On the number of connected convex subgraphs of a connected acyclic digraph

Gregory Gutin* Anders Yeo†

Abstract

A digraph D is connected if the underlying undirected graph of D is connected. A subgraph H of an acyclic digraph D is convex if there is no directed path between vertices of H which contains an arc not in H. We find the minimum and maximum possible number of connected convex subgraphs in a connected acyclic digraph of order n. Connected convex subgraphs of connected acyclic digraphs are of interest in the area of modern embedded processors technology.

1 Introduction

To speed up computations, modern embedded processors often accommodate both a conventional processor core and a large amount of special purpose hardware. The current trend is to transfer computation from software to special purpose hardware. To do that, one should analyze the dataflow graph G of the program under consideration and generate some special subgraphs of G (see, e.g., [1, 2, 4] and a large number of references there). Notice that a dataflow graph G is always a connected acyclic digraph and the main desired property of a subgraph H of G of interest is convexity. A subgraph H is *convex* if there is no directed path between vertices of H which contains an arc not in H. Clearly, every convex subgraph is an induced subgraph. Often the connectivity property is also imposed: a subgraph H is *connected* if its underlying undirected graph is connected.

As a result, mainly connected convex subgraphs (*cc-subgraphs*) of G are of interest and should be generated and analyzed. When one designs algorithms to generate cc-subgraphs (see, e.g., [2, 4]), one arrives at the following natural question: what are the smallest and largest possible numbers of cc-subgraphs in a connected acyclic digraph on n vertices? In this short paper, we answer this question. In fact, we prove that the minimum possible

*Corresponding author. Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, gutin@cs.rhul.ac.uk

†Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, anders@cs.rhul.ac.uk

number of cc-subgraphs in a connected acyclic digraph of order n is $n(n+1)/2$. The maximum possible number of cc-subgraphs in a connected acyclic digraph of order n is $2^n + n + 1 - d_n$, where $d_n = 2 \cdot 2^{n/2}$ for every even n and $d_n = 3 \cdot 2^{(n-1)/2}$ for every odd n.

Notice that a key idea in the proof of Theorem 2.1 is used to design an algorithm for generating all cc-subgraphs in a connected acyclic digraph D in time $O(n \cdot cc(D))$ in the very recent paper [2], where n is the order of D and the claim of Theorem 2.1 is used to show the complexity of the algorithm in [2]. (The authors of [2] refer to this paper for the proof of Theorem 2.1.)

It is easy to evaluate the maximum and minimum possible number of convex (but not necessarily connected) subgraphs in a connected acyclic digraph of order n. The maximum number is $2^n - 1$ due to the digraph obtained from $K_{1,n-1}$ by orienting all edges from the partite set of cardinality 1 to the partite set of cardinality $n-1$. Let D be a connected acyclic digraph and let $x_1, x_2, \ldots, x_n$ be an acyclic ordering of the vertices in D (i.e., if $x_i x_j$ is an arc, then $i < j$) [3]. Observe that all subgraphs induced by the sets of the form $\{x_i, x_{i+1}, \ldots, x_j\}$ ($i \leq j$) are convex. Thus, the minimum number of convex subgraphs is at least $n(n+1)/2$. Clearly, if $x_1 x_2 \ldots x_n$ is a Hamilton directed path, then only the subgraphs described above are convex. Thus, the minimum number of convex subgraphs is $n(n+1)/2$. Our result on the minimum number of cc-subgraphs strengthens this simple result.

For a digraph D and a vertex x, $cc(D)$ ($cc(D,x)$) denotes the number of cc-subgraphs (cc-subgraphs containing x). The number of out-neighbors (in-neighbors) of x in D, is called the *out-degree* (*in-degree*) of x and is denoted by $d^+_D(x)$ ($d^-_D(x)$). For a digraph D, $V(D)$ and $A(D)$ denote the vertex and arcs sets of D and for $X \subseteq V(D)$, $D[X]$ is the subgraph of D induced by X. For further basic terminology and notation in digraph theory, see [3].

2 Lower Bound

Let D_n be a connected acyclic digraph and let D_n have a Hamilton directed path $x_1 x_2 \ldots x_n$. Observe that all cc-subgraphs of D_n are of the form $D_n[\{x_i, x_{i+1}, \ldots, x_j\}]$, where $i \leq j$. Thus, $cc(D_n) = n(n+1)/2$. In the following theorem, we show that no connected acyclic digraphs of order n has less cc-subgraphs than D_n.

