Reconstruction from smaller cards

Groenland, Carla; Johnston, Tom; Scott, Alex; Tan, Jane

DOI
10.1007/s11856-025-2858-3

Publication date
2025

Document Version
Final published version

Published in
Israel Journal of Mathematics

ISRAEL JOURNAL OF MATHEMATICS **TBD** (2025), 1–38
DOI: 10.1007/s11856-025-2858-3

RECONSTRUCTION FROM SMALLER CARDS

BY

Carla Groenland

Delft Institute of Applied Mathematics, Technische Universiteit Delft
2628 CD Delft, Netherlands
e-mail: c.e.groenland@tudelft.nl

AND

Tom Johnston, Alex Scott* and Jane Tan

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
e-mail: tom.johnston@bristol.ac.uk, scott@maths.ox.ac.uk, jane.tan@maths.ox.ac.uk

ABSTRACT

The ℓ-deck of a graph G is the multiset of all induced subgraphs of G on ℓ vertices. We say that a graph is reconstructible from its ℓ-deck if no other graph has the same ℓ-deck. In 1957, Kelly showed that every tree with $n \geq 3$ vertices can be reconstructed from its $(n-1)$-deck, and Giles strengthened this in 1976, proving that trees on at least 6 vertices can be reconstructed from their $(n-2)$-decks. Our main theorem states that trees are reconstructible from their $(n-r)$-decks for all $r \leq n/9 + o(n)$, making substantial progress towards a conjecture of Nýdl from 1990. In addition, we can recognise the connectedness of a graph from its ℓ-deck when $\ell \geq 9n/10$, and reconstruct the degree sequence when $\ell \geq \sqrt{2n\log(2n)}$. All of these results are significant improvements on previous bounds.

* Supported by EPSRC grant EP/V007327/1.
Received August 17, 2021 and in revised form July 28, 2024

1. Introduction

Throughout this paper, all graphs are finite and undirected with no loops or multiple edges. Given a graph G on n vertices and any vertex $v \in V(G)$, the *card* $G-v$ is the subgraph of G obtained by removing the vertex v together with all edges incident to v. The **deck** $\mathcal{D}(G)$ is then the multiset of all unlabelled cards of G. A graph G is said to be **reconstructible** from its deck if any graph with the same deck is isomorphic to G.

The graph reconstruction conjecture of Kelly and Ulam [18, 19, 37] states that all graphs on at least three vertices are reconstructible. While this classical conjecture has been verified for certain classes such as trees (Kelly [19]), outerplanar graphs (Giles [13]) and maximal planar graphs (Lauri [24]), it remains open even for simple classes such as planar graphs with maximum degree three. However, various graph parameters, such as the degree sequence and connectedness, are known to be reconstructible for general graphs in the sense that they are determined by the deck (i.e., if two graphs have the same deck, then the parameter takes the same value for both graphs).

There is a significant body of research on the problem of reconstructing graphs and graph parameters from smaller cards. Instead of taking induced subgraphs on $n-1$ vertices, it is natural to consider cards which are the induced subgraphs on ℓ vertices where ℓ may be much smaller than $n-1$. The **ℓ-deck** of G, denoted by $\mathcal{D}_\ell(G)$, is the multiset of the isomorphism classes of all $\binom{n}{\ell}$ induced subgraphs of G on ℓ vertices (in this notation $\mathcal{D}(G) = \mathcal{D}_{n-1}(G)$).

Extending the terminology from the classical case, a graph G is **reconstructible from the ℓ-deck** if it is uniquely determined up to isomorphism by its ℓ-deck: that is, if $\mathcal{D}_\ell(G) = \mathcal{D}_\ell(G')$ for a graph G', then $G \cong G'$. A graph parameter (or property) is **reconstructible from the ℓ-deck** if the value it takes for any graph (or whether or not the property holds) is determined by the ℓ-deck of that graph. If $\mathcal{C}$ is a class of graphs, we say that $\mathcal{C}$ is **recognisable from the ℓ-deck** if the property of belonging to $\mathcal{C}$ is reconstructible from the ℓ-deck. We say that a graph G is **reconstructible amongst graphs in $\mathcal{C}$ from its ℓ-deck** if any other graph in $\mathcal{C}$ with the same ℓ-deck is isomorphic to G. The class of graphs $\mathcal{C}$ is **weakly reconstructible from the ℓ-deck** if any two graphs in $\mathcal{C}$ with the same ℓ-deck are isomorphic, and if $\mathcal{C}$ is also recognisable from the ℓ-deck, it is said to be **reconstructible from the ℓ-deck**.

Intuitively, individual cards that are smaller carry less information. Indeed, the $(\ell-1)$-deck is determined by the ℓ-deck for each ℓ, since each $(\ell-1)$-vertex subgraph occurs in exactly $n-\ell+1$ members of the ℓ-deck (see Lemma 8). Thus, a graph that is reconstructible from its ℓ'-deck is also reconstructible from its ℓ-deck for all $\ell \geq \ell'$. The main question is then to determine the threshold; that is, find the smallest ℓ for which a given class of graphs or a property is reconstructible from the ℓ-deck.

As far as we are aware, the earliest mention of reconstruction from small cards is a brief suggestion in the final sentence of Kelly's paper on reconstructing trees [19]. The extension of the Reconstruction Problem that follows seems to have been formulated by Manvel, who called it "Kelly's Conjecture".

