Graphs and Combinatorics (2012) 28:333–345
DOI 10.1007/s00373-011-1049-7

ORIGINAL PAPER

On the Convexity Number of Graphs

Mitre C. Dourado · Fábio Protti · Dieter Rautenbach · Jayme L. Szwarcfiter

Received: 6 March 2009 / Revised: 25 March 2011 / Published online: 22 April 2011

Abstract A set of vertices S in a graph is convex if it contains all vertices which belong to shortest paths between vertices in S. The convexity number $c(G)$ of a graph G is the maximum cardinality of a convex set of vertices which does not contain all vertices of G. We prove NP-completeness of the problem to decide for a given bipartite graph G and an integer k whether $c(G) \geq k$. Furthermore, we identify natural necessary extension properties of graphs of small convexity number and study the interplay between these properties and upper bounds on the convexity number.

Keywords Convexity number · Convex hull · Convex set · Graph · Shortest path

M. C. Dourado (✉)
Instituto de Matemática, Universidade Federal do Rio de Janeiro,
Rio de Janeiro, RJ, Brazil
e-mail: mitre@dcc.ufrj.br

F. Protti
Instituto de Computação, Universidade Federal Fluminense,
Niterói, RJ, Brazil
e-mail: fabio@ic.uff.br

D. Rautenbach
Institut für Optimierung und Operations Research, Universität Ulm,
89069 Ulm, Germany
e-mail: dieter.rautenbach@uni-ulm.de

J. L. Szwarcfiter
Instituto de Matemática, NCE and COPPE, Universidade Federal do Rio de Janeiro,
Rio de Janeiro, RJ, Brazil
e-mail: jayme@nce.ufrj.br

1 Introduction

We consider finite, undirected and simple graphs G with vertex set $V(G)$ and edge set $E(G)$. For two vertices u and v of a graph G, let $I[u, v]$ denote the set of vertices of G which belong to a shortest path between u and v in G. For a set of vertices S, let $I[S]$ denote the union of the sets $I[u, v]$ over all pairs of vertices u and v in S. A set of vertices S is convex if $I[S] = S$. The convex hull $H[S]$ of a set S of vertices is the smallest convex set of vertices which contains S. Since the intersection of two convex sets is convex, the convex hull is well defined.

Chartrand, Wall, and Zhang [4] define the convexity number $c(G)$ of a graph G as the largest cardinality of a convex set of vertices which does not contain all vertices of Gimbel [8] proved that the decision problem associated to the convexity number is NP-complete. For further related results, we refer the reader to [2,3,5,6,9,10].

Our contributions in the present paper concern the algorithmic complexity of the convexity number and the structure of graphs of small convexity number. In Sect. 2, we refine Gimbel's hardness result [8] by proving NP-completeness for the class of bipartite graphs. Furthermore, we describe how to efficiently decide whether the convexity number is at least k for some fixed k and how to determine the convexity number for cographs in linear time. In Sect. 3, we study graphs of small convexity number. We identify natural necessary extension properties of such graphs and prove best possible upper bounds on the convexity number implied by these necessary conditions.

2 NP-Completeness for Bipartite Graphs

Our main result in this section is the NP-completeness of the following decision problem restricted to bipartite graphs.

Convexity Number
Instance: A graph G and an integer k.
Question: Is $c(G) \geq k$?

We start by showing how to solve the above problem in polynomial time, for fixed k. Let G be a graph and let S be a set of vertices of G. By definition, S is not convex if and only if there are two vertices x and y in S such that $I[x, y] \not\subseteq S$. Choosing such a pair of vertices at minimum distance, we obtain that S is not convex if and only if there are two vertices x and y in S such that there exists a shortest path P between x and y which is of length at least 2 and whose internal vertices all belong to $V(G)\setminus S$. Applying a shortest path algorithm to the induced subgraphs $G-(S\setminus\{x, y\}) = G[\{x, y\}\cup(V(G)\setminus S)]$ of G for all pairs of distinct vertices x and y in S, such paths can be found in polynomial time. Furthermore, iteratively extending a non-convex set by the internal vertices of such paths, one can determine the convex hull of a set of vertices in polynomial time.

By definition, the convexity number of a graph G is less than some integer k if and only if the convex hull of every set of exactly k vertices contains all vertices of G. Hence for fixed k, it can be decided in polynomial time whether the convexity number of a graph is at least k.

We proceed to our main result in this section.

Theorem 1 Convexity Number *restricted to bipartite graphs is NP-complete.*

Proof Since the convex hull of a set can be determined in polynomial time, Convexity Number is in NP. In order to prove NP-completeness, we reduce an instance (H, k) of the well-known NP-complete problem Clique (see [7], p. 194) to an instance (G, k') of Convexity Number such that the graph H has a clique of order at least k if and only if $c(G) \geq k'$, the encoding length of (G, k') is polynomially bounded in terms of the encoding length of (H, k), and G is bipartite.

