Abstract

This paper studies algorithms for computing a Gomory-Hu tree, which is a classical data structure that compactly stores all minimum s-t cuts of an undirected weighted graph. We consider two classes of algorithms: the original method by Gomory and Hu and the method based on “OrderedCuts” that we recently proposed. We describe practical implementations of these methods, and compare them experimentally with the algorithms from the previous experimental studies by Goldberg and Tsioutsiouliklis (2001) and by Akibo et al. (2016) (designed for unweighted simple graphs). Results indicate that the method based on OrderedCuts is the most robust, and often outperforms other implementations by a large factor.

1 Introduction

We study the problem of computing the Gomory-Hu tree [18] (aka cut tree) in an undirected weighted graph $G = (V, E, w)$ with non-negative weights. This is a data structure that allows to efficiently compute the cost of a minimum s-t cut (denoted as $f_G(s,t)$) for any pair of nodes $s, t$. Formally, the Gomory-Hu tree can be defined as a weighted spanning tree $T$ with the property that $f_T(s,t) = f_G(s,t)$ for all pairs $s, t$. It has numerous applications in various domains (see e.g. Section 1.4 in [5]).

Gomory and Hu [18] showed how to solve the problem using $n-1$ maximum flow computations and graph contractions. (As usual, we denote $n = |V|$ and $m = |E|$). An alternative simple algorithm that does not involve contractions has been proposed by Gusfield [19]; it performs all $n-1$ maxflow computations on the original graph. This bound has been improved upon only very recently: in a breakthrough result Abboud et al. showed in [2] how to solve the problem w.h.p. in time $\tilde{O}(n^2)$. Better bounds are known for unweighted simple graphs [9, 5, 4, 24, 6] and for the approximate version of the problem [3, 23, 22].

This paper focuses on a computational study of algorithms for computing a Gomory-Hu tree. The first such study by Goldberg and Tsioutsiouliklis appeared in [15]. It relied on the preflow maximum flow algorithm [14] for which efficient implementations had been developed [11]. A more recent study by Akibo et al. [7] focused on real sparse graphs such as social networks and web graphs. It described an implementation that could handle graphs with billions of nodes. This implementation, however, is restricted to unweighted simple graphs. Its features include a tailored bidirectional Dinitz algorithm and graph reduction techniques such as tree packing (which works only for unweighted graphs), contracting degree-2 nodes and elimination of bridges.

Our contributions We recently proposed in [20] a new approach for computing the cut tree based on a reduction to the procedure that we called OrderedCuts. Specifically, we showed in [20] that (i) the problem can be solved via $O(\log^4 n)$ expected calls to OrderedCuts on graphs of size $(n, m)$, and (ii) proved results that allow divide-and-conquer implementations of OrderedCuts. In this paper we describe a practical implementation of this approach (that we term the “OC approach”),
and compare it with several implementations of the classical algorithm of Gomory and Hu [18] ("GH approach"), namely with the codes from [15], [7], and with our own implementation.

Our reduction to the OrderedCuts problem does not follow exactly the scheme in [20], and does not have the guarantee of $O(\log^4 n)$ expected calls to OrderedCuts. Nonetheless, experiments suggest that our choice is very effective: on all of our test instances the number of calls to OrderedCuts was a small constant $c < 10$ (in fact, $c < 4$ for most of the families). We also propose a concrete implementation of the OrderedCuts procedure. To evaluate its effectiveness, we measured the total size of graphs on which the maxflow problem is solved (which we denote as $\text{size}(MF)$). On all of our test instances (except for one family of graphs, namely cycle graphs) $\text{size}(MF)$ was significantly smaller in the OC approach compared to the GH approach.

Note that in the GH approach consecutive maxflow problems are usually similar to each other, which is not the case in the OC approach. Thus, the large number of calls to the maxflow algorithm could potentially be compensated by using warm-starts in the GH approach. We put an effort in trying to make these warm-starts efficient. First, we modified the two maxflow algorithms that we used, namely Boykov-Kolmogorov (BK) [10] and the Incremental Breadth-First Search (IBFS) [17, 16]. Both of them support warm-starts, but we had to add the functionality of contracting a subset of nodes. We also tried to avoid traversing the entire graph when the minimum $s$-$t$ cut is very non-balanced (which is often the case in practice). Alternative techniques for exploiting graph similarities were described in [7] (tailored to their bidirectional Dinitz algorithm).

With these tricks some implementations of the GH approach were sometimes on par or slightly faster than the OC approach, despite having to solve much larger maxflow problems. However, for the majority of our test instances the OC approach was a clear winner in terms of the runtime.

Our implementations are available at [http://pub.ist.ac.at/~vnk/GH.html](http://pub.ist.ac.at/~vnk/GH.html).

The rest of the paper is organized as follows. Section 2 gives background on the GH and OC approaches. Section 3 describes details of our implementation. Experimental results are presented in Section 4.

## 2 Background and notation

Consider an undirected weighted graph $G = (V_G, E_G, w_G)$. A cut of $G$ is a set $U$ with $\emptyset \subsetneq U \subsetneq V_G$. It is an $S$-$T$ cut for disjoint subsets $S, T$ of $V_G$ if $T \subseteq U \subseteq V_G - S$. The cost of $U$ is defined as

$$\text{cost}_G(U) = \sum_{uv \in E_G : |\{u, v\} \cap U| = 1} w_G(uv).$$

The cost of a minimum $S$-$T$ cut in $G$ is denoted as $f_G(S, T)$. If one of the sets $S, T$ is singleton, e.g. $T = \{t\}$, then we say "$S$-$t$ cut" and write $f_G(S, t)$ for brevity.

When graph $G$ is clear from the context we omit subscript $G$ and write $V, E, \text{cost}(U), f(S, T)$.

### 2.1 Gomory-Hu algorithm

In this section we review the classical Gomory-Hu algorithm [18] for computing the cut tree. It works with a partition tree for $G$ which is a spanning tree $T = (V_T, E_T)$ such that $V_T$ is a partition of $V$. An element $X \in V_T$ is called a supernode of $T$. Each edge $XY \in E_T$ defines a cut in $G$ in a natural way; it will be denoted as $C_{XY}$ where we assume that $Y \subseteq C_{XY}$. We view $T$ as a weighted tree where the weight of $XY$ (equivalently, $f_T(X, Y)$) equals $\text{cost}_G(C_{XY})$. Tree $T$ is complete if all supernodes are singleton subsets of the form $\{v\}$; such $T$ can be identified with a spanning tree on $V$ in a natural way.

We use the following notation for a graph $G$, partition tree $T$ on $V$ and supernode $X \in V_T$:

- $H = G[T, X]$ is the auxiliary graph obtained from $G$ as follows: (i) initialize $H := G$, let $F$ be the forest on $V_T - \{X\}$ obtained from tree $T$ by removing node $X$; (ii) for each edge $XY \in E_T$ find the connected component $C_Y$ of $F$ containing $Y$, and then modify $H$ by contracting nodes in $\bigcup_{C \in C_Y} C$ to a single node called $v_Y$. Note that $V_H = X \cup \{v_Y : XY \in E_T\}$.
Algorithm 1: Gomory-Hu algorithm for graph $G$.

1. set $T = (\{V\}, \emptyset)$
2. while exists $X \in V_T$ with $|X| \geq 2$ do
3.  pick supernode $X \in V_T$ with $|X| \geq 2$ and distinct $s, t \in X$
4.  form auxiliary graph $H = G[T, X]$
5.  compute minimum $s$-$t$ cut $S$ in $H$
6.  define $(A, B) = (X - S, X \cap S)$, update $V_T := (V_T - \{X\}) \cup \{A, B\}$ and $E_T := E_T \cup \{AB\}$
7.  for each $XY \in E_T$ update $E_T := (E_T - \{XY\}) \cup \{CY\}$ where $C = \begin{cases} A & \text{if } v_Y \notin S \\ B & \text{if } v_Y \in S \end{cases}$

Note that at every step Algorithm 1 splits some supernode $X$ into two smaller supernodes, $A$ and $B$. Abboud et al. proposed in [3] a generalized version in which one iteration may split $X$ into more than two supernodes.

Algorithm 2: Generalized Gomory-Hu algorithm for graph $G$.

1. set $T = (\{V\}, \emptyset)$
2. while exists $X \in V_T$ with $|X| \geq 2$ do
3.  pick supernode $X \in V_T$ with $|X| \geq 2$, form auxiliary graph $H = G[T, X]$
4.  find node $s \in X$ and non-empty laminar family $\Pi$ of subsets of $V_H - \{s\}$ such that each $S \in \Pi$ is a minimum $s$-$t$ cut in $H$ for some $t \in X$
5.  while $\Pi \neq \emptyset$ do
6.    pick minimal set $S \in \Pi$
7.    update $T$ as in lines 6-7 of Alg. 1 for given $S$ (with $(A, B) = (X - S, X \cap S)$)
8.    update $X := A$, update $H$ accordingly to restore $H = G[T, X]$ /* now $v_B \in V_H$ */
9.    remove $S$ from $\Pi$, for each $S' \in \Pi$ with $S \subseteq S'$ replace $S'$ with $(S' - S) \cup v_B$ in $\Pi$

Note that graph $H$ at line 8 is updated by contracting set $S$ to a single vertex named $v_B$.