Conjecture 1 ([27]): *For every $r \in \mathbb{N}$, there is an integer N_r such that every graph with at least N_r vertices is reconstructible from its $(n-r)$-deck.*

Kelly and Ulam's conjecture posits that $N_1 = 3$. In the same paper where they posed this extension, Manvel [27] showed that several classes of graphs, such as connected graphs, trees, regular graphs and bipartite graphs, can be recognised from the $(n-2)$-deck where $n \geq 6$ is the number of vertices. Since then, recognition and reconstruction problems of this type have been widely studied. Recent developments include the reconstructibility of 3-regular n-vertex graphs from the $(n-2)$-deck (Kostochka, Nahvi, West and Zirlin [21]), and that almost all graphs are reconstructible from the $(n-r)$-deck when $r \leq (1/2 - o(1))n$ (Spinoza and West [33], building on results of Müller [28] and Bollobás [3]). For further background, we refer to the survey of Kostochka and West [23].

For general graphs, it is not possible to guarantee reconstructibility from the $(n-r)$-deck unless $r = o(n)$, as shown by the following theorem of Nýdl.

Theorem 2 (Nýdl [32]): *For any integer n_0 and $0 < \alpha < 1$, there exists an integer $n > n_0$ such that there are two non-isomorphic graphs on n vertices which share the same multiset of subgraphs of order at most αn.*

However, Nýdl's theorem may not hold for specific families of graphs. Indeed, Nýdl conjectured in 1990 that all trees are weakly reconstructible from their ℓ-deck when ℓ is slightly larger than $n/2$.

Conjecture 3 (Nýdl [31]): *For any $n \geq 4$ and $\ell \geq \lfloor n/2 \rfloor + 1$, if two trees on n vertices have the same ℓ-deck then they are isomorphic.*

The conjectured bound on ℓ would be sharp: Nýdl [31] presented trees for which $\ell \geq \lfloor n/2 \rfloor + 1$ is necessary (see [23] for a short proof).

There has been no progress on Nýdl's conjecture since it was made. Indeed, the best previous result is an earlier theorem of Giles [14] from 1976, which states that for $n \geq 5$ no two non-isomorphic n-vertex trees have the same $(n-2)$-deck. That is, this gives the cases of the conjecture where $n \geq 5$ and $\ell \geq n-2$. Using the result of Manvel [27] that the class of n-vertex trees is recognisable from the $(n-2)$-deck when $n \geq 6$, Giles' result confirms that trees are reconstructible (not just weakly) from their $(n-2)$-deck for any $n \geq 6$.

Our main theorem improves very substantially on the result of Giles and takes a significant step towards Conjecture 3, showing that we can reconstruct trees from the $(n-r)$-deck for r with linear size.

Theorem 4: *Any n-vertex tree T can be reconstructed from $\mathcal{D}_{n-r}(T)$ when $r < \frac{n}{9} - \frac{4}{9}\sqrt{8n+5} - 1$.*

In particular, it follows that Nýdl's theorem (Theorem 2) does not hold when restricted to the class of trees. We remark that Conjecture 3 is false in the case $n = 13$, as demonstrated by the two graphs in Figure 1 which have been verified to have the same deck by computer. However, our computer search has also shown that the conjecture is true for all other n in the range $4 \leq n \leq 25$. It remains open for large n.

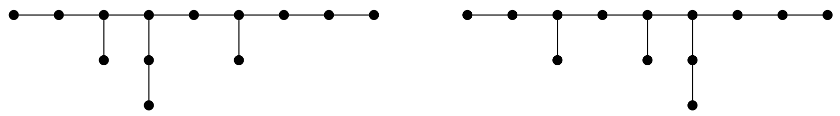

Figure 1. Two non-isomorphic trees on 13 vertices which have the same 7-deck.

It is worth noting that the class of trees, being one of the first non-trivial classes shown to be reconstructible in the classical sense, is very prominent in reconstruction literature. For example, assuming we know *a priori* that the graph is a tree (as in weak reconstruction), Harary and Palmer [16] showed how to reconstruct a tree using only the cards that are subtrees, Bondy [4] showed that only the cards where peripheral vertices (that is, leaves with maximum eccentricity) have been removed are needed, and Manvel [26] subsequently showed that the set (as opposed to the multiset) of cards that are subtrees suffices except in two cases. Lauri [25] also showed that trees with at least three cut-vertices can be reconstructed (amongst all graphs) from the cards corresponding to removing a cut-vertex. Indeed, Myrvold [29] proved that only three carefully chosen cards are needed to reconstruct a tree when $n \geq 5$. Related problems have also been investigated extensively for infinite trees (see, for example, [1, 2, 6, 17, 30, 35]), and it was recently shown by Bowler, Erde, Heinig, Lehner and Pitz that there are non-reconstructible locally finite trees [10].

