Convex sets in acyclic digraphs

Paul Balister*§ Stefanie Gerke† Gregory Gutin‡

September 29, 2008

Abstract

A non-empty set X of vertices of an acyclic digraph is called connected if the underlying undirected graph induced by X is connected and it is called convex if no two vertices of X are connected by a directed path in which some vertices are not in X. The set of convex sets (connected convex sets) of an acyclic digraph D is denoted by $\mathcal{CO}(D)$ ($\mathcal{CC}(D)$) and its size by $\mathrm{co}(D)$ ($\mathrm{cc}(D)$). Gutin, Johnstone, Reddington, Scott, Soleimanfallah, and Yeo (Proc. ACiD'07) conjectured that the sum of the sizes of all convex sets (connected convex sets) in D equals $\Theta(n \cdot \mathrm{co}(D))$ ($\Theta(n \cdot \mathrm{cc}(D))$) where n is the order of D.

In this paper we exhibit a family of connected acyclic digraphs with $\sum_{C \in \mathcal{CO}(D)} |C| = o(n \cdot \mathrm{co}(D))$ and $\sum_{C \in \mathcal{CC}(D)} |C| = o(n \cdot \mathrm{cc}(D))$. We also show that the number of connected convex sets of order k in any connected acyclic digraph of order n is at least $n - k + 1$. This is a strengthening of a theorem of Gutin and Yeo.

1 Introduction

Let D be an acyclic digraph of order n. A non-empty set X of vertices in D is *connected* if the underlying undirected graph of $D[X]$, the subgraph of D induced by X, is connected. A non-empty set X of vertices in D is *convex* if there is no directed path in D between vertices of X containing a vertex not in X. The set of all convex sets of D is denoted by $\mathcal{CO}(D)$ and its size by $\mathrm{co}(D)$. The set of all connected convex sets of D is denoted by $\mathcal{CC}(D)$ and

*Department of Mathematical Sciences, University of Memphis, TN 38152-3240, USA. E-mail: pbalistr@memphis.edu

†Department of Mathematics, Royal Holloway, University of London, Egham, TW20 0EX, UK, E-mail: stefanie.gerke@rhul.ac.uk

‡Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, E-mail: gregory.gutin@rhul.ac.uk

its size by $\mathrm{cc}(D)$. Convex sets and connected convex sets in acyclic digraphs are of interest in the field of custom computing in which central processor architectures are parameterized for particular applications, see, e.g., [3, 4]. Gutin, Johnstone, Reddington, Scott, Soleimanfallah, and Yeo [4] introduced an algorithm $\mathcal{A}$ determining all connected convex sets of D in time $O(n \cdot \mathrm{cc}(D))$. They observed that $\mathcal{A}$ can be modified to produce all convex sets in time $O(n \cdot \mathrm{co}(D))$. The authors of [4] conjectured that the sum of the sizes of all convex sets (all connected convex sets, respectively) in D equals $\Theta(n \cdot \mathrm{co}(D))$ ($\Theta(n \cdot \mathrm{cc}(D))$, respectively). If the conjecture were true, then their algorithms would be optimal. The conjecture can be formulated differently. Let $\bar{s}_{\mathrm{co}}(D)$ and $\bar{s}_{\mathrm{cc}}(D)$ be the average size of a convex set and the average size of a connected convex set in D. The conjecture claims that $\bar{s}_{\mathrm{co}}(D) = \Theta(n)$ and $\bar{s}_{\mathrm{cc}}(D) = \Theta(n)$.
In this paper we disprove both parts of the conjecture by exhibiting a family $\mathcal{F}$ of digraphs for which $\bar{s}_{\mathrm{co}}(D) = O(\sqrt{n})$ and $\bar{s}_{\mathrm{cc}}(D) = O(\sqrt{n})$; see Section 2. In Section 3 we show that each connected digraph of order n contains at least $n-k+1$ connected convex sets of size k for each $1 \le k \le n$. This extends a result of Gutin and Yeo [5] who showed that each connected acyclic digraph of order n has at least $n(n+1)/2$ connected convex sets.
To simplify notation in the rest of the paper, we use n for the order of the digraph under consideration; $[m]$ will denote the set $\{1, 2, \dots, m\}$ (m is a positive integer). A vertex x of D is a *source* (*sink*) if its in-degree $d^-(x)$ (out-degree $d^+(x)$) equals zero. A vertex v of a connected digraph D is a *cut-vertex* if $D - v$ is not connected, i.e., $V(D) - v$ is not connected.