The main computational bottleneck in Algorithm 2 is line 4. The next section describes one particular way for implementing this line.

2.2 Computing Gomory-Hu tree via OrderedCuts

Consider undirected weighted graph $G = (V, E, w)$. We will use letters $\varphi, \alpha, \beta, \ldots$ to denote sequences of distinct nodes in $V$. For a sequence $\varphi$ and a set $X \subseteq V$ we denote $\varphi \cap X$ to be the subsequence of $\varphi$ containing only nodes in $X$. With some abuse of notation we will often view sequence $\varphi = v_0 \ldots v_\ell$ as a set $\{v_0, \ldots, v_\ell\}$; such places should always be clear from the context.

The OrderedCuts problem for graph $G = (V, E, w)$ and sequence $\varphi = sv_1 \ldots v_\ell$ of distinct nodes in $V$ can be formulated as follows: compute minimum $s, v_1, \ldots, v_k \}$ cut $S_{v_k}$ for each $k \in [\ell]$. (This definition is slightly different from the one given in [20], but suffices for our purposes). It was shown in [20] that all cuts $s, v_1, \ldots, S_{v_\ell}$ can be stored compactly in $O(|V|)$ space using a data structure called an $\mathcal{OC}$ tree.

Definition 1. Consider undirected weighted graph $G = (V, E, w)$ and sequence $\varphi = sv_1 \ldots v_\ell$. An $\mathcal{OC}$ tree for $(\varphi, G)$ is a pair $(\Omega, \mathcal{E})$ satisfying the following.

- $(\varphi, \mathcal{E})$ is a rooted tree with the root $s$ and edges oriented towards the root such that for any $v_i v_j \in \mathcal{E}$ we have $i > j$. We write $u \preceq v$ for nodes $u, v \in \varphi$ if this tree has a directed path from $u$ to $v$.

- $\Omega$ is a partition of some set $V$ with $|\Omega| = |\varphi|$ such that $|A \cap \varphi| = 1$ for all $A \in \Omega$. For node $v \in V$ we denote $[v]$ to be the unique component in $\Omega$ containing $v$ (then $[u] \neq [v]$ for distinct $u, v \in \varphi$). We also denote $[v] \dagger = \bigcup_{u \in \varphi : u \preceq v} [u]$ for $v \in \varphi$. 

Some of the cuts stored in an OC tree are minimum s-t cuts for some $t \in \varphi$; for example, by definition $[v_1]^{\dag}$ is a minimum s-$v_1$ cut. \cite{20} gave a sufficient criterion for extracting minimum s-$t$ cuts from OC tree. To formulate it, we need some notation. Let $(\Omega, \mathcal{E})$ be an OC tree for sequence $\varphi = v_0v_1\ldots v_\ell$ and graph $G$, and consider index $k \in [\ell]$. Let $p$ be the parent of $v_k$ in $(\varphi, \mathcal{E})$. Define $\pi(v_k) = v_i$ where $i$ is the largest index in $[0, k - 1]$ such that $v_i \in \{p\} \cup \{w : wp \in \mathcal{E}\}$. Clearly, by following pointers $\pi(\cdot)$ we will eventually arrive at the root node $s$ (visiting $p$ on the way). Let $\pi^\ast(v_k) = s \ldots p \ldots$ be the reversed sequence of nodes traversed during this process.

**Lemma 2** (\cite{20}). Let $(\Omega, \mathcal{E})$ be an OC tree for $(\varphi, G)$, and $u$ be a non-root node in $\varphi$. If $\text{cost}([u]^{\dag}) \geq \text{cost}([w]^{\dag})$ for all $w \in \pi^\ast(u) - \{s\}$ then $[u]^{\dag}$ is a minimum s-$u$ cut.

The lemma suggests the following approach for implementing line 4 of Alg. 2: select subset $Y \subseteq X - \{s\}$, select an ordering $v_1 \ldots v_\ell$ of nodes in $Y$, compute OC tree $(\Omega, \mathcal{E})$ for sequence $sv_1\ldots v_\ell$, and then use Lemma 2 to extract minimum s-$t$ cuts from $(\Omega, \mathcal{E})$; clearly, these cuts will form a laminar family. Ideally, nodes $v \in Y$ should be ordered accordingly to $f(s, v)$; if, for example, all values $f(s, v)$ are unique then the criterion in Lemma 2 would produce minimum s-$t$ cuts for all $t \in Y$. Since values $f(s, v)$ are unknown, \cite{20} uses instead upper bounds $\mu(v)$ on $f(s, u)$ that are computed from previously found cuts. We refer to Section 3.2 for further details of our implementation and its comparison with the scheme in \cite{20}.

## 3 Our implementation

We have implemented both the classical Gomory-Hu algorithm (Alg. 1) and its generalized version (Alg. 2) based on the OrderedCuts procedure. We refer to these two methods as GH and OC algorithms, respectively. We describe details of the GH method in Section 3.1 and details of the OC method in Sections 3.2 and 3.3. Section 3.4 describes maxflow algorithms that we used.

Below we denote $\text{size}(G) = (n, m)$ to be the size of the original graph, $\text{size}(H)$ to be the total size of graphs $H$ that appear in the algorithms, and $\text{size}(MF)$ to be the total size of graphs on which the maxflow problem is solved (excluding terminals and their incident edges). More precisely, for the GH method we define $\text{size}(H) = \text{size}(MF)$, and for the OC method we let $\text{size}(H)$ be the total size of graphs on which the OrderedCuts procedure is called. Note that in the latter case we have $\text{size}(H) \leq \text{size}(MF)$ in general.

### 3.1 GH algorithm

The main question in implementing Algorithm 1 is how to select nodes $s, t$ at line 3. There are several guiding principles proposed in the literature. \cite{15} proposes two heuristics that try to find a balanced cut: (i) choose two heaviest nodes (i.e. nodes with the highest total capacity of incident edges); (ii) after choosing $t$, choose $s$ that is furthest away from $t$ (with respect to unit edge lengths). \cite{7} uses a somewhat opposite strategy: pick $s, t$ which are adjacent in $G$, if exist (otherwise pick an arbitrary pair). They argue that such choice leads to a smaller time for computing maxflow.

We implemented two heuristics.

1. Choose $s, t$ as two heaviest nodes.

2. In the beginning compute $s, t$ as two heaviest nodes, and for each node $i$ compute distances $D_s(i), D_t(i)$ from these nodes (with respect to the unit edge lengths). After computing minimum s-$t$ cut supernode $X$ is split into supernodes $A, B$. Assume that $|A| \geq |B|$ (if not, swap $A$ and $B$). One of the terminals in $A$ is preserved and the other one gets contracted. Suppose that $s$ is preserved (the other case is symmetric). We do breadth-first search from
(contracted) \(t\), and for each layer check whether it has nodes eligible to become new sink \(t\). If yes, then we choose a node with the the smallest \(D_i(i)\). Thus, this strategy attempts to find new sink \(t\) which is close both to \(s\) and to the old sink \(t\). For supernode \(B\) we compute \(s,t\) from scratch as two heaviest nodes.

The first heuristic is aimed at producing more balanced cuts, while the second one tries to ensure that consecutive maxflow graphs are similar, so maxflow computations can be warm-started. Section 3.4 discusses warm-starts in more detail. We refer to the resulting algorithms as \(GH_h\) and \(GH_r\), respectively; these letters stand for “(h)eaviest nodes” and “(r)euse maxflow computations”.

To minimize memory usage, we use the same scheme as in [20]: if supernode \(X\) is split into two supernodes \(A\) and \(B\) then we first make a recursive call for a supernode of a smaller size, and then proceed with solving the larger supernode (using effectively tail recursion) reusing the memory which is already allocated. (A similar scheme is used for the OC implementation described in the next section.)

### 3.2 OC algorithm

Our implementation of line 4 of Alg. 2 has the following structure.