Returning to the small cards setting, we have already mentioned Manvel's result in [27] that the class of connected graphs is recognisable from the $(n-2)$-deck for $n \geq 6$. Extending this, Kostochka, Nahvi, West and Zirlin [20] showed that the connectedness of a graph on $n \geq 7$ vertices is determined by $\mathcal{D}_{n-3}(G)$. As shown by Spinoza and West [33], if we take $G_1 = P_n$ (the path on n vertices) and $G_2 = C_{\lceil n/2 \rceil + 1} \sqcup P_{\lfloor n/2 \rfloor - 1}$ (the disjoint union of a cycle and a path), we find that $\mathcal{D}_\ell(G_1) = \mathcal{D}_\ell(G_2)$ for all $\ell \leq \lfloor n/2 \rfloor$. However, G_1 is connected and G_2 is not. In light of this construction, Spinoza and West believe that for $n \geq 6$ and $\ell \geq \lfloor n/2 \rfloor + 1$, the connectedness of an n-vertex graph G is determined by $\mathcal{D}_\ell(G)$. This threshold would be sharp.

Spinoza and West proved in [33] that connectedness can be recognised from $\mathcal{D}_\ell(G)$ provided

$$n - \ell \leq (1 + o(1))\sqrt{\frac{2 \log n}{\log(\log n)}}.$$

We significantly improve this bound to allow a linear gap between n and ℓ.

THEOREM 5: *The connectedness of an n-vertex graph G can be recognised from $\mathcal{D}_\ell(G)$ provided $\ell \geq 9n/10$.*

By Theorem 5 (and the fact that we can reconstruct the number of edges), we can recognise trees from the ℓ-deck when $\ell \geq 9n/10$. In order to prove Theorem 4, we need a slightly stronger bound.

THEOREM 6: *For $\ell \geq (2n+4)/3$, the class of trees on n vertices is recognisable from the ℓ-deck.*

As we were completing this paper, Kostochka, Nahvi, West and Zirlin [22] independently announced a similar result to Theorem 6. In fact, they proved that one can recognise if a graph is acyclic from the ℓ-deck when $\ell \geq \lfloor n/2 \rfloor + 1$, which also verifies the believed bound for reconstructing connectedness in the special case of forests. This has the particularly nice consequence that trees can be recognised from their ℓ-deck, and so Conjecture 3 is equivalent to the reconstruction of trees amongst general graphs. Since our proof of Theorem 6 is short and already (more than) sufficient for our use in reconstructing trees, we have retained it for completeness.

The proof of Theorem 5 relies on an algebraic result (Lemma 11) which we also apply to reconstructing degree sequences. The story in the literature here is similar to that of connectedness. Chernyak [12] showed that the degree sequence of an n-vertex graph can be reconstructed from its $(n-2)$-deck for $n \geq 6$, and this was later extended by Kostochka, Nahvi, West, and Zirlin [20] to the $(n-3)$-deck for $n \geq 7$. The best known asymptotic result is due to Taylor [34], and implies that the degree sequence of a graph G on n vertices can be reconstructed from $\mathcal{D}_\ell(G)$ where $\ell \sim (1-1/e)n$. Our improved bound is as follows.

Theorem 7: *The degree sequence of an n-vertex graph G can be reconstructed from $\mathcal{D}_\ell(G)$ for any $\ell \geq \sqrt{2n\log(2n)}$.*

In Section 2, we give ℓ-deck versions of both Kelly's Lemma [19] for counting subgraphs and a result on counting maximal subgraphs by Greenwell and Hemminger [15], as well as an algebraic result of Borwein and Ingalls [9] bounding the number of moments shared by two distinct sequences. These are used to deduce Theorem 7 (Section 3) and Theorem 5 (Section 4). Section 5 contains the proof of Theorem 4, our main result on reconstructing trees, including the tree recognition statement given by Theorem 6. There, we also introduce a new counting tool for reconstruction that may be of independent interest. We conclude with some further discussion in Section 6.

2. Preliminaries

This paper makes extensive use of three key results which we give in this section.

2.1. Kelly's Lemma for small cards. Perhaps the most fundamental tool in graph reconstruction is Kelly's Lemma for reconstructing subgraph counts. The utility of such a result is reflected in the fact that variants of the lemma exist for many different reconstruction problems (see [5]). To formulate Kelly's Lemma, let $\widetilde{n}_H(G)$ and $n_H(G)$ denote the number of subgraphs and induced subgraphs of G isomorphic to H, respectively. That is, $n_H(G)$ is the number of vertex subsets $S \subseteq V(G)$ that induce a subgraph of G isomorphic to H, and $\widetilde{n}_H(G)$ is the number of edge subsets $S \subseteq E(G)$ that induce a subgraph of G isomorphic to H. We will refer to an induced subgraph isomorphic to H as an **induced copy** of H, and say **copy** by itself to mean not necessarily induced. When H is a connected graph, every copy of H in a tree is an induced copy (and vice versa), but the difference between these notions can be relevant for some of our results that apply to a wider class of graphs.

In the classical graph reconstruction problem, Kelly's Lemma states that we can reconstruct $n_H(G)$ and $\widetilde{n}_H(G)$ provided $|V(H)| < |V(G)|$. We will use the following small cards variant, which is a direct generalisation.