Theorem 2.1 *Let H be a connected acyclic digraph of order n and let z be a vertex of H. Then $cc(H,z) \geq n$ and $cc(H) \geq n(n+1)/2$.*

Proof: We will show the theorem by induction on n. It clearly holds for $n = 1$ so let $n > 1$. Let x be any vertex in H with $d^+_H(x) = 0$ and let $H' = H - x$. Let $R_1, R_2, \ldots, R_k$ be the connected components of H', where $k \geq 1$. Let $n_i = |V(R_i)|$ and let $H_i = H[V(R_i) \cup \{x\}]$.

First assume that $k \geq 2$. Let S_i be any cc-subgraph in H_i which contains x. By the induction hypothesis, there are at least n_i+1 such cc-subgraphs. Note that $S_1 \cup S_2 \cup \ldots \cup S_k$ is a cc-subgraph containing x. Since $(n_1+1)(n_2+1)\cdots(n_k+1) \geq n_1+n_2+\ldots+n_k+1 = n$ we have shown that there are at least n cc-subgraphs in H containing x.

Let $w \in V(H) \setminus \{x\}$ be arbitrary. Without loss of generality we may assume that $w \in V(R_1)$. By the induction hypothesis, there are at least n_1 cc-subgraphs containing w which do not contain x (all are in R_1). Note that $H_1 \cup S_2 \cup \ldots \cup S_k$ is a cc-subgraph containing w and x. Since $(n_2+1)\cdots(n_k+1) \geq n_2+\ldots+n_k+1 = n-n_1$, we have shown that there are at least n cc-subgraphs in H containing w.

We will now show that $cc(H) \geq n(n+1)/2$. By the induction hypothesis, there are at least $n_i(n_i+1)/2$ cc-subgraphs in R_i. Furthermore, we saw above that we have at least $(n_1+1)(n_2+1)\cdots(n_k+1)$ cc-subgraphs in H containing x. Thus, we get the following:

$$\begin{aligned} cc(H) &\geq \sum_{i=1}^{k} \frac{n_i(n_i+1)}{2} + \prod_{i=1}^{k}(n_i+1) \\ &\geq \sum_{i=1}^{k}(n_i^2+n_i)/2 + \sum_{1\leq i<j\leq k} n_i n_j + \sum_{i=1}^{k} n_i + 1 \\ &= \frac{1}{2}\left[\left(\sum_{i=1}^{k} n_i\right)^2 + 3\left(\sum_{i=1}^{k} n_i\right) + 2\right] \\ &= [(n-1)^2 + 3(n-1) + 2]/2 = n(n+1)/2. \end{aligned}$$

So now consider the case when $k = 1$.

Let $w \in V(H')$ be arbitrary. By the induction hypothesis, there are $n-1$ cc-subgraphs containing w in H'. Since H is also a cc-subgraph we have at least n cc-subgraphs containing w. Let H^* be the converse of H (H^* is obtained from H by reversing all arcs of H). Let $y \in V(H^*)$ be arbitrary with $d^+_{H^*}(y) = 0$ (i.e., $d^-_H(y) = 0$). By considering $H^* - y$ instead of H' we observe that there are also at least n cc-subgraphs in H containing x (it does not matter whether $H^* - y$ is connected or not since we have already looked at the non-connected case).

Now we are able to show that $cc(H) \geq n(n+1)/2$. By the induction hypothesis, there are at least $(n-1)n/2$ cc-subgraphs in H' and they are all cc-subgraphs in H as $d^+_H(x) = 0$. Since there are also at least n cc-subgraphs containing x, we conclude that $cc(H) \geq (n-1)n/2 + n = n(n+1)/2$. $\diamond$

3 Upper Bound

Consider a complete bipartite graph $K_{a,b}$ with partite sets A, B ($|A| = a$, $|B| = b$) and orient all its edges from A to B. We have obtained the bipartite tournament $\vec{K}_{a,b}$.

Lemma 3.1 *Let $n = a + b$. We have $cc(\vec{K}_{a,b}) = 2^{a+b} - 2^a - 2^b + a + b + 1$ and*

$$\max\{cc(\vec{K}_{a,b}) :\ a + b = n\} = 2^n + n + 1 - d_n,$$

where $d_n = 2 \cdot 2^{n/2}$ for every even n and $d_n = 3 \cdot 2^{(n-1)/2}$ for every odd n.

Proof: Let $g(a, b) = 2^{a+b} - 2^a - 2^b + a + b + 1$. Since all non-empty sets of vertices of $\vec{K}_{a,b}$, excluding those that are subsets of A or B of cardinality at least 2, induce cc-subgraphs, we have $cc(\vec{K}_{a,b}) = g(a, b)$. It remains to observe that $\max\{g(a, b) :\ a + b = n\}$ is obtained when a and b differ by at most 1. $\diamond$

In the following theorem, we will show that the bipartite tournaments $\vec{K}_{a,n-a}$ with $|n - 2a| \le 1$ have the maximum possible number of cc-subgraphs.

