



# Faster algorithms for counting subgraphs in sparse graphs

Marco Bressan<sup>1</sup>

Received: 6 January 2020 / Accepted: 2 February 2021  
© The Author(s) 2021

## Abstract

Given a  $k$ -node pattern graph  $H$  and an  $n$ -node host graph  $G$ , the subgraph counting problem asks to compute the number of copies of  $H$  in  $G$ . In this work we address the following question: can we count the copies of  $H$  faster if  $G$  is sparse? We answer in the affirmative by introducing a novel tree-like decomposition for directed acyclic graphs, inspired by the classic tree decomposition for undirected graphs. This decomposition gives a dynamic program for counting the homomorphisms of  $H$  in  $G$  by exploiting the degeneracy of  $G$ , which allows us to beat the state-of-the-art subgraph counting algorithms when  $G$  is sparse enough. For example, we can count the induced copies of any  $k$ -node pattern  $H$  in time  $2^{O(k^2)} O(n^{0.25k+2} \log n)$  if  $G$  has bounded degeneracy, and in time  $2^{O(k^2)} O(n^{0.625k+2} \log n)$  if  $G$  has bounded average degree. These bounds are instantiations of a more general result, parameterized by the degeneracy of  $G$  and the structure of  $H$ , which generalizes classic bounds on counting cliques and complete bipartite graphs. We also give lower bounds based on the Exponential Time Hypothesis, showing that our results are actually a characterization of the complexity of subgraph counting in bounded-degeneracy graphs.

**Keywords** Subgraph counting · Tree decomposition · Degeneracy · Sparsity

---

A preliminary version of this article, [4], appeared in the proceedings of the 14th International Symposium on Parameterized and Exact Computation (IPEC 2019).

---

✉ Marco Bressan  
marco.bressan@unimi.it

<sup>1</sup> Dipartimento di Informatica, Università Statale di Milano, Milan, Italy

## 1 Introduction

We address the following fundamental subgraph counting problem:

*Input:* an  $n$ -node graph  $G$  (the *host graph*) and a  $k$ -node graph  $H$  (the *pattern*)

*Output:* the number of induced copies of  $H$  in  $G$

If no further assumptions are made, the best possible algorithm for this problem is likely to have running time  $f(k) \cdot n^{\Theta(k)}$ . Indeed, the naive brute-force algorithm has running time  $O(k^2 n^k)$ , and under the Exponential Time Hypothesis [23] any algorithm for counting  $k$ -cliques has running time  $n^{\Omega(k)}$  [8, 9]. The best algorithm known, which was given over 30 years ago by Nešetřil and Poljak [29] and is based on fast matrix multiplication, is only slightly faster than  $O(k^2 n^k)$ . Ignoring poly( $k$ ) factors,<sup>1</sup> the algorithm runs in time  $O(n^{\omega \lfloor \frac{k}{3} \rfloor + 2})$  where  $\omega$  is the matrix multiplication exponent. Since  $\omega \leq 2.373$  [25], this gives a state-of-the-art running time of  $O(n^{0.791k+2})$ .

In this work, we aim at breaking through this “ $n^{\Theta(k)}$  barrier” by assuming that  $G$  is sparse, and in particular, that  $G$  has bounded degeneracy. This assumption is often made for real-world graphs like social networks, since it agrees well with their structural properties [17]. The family of bounded-degeneracy graphs is rich from a theoretical point of view, too: it includes many important classes such as Barabási-Albert preferential attachment graphs, graphs excluding a fixed minor, planar graphs, bounded-treewidth graphs, bounded-degree graphs, and bounded-genus graphs, see [20]. Unfortunately, even when  $G$  has bounded degeneracy, the state of the art remains the  $O(n^{0.791k+2})$ -time algorithm by Nešetřil and Poljak, unless one makes further assumptions. For example, one can count the copies of any given pattern  $H$  in time  $O(n)$ , provided  $G$  is planar [15] or has bounded treewidth [27] or has bounded degree [30]; all conditions that are stricter than bounded degeneracy. Alternatively, if  $G$  has bounded degeneracy,  $O(n)$ -time algorithms exist when  $H$  is the clique [1, 10, 16], or when  $H$  is a complete bipartite graph, if we do not require the copies of  $H$  to be induced [14]. Unfortunately, it is not clear how to extend the techniques behind these results to all patterns  $H$  and all  $G$  with bounded degeneracy. Thus, to what extent a small degeneracy of  $G$  makes subgraph counting easier remains an open question.

In this work we introduce a novel tree-like graph decomposition, to be applied to the pattern graph  $H$ , designed to exploit the degeneracy of  $G$  when counting the homomorphisms from  $H$  to  $G$ . When  $G$  is sparse enough, this decomposition yields subgraph counting algorithms faster than the state of the art. For example, we show how to count the induced copies of *any*  $k$ -node pattern  $H$  in time  $2^{O(k^2)} O(n^{0.25k+2} \log n)$  when  $G$  has bounded degeneracy, and in time  $2^{O(k^2)} O(n^{0.625k+2} \log n)$  when  $G$  has bounded average degree. These results are instantiations of a more general result which says that  $H$  can be counted in time  $f(k) O(d^{k-\tau(H)} n^{\tau(H)} \log n)$ , where  $d$  is the degeneracy of  $G$ , and  $\tau(H)$  is a certain measure of “width” of  $H$  arising from our decomposition. Assuming the Exponential

<sup>1</sup> In this paper we suppress poly( $k$ ) factors by default; if needed, we explicit them in order to emphasize that the dependence on  $k$  is polynomial rather than exponential.

Time Hypothesis, we also show that  $n^{\Omega(\tau(H)/\log \tau(H))}$  operations are required in the worst case, even if  $G$  has degeneracy 2. This provides a novel characterization of the complexity of subgraph counting in bounded-degeneracy graphs.

### 1.1 Results

We divide our results into *bounds* (Sect. 1.1.1) and *techniques* (Sect. 1.1.2). We denote by  $d$  the degeneracy of  $G$ , and we denote by  $\text{hom}(H, G)$ ,  $\text{sub}(H, G)$ ,  $\text{ind}(H, G)$  the number of, respectively, homomorphisms, occurrences, and induced occurrences of  $H$  in  $G$ . See Sect. 1.2 for further definitions and notation. We remark that, unless otherwise specified, our bounds hold for every  $H$  including disconnected ones.

#### 1.1.1 Bounds

Our first results are two running time bounds parameterized by the sparsity of  $G$ .

**Theorem 1** *For any  $k$ -node pattern  $H$  one can compute  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{k}{4} \rfloor + 2} n^{\lfloor \frac{k}{4} \rfloor + 2} \log n)$ , and one can compute  $\text{ind}(H, G)$  in time  $2^{O(k^2)} \cdot O(d^{k - \lfloor \frac{k}{4} \rfloor + 2} n^{\lfloor \frac{k}{4} \rfloor + 2} \log n)$ , where  $d$  is the degeneracy of  $G$ .*

This bound reduces the exponent of  $n$  to  $\lfloor \frac{k}{4} \rfloor + 2 \leq 0.25k + 2$ , down from the state-of-the-art  $\omega \lfloor \frac{k}{3} \rfloor + 2 \leq 0.791k + 2$  of the Nešetřil–Poljak bound. This implies that our polynomial dependence on  $n$  is better whenever  $d = O(n^{0.721})$ , and in any case (that is, even if  $\omega = 2$ ) whenever  $d = O(n^{0.556})$ . As a corollary of Theorem 1, since  $d = O(\sqrt{rn})$  where  $r$  is the average degree of  $G$ , we obtain:

**Theorem 2** *For any  $k$ -node pattern  $H$  one can compute  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(r^{\frac{1}{2}(k - \lfloor \frac{k}{4} \rfloor) - 1} n^{\frac{1}{2}(k + \lfloor \frac{k}{4} \rfloor) + 1} \log n)$ , and one can compute  $\text{ind}(H, G)$  in time  $2^{O(k^2)} \cdot O(r^{\frac{1}{2}(k - \lfloor \frac{k}{4} \rfloor) - 1} n^{\frac{1}{2}(k + \lfloor \frac{k}{4} \rfloor) + 1} \log n)$ , where  $r$  is the average degree of  $G$ .*

This bound has a polynomial dependence on  $n$  better than Nešetřil–Poljak whenever  $r = O(n^{0.221})$ , and in any case (that is, even if  $\omega = 2$ ) whenever  $r = O(n^{0.056})$ . In particular, we have a  $2^{O(k^2)} \cdot O(n^{0.625k+1} \log n)$ -time algorithm when  $r = O(1)$ . These are the first improvements over the Nešetřil–Poljak algorithm for graphs with small degeneracy or small average degree.

As a second result, we give improved bounds for some classes of patterns. The first is the class of quasi-cliques, a typical target pattern for social networks [5–7, 31–33]. We prove:

**Theorem 3** *If  $H$  is the clique minus  $\epsilon$  edges, then one can compute  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{1}{2} + \sqrt{\frac{\epsilon}{2}} \rfloor} n^{\lfloor \frac{1}{2} + \sqrt{\frac{\epsilon}{2}} \rfloor} \log n)$ , and  $\text{ind}(H, G)$  in time  $2^{O(\epsilon + k \log k)} \cdot O(d^{k - \lfloor \frac{1}{2} + \sqrt{\frac{\epsilon}{2}} \rfloor} n^{\lfloor \frac{1}{2} + \sqrt{\frac{\epsilon}{2}} \rfloor} \log n)$ .*

**Table 1** Summary of upper bounds for the problem of counting the number of occurrences of  $H$  in  $G$

| Pattern $H$                               | Time to compute $\text{ind}(H, G)$  | References |
|---|---|------------|
| All (even disconnected)                   | $O(n^{\omega\lfloor \frac{k}{3} \rfloor + 2})$  | [29]       |
| All (even disconnected)                   | $2^{O(k^2)} \cdot O(d^{k - \lfloor \frac{k}{4} \rfloor - 2} n^{\lfloor \frac{k}{4} \rfloor + 2} \log n)$  | This work  |
| All (even disconnected)                   | $2^{O(k^2)} \cdot O(r^{\frac{1}{2}(k - \lfloor \frac{k}{4} \rfloor) - 1} n^{\frac{1}{2}(k + \lfloor \frac{k}{4} \rfloor) + 1} \log n)$  | This work  |
| $K_k$                                     | $O(d^{k-1}n)$   | [10]       |
| $K_k - \epsilon$ edges                    | $2^{O(\epsilon + k \log k)} \cdot O(d^{k - \lfloor \frac{k}{2} \rfloor + \sqrt{\frac{\epsilon}{2}} \lfloor n^{\lfloor \frac{k}{2} \rfloor + \sqrt{\frac{\epsilon}{2}} \rfloor} \log n)$ | This work  |
| $K_{k_1, k_2}$                            | $O(d^3 2^{2d} n)$ (sub( $H, G$ ) only)  | [14]       |
| $K_{k_1, \dots, k_\ell}$                  | $2^{O(k \log k)} \cdot O(d^{k-1} n \log n)$ (sub( $H, G$ ) only)  | This work  |
| $K_{k_1, \dots, k_\ell} + \epsilon$ edges | $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{k}{4} \rfloor - 2} n^{\lfloor \frac{k}{4} \rfloor + 2} \log n)$ (sub( $H, G$ ) only)  | This work  |

The graphs  $G$  and  $H$  have  $n$  and  $k$  vertices respectively, and  $d$  is the degeneracy of  $G$ . All except the last tree bounds hold for counting the number of *induced* occurrences

This generalizes the classic  $O(d^{k-1}n)$  bound for counting cliques by Chiba and Nishizeki [10], at the price of an extra factor  $2^{O(\epsilon + k \log k)} O(\log n)$ . Next, we consider complete quasi-multipartite graphs:

**Theorem 4** *If  $H$  is a complete multipartite graph, then one can compute  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(d^{k-1} n \log n)$ . If  $H$  is a complete multipartite graph plus  $\epsilon$  edges, then one can compute  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{k}{4} \rfloor - 2} n^{\lfloor \frac{k}{4} \rfloor + 2} \log n)$ .*

This generalizes an existing  $O(d^3 2^{2d} n)$  bound for counting the non-induced copies of complete (maximal) bi-partite graphs [14], again at the price of an extra factor  $2^{O(k \log k)} \log n$ .

Table 1 summarizes our upper bounds. We remark that our algorithms work for the colored versions of the problem (count only copies of  $H$  with prescribed vertex and/or edge colors) as well as the weighted versions of the problem (compute the total node or edge weight of copies of  $H$  in  $G$ ). This can be obtained by a straightforward adaptation of our homomorphism counting algorithms.

### 1.1.2 Techniques

The bounds of Sect. 1.1.1 are instantiations of a single, more general result. This result is based on a novel notion of width, the *dag treewidth*  $\tau(H)$  of  $H$ , which captures the relevant structure of  $H$  when counting its copies in a  $d$ -degenerate graph. In a simplified form, the bound is the following:

**Theorem 5** *For any  $k$ -node pattern  $H$  one can compute  $\text{hom}(H, G)$ ,  $\text{sub}(H, G)$ , and  $\text{ind}(H, G)$  in time  $f(k) \cdot O(d^{k - \tau(H)} n^{\tau(H)} \log n)$ .*

Let us briefly explain this result. The heart of the problem is computing  $\text{hom}(H, G)$ ; once we know how to do this, we can obtain  $\text{sub}(H, G)$  and  $\text{ind}(H, G)$  via inclusion-exclusion arguments at the price of an extra multiplicative factor  $f(k)$ , like in [3, 11]. To compute  $\text{hom}(H, G)$ , we give  $G$  an acyclic orientation with maximum outdegree  $d$ . Then, we take every possible acyclic orientation  $P$  of  $H$ , and compute  $\text{hom}(P, G)$  where by  $\text{hom}(P, G)$  we mean the number of homomorphisms from  $P$  to  $G$  that respect the orientations of the arcs. Note that the number of such homomorphisms can be  $n^{\Omega(k)}$  even if  $G$  has bounded degeneracy (for example, if  $P$  is an independent set), so we cannot list them explicitly. At this point we introduce our technical tool, the *dag tree decomposition* of  $P$ . This is a tree  $T$  that captures the relevant reachability relations between the nodes of  $P$ . Given  $T$ , one can compute  $\text{hom}(P, G)$  via dynamic programming in time  $f(k) \cdot O(d^{k-\tau(T)} n^{\tau(T)} \log n)$ , where  $\tau(T) \in \{1, \dots, k\}$  is the *width* of  $T$ . The dynamic program computes  $\text{hom}(P, G)$  by combining carefully the homomorphism counts of certain subgraphs of  $P$ . The dag-treewidth  $\tau(H)$ , which is the parameter appearing in the bound of Theorem 5, is the maximum width of the optimal dag tree decomposition of any acyclic orientation  $P$  of any graph obtainable by identifying nodes of and/or adding edges to  $H$  (this arises from the inclusion-exclusion arguments). With this, our technical machinery is complete. To obtain the bounds of the previous paragraph, we show how to compute efficiently dag tree decompositions of low width, and apply a more technical version of Theorem 5.