LEMMA 8: *Let $\ell \in \mathbb{N}$ and let H be a graph on at most ℓ vertices. For any graph G, the multiset of ℓ-vertex induced subgraphs of G determines both the number of subgraphs of G that are isomorphic to H and the number of induced subgraphs that are isomorphic to H.*

Proof. Suppose we count the number of induced copies of H in each of the ℓ-cards of G, and take the sum over all cards. Each induced copy of H in G is counted exactly $\binom{n-|V(H)|}{\ell-|V(H)|}$ times in this total. Hence, we can reconstruct the number $n_H(G)$ of induced copies of H in G from the ℓ-deck as

$$n_H(G) = \binom{n-|V(H)|}{\ell-|V(H)|}^{-1} \sum_{C \in \mathcal{D}_\ell(G)} n_H(C).$$

The same argument applies with copies in place of induced copies. ∎

In particular, Kelly's Lemma means that $\mathcal{D}_{\ell'}(G)$ can be reconstructed from $\mathcal{D}_\ell(G)$ for all $\ell' \leq \ell$. Foreshadowing later usage of this lemma, we remark that in the displayed formula in the proof, we only need to use the subset of the deck consisting of all cards which contain at least one (possibly induced) copy of the fixed graph H. Thus, we can still reconstruct these subgraph counts if we are handed a subset of the deck and told that the subset includes every card containing a copy of the subgraph.

2.2. COUNTING MAXIMAL $\mathcal{F}$-SUBGRAPHS. Given a class of graphs $\mathcal{F}$, a subgraph F' of some graph G is said to be an **$\mathcal{F}$-subgraph** if F' is isomorphic to some $F \in \mathcal{F}$, and is a **maximal** $\mathcal{F}$-subgraph if the subgraph F' cannot be extended to a larger $\mathcal{F}$-subgraph, that is, there does not exist an $\mathcal{F}$-subgraph F'' of G such that F' is a proper subgraph of F''.

Let $m_{\mathcal{F}}(F,G)$ denote the number of maximal $\mathcal{F}$-subgraphs in G which are isomorphic to F. We suppress the subscript when it is clear from the context.

We give a slight variation of a classical "Counting Theorem" due to Bondy and Hemminger [7] (see also the statement of Greenwell and Hemminger [15]) which reconstructs $m_{\mathcal{F}}(F,G)$ from the ℓ-deck. The following proof is essentially that of Bondy and Hemminger [7], only with a few additional observations used to accommodate our slight changes to the assumptions.

Lemma 9: *Given $\ell, n \in \mathbb{N}$ with $\ell < n$, let $\mathcal{G}$ be a class of n-vertex graphs. Let $\mathcal{F}$ be a class of graphs such that for any $G \in \mathcal{G}$ and for any $\mathcal{F}$-subgraph F of G,*

(i) $|V(F)| \leq \ell$;
(ii) *F is contained in a unique maximal $\mathcal{F}$-subgraph of G.*

Then for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, we can reconstruct $m_{\mathcal{F}}(F,G)$ from the collection of cards in the ℓ-deck that contain an $\mathcal{F}$-subgraph.

Proof. Define an (F,G)**-chain** of length k to be a sequence $(X_0,\dots,X_k)$ of $\mathcal{F}$-subgraphs of G such that

$$F \cong X_0 \subsetneq X_1 \subsetneq \cdots \subsetneq X_k \subsetneq G.$$

The **rank** of F in G is the length of a longest (F,G)-chain, and two chains are called **isomorphic** if they have the same length and the corresponding terms are isomorphic. Following Bondy and Hemminger's argument, we first show that

$$m_{\mathcal{F}}(F,G) = \sum_{k=0}^{\operatorname{rank} F} \sum (-1)^k \widetilde{n}_F(X_1)\widetilde{n}_{X_1}(X_2)\cdots\widetilde{n}_{X_{k-1}}(X_k)\widetilde{n}_{X_k}(G) \tag{1}$$

where the second summation is over all non-isomorphic (F,G)-chains of length k. When $\operatorname{rank} F = 0$, we have

$$m_{\mathcal{F}}(F,G) = \widetilde{n}_F(G).$$

Let $\operatorname{rank} F = r$, and suppose that (1) holds for all graphs $F \in \mathcal{F}$ with rank less than r. The second assumption states that every copy of F has a unique maximal extension X, which implies that

$$\widetilde{n}_F(G) = \sum_X \widetilde{n}_F(X) m_{\mathcal{F}}(X,G),$$

where the sum is over all non-isomorphic $\mathcal{F}$-subgraphs X of G. This gives the expression

$$m_{\mathcal{F}}(F,G) = \widetilde{n}_F(G) - \sum_{X \ncong F} \widetilde{n}_F(X) m_{\mathcal{F}}(X,G).$$

In the summation, we can restrict to X for which $\widetilde{n}_F(X) > 0$. Such a graph X has rank at most $r-1$, so we may apply the induction hypothesis to rewrite each $m_{\mathcal{F}}(X,G)$-term into a double sum. The resulting triple sum can be simplified to obtain (1).

It now suffices to show that the right-hand side of (1) is reconstructible. To see this, we note that the inner summation is over (F,G)-chains for which X_k has size at most ℓ (since X_k is an $\mathcal{F}$-subgraph and by condition (i)), and so all such chains can be seen on cards. The remaining terms can be reconstructed by Kelly's Lemma (again using (i)), and this only requires the cards from $\mathcal{D}_\ell(G)$ that contain an $\mathcal{F}$-subgraph. ∎

Later in this paper, we will apply Lemma 9 with both $\mathcal{G}$ and $\mathcal{F}$ a family of trees. Since every connected subgraph of a tree is an induced subgraph, the lemma can be applied to count maximal induced $\mathcal{F}$-subgraphs.

