# A Lower Bound on the Size of Series-Parallel Graphs Dense in Long Paths

Chris Calabro

November 18, 2008

**Abstract**

One way to quantify how dense a multidag is in long paths is to find the largest $n, m$ such that whichever $\leq n$ edges are removed, there is still a path from an original input to an original output with $\geq m$ edges - the larger we can make $n, m$, the denser is the graph. For a given $n, m$, we would like to lower bound the size such a graph, say in edges, at least when restricting to a particular class of graphs. A bound of $\Omega(n \lg m)$ was provided in [Val77] for one notion of series-parallel graphs. Here we reprove the same result but in greater detail and relate that notion of series-parallel to other popular notions of series-parallel. In particular, we show that that notion is more general than minimal series-parallel and two terminal series-parallel.

## 1 Introduction

Theorem 5.4 appearing in [Val77] states that if a series-parallel graph $G$ has the property that (whichever $\leq n$ edges are removed, there is still a path from an original input node to an original output node with $\geq m$ edges) then $G$ has $\Omega(n \lg m)$ edges. [1] The definition of series-parallel graph there is misstated, yielding a trivial class of graphs; but a slight modification that still allows the proof to work yields a very large class of graphs. (and so this definition is what was probably originally intended) While the proof there was in outline form, the proof here will be much more specific about how to decompose a graph, whether input, output edges are counted in a particular path or graph, and whether input, output are with respect to the original graph or the parts into which it is decomposed. The main purpose of the extra rigor is to test this modified definition, which we will call *Valiant series-parallel* (VSP).

We relate VSP to other popular definitions of series-parallel - minimal series-parallel (MSP), two terminal series-parallel (TTSP), general series-parallel (GSP) - and show

$$\mathrm{MSP} \cup \mathrm{TTSP} \subseteq \mathrm{GSP} \cap \mathrm{VSP},$$

[1] Theorem 5.4 [Val77] actually stated $\Omega(n \lg \lg m)$, but this was a typo.

as well as basic relationships among them. Since GSP and VSP are subset incomparable, despite its name, GSP is therefore not the only reasonable generalization of the MSP and TTSP notions of series-parallel.

## 2 Series-parallel graphs

There are several incompatible definitions of series-parallel graphs. We will explore a few of the most popular ones here. A comprehensive survey can be found in [Val78]. A *multidag* $G = (V, E)$ is a directed-acyclic multigraph. Define

$$\text{input}(G) = \{v \in V \mid \deg_{\text{in}}(v) = 0\}$$
$$\text{output}(G) = \{v \in V \mid \deg_{\text{out}}(v) = 0\}$$

We now define the class of two terminal series-parallel multidags (TTSP). A multidag $G$ is *two terminal* (TT) iff $|\text{input}(G)| = |\text{output}(G)| = 1$ and $\text{input}(G) \neq \text{output}(G)$. The class of *elementary* TTSPs is those multidags with 2 nodes and a single edge between them. For $i = 1, 2$, let $G_i$ be TT with input $a_i$ and output $b_i$. Define the *TT parallel composition* $P_{\text{TT}}(G_1, G_2)$ as the disjoint union of $G_1, G_2$ but with $a_1, a_2$ identified and $b_1, b_2$ identified, and the *TT series composition* $S_{\text{TT}}(G_1, G_2)$ as the disjoint union of $G_1, G_2$ but with $b_1, a_2$ identified. Note that $P_{\text{TT}}(G_1, G_2), S_{\text{TT}}(G_1, G_2)$ are both TT, $S_{\text{TT}}$ is associative but not commutative, and $P_{\text{TT}}$ is associative and commutative. Associativity allows us to extend the notation to more than 2 arguments: $S_{\text{TT}}(G_1, \ldots, G_k), P_{\text{TT}}(G_1, \ldots, G_k)$. The class of *TT series-parallel multidags* (TTSP) is the closure of the elementary TTSPs under the 2 operations $S_{\text{TT}}, P_{\text{TT}}$. By a structural induction, one can see that every TTSP is TT.

Define a *choke point* of a TT as a node other than the input and output through which every path from the input to the output must pass. The result of an $S_{\text{TT}}$ operation has a choke point but the result of a $P_{\text{TT}}$ operation does not. So the last operation used to construct a TTSP can be uniquely identified.