**Algorithm 3:** Line 4 of Alg. 2

1. choose node \(s \in X\) and vector \(\mu : X - \{s\} \rightarrow \mathbb{R}\) so that \(\mu(v) \geq f(s,v)\) for each \(v \in X - \{s\}\)
2. sort nodes in \(X - \{s\}\) as \(v_1, \ldots, v_\ell\) so that \(\mu(v_1) \geq \ldots \geq \mu(v_\ell), \ell = |X - \{s\}|\)
3. compute \(\text{OC}\) tree \((\Omega, E)\) for sequence \(sv_1 \ldots v_\ell\) and graph \(H\) (see Sec. 3.3)
4. use Lemma 2 to find subset \(U \subseteq X - \{s\}\) s.t. \([v]^4\) is a minimum \(s\)-\(v\) cut for each \(v \in U\)
5. let \(\Pi = \left\{[v]^4 : v \in U\right\}\), return \((s, \Pi)\)

We need to specify line 1. For the initial supernode \(X = V\) we let source \(s\) to be a node that maximizes \(\text{cost}([s])\), and set \(\mu(v) = \text{cost}([v])\) for all \(v \in X - \{s\}\). Now consider lines 6-9 of Alg. 2 for set \(S = [v]^4\). Recall that it splits \(X\) into supernodes \(A\) and \(B\) and then replaces \(X := A\). We let \(v\) to be the source for supernode \(B\), and set upper bounds \(\mu\) for \(B\) based on the cuts stored in the \(\text{OC}\) tree \((\Omega, E)\): \(\mu(u) = \min_{w : u \leq w \prec v} \text{cost}([w]^4)\). When \(\Pi\) becomes empty, we update upper bounds \(\mu\) for nodes \(u \in X - \{s\}\) in a similar way (except that we also take into account current upper bounds): \(\mu(u) := \min\{\mu(u), \min_{w : u \leq w \prec s} \text{cost}([w]^4)\}\). Note that the source node \(s\) is not changed for \(X\).

It may easily happen that some nodes \(v \in X - \{s\}\) have identical upper bounds \(\mu(v)\). To break the ties, we use an additional rule when sorting nodes at line 2: if \(\mu(u) = \mu(v)\) then we choose the same order for \(u, v\) that was used in the previous iteration.

We remark that the scheme described above differs from the one in [20]. Most importantly, we compute an \(\text{OC}\) tree for all nodes in \(X - \{s\}\) (sorted according to \(\mu\)), while [20] uses a random subset \(Y \subseteq X - \{s\}\) where each node is included with some probability \(\alpha \in \{1, 2^{-1}, 2^{-2}, \ldots, 2^{-d}\}\) for \(d = \lfloor \log_2 |X| \rfloor\). Another difference is that we choose source \(s\) deterministically, while [20] choose \(s \in X\) uniformly at random.

1[1] showed for their scheme that the expected total size of graphs \(H\) on which the \text{OrderedCuts} procedure is called is \(\text{size}(H) = O(n \log^4 n, m \log^4 n)\). This result does not apply to our scheme. We can only show a trivial bound \(\text{size}(H) = (O(n^2), O(nm))\) on the total size of graphs \(H\) (see Appendix A). Nonetheless, experimental results indicate that our approach is very effective: on all of our test instances \(\text{size}(H)\) was at most \((cn, cm)\) for a small constant \(c < 10\) (in fact, \(c < 4\) for most of the families). Based on these results, we did not investigate the approach in [20] (which we conjecture would be slower in practice).

---

1[1] There are a few other differences as well: (1) [20] performs several iterations with \text{OrderedCuts} before returning the resulting laminar family \(\Pi\); (2) [20] throws away some of the minimum \(s\)-\(t\) cuts that it finds. This allowed [20] to simplify the analysis, but there seems to be no good reason to do so in a practical implementation.
3.3 Implementation of OrderedCuts

We now consider the problem of constructing an OC tree \((\Omega, \mathcal{E})\) for a given graph \(G\) and sequence \(\varphi\). \[20\] gave several results that allow a divide-and-conquer approach. In order to state them, we need some notation. Recall that operation \([\cdot]\) in the definition of OC tree depends on \(\Omega\). To work with multiple OC trees, we use the following convention: different OC trees will be denoted with an additional symbol, e.g. \((\Omega, \mathcal{E}), (\Omega', \mathcal{E}'), \ldots\), and the corresponding operations will be denoted accordingly as \([\cdot], [\cdot]', \ldots\). We also use the following notation for graph \(G = (V, E, w)\), subset \(X \subseteq V\) and node \(s \in X\): \(G[X \cup \{s\}]\) is the graph obtained from \(G\) by contracting set \((V - X) \cup \{s\}\) to \(s\). Note that the set of nodes of \(G[X \cup \{s\}]\) equals \(X\).

**Lemma 3** \([20]\). Consider sequence \(\varphi = sv \ldots\) in graph \(G\). Let \((S, T)\) be a minimum \(s\)-\(v\) cut in \(G\). Suppose that \((\Omega', \mathcal{E}')\) is an OC tree for \((\varphi \cap S, G[S; s])\) and \((\Omega'', \mathcal{E}'')\) is an OC tree for \((\varphi \cap T, G[T; v])\). Then \((\Omega, \mathcal{E})\) is an OC tree for \((\varphi, G)\) where \(\Omega = \Omega' \cup \Omega''\) and \(\mathcal{E} = \mathcal{E}' \cup \mathcal{E}'' \cup \{sv\}\).

**Lemma 4** \([20]\). Consider sequence \(\varphi = s\alpha \ldots\) in graph \(G\) with \(|\alpha| \geq 1\). Let \((S, T)\) be a minimum \(s\)-\(\alpha\) cut in \(G\), and let \(T_s = T \cup \{s\}\). Suppose that \((\Omega', \mathcal{E}')\) is an OC tree for \((\varphi \cap S, G[S; s])\) and \((\Omega'', \mathcal{E}'')\) is an OC tree for \((\varphi \cap T_s, G[T_s; s])\). Then \((\Omega, \mathcal{E})\) is an OC tree for \((\varphi, G)\) where \(\Omega = \{[s]' \cup [s]'\} \cup (\Omega' - \{[s]'\}) \cup (\Omega'' - \{[s]'\})\) and \(\mathcal{E} = \mathcal{E}' \cup \mathcal{E}''\).

The lemmas imply that the following recursive algorithm computes a correct OC tree for a given pair \((\varphi, G)\) (if it terminates).

**Algorithm 4: OrderedCuts** \((\varphi; G)\) with \(\varphi = sv \ldots v_\ell\) and \(G = (V, E, w)\).

1. if \(\ell = 0\) then return OC tree \((\{V\}, \varnothing)\)
2. choose \(k \in [1, \ell]\), compute minimum \(s\)-\(\{v_1, \ldots, v_k\}\) cut \((S, T)\) in \(G\)
3. call \((\Omega', \mathcal{E}') \leftarrow \text{OrderedCuts}(\varphi \cap S, G[S; s])\)
4. if \(k = 1\) then
   5. call \((\Omega'', \mathcal{E}'') \leftarrow \text{OrderedCuts}(\varphi \cap T, G[T; v_1])\)
   6. return \((\Omega' \cup \Omega'', \mathcal{E}' \cup \mathcal{E}'' \cup \{sv\})\)
5. else
   8. call \((\Omega', \mathcal{E}') \leftarrow \text{OrderedCuts}(\varphi \cap T_s, G[T_s; s])\) where \(T_s = T \cup \{s\}\)
9. return \((\{[s]' \cup [s]'\} \cup (\Omega' - \{[s]'\}) \cup (\Omega'' - \{[s]'\}) \cup \mathcal{E}' \cup \mathcal{E}'')\)

Next, we discuss how to choose index \(k\) at line 2. Ideally, we would like to choose \(k\) so that the cut \((S, T)\) is balanced, i.e. \(|S| \approx |T|\). Note that \((S, T)\) depends monotonically on \(k\): as \(k\) grows, set \(S\) shrinks and set \(T\) expands (see e.g. \[13\]). Thus, one possibility would be to do a binary search on \(k\) to find a value that maximizes \(\min\{|S|, |T|\}\). This would mean throwing away intermediate cuts, and keeping only the last one. We used an alternative technique where we use every cut (even if it is unbalanced), but adjust the value of \(k\) for recursive calls. Specifically, we pass integer parameter \(k\) to every call of \text{OrderedCuts}(). For the initial call we set \(k = 1\). At line 2 we set \(k = \max\{1, \min\{[\ell/2], k\}\}\). Then after computing cut \((S, T)\) we update \(k = \text{round}(k2^{1-2|T|/(|V|-1)})\), and pass this value to recursive calls at lines 3, 5, 8. Thus, \(k \in ([k/2], 2k]\), and if \(S = \{s\}\) then \(k = [k/2]\).

We can only prove a rather weak bound \(\text{size}(MF) = O(n^2 \log n), O(nm \log n)\) on the total size of the graphs on which the maxflow problem is called during the execution of \text{OrderedCuts}(\varphi; G)\) for graph \(G\) with \(\text{size}(G) = (n, m)\) (see Appendix A). This complexity could potentially be realized if all cuts \((S, T)\) are unbalanced. However, results in Section 4 suggest that in practice the method performs much better than this bound.