2.3. Shared moments of sequences. We will need a bound on the maximum number of shared moments that two sequences $\alpha, \beta \in \{0, \dots, n\}^m$ can have. This result follows from the following theorem on the number of positive real roots of a polynomial. Here, we use log to mean the natural logarithm.

Theorem 10 ([8, Theorem A]): *Suppose that the complex polynomial*

$$p(z) := \sum_{j=0}^{n} a_j z^j$$

has k positive real roots (counted with multiplicity). Then

$$k^2 \le 2n \log\Big(\frac{|a_0| + |a_1| + \cdots + |a_n|}{\sqrt{|a_0 a_n|}}\Big).$$

This theorem is attributed to Schmidt, but the first published proof is due to Schur and a series of simplifications have followed (see [8]). We shall require a specific application of the theorem given by Borwein and Ingalls [9, Proposition 1]. We shall use the following formulation, which is tailored to our purposes.

LEMMA 11: *Let $\alpha, \beta \in \{0, \dots, n\}^m$ be sequences that are not related to each other by a permutation. If*

$$\binom{\alpha_1}{j} + \cdots + \binom{\alpha_m}{j} = \binom{\beta_1}{j} + \cdots + \binom{\beta_m}{j} \quad \text{for all } j \in \{0, \dots, \ell\}, \tag{2}$$

then $\ell + 1 \leq \sqrt{2n \log(2m)}$.

Proof. Since $\alpha_i, \beta_j \in \{0, \dots, n\}$ for all $i, j \in \{1, \dots, m\}$, the polynomial $p_{\alpha,\beta}$ defined by

$$p_{\alpha,\beta}(x) := \sum_{i=1}^{m} x^{\alpha_i} - \sum_{i=1}^{m} x^{\beta_i} \tag{3}$$

is of degree at most n. For $c \in \mathbb{C}$, let $\mathrm{mult}_c(p_{\alpha,\beta})$ denote the multiplicity of the root at c, or 0 if c is not a root of $p_{\alpha,\beta}$. We will show that

$$\ell + 1 \leq \mathrm{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n \log(2m)}.$$

Since α and β are not related by a permutation, the polynomial $p_{\alpha,\beta}$ is non-zero. We may write (with $r = \mathrm{mult}_0(p_{\alpha,\beta})$)

$$p_{\alpha,\beta}(x) = x^r \left(\sum_{j=0}^{n'} a_j x^j \right)$$

where a_0 and $a_{n'}$ are non-zero and $n' \leq n$. The coefficients are all integral, so $\sqrt{|a_0 a_{n'}|} \geq 1$. Moreover, from the definition of the polynomial in (3) there are at most $2m$ contributions of ± 1 to the coefficients, so we have

$$\sum_{i=0}^{n'} |a_i| \leq 2m.$$

By Theorem 10, the number of positive real roots of $\sum_{j=0}^{n'} a_j x^j$ is at most

$$\sqrt{2n' \log\left(\frac{|a_0| + |a_1| + \cdots + |a_{n'}|}{\sqrt{|a_0 a_{n'}|}} \right)} \leq \sqrt{2n \log(2m)}$$

and in particular, $\mathrm{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n \log(2m)}$. On the other hand, for all $j \in \{0, \dots, \ell\}$, equation (2) shows that

$$\left| \left(\frac{d}{dx^j} \left[\sum_{i=1}^{m} x^{\alpha_i} - \sum_{i=1}^{m} x^{\beta_i} \right] \right) \right|_{x=1} = \sum_{i=1}^{m} j! \binom{\alpha_i}{j} - \sum_{i=1}^{m} j! \binom{\beta_i}{j} = 0.$$

Hence, $\ell + 1 \leq \mathrm{mult}_1(p_{\alpha,\beta})$, and $\ell + 1 \leq \sqrt{2n \log(2m)}$ as desired. ∎

Condition (2) is equivalent to the condition that the first ℓ moments of α and β agree. To see this, observe that $\{x^i : i \in \{0,\dots,\ell\}\}$ and $\{\binom{x}{i} : i \in \{0,\dots,\ell\}\}$ both form a basis for the polynomials of degree at most ℓ. When α, β can be arbitrary integer sequences (instead of taking values in $\{0,\dots,n\}$) this variant is sometimes called the Prouhet-Tarry-Escott problem, and sequences are known with the first $\Omega(\sqrt{m})$ moments in common (see [9, Proposition 3] for a simple counting argument).

3. Reconstructing the degree sequence

The tools of the preceding section allow us to prove that the degree sequence of an n-vertex graph G can be reconstructed from the ℓ-deck of G whenever $\ell \geq \sqrt{2n\log(2n)}$. The proof is essentially identical to that given by Taylor [34], except for the use of the stronger bounds provided by Lemma 11.