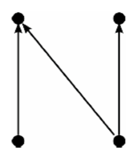
A second notion of series-parallel is the class of minimal series-parallel dags (MSP). The *elementary* MSPs are those dags with a single node. Let $G_1, G_2$ be multidags. Define the *parallel composition* $P(G_1, G_2)$ as the disjoint union of $G_1, G_2$, and the *series composition* $S(G_1, G_2)$ as the disjoint union of $G_1, G_2$ together with the edges $\text{output}(G_1) \times \text{input}(G_2)$. Note that $S$ is associative but not commutative and $P$ is associative and commutative. Associativity allows us to extend the notation to more than 2 arguments: $S(G_1, \ldots, G_k), P(G_1, \ldots, G_k)$. The class of *minimal series-parallel dags* (MSP) is the closure of the elementary MSPs under the 2 operations $S, P$. A structural induction shows that every MSP is a dag with no parallel edges.

The *underlying graph* of a digraph is obtained by replacing each directed edge by an undirected edge. A digraph is *weakly connected* iff the underlying graph is connected. The result of an $S$ operation is weakly connected and the result of a $P$ operation is not. So the last operation used to construct a MSP can be uniquely identified.

A third notion of series-parallel is the general series-parallel multidags. Let $G = (V, E)$ be a multidag. The *transitive reduction* $\mathrm{tr}(G)$ of $G$ is the unique minimal subgraph of $G$ with the same transitive closure as $G$. Existence and uniqueness are due to $G$ being finite and acyclic. By a structural induction, one can see that every MSP is transitively reduced. A multidag is *general series-parallel* (GSP) iff its transitive reduction is MSP.

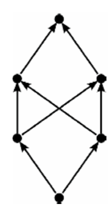
To test whether a multidag $G$ is TTSP, we can repeatedly apply the following 2 operations: replace 2 parallel edges by a single edge, replace a node $b$ with indegree and outdegree 1 and its two neighboring edges $(a, b), (b, c)$ by a single edge $(a, c)$. The result is unique and is an elementary TTSP iff $G$ is TTSP.

To test whether a dag $G$ is MSP, we use a fact from [Val78]: a multidag $G$ is GSP iff the transitive closure of $G$ does not contain the following $N$ graph as a node-induced subgraph.

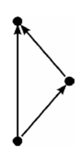


This means that it is not the case that $G$ contains 4 distinct nodes $a, b, c, d$ such that there are paths from $a$ to $b$, from $c$ to $d$, from $c$ to $b$, and between no other pairs. Then multidag $G$ is MSP iff, in addition to forbidding the $N$ graph, it is transitively reduced.

To demonstrate the incompatible nature of these definitions, note that the following dag



is MSP, TT but not TTSP; and



is TTSP but not MSP since it includes a transitive edge.

However, there is a dual relationship between 2 of the definitions. The *line graph* $L(G)$ of $G = (V, E)$ is $(E, \{(e_1, e_2) \in E \times E \mid \text{the head of } e_1 \text{ is the tail of } e_2\})$.

**Theorem 1.** *Let $G$ be* TT. *Then $G$ is* TTSP *iff $H = L(G)$ is* MSP.

*Proof.* This is lemma 1 from [VTL79]. The proof there is cursory, so we give more details.

(⇒) Follows easily by structural induction.

(⇐) We use a structural induction on $H$. If $H$ is an elementary MSP, then $G$ must be an elementary TTSP, and so is TTSP.

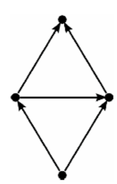
If $H = S(H_1, H_2)$, then $G$ must have a corresponding choke point $v$ and can be constructed as $S_{\mathrm{TT}}(G_1, G_2)$ where $G_1$ is the multidag inclusively between the input of $G$ and $v$ and $G_2$ is the multidag inclusively between $v$ and the output of $G$. Furthermore, $L(G_i) = H_i$, and so by the induction hypothesis, each $G_i$ is TTSP, and so $G$ is TTSP.

If $H = P(H_1, \ldots, H_l)$ where each $H_i$ is either an elementary MSP or produced by a series operation, then each $H_i$ is weakly connected. By removing the terminals of $G$ we can divide what remains into weak components and add the terminals of $G$ back to each weak component to obtain a sequence of TT multidags $G_1, \ldots, G_k$. Any edge that goes directly between the terminals of $G$ we will also represent in this list as an elementary TTSP. Note that each $G_i$ is weakly connected even when its terminals are removed.

We claim that each $L(G_i)$ is some weak component of $H$. To see this, let $e_1, e_2$ be 2 edges in $G_i$. There is a weak path $v_1, \ldots, v_k$ in $G_i$ not using any terminal that connects an endpoint of $e_1$ to an endpoint of $e_2$. Because $G_i$ is TT, there are paths from each $v_j$ to the terminals of $G_i$. Consider the undirected edges $\{v_j, v_{j+1}\}, \{v_{j+1}, v_{j+2}\}$. If the corresponding directed edges $f_1, f_2$ in $G_i$ have that the head of one is the tail of the other then the corresponding nodes in $H$ are connected. Otherwise, suppose their tails (heads) are both $v_{j+1}$. Let $x$ be the input (output) of $G_i$. There is a path in $G_i$ from $x$ to $v_{j+1} \neq x$ (from $v_{j+1} \neq x$ to $x$). So $f_1, f_2$ are weakly connected in $H$. So $e_1, e_2$ are weakly connected in $H$.