2 Counterexample

Theorem 2.1. *There is a family $\mathcal{F}$ of digraphs such that $\bar{s}_{\mathrm{co}}(D) = O(\sqrt{n})$ and $\bar{s}_{\mathrm{cc}}(D) = O(\sqrt{n})$ for each $D \in \mathcal{F}$.*

Proof. For $t = 1, 2, \dots$ and $r = \lceil \sqrt{t} \rceil$, the acyclic digraph D_t consists of vertex set $V(D_t) = X \cup Y \cup \{z\} \cup Y' \cup X'$, where

$$X = \{x_i : \; i \in [t]\}, \qquad X' = \{x'_i : \; i \in [t]\},$$
$$Y = \{y_j : \; i \in [r]\}, \qquad Y' = \{y'_j : \; i \in [r]\},$$

and arc set

$$A(D_t) = \{x_i x_{i+1}, x'_i x'_{i+1} : \; i \in [t-1]\} \cup \{x_t y_j, y_j z, z y'_j, y'_j x'_1 : j \in [r]\}.$$

For illustration, see Figure 1.

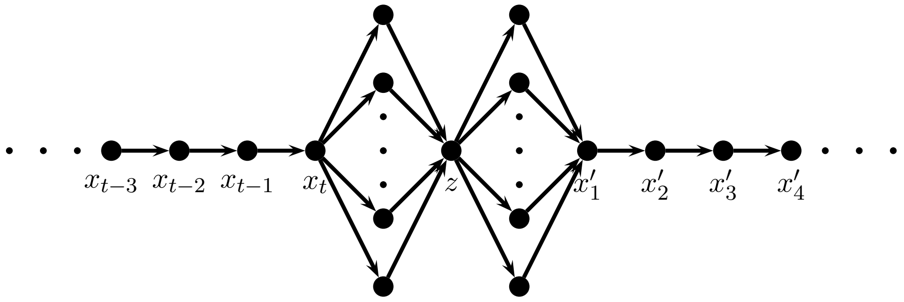

Figure 1: Digraphs from $\mathcal{F}$

Consider the family $\mathcal{C}$ of convex sets of D_t of size at least $2r+2$. Observe that each set in $\mathcal{C}$ contains a vertex in $X \cup X'$. Thus, $|\mathcal{C}| = |\mathcal{C}_X| + |\mathcal{C}_{X'}| + |\mathcal{C}_{X,X'}|$, where $\mathcal{C}_X$ ($\mathcal{C}_{X'}$, $\mathcal{C}_{X,X'}$, respectively) is the family of sets in $\mathcal{C}$ containing a vertex in X but not in X' (a vertex in X' but not in X, vertices in both X and X', respectively). By symmetry, $|\mathcal{C}| = 2 \cdot |\mathcal{C}_X| + |\mathcal{C}_{X,X'}|$. Observe that $\mathcal{C}_X$ consists of
(a) $\Theta(t^2)$ sets in $\mathcal{C}$ containing only vertices in X;
(b) $\Theta(t2^r)$ sets in $\mathcal{C}$ containing only vertices in $X \cup Y$ but at least one vertex in Y;
(c) $\Theta(t2^r)$ sets in $\mathcal{C}$ containing z and possibly some vertices in Y'.
Thus, $|\mathcal{C}_X| = \Theta(t2^r)$. In addition, $|\mathcal{C}_{X,X'}| = \Theta(t^2)$ and, hence, $|\mathcal{C}| = \Theta(t2^r)$. For each $Q \subseteq Y$ and $Q' \subseteq Y'$, the set $Q \cup \{z\} \cup Q'$ is connected and convex. So, there are 2^{2r} connected convex sets contained in the set $Y \cup \{z\} \cup Y'$., and thus there are $\Omega(2^{2r})$ connected convex sets in D_t. Hence,