Theorem 7: *The degree sequence of an n-vertex graph G can be reconstructed from $\mathcal{D}_\ell(G)$ for any $\ell \geq \sqrt{2n\log(2n)}$.*

Proof. Let G be an n-vertex graph with vertices $v_1,\dots,v_n$, and let $\ell \geq \sqrt{2n\log(2n)}$ be an integer. By Lemma 8, we can reconstruct the number of subgraphs of G isomorphic to the star $K_{1,j}$ for all $j \in \{2,\dots,\ell-1\}$. Since vertex v lies at the centre of $\binom{d(v)}{j}$ copies of $K_{1,j}$, we can compute the quantity

$$\widetilde{n}_{K_{1,j}}(G) = \sum_{v\in V(G)} \binom{d(v)}{j}$$

from the ℓ-deck. We can also reconstruct

$$\sum_{v\in V(G)} \binom{d(v)}{0} = n \quad \text{and} \quad \sum_{v\in V(G)} \binom{d(v)}{1} = 2\cdot|E(G)|$$

from the 2-deck. Write $\alpha_i = d(v_i)$ for $i \in [n]$ where we may assume

$$d(v_1) \leq \cdots \leq d(v_n).$$

Suppose, for a contradiction, that a graph with a different degree sequence $\beta_1 \leq \cdots \leq \beta_n$ gives the same counts. Then, for $j \in \{0,\dots,\ell-1\}$,

$$\sum_{i=1}^{n} \binom{\alpha_i}{j} = \sum_{i=1}^{n} \binom{\beta_i}{j}.$$

Since $\alpha, \beta \in \{0,\dots,n-1\}^n$ are not permutations of each other, Lemma 11 applies to show $\ell \leq \sqrt{2(n-1)\log(2n)}$ as desired. ∎

4. Recognising connectedness

In this section, we prove our theorem on reconstructing connectedness from the ℓ-deck. Recall that throughout this paper, an induced copy H' of H in some graph G refers to an induced subgraph of G that is isomorphic to H.

The main idea of the proof is that a graph G has a connected component isomorphic to some graph H if and only if it has an induced subgraph isomorphic to H 'without any neighbours'. By a similar approach to the previous section, when $|V(H)|$ is small we can actually compute the entire 'degree sequence', that is, for each k we can find the number of induced copies of H with k 'neighbours'. This will handle the case where G has a small component. If G has no small components, then either G is connected or only has medium-sized components, in which case we will recognise that it has no large connected subgraphs.

THEOREM 5: *The connectedness of an n-vertex graph G can be recognised from $\mathcal{D}_\ell(G)$ provided $\ell \geq 9n/10$.*

Proof. Let G be an n-vertex graph and let $\varepsilon = 1/10$, so our assumption is that $\ell \geq 9n/10 = (1-\varepsilon)n$. We begin by making an additional assumption on the size of n; it was shown by Kostochka, Nahvi, West and Zirlin [20] that the connectedness of a graph can be recognised from the $(n-3)$-deck for $n \geq 7$, so we can assume that $n \geq 39$.

Using Lemma 8 we can compute the number of connected subgraphs of G on ℓ vertices. If there are no such subgraphs, the graph must be disconnected and we are done. We may therefore assume that either G is connected, or its largest component has order at least ℓ. In particular, if G is not connected then it has a component of order at most $n-\ell$.

We will reconstruct all components that have at most $n-\ell$ vertices using the ℓ-deck. Let H be a connected graph with h vertices, where $1 \leq h \leq \varepsilon n$. Since $h \leq \ell$, we may compute $n_H(G)$ from the ℓ-deck by Lemma 8. Suppose $m = n_H(G) > 0$. Write $H_1, \ldots, H_m$ for the induced copies of H in G, and define the **neighbourhood** of H_i by

$$\Gamma(H_i) = \{v \in V(G) \setminus V(H_i) : vu \in E(G) \text{ for some } u \in H_i\}.$$

Define the **degree** of H_i to be $|\Gamma(H_i)|$, and denote it by α_i. Note that G has a component isomorphic to H if and only if $\alpha_i = 0$ for some $i \in [m]$. Thus, $(\alpha_1, \ldots, \alpha_m) \in \{0, \ldots, n-h\}^m$ determines the number of components isomorphic to H.

We now show that we can reconstruct $(\alpha_1, \dots, \alpha_m)$ up to permutation. Since $1 \le h \le \varepsilon n$ and $m \le \binom{n}{h} \le (\frac{en}{h})^h$, we have

$$\begin{aligned}\sqrt{2(n-h)\log(2m)} &\le \sqrt{2(n-h)h\log(en/h) + 2n\log 2}\\ &\le n\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2(\log 2)/n},\end{aligned}$$

where we have also used that $(n-h)h\log(en/h)$ is increasing in h within the given range. Hence, by Lemma 11, it suffices to show that we can reconstruct

$$\sum_{i=1}^{m} \binom{\alpha_i}{j} \quad \text{for all integers } 0 \le j \le N, \tag{4}$$

where

$$N = n\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2(\log 2)/n}.$$

For $j \ge 0$, let P_j denote the set of pairs of vertex sets (A, B) where $A \subseteq B \subseteq V(G)$, $G[A] \cong H$, $|B| = |A| + j$ and A is **dominating** in $G[B]$—that is, each vertex in $B \setminus A$ is adjacent to some vertex in A. Each $(A, B) \in P_j$ has some $i \in [m]$ for which $G[A] \cong H_i$ and B is contained in the neighbourhood of H_i, so

