

# Amortized $\tilde{O}(|V|)$ -Delay Algorithm for Listing Chordless Cycles in Undirected Graphs

Rui Ferreira, Roberto Grossi, Romeo Rizzi, Gustavo Sacomoto, Marie-France  
Sagot

► **To cite this version:**

Rui Ferreira, Roberto Grossi, Romeo Rizzi, Gustavo Sacomoto, Marie-France Sagot. Amortized  $\tilde{O}(|V|)$  -Delay Algorithm for Listing Chordless Cycles in Undirected Graphs. 22th Annual European Symposium on Algorithms, Sep 2014, Wroclaw, Poland. 8737, pp.418-429, 2014, Algorithms ESA 2014, Lecture Notes in Computer Science. <10.1007/978-3-662-44777-2\_35>. <hal-01081031>

**HAL Id: hal-01081031**

**<https://hal.inria.fr/hal-01081031>**

Submitted on 30 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Amortized $\tilde{O}(|V|)$ -Delay Algorithm for Listing Chordless Cycles in Undirected Graphs

Rui Ferreira<sup>1</sup>, Roberto Grossi<sup>2</sup>, Romeo Rizzi<sup>3</sup>, Gustavo Sacomoto<sup>4</sup>, and Marie-France Sagot<sup>4</sup>

<sup>1</sup> Microsoft Bing, UK

<sup>2</sup> Università di Pisa, Italy

<sup>3</sup> Università di Verona, Italy

<sup>4</sup> INRIA Grenoble Rhône-Alpes, France & Université Lyon 1, France

**Abstract.** Chordless cycles are very natural structures in undirected graphs, with an important history and distinguished role in graph theory. Motivated also by previous work on the classical problem of listing cycles, we study how to list chordless cycles. The best known solution to list all the  $C$  chordless cycles contained in an undirected graph  $G = (V, E)$  takes  $O(|E|^2 + |E| \cdot C)$  time. In this paper we provide an algorithm taking  $\tilde{O}(|E| + |V| \cdot C)$  time. We also show how to obtain the same complexity for listing all the  $P$  chordless  $st$ -paths in  $G$  (where  $C$  is replaced by  $P$ ).

## 1 Introduction

A *chordless (induced) cycle*  $c$  in an undirected graph  $G$  is a cycle such that the subgraph induced by its vertices contains exactly the edges of  $c$ . A chordless cycle is called a *hole* when its length is at least 4. Similarly, a *chordless (induced) path*  $\pi$  in  $G$  is such that the subgraph of  $G$  induced by  $\pi$  contains exactly the edges of  $\pi$ . Both chordless cycles and paths are very natural structures in undirected graphs with an important history, appearing in many papers in graph theory related to chordal graphs, perfect graphs and co-graphs (e.g. [11, 6, 3]), as well as many NP-complete problems involving them (e.g. [2, 7, 9]).

In this paper we consider algorithms for listing chordless cycles and  $st$ -paths in a undirected graph  $G = (V, E)$ , with  $n = |V|$  vertices and  $m = |E|$  edges, motivated by the algorithms for listing cycles and  $st$ -paths that have been produced by an active area of research since the early 70s [10, 13, 1]. In this paper we present an algorithm for listing all the  $C$  chordless cycles in an undirected graph  $G = (V, E)$  in  $\tilde{O}(m + n \cdot C)$  time, hence with an amortized  $\tilde{O}(n)$  time delay, where  $\tilde{O}(f(n, m))$  is used as a shorthand for  $O(f(n, m) \text{ polylog } n)$ . We also show that the same algorithm may be used to list all the  $P$  chordless  $st$ -paths  $\tilde{O}(m + n \cdot P)$  time, hence amortized  $\tilde{O}(n)$  time delay.

There are very few algorithms in the literature for listing chordless cycles and/or paths, where some of them have no guaranteed performance [12, 16]. The most notable and elegant listing algorithm is by Uno [15], with a cost of  $O(m^2 + m \cdot C)$  time for chordless cycles and  $O(m^2 + m \cdot P)$  time for chordless  $st$ -paths, hence amortized  $O(m)$  time delay.

## 2 Preliminaries

Our graphs are finite, undirected, and *simple*, *i.e.* without self-loops or parallel edges. Given a graph  $G = (V, E)$  with  $n = |V|$  vertices and  $m = |E|$  edges, our task is to list out fast all its chordless cycles. We hence assume that  $G$  is connected. Where  $V' \subseteq V$ , we denote by  $E\langle V' \rangle := \{uv \in E \mid u, v \in V'\}$  the set of those edges which are contained in  $V'$ . A graph  $G' = (V', E')$  is called a *subgraph* of  $G$  if  $V' \subseteq V$  and  $E' \subseteq E$ . The subgraph  $G'$  is called *induced* (or *chordless*) if  $E' = E\langle V' \rangle$ . For any  $V' \subseteq V$ , we denote by  $G[V'] := (V', E\langle V' \rangle)$  the *subgraph of  $G$  induced by  $V'$* . Where  $e \in E$ , we denote by  $G \setminus e := (V, E \setminus \{e\})$  the subgraph obtained from  $G$  by *deleting* the edge  $e$ . Where  $v \in V$ , we denote by  $G \setminus v := G[V \setminus \{v\}]$  the subgraph obtained from  $G$  by first deleting all the edges incident to  $v$ , and then removing the isolated vertex  $v$ . Given a vertex  $u \in V$ , we denote by  $N_G(u) := \{v \in V \mid uv \in E\}$  the *neighbourhood* of  $u$ , the subscript is omitted whenever the graph is clear from the context.