Let (H, k) be an instance of Clique. Clearly, we may assume that H is connected and that $k \geq 3$. We construct G as follows. For every vertex u of H, we create four vertices w_u, x_u, y_u, and z_u in G and add the three edges $x_u z_u$, $y_u z_u$, and $w_u z_u$ as shown in Fig. 1a. For every edge uv of H, we create a set $V_{uv} = \{a_{uv}, b_{uv}, c_{uv}, d_{uv}, e_{uv}\} \cup I_{uv}$ of $n + 5$ further vertices in G, where n denotes the order of H and I_{uv} denotes an independent set of n vertices, and add edges such that z_u, w_u, z_v, and w_v together with the vertices in V_{uv} induce the graph G_{uv} as shown in Fig. 1b, where all the vertices in I_{uv} have exactly the same four neighbors. In other words,

$$
\begin{aligned}
V(G_{uv}) &= \{w_u, z_u, w_v, z_v\} \cup \{a_{uv}, b_{uv}, c_{uv}, d_{uv}, e_{uv}\} \cup I_{uv} \text{ and} \\
E(G_{uv}) &= \{a_{uv}b_{uv}, b_{uv}c_{uv}\} \cup \{w_u b_{uv}, w_v b_{uv}\} \\
&\quad \cup \{z_u a_{uv}, z_v a_{uv}, z_u c_{uv}, z_v c_{uv}, z_u d_{uv}, z_v d_{uv}, z_u e_{uv}, z_v e_{uv}\} \\
&\quad \cup \bigcup_{r \in I_{uv}} \{r w_u, r w_v, r d_{uv}, r e_{uv}\}.
\end{aligned}
$$

To complete the construction, we create two vertices x and y in G and add the edges xx_u and yy_u for all vertices u of H.

Note that G is bipartite with bipartition (V_1, V_2), where

$$
\begin{aligned}
V_1 &= \{z_u \mid u \in V(H)\} \cup \{b_{uv} \mid uv \in E(H)\} \cup \{x, y\} \cup \bigcup_{uv \in E(H)} I_{uv} \text{ and} \\
V_2 &= V(G) \setminus V_1.
\end{aligned}
$$

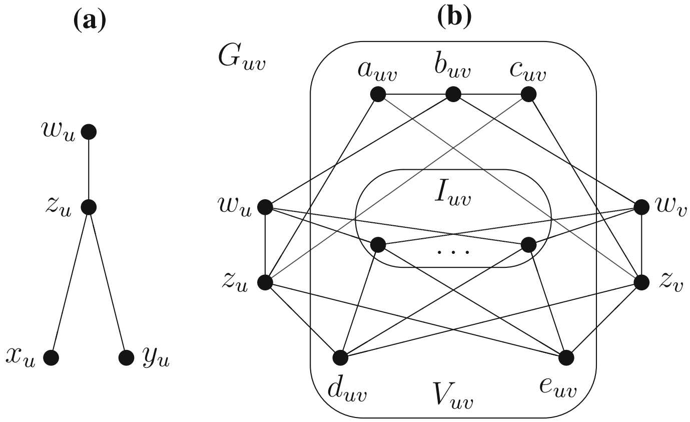


Fig. 1 Gadgets for the construction of G

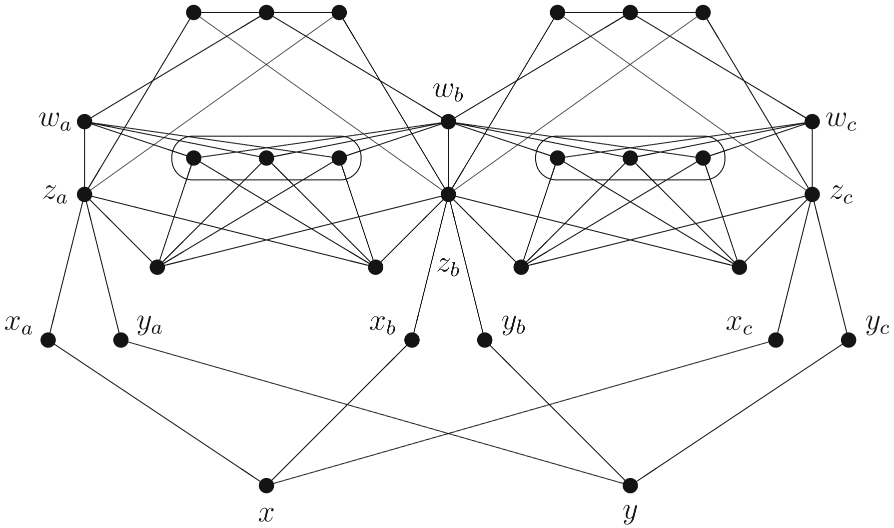


Fig. 2 The graph G constructed from $H = P_3$

Figure 2 illustrates the complete construction of G for the case that H is a path P_3 with the three vertices a, b and c.