We conclude by complementing Theorem 5 with a lower bound based on the Exponential Time Hypothesis. This lower bound shows that in the worst case the dag-treewidth  $\tau(H)$  cannot be beaten, and therefore our decomposition captures, at least in part, the complexity of counting subgraphs in  $d$ -degenerate graphs.

**Theorem 6** *Under the Exponential Time Hypothesis [23], no algorithm can compute  $\text{sub}(H, G)$  or  $\text{ind}(H, G)$  in time  $f(d, k) \cdot n^{o(\tau(H)/\log \tau(H))}$  for all  $H$ .*

### 1.2 Preliminaries and notation

Both  $G = (V, E)$  and  $H = (V_H, E_H)$  are simple graphs, possibly disconnected. For any subset  $V' \subseteq V$  we denote by  $G[V']$  the subgraph of  $G$  induced by  $V'$ ; the same notation applies to any graph. A *homomorphism* from  $H$  to  $G$  is a map  $\phi : V_H \rightarrow V$  such that  $\{u, u'\} \in E_H$  implies  $\{\phi(u), \phi(u')\} \in E$ . We write  $\phi : H \rightarrow G$  to highlight the edges that  $\phi$  preserves. When  $H$  and  $G$  are oriented,  $\phi$  must preserve the direction of the arcs. If  $\phi$  is injective then we have an injective homomorphism. We denote by  $\text{hom}(H, G)$  and  $\text{inj}(H, G)$  the number of homomorphisms and injective homomorphisms from  $H$  to  $G$ . To avoid confusion, we will use the symbol  $\psi$  to denote maps that are not necessarily homomorphisms. The symbol  $\simeq$  denotes isomorphism. A *copy* of  $H$  in  $G$  is a subgraph  $F \subseteq G$  such that  $F \simeq H$ . If moreover  $F \simeq G[V_F]$  then  $F$  is an induced copy. We denote by  $\text{sub}(H, G)$  and  $\text{ind}(H, G)$  the number of copies and induced copies of  $H$  in  $G$ ; we may omit  $G$  if clear from the context. When we give an acyclic orientation to the edges of  $H$ , we denote the resulting dag by  $P$ . All the notation described above applies to directed graphs in the natural way.

The *degeneracy* of  $G$  is the smallest integer  $d$  such that there is an acyclic orientation of  $G$  with maximum outdegree bounded by  $d$ . Such an orientation can be found in time  $O(|E|)$  by repeatedly removing from  $G$  a minimum-degree node [27]. From now on we assume that  $G$  has this orientation. Equivalently,  $d$  is the smallest integer that bounds from above the minimum degree of every subgraph of  $G$ .

We assume the following operations take constant time: accessing the  $i$ -th arc of any node  $u \in V$ , and checking if  $(u, v)$  is an arc of  $G$  for any pair  $(u, v)$ . Our upper bounds still hold if checking an arc takes time  $O(\log n)$ , which can be achieved via binary search if we first sort the adjacency lists of  $G$ . The  $\log n$  factor in our bounds appears since we assume logarithmic access time for our dictionaries, each of which holds  $O(n^k)$  entries. This factor can be removed by using dictionaries with worst-case  $O(1)$  access time (e.g., hash maps), at the price of obtaining probabilistic/amortized bounds rather than deterministic ones.

Finally, we recall the tree decomposition and treewidth of a graph. For any two nodes  $X, Y$  in a tree  $T$ , we denote by  $T(X, Y)$  the unique path between  $X$  and  $Y$  in  $T$ .

**Definition 1** (see [13], Ch. 12.3) Given a graph  $G = (V, E)$ , a tree decomposition of  $G$  is a tree  $T = (V_T, E_T)$  such that each node  $X \in V_T$  is a subset  $X \subseteq V$ , and that:<sup>2</sup>

1.  $\cup_{X \in V_T} X = V$
2. for every edge  $e = \{u, v\} \in G$  there exists  $X \in T$  such that  $u, v \in X$
3.  $\forall X, X', X'' \in V_T$ , if  $X \in T(X', X'')$  then  $X' \cap X'' \subseteq X$

The width of a tree decomposition  $T$  is  $t(T) = \max_{X \in V_T} |X| - 1$ . The treewidth  $t(G)$  of a graph  $G$  is the minimum of  $t(T)$  over all tree decompositions  $T$  of  $G$ .

### 1.3 Related work

As discussed above, the fastest algorithm known for computing  $\text{ind}(H, G)$  is the one by Nešetřil and Poljak [29] that runs in time  $O(n^{\omega \lfloor \frac{k}{3} \rfloor + (k \bmod 3)})$  where  $\omega$  is the matrix multiplication exponent. With the current bound  $\omega \leq 2.373$ , this running time is in  $O(n^{0.791k+2})$ . Unfortunately, the algorithm is based on fast matrix multiplication, which makes it oblivious to the sparsity of  $G$ .

Under certain assumptions on  $G$ , faster algorithms are known. If  $G$  has bounded maximum degree,  $\Delta = O(1)$ , then we can compute  $\text{ind}(H, G)$  in time  $c^k \cdot O(n)$  for some  $c = c(\Delta)$  via multivariate graph polynomials [30]. If  $G$  has treewidth  $t(G) \leq k$ , and we are given a tree decomposition of  $G$  of such width, then we can compute  $\text{ind}(H, G)$  in time  $2^{O(k \log k)} O(n)$ ; see Lemma 18.4 of [27]. When  $G$  is planar, we obtain an  $f(k) O(n)$  algorithm where  $f$  is exponential in  $k$  [15]. All these assumptions are stronger than bounded degeneracy, and the techniques cannot be extended easily. A more general class that captures all these cases is that of nowhere-dense graphs

<sup>2</sup> Formally, we should define a tree together with a mapping between its nodes and the subsets of  $V$ . However, the definition adopted here is sufficient for our purposes and lightens the notation.

[28], for which there exist fixed-parameter-tractable subgraph counting algorithms [22]. Nowhere dense graphs however do not include all bounded degeneracy graphs or all graphs with bounded average degree.

Even assuming  $G$  has bounded degeneracy, algorithms faster than Nešetřil-Poljak are known only when  $H$  belongs to special classes. The earliest result of this kind is the classic algorithm by Chiba and Nishizeki [10] to list all  $k$ -cliques in time  $O(d^{k-1}n)$ . Eppstein showed that one can list all maximal cliques in time  $O(d3^{d/3}n)$  [16] and all non-induced complete bipartite subgraphs in time  $O(d^32^{2d}n)$  [14]. These algorithms exploit the degeneracy ordering of  $G$  in a way similar to ours. In fact, our techniques can be seen as a generalization of [10] that takes into account the structure of  $H$ . We note that a fundamental limitation of [10, 14, 16] is that they list all the copies of  $H$ , which for a generic  $H$  might be  $\Theta(n^k)$  even if  $G$  has bounded degeneracy (for example if  $H$  is the independent set). In contrast, we list the copies of subgraphs  $H$ , and combine them to infer the number of copies of  $H$ . To be more precise, we list the homomorphisms of  $H$ , which is another difference we have with [10, 14, 16] and a point we have in common with previous work [11].

Regarding our “dag tree decomposition”, it is inspired by the standard notion of tree decomposition of a graph, and it yields a similar dynamic program. Yet, the similarity between the two decompositions is rather superficial; indeed, our dag-treewidth can be  $O(1)$  when the treewidth is  $\Omega(k)$ , and vice versa. Our decomposition is unrelated to the several notions of tree decomposition for directed graphs already known [19]. Finally, our lower bounds are novel; no general lower bound in terms of  $d$  and of the structure of  $H$  was available before.

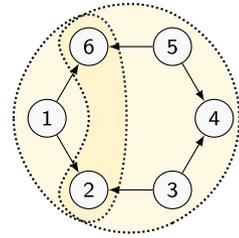
## 1.4 Manuscript organisation

In Sect. 2 we build the intuition with a gentle introduction to our approach. In Sect. 3 we give our dag tree decomposition and the dynamic program for counting homomorphisms. In Sect. 4 we show how to compute good dag tree decompositions. Finally, in Sect. 5 we prove the lower bounds.

## 2 Exploiting degeneracy orientations

We build the intuition behind our approach, starting from the classic algorithm for counting cliques by Chiba and Nishizeki [10]. The algorithm begins by orienting  $G$  acyclically so that  $\max_{v \in G} d_{\text{out}}(v) \leq d$ , which takes time  $O(|E|)$ . With  $G$  oriented acyclically, we take each  $v \in G$  in turn, enumerate every subset of  $(k-1)$  out-neighbors of  $v$ , and check its edges. In this way we can explicitly find all  $k$ -cliques of  $G$  in time  $O(k^2d^{k-1}n)$ . Observe that the crucial fact here is that an acyclically oriented clique has exactly one *source*, that is, a node with no incoming arcs. We would like to extend this approach to an arbitrary pattern  $H$ . Since every copy of  $H$  in  $G$  appears with exactly one acyclic orientation, we take every possible acyclic orientation  $P$  of  $H$ , count the copies of  $P$  in  $G$ , and sum all the counts. Thus, the problem reduces to counting the copies of an arbitrary dag  $P$  in our acyclic orientation of  $G$ .

**Fig. 1** Toy example: an acyclic orientation  $P$  of  $H = C_6$ , decomposed into two pieces



Let us start in the naive way. Suppose  $P$  has  $s$  sources. Fix a directed spanning forest  $F$  of  $P$ . This is a collection of  $s$  directed disjoint trees rooted at the sources of  $P$  (arcs pointing away from the roots). Clearly, each copy of  $P$  in  $G$  contains a copy of  $F$ . Hence, we can enumerate the copies of  $F$  in  $G$ , and for each one check if it is a copy of  $P$ . To this end, first we enumerate the  $O(n^s)$  possible  $s$ -tuples of  $V$  to which the sources of  $P$  can be mapped. For each such  $s$ -uple, we enumerate the possible mappings of the remaining  $k - s$  nodes of the forest. This can be done in time  $O(d^{k-s})$  by a straightforward extension of the out-neighbor listing algorithm above. Finally, for each mapping we check if its nodes induce  $P$  in  $G$ , in time  $O(k^2)$ . The total running time is  $O(k^2 d^{k-s} n^s)$ . Unfortunately, if  $P$  is an independent set then  $s = k$  and the running time is  $O(k^2 n^k)$ , so we have made no progress over the naive algorithm.

At this point we introduce our first idea. For reference we use the toy pattern  $P$  in Fig. 1. Instead of enumerating the copies of  $P$  in  $G$ , we decompose  $P$  into two pieces,  $P(1)$  and  $P(3, 5)$ . Here,  $P(1)$  denotes the subgraph of  $P$  reachable from 1 (that is, the transitive closure of 1 in  $P$ ). The same for  $P(3)$  and  $P(5)$ , and we let  $P(3, 5) = P(3) \cup P(5)$ . Now we count the copies of  $P(1)$ , and then the copies of  $P(3, 5)$ , hoping to combine the result in some way to obtain the count of  $P$ . To simplify the task, we focus on counting homomorphisms rather than copies (see below). Thus, we want to compute  $\text{hom}(P, G)$  by combining  $\text{hom}(P(1), G)$  and  $\text{hom}(P(3, 5), G)$ .

Now, clearly, knowing  $\text{hom}(P(1), G)$  and  $\text{hom}(P(3, 5), G)$  is not sufficient to infer  $\text{hom}(P, G)$ . Thus, we need to solve a slightly more complex problem. For every pair  $x, y \in V(G)$ , let  $\phi : \{2, 6\} \mapsto V(G)$  be the map given by  $\phi(2) = x$  and  $\phi(6) = y$ . We let  $\text{hom}(P, G, (x, y))$  be the number of homomorphisms of  $P$  in  $G$  whose restriction to  $\{2, 6\}$  is  $\phi$ . By a counting argument one can immediately see that:

$$\text{hom}(P, G) = \sum_{\phi: \{2,6\} \rightarrow V(G)} \text{hom}(P, G, \phi) \tag{1}$$

Thus, to compute  $\text{hom}(P, G)$  we only need to compute  $\text{hom}(P, G, \phi)$  for all possible  $\phi$ . Now, define  $\text{hom}(P(1), G, \phi)$  and  $\text{hom}(P(3, 5), G, \phi)$  with the same meaning as above. A crucial observation is that  $\{2, 6\}$ , the domain of  $\phi$ , is precisely the set of nodes in  $P(1) \cap P(3, 5)$ . It is not difficult to see that this implies:

$$\text{hom}(P, G, \phi) = \text{hom}(P(1), G, \phi) \cdot \text{hom}(P(3, 5), G, \phi) \tag{2}$$

Thus, now our goal is to compute  $\text{hom}(P(1), G, \phi)$  and  $\text{hom}(P(3, 5), G, \phi)$  for all  $\phi : \{2, 6\} \rightarrow V(G)$ . To this end, we list all  $\phi_{P(1)} : P(1) \rightarrow G$  with

the technique above, and for each such  $\phi_{P(1)}$  we increment a counter associated to  $(\phi_{P(1)}(2), \phi_{P(1)}(6))$  in a dictionary with default value 0. Thus, we obtain  $\text{hom}(P(1), G, \phi)$  for all  $\phi : \{2, 6\} \rightarrow V$ . Since  $P(1)$  has one source, we enumerate  $O(k^2 d^{k-1} n)$  maps. If the dictionary takes time  $O(\log n)$  to access an entry, the total running time is  $O(k^2 d^{k-1} n \log n)$ . The same technique applied to  $P(3, 5)$  yields a running time of  $O(k^2 d^{k-2} n^2 \log n)$ , since  $P(3, 5)$  has two sources. Finally, we apply Eq. 1 by running over all entries in the first dictionary and retrieving the corresponding value from the second dictionary. The total running time is  $O(k^2 d^{k-2} n^2 \log n)$ , while enumerating the homomorphisms of  $P$  would have required time  $O(k^2 d^{k-3} n^3)$ .

Let us abstract the general approach from this toy example. We want to decompose  $P$  into a set of pieces  $P_1, P_2, \dots$  with the following properties: (i) Each piece  $P_i$  has a small number of sources  $s(P_i)$ , and (ii) we can obtain  $\text{hom}(P, G, \phi)$  by combining the homomorphism counts of the  $P_i$ . This is achieved by the *dag tree decomposition*, which we introduce in Sect. 3. Like the tree decomposition for undirected graphs, the dag tree decomposition leads to a dynamic program to compute  $\text{hom}(P, G)$ .