A *cycle* is a connected graph in which every vertex has degree 2. A *path* is a connected graph in which every vertex has degree 2 except for two degree 1 vertices,  $s$  and  $t$ , called the *endvertices* of the path. This is also called an *st-path* and denoted by  $\pi = s \rightsquigarrow t$ . Indeed, when buiding a path from  $s$  to  $t$  edge after edge, it will be most natural, and more precise, to think like we are orienting the traversed edges. For this reason, we will also write  $(u, v)$  for an edge that, when buiding a path, has been traversed from  $u$  to  $v$ .

A (chordless) path (or cycle) of  $G$  is a (chordless) subgraph of  $G$  which is a path (or cycle). We denote by  $\mathcal{C}(G)$  the set of all chordless cycles in  $G$ . We denote by  $\mathcal{P}(G)$  (by  $\mathcal{P}_{st}(G)$ ) the set of all chordless paths (*st-paths*) in  $G$ . When  $s = t$ , we get those cycles visiting  $s$ . We refer to a path  $\pi \in \mathcal{P}(G)$  by its natural sequence of vertices or edges. An *hole* is a chordless cycle of size at least 4. Thus  $\mathcal{C}(G)$  comprises holes and triangles. Since the triangles are at most  $mn$ , our algorithm can be used to list the holes of  $G$  in  $\tilde{O}(n)$  time each, with an overall  $\tilde{O}(mn^2)$  additive time cost.

Uno [15] proposed an algorithm that lists each chordless cycle in an undirected graph  $G = (V, E)$  in  $O(m)$  time while using  $O(m)$  space. The first step is the following reduction to the enumeration of chordless *st-path* in  $G$ . Based on the fact that for any vertex  $s \in V$  the chordless cycles in  $G \setminus s$  are also chordless cycles in  $G$ , the algorithm proceeds by listing all chordless cycles passing through  $s$ ; and repeating the process in  $G \setminus s$ , until the graph is empty. Then, to list all chordless cycles passing through  $s$  in  $G' = G$ , the algorithm follows the approach of listing the chordless paths  $s \rightsquigarrow t$  in  $G' \setminus (s, t)$ , for each  $t \in N_{G'}(s)$ ; and to avoid duplications, at the end of iteration the graph is updated  $G' = G' \setminus t$ .

Given a previously computed chordless *st-path*  $\pi = v_0 v_1 \dots v_k$ , Uno's algorithm identifies the set of vertices  $U \subseteq V$  such that each  $u \in U$  is adjacent to some  $v_i \in \pi$ , and the edge  $(v_j, u)$  is contained in a chordless *st-path* extending the prefix  $\pi_j = v_0 v_1 \dots v_j$  different from  $\pi$ . The algorithm is kick-started by taking the shortest *st-path* (as the shortest path has the property of also being a chordless path) and employs a recursive strategy of vertex removal to avoid listing the same chordless path multiple times. This ensures that each chordless

path is listed once. Uno’s algorithm takes  $O(m)$  time to compute  $U$  and prepare the recursive calls before it either outputs a new path or stops. The total time is therefore  $O(m^2 + m \cdot |\mathcal{C}(G)|)$ .

### 3 Our Approach and Key Ideas

We outline the main ideas which allow us to reduce the amortized cost for a chordless path from  $O(m)$  to  $\tilde{O}(n)$ , giving a total  $\tilde{O}(m + n \cdot |\mathcal{C}(G)|)$  time to list all the chordless cycles. Our approach relies on a variant of the *cleaning* operation introduced in [6] to recognize linear balanced matrices and even holes in graphs [4, 5].

#### 3.1 Certificates for chordless $st$ -path

A listing algorithm usually takes the form of a recursive procedure exploring the space of all solutions. A key idea employed since the first listing papers [10] is to check for the existence of at least one solution before branching, *i.e.* before partitioning the solution space in subspaces to be assigned to the children. This avoids unproductive recursive calls as they do not list any solutions but their overhead cost can completely dominate the cost of reporting the solutions (e.g. see [14]). In a previous work [1], we stressed the notion of certificate since, in a more refined recursive scheme, passing a certificate of existence as an extra parameter may facilitate the work of the children which may avoid running the existence check: if they have a single child, they could be done by just passing the certificate received in input or a small adaptation of it. We also saw that more structural facts around the certificate could be useful. For the case of  $st$ -paths [1], the certificate is a DFS tree rooted in  $s$  and reaching  $t$ , which contained an  $st$ -path and also helped in other ways. Until now, the certificate was also a solution itself or contained one.

Here we try out something new: what if our certificate guarantees the existence of a solution but is *not* itself a solution? The following fact suggests that the certificate for the existence of a chordless  $st$ -path might be just any  $st$ -path.

**Fact 1** *Given two vertices  $s, t$  in  $G$ , there is a chordless  $st$ -path in  $G$  iff there is an  $st$ -path in  $G$ .*

Thus we allow for certificates which are somewhat less refined than actual solutions, in the same spirit that a binary heap demands a less strict and lazy notion of order. This is a new asset of the notion of certificate and opens up new possibilities.

#### 3.2 From chordless cycles to chordless $st$ -paths

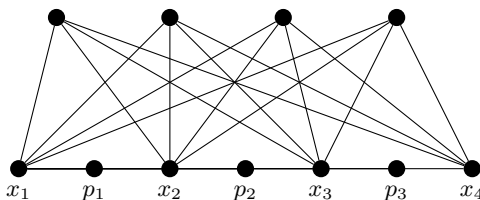
Uno [15] shows how to reduce listing chordless cycles in a graph to listing chordless  $st$ -paths for all edges  $(s, t)$  chosen in a specific order (see Section 2), which

is necessary to avoid duplications in the output. The initialization step for each edge  $(s, t)$  takes  $O(m)$  time as it requires to find one chordless  $st$ -path. This gives the  $m^2$  term in the total cost of  $O(m^2 + m \cdot |\mathcal{C}(G)|)$  for chordless cycles.