$$
\begin{aligned}
\bar{s}_{\mathrm{co}}(D) = \frac{\sum_{C \in \mathcal{CO}(D)} |C|}{\mathrm{co}(D)} &\leq \frac{1}{\mathrm{cc}(D)} \left(\sum_{\substack{C \in \mathcal{CO}(D) \\ |C| > 2r+1}} |C| \right) + (2r+2) \\
&\leq O\left(t \frac{t2^r}{2^{2r}} \right) + (2r+2) = O(r),
\end{aligned}
$$

and similarly

$$
\bar{s}_{\mathrm{cc}}(D) \leq O\left(t \frac{t2^r}{2^{2r}} \right) + (2r+2) = O(r).
$$

□

Let us remark that the digraph in the family $\mathcal{F}$ in the previous proof have asymptotically the same number of convex sets and connected convex sets.

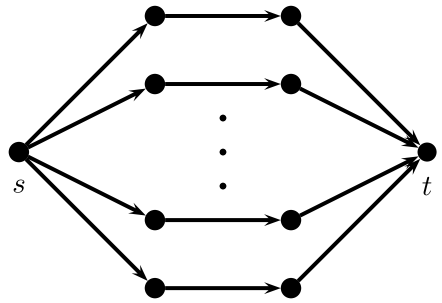

Figure 2: Digraphs from $\mathcal{G}$

This is generally not true as can be seen by considering the family of digraphs $\mathcal{G} = \{G_1, G_2, \ldots\}$, where the digraph G_i consists of a source s, a sink t and i internally-disjoint directed paths with two internal vertices each; see Figure 2. Then any non-empty subset of $V(G_i) \setminus \{s, t\}$ is convex but only $3i$ of these $4^i - 1$ sets are connected. There are 3^i connected convex sets containing s but not t, 3^i connected convex sets containing t but not s and only one connected convex set containing s and t (namely $V(G_i)$). Hence G_i contains at least $4^i - 1$ convex sets but at most $2 \cdot 3^i + 3i + 1$ connected convex sets and thus $\mathrm{cc}(G_i)/\mathrm{co}(G_i) \to 0$ as $i \to \infty$.

3 Connected convex sets of size k

In this section we show that any acyclic digraph D of order n contains at least $n - i + 1$ connected convex sets of order i for each $i \in [n]$. To do so we first show that every connected convex set that is not the entire graph can be extended by one vertex, and that there always exists a source or a sink that is not a cut-vertex.

Lemma 3.1. *Let $D = (V, A)$ be a connected acyclic digraph and let $H \neq V$ be a connected convex set in D. Then there exists a vertex v in $V \setminus H$ such that $H \cup \{v\}$ is a connected convex set in D.*

Proof. Since D is connected there is an arc uv with either $u \in V \setminus H$ and $v \in H$, or $u \in H$ and $v \in V \setminus H$. We may assume that $u \in H$ and $v \in V \setminus H$ as the other case can be treated similarly. Consider a longest directed path P in which the initial vertex is in H, the terminal vertex is v and all other vertices are in $V \setminus H$. Let y and w be the initial and second vertices in P, respectively. We will prove that $H \cup \{w\}$ is a connected convex set. It is

easily seen that $H \cup \{w\}$ is connected as H is connected and there is an arc yw. Let x be an arbitrary vertex of H. Observe that it suffices to show that there is no directed path between x and w having intermediate vertices that are all not in $H \cup \{w\}$.
Suppose that there is a directed path Q from w to x. Since D is acyclic, $y \neq x$ and we obtain a directed path yQ between y and x containing w and thus vertices outside H, a contradiction with the convexity of H. Suppose that there is a directed path R from x to w having intermediate vertices that are not in $H \cup \{w\}$. If R and P have a vertex $z \neq w$ in common then there is a closed walk from z to itself via w, which is impossible as D is acyclic. If R and P have no vertex other than w in common, then $R(P-y)$ is a directed path from x to v with all intermediate vertices in $V \setminus H$, and it is longer than P. This contradicts the choice of P. □