Let $k' = 3k + (n + 5)\binom{k}{2} + 1$.

Clearly, the encoding length of (G, k') is polynomially bounded in terms of the encoding length of (H, k).

It remains to prove that H has a clique of order at least k if and only if the convexity number of G is at least k'.

First, we assume that H has a clique C of order at least k and construct a set S as follows. For every two vertices u and v in C, we add all vertices of G_{uv} to S. For every vertex u in C, we add the vertex x_u to S. Finally, we add x to S. It is easy to check that S is a convex set of at least k' vertices which does not contain y, i.e. $c(G) \geq k'$.

Next, we assume that G has a convex set of vertices S of order at least k' which does not contain all vertices of G.

Let

$$\begin{aligned} V_x &= \{x_u \mid u \in V(H)\}, \\ V_y &= \{y_u \mid u \in V(H)\}, \\ V_z &= \{z_u \mid u \in V(H)\}, \quad \text{and} \\ V_w &= \{w_u \mid u \in V(H)\}. \end{aligned}$$

Since $H[x, y]$ contains all vertices of G, at most one of the two vertices x and y belongs to S. If S contains more than n vertices from $V_x \cup V_y$, then there are distinct vertices u and v in H such that x_u and y_v both belong to S. Since $x, y \in I[x_u, y_v]$, we obtain $x, y \in S$ which is a contradiction. Hence S contains at most n vertices from $V_x \cup V_y$.

Claim A If S contains three vertices of G_{uv} for some edge uv of H, then S contains all vertices of G_{uv}.

Proof This property is easily verified. □

Claim B S contains at least two vertices from V_z.

Proof of Claim B For contradiction, we assume that S contains at most one vertex from V_z.

Using Claim A it follows easily that there is no edge uv of H such that

- either $|S \cap \{z_u, w_u\}| + |S \cap V_{uv}| \geq 3$,
- or $|S \cap \{z_u, w_u\}| \geq 1$ and $|S \cap \{z_v, w_v\}| \geq 1$.

Similarly, there are no three vertices u, v and w of H such that uv and vw are edges of H and

- $|S \cap \{z_v, w_v\}| = 0$, $|S \cap V_{uv}| \geq 1$, and $|S \cap V_{vw}| \geq 1$.

These observations imply an upper bound on the number of vertices in S from

$$V_z \cup V_w \cup \bigcup_{uv \in E(H)} V_{uv}$$

as follows: We assign to every vertex u of H, the weight $w(u) = |S \cap \{z_u, w_u\}|$ and to every edge uv of H, the weight $w(uv) = |S \cap V_{uv}|$. All weights of vertices are at most 1 except for at most one weight of a vertex that is 2. All weights of edges are at most 2. Let H' be a component of the spanning subgraph of H that contains the edges of H of positive weight. We obtain that H' is

- either an isolated vertex,
- or a star of order at least 3 whose central vertex has weight 1, whose edges have weight 1, and whose endvertices have weight 0,
- or a complete graph on two vertices u and v such that $w(u) + w(uv) + w(v) \leq 2$.

This implies that the total weight of all vertices and edges of H, which, by definition, equals the number of vertices in S from

$$V_z \cup V_w \cup \bigcup_{uv \in E(H)} V_{uv},$$

is at most $n + 1$.

Together with the remarks preceeding Claim A, we obtain that $|S| \leq 2n + 2$. Since $k \geq 3$, this is a contradiction. □

Let $C = \{u \in V(H) \mid z_u \in S\}$.

By Claim B, the set C contains at least two elements.

If S contains two vertices z_u and z_v such that u and v are not adjacent in H, then the distance of z_u and z_v in G is 4. Hence x and y belong to S which is a contradiction. Hence C is a clique of H.

For contradiction, we assume that $|C| = t < k$.

Let S' denote the union of the vertex sets of the graphs G_{uv} for all pairs of distinct vertices u and v in C. Note that S' contains exactly $2t + (n+5)\binom{t}{2}$ vertices. Since S is convex, S' is a subset of S.

Claim C $S \setminus S'$ contains no vertex from $V_w \cup \bigcup_{uv \in E(H)} V_{uv}$.