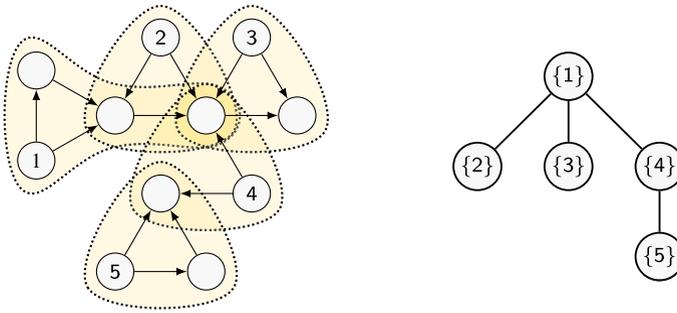
### 3 DAG tree decompositions

Let  $P = (V_P, A_P)$  be a directed acyclic graph. We denote by  $S_P$ , or simply  $S$ , the set of nodes of  $P$  having no incoming arc. These are the *sources* of  $P$ . We denote by  $V_P(u)$  the transitive closure of  $u$  in  $P$ , i.e. the set of nodes of  $P$  reachable from  $u$ , and we let  $P(u) = P[V_P(u)]$  be the corresponding subgraph of  $P$ . For a subset of sources  $B \subseteq S$  we let  $V_P(B) = \cup_{u \in B} V_P(u)$  and  $P(B) = P[V_P(B)]$ . Thus,  $P(B)$  is the subgraph of  $P$  induced by all nodes reachable from  $B$ . We call  $B$  a *bag* of sources. We can now formally introduce our decomposition.

**Definition 2** (*Dag tree decomposition*) Let  $P = (V_P, A_P)$  be a dag. A dag tree decomposition (d.t.d.) of  $P$  is a (rooted) tree  $T = (\mathcal{B}, \mathcal{E})$  with the following properties:

1. each node  $B \in \mathcal{B}$  is a bag of sources  $B \subseteq S_P$
2.  $\cup_{B \in \mathcal{B}} B = S_P$
3. for all  $B, B_1, B_2 \in T$ , if  $B \in T(B_1, B_2)$  then  $V_P(B_1) \cap V_P(B_2) \subseteq V_P(B)$

One can see the similarity with the tree decomposition of an undirected graph (Definition 1). However, our dag tree decomposition differs crucially in two aspects. First, the bags are subsets of  $S$  rather than subsets of  $V_P$ . This is because the time needed to list the homomorphisms between  $P(B_i)$  and  $G$  is driven by  $n^{|B_i|}$ . Second, the path-intersection property (3) concerns the pieces reachable from the bags rather than the bags themselves. The reason is that, to combine the counts of two pieces together, their intersection must form a separator in  $P$  (similarly to the tree decomposition of an undirected graph). The dag tree decomposition induces the following notions of *width*, used throughout the rest of the article.



**Fig. 2** Left: a dag  $P$  formed by five pieces. Right: a dag tree decomposition  $T$  for  $P$ . Since  $\tau(T) = 1$  and the largest piece contains 4 nodes, we can compute  $\text{hom}(P, G)$  in time  $O(d^3 n \log n)$

**Definition 3** The *width* of  $T$  is  $\tau(T) = \max_{B \in \mathcal{B}} |B|$ . The *dag treewidth*  $\tau(P)$  of  $P$  is the minimum of  $\tau(T)$  over all dag tree decompositions  $T$  of  $P$ .

Clearly  $\tau(P) \in \{1, \dots, k\}$  for any  $k$ -node dag  $P$ . Figure 2 shows a pattern  $P$  together with a d.t.d. of width 1. We observe that  $\tau(P)$  has no obvious relation to the treewidth  $t(H)$  of  $H$ ; see the discussion in Sect. 3.2.

### 3.1 Counting homomorphisms via dag tree decompositions

For any  $B \in \mathcal{B}$  let  $T(B)$  be the subtree of  $T$  rooted at  $B$ . We let  $\Gamma[B]$  be the down-closure of  $B$  in  $T$ , that is, the union of all bags in  $T(B)$ . Consider  $P(\Gamma[B])$ , the subgraph of  $P$  induced by the nodes reachable from  $\Gamma[B]$  (note the difference with  $P(B)$ , which contains only the nodes reachable from  $B$ ). We compute  $\text{hom}(P(\Gamma[B]), G)$  in a bottom-up fashion over all  $B$ , starting with the leaves of  $T$  and moving towards the root. This is similar to the dynamic program given by the standard tree decomposition (see [18]).

As anticipated, we actually compute  $\text{hom}(P(\Gamma[B]), \phi)$ , the number of homomorphisms that extend a fixed mapping  $\phi$ . We need the following concept:

**Definition 4** Let  $P_1 = (V_{P_1}, A_{P_1}), P_2 = (V_{P_2}, A_{P_2})$  be two subgraphs of  $P$ , and let  $\phi_1 : P_1 \rightarrow G$  and  $\phi_2 : P_2 \rightarrow G$  be two homomorphisms. We say  $\phi_1$  and  $\phi_2$  *respect* each other if  $\phi_1(u) = \phi_2(u)$  for all  $u \in V_{P_1} \cap V_{P_2}$ .

Given some  $\phi$ , we denote by  $\text{hom}(P_1, G, \phi_2)$  the number of homomorphisms from  $P_1$  to  $G$  that respect  $\phi_2$ . We can now present our main algorithmic result.

**Theorem 7** Let  $P$  be any  $k$ -node dag, and  $T = (\mathcal{B}, \mathcal{E})$  be a d.t.d. for  $P$ . Fix any  $B \in \mathcal{B}$  as the root of  $T$ . There is a dynamic programming algorithm  $\text{HomCount}(P, T, B)$  that in time  $O(|\mathcal{B}|k^2 d^{k-\tau(T)} n^{\tau(T)} \log n)$  computes  $\text{hom}(P(\Gamma[B]), G, \phi_B)$  for all  $\phi_B : P(B) \rightarrow G$ . This is also a bound on the time needed to compute  $\text{hom}(P, G)$ .

The proof of Theorem 7 is given in the next subsection. Before continuing, let  $f_T(k)$  be an upper bound on the time needed to compute a d.t.d. of minimum width with at most  $2^k$  bags for a pattern on  $k$  nodes. We can show that such a d.t.d. always exists:

**Lemma 1** *Any  $k$ -node dag  $P$  has a minimum-width d.t.d. on at most  $2^k$  bags.*

**Proof** We show that, if a d.t.d.  $T = (\mathcal{B}, \mathcal{E})$  has two bags containing exactly the same sources, then one of the two bags can be removed. This implies that there exists a minimum-width d.t.d. where every bag contains a distinct source set, which therefore has at most  $2^k$  bags. Suppose indeed  $T$  contains two bags  $X$  and  $X'$  formed by the same subset of sources. Let  $B$  be the neighbor of  $X$  on the unique path  $T(X, X')$ . Let  $T^* = (\mathcal{B}^*, \mathcal{E}^*)$  be the tree obtained from  $T$  by replacing the edge  $\{B', X\}$  with  $\{B', B\}$  for every neighbor  $B' \neq B$  of  $X$  and then deleting  $X$ . Clearly  $|\mathcal{B}^*| = |\mathcal{B}| - 1$ , and properties (1) and (2) of Definition 2 are satisfied. Let us then check property (3). Consider a generic path  $T^*(B_1, B_2)$  and look at the corresponding path  $T(B_1, B_2)$ . If  $T(B_1, B_2)$  does not contain edges that we deleted, then  $T^*(B_1, B_2) = T(B_1, B_2)$ . In this case the property holds for any bag in  $T^*(B_1, B_2)$  since it holds in  $T$ . Suppose instead  $T(B_1, B_2)$  contains edges that we deleted. Then  $T^*(B_1, B_2)$  contains the same bags of  $T(B_1, B_2)$  save that  $X$  is replaced by  $B$ . Thus we only need to check that  $V_P(B_1) \cap V_P(B_2) \subseteq V_P(B)$ . By property (3),  $V_P(B_1) \cap V_P(B_2) \subseteq V_P(X)$ . Moreover, since by construction  $B \in T(X, X')$ , property (3) also gives  $V_P(X) = V_P(X) \cap V_P(X') \subseteq V_P(B)$ . Thus  $V_P(B_1) \cap V_P(B_2) \subseteq V_P(B)$ . Therefore  $T^*$  is a d.t.d. for  $P$ .  $\square$

Then, as an immediate corollary of Theorem 7, we have:

**Theorem 8** *Let  $P$  be any  $k$ -node dag, and  $T = (\mathcal{B}, \mathcal{E})$  be a d.t.d. for  $P$ . We can compute  $\text{hom}(P, G)$  in time  $f_T(k) + O(k^2 2^k d^{k-\tau(P)} n^{\tau(P)} \log n)$ .*

Theorem 8 will be used in Sect. 3.2 to prove the bounds for our original problem of counting the copies of  $H$  via inclusion-exclusion arguments.

### 3.1.1 Proof of Theorem 7

The algorithm behind Theorem 7 is similar to the one for counting homomorphisms using a tree decomposition. To start, we prove that our dag tree decomposition enjoys a separator property similar to the one enjoyed by tree decompositions.

**Lemma 2** *Let  $T$  be a rooted d.t.d. and let  $B_1, \dots, B_l$  be the children of  $B$  in  $T$ . Then for all  $i \in [l]$ :*

- a.  $V_P(\Gamma[B_i]) \cap V_P(\Gamma[B_j]) \subseteq V_P(B)$  for all  $j \neq i$
- b. for any arc  $(u, u') \in P(\Gamma[B])$ , if  $u \in V_P(\Gamma[B_i]) \setminus V_P(B)$  then  $u' \in V_P(\Gamma[B_i])$
- c. for any arc  $(u', u) \in P(\Gamma[B])$ , if  $u \in V_P(\Gamma[B_i]) \setminus V_P(B)$  then  $u' \in V_P(\Gamma[B_i])$

**Proof** We prove (a). Suppose for some  $i \neq j$  we have  $V_P(\Gamma[B_i]) \cap V_P(\Gamma[B_j]) \not\subseteq V_P(B)$ . So there exists some node  $u \in V_P$  such that  $u \in V_P(\Gamma[B_i])$ ,  $u \in V_P(\Gamma[B_j])$ , and  $u \notin V_P(B)$ . By definition of  $\Gamma[\cdot]$ , this implies  $u \in V_P(B'_i)$  and  $u \in V_P(B'_j)$  for some bags  $B'_i \in T(B_i)$  and  $B'_j \in T(B_j)$ . Observe however that  $B \in T(B'_i, B'_j)$ . Thus, by point (3) of Definition 2, we have  $u \in V_P(B)$ . This contradicts the third inclusion,  $u \notin V_P(B)$ .

Now we prove (b) and (c). For (b), since  $u \in V_P(\Gamma[B_i])$  and  $(u, u') \in P$ , then  $u' \in V_P(\Gamma[B_i])$  too. For (c), suppose by contradiction  $u' \notin V_P(\Gamma[B_i])$ . Therefore, either  $u' \in V_P(B)$ , or  $u' \in V_P(B'_j)$  for some  $B'_j \in \Gamma[B_j]$  with  $j \neq i$ . In both cases however we have  $u \in V_P(B)$ : in the first case this holds since  $u$  is reachable from  $u'$ , and in the second case since  $B \in T(B_i, B'_j)$  and by point (3) of Definition 2. Thus in any case  $u \in V_P(B)$ , which contradicts again  $u \notin V_P(B)$ .  $\square$

Lemma 2 says that  $V_P(B)$  is a separator for the sub-patterns  $P(\Gamma[B_i])$  in  $P$ . This allows us to compute  $\text{hom}(P(\Gamma[B]))$  by combining  $\text{hom}(P(\Gamma[B_1])), \dots, \text{hom}(P(\Gamma[B_l]))$ .

Next, we show that each homomorphism  $\phi$  of  $P(\Gamma[B])$  is the *juxtaposition* (definition below) of some homomorphism  $\phi_B$  of  $B$  and some homomorphisms  $\phi_1, \dots, \phi_l$  of  $\Gamma[B_1], \dots, \Gamma[B_l]$ , provided they respect  $\phi_B$ . This establishes a bijection, implying that we can count the homomorphisms  $\phi$  by multiplying the counts of the homomorphisms  $\phi_B, \phi_1, \dots, \phi_l$ .

**Definition 5** Let  $\{\phi_1, \dots, \phi_\ell\}$  be any set of homomorphisms, where for all  $i = 1, \dots, \ell$  we have  $\phi_i : X_i \rightarrow G$  and  $\phi_i$  respects  $\phi_j$  for all  $j = 1, \dots, \ell$ . The *juxtaposition* of  $\phi_1, \dots, \phi_\ell$ , denoted by  $\phi_1 \dots \phi_\ell$ , is the homomorphism  $\phi : \cup_{i=1}^\ell X_i \rightarrow G$  such that  $\phi(u) = \phi_i(u)$  whenever  $u \in X_i$ .

Note that the juxtaposition is always well-defined and unique, since the  $\phi_i$  respect each other and the image of every  $u$  is determined by at least one among  $\phi_1, \dots, \phi_\ell$ .

**Lemma 3** Let  $T$  be a d.t.d. and let  $B_1, \dots, B_l$  be the children of  $B$  in  $T$ . Fix any  $\phi_B : P(B) \rightarrow G$ . Let  $\Phi(\phi_B) = \{\phi : P(\Gamma[B]) \rightarrow G \mid \phi \text{ respects } \phi_B\}$ , and for  $i = 1, \dots, l$  let  $\Phi_i(\phi_B) = \{\phi : P(\Gamma[B_i]) \rightarrow G \mid \phi \text{ respects } \phi_B\}$ . Then there is a bijection between  $\Phi(\phi_B)$  and  $\Phi_1(\phi_B) \times \dots \times \Phi_l(\phi_B)$ , and therefore:

$$\text{hom}(P(\Gamma[B]), G, \phi_B) = \prod_{i=1}^l \text{hom}(P(\Gamma[B_i]), G, \phi_B) \tag{3}$$

**Proof** First, we show there is an injection between  $\Phi(\phi_B)$  and  $\Phi_1(\phi_B) \times \dots \times \Phi_l(\phi_B)$ . Fix any  $\phi \in \Phi(\phi_B)$ , and consider the tuple  $(\phi_1, \dots, \phi_l)$  where each  $\phi_i$  is the restriction of  $\phi$  to  $P(\Gamma[B_i])$ . Note that  $\phi_i$  is unique, and that it respects  $\phi_B$  since  $\phi$  does. Thus  $\phi_i \in \Phi_i(\phi_B)$ . It follows that  $(\phi_1, \dots, \phi_l) \in \Phi_1(\phi_B) \times \dots \times \Phi_l(\phi_B)$ . Now we show there is an injection between  $\Phi_1(\phi_B) \times \dots \times \Phi_l(\phi_B)$  and  $\Phi(\phi_B)$ . Consider any tuple  $(\phi_1, \dots, \phi_l) \in \Phi_1(\phi_B) \times \dots \times \Phi_l(\phi_B)$ , and consider the juxtaposition

$\phi = \phi_B \phi_1 \dots \phi_l$ . Then  $\phi : P(\Gamma[B]) \rightarrow G$  and  $\phi$  respects  $\phi_B$ . It follows that  $\phi \in \Phi(\phi_B)$ . □

Last, we bound the cost of enumerating the homomorphisms of a piece of  $P$ .