**Remark 1.** We have also implemented a different procedure: (i) compute \(s\)-\(\{v_1, \ldots, v_\ell\}\) for all \(i \in [\ell]\); the result is a nested family of cuts \(T_1 \subset T_2 \subset \ldots \subset T_k\) where \(T_1\) is a minimum \(s\)-\(v_1\) cut and \(T_k\) is a minimum \(s\)-\(\{v_1, \ldots, v_\ell\}\) cut; (ii) solve recursively problems on subsets \(T_1 - \{v_1\}, \ldots, T_k - \{v_\ell\}\).
\( T_2 - T_1, \ldots, T_k - T_{k-1} \) (contracting other nodes to a single vertex); (iii) merge the results. Note that the problem (i) is an instance of the parametric maxflow problem (cf. [13]); to solve it, we used a natural divide-and-conquer strategy with a binary search (except that the initial index \( i \) was taken as \( \min\{2^d, \lceil \ell/2 \rceil \} \) where \( d \) is depth of the recursion).

The procedure described earlier performed better than this parametric scheme, so we settled for the former.

### 3.4 Maxflow algorithms

We incorporated two maxflow algorithms in our implementations: BK [10] and IBFS [16]. Importantly, these algorithms support dynamic graph updates (for the BK algorithm such updates were described in [8]). This is essential for the GH algorithm (and is not used for the OC algorithm). Recall that the GH algorithm calls maxflow algorithm for different \( s-t \) pairs on graphs that are often similar. The functionality that we need is to change the capacity of arcs from the source \( c_{si} \) and to the sink \( c_{ti} \). (In fact, both implementations do not store these arcs implicitly, but instead maintain the difference \( c_{si} - c_{ti} \), which is the “flow excess” at node \( i \)). We added the following operation to BK and IBFS codes: contract a given list of nodes (given in an array) to a single node. Importantly, this is done “locally” by traversing only nodes in the given list and their incident arcs (and without traversing the entire graph).

As observed in [16], the data structure used for storing the graph may have a significant impact on the performance. The original BK implementation used the “adjacency list” (AL) representation, while the IBFS code used the “forward star” (FS) representation. [16] changed the BK code to use FS representation and showed a significant speed-up on some instances, which it attributed to a better cache utilization (the FS representation stores incident outgoing arcs of nodes in a consecutive array).

We modified BK and IBFS codes to use a common base graph class, which can be either AL or FS. Note that with the FS representation we may need to allocate new memory for arcs during graph contraction, while in AL we can simply replace previously allocated arcs. Consistent with [16], we found that FS performed better, sometimes significantly. We report results for both AL and FS implementations to facilitate comparison with [15], which used the AL representation.

On some classes of problems most of the \( s-t \) cuts in GH algorithm are very unbalanced (e.g. containing a single node). It is important that in such cases we do not traverse the entire graph during dynamic graph updates. In particular, a minimum \( s-t \) cut should ideally be computed without traversing all nodes. We implemented the following approach: we maintain the lists of nodes \( V^+, V^- \) with the positive excess and the negative excesses, respectively. (Each flow pushing operation updates these lists if necessary). After the maxflow computation is finished, we start two breadth-first searches (one from \( V^+ \) and one from \( V^- \)), processing the component that currently has the smaller size.

Note that the BK code maintains a single list of “active” nodes in the source and the sink search trees. Guided by the same motivation as above, we modified this to maintain separate lists. When growing search trees, we then pick an active node from the list that has the smaller size. In addition, we terminate the maxflow algorithm if either list becomes empty (while the original BK implementation stops when no active nodes are left).

### 4 Experimental results

We used a single 8-core machine with the processor Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz and 16Gb RAM. All codes were written in C / C++. For each instance we reordered nodes and

---

\(^2\)We believe that IBFS code already maintains buckets with lists \( V^+ \) and \( V^- \), but we have not exploited them in the current version.
edges by performing a breadth-first search (in order to improve locality). Next, we describe baseline methods, test instances, and discuss the results.

4.1 Baseline methods

The study [15] reported 4 algorithms named gus, gh, ghs and ghg. The first one is an implementation of the Gusfield’s algorithm [19] that works on the original graph, while the other three perform graph contractions. gh and ghs differ in the source-sink selection strategy: the former picks an s-t pair at random while the latter chooses s which is furthest away from the current t. ghg uses the Hao-Orlin algorithm instead of the standard preflow algorithm. [15] shows that a single Hao-Orlin run can identify several valid cuts (1 or 2 in their implementation).

We also evaluated the algorithm from [7] that we term MG (which stands for “Massive Graphs” from the title of the paper). Their code supports only unweighted simple graphs. It implements several preprocessing heuristics for graph reduction, namely tree packing for identifying singleton valid cuts, removing bridge edges, and contracting degree-2 nodes. In our experiments these heuristics did not significantly affect the running time (usually by a small factor less than two), so we report results for the default version that includes all heuristics. It uses a bidirectional Dinitz algorithm with a “goal-oriented search”: it precomputes a shortest path tree from a fixed sink t and then uses this tree for multiple sources s. It also uses heuristics such as augmenting non-shortest paths with “detour edges” (see [7] for more details).

Finally, we experimented with the Gomory-Hu algorithm from the Lemon graph library [12].

4.2 Performance measures

For each family of graphs we present two tables (see e.g. Table 1). The first table contains the following information (see Section 3 for the description of size(G), size(H), size(MF)):

- column 2: size(G) = (n, m)
- column 3: GH tree diameter (or a range of diameters, if different algorithms produced different trees)
- column 4: size(H) / size(G) for the OC method. (The ratios are computed component-wise, i.e. separately for nodes and for edges). This measures the effectiveness of our reduction technique to the OrderedCuts problems.
- column 5: size(MF) / size(H) for the OC method. This can be viewed as a measure of effectiveness of our implementation of the OrderedCuts procedure.
- columns 6-8: size(MF) / size(G) for OC, GHb and GHr methods.

The second table contains runtimes in seconds. We report two numbers in the format t_{total} / t_{MF} where t_{total} is the overall runtime and t_{MF} ≤ t_{total} is the time spent in maxflow computations. (If t_{MF} is not available then we report only t_{total}). Note that there are two rows for each instance; the first and second rows correspond to implementations that use the adjacency list (AL) and the forward star (FS) representations, respectively. Consistent with the results in [16], the FS representation is generally faster than AL. We report the results of both to facilitate comparison with the codes of [15] which used the AL representation.

The caption in each table (except for Table 1) also gives the time of the Lemon’s implementation for the first instance in the corresponding table.
4.3 Test instances

Weighted graphs  First, we tested the algorithms on synthetic classes of graphs used in [15] and in earlier studies (only of bigger sizes). We used the generators that come with the code in [15]. In tables 6–23 we specify the exact commands that were used to generate the instances. The generators are randomized (except for CYC1 and DBLCYC1); as in [15], we report the numbers averaged over 5 runs with different random seeds.

Second, we tested TSP instances from the TSPLIB dataset [1]. Each instance is given by a set of points and a function that allows to compute a distance between two points. To obtain a sparse instance, we took the $kn$ edges with the smallest weight for $k = 2, 4, 8$.

Results for weighted graphs are presented in tables 2 (TSP instances) and 6–23 (synthetic instances). Table 1 summarizes the last entries in tables 6–23 (In general, these entries are the hardest instances in the corresponding family, except for NOI5 and NOI6).

We do not report results of $ghs$ on TSP instances since for most of them $ghs$ crashed with a segmentation fault.

Unweighted simple graphs  We also compared the algorithms with MG on unweighted simple graphs. MG stores arcs incident to each node in a consecutive array; accordingly, we use the forward star versions of our implementations.

First, we tested the algorithms on synthetic instances by taking the last instance in each of the tables 6–23 and changing the weight of all edges to one. Results are given in Table 3. Table 4 shows the results on TSP instances from Table 2 converted to unit weights.

Finally, we took large social and web graphs from [21] used in [7] (Table 5). Some of them are directed; we converted them to undirected. We also removed bridge edges and took the largest remaining connected component (motivated by the preprocessing technique in [7]).

4.4 Discussion of results

Performance of the reduction to OrderedCuts  On all test instances the ratios $\frac{\text{size}(H)}{\text{size}(G)}$ is small constants ($< 10$ on CYC1, $< 6$ on PATH and $< 4$ on all other instances). This indicates that our reduction to OrderedCuts is very effective, despite the lack the theoretical guarantees.

Sizes of maxflow graphs  Let us now discuss the ratio $\frac{\text{size}(MF)}{\text{size}(G)}$. For most of the problems this ratio was dramatically smaller for the OC approach compared to the GH approach. The latter had $\text{size}(MF) \approx n \cdot \text{size}(G)$ for many classes of problems, indicating that most $s$-$t$ cuts computed during the algorithm were very unbalanced. However, there was one single exception: for CYC1 problems (which are cycle graphs on $n$ nodes and $n$ edges) $\text{size}(MF)$ was much larger for the OC approach.