$$|P| = \sum_{i=1}^{m} \binom{\alpha_i}{j}.$$

For $j \ge 0$, let $\mathcal{H}_j$ denote the set of $(h+j)$-vertex graphs that consist of H along with j additional vertices, all of which are adjacent to at least one vertex in the induced copy of H (we include each isomorphism type once). If $(A, B) \in P$, then B corresponds to some $H' \in \mathcal{H}_j$. By definition, there are $n_{H'}(G)$ vertex sets $B \subseteq V(G)$ with $G[B] \cong H'$. Both $\mathcal{H}_j$ and H are known to us, so for each $H' \in \mathcal{H}_j$ we can calculate the number $n_{\mathrm{dom}}(H, H')$ of dominating induced copies of H in H'. Since

$$\sum_{H' \in \mathcal{H}_j} n_{\mathrm{dom}}(H, H') n_{H'}(G) = |P| = \sum_{i=1}^{m} \binom{\alpha_i}{j},$$

it only remains to show that we can determine $n_{H'}(G)$ from the ℓ-deck. We may use Lemma 8 to reconstruct $n_{H'}(G)$ if $|H'| = h + j \le \ell$. For $j \le N$ and $n \ge 39$, we find that

$$h + j \le \varepsilon n + N \le n - \varepsilon n \le \ell,$$

where the middle inequality follows from the fact that, using $\varepsilon = 1/10$, we have

$$\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2(\log 2)/39} \le 1 - 2\varepsilon.$$

This shows that we can reconstruct (4), and hence the number of components isomorphic to H. In particular, doing so for every graph H with at most $n-\ell$ vertices allows us to determine whether any component of G has at most $n-\ell$ vertices, which we saw would hold if and only if G is disconnected. ∎

We remark that the constant $9/10$ in the proof above can be improved slightly provided n is large enough. Indeed, the proof holds for any n and ε such that

$$\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon)+2(\log 2)/n} \leq 1-2\varepsilon,$$

and, for large enough n, we can take $\varepsilon \approx 0.1069$.

5. Reconstructing trees

We now work toward proving our main theorem on reconstructing trees, which we recall below.

THEOREM 4: *Any n-vertex tree T can be reconstructed from $\mathcal{D}_{n-r}(T)$ when $r < \frac{n}{9} - \frac{4}{9}\sqrt{8n+5} - 1$.*

The proof of Theorem 4 is spread across the following four subsections. First, we introduce a general technique for counting balls around a subgraph, which may be of independent interest. This strategy allows us to keep track of copies of fixed graphs in T that have a specified distinguished subgraph, which is a crucial ingredient of our proofs. This is done in Section 5.1.

In Section 5.2, we prove Theorem 6 which shows that the family of n-vertex trees is recognisable from the ℓ-deck when ℓ is in the assumed range. This allows us to proceed with the assumption that we have already recognised that every reconstruction from the deck is a tree.

The remaining parts contain the proof of reconstruction, which is split into two cases depending on whether or not the tree T contains a path that is long relative to the order of the graph n and the number ℓ of vertices on each card. Let the **length** of a path P be the number of edges in P, or equivalently $|V(P)|-1$. The **diameter** of a graph G is the maximum distance between two vertices in G, and for a tree T this is the same as the length of a longest path. When the diameter is less than about $\ell - 2n/3$, we can apply an argument based on reconstructing branches off the centre. For trees with diameter higher than this (in fact there is some overlap between the two cases), we will split the tree into two parts by removing a central edge, and then reconstruct these parts together with the information of how they glue together.

Having recognised that every reconstruction from the deck is a tree, the high diameter case is handled by the following lemma, which we prove in Section 5.3.

Lemma 12: *Let $\ell, k \in \{1, \dots, n\}$ with $k > 4\sqrt{\ell} + 2(n - \ell)$. If T is an n-vertex tree with diameter $k-1$, then T can be reconstructed amongst connected graphs from its ℓ-deck provided $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n+7} + \frac{11}{9}$.*

If T has low diameter, then we instead use the next lemma, which we prove in Section 5.4.

Lemma 13: *Let $\ell, k \in \{1, \dots, n\}$ with $k < \ell - \frac{2n+1}{3}$. If T is an n-vertex tree with diameter $k - 1$, then T is reconstructible amongst trees from its ℓ-deck.*

The proof of Theorem 4 then amounts to verifying that the condition on ℓ is sufficient to apply our result for recognising trees, and that our definitions of high and low diameter together cover the full range. The latter calculation is the source of the threshold on card size in the statement of Theorem 4.

Proof of Theorem 4. The conditions on ℓ and n imply that $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n+7} + \frac{11}{9}$. This bound on ℓ suffices to apply Theorem 6 in order to recognise that T is a tree.

Let k be the number of vertices in the longest path in T. When $k > 4\sqrt{\ell} + 2(n - \ell)$, T is reconstructible by Lemma 12 (amongst all connected graphs, without needing to know k).