So the $G_i$ and $H_j$ are in bijective correspondence. The induction hypothesis implies that each $G_i$ is TTSP. $G = P_{\mathrm{TT}}(G_1, \ldots, G_k)$ and so is TTSP. □

It is shown in [Val78] that for each TT $G$, $G$ is TTSP iff it does not contain a subgraph homeomorphic to the $W$ graph:



2 digraphs are *homeomorphic* iff one can be obtained from the other by a sequence of operations that either replace a vertex $b$ with indegree and outdegree 1, say $(a, b), (b, c)$ are its neighboring edges, by the single edge $(a, c)$; or replace an edge $(a, c)$ by 2 edges $(a, b), (b, c)$ where $b$ is a new vertex.

A *labeling* of a multidag $G = (V, E)$ is a function $l : V \to \mathbb{N}$ such that $\forall (a, b) \in E \; l(a) < l(b)$. It is said in [Val77] that $G$ is series-parallel iff it has the property that there is some labeling $l$ such that

$$\forall (a, b), (c, d) \in E \; (l(a) - l(c))(l(d) - l(b)) \geq 0, \tag{1}$$

which we can see as a forbidden subgraph characterization. But (1) is uninteresting because of the following.

**Lemma 2.** *$G$ satisfies (1) iff $G$ is bipartite.*

*Proof.* Suppose $G$ satisfies (1) with labeling $l$. We claim there is no node $b$ with indegree $\geq 1$ and outdegree $\geq 1$. Otherwise, let $(a, b), (b, c)$ be edges. Since $l$ is a labeling, $l(a) < l(b) < l(c)$. So $(l(a)-l(b))(l(c)-l(b)) < 0$, a contradiction. So all edges go from input nodes to output nodes, and so $G$ is bipartite. Conversely, if $G$ is bipartite, label all the input nodes 0 and all the output nodes that are not also input nodes 1. This shows $G$ satisfies (1). $\square$

In a recent personal communication, Valiant fixed this by giving what was originally intended: say multidag $G = (V, E)$ is *Valiant series-parallel* (VSP) iff there is a normal labeling $l$ s.t.

$$\neg\exists(a, b), (c, d) \in E \; l(a) < l(c) < l(b) < l(d), \tag{2}$$

where a labeling $l$ is *normal* iff $\forall a \in \text{input}(G) \; l(a) = 0$ and $\exists d \in \mathbb{N} \; \forall b \in \text{output}(G) - \text{input}(G) \; l(b) = d$.

This definition will allow us to prove theorem 5.4 [Val77]. In the next section, we will show that MSP, TTSP $\subseteq$ VSP, which shows that theorem 5.4 applies to several of the definitions of series-parallel that appear in [Val78], although not to GSP.

## 2.1 Generality of VSP and GSP

**Lemma 3.** MSP $\subseteq$ VSP*.*

*Proof.* It is sufficient to show that VSP contains the elementary MSPs and is closed under $S, P$. If $G$ is an elementary MSP, then label its only node 0. So $G \in$ VSP. Let $G_1, G_2$ be VSP and for $j \in \{1, 2\}$, let $l_j$ be a normal labeling satisfying (2) for $G_j$. If one of $G_1, G_2$ is empty, then obviously $S(G_1, G_2) = P(G_1, G_2) \in$ VSP, so suppose both are nonempty. Let $d$ be the largest label $l_1$ assigns. Then $l_1 \cup (l_2 + d + 1)$ is the required normal labeling of $S(G_1, G_2)$. To show that $P(G_1, G_2) \in$ VSP will require much more effort.

We consider an alternative formulation of (2). For each edge $e = (a, b)$, let $I_e = (l(a), l(b)]$ be the half-open real interval from $l(a)$ to $l(b)$. (2) is equivalent to saying that for any 2 edges $e_1, e_2$, either $I_{e_1}, I_{e_2}$ are disjoint or one is a subset of the other. We can say 2 edges are equivalent iff they define the same interval, and then $\subseteq$ induces a partial order on the equivalence classes: $[e_1] \leq [e_2]$ iff $I_{e_1} \subseteq I_{e_2}$. We claim that the transitive reduction of $\leq$, represented as a graph $F$, is a forest. To see this, let $[e_1]$ be a node in $F$ and suppose indirectly that $[e_1]$ has 2 parents $[e_2], [e_3]$. If there were a path between $[e_2], [e_3]$ in $F$, then one of the edges $([e_1], [e_2]), ([e_1], [e_3])$ in $F$ would be a transitive edge. So $[e_2], [e_3]$ are incomparable in $F$. So the intervals $I_{e_2}, I_{e_3}$ are disjoint. But $\emptyset \neq I_{e_1} \subseteq I_{e_2} \cap I_{e_3}$, a contradiction.