On most instances $\text{size}(MF)$ was roughly similar for $\text{GH}_h$ and $\text{GH}_r$, except for TSP instances where $\text{size}(MF)$ for $\text{GH}_h$ was 2-3 times smaller than for $\text{GH}_r$. However, the runtimes of $\text{GH}_r$ was much smaller than of $\text{GH}_h$ on these instances. We believe that this was because consecutive graphs in $\text{GH}_h$ are more similar to each other, and so warm-starts are more effective in $\text{GH}_r$. Accordingly, we report runtimes of only $\text{GH}_r$.

Runtimes on weighted graphs  Maxflow computations of OC-IBFS and OC-BK (with the AS representation) are consistently faster than maxflow computations in the codes of [15]. On top of it, on some families the codes of [15] have a very large overhead over maxflow computations (e.g. for TSP instances), whereas the overhead in our codes is usually below 50%. Thus, our OC implementation (especially with the FS representation) strictly dominates the codes of [15].

Next, we discuss the comparison of OC-IBFS / OC-BK to $\text{GH}_r$-IBFS / $\text{GH}_r$-BK. In general, OC is either significantly faster than $\text{GH}_r$ (e.g. on NOI*, PR*, TREE families) or on par with $\text{GH}_r$ (with runtime difference by a factor less than two). The only exception is the CYC1 family (Table 7), where smaller maxflow sizes translate to much faster runtimes for $\text{GH}_r$. In some cases
the performance also significantly depends on the maxflow algorithm used (IBFS or BK). For example, on DBLCYC and TSP instances \(GH_r\)-BK is much faster than \(GH_r\)-IBFS (and on par with OC-IBFS and OC-BK), whereas on TREE and WHEEL instances \(GH_r\)-IBFS is much faster than \(GH_r\)-BK.

The implementation of Lemon does not appear to be competitive.

Overall, OC-IBFS is the most robust and most often the fastest algorithm, followed by OC-BK.

**Results on unweighted simple graphs** For these types of graphs we added the MG algorithm and excluded the algorithms from [15]. The OC codes are again the most robust: on almost all instances they either significantly outperform other algorithms or on par with other codes. The CYC1 family is again the exception: OC is slower than \(GH_r\) and MG, even though \(size(MF)\) is now smaller for OC than for \(GH_r\). Furthermore, on large social and web graphs (Table 5) MG is consistently faster than OC, though by a small constant (less than 2). Note that the MG algorithm was specifically designed for such types of graphs.

**Acknowledgements**

I thank Sagi Hed for kindly providing the IBFS code from [16].
| Instance | size(GC) | D | OC: size(OF) | OC: size(OF)? | OC: size(F) | GC: size(OF) | GC: size(F) | GB: size(OF) | GB: size(F) |
|----------|----------|---|--------------|--------------|------------|-------------|-------------|-------------|-------------|
| BIKEWHE  | 4196, 8389 | 2 | 1.0 1.0 | 41.4, 27.0 | 41.4, 27.0 | 4195, 4195 | 4195, 4195 | 4195, 4195 |
| CYC1     | 16784, 16784 | 5 | 2.0 2.0 | 27.8, 25.3 | 55.7, 50.6 | 4194, 4194 | 4194, 4194 | 4194, 4194 |
| DBCYC1   | 8392, 16784 | 6.2 | 3.6 2.6 | 12.3, 9.7 | 44.5, 25.3 | 533, 445 | 540, 453 | 543, 445 | 540, 453 |
| NOI1     | 1000, 249750 | 2 | 1.0 1.0 | 19.1, 18.1 | 162.3, 153.5 | 19, 19 | 23, 23 | 19, 19 | 23, 23 |
| NOI2     | 1000, 249750 | 3 | 2.1 1.6 | 16.3, 7.3 | 34.3, 11.1 | 504, 220 | 503, 219 | 504, 220 | 503, 219 |
| NOI3     | 1000, 499500 | 2 | 1.0 1.0 | 17.1, 8.8 | 34.3, 11.1 | 504, 220 | 503, 219 | 504, 220 | 503, 219 |
| NOI4     | 1000, 499500 | 3 | 2.1 1.5 | 16.5, 7.3 | 34.3, 11.1 | 504, 220 | 503, 219 | 504, 220 | 503, 219 |
| NOI5     | 1000, 249750 | 6.2 | 3.6 2.6 | 12.3, 9.7 | 44.5, 25.3 | 533, 445 | 540, 453 | 543, 445 | 540, 453 |
| NO6      | 1000, 249750 | 3 | 2.0 1.5 | 18.2, 8.5 | 34.3, 11.1 | 504, 220 | 503, 219 | 504, 220 | 503, 219 |
| PATH     | 2000, 21990 | 40.4 | 5.3 4.2 | 13.7, 10.0 | 73.4, 41.8 | 103, 151 | 104, 153 | 103, 151 | 104, 153 |
| PR1      | 2000, 41731 | 2 | 1.0 1.0 | 19.1, 9.6 | 19.1, 9.6 | 1999, 1999 | 1999, 1999 | 1999, 1999 | 1999, 1999 |
| PR5      | 2000, 41731 | 3 | 2.2 1.7 | 16.5, 6.7 | 34.3, 11.1 | 1002, 517 | 1003, 517 | 1002, 509 | 1003, 517 |
| PR6      | 2000, 201455 | 3 | 2.0 1.5 | 18.4, 7.4 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| PR7      | 2000, 1000790 | 3 | 2.0 1.5 | 18.4, 7.4 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| PR8      | 2000, 1000790 | 3 | 2.0 1.5 | 18.4, 7.4 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| REG1     | 1000, 128000 | 2 | 2.0 1.0 | 18.7, 10.5 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| REG2     | 1000, 128000 | 2 | 2.0 1.0 | 18.7, 10.5 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| REG3     | 1000, 128000 | 3 | 2.0 1.0 | 18.7, 10.5 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |
| WHE      | 4196, 8390 | 2 | 1.0 1.0 | 16.5, 8.6 | 34.3, 11.1 | 1002, 509 | 1003, 517 | 1002, 509 | 1003, 517 |

Table 1: Synthetic instances: summary.
Table 2: TSP instances (for average degree $k = 2, 4, 8$). (Lemon: 274).
|         | size(G) | D | OC | size(H) | OC | size(MF) | OC | size(MF) | GH | size(MF) | GH | size(MF) |
|--------|---------|---|-----|---------|-----|----------|-----|----------|-----|----------|-----|----------|
| BIKEWHE| 4196, 8389 | 2 | 1.0, 1.0 | 24.3, 11.3 | 24.3, 11.3 | 4195, 4195 | 4195, 4195 |
| CYC1   | 16784, 16784 | 2 | 1.0, 1.0 | 132.5, 130.7 | 132.5, 130.7 | 16783, 16783 | 16783, 16783 |
| DBCYC1 | 8392, 16784 | 2 | 1.0, 1.0 | 121.5, 118.6 | 121.5, 118.6 | 8391, 8391 | 8391, 8391 |
| NOI1   | 1000, 196507 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| NOI2   | 1000, 196507 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| NOI3   | 1000, 315667 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| NOI4   | 1000, 315667 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| NOI5   | 1000, 196507 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| NOI6   | 1000, 196507 | 2 | 1.0, 1.0 | 17.1, 8.8 | 17.1, 8.8 | 999, 999 | 999, 999 |
| PATH   | 2000, 21868 | 2 | 1.0, 1.0 | 19.1, 9.5 | 19.1, 9.5 | 1999, 1999 | 1999, 1999 |
| PR1    | 2000, 41731 | 2 | 1.0, 1.0 | 19.1, 9.6 | 19.1, 9.6 | 1999, 1999 | 1999, 1999 |
| PR5    | 2000, 41731 | 2 | 1.0, 1.0 | 19.1, 9.6 | 19.1, 9.6 | 1999, 1999 | 1999, 1999 |
| PR6    | 2000, 201455 | 2 | 1.0, 1.0 | 19.1, 9.7 | 19.1, 9.7 | 1999, 1999 | 1999, 1999 |
| PR7    | 2000, 1000790 | 2 | 1.0, 1.0 | 19.1, 9.8 | 19.1, 9.8 | 1999, 1999 | 1999, 1999 |
| PR8    | 2000, 1978985 | 2 | 1.0, 1.0 | 19.1, 9.9 | 19.1, 9.9 | 1999, 1999 | 1999, 1999 |
| REG1   | 1000, 112991 | 2 | 1.0, 1.0 | 17.5, 9.2 | 17.5, 9.2 | 999, 999 | 999, 999 |
| TREE   | 800, 126250 | 2 | 1.0, 1.0 | 16.4, 8.5 | 16.4, 8.5 | 799, 799 | 799, 799 |
| WHE    | 4196, 8390 | 2 | 1.0, 1.0 | 21.2, 9.6 | 21.2, 9.6 | 4195, 4195 | 4195, 4195 |