Proof of Claim C For contradiction, we assume that $S \setminus S'$ contains a vertex a from this set.

First, we assume that $a = w_u$ for some vertex u of H. By the definition of S', $u \notin C$. Let v be some vertex in C. Now $I[w_u, z_v]$ contains z_u which is a contradiction. Hence a belongs to V_{uv} for some edge uv of H.

If $v \in C$, then, by the definition of S', $u \notin C$ and Claim A implies that S contains all vertices of G_{uv} which is a contradiction. Hence $u, v \notin C$.

Let w be some vertex in C. Now $H[a, z_w]$ contains either z_u or z_v which is a contradiction.

This completes the proof of the claim. □

Together with the remarks preceeding Claim A, we obtain that S contains at most $2t + (n+5)\binom{t}{2} + n + 1 < k'$ elements which is a contradiction.

This completes the proof. □

We close this section with a positive result concerning the computation of the convexity number of cographs [1].

Theorem 2 *Let G be a cograph of order n.*

(i) *If G is connected, $G_1, \ldots, G_k, G_{k+1}, \ldots, G_t$ are the subgraphs of G induced by the vertex sets of the connected components of the complement of G where $|V(G_i)| \geq 2$ if and only if $i \leq k$, and ω denotes the clique number of G, then*

$$c(G) = \begin{cases} n-1, & \text{if } k = 0, \\ c(G_1) + t - 1, & \text{if } k = 1, \quad \text{and} \\ \omega, & \text{if } k \geq 2. \end{cases}$$

(ii) *If G is disconnected, then*

$$c(G) = n - \min\{|V(H)| - c(H) \mid H \text{ is a connected component of } G\}.$$

Proof (i) First, let $k = 0$. In this case, G is a complete graph and $c(G) = n - 1$.

Next, let $k = 1$. In this case, every vertex u in $G_2, \ldots, G_t$ is adjacent to all vertices in $V(G) \setminus \{u\}$. Let S be a convex set of vertices of cardinality $c(G)$. Let S_1 be the intersection of S and the vertex set of G_1. Clearly, S_1 is a convex set with respect to G_1. If S_1 is a clique, then S_1 does not contain all vertices of the graph G_1, because G_1 is not complete. By the choice of S, S contains all vertices in $G_2, \ldots, G_t$. If S_1 is not a clique, then S contains all vertices in $G_2, \ldots, G_t$, because S is convex. Therefore, $c(G) = c(G_1) + t - 1$.

Finally, let $k \geq 2$. Let S be a convex set of vertices of cardinality $c(G)$. If S contains two non-adjacent vertices from some G_{i^*}, then S contains all vertices of G outside of G_{i^*}. Hence S contains two non-adjacent vertices outside of G_{i^*} which implies that S contains all vertices of G_i, i.e. S contains all vertices of G which is a contradiction. Hence S is complete and $c(G) = \omega$.

(ii) This follows directly from the fact that a convex set of vertices of G of cardinality $c(G)$ contains all but one of the connected components of G. □

Using Theorem 2 and modular decompositions [11,12], one can easily compute the convexity number of a cograph in linear time.

3 Graphs of Small Convexity Number

A subgraph H of a graph G is called distance-preserving if for every two vertices x and y in H, the distance between x and y with respect to H equals the distance between x and y with respect to G. Clearly, every distance-preserving subgraph is induced and every subgraph induced by a convex set of vertices is distance-preserving. For some integer k, we say that a graph G has the property $\mathcal{E}(k)$ if G has no distance-preserving subgraph H of order k for which $V(H)$ is convex in G.

The properties $\mathcal{E}(k)$ represent natural necessary extension properties of graphs with small convexity number. In the present section we study the interplay between these properties and upper bounds on the convexity number. In fact, the exact value of the convexity number can easily be characterized using these properties.

Proposition 3 *If G is a graph of order n and k is such that $2 \leq k \leq n-1$, then $c(G) = k$ if and only if G does not have property $\mathcal{E}(k)$ but has property $\mathcal{E}(i)$ for every integer i with $k+1 \leq i \leq n-1$.*

Proof This follows immediately from the observation, that a graph G with $c(G) = k$ has a convex set of k vertices which induces a distance-preserving subgraph and that no set of vertices of cardinality between $k+1$ and $n-1$ is convex. □

For $k = 2$, Proposition 3 can be improved in the sense that we do not need the properties $\mathcal{E}(i)$ for $i > \max\left\{4, \left\lfloor \frac{n-2}{2} \right\rfloor\right\}$ (cf. Corollary 8 below).

The two properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$ will play a central role and the next proposition contains an alternative description of them.