To make the following discussion simpler, consider the maximal nodes of $F$ to be siblings, as well as the children of internal nodes. We can order sibling nodes of $F$: if $[e_1] = [(a, b)], [e_2] = [(c, d)]$ are siblings, then $I_{e_1}, I_{e_2}$ are disjoint, so we can say $[e_1]$ comes before $[e_2]$ iff $l(a) < l(c)$. Do not confuse this sibling ordering with the ancestor partial ordering induced by $F$.

Say $F$ is *gap free* iff for each node $[e]$ with children $[e_1], \ldots, [e_k]$, we have that $I_{e_1}, \ldots, I_{e_k}$ is a partition of $I_e$, and for the maximal nodes $[e_1], \ldots, [e_k]$, we have that $I_{e_1} \cup \cdots \cup I_{e_k}$ is an interval. If $F$ is not gap free, we can complete it by adding new nodes representing new edges in $G$ with appropriate labels.

Note that we can construct a normal labeling $l_F$ satisfying (2) from a gap free $F$ as follows: for any edge $e = (a, b)$ in the equivalence class of the $i$th leaf (starting at 1) of $F$, set $l_F(a) = i - 1, l_F(b) = i$. For an edge $e = (a, b)$ in the equivalence class of an internal node of $F$, set $l_F(a)$ ($l_F(b)$) to be the left (right) label of the leftmost (rightmost) leaf descended from $[e]$. It is not hard to see that $l_F$ is well defined.

For $j \in \{1, 2\}$, we can build a forest $F_j$ representing $G_j$. Since we are in the business of proving that $G = S(G_1, G_2) \in$ VSP and adding extra edges to $G_j$ only makes this harder, we can assume wlog that each $F_j$ is gap free.

The following algorithm will graft $F_1, F_2$ together to make a forest representing an appropriate labeling for $G$.

```
graft(F_1, F_2)
  for j ∈ {1,2}, let F_{j,1}, ..., F_{j,k_j} be the trees of F_j
  if k_1 = 0, return F_2
  if k_2 = 0, return F_1
  if k_1 ≤ k_2
    for i ← 1, ..., k_1 − 1
      H_i ← graft(F_{1,i}, F_{2,i})
    remove root r from F_{1,k_1} and let F'_{1,k_1} be the remaining forest
    H ← graft(F'_{1,k_1}, union{F_{2,k_1}, ..., F_{2,k_2}})
    add r as a root to H and call resulting tree H'
    return union{H_1, ..., H_{k_1−1}, H'}
  else
    for i ← 1, ..., k_2 − 1
      H_i ← graft(F_{1,i}, F_{2,i})
    remove root r from F_{2,k_2} and let F'_{2,k_2} be the remaining forest
    H ← graft(union{F_{1,k_2}, ..., F_{1,k_1}}, F'_{2,k_2})
    add r as a root to H and call resulting tree H'
    return union{H_1, ..., H_{k_2−1}, H'}
```

Let $F = \text{graft}(F_1, F_2)$. It can be shown by induction on the number of nodes of $F_1, F_2$ that $l_F$ is normal and induces the same order on the nodes of each $G_j$ as $l_{F_j}$ does - i.e. for nodes $a, b$ in $G_j$, $l_F(a) \leq l_F(b)$ iff $l_{F_j}(a) \leq l_{F_j}(b)$, and $l_F(a) = 0$ iff $l_{F_j}(a) = 0$, and $l_F(a) =$ the largest label $l_F$ assigns iff $l_{F_j}(a) =$ the largest label $l_{F_j}$ assigns. So $G \in$ VSP. □

**Lemma 4.** TTSP $\subseteq$ VSP.

*Proof.* The elementary TTSPs are in VSP. It is sufficient to show that if $G_1, G_2 \in$ VSP are TT, then $S_{\mathrm{TT}}(G_1, G_2), P_{\mathrm{TT}}(G_1, G_2) \in$ VSP. For $j \in \{1, 2\}$, let $l_j$ be a normal labeling of $G_j$ satisfying (2). Let $d$ be the largest label assigned by $l_1$. Then $l = l_1 \cup (l_2 + d)$ is a normal labeling satisfying (2) for $S_{\mathrm{TT}}(G_1, G_2)$. In the proof of lemma 3, we showed that VSP is closed under disjoint union. So there is a normal labeling $l$ of union$\{G_1, G_2\}$ satisfying (2). After identifying the terminals, this same labeling will suffice to show $P_{\mathrm{TT}}(G_1, G_2) \in$ VSP. □

We now show that GSP is a generalization of both MSP and TTSP, not just of MSP alone. If $G = (V, E)$ is a graph and $V' \subseteq V$, let $G - V'$ be the subgraph induced by the nodes $V - V'$.