Table 3: Unweighted simple graphs: synthetic instances.
Table 4: Unweighted simple graphs: TSP instances.

| size(G) | D | DC | size(H) | GHR-BF | size(H) | GHR-BK | size(G) |
|---------|---|----|---------|--------|---------|--------|---------|
| brd14051 | 0.28 | 3.5 | 15.7 | 5.9 | 11.3 | 12.6 | 59.2 |
| usa13509 | 0.40 | 8.3 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |
| d85900 | 0.28 | 3.5 | 15.7 | 5.9 | 11.3 | 12.6 | 59.2 |
| plas85900 | 0.40 | 8.3 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |

Table 5: Unweighted simple graphs: social and web graphs.

| size(G) | D | DC | size(H) | GHR-BF | size(H) | GHR-BK | size(G) |
|---------|---|----|---------|--------|---------|--------|---------|
| facebook | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 | 85.9 |
| wiki-Vote | 0.28 | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |
| ca-CondMat | 0.28 | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |
| soc-Epinions1 | 0.28 | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |
| twitter | 0.28 | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |
| com-dblp | 0.28 | 3.5 | 11.0 | 5.9 | 11.3 | 12.6 | 59.2 |

Table 6: Unweighted bipartite graphs: TSP instances.
Table 6: BIKEWHE: bikewheelgen 1024, ..., 4196. (Lemon: 296).

| size(G) | D | OC: size(H) | OC: size(MF) | Gh: size(H) | Gh: size(MF) |
|---------|---|-------------|--------------|-------------|--------------|
| 1024, 2045 | 2 | 1.0, 1.0 | 27.3, 18.7 | 1023, 1023 | 1023, 1023 |
| 2048, 4093 | 2 | 1.0, 1.0 | 35.2, 23.5 | 2047, 2047 | 2047, 2047 |
| 4196, 8389 | 2 | 1.0, 1.0 | 41.4, 27.0 | 4195, 4195 | 4195, 4195 |

Table 7: CYC1: cyclegen 4196, ..., 16784. (Lemon: 22.5).

| size(G) | D | OC: size(H) | OC: size(MF) | Gh: size(H) | Gh: size(MF) |
|---------|---|-------------|--------------|-------------|--------------|
| 4196, 4196 | 5 | 2.0, 2.0 | 21.6, 19.0 | 1022, 1022 | 1021, 1020 |
| 4196, 8392 | 5 | 2.5, 2.5 | 24.1, 22.0 | 2096, 2096 | 2095, 2094 |
| 8392, 16784 | 5 | 2.0, 2.0 | 27.8, 25.3 | 4194, 4194 | 4194, 4193 |

Table 8: DBLCYC1: dblcyclegen 2048, ..., 8392. (Lemon: 4223).

| size(G) | D | OC: size(H) | OC: size(MF) | Gh: size(H) | Gh: size(MF) |
|---------|---|-------------|--------------|-------------|--------------|
| 400, 39900 | 2 | 1.0, 1.0 | 14.4, 7.3 | 399, 340 | 399, 343 |
| 600, 89850 | 2 | 1.0, 1.0 | 15.6, 8.0 | 599, 511 | 599, 514 |
| 800, 159800 | 2 | 1.0, 1.0 | 16.4, 8.5 | 799, 682 | 799, 686 |
| 1000, 249750 | 2 | 1.0, 1.0 | 17.1, 8.8 | 999, 853 | 999, 857 |

Table 9: NOII: noigen 400 50 1 400, ..., 1000 50 1 1000. (Lemon: 3.53).
| size(G) | D  | OC: size(H) | size(C) | OC: size(MF) | size(H) | OC: size(MF) | size(C) | GH_h: size(MF) | size(C) | GH_h: size(MF) | size(C) |
|--------|----|-------------|--------|-------------|--------|-------------|--------|--------------|--------|--------------|--------|
| 400, 39900 | 3 | 2.0, 1.5    | 14.7, 6.7 | 29.4, 10.1 |        | 204, 93    |       | 203, 91    |       |
| 600, 89850 | 3 | 2.1, 1.6    | 15.3, 7.0 | 32.1, 10.9 |        | 306, 138   |       | 303, 133   |       |
| 800, 159800 | 3 | 2.0, 1.5    | 17.0, 8.0 | 34.1, 11.9 |        | 408, 183   |       | 403, 177   |       |
| 1000, 249750 | 3 | 2.1, 1.6    | 16.3, 7.2 | 34.3, 11.1 |        | 504, 220   |       | 503, 219   |       |

Table 10: NOI2: noigen 400 50 2 400, ..., 1000 50 2 1000. (Lemon: 3.66).

| size(G) | D  | OC: size(H) | size(C) | OC: size(MF) | size(H) | OC: size(MF) | size(C) | GH_h: size(MF) | size(C) | GH_h: size(MF) | size(C) |
|--------|----|-------------|--------|-------------|--------|-------------|--------|--------------|--------|--------------|--------|
| 1000, 24975 | 2 | 1.0, 1.0    | 17.1, 8.7 | 17.1, 8.7   | 999, 982 | 999, 983   |       |
| 1000, 49950 | 2 | 1.0, 1.0    | 17.1, 8.7 | 17.1, 8.7   | 999, 965 | 999, 967   |       |
| 1000, 124875 | 2 | 1.0, 1.0   | 17.1, 8.8 | 17.1, 8.8   | 999, 919 | 999, 922   |       |
| 1000, 249750 | 2 | 1.0, 1.0   | 17.1, 8.8 | 17.1, 8.8   | 999, 953 | 999, 957   |       |
| 1000, 499500 | 2 | 1.0, 1.0   | 17.1, 8.8 | 17.1, 8.8   | 999, 797 | 999, 801   |       |

Table 11: NOI3: noigen 1000 d 1 1000, d=5,10,25,50,75,100. (Lemon: 3.88).

| size(G) | D  | OC: size(H) | size(C) | OC: size(MF) | size(H) | OC: size(MF) | size(C) | GH_h: size(MF) | size(C) | GH_h: size(MF) | size(C) |
|--------|----|-------------|--------|-------------|--------|-------------|--------|--------------|--------|--------------|--------|
| 1000, 24975 | 3 | 2.2, 1.0    | 14.9, 6.1 | 32.7, 10.0  | 502, 268 | 502, 268   |       |
| 1000, 49950 | 3 | 2.0, 1.5    | 17.4, 7.5 | 34.8, 11.4  | 503, 255 | 503, 254   |       |
| 1000, 124875 | 3 | 2.2, 1.6   | 15.0, 6.4 | 33.0, 10.2  | 503, 238 | 502, 236   |       |
| 1000, 249750 | 3 | 2.1, 1.6   | 16.3, 7.2 | 34.3, 11.1  | 504, 220 | 503, 219   |       |
| 1000, 374625 | 2 | 2.0, 1.5    | 17.0, 7.4 | 33.9, 11.1  | 505, 208 | 502, 204   |       |
| 1000, 499500 | 3 | 2.1, 1.5    | 16.5, 7.5 | 34.7, 11.6  | 508, 200 | 502, 192   |       |

Table 12: NOI4: noigen 1000 d 2 1000, d=5,10,25,50,75,100. (Lemon: 4.34).
Table 13: NOI5: noigen 1000 50 k 1000, k=1,2,5,10,30,50,100,200,400,500. (Lemon: 67.0).