Lemma 3.2. *Let D be a connected acyclic digraph of order $n \geq 2$. Then there exist at least two non-cut-vertices that are a source or a sink.*

Proof. We prove the result by induction on n. For $n = 2$ the result clearly holds as both vertices are non-cut-vertices and one is a source and one is a sink. Now assume that the lemma is true for all connected acyclic digraphs of order at most $n_0 \geq 2$, and let D be a connected acyclic digraph of order $n_0 + 1$. Since D is acyclic there exists a source s and sink t of D; see for example [2]. If s and t are non-cut-vertices, then the claim of the lemma is satisfied. So assume that at least one of s and t, say s, is a cut-vertex. Let $C_1, \ldots, C_k$ be the vertex sets of the components of $D - s$. Note that $k \geq 2$ and that for all $i = 1, \ldots, k$, $D[C_1 + s]$ is a connected acyclic digraph of order at least 2 and at most n_0. Thus, by the induction hypothesis each of the digraphs $D[C_1 + s]$ and $D[C_2 + s]$ has at least two non-cut-vertices that are either a source or a sink. Choose two of these vertices that are not equal to s. It is easily verified that these vertices are non-cut-vertices and are sources or sinks of D. □

Theorem 3.3. *Let D be a connected acyclic digraph of order n. Then there exist at least $n - i + 1$ connected convex sets in D of order i for each $i \in [n]$.*

Proof. We prove the result by induction on n. For $n = 1$ the result is trivial. Assume that the claim of the theorem holds for all digraphs of order $1 \leq n \leq n_0$. We will show the result for an arbitrary connected acyclic digraphs D of order $n_0 + 1$. Since D is connected and acyclic, by Lemma 3.2 there exists a source or a sink s of D that is not a cut-vertex. By the induction hypothesis $D - s$ contains $n_0 - i + 1$ connected convex sets of size

$i = 1, 2, \ldots, n_0$. These sets are also connected convex sets of D as s is a source or sink and thus no directed path between vertices in $D - s$ can use s. In addition by a using Lemma 3.1 $i - 1$ times we can deduce that there is at least one connected convex set in D of size i containing s. Thus, there exist at least $n_0 - i + 1 + 1 = (n_0 + 1) - i + 1$ sets of size i for each $i \in [n_0]$. Since the vertex set of the digraph is a connected convex set there is also (exactly) one such set of order $n_0 + 1$. $\square$

Let us remark that the directed path of order n has exactly $n-i+1$ connected convex subgraphs of order i for each $i \in [n]$.

4 Open Question

Recall that the authors of [4] designed an $O(n \cdot cc(D))$-time algorithm to generate all connected convex sets of a connected acyclic digraph D. As we showed in Section 2 there could be an asymptotically faster algorithm. Is there an $O(\sum_{C \in \mathcal{CC}(D)} |C|)$-time algorithm? It is known [1] that there is an algorithm for generating all (not necessarily connected) convex sets of a connected acyclic digraph D in $O(\sum_{C \in \mathcal{CO}(D)} |C|)$ time.

Acknowledgement Research of Gutin was supported in part by an EPSRC grant.

References

[1] 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. To appear in *J. Discrete Algorithms.*

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

[3] X. Chen, D.L. Maskell, and Y. Sun, Fast identification of custom instructions for extensible processors. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.* **26** (2007), 359–368.

[4] G. Gutin, A. Johnstone, J. Reddington, E. Scott, A. Soleimanfallah, and A. Yeo, An algorithm for finding connected convex subgraphs of an acyclic digraph, in Proc. ACiD'07 (Algorithms and Complexity in Durham), College Publications, 2008.

[5] G. Gutin and A. Yeo, On the number of connected convex subgraphs of a connected acyclic graph. To appear in *Discrete Appl. Math.*