**Lemma 5.** *Let $G$ be* MSP *and* TT *with input $x$ and output $y$. Then $H = G - \{x, y\}$ is* MSP *or empty, and $G = S(x, H, y)$.*

*Proof.* Since $G$ is TT, $G$ cannot be an elementary MSP nor the result of a parallel operation. Say $G = S(G_1, \ldots, G_k)$ where each $G_i$ is an elementary MSP or the result of a parallel operation. Clearly $x$ is in $G_1$. We claim that $G_1$ is just $x$. Otherwise, $G_1$ is the result of a parallel operation and so is not weakly connected, contradicting that in $G_1$ every node is reachable from $x$. A similar argument shows $G_k$ is just $y$. So $H = S(G_2, \ldots, G_{k-1})$ is MSP or empty. □

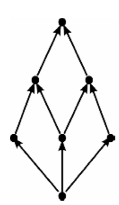
**Lemma 6.** TTSP, MSP $\subseteq$ GSP.

*Proof.* MSP $\subseteq$ GSP follows from the definitions. To show TTSP $\subseteq$ GSP, let $G$ be TTSP. We will show by a structural induction on $G$ that $\mathrm{tr}(G)$ is MSP. If $G$ is an elementary TTSP, then $\mathrm{tr}(G) = G$ is MSP.

If $G = S_{\mathrm{TT}}(G_1, G_2)$, then $\mathrm{tr}(G) = S_{\mathrm{TT}}(\mathrm{tr}(G_1), \mathrm{tr}(G_2))$. By the induction hypothesis, $\mathrm{tr}(G_1), \mathrm{tr}(G_2)$ are MSP. Let $x$ be the input of $\mathrm{tr}(G_2)$. By lemma 5, $\mathrm{tr}(G_2) - \{x\}$ is MSP. So $\mathrm{tr}(G) = S(\mathrm{tr}(G_1), \mathrm{tr}(G_2) - \{x\})$ is MSP.

Suppose $G = P_{\mathrm{TT}}(G_1, G_2)$. Let $x$ be the common input of $G_1, G_2$ and $y$ be the common output. If one of the $G_i$ has only 2 nodes, then $\mathrm{tr}(G) = \mathrm{tr}(G_{3-i})$ is MSP, by the induction hypothesis. Otherwise, $\mathrm{tr}(G) = P_{\mathrm{TT}}(\mathrm{tr}(G_1), \mathrm{tr}(G_2))$. By the induction hypothesis and lemma 5, $\mathrm{tr}(G_1) - \{x, y\}, \mathrm{tr}(G_2) - \{x, y\}$ are each MSP, and $\mathrm{tr}(G) = S(x, P(\mathrm{tr}(G_1) - \{x, y\}, \mathrm{tr}(G_2) - \{x, y\}), y)$. □

Note that a total ordering on 4 nodes is GSP but not VSP, and that



is VSP but not GSP. So VSP is in some sense a competing, different generalization of MSP and TTSP than GSP.

## 2.2 Decomposability

In the proof of the main theorem 8, given a graph $G$ from some class of graphs, we will make use of an inductive step that decomposes $G$ into 2 parts $G_1, G_2$ based on a labeling of our choice. It is critical to the proof that $G_1, G_2$ are both from the same class of graphs as $G$ is (which favors choosing a large class of graphs), and that there are not simultaneously edges from the inputs of $G_1$ to some reasonable notion of the interior of $G_2$ and from some reasonable notion of the interior of $G_1$ to the outputs of $G_2$ (which favors choosing a small class of graphs). The purpose of this section is to abstract out these requirements and make them technically precise.

Let $G = (V, E)$ be a multidag. The *neighbors* of a set of nodes $V' \subseteq V$ in $G$ is $N_G(V') = \{v \in V \mid \exists u \in V'\ (u, v) \in E \vee (v, u) \in E\}$. Given a labeling $l$, $i \in \mathbb{N}, j \in \{1, 2\}$, define

$$\begin{aligned} I^{l,i} &= \{v \in V \mid l(v) = i\} & V_j^{l,i} &= I_j^{l,i} \cup N_G(I_j^{l,i}) \\ I_1^{l,i} &= \{v \in V \mid l(v) < i\} & E_1^{l,i} &= E \cap I_1^{l,i} \times V_1^{l,i} \\ I_2^{l,i} &= \{v \in V \mid l(v) > i\} & E_2^{l,i} &= E \cap V_2^{l,i} \times I_2^{l,i} \\ & & G_j^{l,i} &= (V_j^{l,i}, E_j^{l,i}). \end{aligned}$$

Expressions like $E \cap A \times B$ are to be interpreted as those edges in $E$ with endpoints in $A \times B$ - the multiplicity of the edge comes from $E$.