We observed in Section 3.1 that any  $st$ -path will suffice as a starter, as they are our certificates of choice. This makes a difference for the above reduction since using dynamic graph connectivity algorithms [8], it is possible to maintain a spanning tree in  $O(\text{polylog } n)$  time per edge deletion, perform connectivity queries in  $O(\text{polylog } n)$ , and more importantly obtain an  $st$ -path in  $\tilde{O}(n)$ . It is worth noting that it is not known how to obtain a chordless  $st$ -path faster than  $O(m)$ . Hence, we first build the dynamic connectivity structure as preprocessing step. Then, for each edge  $(s, t)$ , in the same order as Uno's reduction, we list the chordless  $st$ -paths. Before calling our path listing algorithm for edge  $(s, t)$ , we test if  $s$  and  $t$  are connected (Fact 1): if so, we call our path listing algorithm, paying  $\tilde{O}(n)$  to find one initial  $st$ -path; otherwise, we skip the edge  $(s, t)$  and take the next in order. As a result, the total initialization cost is  $\tilde{O}(m + kn)$  for all edges instead of  $O(m^2)$ , where  $k$  is the number of edges for which we find one initial  $st$ -path. Note that  $k \leq |\mathcal{C}(G)|$  as each of them surely gives rise to a chordless  $st$ -path and, so a distinct chordless cycle. We obtain in this way an  $\tilde{O}(m + n \cdot |\mathcal{C}(G)|)$  time algorithm to list chordless cycles, if we can list  $st$ -paths in amortized  $\tilde{O}(n)$  time each.

### 3.3 Difficulty of cleaning $st$ -paths

Given any  $st$ -path, we can clean it as stated in Fact 1 to obtain a chordless  $st$ -path in a greedy fashion: start from  $u = s$  and iteratively take a neighbour of  $u$  that is closer to  $t$  along the path. The process stops when  $u = t$ . The vertices taken in this way form a chordless path. The problem is that the cost of such a greedy traversal of the path is upper bounded by the sum of the degrees of the vertices along it. Unfortunately, this sum could be  $\Theta(m)$  in the worst case.



**Fig. 1.** Sum of degrees on chordless path  $x_1, p_1, x_2, p_2, \dots, x_{r-1}, p_{r-1}, x_r$  is  $\Theta(m)$ .

Even worse, this is still true when the initial path is already chordless, as shown in Fig. 1. Consider the complete bipartite clique  $K_{r,r} = (V_1 \cup V_2, E_{12})$ , where  $V_1 = \{x_1, x_2, \dots, x_r\}$ . Build a new graph  $G = (V, E)$  where the vertex set is  $V = V_1 \cup V_2 \cup \{p_1, \dots, p_{r-1}\}$  for some new vertices  $p_1, \dots, p_{r-1}$ , and the edge set is  $E = E_{12} \cup \{(x_1, p_1), (p_1, x_2), (x_2, p_2), \dots, (x_{r-1}, p_{r-1}), (p_{r-1}, x_r)\}$ . Now,

the path  $x_1, p_1, x_2, p_2, \dots, x_{r-1}, p_{r-1}, x_r$  is chordless but each edge is incident to at least one vertex in that path, so the sum of the degrees is  $m = |E| = \Theta(r^2) = \Theta(|V|^2) = \Theta(n^2)$ .

What we would like to do: recursively extend a given chordless path  $\pi_{su}$  into a chordless  $st$ -path, while maintaining as a certificate an  $st$ -path. The recursive extension can be seen as an implicit cleaning of our  $st$ -path certificate. Consider a vertex  $u$  along a given  $st$ -path (our certificate), where initially  $u = s$ . Our certificate guarantees that there is at least one chordless  $st$ -path going through a neighbour of  $u$ , say  $a$ . However, we cannot explore all of  $u$ 's neighbours, so consider any neighbour  $b \neq a$ , the following two situations may occur. (1)  $a$  and  $b$  are both *good*, meaning that  $(u, a)$  and  $(u, b)$  are on two distinct chordless  $st$ -paths. In this case, the chordless  $st$ -paths traversing  $(u, a)$  cannot go through  $b$  too, as otherwise it would not be chordless (see Remark 1 below), so  $b$  should be removed. (2)  $b$  is not on any chordless  $st$ -path, so it is either disconnected from  $t$  or every  $st$ -path going through  $b$  passes through  $a$ . In this case, as it will be clear later, we need neither to explore nor to remove  $b$ .

In other words, we can treat the neighbours of  $u$  as described above, and they will not interfere when cleaning the  $st$ -path in the next recursive calls. We make this statement more precise below.

### 3.4 Reduced degree property

We introduce a notion of reduced degree with a stronger property in mind. Consider a chordless  $st$ -path  $\pi_{st} = v_0 v_1 \dots v_\ell$  in the graph  $G$ , for some integer  $\ell > 1$ , where  $v_0 = s$  and  $v_\ell = t$ . For a vertex  $v_i$ , a neighbour  $v \in N(v_i)$  is *good* if there exists a chordless  $st$ -path in  $G$  with prefix  $v_0 v_1 \dots v_i v$  (*i.e.* it extends  $v_0 v_1 \dots v_i$  by adding the edge  $(v_i, v)$  as illustrated in Fig. 2). We denote by  $N^{good}(v_i) \subseteq N(v_i)$  the set of *good* neighbours of  $v_i$ , noting that  $v_{i+1} \in N^{good}(v_i)$ . For each  $v_i$ , its *reduce degree*  $d_i$  is given by the number of non-good neighbours, namely,  $d_i = |N(v_i) \setminus (N^{good}(v_i) \setminus \{v_{i+1}\})|$ .