Theorem 3.2 *Let H be a connected acyclic digraph of order n and let $f(n) = 2^n + n + 1 - d_n$, where $d_n = 2 \cdot 2^{n/2}$ for every even n and $d_n = 3 \cdot 2^{(n-1)/2}$ for every odd n. Then $cc(H) \ge f(n)$.*

Proof: Clearly, we may assume that $n \ge 3$. Suppose that H has a directed path of length 2. We will prove that $cc(H) \le f(n)$. If xyz is a directed path of length 2 in H, then we have the following:

(C1) There are at most 2^{n-2} cc-subgraphs containing x but not z.

(C2) There are at most 2^{n-2} cc-subgraphs containing z but not x.

(C3) There are at most $2^{n-2} - 1$ cc-subgraphs containing neither x nor z.

(C4) There are at most 2^{n-3} cc-subgraphs containing x and z.

(C4) is true as if x and z belong to a cc-subgraph, then y has to belong to it as well. Therefore there are at most $7 \cdot 2^{n-3} - 1$ cc-subgraphs. Observe that $7 \cdot 2^{n-3} - 1 \le f(n)$ for every $n \ge 3$ apart from $n = 5$. Indeed, it is not difficult to prove that $f(n) - 7 \cdot 2^{n-3} + 1 > 2^{\frac{n}{2}+1}(2^{\frac{n}{2}-4} - 1)$ for every even n and that $f(n) - 7 \cdot 2^{n-3} + 1 > 2^{\frac{n-1}{2}}(2^{\frac{n-5}{2}} - 3)$ for every odd n. These two inequalities imply $7 \cdot 2^{n-3} - 1 \le f(n)$ for each $n \ge 8$. The cases $n = 3, 4, 6, 7$ can be easily checked separately. Thus, it remains to consider the case $n = 5$.

Suppose that H has a directed path P with $n-1$ vertices and let u be the vertex not on P. Then by the discussion before Theorem 2.1, $cc(H-u) = n(n-1)/2$. There are at most 2^{n-1} induced subgraphs of H containing u. Thus, $cc(H) \leq 2^{n-1} + n(n-1)/2$. Observe that $2^{n-1} + n(n-1)/2 \leq f(n)$ for every $n \geq 5$. Thus, we may assume that if $n \geq 5$, then H has no directed path with $n-1$ vertices.

Let $n = 5$ and let $u \in V(H) \setminus \{x, y, z\}$. By (C4), 2^{n-3} subgraphs containing x and z are not cc-subgraphs. Observe that $(2^n - 1 - 2^{n-3}) - f(n) = 1$ for $n = 5$. Thus, to show that $cc(H) \leq f(5)$, it suffices to find a non-cc-subgraph of H that does not contain at least one of the vertices x and z. Since H has no directed path of length 3, u is not adjacent with at least one of the vertices x, y, z. The subgraph induced by any such pair of non-adjacent vertices is not a cc-subgraph.

So we may now assume that there is no directed path of length 2. This means that the vertices can be partitioned into sets A and B such that A contains all vertices with in-degree zero and B contains all the vertices with out-degree zero. Observe that now every connected induced subgraph of H is a cc-subgraph. This implies that $cc(H)$ is maximum when there is an arc from a to b for each $a \in A$, $b \in B$. Now our result follows from Lemma 3.1. $\diamond$

Acknowledgements We are thankful to Adrian Johnstone, Joseph Reddington and Elizabeth Scott for useful discussions on the topic of the paper. Research of both authors was supported in part by an EPSRC grant.

References

[1] K. Atasu, L. Possi and P. Ienne, Automatic application-specific instruction-set extensions under microarchitectural constraints. In *Proc. 40th Conf. Design Automation*, ACM Press (2003), 256–261.

[2] P. Balister, S. Gerke, G. Gutin, A. Johnstone, J. Reddington, E. Scott, A. Soleimanfallah and A. Yeo, Algorithms for generating convex sets in acyclic digraphs. Submitted.

[3] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications.* Springer-Verlag, London, 2000, 754 pp.

[4] P. Yu and T. Mitra, Scalable custom instructions identification for instruction-set extensible processors. In *Proc. 2004 Int. Conf. Compilers, Architecture, and Synthesis for Embedded Systems* (2004), 69–78.