A class $C$ of multidags is *decomposable* iff $\forall G \in C\ \exists$ a normal labeling $l$ of $G$ s.t. $\forall i \in \mathbb{N}, j \in \{1, 2\}$, we have

$$G_j^{l,i} \in C \text{ and} \tag{3}$$

$$\begin{aligned} &E \cap \text{input}(G) \times (I_2^{l,i} - \text{output}(G)) = \emptyset \\ \vee &E \cap (I_1^{l,i} - \text{input}(G)) \times \text{output}(G) = \emptyset, \end{aligned} \tag{4}$$

in which case we say $l$ *decomposes* $G$.

**Lemma 7.** VSP *is decomposable.*

*Proof.* Let $G = (V, E) \in$ VSP have normal labeling $l$ satisfying (2) and let $i \in \mathbb{N}$. We will show that $G_1^{l,i} \in$ VSP (that $G_2^{l,i} \in$ VSP is symmetric). Let $D = \max\{l(a) \mid a \in V\}$. Let $l_1$ be $l|V_1^{l,i}$ but with each $a \in \text{input}(G_1^{l,i})$ relabeled 0 and each $a \in \text{output}(G_1^{l,i}) - \text{input}(G_1^{l,i})$ relabeled $D$. Then $l_1$ is normal.

If $l_1$ violates (2) for $G_1^{l,i}$, then $\exists (a, b), (c, d) \in E_1^{l,i}$ s.t. $l_1(a) < l_1(c) < l_1(b) < l_1(d)$ and at least one $x \in \{a, b, c, d\}$ has $l(x) \neq l_1(x)$. $x \notin \{a, c\}$ since if $x \in \text{input}(G_1^{l,i})$, then $x \in \text{input}(G)$ and so $l(x) = 0 = l_1(x)$. It cannot be the case that $l, l_1$ disagree on both $b, d$ since then both would assign $b, d$ to $D$. So we must have that $l, l_1$ agree on $a, b, c$; and $d \in \text{output}(G_1^{l,i}) - \text{output}(G)$; and $l(b) \geq l(d)$. So $l(d) \geq i$, otherwise, $d$ would not be an output of $G_1^{l,i}$. But then $l(b) \geq l(d) \geq i$, which implies that $b \in \text{output}(G_1^{l,i}) - \text{input}(G_1^{l,i})$, which implies that $l_1(b) = D \geq l_1(d)$, a contradiction.

This shows (3). (4) is somewhat more obvious. □

A careful analysis of the proof above shows that we can actually make the definition of VSP a little looser, forbidding only edges $(a, b), (c, d)$ with $l(a)+1 < l(c) + 1 < l(b) < l(d)$.

## 3 Main theorem

Let $C$ be a class of multidags, $G \in C$. An $\text{io}_G$ path $p$ is a path in $G$ from some $a \in \text{input}(G)$ to some $b \in \text{output}(G)$. We let $|p|$ be the number of *edges* in $p$. Define

$$\begin{aligned} R_C(n,m) = \{G = (V,E) \in C \mid \forall E' \subseteq E, |E'| \leq n \\ \exists \text{io}_G \text{ path } p \text{ in } (V, E - E') \text{ s.t. } |p| \geq m\} \\ S_C(n,m) = \min\{|E| \mid \exists G = (V,E) \in R_C(n,m)\}. \end{aligned}$$

**Theorem 8.** *Let $C$ be a decomposable class of multidags. $\exists c > 0\ \forall m \geq 1, n \geq 0$, if $S_C(n,m)$ exists, then*

$$S_C(n,m) \geq cn \lg m.$$

*Proof.* For $G = (V,E) \in C$, define the *interior edges* of $G$ as

$$\text{interior}(G) = E - E \cap (\text{input}(G) \times V \cup V \times \text{output}(G)).$$

It suffices to lower bound

$$S'_C(n,m) = \min\{|\text{interior}(G)| \mid G \in R_C(n,m)\},$$

since $S_C(n,m) \geq S'_C(n,m)$. We work with $S'_C$ since it will be slightly easier to formulate a recursive lower bound for $S'_C$ than for $S_C$. We will show by induction on $m$ that $\forall m \geq 4$ where $m$ is a power of 2, if it exists, $S'_C(n,m) \geq cn \lg m$.