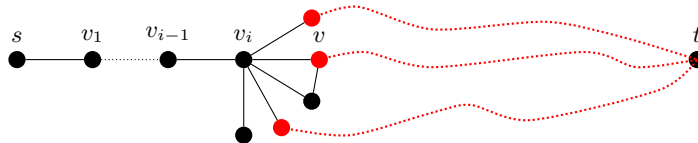


Fig. 2. Good neighbours (in red) of vertex  $v_i$  in  $G_i$ .

The rationale is that exploring the good neighbours of  $v_i$  will list further chordless paths while examining its neighbours that are not good is a waste of computation. The reduced degree of  $v_i$  is actually an upper bound on the number of not-good vertices examined when exploring  $v_i$  to produce the chordless  $st$ -path  $\pi_{st}$  and gives an upper bound on the waste. We now prove Lemma 1 below, as it implies that while examining the neighbours of the vertices along a chordless

path still takes  $O(m)$  time, only  $O(n)$  neighbours are a waste while the remaining ones lead to further chordless paths (which is a good argument for amortization).

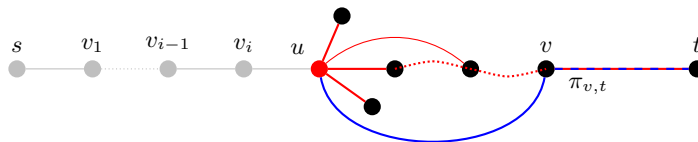
**Lemma 1.** *For a chordless path  $\pi_{st}$ , we have  $\sum_{v_i \in \pi_{st}} d_i \leq 2n$ , where  $d_i$  is the reduced degree of  $v_i \in \pi_{st}$ .*

*Proof.* Consider a vertex  $x \in G$  that is a non-good neighbour of both  $v_i$  and  $v_j$  for  $i \neq j$ . We can assume wlog that  $i < j$ . Moreover, we choose the vertices  $v_i, v_j$  and  $x$  such that the difference  $j - i$  is the largest. Let us assume by contradiction that  $v_i$  and  $v_j$  are not adjacent in  $\pi_{st}$ . We have that  $(v_i, v_j)$  is not an edge of  $G$ , and  $x \notin \pi_{st}$ , since  $\pi_{st}$  is chordless. We claim that the path  $\pi^* = v_0 \dots v_i x v_j \dots v_l$  is a chordless  $st$ -path, contradicting the fact that  $x$  is not a good neighbour of  $v_i$ . Clearly,  $\pi^*$  contains no repeated vertices, it is indeed an  $st$ -path. Let us prove that it is also chordless. This follows from the fact that there is no  $v_k \in \pi^*$ ,  $k \neq i$  and  $k \neq j$ , such that  $(v_k, x)$  is an edge of  $G$ , otherwise  $j - k$  or  $k - i$  would be strictly larger than  $j - i$ , contradicting our choice of  $v_i, v_j$  and  $x$ . Hence,  $v_i$  and  $v_j$  are adjacent in  $\pi_{st}$ . Each vertex of  $G$  is therefore a non-good neighbour of at most two vertices in  $\pi_{st}$ .  $\square$

### 3.5 Cleanup of current vertex

Suppose that we are exploring vertex  $u$  and we want to clean it as described in the previous sections. We identify a good vertex  $v$ , which is closer to  $t$  along the  $st$ -path than the other neighbours. Ideally, we would like to throw away all the other neighbours of  $u$  but this cost gives  $O(m)$  per chordless path as illustrated in Fig. 1 and discussed in Section 3.3. We perform a partial cleaning, called *cleanup*, which consists in identifying, among all neighbours of  $u$  (i.e.  $|N(u)|$  elements) only its set  $N^{good}(u)$  of good ones.

For a given  $u$  in a chordless  $st$ -path  $\pi_{st} = v_0 \dots v_i u \dots v_l$ , we let emerge the good neighbours in  $N^{good}(u)$  one by one as follows. Consider the graph  $G'$  where the vertices  $v_0 \dots v_i$  and its good neighbours were removed. If  $u$  and  $t$  are not connected, then there cannot be further chordless paths from  $u$  and so there cannot be further good neighbours. Otherwise, if  $u$  and  $t$  are connected, we take any path from  $u$  to  $t$ , and select its neighbour  $v$  that appears along the path and is closer to  $t$ , as illustrated in Fig. 3. After that, we remove  $v$  and its incident edges, and iterate what described above until  $u$  is disconnected from  $t$ . The vertices  $v$  thus selected form the set  $N^{good}(u)$  of good neighbours.



**Fig. 3.** Cleanup of the neighbours of vertex  $u$ .

**Lemma 2.** *For a chordless path  $\pi_{st}$ , the cleanup of vertex  $u \in \pi_{st}$  correctly produces the set  $N^{good}(u)$  of its good neighbours.*

## 4 Listing Algorithm

We blend the key ideas discussed in Section 3 to get Algorithm 1, which has four parameters as input and lists all the chordless  $st$ -paths: the first parameter is the chordless path  $\pi_{su}$  partially built from  $s$  to the current vertex  $u$  (initially,  $u = s$ ), which is the second parameter; the third parameter is a  $ut$ -path  $\pi_{ut}$  that is introduced following Fact 1; the fourth parameter is the reduced graph  $G$ , which changes with the recursive calls.