Proposition 4 *Let G be a connected graph.*

(i) *G has property $\mathcal{E}(3)$ if and only if it is triangle-free and every induced P_3 is contained in an induced C_4.*
(ii) *G has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$ if and only if it is triangle-free, every induced P_3 is contained in an induced C_4, and every induced C_4 is contained in an induced $K_{2,3}$.*

Proof Since the proofs of (i) and (ii) are very similar, we only give details for the proof of (ii).

First, we assume that G has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$. By $\mathcal{E}(3)$, G can not contain a triangle T, because T would be distance-preserving and $V(T)$ convex. Furthermore, by $\mathcal{E}(3)$, every induced P_3 must be contained in an induced C_4 and, by $\mathcal{E}(4)$, every induced C_4 must be contained in an induced $K_{2,3}$.

Next, we assume that G is such that it is triangle-free, every induced P_3 is contained in an induced C_4, and every induced C_4 is contained in an induced $K_{2,3}$. If H is a distance-preserving subgraph of G of order 3, then, since G is connected, H is connected. Since G is triangle-free, H is a P_3. Since every induced P_3 is contained in an induced C_4, $V(H)$ is not convex in G. Similarly, if H is a distance-preserving subgraph of G of order 4, then, since G is connected and triangle-free, H is either an induced P_4, or an induced claw $K_{1,3}$, or an induced C_4. In the first two cases $V(H)$ is not convex, because every induced P_3 is contained in an induced C_4, and in the last case $V(H)$ is not convex, because every induced C_4 is contained in an induced $K_{2,3}$. Hence G has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$.

This completes the proof. □

Our next result shows that the extension property $\mathcal{E}(3)$ already implies a non-trivial upper bound on the convexity number.

Theorem 5 *If G is a connected graph of order n which has property $\mathcal{E}(3)$, then*

$$c(G) \leq \frac{n}{2}$$

with equality if and only if G arises from a graph H of order $\frac{n}{2}$ which has property $\mathcal{E}(3)$ by adding a disjoint isomorphic copy H' of H and adding a new edge between every vertex $u \in V(H)$ and its copy $u' \in V(H')$.

Proof Let G be a connected graph of order n which has property $\mathcal{E}(3)$. By Proposition 4 (i), G is triangle-free and every induced P_3 is contained in an induced C_4.

Let C be a convex set of vertices of cardinality $c(G)$. Let $R = V(G)\setminus C$.

If a vertex $v \in R$ has two neighbours u and w in C, then u and w are not adjacent and $v \in I[u, w]$, which contradicts the convexity of C. Hence every vertex in R has at most one neighbour in C.

Since G is connected, C induces a connected subgraph $G[C]$ of G and there is an edge $u_0u_0' \in E(G)$ with $u_0 \in C$ and $u_0' \in R$.

If $uv, uu' \in E(G)$ with $u, v \in C$ and $u' \in R$, then u' and v are not adjacent and vuu' is an induced P_3. By $\mathcal{E}(3)$, u' and v have a common neighbour v' such that $u'uvv'u'$ is an induced C_4. Since u' has at most one neighbour in C, $v' \in R$. Iteratively applying this observation to the edges of a spanning tree of $G[C]$ rooted at u_0 implies the existence of a matching $M = \{e_u \mid u \in C\}$ such that for every $u \in C$, $e_u = uu'$ for some $u' \in R$. This already implies $|C| \leq |R|$ and hence $c(G) \leq \frac{n}{2}$.
If $c(G) = \frac{n}{2}$, then M is a perfect matching.

If $uu', vv' \in M$ are such that $uv \in E(G)$ and $u'v' \notin E(G)$, then $u'uv$ is an induced P_3 and u' and v have a common neighbour $v'' \neq u$. Since $u'v' \notin E(G)$, $v'' = w'$ for some $ww' \in M$ with $w \in C\setminus\{v\}$. Now $w' \in R$ has two neighbours w and v in C which is a contradiction. Similarly, if $uu', vv' \in M$ are such that $uv \notin E(G)$ and

$u'v' \in E(G)$, then $u'v'v$ is an induced P_3 and u' and v have a common neighbour $v''' \neq v'$. Since $uv \notin E(G)$, $v''' = w'$ for some $ww' \in M$ with $w \in C \setminus \{u\}$. Now $w' \in R$ has two neighbours w and v in C which is a contradiction. Altogether, this implies that the mapping defined by $u \mapsto u'$ for every $e_u = uu' \in M$ is an isomorphism between $G[C]$ and $G[R]$. Since G is triangle-free, the graph $G[C]$ is triangle-free. If uvw is an induced P_3 in $G[C]$, then property $\mathcal{E}(3)$ implies the existence of a vertex x with $xu, xw \in E(G)$ and $xv \notin E(G)$. Since C is convex, $x \in C$. Altogether, this implies that $G[C]$ has property $\mathcal{E}(3)$, i.e. G is as described in the statement of the theorem.