**Lemma 4** *Given any  $B \subseteq S$ , the set of homomorphisms  $\Phi = \{\phi : P(B) \rightarrow G\}$  has size  $O(d^{k-|B|}n^{|B|})$  and can be enumerated in time  $O(k^2 d^{k-|B|}n^{|B|})$ .*

**Proof** We prove the bound on the enumeration time; the proof gives immediately also the bound on  $|\Phi|$ . Let  $B = \{u_1, \dots, u_b\}$  where  $b = |B|$ . Fix a spanning forest  $\{T_1, \dots, T_b\}$  of  $P(B)$ , where each  $T_i = (V_i, A_i)$  is a directed tree rooted at  $u_i$  (arcs pointing away from the root). Consider any  $\phi \in \Phi$ , and let  $\phi_i$  be its restriction to  $V_i$ . Clearly,  $\phi = \phi_1 \dots \phi_b$ . Note that  $\phi_i$  is a homomorphism of  $T_i$  in  $G$ . Thus, to enumerate  $\Phi$  we can enumerate each possible tuples  $(\phi_1, \dots, \phi_b)$  where  $\phi_i$  is a homomorphism of  $T_i$  in  $G$  for all  $i$ . Note that not all such tuples give a valid juxtaposition that is a homomorphism  $\phi \in \Phi$ . However, we can check if  $\phi \in \Phi$  in time  $O(k^2)$  by checking the arcs between the vertices of  $G$  in the image of  $\phi$ .

Let then  $\Phi_{T_i}$  be the set of homomorphisms of  $T_i$  in  $G$ . We show how to enumerate  $\Phi_{T_i}$  in time  $O(d^{|V_i|-1}n)$ , and thus all tuples  $(\phi_1, \dots, \phi_b) \in \Phi_{T_1} \times \dots \times \Phi_{T_b}$  in time  $\prod_{i=1}^b O(d^{|V_i|-1}n) = O(d^{k-b}n^b)$ . Together with the check on the arcs, this gives a total running time of  $O(k^2 d^{k-b}n^b)$  for enumerating  $\Phi$ , as desired. To enumerate  $\Phi_{T_i}$ , we take each  $v \in G$  and enumerate all  $\phi_i \in \Phi_{T_i}$  such that  $\phi_i(s_i) = v$ . To this end note that, once we have fixed  $\phi_i(x)$ , for each arc  $(x, y) \in T_i$  we have at most  $d$  choices for  $\phi_i(y)$ . Thus we can enumerate all  $\phi_i \in \Phi_{T_i}$  that map  $s_i$  to  $v$  in time  $d^{|V_i|-1}$ . The total time to enumerate  $\Phi_{T_i}$  is therefore  $O(d^{|V_i|-1}n)$ , as claimed. □

We can now describe our dynamic programming algorithm, HomCount, to compute  $\text{hom}(P(\Gamma[B]), G)$ . Given a d.t.d.  $T$  of  $P$ , the algorithm goes bottom-up from the leaves of  $T$  towards the root, combining the counts using Lemma 3. For readability, we write the algorithm in a recursive fashion. We prove:

**Algorithm 1** HomCount( $P, T, B$ )

```

1:  $C_B =$  empty dictionary with default value 0
2: if  $B$  is a leaf then ▷ base case
3:   for every homomorphism  $\phi_B : P(B) \rightarrow G$  do
4:      $C_B(\phi_B) = 1$ 
5: else
6:   let  $B_1, \dots, B_l$  be the children of  $B$  in  $T$ 
7:   for  $i = 1, \dots, l$  do
8:      $C_{B_i} =$  HomCount( $P, T, B_i$ ) ▷ recurse on the subtree of  $T$  rooted at  $B_i$ 
9:      $AGG_{B_i} =$  empty dictionary with default value 0
10:    for every key  $\phi$  in  $C_{B_i}$  do ▷ we aggregate the values of  $C$ 
11:      let  $\phi_r$  be the restriction of  $\phi$  to  $V_P(B) \cap V_P(\Gamma[B_i])$ 
12:       $AGG_{B_i}(\phi_r) += C_{B_i}(\phi)$ 
13:   for every homomorphism  $\phi : P(B) \rightarrow G$  do ▷ Lemma 3 in action
14:     for  $i = 1, \dots, l$  let  $\phi_i$  be the restriction of  $\phi$  to  $V_P(B) \cap V_P(\Gamma[B_i])$ 
15:      $C_B(\phi) = \prod_{i=1}^l AGG_{B_i}(\phi_i)$ 
16: return  $C_B$ 

```

**Lemma 5** Let  $P$  be any dag,  $T = (\mathcal{B}, \mathcal{E})$  any d.t.d. for  $P$ , and  $B$  any element of  $\mathcal{B}$ . HomCount( $P, T, B$ ) in time  $O(|\mathcal{B}| \text{poly}(k)d^{k-\tau(T)}n^{\tau(T)} \log n)$  returns a dictionary  $C_B$  that for all  $\phi_B : P(B) \rightarrow G$  satisfies  $C_B(\phi_B) = \text{hom}(P(\Gamma[B]), G, \phi_B)$ .

**Proof** We first prove the correctness, by induction on the nodes of  $T$ . The base case is when  $B$  is a leaf of  $T$ . In this case  $P(B) = P(\Gamma[B])$ , and the algorithm sets  $C_B(\phi_B) = 1$  for each  $\phi_B : P(B) \rightarrow G$ . Therefore  $C_B(\phi_B) = \text{hom}(P(\Gamma[B]), G, \phi_B)$  as desired. The inductive case is when  $B$  is an internal node of  $T$ . As inductive hypothesis we assume that, for every child  $B_i$  of  $B$ , the dictionary  $C_{B_i}$  computed at line 8 satisfies  $C_{B_i}(\phi) = \text{hom}(P(\Gamma[B_i]), G, \phi)$  for every  $\phi : P(B_i) \rightarrow G$ . Let  $\Phi_{P(\Gamma[B_i])}$  be the set of homomorphisms from  $P(\Gamma[B_i])$  to  $G$ , and let  $\Phi_{P(\Gamma[B_i])}(\phi)$  be the subset of elements of  $\Phi_{P(\Gamma[B_i])}$  that respect  $\phi$ . Thus, the inductive hypothesis says that  $C_{B_i}(\phi) = |\Phi_{P(\Gamma[B_i])}(\phi)|$  for every  $\phi : P(B_i) \rightarrow G$ .

Now consider the loop at lines 10–12. We claim that, after that loop, we have:

$$\begin{aligned}
 AGG_{B_i}(\phi_r) &= \sum_{\substack{\phi \text{ in } C_{B_i} \\ \phi \text{ respects } \phi_r}} C_{B_i}(\phi) = \sum_{\substack{\phi : P(B_i) \rightarrow G \\ \phi \text{ respects } \phi_r}} C_{B_i}(\phi) \\
 &= \sum_{\substack{\phi : P(B_i) \rightarrow G \\ \phi \text{ respects } \phi_r}} |\Phi_{P(\Gamma[B_i])}(\phi)| = |\Phi_{P(\Gamma[B_i])}(\phi_r)|
 \end{aligned} \tag{4}$$

The first equality holds since the loop adds  $C_{B_i}(\phi)$  to  $AGG_{B_i}(\phi_r)$  if and only if the restriction of  $\phi$  to  $V_P(B) \cap V_P(\Gamma[B_i])$  is  $\phi_r$ , that is, if and only if  $\phi$  respects  $\phi_r$ . The second equality holds since the keys of  $C_{B_i}$  are a subset of  $\{\phi : P(B_i) \rightarrow G\}$  and  $C_{B_i}(\phi) = 0$  if  $\phi$  is not in  $C_{B_i}$ . The third equality holds by the inductive hypothesis above. The fourth equality holds since the sets  $\Phi_{P(\Gamma[B_i])}(\phi)$  form a partition of  $\Phi_{P(\Gamma[B_i])}(\phi_r)$ .

Finally, consider the loop at lines 13–15. We claim that line 15 sets:

$$C_B(\phi) = \prod_{i=1}^l \left| \Phi_{P(\Gamma[B_i])}(\phi_i) \right| = \prod_{i=1}^l \left| \Phi_{P(\Gamma[B_i])}(\phi) \right| = \text{hom}(P(\Gamma[B]), G, \phi) \quad (5)$$

The first equality holds by coupling line 15 and Eq. (4). The second equality holds since any element of  $\Phi_{P(\Gamma[B_i])}$  respects  $\phi$  if and only if it respects its restriction  $\phi_i$  to  $V_P(B) \cap V_P(\Gamma[B_i])$ , thus  $\Phi_{P(\Gamma[B_i])}(\phi_i) = \Phi_{P(\Gamma[B_i])}(\phi)$ . The last equality holds by definition of  $\Phi_{P(\Gamma[B_i])}(\phi_i)$  and by Lemma 3. This proves the correctness.

Let us turn to the running time. For each dictionary  $C$  we let  $|C|$  be the number of distinct keys in  $C$ . Recall that reading or writing an entry in our dictionary takes time  $O(\text{poly}(k) \log n)$ . We split the running time as follows:

- (i) the cost of the base case (lines 2–4). Since the loop has  $|C_B|$  iterations, and each one costs  $O(\text{poly}(k) \log n)$ , this cost is in  $O(\text{poly}(k)|C_B| \log n)$ .
- (ii) the cost of the iteration at lines 7–12, performed by the parent of  $B$  in  $T$ , where the considered child is  $B$ , excluding obviously the cost of the recursive call at line 8. Similarly to (i), this cost is bounded by  $O(\text{poly}(k)|C_B| \log n)$ .
- (iii) the same as (ii), but for the loop at lines 13–15. This cost is in  $O(\text{poly}(k)|C_{B'}| \log n)$  where  $B'$  is the parent of  $B$  in  $T$ .

We charge  $B$  with every cost among (i), (ii), (iii) that applies (this depends on whether  $B$  is the root, a leaf, or an internal node of  $T$ ). It is easy to check that the sum of all charged costs accounts for the total running time of  $\text{HomCount}(P, T, B)$  when  $B$  is the root of  $T$ . Now, by Lemma 4, for every  $B \in \mathcal{B}$  we have  $|C_B| = O(d^{k-|B|}n^{|B|})$ , which is in  $O(d^{k-\tau(T)}n^{\tau(T)})$  by definition of  $\tau(T)$ . Thus each one of the three costs above is in  $O(\text{poly}(k)d^{k-\tau(T)}n^{\tau(T)} \log n)$ . Summing over  $B \in \mathcal{B}$  yields the claimed bound. □

### 3.2 Inclusion–exclusion arguments and the dag-treewidth of undirected graphs

We turn to computing  $\text{hom}(H, G)$ ,  $\text{sub}(H, G)$  and  $\text{ind}(H, G)$ . We do so via standard inclusion-exclusion arguments, using our algorithm for computing  $\text{hom}(P, G)$  as a primitive. To this end we shall define appropriate notions of width for undirected pattern graphs. Let  $\Sigma(H)$  be the set of all dags  $P$  that can be obtained by orienting  $H$  acyclically. Let  $\Theta(H)$  be the set of all equivalence relations on  $V_H$  (that is, all the partitions of  $V_H$ ), and for  $\theta \in \Theta(H)$  let  $H/\theta$  be the pattern obtained from  $H$  by identifying equivalent nodes according to  $\theta$  and removing loops and multiple edges. Let  $D(H)$  be the set of all supergraphs of  $H$  on the node set  $V_H$ , including  $H$ .

**Definition 6** The dag treewidth of  $H$  is  $\tau(H) = \tau_3(H)$ , where:

$$\tau_1(H) = \max\{\tau(P) : P \in \Sigma(H)\} \quad (6)$$

$$\tau_2(H) = \max\{\tau_1(H/\theta) : \theta \in \Theta(H)\} \tag{7}$$

$$\tau_3(H) = \max\{\tau_2(H') : H' \in D(H)\} \tag{8}$$

Note that  $\tau(H)$  is unrelated to the treewidth  $t(H)$ . For example, when  $H$  is a clique we have  $t(H) = k$  and  $\tau(H) = 1$ ; when  $H$  is the independent set we have  $t(H) = 1$  and  $\tau(H) = \Theta(k)$ , see Lemma 14; and when  $H$  is an expander we have  $t(H), \tau(H) \in \Theta(k)$ , see again Lemma 14. In fact,  $\tau(H)$  is within constant factors of the independence number  $\alpha(H)$  of  $H$  (see Sect. 4.4), and thus decreases as  $H$  becomes denser. This happens because adding arcs increases the number of nodes reachable from the sources of  $P \in \Sigma(H)$ , so we may need fewer sources to reach a given piece of  $P$ . When  $H$  is a clique,  $P$  is reachable from just one source and thus  $\tau(H) = 1$ .

Clearly,  $\tau_1(H) \leq \tau_2(H) \leq \tau(H)$ . The intuition behind  $\tau_1(H)$  is that, in  $G$ , each homomorphism of  $H$  corresponds to a homomorphism of some acyclic orientation  $P$  of  $H$ . Thus to compute  $\text{hom}(H, G)$  we sum  $\text{hom}(P, G)$  over all orientation  $P$  of  $H$ , and the running time is dominated by the  $P$  with largest treewidth. The intuition behind  $\tau_2(H)$  is similar but now we look at computing  $\text{sub}(H, G)$ . Since homomorphisms can map different nodes of  $H$  to the same node of  $G$ , to recover  $\text{sub}(H, G)$  we must combine  $\text{hom}(H', G)$  for all possible  $H' = H/\theta$  through inclusion-exclusion arguments. The intuition behind  $\tau_3(H)$  is that to compute  $\text{ind}(H, G)$  we must remove from  $\text{sub}(H, G)$  the counts of  $\text{sub}(H', G)$  for certain supergraphs  $H'$  of  $H$ . Indeed, the three measures  $\tau_1(H), \tau_2(H), \tau(H)$  yield:

**Theorem 9** Consider any  $k$ -node pattern graph  $H = (V_H, E_H)$ , and let  $f_T(k)$  be an upper bound on the time needed to compute a d.t.d. of minimum width on  $2^{O(k \log k)}$  bags for any  $k$ -node dag. Then one can compute:

- $\text{hom}(H, G)$  in time  $2^{O(k \log k)} \cdot O(f_T(k) + d^{k-\tau_1(H)} n^{\tau_1(H)} \log n)$ ,
- $\text{sub}(H, G)$  in time  $2^{O(k \log k)} \cdot O(f_T(k) + d^{k-\tau_2(H)} n^{\tau_2(H)} \log n)$ ,
- $\text{ind}(H, G)$  in time  $2^{O(k^2)} \cdot O(f_T(k) + d^{k-\tau(H)} n^{\tau(H)} \log n)$ .