---

**Algorithm 1:** `list_induced_pathss,t( $\pi_{su}, u, \pi_{ut}, G$ )`

---

```

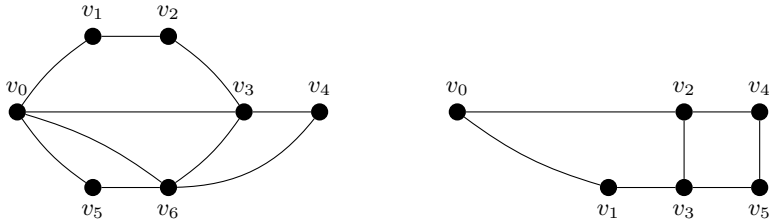
1 if  $u = t$  then
2   |   output( $\pi_{su}$ )
3 else
4   |    $S := \emptyset$ 
5   |   while true do
6     |    $v :=$  the vertex in  $\pi_{ut} \cap N(u)$  that is closer to  $t$ 
7     |    $\pi_{vt} :=$  the subpath of  $\pi_{ut}$  from  $v$  to  $t$ 
8     |    $S := S \cup \{(v, \pi_{vt})\}$ 
9     |   remove  $v$  and its incident edges from  $G$ 
10    |   if  $u$  and  $t$  are not connected then break
11    |    $\pi_{ut} :=$  any path from  $u$  to  $t$ 
12   |   end
13   |   foreach  $(v, \pi_{vt}) \in S$  do
14     |   adds back  $v$  and its incident edges to  $G$ 
15     |   list_induced_pathss,t( $\pi_{su} \cdot (u, v), v, \pi_{vt}, G$ )
16     |   remove  $v$  and its incident edges from  $G$ 
17   |   end
18 end

```

---

The algorithm outputs a chordless  $st$ -path if  $u = t$  (line 2). Otherwise, it performs a cleanup of  $u$  (the loop at lines 5–12). After that, it explores only the good neighbours recursively as they will surely lead to further chordless paths (the other loop at lines 13–17). Observe that  $S$  stores the good neighbours  $v$  of  $u$  and a  $vt$ -path for each of them: when performing the recursive call at line 15, only one of the vertices in  $S$  appears in the reduced graph  $G$  passed as a parameter to the recursive call (see lines 14 and 16 that guarantee this, and Remark 1 below). Hence, the recursive call now has as parameters the chordless  $sv$ -path  $\pi_{su} \cdot (u, v)$  ending in  $v$ , and a  $vt$ -path that guarantees that a chordless  $st$ -path exists and has  $\pi_{su} \cdot (u, v)$  as a prefix: all the chordless  $st$ -paths in the reduced graph share this prefix.





**Fig. 4.** Two example graphs where  $s = v_0$  and  $t = v_4$ .

For example, let us run Algorithm 1 on the input graph shown on the left of Fig. 4, with  $u = s = v_0$  and the initial path  $\pi_{ut} = v_0v_1v_2v_3v_4$ . It computes the pairs  $(v, \pi_{vt})$  in  $S$  as follows. First,  $(v_3, v_3v_4)$  is added to  $S$  as  $v_3$  is a good neighbour for  $v_0$  (the neighbour closer to  $t$  in the path), and the edges incident to  $v_3$  are removed. After this removal,  $s = v_0$  is still connected to  $t = v_4$  through the path  $v_0v_5v_6v_4$ , which becomes the input for the next iteration of the while loop. Next,  $(v_6, v_6v_4)$  is added to  $S$  as  $v_6$  is another good neighbour, and the edges incident to  $v_6$  are removed and disconnect  $s$  from  $t$ , so the while loop ends. The recursive calls in the foreach loop give the two chordless paths  $v_0v_3v_4$  and  $v_0v_6v_4$  contained in the graph.

*Remark 1.* It is important to run the recursive calls with all good neighbours in  $S$  removed except one. If we left two or more good neighbours in the recursive call of line 15, they could interfere with each other and we might not obtain the chordless paths correctly. A very simple example is given in the graph shown on the right of Fig. 4. Consider for instance the case where Algorithm 1 would be given as input the path  $v_0v_1v_3v_2v_4$ . The pair  $(v_2, v_2v_4)$  is added to  $S$  and  $v_2$  removed. After that, the path  $v_0v_1v_3v_5v_4$  is found and the pair  $(v_1, v_1v_3v_5v_4)$  is added to  $S$  and  $v_1$  removed. Since  $v_0$  and  $v_4$  become disconnected,  $S$  contains all the good neighbours of  $v_0$ . Algorithm 1 executes the recursive calls with  $S$ . Suppose that we keep both good neighbours  $v_1$  and  $v_2$  in  $G$  during these calls, in particular for the call with the pair  $(v_1, v_1v_3v_5v_4)$  from  $S$ . This call will extend in a nested call the chordless path to  $\pi_{su} = v_0v_1v_3$  for  $u = v_3$ , and will claim that the good neighbours of  $v_3$  are  $v_2$  and  $v_5$ , which is incorrect since  $v_0v_1v_3v_2$  is not chordless. This situation does not arise if  $v_2$  is kept deleted in  $G$  when the recursive call on  $v_1$  is performed as done in Algorithm 1.

The correctness of Algorithm 1 follows mostly from Lemma 2. Recall that, it guarantees that for a given path prefix  $\pi_{su}$  the set  $S$  contains the good neighbours of  $u$ , *i.e.* the neighbours of  $u$  that belong to at least one chordless  $st$ -path extending  $\pi_{su}$ . Clearly, we only have to recursively call the algorithm for these neighbours, the others certainly lead to no solution. This implies that Algorithm 1 tries all the possibilities to extend  $\pi_{su}$ , so all chordless  $st$ -paths are output. Moreover, since each good neighbour of  $u$  lead to a different extension, we have that no  $st$ -path is output more than once.

Certainly only  $st$ -paths are output by Algorithm 1, but at this point we have no guarantee that the paths are indeed chordless. In fact, after building  $S$ , the algorithm proceeds to recursively extend the prefix  $\pi_{su}(u, v)$  for each  $v \in S$  in the graph  $G' = G \setminus (S \setminus \{v\})$ . However, since  $u$  was included in current path none of its neighbours can be used later in the recursion. The algorithm removes the good neighbours of  $u$  from  $G$ , but the other neighbours,  $N_G(u) \setminus S$ , are still present in  $G'$ . They could thus be used to extend the path later in the recursion, resulting in a non-chordless  $st$ -path. Lemma 3 shows that this cannot happen.