Conversely, let G arises from a graph H of order $\frac{n}{2}$ which has property $\mathcal{E}(3)$ by adding a disjoint isomorphic copy H' of H and adding a new edge between every vertex $u \in V(H)$ and its copy $u' \in V(H')$. Since H has property $\mathcal{E}(3)$ and every induced P_3 which intersects $V(H)$ as well as $V(H')$ is contained in an induced C_4 by construction, G has property $\mathcal{E}(3)$ which implies $c(G) \leq \frac{n}{2}$. Furthermore, by construction, every path of length l in G between two vertices x and y in $V(H)$ which intersects $V(H')$ corresponds to a walk of length at most $l-2$ in H between x and y. Hence $V(H)$ is convex and $c(G) \geq |V(H)| = \frac{n}{2}$ which completes the proof. □

Adding the next extension property $\mathcal{E}(4)$, the upper bound from Theorem 5 improves only by 1 but the structure of the extremal graphs becomes far more restricted.

Theorem 6 *If G is a connected graph of order $n \geq 6$ which has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$, then*

$$c(G) \leq \frac{n-2}{2}$$

with equality if and only if either $n = 6$ or $n \geq 12$ and the vertex set $V(G)$ can be partitioned into four independent sets X, X', Y and Y' such that (cf. Fig. 3)

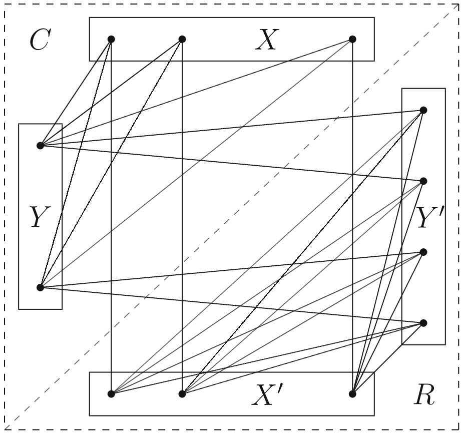
- $|X'| = |X|$, $|Y'| = 2|Y| = 4$,
- *$G[X \cup Y]$ and $G[X' \cup Y']$ are complete bipartite graphs,*
- *the edges between X and X' form a perfect matching,*
- *the edges between Y and Y' form two disjoint P_3s,*
- *and there are no edges between X and Y' or X' and Y.*

Proof Let G be a connected graph of order $n \geq 6$ which has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$. Let C be a convex set of vertices of cardinality $c(G)$. Let $R = V(G) \setminus C$. If $n = 6$ or $n = 7$, then the result easily follows. Hence we may assume $n \geq 8$ and $|C| \geq 3$.

Since G has property $\mathcal{E}(3)$, we obtain as in the proof of Theorem 5 that every vertex in R has at most one neighbour in C and that there is a matching $M = \{e_u \mid u \in C\}$ such that for every $u \in C$, $e_u = uu'$ for some $u' \in R$. Since C is convex and G is connected and has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$, the graph $G[C]$ is connected and has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$. Since $|C| \geq 3$, we obtain that $G[C]$ contains an induced $K_{2,3}$.

If $uv \in E(G)$ for $u, v \in C$, then $uvv'u'u$ is an induced C_4. By $\mathcal{E}(4)$, every induced C_4 is contained in an induced $K_{2,3}$. This implies that either u and v' have a common

Fig. 3 The structure of the extremal graphs for Theorem 6 and Corollary 7



neighbour $u'' \notin \{u', v\}$ or v and u' have a common neighbour $v'' \notin \{u, v'\}$, i.e. for every edge of $G[C]$ at least one of the two incident vertices has at least two neighbours in R. This implies that the number of vertices in C that have at least two neighbours in R, is at least $|C| - \alpha$, where α denotes the independence number of $G[C]$. Since $G[C]$ contains an induced $K_{2,3}$, α is at most $|C| - 2$ and we obtain

$$n = |C| + |R| = \alpha + (|C| - \alpha) + |R| \geq 2\alpha + 3(|C| - \alpha) = 3|C| - \alpha \geq 2|C| + 2$$

which implies $c(G) \leq \frac{n-2}{2}$.