The claim still holds if we replace  $\tau_1, \tau_2, \tau$  with upper bounds, and  $f_T(k)$  with the time needed to compute a d.t.d. on  $2^{O(k \log k)}$  bags that satisfies those upper bounds.

**Proof** We prove the three bounds in three separated steps. The last claim follows straightforwardly.

*From dags to undirected patterns.* Let  $H$  be any undirected pattern. First, note that:

$$\text{hom}(H, G) = \sum_{P \in \Sigma(H)} \text{hom}(P, G) \tag{9}$$

Let indeed  $\Phi(H) = \{\phi : H \rightarrow G\}$  be the set of homomorphisms from  $H$  to  $G$ . Similarly, for any  $P \in \Sigma(H)$  define  $\Phi(P) = \{\phi_P : P \rightarrow G\}$  (note that  $\phi_P$  must preserve the direction of the arcs). Then, there is a bijection between  $\Phi(H)$  and  $\cup_{P \in \Sigma(H)} \Phi(P)$ . Consider indeed any  $\phi \in \Phi(H)$ . Let  $\sigma$  be the orientation of  $H$  that assigns to  $\{u, v\} \in E_H$  the orientation of  $\{\phi(u), \phi(v)\}$  in  $G$ , and let  $P = H_\sigma$ . Then  $\phi$  is a homomorphism of

$P$  in  $G$ . On the other hand consider any homomorphism  $\phi \in \Phi(P)$  for some acyclic orientation  $P$  of  $H$ . By ignoring the orientation of the edges,  $\phi \in \Phi(H)$ , too.

Thus, to compute  $\text{hom}(H, G)$  we compute  $\text{hom}(P, G)$  for all  $P \in \Sigma(H)$  and apply Eq. 9. Clearly, enumerating  $|\Sigma(H)|$  takes time  $O(k!) = 2^{O(k \log k)}$ . For each  $P$ , by Lemma 1 in time  $f_T(k)$  we compute a d.t.d.  $T = (\mathcal{B}, \mathcal{E})$  of width  $\tau(P)$  such that  $|\mathcal{B}| \leq 2^k$ . Then, by Lemma 5 we compute  $\text{hom}(P, G)$  in time  $O(2^k \text{poly}(k) d^{k-\tau(P)} n^{\tau(P)} \log n)$ . Thus, we can compute every  $\text{hom}(P, G)$  in time  $O(f_T(k) + 2^{O(k)} d^{k-\tau(P)} n^{\tau(P)} \log n)$ . Multiplying by  $2^{O(k \log k)}$  gives the first bound of the theorem.

*From homomorphisms to non-induced copies* Recall that  $H/\theta$  is the graph obtained from  $H$  by identifying the nodes in the same equivalence class and removing loops and multiple edges, where  $\theta \in \Theta(H)$  is an equivalence relation (or partition) over  $V_H$ . Then, by Equation 15 of [3]:

$$\text{inj}(H, G) = \sum_{\theta \in \Theta(H)} \mu(\theta) \text{hom}(H/\theta, G) \tag{10}$$

where  $\mu(\theta) = \prod_{A \in \theta} (-1)^{|A|-1} (|A| - 1)!$ , where  $A$  runs over the equivalence classes (the sets) in  $\theta$ . Thus, to compute  $\text{inj}(H, G)$ , we enumerate all  $\theta \in \Theta(H)$ , compute  $\text{hom}(H/\theta, G)$ , and apply Eq. 10. It is known that  $|\Theta| = 2^{O(k \ln k)}$  (see e.g. [2]), and clearly for each  $\theta$  we can compute  $\mu(\theta)$  and  $H/\theta$  in  $O(\text{poly}(k))$ . Thus, the first bound of the theorem holds for computing  $\text{inj}(H, G)$  too. Finally, we compute  $\text{sub}(H, G) = \frac{\text{inj}(H, G)}{\text{aut}(H)}$ , where  $\text{aut}(H)$  is the number of automorphisms of  $H$ , which can be computed in time  $2^{O(\sqrt{k \ln k})}$  [26]. This proves the second bound of the theorem.

*From non-induced to induced.* Finally, let  $D(H)$  be the set of all supergraphs of  $H$  on the same node set. Then from Equation 14 of [3]:

$$\text{ind}(H, G) = \sum_{H' \in D(H)} (-1)^{|E_{H'} \setminus E_H|} \text{inj}(H', G) \tag{11}$$

To compute  $\text{ind}(H, G)$ , we take every  $H' \in D(H)$ , compute  $\text{inj}(H', G)$ , and apply Eq. 11. Since  $|D(H)| \leq 2^{k^2}$ , the third bound of the theorem follows.  $\square$

The algorithmic part of our work is complete. We shall now focus on computing good dag tree decompositions, so to instantiate Theorem 9 and obtain the upper bounds of Sect. 1.1.

### 4 Computing good dag tree decompositions

In this section we show how to compute dag tree decompositions of low width. First, we show that for every  $k$ -node dag  $P$  we can compute in time  $2^{O(k)}$  a dag tree decomposition  $T$  that satisfies  $\tau(T) \leq \lfloor \frac{k}{4} \rfloor + 2$ . This result requires a nontrivial proof. As a corollary, we prove Theorems 1 and 2. Second, we give improved bounds for cliques minus  $\epsilon$  edges; as a corollary, we prove Theorem 3. Third, we give improved bounds for complete multipartite graphs plus  $\epsilon$  edges; as a corollary, we prove Theorem 4. Finally, we show that  $\Omega(\alpha(H)) \leq \tau(H) \leq \alpha(H)$  where  $\alpha(H)$  is the independence

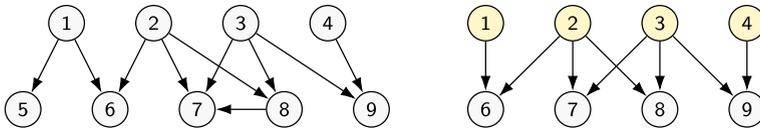


Fig. 3 Left: a dag  $P$ . Right: its skeleton  $\Lambda(P)$  (sources  $S$  above, joints  $J$  below)

number of  $H$ , which is of independent interest. This implies that the trivial decomposition on one bag has width that is asymptotically optimal, since in any orientation of  $H$ , the set of sources is an independent set.

To proceed, we need some additional notation. For a dag  $P$ , we say  $v \in V_P$  is a *joint* if it is reachable from at least two sources, i.e., if  $v \in V_P(u) \cap V_P(u')$  for some  $u, u' \in S$  with  $u \neq u'$ . Let  $J$  be the set of joints of  $P$ . We write  $J(u)$  for the set of joints reachable from  $u$ , and for any  $X \subseteq V_P$  we let  $J(X) = \cup_{u \in X} J(u)$ . Similarly, we denote by  $S(y)$  the sources from which  $y$  is reachable, and we let  $S(X) = \cup_{u \in X} S(u)$ .

### 4.1 A bound for all patterns

This subsection is devoted to prove:

**Theorem 10** *For any dag  $P = (V_P, A_P)$ , in time  $O(1.7549^k)$  we can compute a dag tree decomposition  $T = (\mathcal{B}, \mathcal{E})$  with  $\tau(T) \leq \min(\lfloor \frac{e}{4} \rfloor, \lfloor \frac{k}{4} \rfloor) + 2$  and  $|\mathcal{B}| = O(k)$ , where  $k = |V_P|$  and  $e = |A_P|$ .*

By combining Theorem 10 with Definition 6 and Theorem 9, we obtain as a corollary Theorem 1. The proof of Theorem 10 is divided in four steps, as follows. First (step 1), we remove the “easy” pieces of  $P$ ; this can break  $P$  into several connected components, and we show that their d.t.d.’s can be composed into a d.t.d. for  $P$ . Next, we show that if the  $i$ -th component has  $k_i$  nodes, then it admits a d.t.d. of width  $\frac{k_i}{4} + 2$ . This requires to “peel” the component to remove its tree-like parts (step 2) and decomposing the remainder using a reduction to standard tree decompositions (step 3). Finally, we wrap up our results and conclude the proof (step 4).

Throughout the proof, the relevant structure of  $P$  is encoded by a graph that we call the *skeleton* of  $P$ , defined as follows.

**Definition 7** The *skeleton* of a dag  $P = (V_P, A_P)$  is the bipartite dag  $\Lambda(P) = (V_\Lambda, E_\Lambda)$  where  $V_\Lambda = S \cup J$  and  $E_\Lambda \subseteq S \times J$ , and  $(u, v) \in E_\Lambda$  if and only if  $v \in J(u)$ .

Figure 3 gives an example. Note that  $\Lambda(P)$  does not contain nodes that are neither sources nor joints, as they are irrelevant to the d.t.d.. Note also that computing  $\Lambda(P)$  takes time  $O(\text{poly}(k))$ .

Let us now delve into the proof. For any node  $x$ , we denote by  $d_x$  the current degree of  $x$  in the skeleton.

*Step 1: Greedy bag construction* Set  $B^{(0)} = \emptyset$  and let  $\Lambda^{(0)} = (V_\Lambda^{(0)}, E_\Lambda^{(0)}) = \Lambda$ . Set  $j = 0$  and proceed iteratively as follows, recalling that  $V_\Lambda^{(j)} = S_\Lambda^{(j)} \cup J_\Lambda^{(j)}$ . For any

source  $u \in S_\Lambda^{(j)}$  let  $V_\Lambda^{(j)}(u)$  be the transitive closure of  $u$  in  $\Lambda^{(j)}$ . If there exists a source  $u \in S_\Lambda^{(j)}$  such that  $|V_\Lambda^{(j)}(u)| \geq 4$ , then let  $B^{(j+1)} = B^{(j)} \cup \{u\}$ , and let  $\Lambda^{(j+1)}$  be obtained from  $\Lambda^{(j)}$  by removing  $V_\Lambda^{(j)}(u)$  from  $V_\Lambda^{(j)}$ . Repeat the procedure until  $|V_\Lambda^{(j)}(u)| \leq 3$  for all  $u$ . Suppose the procedure stops at  $j = j^*$ , producing the subset  $B^* = B^{(j^*)}$  and the residual skeleton  $\Lambda^* = \Lambda^{(j^*)} = (V_\Lambda^*, E_\Lambda^*)$ .

**Lemma 6**  $|B^*| \leq \min\left(\frac{k - |V_\Lambda^*|}{4}, \frac{e - |E_\Lambda^*|}{4}\right)$ , where  $k = |V_P|$  and  $e = |A_P|$ .

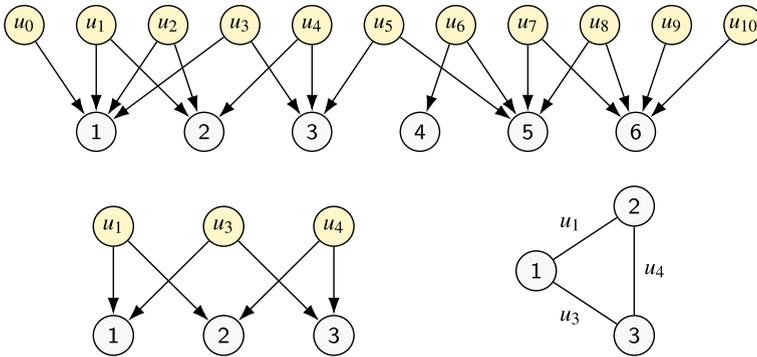
**Proof** For the first term of the  $\min()$ , note that at each step we remove at least 4 nodes from  $\Lambda^{(j)}$  and add one node to  $B^{(j)}$ . Hence  $4|B^*| \leq |V_\Lambda \setminus V_\Lambda^*| \leq k - |V_\Lambda^*|$ , which implies  $|B^*| \leq \frac{k - |V_\Lambda^*|}{4}$ . For the second term of the  $\min()$ , we show that the set of nodes  $V_\Lambda^{(j)}(u)$  removed at step  $j$  identifies at least 4 unique arcs of  $P$ . To this end, consider the sub-pattern  $P^{(j)} = P \setminus P(B^{(j)})$  containing all nodes not reachable from  $B^{(j)}$ . Note that  $\Lambda^{(j)}$  is the skeleton of  $P^{(j)}$ . Indeed, if  $v \in J^{(j)}$  then  $v$  is not reachable from any source in  $B^{(j)}$ , since otherwise  $v$  would have been removed before step  $j$ . Thus, at step  $j$  we are removing at least 3 joint nodes of  $P^{(j)}(u)$ . Therefore,  $P^{(j)}(u)$  contains at least three arcs pointing to its joints. In addition, by definition, the joints of  $P^{(j)}(u)$  must be reachable from some node in  $P^{(j)} \setminus P^{(j)}(u)$ . Thus there is an arc from  $P^{(j)} \setminus P^{(j)}(u)$  to a node of  $P^{(j)}(u)$ , and this node is therefore a joint itself. Thus,  $P^{(j)}$  contains at least 4 arcs pointing to the joints of  $P^{(j)}(u)$ . Since the joints of  $P^{(j)}(u)$  are then removed from  $P^{(j)}$ , these arcs are counted only once. Hence  $e \geq 4|B^*| + |E_\Lambda^*|$ , and  $|B^*| \leq \frac{e - |E_\Lambda^*|}{4}$ .  $\square$

Now, if  $B^* = S$ , then  $T = (\{B^*\}, \emptyset)$  is a d.t.d. of  $P$  whose width is  $\tau(T) = |B^*|$ . By Lemma 6,  $|B^*| \leq \min\left(\frac{k}{4}, \frac{e}{4}\right)$ , so Theorem 10 is proven.

If instead  $B^* \subset S$ , then  $\Lambda^*$  has  $\ell \geq 1$  nonempty connected components. For each  $i = 1, \dots, \ell$  let  $\Lambda_i = (S_i \cup J_i, E_i)$  be the  $i$ -th component of  $\Lambda^*$ . Let  $P^* = P \setminus P(B^*)$ , and let  $P_i = P^*(S_i)$ . Then,  $\Lambda_i$  is the skeleton of  $P_i$ ; this follows from the same argument used in the proof of Lemma 6. We shall now see that we can obtain a d.t.d. for  $P$  by arranging the d.t.d.'s of the  $P_i$  into a tree, and adding  $B^*$  to all bags.