**Lemma 3.** *The  $st$ -paths output by Algorithm 1 are chordless.*

The previous lemma leads to the following theorem.

**Theorem 1.** *The algorithm correctly outputs all chordless  $st$ -paths of  $G$ .*

**Theorem 2.** *The algorithm takes  $O(m + |\mathcal{P}_{st}(G)|(t_p + nt_q + nt_u))$  time, where  $t_p$  is the cost of choosing any path from any two given vertices,  $t_q$  is the cost of checking if any given two vertices are connected or not, and  $t_u$  is the cost of removing/adding back any given edge.*

*Proof.* See Section 5.

There are several dynamic data structures in the literature [8] that maintain a spanning forest for a dynamic graph, supporting insertions and deletions of edges in polylogarithmic time. Consequently,  $t_p = O(n \text{ polylog}(n))$ ,  $t_q = O(\text{polylog}(n))$ , and  $t_u = O(\text{polylog}(n))$ , thus giving the following bound.

**Corollary 1.** *The algorithm takes  $\tilde{O}(m + |\mathcal{P}_{st}(G)| \cdot n)$  time to report all the chordless  $st$ -paths.*

## 5 Amortized Analysis

Before starting our analysis, we observe some simple properties of the recursion tree generated by Algorithm 1.

**Fact 2** *The recursion tree  $R$  of Algorithm 1 has the following properties:*

1. *There is a one-to-one correspondence between paths in  $\mathcal{P}_{st}(G)$  and leaves in the recursion tree.*
2. *There is a one-to-one correspondence between proper prefixes of paths in  $\mathcal{P}_{st}(G)$  and internal nodes in the recursion tree.*
3. *The number of branching nodes is  $|\mathcal{P}_{st}(G)| - 1$ .*
4. *The length of a root-to-leaf path is equal to the length of the chordless  $st$ -path corresponding to the leaf. In particular, the height of the tree is  $\leq n$ .*

Fact 2 suggests us to follow the following overall strategy.

1. We analyze the cost of each type (leaf, unary and branching) of node separately.

2. We consider all branching nodes together, and show that their amortized cost is  $O(t_p + t_q + nt_u + n) = \tilde{O}(n)$  per solution.
3. We consider all unary nodes together, and show that their amortized cost is  $O(|\pi_{st}|t_q + nt_u) = \tilde{O}(n)$  per solution.
4. We deduce that the cost of each solution is  $O(t_p + nt_q + nt_u) = \tilde{O}(n)$ .

**Lemma 4.** *The cost of a leaf is  $O(|\pi_{st}|)$ .*

Let us now analyze the cost of the unary nodes. Let  $r = \langle \pi_{su}, u, \pi_{ut}, G \rangle$  be a unary node. The vertex  $v \in N(u)$  is the only neighbour of  $u$  that can extend the prefix  $\pi_{su}$  into a chordless  $st$ -path. Thus, removing  $v$  from  $G$  disconnects  $u$  from  $t$ , and the algorithm performs a single iteration of the loop in line 5, not executing line 11. In this case, the algorithm performs the following operations: (i) one connectivity query (line 10), (ii)  $|N(v)|$  edge update operations on  $G$  (lines 9, 14 and 16), and (iii) a scan in the intersection of  $N(u)$  and  $\pi_{ut}$  to find  $v$  (line 6). The cost of (i) and (ii) is  $O(t_q + |N(v)|t_u)$ .

A naive implementation of (iii) takes  $O(|N(u)| + |\pi_{ut}|)$  time, which is too large to fit in our amortization strategy. In order to reduce this cost to  $O(|N(u)|)$  we therefore maintain, as an extra invariant, for each vertex in the current graph its distance to  $t$  in the path  $\pi_{ut}$ . In this way, we can find  $v$  simply scanning  $N(u)$ . Thus, assuming the distance information is correctly maintained, we complete the proof of Lemma 5.

**Lemma 5.** *The cost of a unary node is  $O(t_q + |N(v)|t_u + |N(u)|)$ , where  $(u, v)$  is the edge added to the chordless path.*

It is not hard to maintain the distance information for  $\langle \pi_{su}(u, v), v, \pi_{vt}, G' \rangle$ , the only child of the unary node  $\langle \pi_{su}, u, \pi_{ut}, G \rangle$ . As the path  $\pi_{vt}$  is a suffix of  $\pi_{ut}$ , the distance of the vertices in  $\pi_{ut}$  do not change. On the other hand, the only vertices that the distances can change are the ones in  $\pi_{vt}$  but not in  $\pi_{ut}$ . These vertices can be identified when scanning  $N(v)$  in the child node  $\langle \pi_{su}(u, v), v, \pi_{vt}, G' \rangle$ , since their distance is strictly larger than  $|\pi_{vt}|$ . It remains to show that the distance information can be maintained in the branching nodes. Lemma 6 shows that this is indeed the case, and gives an upper bound on the cost of the branching nodes.

**Lemma 6.** *The cost of a branching node  $r \in R$  is  $O(\beta(r)(t_p + t_q + nt_u))$ , where  $\beta(r)$  is the number of children of  $r$ .*

It remains to prove that we can maintain the distance information in branching nodes in the same time bound of Lemma 6. This follows from the fact that in each iteration of the loop (line 5) we already paying  $O(|\pi_{ut}|)$ , *i.e.* a full traversal of the path  $\pi_{ut}$ . Before each recursive call in line 15 we can traverse the path  $\pi_{ut}$  adding for each vertex the distance information, *i.e.* their position in the path.