For $m \geq 3$, if it exists, $S'_C(n,m) \geq n + m - 2$. If we choose $c \leq \frac{1}{3}$, then for $m \in [3, 8]$, if it exists, $S'_C(n,m) \geq n + m - 2 \geq cn \lg m$. Suppose $m \geq 8$.

Let $G = (V,E) \in R_C(n,m)$, $|\text{interior}(G)| = S'_C(n,m)$, and let $l : V \to \mathbb{N}$ be a normal labeling of $G$ with maximum label $d$. Let

$$i = \min\{i \in \mathbb{N} \mid G_1^{l,i} \in R_C\left(\frac{n}{2}, \frac{m}{2}\right)\}.$$

Since $m \geq 1$, $G_1^{l,0}$ is empty and $G_1^{l,0} \notin R_C(\frac{n}{2}, \frac{m}{2})$. Also $G_1^{l,d} = G \in R_C(n,m) \subseteq R_C(\frac{n}{2}, \frac{m}{2})$. So $1 \leq i \leq d$. For $j \in \{1, 2\}$, define

$$\begin{aligned} I &= I^{l,i} & V_j &= V_j^{l,i} \\ I_j &= I_j^{l,i} & E_j &= E_j^{l,i} \\ & & G_j &= G_j^{l,i}. \end{aligned}$$

**Lemma 9.** $G_1 \in R_C(\frac{n}{2}, \frac{m}{2}) - R_C(\frac{n}{2}, \frac{m}{2} + 1)$.

*Proof.* $G_1 \in R_C(\frac{n}{2}, \frac{m}{2})$ by definition. Suppose indirectly that $G_1 \in R_C(\frac{n}{2}, \frac{m}{2} + 1)$. We claim $G_1^{l,i-1} \in R_C(\frac{n}{2}, \frac{m}{2})$. Let $E' \subseteq E_1^{l,i-1}, |E'| \leq \frac{n}{2}$. Then $\exists$ $\text{io}_{G_1}$ path $p$ in $(V_1, E_1 - E')$ s.t. $|p| \geq \frac{m}{2} + 1$. Let $v_0, \ldots, v_k$ be the sequence of nodes in $p$. Since $m \geq 1$, $k \geq 2$ and $v_{k-2} \in I_1^{l,i-1}$. So all of $p$ except possibly the last edge is in $G_1^{l,i-1}$. Let $p'$ be $p$ restricted to $G_1^{l,i-1}$. If $p' = p$ then $p$ is an $\text{io}_{G_1^{l,i-1}}$ path. Otherwise, $l(v_{k-1}) = i - 1$ and so $v_{k-1} \in \text{output}(G_1^{l,i-1})$. So $G_1^{l,i-1} \in R(\frac{n}{2}, \frac{m}{2})$, contradicting the minimality of $i$. □

**Lemma 10.** *For $m$ even, $G_2 \in R_C(\frac{n}{2}, \frac{m}{2})$.*

*Proof.* By lemma 9, $\exists E' \subseteq E_1, |E'| \leq \frac{n}{2}$ s.t.

$$\neg\exists \text{io}_{G_1} \text{ path } p \text{ in } (V_1, E_1 - E') \text{ s.t. } |p| \geq \frac{m}{2} + 1. \tag{5}$$

Let $E'' \subseteq E_2, |E''| \leq \frac{n}{2}$. We want to show that $\exists$ $\text{io}_{G_2}$ path of size $\geq \frac{m}{2}$ in $(V_2, E_2 - E'')$. $\exists$ $\text{io}_G$ path $p$ in $(V, E - (E' \cup E''))$ with node sequence $v_0, \ldots, v_k$ where $k \geq m$.

Let $q_1 = \max\{q \mid v_q \in V_1\}$. $i \geq 1$ implies $v_0 \in I_1$, which implies $q_1 \geq 1$. Note that $v_0, \ldots, v_{q_1}$ is a path in $G_1$. So (5) implies that $q_1 < k$. The maximality of $q_1$ implies that $v_{q_1} \notin I_1$. So $l(v_{q_1}) \geq i$ and $v_{q_1} \in \text{output}(G_1)$. So $v_0, \ldots, v_{q_1}$ is an $\text{io}_{G_1}$ path. (5) implies that $q_1 < \frac{m}{2} + 1$, which implies $q_1 \leq \frac{m}{2}$. (n.b. this is the only place where we use that $m$ is even.)