**Lemma 7** For each  $i = 1, \dots, \ell$  let  $T_i = (\mathcal{B}_i, \mathcal{E}_i)$  be a d.t.d. of  $P_i$ . Consider the tree  $T$  obtained as follows. The root of  $T$  is the bag  $B^*$ , and the subtrees below  $B^*$  are  $T_1, \dots, T_\ell$ , where each bag  $B \in T_i$  has been replaced by  $B \cup B^*$ . Then  $T = (\mathcal{B}, \mathcal{E})$  is a d.t.d. of  $P$  with  $\tau(T) \leq |B^*| + \max_{i=1, \dots, \ell} \tau(T_i)$  and  $|\mathcal{B}| = 1 + \sum_{i=1}^\ell |\mathcal{B}_i|$ .

**Proof** The claims on  $\tau(T)$  and  $|\mathcal{B}|$  are straightforward. Let us check that  $T$  is a d.t.d. of  $P$ , via Definition 2. Property (1) is immediate. For property (2), note that  $\cup_{B \in \mathcal{B}_i} B = S_i$  because  $T_i$  is by hypothesis a d.t.d. of  $\Lambda_i$ . Thus  $\cup_{B \in \mathcal{B}} B = B^* \cup (\cup_{i=1}^\ell S_i) = S_P$ . We turn to property (3). Choose any two bags  $B' \cup B^*$  and  $B'' \cup B^*$  of  $T$ , where  $B' \in T_i$  and  $B'' \in T_j$  for some  $i, j \in \{1, \dots, \ell\}$ , and any bag  $B \cup B^* \in T(B' \cup B^*, B'' \cup B^*)$ . Suppose first  $i = j$ ; thus by construction  $B \in T(B', B'')$ . Since  $T_i$  is a d.t.d., then  $J_i(B') \cap J_i(B'') \subseteq J_i(B)$ , and in  $T$  this implies  $V_P(B' \cup B^*) \cap V_P(B'' \cup B^*) \subseteq V_P(B \cup B^*)$ . Suppose instead  $i \neq j$ . Thus  $J_i(S_i) \cap J_j(S_j) = \emptyset$  and this means that  $J(S_i) \cap J(S_j) \subseteq J(B^*)$ .



**Fig. 4** Above: example of a skeleton component  $\Lambda_i$ . Below: the core  $\Lambda_i^*$  obtained from  $\Lambda_i$  after peeling (left), and its encoding as  $C_i$  (right)

But  $V_P(B_i) \cap V_P(B_j) \subseteq J(S_i) \cap J(S_j)$  and  $J(B^*) \subseteq V_P(B^*)$ , thus  $V_P(B_i) \cap V_P(B_j) \subseteq V_P(B^*)$ . It follows that for every bag  $B \cup B^*$  of  $T$  we have  $V_P(B_i \cup B^*) \cap V_P(B_j \cup B^*) \subseteq V_P(B \cup B^*)$ .  $\square$

*Step 2: Peeling  $\Lambda_i$ .* We now remove the tree-like parts of  $\Lambda_i$ . These include, for instance, sources that have only one reachable joint. For each such source, we create a dedicated bag which becomes the child of another bag that reaches the same joint. This removes a source without increasing the width of the decomposition.

The construction is recursive. Let  $P_i^{(0)} = P_i$  and  $\Lambda_i^{(0)} = (S_i^{(0)} \cup J_i^{(0)}, E_i^{(0)}) = \Lambda_i$ . Set  $j = 0$ . For any node  $x \in \Lambda_i^{(j)}$ , we denote by  $d_u^{(j)}$  its degree in  $\Lambda_i^{(j)}$ . We will show that the tree  $T^{(0)}$  returned by our recursive construction is a d.t.d. for  $P_i^{(0)} = P_i$ .

The base case is  $|S_i^{(j)}| = 1$ . In this case we set  $T_i^{(j)} = (\{S_i^{(j)}\}, \emptyset)$ . Clearly,  $T_i^{(j)}$  is a d.t.d. for  $P_i^{(j)}$  of width 1 and we are done. Suppose instead  $|S_i^{(j)}| > 1$ . Recall that  $d_u^{(j)} \leq 2$  for all  $u \in S_i^{(j)}$ . Consider the first one of these three cases that applies (if none of them does, then we stop):

1.  $\exists u \in S_i^{(j)} : d_u^{(j)} = 1$ . Then choose any such  $u$ , and we choose any  $u' \in S_i^{(j)} \setminus \{u\}$  with  $J_i^{(j)}(u) \cap J_i^{(j)}(u') \neq \emptyset$ .
2.  $J_i^{(j)}(u) = J_i^{(j)}(u')$  for some  $u, u' \in S_i^{(j)}$  with  $u \neq u'$ . Then, choose any such  $u, u'$ .
3.  $\exists v \in J_i^{(j)} : d_v^{(j)} = 1$ . Then choose any such  $v$ , let  $u$  be the unique source such that  $v \in J_i^{(j)}(u)$ , and let  $u' \neq u$  be any source with  $J_i^{(j)}(u) \cap J_i^{(j)}(u') \neq \emptyset$ .

Then, we define  $T_i^{(j)}$  recursively as follows. Let  $P_i^{(j+1)} = P_i(S_i^{(j)} \setminus \{u\})$ , and let  $\Lambda_i^{(j+1)}$  be the skeleton graph obtained from  $\Lambda_i^{(j)}$  by removing  $u$  and (for the third case) the node  $v \in J_i^{(j)}(u)$  that is reachable only from  $u$ . We invoke the procedure recursively on  $\Lambda_i^{(j+1)}$ . Suppose the recursive procedure returns a d.t.d.  $T_i^{(j+1)}$  of  $P_i^{(j+1)}$ . Then,  $T_i^{(j+1)}$  must contain a bag  $B'$  such that  $u' \in B'$ . Create the bag  $B_u = \{u\}$ , set it as a child of  $B'$  in  $T_i^{(j+1)}$ , and let the resulting tree be  $T_i^{(j)}$ . Let us check that  $T_i^{(j)}$  is a d.t.d. for  $P_i^{(j)}$ . Properties (1) and (2) of Definition 2 are obviously satisfied. For property

(3), since  $B_u$  is a leaf, we only need to check that  $J_i^{(j)}(B_u) \cap J_i^{(j)}(B'') \subseteq J_i^{(j)}(B')$  for all  $B'' \in T_i^{(j)}$  and all  $B' \in T_i^{(j)}(B_u, B'')$ . To this end note that, by the choice of  $u$  and  $u'$ , for any  $u'' \in S_i^{(j)} \setminus \{u, u'\}$  we have  $J_i^{(j)}(u) \cap J_i^{(j)}(u'') \subseteq J_i^{(j)}(u')$ . We repeat the entire procedure until we reach the base case, or until  $|S_i^{(j)}| > 1$  and none of the three cases above holds, in which case we move to the next phase.

Before continuing, we make sure that the procedure above is well defined; we must guarantee that, in each of the three cases, the node  $u'$  exists. One can see that  $u'$  exists whenever  $|S_i^{(j)}| > 1$  (which is true by hypothesis) and  $\Lambda_i^{(j)}$  is connected. To see that  $\Lambda_i^{(j)}$  is connected, note that if this was not the case then in some step  $h < j$  we removed a source  $u$  with  $d_u^{(h)} = 2$  such that no other source  $u'$  has  $J_i^{(h)}(u) = J_i^{(h)}(u')$ . However, this cannot happen by construction of the procedure.

*Step 3: Decomposing the core* Suppose the peeling phase stops at  $j = j^*$ . Let  $P_i^* = P_i^{(j^*)}$  and  $\Lambda_i^* = (S_i^* \cup J_i^*, E_i^*) = \Lambda_i^{(j^*)}$ . We say  $P_i^*$  is the *core* of  $P_i$ ; this is the part that determines the dag treewidth. Now, since  $\Lambda_i^*$  violates all three conditions of the peeling step, we have  $d_u^* = 2$  for every source  $u$  and  $d_v^* \geq 2$  for every joint  $v$ . Thus  $\Lambda_i^*$  can be encoded as a simple graph. Formally, let  $C_i = (V_{C_i}, E_{C_i})$  where  $V_{C_i} = J_i^*$  and  $E_{C_i} = \{e_u : u \in S_i^*\}$ , where  $e_u = J_i^*(u)$  for each  $u \in S_i^*$ . To ease the discussion, for the edges we use  $u$  and  $e_u$  interchangeably. Figure 4 gives an example. Note that  $\Lambda_i^*$  is the skeleton of  $P_i^*$ , since  $S_i^*$  are the sources of  $P_i^*$  and the degree bound above implies that each  $v \in J_{C_i}$  is reachable from at least two sources of  $P_i^*$ . In what follows we let  $k_i = |S_i^* \cup J_i^*| = |E_{C_i}| + |V_{C_i}|$ .

We use  $C_i$  to compute a good d.t.d.  $T_i^* = (B_i^*, \mathcal{E}_i^*)$  of  $P_i^*$  via tree decompositions. First, we show that for our purposes it is sufficient to find a d.t.d. of width at most  $\frac{|E_{C_i}|}{5} + 3$ .

**Lemma 8** *If  $\tau(P_i^*) \leq \frac{|E_{C_i}|}{5} + 3$  then  $\tau(P_i^*) \leq \frac{k_i}{4} + 2$ .*

**Proof** First, suppose that  $|V_{C_i}| \leq 4$ . Since all nodes of  $C_i$  have degree at least 2, then  $C_i$  contains a 4-cycle, and thus an edge cover  $B^{\text{cov}}$  of size 2. We then build  $T_i^*$  by setting  $B^{\text{cov}}$  as root, and  $B_u = \{u\}$  for every  $u \in E_{C_i} \setminus B^{\text{cov}}$  as child of  $B^{\text{cov}}$ . This is clearly a d.t.d. for  $P_i^*$  of width  $2 < \frac{k_i}{4} + 2$ , and thus  $\tau(P_i^*) < \frac{k_i}{4} + 2$ .

Suppose instead that  $|V_{C_i}| \geq 5$ . Note that  $|E_{C_i}| \geq |V_{C_i}|$  by construction of  $C_i$ . One can check that these conditions imply  $\frac{|E_{C_i}|}{20} + \frac{|V_{C_i}|}{4} > 1$ , which in turn gives:

$$\tau(P_i^*) \leq \frac{|E_{C_i}|}{5} + 3 \leq \frac{|E_{C_i}| + |V_{C_i}|}{4} + 2 = \frac{k_i}{4} + 2 \tag{12}$$

concluding the proof. □

Therefore, we compute a d.t.d. of width at most  $\frac{|E_{C_i}|}{5} + 3$ . We do this in two steps.

**Lemma 9** *In time  $O(1.7549^{k_i})$  one can compute a tree decomposition  $D = (V_D, E_D)$  for  $C_i$  with treewidth at most  $\frac{|E_{C_i}|}{5} + 2$  and  $O(k_i)$  bags.*

**Proof** By Theorem 2 of [24], the treewidth of a graph  $G = (V, E)$  is at most  $\frac{|E|}{5} + 2$ . By Theorems 5.23-5.24 of [18], we can compute a minimum-width tree decomposition of an  $n$ -node graph in time  $O(1.7549^n)$ . By Lemma 5.16 of [18], in time  $O(n)$  we can transform such a decomposition into one that contains at most  $4n$  bags, leaving its width unchanged. Therefore, in time  $O(1.7549^{k_i})$  we can build a tree decomposition  $D$  for  $C_i$  with  $O(k_i)$  bags that satisfies  $t(D) \leq \frac{|E_{C_i}|}{5} + 2$ .  $\square$

**Lemma 10** *Let  $D = (V_D, E_D)$  be a tree decomposition of  $C_i$ . In time  $\text{poly}(k)$  we can build a d.t.d.  $T_i^* = (\mathcal{B}_i^*, \mathcal{E}_i^*)$  for  $P_i^*$  such that  $\tau(T_i^*) \leq t(D) + 1$  and  $|\mathcal{B}_i^*| = O(|V_D| + k_i)$ .*

**Proof** To simplify the notation let us write  $T, \mathcal{B}, \mathcal{E}$  in place of  $T_i^*, \mathcal{B}_i^*, \mathcal{E}_i^*$ . We first show how to build the tree  $T$ . The tedious part is proving it is a valid d.t.d. for  $P_i^*$ .

The intuition is that  $D$  covers the edges of  $C_i$ , which correspond to the sources of  $P_i^*$ . This gives a way to “convert” the bags of  $D$  into bags for  $T_i^*$ . For every  $v \in V_{C_i}$  choose an arbitrary incident edge  $u_v = \{v, z\} \in E_{C_i}$ . Replace each bag  $Y \in D$  by  $B(Y) = \{u_v : v \in Y\}$ , and for every  $u \in S_i^* \setminus \cup_{Y \in D} B(Y)$ , choose a bag  $B(Y) : J(u) \subseteq Y$ , and set the bag  $B_u = \{u\}$  as child of  $B(Y)$ . Let  $T$  be the resulting tree. To see that the construction is well-defined, note that, by point (2) of Definition 1, for any  $u \in E_{C_i}$  there exists some  $Y \in D$  such that  $u = \{x, y\} \subseteq Y$ . Therefore assigning  $B_u$  as child of some  $B(Y)$  with  $u \subseteq Y$  is licit. Now,  $\tau(T) \leq t(D) + 1$  follows immediately by the facts that  $|B(Y)| \leq |Y|$  for all  $Y \in D$  and that  $|B_u| = 1$  for each of the bags  $B_u$  above, and by Definitions 1 and 2. The bound  $|\mathcal{B}_i^*| = O(|V_D| + k_i)$  holds since  $T$  contains a bag for each bag of  $D$ , plus at most one bag for each node in  $S_i^*$ , and  $|S_i^*| \leq k_i$ .

Let us then check that  $T$  is a d.t.d. for  $P_i^*$  via Definition 2. Clearly,  $T$  is a tree and satisfies property (1). For property (2), let  $\hat{E}_{C_i}(D) = \cup_{Y \in D} B(Y)$ . Observe that by construction  $\cup_{B \in T} B = E_{C_i}(D) \cup (\cup_{u \in E_{C_i} \setminus E_{C_i}(D)} B_u)$ . The right-hand expression is  $E_{C_i}$ .