At this point we have bounds for the cost of each node in the recursion tree. However, by directly applying them we cannot achieve our goal of  $\tilde{O}(n)$  time per solution. For instance, consider the particular case where all internal

nodes of the recursion tree are branching. The cost of each internal node is  $O(\beta(r)(t_p + t_q + nt_u)) = \tilde{O}(n^2)$ , since  $\beta(r) = O(n)$  in the worst case. Then, from item 3 of Fact 2, the number of branching nodes is  $|\mathcal{P}_{st}(G)| - 1$ . The total cost for the tree is thus  $\tilde{O}(|\mathcal{P}_{st}(G)|n^2)$  or  $\tilde{O}(n^2)$  per solution.

In order to get a tighter bound for the total cost of the branching nodes, we use the following amortization strategy. Let  $r \in R$  be a branching node. We divide the cost  $O(\beta(r)(t_p + t_q + nt_u))$  among the closer descendants that are branching nodes or leaves (no unary nodes), each being charged  $O(t_p + t_q + nt_u)$ . This can always be done since  $r$  has  $\beta(r)$  children and the subtree of each child contains at least one leaf, *i.e.* the node  $r$  has at least  $\beta(r)$  non-unary descendants. In this way, the original cost of node  $r$  is completely charged to its non-unary descendants, and the only cost that remains associated to  $r$  is the one received from its ancestors. Finally, each branching node can only be charged once, by its lowest non-unary ancestor. Each branching node and each leaf is therefore charged with  $O(t_p + t_q + nt_u)$ . Thus, the total cost of the branching nodes is  $O(|\mathcal{P}_{st}(G)|(t_p + t_q + nt_u))$ , completing the proof of Lemma 7.

**Lemma 7.**  $\sum_{r:\text{branching}} T(r) = O(|\mathcal{P}_{st}(G)|(t_p + t_q + nt_u))$ .

Let us now bound the total cost of the unary nodes. Similarly to the branching nodes case, a straightforward use of the bound given by Lemma 5 leads to an  $\tilde{O}(n^2)$  cost per solution, since in the worst case the recursion tree can have  $O(n)$  unary nodes for each leaf. The key idea to obtain a better amortized cost is to consider the bound on the reduced degrees given by Lemma 1.

We first observe that each unary node is contained in some root-to-leaf path in the recursion tree. Thus,

$$\sum_{r:\text{unary}} T(r) \leq \sum_{l:\text{leaf}} \sum_{r \in \text{root} \rightsquigarrow l} T(r). \quad (1)$$

Fact 2 implies that there is a one-to-one correspondence between the prefixes of paths in  $\mathcal{P}_{st}(G)$  and nodes in the recursion tree. That is, the each leaf corresponds to a solution, and the root-to-leaf path  $\text{root} \rightsquigarrow l$  corresponds to the chordless  $st$ -path associated to the leaf  $l$ . We can thus rewrite the double sum as

$$\sum_{l:\text{leaf}} \sum_{r \in \text{root} \rightsquigarrow l} T(r) = \sum_{\pi \in \mathcal{P}_{st}(G)} \sum_{v_i \in \pi} (t_q + |N(v_i)|t_u + |N(v_{i+1})|). \quad (2)$$

For each chordless  $st$ -path  $\pi_{st} = v_0 \dots v_i v_{i+1} \dots v_k$  in the internal sum of Eq. 2, we have that the degrees are actually the reduced degrees of Section 3.4, since the good neighbours (*i.e.* the set  $S$  in Algorithm 1) are always removed. Using Lemma 1 we can thus bound the sum of the degrees by  $2n$ . Therefore,

$$\sum_{r:\text{unary}} T(r) \leq \sum_{\pi \in \mathcal{P}_{st}(G)} (|\pi|t_q + 2nt_u + 2n), \quad (3)$$

completing the proof of Lemma 8.

**Lemma 8.**  $\sum_{r:\text{unary}} T(r) = O(\sum_{\pi \in \mathcal{P}_{st}(G)} |\pi|t_q + nt_u)$ .

As a corollary of Lemmas 8 and 7, we obtain Theorem 2.

## References

1. Etienne Birmelé, Rui A. Ferreira, Roberto Grossi, Andrea Marino, Nadia Pisanti, Romeo Rizzi, and Gustavo Sacomoto. Optimal listing of cycles and st-paths in undirected graphs. In *SODA 2013*, 1884–1896. ACM/SIAM, 2013.
2. Yijia Chen and Jörg Flum. On parameterized path and chordless path problems. In *IEEE Conference on Computational Complexity*, pages 250–263, 2007.
3. Maria Chudnovsky, Neil Robertson, Paul Seymour, and Robin Thomas. The strong perfect graph theorem. *Annals of Mathematics*, 164:51–229, 2006.
4. Michele Conforti, Gérard Cornuéjols, Ajai Kapoor, and Kristina Vuskovic. Recognizing balanced 0, +/- matrices. In *SODA 1994*, 103–111. ACM/SIAM, 1994.
5. Michele Conforti, Gérard Cornuéjols, Ajai Kapoor, and Kristina Vuskovic. Finding an even hole in a graph. In *FOCS 1997*, 480–485. IEEE Computer Society, 1997.
6. Michele Conforti and M. R. Rao. Structural properties and decomposition of linear balanced matrices. *Math. Program.*, 55:129–168, 1992.
7. Robert Haas and Michael Hoffmann. Chordless paths through three vertices. *Theoretical Computer Science*, 351(3):360 – 371, 2006.
8. Bruce M. Kapron, Valerie King, and Ben Mountjoy. Dynamic graph connectivity in polylogarithmic worst case time. In *SODA*, pages 1131–1142, 2013.
9. Ken-ichi Kawarabayashi and Yusuke Kobayashi. The induced disjoint paths problem. In Andrea Lodi, Alessandro Panconesi, and Giovanni Rinaldi, editors, *IPCO*, volume 5035 of *Lecture Notes in Computer Science*, pages 47–61. Springer, 2008.
10. R C Read and Robert E Tarjan. Bounds on backtrack algorithms for listing cycles, paths, and spanning trees. *Networks*, 5(3):237252, 1975.
11. D. Seinsche. On a property of the class of n-colorable graphs. *Journal of Combinatorial Theory, Series B*, 16(2):191 – 193, 1974.
12. Nayla Sokhn, Richard Baltensperger, Louis-Felix Bersier, Jean Hennebert, and Ulrich Ultes-Nitsche. Identification of chordless cycles in ecological networks. In *COMPLEX*, 2012.
13. Maciej M. Syslo. An efficient cycle vector space algorithm for listing all cycles of a planar graph. *SIAM J. Comput.*, 10(4):797–808, 1981.
14. Takeaki Uno. Algorithms for enumerating all perfect, maximum and maximal matchings in bipartite graphs. In *ISAAC 1997*, LNCS, 92–101. Springer, 1997.
15. Takeaki Uno. An output linear time algorithm for enumerating chordless cycles. In *92nd SIGAL of Information Processing Society Japan*, pages 47–53, 2003. (in Japanese).
16. Marcel Wild. Generating all cycles, chordless cycles, and hamiltonian cycles with the principle of exclusion. *J. of Discrete Algorithms*, 6(1):93–102, March 2008.