Let $q_2 = \min\{q \mid v_q \in V_2\}$. We claim that $i < d$. Otherwise, $G_1 = G \in R_C(n, m) \subseteq R_C(\frac{n}{2}, \frac{m}{2} + 1)$, contradicting lemma 9. $i < d$ implies $v_k \in I_2$, which implies $q_2 < k$. Note that $v_{q_2}, \ldots, v_k$ is a path in $G_2$. If $q_2 = 0$, then we are done, since then $p$ is an $\text{io}_{G_2}$ path. So suppose $q_2 \geq 1$. The minimality of $q_2$ implies $v_{q_2} \notin I_2$. So $l(v_{q_2}) \leq i$ and $v_{q_2} \in \text{input}(G_2)$. $l(v_{q_2}) \leq i \leq l(v_{q_1})$ implies $q_2 \leq q_1$. So $v_{q_2}, \ldots, v_k$ is an $\text{io}_{G_2}$ path of size $k - q_2 \geq k - q_1 \geq \frac{m}{2}$. □

Let

$$\begin{aligned} F = E \cap \Big( &(I_1 - \text{input}(G)) \times I \\ &\cup (I_1 - \text{input}(G)) \times (I_2 - \text{output}(G)) \\ &\cup I \times (I_2 - \text{output}(G)) \Big). \end{aligned}$$

Suppose $|F| < \frac{n}{4}$. Also suppose $E \cap \text{input}(G) \times (I_2 - \text{output}(G)) = \emptyset$.

We claim that $G_1 \in R_C(\frac{3n}{4}, m)$. To see this, let $E' \subseteq E_1, |E'| \leq \frac{3n}{4}$. $\exists$ an $\text{io}_G$ path $p$ in $(V, E - (E' \cup F))$ s.t. $|p| \geq m$. Since $m \geq 3$, $p$ does not go from input($G$) directly to $I$ directly to output($G$), it must use a node in $I_1 - \text{input}(G)$ or $I_2 - \text{output}(G)$. But there is no path in $(V, E - (E' \cup F))$ from input($G$) to $I_2 - \text{output}(G)$. So $p$ uses a node from $I_1 - \text{input}(G)$, which implies $p$ does not use any node from $I$. So the penultimate node in $p$ is in $I_1$ and we conclude that $p$ is in $G_1$. So $G_1 \in R_C(\frac{3n}{4}, m)$, contradicting lemma 9. So $E \cap \text{input}(G) \times (I_2 - \text{output}(G)) \neq \emptyset$.

By (4), we have $E \cap (I_1 - \text{input}(G)) \times \text{output}(G) = \emptyset$. A similar argument as above shows that $G_2 \in R_C(\frac{3n}{4}, m)$.

**Lemma 11.** $\forall m \geq 3, n \geq 0, k \in [0, n]$, *if it exists,*

$$S'_C(n, m) \geq S'_C(n - k, m) + k.$$

*Proof.* Let $G = (V, E) \in R_C(n, m)$. $m \geq 3$ implies that $|\text{interior}(E)| \geq n$. Remove $E' \subseteq \text{interior}(E)$ where $|E'| = k$ from $G$ to get $G' \in R_C(n - k, m)$. $\square$

Using lemma 11, we have

$$\begin{aligned} S'_C(n, m) &= |\text{interior}(G)| \\ &\geq |\text{interior}(G_1)| + |\text{interior}(G_2)| + |F| \\ &\geq S'_C\left(\frac{n}{2}, \frac{m}{2}\right) + S'_C\left(\frac{3n}{4}, m\right) \\ &\geq 2S'_C\left(\frac{n}{2}, \frac{m}{2}\right) + \frac{n}{4}. \end{aligned}$$

If $|F| \geq \frac{n}{4}$, then

$$\begin{aligned} S'_C(n, m) &= |\text{interior}(G)| \\ &\geq |\text{interior}(G_1)| + |\text{interior}(G_2)| + |F| \\ &\geq 2S'_C\left(\frac{n}{2}, \frac{m}{2}\right) + \frac{n}{4}. \end{aligned}$$

Either way, $S'_C(n, m) \geq 2S'_C(\frac{n}{2}, \frac{m}{2}) + \frac{n}{4}$. Solving this recurrence gives $S_C(n, m) \geq S'_C(n, m) \geq cn \lg m$ for some $c > 0$. If $m$ is not a power of 2, let $m'$ be the largest power of 2 smaller than $m$. Then $S_C(n, m) \geq S_C(n, m') \geq cn \lg m' \geq \frac{1}{2} cn \lg m$. $\square$

# 4 Thanks

This note is a result of discussions with Mohan Paturi and Leslie Valiant, whom I thank.

# References

[Val77] L. Valiant. Graph-theoretic arguments in low level complexity. *Proc. 6th MFCS*, Springer-Verlag LNCS 53 (1977)

[Val78] J. Valdes. Parsing flowcharts and series-parallel graphs. Stanford University (1978)

[VTL79] J. Valdes, R. Tarjan, E. Lawler. The recognition of series parallel digraphs. *Proc. 11th STOC*, p 1-12 (1979)