It remains to check property (3). First, if we have set  $B_u$  as child of  $B_Y$  then by construction  $J(B_u) \subseteq J(B_Y)$ . Thus we can ignore any such  $B_u$  and focus on the remaining bags of  $T$ , proving that every  $B, B', B''$  such that  $B \in T(B', B'')$  satisfy  $J(B') \cap J(B'') \subseteq J(B)$ . Let  $Y, Y', Y''$  the three bags of  $D$  from which the construction produced respectively  $B, B', B''$ . Observe that  $B \in T(B', B'')$  implies  $Y \in D(Y', Y'')$ . Now suppose that, by contradiction, there exists  $v \in J(B') \cap J(B'')$  such that  $v \notin J(B)$ . Note that, by construction, we must have put some  $u'$  with  $e_{u'} = \{v, z'\}$  in  $B'$  and some  $u''$  with  $e_{u''} = \{v, z''\}$  in  $B''$ , for some  $z', z'' \in V_i$ . Moreover,  $Y' \cap \{v, z'\} \neq \emptyset$  and  $Y'' \cap \{v, z''\} \neq \emptyset$ , else we could not have  $u' \in B'$  and  $u'' \in B''$ . Finally, bear in mind that  $v \notin Y$  and  $u', u'' \notin B$ , otherwise  $v \in J(B)$ , contradicting the hypothesis. Now we consider three cases. We use repeatedly properties (2) and (3) of Definition 1.

**Case 1**  $v \in Y' \cap Y''$ . Then  $v \in Y$ , a contradiction.

**Case 2**  $v \in Y'$  and  $v \notin Y''$ . Then  $z'' \in Y''$  and  $u''$  with  $e_{u''} = \{v, z''\}$  is the edge chosen to cover  $z''$ , else we would not put  $u'' \in B_{Y''}$ . Moreover there must be  $\hat{Y} \in D$  such that  $e_{u''} = \{v, z''\} \subseteq \hat{Y}$ . For the sake of the proof root  $D$  at  $Y$ , so  $Y'$  and  $Y''$  are in distinct subtrees. If  $\hat{Y}$  and  $Y''$  are in the same subtree then  $Y \in D(Y', \hat{Y})$ , but  $v \in Y' \cap \hat{Y}$  and thus  $v \in Y$ , a contradiction. Otherwise  $Y \in D(Y'', \hat{Y})$ , and since  $z'' \in Y'' \cap \hat{Y}$  then  $z'' \in Y$  and then  $u'' \in B(Y)$ , a contradiction.

**Case 3**  $v \notin Y'$  and  $v \notin Y''$ . Then  $z' \in Y'$ ,  $z'' \in Y''$ , and  $u', u''$  are the sources chosen to cover respectively  $z', z''$ . Moreover there must be  $\hat{Y}, \hat{Y}' \in D$  such that  $e_{u'} = \{v, z'\} \subseteq \hat{Y}$  and  $e_{u''} = \{v, z''\} \subseteq \hat{Y}'$ . Root again  $D$  at  $Y$ . If  $Y \in D(\hat{Y}, \hat{Y}')$  then since  $v \in \hat{Y} \cap \hat{Y}'$  it holds  $v \in Y$ , a contradiction. Otherwise  $\hat{Y}, \hat{Y}'$  are in the same subtree of  $D$ . If the subtree is the same as  $Y''$ , then  $Y \in D(Y', \hat{Y})$ , but  $z \in Y' \cap \hat{Y}$  and thus  $z' \in Y$  and thus  $u' \in B(Y)$ , a contradiction. Otherwise we have  $Y \in D(Y'', \hat{Y})$ ; but  $z'' \in Y'' \cap \hat{Y}$ , thus  $z'' \in Y$  and  $u'' \in B(Y)$ , again a contradiction.  $\square$

Combining Lemma 8, Lemma 9, and Lemma 10, we obtain:

**Lemma 11** *In time  $O(1.7549^{k_i} \text{poly}(k_i))$  we can compute a d.t.d.  $T_i^* = (\mathcal{B}_i^*, \mathcal{E}_i^*)$  for  $P_i^*$  such that  $\tau(T_i^*) \leq \lfloor \frac{k_i}{4} \rfloor + 2$  and  $|\mathcal{B}_i^*| = O(k_i)$ .*

With Lemma 11, we are almost done. It remains to wrap all our bounds together.

*Step 4: Assembling the tree* Recall the sub-patterns  $P_i$  obtained after the greedy bag construction (step 1). Let  $T_i = (\mathcal{B}_i, \mathcal{E}_i)$  be the d.t.d. for  $P_i$  as returned by the recursive peeling and the core decomposition. Since the peeling phase only adds bags of size 1, then  $\tau(T_i) = \tau(T_i^*)$ . Therefore, by Lemma 9,  $\tau(T_i) \leq \lfloor \frac{k_i}{4} \rfloor + 2$ . Moreover, since each bag added in the peeling phase corresponds to a unique source, then  $|\mathcal{B}_i| = O(k_i + |V(P_i)|) = O(|V(P_i)|)$ .

Let now  $T = (\mathcal{B}, \mathcal{E})$  be the d.t.d. for  $P$  obtained by assembling the trees  $T_1, \dots, T_\ell$  as in Lemma 7. By Lemma 7 itself,  $\tau(T) \leq |\mathcal{B}^*| + \max_{i=1, \dots, \ell} \tau(T_i)$ , thus:

$$\tau(T) \leq |\mathcal{B}^*| + \max_{i=1, \dots, \ell} \left\lfloor \frac{k_i}{4} \right\rfloor + 2 \tag{13}$$

Now, by Lemma 6 we know that  $P(\mathcal{B}^*)$  has at least  $4|\mathcal{B}^*|$  nodes and  $4|\mathcal{B}^*|$  arcs. Similarly, since each  $\Lambda_i$  has at least  $k_i$  nodes and  $k_i$  arcs, then  $P \setminus P(\mathcal{B}^*)$  has at least  $\sum_{i=1}^\ell k_i$  nodes and  $\sum_{i=1}^\ell k_i$  arcs. Then  $\tau(T) \leq \lfloor \frac{k}{4} \rfloor + 2$  and  $\tau(T) \leq \lfloor \frac{\epsilon}{4} \rfloor + 2$ , so  $\tau(T) \leq \min(\lfloor \frac{k}{4} \rfloor, \lfloor \frac{\epsilon}{4} \rfloor) + 2$ . Moreover,  $|\mathcal{B}| = 1 + \sum_{i=1}^\ell |\mathcal{B}_i| = O(\sum_{i=1}^\ell |V(P_i)|) = O(k)$ . Finally, by Lemma 9 the time to build  $T_i$  is  $O(1.7549^{k_i} \text{poly}(k_i))$ , since the peeling phase clearly takes time  $\text{poly}(k_i)$ . The total time to build  $T$  is therefore  $O(1.7549^k \text{poly}(k))$ . This concludes the proof of Theorem 10.

### 4.2 Bounds for quasi-cliques (Theorem 3)

**Lemma 12** *If a  $k$ -node dag  $P$  has  $\binom{k}{2} - \epsilon$  edges, then in time  $O(\text{poly}(k))$  one can compute a d.t.d.  $T$  for  $P$  on two bags such that  $\tau(T) \leq \lceil 1/2 + \sqrt{\epsilon/2} \rceil$ .*

**Proof** The source set  $S$  of  $P$  is an independent set. Hence  $\epsilon \geq \binom{|S|}{2}$ , and  $|S| \leq 1 + \sqrt{2\epsilon}$ . Consider any tree  $T$  on two bags  $B_1, B_2$  such that  $B_1 \cup B_2 = S$ ,  $|B_1| = \lfloor |S|/2 \rfloor$ , and  $|B_2| = \lceil |S|/2 \rceil$ . It is immediate to check that  $T$  satisfies the claim.  $\square$

By coupling Lemma 12 and Theorem 9, for computing  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  we obtain a running time bound of  $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{\epsilon}{2} \rfloor + \sqrt{\lfloor \frac{\epsilon}{2} \rfloor}} n^{\lfloor \frac{\epsilon}{2} \rfloor + \sqrt{\lfloor \frac{\epsilon}{2} \rfloor}} \log n)$ . For  $\text{ind}(H, G)$ , we refine the bound of Theorem 9 by observing that  $|D(H)| \leq 2^\epsilon$ . This yields a running time bound of  $2^{O(\epsilon + k \log k)} \cdot O(d^{k - \lfloor \frac{\epsilon}{2} \rfloor + \sqrt{\lfloor \frac{\epsilon}{2} \rfloor}} n^{\lfloor \frac{\epsilon}{2} \rfloor + \sqrt{\lfloor \frac{\epsilon}{2} \rfloor}} \log n)$ . This concludes the proof of Theorem 3.

### 4.3 Bounds for quasi-multipartite graphs (Theorem 4)

**Lemma 13** *If  $H$  is a complete multipartite graph, then  $\tau_2(H) = 1$ . If  $H$  is a complete multipartite graph plus  $\epsilon$  edges, then  $\tau_2(H) \leq \lfloor \frac{\epsilon}{4} \rfloor + 2$ . In either case, for any  $\theta \in \Theta(H)$ , for any acyclic orientation  $P$  of  $H/\theta$  we can compute in time  $2^{O(k)}$  a d.t.d. of  $P$  on  $O(k)$  bags whose width satisfies the bounds above.*

**Proof** First, suppose  $H = (V_H, E_H)$  is complete multipartite, so  $V_H = V_H^1 \cup \dots \cup V_H^\kappa$  where each  $V_H^j$  is a maximal independent set in  $H$ . In any acyclic orientation  $P$  of  $H$ , the source set  $S$  satisfies  $S \subseteq V_H^j$  for some  $j \in \{1, \dots, \kappa\}$ . Moreover,  $V_P(u) = V_P(u')$  for any  $u, u' \in S$ . A d.t.d.  $T$  for  $P$  of width  $\tau(T) = 1$  is the tree on  $|S|$  bags with one source per bag, which can be computed in time  $O(\text{poly}(k))$ .

Suppose now we add  $\epsilon$  arcs to  $P$ , with any orientation; this means  $H$  is a complete multipartite graph plus  $\epsilon$  edges. Again we have  $S \subseteq V_H^j$ , but now for some  $u, u' \in S$  we might have  $V_P(u) \neq V_P(u')$ , so the d.t.d above might not be valid anymore. Let  $P_j = P[V_H^j]$  and consider any d.t.d.  $T$  for  $P_j$ . We argue that  $T$  is a valid d.t.d. for  $P$  as well. First, the source set of  $P_j$  is same of  $P$  (that is,  $S$ ). Thus, since  $T$  satisfies properties (1) and (2) of Definition 2 for  $P_j$ , then it does so for  $P$ , too. For property (3), note that every node  $v \in V_H \setminus V_H^j$  is reachable from every  $u \in S$ . Thus, all bags  $B$  of  $T$  satisfy  $V_P(B) = (V_H \setminus V_H^j) \cup (V_P(B) \cap V_H^j)$ . As a consequence, for any three bags  $B, B_1, B_2$ , if  $V_{P_j}(B_1) \cap V_{P_j}(B_2) \subseteq V_{P_j}(B)$  then  $V_P(B_1) \cap V_P(B_2) \subseteq V_P(B)$ . Thus  $T$  satisfies property (3) and is a d.t.d. for  $P$ . Therefore, any d.t.d. for  $P_j$  is a d.t.d. for  $P$ . Now, since  $P_j$  has at most  $\epsilon$  edges, by Theorem 10 in time  $2^{O(k)}$  we can compute a d.t.d. for it of width at most  $\lfloor \frac{\epsilon}{4} \rfloor + 2$  on  $O(k)$  bags.

Consider now any  $\theta \in \Theta(H)$ . For any  $v \in V_H$ , we denote by  $\theta(v)$  the node of  $H/\theta$  corresponding to  $v$ , and for any  $x \in H/\theta$  we let  $\theta^{-1}(x) = \{u \in V_H : \theta(u) = x\}$  be the set of nodes of  $H$  identified in  $x$ . Let  $P$  be any acyclic orientation of  $H/\theta$ . Since the sources  $S$  of  $P$  form an independent set, then  $\cup_{x \in S} \theta^{-1}(x) \subseteq V_H^j$  for some  $j$ . Moreover, for any node  $x$  of  $P$ , if  $v \in \theta^{-1}(x)$  for some  $v \in V_H^i$  with  $i \neq j$  then  $x$  reachable from every node in  $S$ . Therefore, if we let  $V_P^j = \cup_{v \in V_H^j} \theta(v)$  and  $P_j = P[V_P^j]$ , the arguments above apply and we obtain the same bound. □

By coupling Lemma 13 and Theorem 9, when  $H$  is complete multipartite, for computing  $\text{hom}(H, G)$  and  $\text{sub}(H, G)$  we obtain a time bound of  $2^{O(k \log k)} \cdot O(d^{k-1} n \log n)$ . Similarly, when  $H$  is complete multipartite plus  $\epsilon$  edges, we obtain a time bound of  $2^{O(k \log k)} \cdot O(d^{k - \lfloor \frac{\epsilon}{4} \rfloor - 2} n^{\lfloor \frac{\epsilon}{4} \rfloor - 2} \log n)$ . This proves Theorem 4.

### 4.4 Independence number and dag treewidth

Recall that  $\alpha(H)$  is the independence number of  $H$ . We show:

**Lemma 14** *Any  $k$ -node graph  $H$  satisfies  $\Omega(\alpha(H)) \leq \tau(H) \leq \alpha(H)$ .*

**Proof** For the upper bound, note that  $\alpha(H'/\theta) \leq \alpha(H)$  for any  $H' \in D(H)$  and any  $\theta \in \Theta(H')$ . Moreover, in any acyclic orientation  $P$  of  $H'/\theta$  the sources form an independent set. Thus  $\tau(P) \leq \alpha(H)$ . The bound follows by Definition 6.

For the lower bound, we exhibit a pattern  $H'$  obtained by adding edges to  $H$  such that  $\tau(P) = \Omega(\alpha(H))$  for all its acyclic orientations  $P$ . Let  $I \subseteq V_H$  be an independent set of  $H$  with  $|I| = \Omega(\alpha(H))$  and  $|I| \bmod 5 \equiv 0$ . We add edges to  $I$ , so as to obtain the 1-subdivision of an expander. Partition  $I$  into  $I_J, I_S$  where  $|I_J| = \frac{2}{5}|I|$  and  $|I_S| = \frac{3}{5}|I|$ . Consider a 3-regular expander  $\mathcal{E} = (I_J, E_{\mathcal{E}})$  of linear treewidth  $t(\mathcal{E}) = \Omega(|I_J|)$ . It is well known that such expanders exist (see e.g. Proposition 1 and Theorem 5 of [21]). Note that  $|E_{\mathcal{E}}| = \frac{3}{2}|I_J| = |I_S|$ . For each edge  $\{u, v\} \in E_{\mathcal{E}}$ , we choose a distinct node in  $I_S$ , denoted by  $e_{uv}$ , and we add to  $H$  the edges  $\{e_{uv}, u\}$  and  $\{e_{uv}, v\}$ . Let  $H'$  be the resulting pattern. Observe that  $H'[I]$  is the 1-subdivision of  $\mathcal{E}$ , and that  $t(\mathcal{E}) = \Omega(\alpha(H))$  since  $|I_J| = \Omega(\alpha(H))$ .

