel, J. Kratochvil, F. Protti, J. L. Szwarcfiter. On the computation of the hull number of a graph. *Discrete Mathematics* **309** (2010), 5668–5674. MR2567969

[18] M. C. Dourado, F. Protti, D. Rautenbach, J. L. Szwarcfiter. On the hull number of triangle-free graphs. *SIAM J. on Discrete Mathematics* **23** (2010), 2163–2172. MR2594978

[19] M. C. Dourado, F. Protti, D, J. L. Szwarcfiter. Complexty results related to monophonic convexity. *Discrete Applied Mathematics* **158** (2010), 1268–1274. MR2652003

[20] F. F. Dragan, F. Nicolai, A. Bransdstadt. Convexity and HHD-free graphs. *SIAM Journal on Discrete Mathematics* **12** (1999), 119–135. MR1652261

[21] P. Duchet. Convex sets in graphs II. *Journal Combinaotorial Theory B* **44** (1988), 307–316. MR0941439

[22] P. Erdős, E. Fried, A. Hajnal, and E. C. Milner. Some remarks on simple tournaments. *Algebra Univers.* **2** (1972), 238–245. MR0306034

[23] M. Farber and R. E. Jamison. Convexity in graphs and hypergraphs. *SIAM J. on Algebraic and Discrete Methods* **7** (1986), 433–441. MR0844046

[24] M. Farber, R. E. Jamison. Local convexity in graphs. *Discrete Mathematics* **66** (1987), 231–247. MR0900046

[25] M. C. Golumbic, U. Rotics. On the clique-width of some perfect graph classes. *Int. J. Found. Comput. Sci.* **11** (2000) 423–443. MR1792124

[26] B. Jamison, S. Olariu. A tree representation for P_4-sparse graphs. *Discrete Applied Mathematics* **35** (1992), 115–129. MR1137806

[27] M. M. Kanté, L. Nourine. Polynomial time algorithms for computing a minimum hull set in distance-hereditary and chordal graphs. *Proc. 39th International Conference on Current Trends in Theory and Practice of Computer Science* (SOFSEM-2013), 268–279. MR3074097

[28] E. L. Lawler. *Combinatorial Optmization: Networks and Matroids*, Holt, Rinehart and Winston, New York, 1976. MR0439106

[29] J. W. Moon. Embedding tournaments in simple tournaments. *Discrete Math.* **2** (1972), 389–395. MR0302492

[30] I. F. Ramos, V. F. dos Santos, J. L. Szwarcfiter. Complexity aspects of the computation of the rank of a graph. *Discrete Mathematics and Theoretical Computer Science* **16** (2014), 73–86. MR3259607

[31] J. C. Varlet. Convexity in tournaments. *Bull. Soc. R. Sci. Liège* **45** (1976), 570–586. MR0450120

[32] L. J. Van de Vel. *Theory of Convex Structures*, North-Holland, 1995.

[33] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Proc. 38th ACM Symp. Theory of Computing* (STOC-2006), 681–690. MR2277193

Mamadou Moustapha Kanté
Université Clermont Auvergne
Université Blaise-Pascal
LIMOS, CNRS
Aubière
France
E-mail address: mamadou.kante@isima.fr

Rudini M. Sampaio
Universidade Federal do Ceará
Brazil
E-mail address: rudini@lia.ufc.br

Vinícius F. dos Santos
Centro Federal de Educação Tecnológica de Minas Gerais
Departamento de Computação
Belo Horizonte
Brazil

Universidade Federal de Minas Gerais
Departamento de Ciência da Computação
Belo Horizonte
Brazil
E-mail address: viniciussantos@dcc.ufmg.br

Jayme L. Szwarcfiter
Universidade Federal do Rio de Janeiro
IM. COPPE, NCE
Rio de Janeiro
Brazil

Universidade do Estado do Rio de Janeiro
Instituto de Matemática e Estatística
Rio de Janeiro
Brasil
E-mail address: jayme@nce.ufrj.br

Received 14 January 2015