Table 14: NOI6: noigen 1000 50 2 P, P=1,10,50,100,150,200,500,1000,5000. (Lemon: 66.6).
| size(G)   | D  | OC: size(G) / size(H) | OC: size(MF) / size(H) | OC: size(MF) / size(G) | GH: size(MF) / size(H) | GH: size(MF) / size(G) |
|----------|----|-----------------------|------------------------|------------------------|------------------------|------------------------|
| 2000, 21990 | 2  | 1.0, 1.0              | 19.1, 9.0              | 19.1, 9.0              | 1999, 1992             | 1999, 1992             |
| 2000, 21990 | 4  | 1.7, 1.6              | 10.5, 2.7              | 18.5, 4.3              | 503, 229               | 531, 257               |
| 2000, 21990 | 4  | 2.0, 2.0              | 11.7, 4.0              | 23.4, 8.1              | 143, 51                | 179, 90                |
| 2000, 21990 | 6.6 | 2.5, 2.4              | 13.6, 5.3              | 34.7, 12.6             | 70, 35                 | 72, 45                 |
| 2000, 21990 | 13 | 3.8, 3.0              | 14.2, 7.7              | 54.2, 22.8             | 53, 47                 | 50, 48                 |
| 2000, 21990 | 26.4 | 4.9, 3.7            | 13.7, 9.1              | 66.9, 33.7             | 69, 88                 | 70, 91                 |
| 2000, 21990 | 40.4 | 5.3, 4.2            | 13.7, 10.0             | 73.4, 41.8             | 103, 151               | 104, 153               |

Table 15: PATH: pathgen 2000 1 k 1000, k=1,4,15,50,200,800,2000. (Lemon: 6.48).

| size(G)   | D  | OC: size(H) / size(G) | OC: size(MF) / size(H) | OC: size(MF) / size(G) | GH: size(MF) / size(H) | GH: size(MF) / size(G) |
|----------|----|-----------------------|------------------------|------------------------|------------------------|------------------------|
| 400, 1987 | 3  | 1.8, 1.8              | 10.6, 6.5              | 19.1, 11.7             | 397, 398               | 397, 398               |
| 800, 7151 | 2  | 1.0, 1.0              | 16.8, 8.5              | 16.8, 8.5              | 799, 799               | 799, 799               |
| 1200, 15522 | 2  | 1.0, 1.0              | 17.6, 8.8              | 17.6, 8.8              | 1129, 1199             | 1199, 1198             |
| 1600, 27020 | 2  | 1.0, 1.0              | 18.4, 9.2              | 18.4, 9.2              | 1599, 1599             | 1599, 1599             |
| 2000, 41731 | 2  | 1.0, 1.0              | 19.1, 9.6              | 19.1, 9.6              | 1999, 1999             | 1999, 1999             |

Table 16: PR1: prgen 400 2 1, ..., 2000 2 1. (Lemon: 0.16).

| size(G)   | D  | OC: size(H) / size(G) | OC: size(MF) / size(H) | OC: size(MF) / size(G) | GH: size(MF) / size(H) | GH: size(MF) / size(G) |
|----------|----|-----------------------|------------------------|------------------------|------------------------|------------------------|
| 400, 1987 | 5  | 2.3, 2.0              | 11.4, 5.2              | 26.3, 10.5             | 187, 151               | 187, 151               |
| 800, 7151 | 3.2 | 2.4, 1.9              | 12.9, 5.7              | 31.1, 11.0             | 401, 269               | 403, 272               |
| 1200, 15522 | 3  | 2.2, 1.7              | 14.9, 6.0              | 32.8, 10.4             | 601, 369               | 601, 369               |
| 1600, 27020 | 3  | 2.1, 1.6              | 16.6, 6.6              | 34.8, 10.9             | 801, 472               | 801, 472               |
| 2000, 41731 | 3  | 2.2, 1.7              | 16.5, 6.7              | 36.2, 11.3             | 1002, 573              | 1003, 576              |

Table 17: PR5: prgen 400 2 2, ..., 2000 2 2. (Lemon: 0.21).
| size(G) | D | OC - IBFS | GC - BK | GH - IBFS | GH - BK | gus | gh | ghs | ghg |
|--------|---|----------|--------|-----------|---------|-----|----|-----|-----|
| 400    | 3 | 2.0      | 1.5    | 13.4      | 5.3     | 26.8 | 8.1| 201  | 103 |
| 800    | 3 | 2.0      | 1.5    | 15.4      | 6.2     | 30.9 | 9.4| 402  | 204 |
| 1200   | 3 | 2.0      | 1.5    | 16.6      | 6.5     | 33.2 | 9.8| 602  | 304 |
| 1600   | 3 | 2.0      | 1.5    | 17.3      | 6.7     | 34.6 | 10.1| 801  | 403 |
| 2000   | 3 | 2.0      | 1.5    | 18.6      | 7.5     | 37.2 | 11.3| 1002 | 504 |

Table 18: PR6: prgen 400 10 2, ..., 2000 10 2. (Lemon: 0.67).

| size(G) | D | OC - IBFS | GC - BK | GH - IBFS | GH - BK | gus | gh | ghs | ghg |
|--------|---|----------|--------|-----------|---------|-----|----|-----|-----|
| 400    | 3 | 2.0      | 1.5    | 14.1      | 6.3     | 28.3 | 9.5| 202  | 103 |
| 800    | 3 | 2.0      | 1.5    | 16.0      | 6.8     | 32.0 | 10.3| 402  | 203 |
| 1200   | 3 | 2.0      | 1.5    | 16.5      | 6.5     | 33.1 | 9.8| 601  | 303 |
| 1600   | 3 | 2.0      | 1.5    | 17.8      | 7.3     | 35.6 | 10.9| 802  | 403 |
| 2000   | 3 | 2.0      | 1.5    | 18.8      | 7.8     | 37.6 | 11.7| 1002 | 503 |

Table 19: PR7: prgen 400 50 2, ..., 2000 50 2. (Lemon: 4.29).

| size(G) | D | OC - IBFS | GC - BK | GH - IBFS | GH - BK | gus | gh | ghs | ghg |
|--------|---|----------|--------|-----------|---------|-----|----|-----|-----|
| 400    | 3 | 2.0      | 1.5    | 11.6      | 5.0     | 26.4 | 8.2| 201  | 114 |
| 800    | 3 | 2.0      | 1.5    | 15.1      | 6.3     | 31.7 | 10.0| 402  | 216 |
| 1200   | 3 | 2.0      | 1.5    | 17.0      | 6.9     | 34.1 | 10.6| 602  | 316 |
| 1600   | 3 | 2.0      | 1.5    | 17.4      | 6.8     | 34.8 | 10.3| 801  | 415 |
| 2000   | 3 | 2.0      | 1.5    | 18.4      | 7.4     | 36.8 | 11.2| 1002 | 517 |

Table 20: PR8: prgen 400 100 2, ..., 2000 100 2. (Lemon: 12.9).
| size(G) | D | OC: size(MF) | OC: size(MF) | OC: size(MF) | OC: size(MF) | GH: size(MF) | GH: size(MF) | GH: size(MF) |
|---------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1000, 4000 | 2 | 1.0, 1.0 | 41.0, 38.9 | 41.0, 38.9 | 999, 999 | 999, 999 | 999, 999 |
| 1000, 4000 | 2 | 1.0, 1.0 | 17.7, 10.0 | 17.7, 10.0 | 999, 999 | 999, 999 | 999, 999 |
| 1000, 16000 | 2 | 1.0, 1.0 | 18.2, 10.0 | 18.2, 10.0 | 999, 999 | 999, 999 | 999, 999 |
| 1000, 64000 | 2 | 1.0, 1.0 | 18.3, 10.1 | 18.3, 10.1 | 999, 999 | 999, 999 | 999, 999 |
| 1000, 128000 | 2 | 1.0, 1.0 | 18.7, 10.5 | 18.7, 10.5 | 999, 999 | 999, 999 | 999, 999 |

**Table 21:** REG1: regularen 1000 k -w1000, k=1,4,16,64. (Lemon: 0.82).

| size(G) | D | OC: size(MF) | OC: size(MF) | OC: size(MF) | OC: size(MF) | GH: size(MF) | GH: size(MF) | GH: size(MF) |
|---------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 800, 16000 | 2 | 1.0, 1.0 | 16.4, 8.6 | 16.4, 8.5 | 799, 685 | 799, 685 | 799, 685 |
| 800, 16000 | 4 | 1.0, 1.0 | 15.0, 4.9 | 15.0, 4.9 | 397, 203 | 463, 270 | 463, 270 |
| 800, 16000 | 4 | 1.0, 1.0 | 14.6, 4.1 | 14.6, 4.1 | 280, 111 | 357, 185 | 357, 185 |
| 800, 16000 | 4 | 1.3, 1.1 | 11.4, 3.2 | 14.7, 3.6 | 180, 55 | 185, 60 | 185, 60 |
| 800, 16000 | 4 | 1.5, 1.2 | 11.4, 3.6 | 16.9, 4.3 | 126, 36 | 128, 39 | 128, 39 |
| 800, 16000 | 6 | 1.6, 1.3 | 12.5, 4.4 | 20.2, 5.8 | 97, 36 | 101, 39 | 101, 39 |
| 800, 16000 | 6 | 1.8, 1.5 | 12.7, 4.9 | 22.6, 7.3 | 98, 46 | 102, 51 | 102, 51 |
| 800, 16000 | 7.2 | 2.2, 1.8 | 12.2, 5.8 | 27.5, 10.2 | 119, 71 | 130, 81 | 130, 81 |
| 800, 16000 | 7.8 | 2.6, 2.0 | 12.8, 6.5 | 33.0, 12.9 | 155, 108 | 163, 115 | 163, 115 |

**Table 22:** TREE: treelen 800 50 k 1000, k=1,3,5,10,20,50,100,200,400. (Lemon: 28.9).

| size(G) | D | OC: size(MF) | OC: size(MF) | OC: size(MF) | OC: size(MF) | GH: size(MF) | GH: size(MF) | GH: size(MF) |
|---------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1024, 2046 | 2 | 1.0, 1.0 | 31.5, 21.1 | 31.5, 21.1 | 1023, 1023 | 1023, 1023 | 1023, 1023 |
| 2048, 4094 | 2 | 1.0, 1.0 | 38.7, 25.5 | 38.7, 25.5 | 2047, 2047 | 2047, 2047 | 2047, 2047 |
| 4196, 8390 | 2 | 1.0, 1.0 | 46.8, 30.6 | 46.8, 30.6 | 4195, 4195 | 4195, 4195 | 4195, 4195 |

**Table 23:** WHE: wheelgen 1024, ..., 4196. (Lemon: 274).
A Complexity of the OC algorithm

In this section we give bounds on the complexity of the OC algorithm.