Let now  $P = (V_P, A_P)$  be any acyclic orientation of  $H'$  where  $I_S \subseteq S_P$  where  $S_P$  are the sources of  $P$ . Such an orientation exists since  $I_S$  is an independent set in  $H'$ . Let  $T$  be any d.t.d. of  $P$ . We show that  $\tau(T) \geq \frac{1}{2}(t(\mathcal{E}) + 1) = \Omega(\alpha(H))$ , which implies  $\tau(P) = \Omega(\alpha(H))$  and therefore the thesis  $\tau(H) = \Omega(\alpha(H))$ . To this end, consider the tree  $D$  obtained from  $T$  by replacing each bag of sources  $B$  with the bag of nodes  $J(B) \cap I_J$ . We claim that  $D$  is a tree decomposition of  $\mathcal{E}$  of width at most  $2\tau(T) - 1$ . Let us start by checking the properties of Definition 1.

**Property (1)** By point (2) of Definition 2, the d.t.d.  $T$  satisfies  $\cup_{B \in T} B = S_P$ . Therefore, by construction of  $D$ , we have  $\cup_{X \in D} X = \cup_{B \in T} (J(B) \cap I_J) = I_J$ .

**Property (2)** Let  $\{v, w\}$  be any edge of  $\mathcal{E}$  where we recall that  $\{v, w\} \subseteq I_J$ . By construction of  $H'$ , there exists  $u \in I_S$  such that  $J(u) = \{v, w\}$  in  $P$ . Since  $T$  is a d.t.d., it satisfies point (2) of Definition 2, hence  $u \in B$  for some  $B \in T$ . By construction of  $D$  this implies there is some bag  $X \in D$  such that  $X = J(u) = \{v, w\}$ .

**Property (3)** Fix any three bags  $X_1, X_2, X_3 \in D$  such that  $X_1 \in D(X_2, X_3)$ . By construction,  $X_1 = J(B_1) \cap I_J, X_2 = J(B_2) \cap I_J, X_3 = J(B_3) \cap I_J$  for some  $B_1, B_2, B_3 \in T$  such that  $B_1 \in T(B_2, B_3)$ . Consider any  $v \in X_2 \cap X_3$ ; we need to show that  $v \in X_1$ . By construction of  $D$ , we have  $v \in X_2 \cap X_3 = J(B_2) \cap J(B_3) \cap I_J$ . Thus, there exist  $u \in B_2$  and  $u' \in B_3$  such that  $v \in J(u) \cap I_J$  and  $v \in J(u') \cap I_J$ . However, since  $B_1 \in T(B_2, B_3)$ , point (3) of Definition 1 implies  $J(u) \cap J(u') \subseteq J(B_1)$ . Therefore,  $v \in J(B_1)$  as well. Moreover  $v \in I_J$ , and thus  $v \in J(B_1) \cap I_J$ . But  $J(B_1) \cap I_J = X_1$ , so  $v \in X_1$ .

Hence,  $D$  is a tree decomposition of  $\mathcal{E}$ . Finally, note that any bag  $X \in D$  by construction satisfies  $|X| = |J(B) \cap I_J| \leq 2|B|$  since any source  $u \in B$  has at most 2 arcs towards  $I_J$ . Then by Definition 1 and Definition 3 we have  $t(\mathcal{E}) \leq 2\tau(P) - 1$ , that is,  $\tau(P) \geq \frac{1}{2}(t(\mathcal{E}) + 1)$ , as claimed. □

## 5 Lower bounds

We prove Theorem 6, in a more technical form. Note that, since  $\tau(H) = \Theta(\alpha(H))$  by Lemma 14, the bound still holds if one replaces  $\tau(H)$  by  $\alpha(H)$ . The proof uses the following result:

**Theorem 11** ([12], Theorem I.2) *The following problems are #W[1]-hard and, assuming ETH, cannot be solved in time  $f(k) \cdot n^{o(k/\log k)}$  for any computable function  $f$ : counting (directed) paths or cycles of length  $k$ , and counting edge-colorful or uncolored  $k$ -matchings in bipartite graphs.*

Let us now state the lower bound.

**Theorem 12** *Choose any function  $a : \mathbb{N} \rightarrow \mathbb{N}$  such that  $a(k) \in [1, k]$  for all  $k \in \mathbb{N}$ . There exists an infinite family  $\mathcal{H}$  of patterns such that (1) for all  $H \in \mathcal{H}$  we have  $\tau(H) = \Theta(a(|V(H)|))$ , and (2) if there exists an algorithm that for all  $H \in \mathcal{H}$  computes  $\text{ind}(H, G)$  or  $\text{sub}(H, G)$  in time  $f(d, k) \cdot n^{o(a(\tau(H))/\log a(\tau(H)))}$ , where  $d$  is the degeneracy of  $G$ , then ETH fails.*

**Proof** We reduce counting cycles in an arbitrary graph to counting a gadget pattern on  $k$  nodes and dag treewidth  $O(a(k))$  in a  $d$ -degenerate graph.

First, fix a function  $d : \mathbb{N} \rightarrow \mathbb{N}$  such that  $d(k) \in \Omega(\frac{k}{a(k)})$ . Now consider a simple cycle on  $k_0 \geq 3$  nodes and any arbitrarily large  $k \geq 3$ . Our gadget pattern on  $k$  nodes is the following. For each edge  $e = uv$  of the cycle create a clique  $C_e$  on  $d(k) - 1$  nodes; delete  $e$  and connect both  $u$  and  $v$  to every node of  $C_e$ . The resulting pattern  $H$  has  $k = k_0 d(k)$  nodes. Let us prove that  $\tau(H) \leq k_0$ ; since  $k_0 = \frac{k}{d(k)} \in O(a(k))$ , this implies  $\tau(H) = O(a(k))$  as desired. Consider again the generic edge  $e = uv$ . In any acyclic orientation  $P$  of  $H$ , the set  $C_e \cup u$  induces a clique, and thus can contain at most one source. Applying the argument to all  $e$  shows that  $|S(P)| \leq k_0$ , hence  $\tau(P) \leq k_0$ . This holds also if we add edges and/or identify nodes of  $P$ , hence  $\tau(H) \leq k_0$ .

Now consider a simple graph  $G_0$  on  $n_0$  nodes and  $m_0$  edges. We replace each edge of  $G_0$  as described above, which takes time  $O(\text{poly}(n_0))$ . The resulting graph  $G$  has  $n = m_0(d - 1) + n_0 = O(dn_0^2)$  nodes and degeneracy  $d$ . Every  $k_0$ -cycle of  $G_0$  is univocally associated to a copy of  $H$  in  $G$  (note that every non-induced copy of  $H$  in  $G$  is induced too). Suppose we have an algorithm that computes  $\text{ind}(H, G)$  or  $\text{sub}(H, G)$  in time  $f(d, k) \cdot n^{o(a(\tau(H))/\log a(\tau(H)))}$ . Since  $\tau(H) \leq k_0$ ,  $n = O(dn_0^2)$ ,  $k = f_1(d, k_0)$ , and  $d = f_2(k_0)$ , for some functions  $f_1, f_2$ , then the running is time  $f(k_0) \cdot (n_0)^{o(k_0/\log k_0)}$ . Invoking Theorem 11 concludes the proof.  $\square$

## 6 Conclusions

We have shown how, by introducing a novel tree-like decomposition for directed acyclic graphs, one can improve on the decades-old state-of-the-art subgraph counting algorithms when the host graph is sufficiently sparse. This decomposition seems to capture the structure of the problem, and may therefore be of independent interest.

We leave open the (important) problem of finding a *tight* characterization of the class of patterns that are fixed-parameter tractable with respect to homomorphism, subgraph, and induced subgraph counting, when the complexity is parameterized by the degeneracy of the host graph. Our results represent one step, outlining properties (the boundedness of our notions of width) that are *sufficient*, but maybe not necessary. Indeed, our lower bounds only give evidence that such properties are necessary for counting induced or non-induced occurrences, and only for *some* classes of patterns. Finding a complete characterization, and therefore establishing a class pattern dichotomy for all three problems, would be an exciting development.

**Acknowledgements** Most of this work was done while the author was affiliated with the Department of Computer Science of the Sapienza University of Rome. The author was supported in part by: BICI, the Bertinoro International Center for Informatics; a Google Focused Award “Algorithms and Learning for AI” (ALL4AI); the ERC Starting Grant DMAP 680153; the “Dipartimenti di Eccellenza 2018–2022” Grant awarded to the Department of Computer Science of the Sapienza University of Rome. The author thanks the anonymous reviewers for their precious feedback.

**Funding** Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Alon, N., Dao, P., Hajirasouliha, I., Hormozdiari, F., Sahinalp, S.C.: Biomolecular network motif counting and discovery by color coding. *Bioinformatics* **24**(13), i241–249 (2008)
2. Berend, D., Tassa, T.: Improved bounds on Bell numbers and on moments of sums of random variables. *Probab. Math. Stat.* **30**(2), 185–205 (2010)
3. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztegombi, K.: Counting Graph Homomorphisms, pp. 315–371 (2006)
4. Bressan, M.: Faster subgraph counting in sparse graphs. In: Proceedings of IPEC, vol. 148, pp. 6:1–6:15 (2019)
5. Bressan, M., Chierichetti, F., Kumar, R., Leucci, S., Panconesi, A.: Counting graphlets: space vs. time. In: Proceedings of ACM WSDM, pp. 557–566 (2017)

6. Bressan, M., Chierichetti, F., Kumar, R., Leucci, S., Panconesi, A.: Motif counting beyond five nodes. *ACM Trans. Knowl. Discov. Data* **20**(2), 1–25 (2018)
7. Bressan, M., Leucci, S., Panconesi, A.: Motivo: fast motif counting via succinct color coding and adaptive sampling. *Proc. VLDB Endow.* **12**(11), 1651–1663 (2019)
8. Chen, J., Chor, B., Fellows, M., Huang, X., Juedes, D., Kanj, I.A., Xia, G.: Tight lower bounds for certain parameterized NP-hard problems. *Inf. Comput.* **201**(2), 216–231 (2005)
9. Chen, J., Huang, X., Kanj, I.A., Xia, G.: Strong computational lower bounds via parameterized complexity. *J. Comput. Syst. Sci.* **72**(8), 1346–1367 (2006)
10. Chiba, N., Nishizeki, T.: Arboricity and subgraph listing algorithms. *SIAM J. Comput.* **14**(1), 210–223 (1985)
11. Curticapean, R., Dell, H., Marx, D.: Homomorphisms are a good basis for counting small subgraphs. In: *Proceedings of ACM STOC*, pp. 210–223 (2017)
12. Curticapean, R., Marx, D.: Complexity of counting subgraphs: only the boundedness of the vertex-cover number counts. In: *Proceedings of IEEE FOCS*, pp. 130–139 (2014)
13. Diestel, R.: *Graph Theory*, 5th edn. Springer, Berlin (2017)
14. Eppstein, D.: Arboricity and bipartite subgraph listing algorithms. *Inf. Process. Lett.* **51**(4), 207–211 (1994)
15. Eppstein, D.: Subgraph isomorphism in planar graphs and related problems. *J. Graph Algorithms Appl.* **3**(3), 1–27 (1999)
16. Eppstein, D., Löffler, M., Strash, D.: Listing all maximal cliques in sparse graphs in near-optimal time. In: *Algorithms and Computation*, pp. 403–414. Springer, Berlin (2010)
17. Eppstein, David, Strash, Darren: Listing all maximal cliques in large sparse real-world graphs. In: *Experimental Algorithms*, pp. 364–375. Springer, Berlin (2011)
18. Fedor, V.: Fomin and Dieter Kratsch, 1st edn. *Exact Exponential Algorithms*. Springer, Berlin (2010)
19. Ganian, R., Hliněný, P., Kneis, J., Meister, D., Obdržálek, J., Rossmanith, P., Sikdar, S.: Are there any good digraph width measures? *J. Comb. Theory Ser. B* **116**, 250–286 (2016)
20. Grohe, M., Kreuzer, S., Siebertz, S.: Characterisations of nowhere dense graphs (invited talk). *Proc. FSTTCS* **24**, 21–40 (2013)
21. Grohe, M., Marx, D.: On tree width, bramble size, and expansion. *J. Comb. Theory Ser. B* **99**(1), 218–228 (2009)
22. Grohe, M., Schweikardt, N.: First-order query evaluation with cardinality conditions. In: *Proceedings of ACM SIGMOD*, pp. 253–266 (2018)
23. Impagliazzo, R., Paturi, R., Zane, F.: Which problems have strongly exponential complexity? In: *Proceedings of IEEE FOCS*, pp. 653–662 (1998)
24. Kneis, J., Mölle, D., Richter, S., Rossmanith, P.: Algorithms based on the treewidth of sparse graphs. In: *Graph-Theoretic Concepts in Computer Science*, pp. 385–396. Springer, Berlin (2005)
25. Le Gall, F.: Powers of tensors and fast matrix multiplication. In: *Proceedings of ISSAC*, pp. 296–303 (2014)
26. Mathon, R.: A note on the graph isomorphism counting problem. *Inf. Process. Lett.* **8**(3), 131–136 (1979)
27. Nešetřil, J., de Mendez, P.O.: Sparsity: graphs, structures, and algorithms. In: *Algorithms and Combinatorics*. Springer, Berlin (2012)
28. Nešetřil, J., de Mendez, P.O.: On nowhere dense graphs. *Eur. J. Combin.* **32**(4), 600–617 (2011)
29. Nešetřil, J., Poljak, S.: On the complexity of the subgraph problem. *Comment. Math. Univ. Carol.* **026**(2), 415–419 (1985)
30. Patel, V., Regts, G.: Computing the number of induced copies of a fixed graph in a bounded degree graph. *Algorithmica* **81**(5), 1844–1858 (2018)
31. Sariyüce, A.E., Pinar, A.: Peeling bipartite networks for dense subgraph discovery. In: *Proceedings of ACM WSDM*, pp. 504–512 (2018)
32. Sariyüce, A.E., Seshadhri, C., Pinar, A., Çatalyürek, Ü.V.: Nucleus decompositions for identifying hierarchy of dense subgraphs. *ACM Trans. Web* **11**(3), 16:1–16:27 (2017)
33. Tsourakakis, C.E., Pachocki, J., Mitzenmacher, M.: Scalable motif-aware graph clustering. In: *Proceedings of WWW*, pp. 1451–1460 (2017)