If $c(G) = \frac{n-2}{2}$, then either $c(G) = 2$ and $n = 6$ or $c(G) \geq 5$ and equality holds throughout the above inequality chain. This implies that the independence number α of $G[C]$ equals $|C|-2$ and that there is an independent set $X \subseteq C$ of order $|C|-2$ such that every vertex in X has exactly one neighbour in R and every vertex in $Y = C \setminus X$ has exactly two neighbours in R. Since $G[C]$ is connected, has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$, and contains an induced $K_{2,3}$, we obtain that C induces a complete bipartite graph with partite sets X and Y.

If $u \in X$ and $v \in Y$, then let $u', v', v'' \in R$ be such that $uu', vv', vv'' \in E(G)$. Note that $uv \in E(G)$ and that vuu' is an induced P_3. By $\mathcal{E}(3)$ and $\mathcal{E}(4)$, we have $u'v', u'v'' \in E(G)$ and $v'v'' \notin E(G)$. This easily implies that R induces a complete bipartite graph with partite sets X' and Y' such that $|X'| = |X|$, $|Y'| = 2|Y| = 4$, the edges between X and X' form a perfect matching, and the edges between Y and Y' form two disjoint P_3's, i.e. G is as described in the statement of the theorem.

Conversely, if G be as described in the statement of the theorem, then it is easy to verify that G has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$, and $c(G) = \frac{n-2}{2}$ which completes the proof. □

Adding further extension properties, the upper bound from Theorem 6 does no longer improve. Only the lower bound on the order of the extremal graphs increases.

Corollary 7 *Let $k \geq 4$. If G is a connected graph of order $n \geq 2k + 4$ which has property $\mathcal{E}(i)$ for $3 \leq i \leq k$, then*

$$c(G) \leq \frac{n-2}{2}$$

with equality if and only if the vertex set $V(G)$ can be partitioned into four independent sets X, X', Y and Y' such that the conditions stated in Theorem 6 are satisfied (cf. Fig. 3).

Proof In view of Proposition 4 and Theorem 6, it remains to prove that the graphs G with $c(G) = \frac{n-2}{2}$ described in the statement of the result have property $\mathcal{E}(i)$ for $5 \leq i \leq k$. Note that $|X| = |X'| \geq k-1$.

Let H be a distance-preserving subgraph of G of order between 5 and k. If H contains the two vertices y_1 and y_2 in Y, then it does not contain at least one vertex x in X. Since $x \in I[y_1, y_2]$, H is not convex. Similarly, if H contains two vertices y_1' and y_2' in Y', then it does not contain at least one vertex x' in X'. Since $x' \in I[y_1', y_2']$, H is not convex. Hence H contains at most one vertex from Y and at most one vertex from Y'. If H contains two vertices x_1 and x_2 in X, then it does not contain at least one vertex y in Y. Since $y \in I[x_1, x_2]$, H is not convex. Similarly, if H contains two vertices x_1' and x_2' in X', then it does not contain at least one vertex y' in Y'. Since $y' \in I[x_1', x_2']$, H is not convex. Hence H contains at most one vertex from X and at most one vertex from X'.

This implies the contradiction that the order of H is at most four which completes the proof. □

For $k = 2$, Proposition 3 can be improved as follows.

Corollary 8 *If G is a graph of order $n \geq 5$, then $c(G) = 2$ if and only if G is connected and has property $\mathcal{E}(i)$ for every integer i with $3 \leq i \leq \max\left\{4, \left\lfloor \frac{n-2}{2} \right\rfloor\right\}$.*

Proof Let G be a graph of order at least 5 with $c(G) = 2$. If G has exactly two components, then two adjacent vertices from one component together with one further vertex from the other component form a convex set. Similarly, if G has at least three components, then three vertices each belonging to a different component form a convex set. Hence G is connected. By Proposition 3, G has property $\mathcal{E}(i)$ for every integer i with $3 \leq i \leq n-1$. Since $\max\left\{4, \left\lfloor \frac{n-2}{2} \right\rfloor\right\} \leq n-1$, this completes the proof of one implication.

Conversely, let G be a connected graph of order $n \geq 5$ and let G have property $\mathcal{E}(i)$ for every integer i with $3 \leq i \leq \max\left\{4, \left\lfloor \frac{n-2}{2} \right\rfloor\right\}$. This implies that G has properties $\mathcal{E}(3)$ and $\mathcal{E}(4)$.

If $n \geq 6$, then, by Theorem 6, $c(G) \leq \left\lfloor \frac{n-2}{2} \right\rfloor$. Since G has property $\mathcal{E}(i)$ for every integer i with $3 \leq i \leq \left\lfloor \frac{n-2}{2} \right\rfloor$, this implies $c(G) = 2$. If $n = 5$, then Proposition 3 implies $c(G) = 2$ which completes the proof. □

The graph which arises by identifying an edge from a complete graph K_k with $k \geq 2$ with an edge from a complete bipartite graph $K_{r,s}$ with $r, s \geq 2$, does not have property $\mathcal{E}(k)$ but has properties $\mathcal{E}(i)$ for every integer i with $k+1 \leq i \leq r+s-1$. This implies that for $k \geq 3$, there is no improvement of Proposition 3 similar to Corollary 8.