Theorem 5. The OC algorithm described in Section 3.2 for graph G with size(G) = (n,m) satisfies size(H) = (O(n^2),O(nm)).

Proof. Recall that lines 6-9 of Alg. 2 process sets in Π starting from the minimal sets in Π. Let us consider a modification that at line 6 picks a maximal set S ∈ Π instead of a minimal set. Processing such set will create supernodes A = X − S and B = X ∩ S. Consider family of sets Π′ = {S′ ∈ Π : S′ ⊆ S}. It can be seen that Π′ is a laminar family of sets with the property that each S′ ∈ Π′ is a minimum v-t cut where v,t are the maximal nodes w.r.t. ≤ in B and in S′ ∩ X, respectively. Let us further modify the execution as follows: when processing supernode B, we first run steps 6-9 for the laminar family Π′ defined above (if it is non-empty). Clearly, this gives an execution that is equivalent to the original execution in the following sense: it uses the same graphs H on which the OrderedCuts procedure is called, possibly in a different order. Note that in the new execution some supernodes may be split without calling OrderedCuts (using instead the laminar family Π′ provided by the parent call). We will refer to the iterations defined above as OC iterations and dummy iterations, respectively. (An iteration is one pass through the loop at lines 3-9).

By repeatedly applying such modification we obtain a valid execution of Algorithm 2 such that every family Π produced at line 4 is a set of disjoint subsets. We can now apply a standard argument. Each supernode X that appears during the execution of the modified algorithm can be assigned a depth using natural rules: (i) the initial supernode V is at depth 0; (ii) if supernode X at depth d is split into supernodes X₀,...,Xₚ then the latter supernodes are assigned depth d+1. Let X₁,...,Xᵣ(d) be the supernodes at depth d, and let H₁,...,Hₖ(d) be the corresponding auxiliary graphs. It is known that the total size of graphs H₁,...,Hₖ(d) is (O(n),O(m)) (see [3, 20]). Clearly, the maximum depth cannot exceed n (since the size of each supernode is smaller than the size of its parent). The claim follows.

Theorem 6. Denote size(G) = (n,m). The total size of graphs on which the maxflow problem is solved during the call OrderedCuts(φ;G) (excluding terminals and their incident edges) is (O(n^2 log n),O(nm log n)).

Proof. Given a pair (φ,G) with φ = s... and G = (V,E,w), we denote G to be the graph obtained from G by removing s and all incident edges. Each call to OrderedCuts(φ;G) with |φ| > 2 makes two recursive calls OrderedCuts(φ';G') (at line 3) and OrderedCuts(φ'';G'') (at lines 5 or 8). We say this call performs a trivial split (or it is a trivial call) if S = {s} and k > 1, in which case G' has a single vertex and G'' = G. If we remove all trivial calls, we obtain a sequence of computations that still fits the structure of Algorithm 1 (and each call is non-trivial). Let us analyze the size of maxflow graphs for such sequence. Clearly, we |V'| ≤ |V| − 1 and |V''| ≤ |V| − 1. Thus, the maximum depth will be at most n. A straightforward induction argument shows that graphs G at a fixed depth are disjoint. Therefore, the total size of graphs G is at most (n², nm).

Now let us consider trivial calls. A consecutive sequence of such calls may have a length of at most O(log n) (since after each such call the value of k is halved). Thus, to each trivial call we can assign a non-trivial call on the same graph such that at most O(log n) trivial calls are assigned to each non-trivial call. This proves the theorem.

References

[1] TSPLIB dataset. [http://elib.zib.de/pub/mp-testdata/tsp/tsplib/tsplib.html]
[2] Amir Abboud, Robert Krauthgamer, Jason Li, Debmalya Panigrahi, Thatchaphol Saranurak, and Ohad Trabelsi. Breaking the cubic barrier for all-pairs max-flow: Gomory-Hu tree in nearly quadratic time. In FOCS, 2022.

[3] Amir Abboud, Robert Krauthgamer, and Ohad Trabelsi. Cut-equivalent trees are optimal for min-cut queries. In FOCS, 2020. Full version: arXiv:2009.06090.

[4] Amir Abboud, Robert Krauthgamer, and Ohad Trabelsi. APMF < APSP? Gomory-Hu tree for unweighted graphs in almost-quadratic time. In FOCS, pages 1135–1146, 2021.

[5] Amir Abboud, Robert Krauthgamer, and Ohad Trabelsi. Subcubic algorithms for Gomory-Hu tree in unweighted graphs. In STOC, pages 1725–1737, 2021. Full version at arXiv:2012.10281.

[6] Amir Abboud, Robert Krauthgamer, and Ohad Trabelsi. Friendly cut sparsifiers and faster Gomory-Hu trees. In SODA, pages 3630–3649, 2022.

[7] Takuya Akiba, Yoichi Iwata, Yosuke Sameshima, Naoto Mizuno, and Yosuke Yano. Cut tree construction from massive graphs. In Proceedings of the IEEE 16th International Conference on Data Mining (ICDM), pages 775–780, 2016.

[8] Kartik Alahari, Pushmeet Kohli, and P. H. S. Torr. Dynamic hybrid algorithms for MAP inference in discrete MRFs. PAMI, 32(10):1846–1857, 2010.

[9] A. Bhalgat, R. Hariharan, T. Kavitha, and D. Panigrahi. An o(mn) Gomory-Hu tree construction algorithm for unweighted graphs. In STOC, pages 605–614, 2007.

[10] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI, 26(9), September 2004.

[11] B. V. Cherkassky and A. V. Goldberg. On implementing push-relabel method for the maximum flow problem. Algorithmica, 19:390–410, 1997.

[12] B. Dezs, A. Jüttner, and P. Kovács. Lemon - an open source C++ graph template library. Electron. Notes Theor. Comput. Sci., 264(5):23–45, 2011.

[13] G. Gallo, M. D. Grigoriadis, and R. E. Tarjan. A fast parametric maximum flow algorithm and applications. SIAM J. Computing, 18:30–55, 1989.

[14] A. V. Goldberg and R. E. Tarjan. A new approach to the maximum flow problem. J. Assoc. Comput. Mach., 35:921–940, 1988.

[15] A. V. Goldberg and K. Tsioutsiouliklis. Cut tree algorithms: an experimental study. Journal of Algorithms, 38(1):51–83, 2001. Code: https://web.archive.org/web/20140115171857/http://www.cs.princeton.edu/~kt/cut-tree/experiments/.

[16] Andrew V. Goldberg, Sagi Hed, Haim Kaplan, Pushmeet Kohli, Robert E. Tarjan, and Renato F. Werneck. Faster and more dynamic maximum flow by incremental breadth-first search. In Algorithms - ESA, pages 619–630, 2015.

[17] Andrew V. Goldberg, Sagi Hed, Haim Kaplan, Robert E. Tarjan, and Renato F. Werneck. Maximum flows by incremental breadth-first search. In Algorithms - ESA, pages 457–468, 2011.

[18] R. E. Gomory and T. C. Hu. Multi-terminal network flows. Journal of the Society for Industrial and Applied Mathematics, 9:551–570, 1961.
[19] D. Gusfield. Very simple methods for all pairs network flow analysis. *SIAM Journal on Computing*, 19(1):143–155, 1990.

[20] Vladimir Kolmogorov. OrderedCuts: A new approach for computing Gomory-Hu tree. *arXiv*, August 2022.

[21] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. [http://snap.stanford.edu/data](http://snap.stanford.edu/data), June 2014.

[22] Jason Li, Danupon Nanongkai, Debmalya Panigrahi, and Thatchaphol Saranurak. Fair cuts, approximate isolating cuts, and approximate gomory-hu trees in near-linear time. *arXiv:2203.00751*, March 2022.

[23] Jason Li and Debmalya Panigrahi. Approximate Gomory-Hu tree is faster than $n - 1$ max-flows. In *STOC*, pages 1738–1748, 2021.

[24] Jason Li, Debmalya Panigrahi, and Thatchaphol Saranurak. A nearly optimal all-pairs min-cuts algorithm in simple graphs. In *FOCS*, pages 1135–1146, 2021.