## Appendix: Omitted Proofs

*Proof (Lemma 2).* Given a chordless  $st$ -path  $\pi_{st} = v_0v_iu\dots v_\ell$  in the graph  $G$ , for some integer  $\ell > 1$ , where  $v_0 = s$  and  $v_\ell = t$ . Let  $G'$  be the subgraph of  $G$  where the vertices  $\{v_0, \dots, v_i\}$  and their good neighbours were removed. Let  $S \subseteq N(v_i)$  be the set of vertices removed by the cleanup procedure for  $u$  in  $G'$ . We divide the proof in two parts. We first show that  $S$  contains  $N^{good}(v_i)$ , and then show that  $N^{good}(v_i)$  contains  $S$ .

Clearly, for the first part, it is enough to show that  $N_{G'}(v_i) \setminus S$  are not good neighbours. The vertices of  $N_{G'}(v_i) \setminus S$  cannot reach  $t$  in  $G'$  without passing through some vertex in  $S$ , otherwise  $u$  would still be connected to  $t$  and the procedure would not stop. This implies that there is no chordless  $ut$ -path in  $G'$  using some vertex of  $N_{G'}(v_i) \setminus S$ , *i.e.* they are not good neighbours of  $v_i$ .

Finally, for the second part, it is enough to show that for each  $w \in S$  there is a chordless  $ut$ -path in  $G'$  passing through  $w$ . Consider the iteration where  $w \in N_{G'}(u)$  was added to  $S$  and let  $S' \subseteq S$  be the set of vertices added in previous iterations. We have that  $w$  is the vertex closer to  $t$  in the path  $\pi = u \rightsquigarrow t$  in  $G' \setminus S'$ . We claim that any subpath of  $(u, w)\pi_{wt}$  in  $G' \setminus S'$  contains  $(u, w)$ , where  $\pi_{wt} = w \rightsquigarrow t$  is a suffix of  $\pi$ . Thus implying that  $w$  is contained in an induced  $ut$ -path in  $G'$ . Indeed, the edge  $(u, w)$  is not contained in a subpath iff there is a vertex  $x \in N_{G'}(u)$  in  $\pi_{wt}$ . By construction,  $\pi_{wt}$  does not contain any vertex of  $S'$ ; and by the choice of  $w$ ,  $\pi_{wt}$  does not contain any vertex of  $N_{G'}(u) \setminus S'$ .  $\square$

*Proof (Lemma 3).* We proceed by contradiction. Suppose  $\pi_{st}$  is output by the algorithm and it is not chordless. This means that there exists a pair of vertices  $x, y \in \pi_{st}$  such that  $x \neq y$ ,  $(x, y)$  is an edge of  $G$ , and  $(x, y) \notin \pi_{st}$ . We can assume the edge  $(x, y)$  is such that  $x$  is the vertex closer to  $s$ . Now, consider the recursive call corresponding to the prefix  $\pi_{sx}$ : let  $G_x$  be the associated graph and  $z$  the next vertex of  $\pi_{st}$ . The suffix  $\pi_{xt}$  of  $\pi_{st}$  must pass through  $z$ , since  $x$  is not connected to  $t$  in  $G_x - S$ . So,  $y$  is closer to  $t$  than  $z$  in  $\pi_{st}$ . Thus, the path  $(x, y)\pi_{yt}$ , where  $\pi_{yt}$  is a suffix of  $\pi_{st}$ , avoids  $S$  in  $G_x$ . This contradicts the test in line 10.  $\square$

*Proof (Lemma 4).* Clearly, when  $u = t$ , the only operation done by the algorithm is to output  $\pi_{st}$ , which takes  $O(|\pi_{st}|)$  time.  $\square$

*Proof (Lemma 6).* The cost of a branching node  $r = \langle \pi_{su}, u, \pi_{ut}, G \rangle$  is dominated by the cost of the loop of line 5. The number of iterations of the loop is equal to the number of neighbours of  $u$  that can extend  $\pi_{su}$  into a chordless  $st$ -path, which is exactly the number of vertices in  $S$  after the loop finishes, *i.e.*  $\beta(r)$ , the number of children of  $r$ . Let us now bound the cost of each iteration. The cost of lines 6 and 7 is bounded by  $O(|N(u)| + |\pi_{ut}|) = O(n)$ : we simply have to traverse the path  $\pi_{ut}$  and scan the set  $N(u)$ . The cost of updating  $G$  is bounded by  $O(nt_u)$ . Finally, the cost for the connectivity query and to find a path is  $t_p + t_q$ . Hence, the total cost for a branching node is  $O(\beta(r)(t_p + t_q + nt_u))$ .  $\square$