Note that the class of graphs with convexity number 2 is structurally quite rich in the sense that every connected graph G is an induced subgraph of a graph G' with

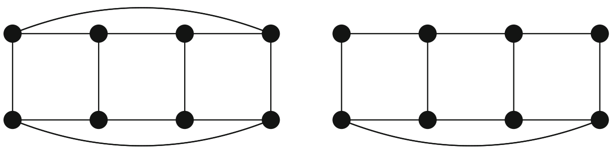


Fig. 4 Q_3 and $Q_3 - e$

$c(G') = 2$. (Such a graph G' can be constructed from G for instance by replacing every vertex u of G with two vertices u_1 and u_2 and replacing every edge uv of G with four edges u_1v_1, u_1v_2, u_2v_1, and u_2v_2. Clearly, G is an induced subgraph of G' and it is easy to check that $c(G') = 2$.)

In our last result, we identify a class of graphs G for which $c(G)$ equals 2 if and only if G has the extension property $\mathcal{E}(3)$.

Theorem 9 *If G is a connected graph of order at least* 2 *which does not contain the cube* Q_3 *or the cube minus an edge* $Q_3 - e$ *as an induced subgraph* (*cf. Fig.* 4)*, then* $c(G) = 2$ *if and only G has property* $\mathcal{E}(3)$.

Proof If $c(G) = 2$, then Proposition 3 implies that G has property $\mathcal{E}(3)$.

Now, let G have property $\mathcal{E}(3)$. For contradiction, we assume that C is a convex set of vertices of cardinality at least 3 which does not contain all vertices of G. Let $R = V(G)\setminus C$. Since G is connected and has property $\mathcal{E}(3)$, C induces a connected triangle-free graph and there are adjacent vertices $u \in C$ and $u' \in V(G)\setminus C$. Let $v \in C$ be a neighbour of u. Since $u'uv$ is an induced P_3, there is a vertex v' different from u such that $vv', u'v' \in E(G)$. Clearly, $v' \in R$. Since C has at least 3 elements, we may assume that there is a vertex $w \in C$ different from u such that $vw \in E(G)$. Since $v'vw$ is an induced P_3, there is a vertex w' different from v such that $ww', v'w' \in E(G)$. Clearly, $w' \in R$. Since uvw is an induced P_3, there is a vertex x different from v such that $ux, wx \in E(G)$. Clearly, $x \in C$. Since $w'wx$ is an induced P_3, there is a vertex x' different from w such that $xx', w'x' \in E(G)$. Clearly, $x' \in R$.

Now the vertices u, v, w, x, u', v', w', and x' induce either Q_3 or $Q_3 - e$ which is a contradiction. □

References

1. Brandstädt, A., Le, V.B., Spinrad, J.P.: Graph classes: a survey. Siam Monogr. Discret. Math. (1999)
2. Cáceres, J., Márquez, A., Oellermann, O.R., Puertas, M.L.: Rebuilding convex sets in graphs. Discret. Math. **297**, 26–37 (2005)
3. Canoy, S.R. Jr., Garces, I.J.L.: Convex sets under some graph operations. Graphs Comb. **4**, 787–793 (2002)
4. Chartrand, G., Wall, C.E., Zhang, P.: The convexity number of a graph. Graphs Comb. **18**, 209–217 (2002)
5. Edelman, P.H., Jamison, R.E.: The theory of convex geometries. Geom. Dedicata. **19**, 247–270 (1985)
6. Farber, M., Jamison, R.E.: Convexity in graphs and hypergraphs. SIAM J. Algebraic Discret. Methods **7**, 433–444 (1986)
7. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-completeness. W.H. Freeman, New York (1979)
8. Gimbel, J.G.: Some remarks on the convexity number of a graph. Graphs Comb. **19**, 357–361 (2003)

9. Harary, F., Nieminen, J.: Convexity in graphs. J. Differ. Geom. **16**, 185–190 (1981)
10. Kim, B.K.: A lower bound for the convexity number of some graphs. J. Appl. Math. Comput. **14**, 185–191 (2003)
11. McConnell, R.M., Spinrad, J.P.: Linear-time modular decomposition and efficient transitive orientation of comparability graphs. In: Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 536–545 (1994)
12. McConnell, R.M., Spinrad, J.P.: Ordered vertex partitioning. Discret. Math. Theor. Comput. Sci. **4**, 45–60 (2000)