So now suppose that $k \leq 4\sqrt{\ell} + 2(n - \ell)$. We will show that $k < \ell - \frac{2n+1}{3}$ as required to apply Lemma 13, and we note that in this case we can deduce the value of k from the ℓ-deck. After rearranging, it suffices to verify that $n - \ell < \frac{n-3k-1}{3}$. Our assumed condition that

$$\ell > \frac{8n}{9} + \frac{4}{9}\sqrt{8n+5} + 1$$

is equivalent to the condition

$$n - \ell < \frac{n - 12\sqrt{\ell} - 6(n - \ell) - 1}{3}.$$

Finally, note that

$$\frac{n - 12\sqrt{\ell} - 6(n - \ell) - 1}{3} \leq \frac{n - 3k - 1}{3}$$

for all $k \leq 4\sqrt{\ell} + 2(n - \ell)$. Thus, we can apply Lemma 13 in this case to reconstruct T. ∎

5.1. Counting extensions. Given a graph H, we define an H**-extension** to be a pair $H_{\text{ext}} = (H^+, A)$ where H^+ is a graph and $A \subseteq V(H^+)$ is a subset of vertices with $H^+[A] \cong H$. The idea is that H^+ may contain multiple induced copies of H, so we are picking out one in particular. One could think of an H-extension as a triple (H^+, A, H), but we suppress H since $H \cong H^+[A]$.

The **order** of $H_{\text{ext}} = (H^+, A)$ is

$$|H_{\text{ext}}| = |V(H^+)|.$$

We will usually work with H-extensions in a setting where H is an induced subgraph of an ambient graph G, and in this case a natural family of H-extensions can be obtained by considering neighbourhoods. Specifically, for $d \in \mathbb{N}$, the (**closed**) d**-ball** of an induced subgraph H of a graph G is defined by

$$B_d(H,G) = G[\{v \in V(G) : d_G(v,H) \leq d\}].$$

That is, $B_d(H,G)$ is the subgraph induced by the set of vertices of distance at most d from H, including the vertices of H itself. It is useful to view the d-ball of H as the H-extension $(B_d(H,G), V(H))$.

Two H-extensions (G_1, A_1) and (G_2, A_2) are **isomorphic** if there is a graph isomorphism $\varphi : G_1 \to G_2$ with $\varphi(A_1) = A_2$. In addition, we say that an H-extension (H^+, A) is a **sub-**H**-extension** of (H^{++}, B) if H^+ is an induced subgraph of H^{++} and $A = B$.

Let $m_d(H_{\text{ext}}, G)$ be the number of induced copies of H in G whose d-ball is isomorphic (as an H-extension) to H_{ext}. The purpose of the notation above is to be able to formalise this notion, which intuitively boils down to counting how often H appears with a particular neighbourhood.

Our key counting result for extensions states that it is possible to reconstruct $m_d(H_{\text{ext}}, G)$ from the ℓ-deck provided the d-balls of all induced copies of H are small enough to appear on the cards as proper subgraphs.

Lemma 14: *Let $\ell, d \in \mathbb{N}$ and let G be a graph on at least $\ell+1$ vertices. Let H be a graph on at most $\ell-1$ vertices. From the ℓ-deck of G, it is possible to recognise whether the d-ball of every induced copy of H in G has fewer than ℓ vertices. If this is the case, then for every H-extension H_{ext} the quantity $m_d(H_{\text{ext}}, G)$ is determined by the ℓ-deck.*

Proof. We first define the set of 'potential d-balls around H'. Let $\mathcal{H}$ denote the set of graphs H^+ such that $|V(H^+)| \leq \ell$ and there is an induced copy H' of H in H^+ such that all the vertices of H^+ are at distance (in H^+) at most d from H'. These represent all possible d-balls of H with at most ℓ vertices, and the ones that appear in G will be a subset of these. Note that in order for H^+ to belong to $\mathcal{H}$, it is not necessary (nor guaranteed) that all induced copies of H in H^+ satisfy the above distance condition, rather only that there is at least one such induced copy.

For any $H^+ \in \mathcal{H}$, we can reconstruct $n_{H^+}(G)$ from the ℓ-deck using Lemma 8. The d-balls of every induced copy of H have fewer than ℓ vertices if and only if $n_{H^+}(G) = 0$ for every $H^+ \in \mathcal{H}$ with $|H^+| = \ell$, and we can tell if this is the case. Suppose that the d-balls around every induced copy of H do indeed have fewer than ℓ vertices and set

$$k = \max\{|V(H^+)| : H^+ \in \mathcal{H},\ n_{H^+}(G) > 0\}.$$

For a fixed $H^+ \in \mathcal{H}$ with $|V(H^+)| = k$, we observe that every induced copy H' of H for which $B_d(H', H^+) \cong H^+$ also satisfies $B_d(H', G) \cong H^+$ by the maximality of k and the definition of $\mathcal{H}$.

Let $\mathcal{H}_{\text{ext}}$ denote the set of isomorphism classes of H-extensions (H^+, A) with $H^+ \in \mathcal{H}$. By the preceding observation, if $H_{\text{ext}} = (H^+, A) \in \mathcal{H}_{\text{ext}}$ with $|H^+| = k$, then the number of induced copies of H in G whose d-balls are isomorphic to H_{ext} is the number of induced copies of H^+ in G multiplied by the number of induced copies of H in H^+ whose d-ball in H^+ is isomorphic to H_{ext} (as an H-extension). That is,

$$m_d(H_{\text{ext}}, G) = n_{H^+}(G) m_d(H_{\text{ext}}, H^+). \tag{5}$$

Both of these quantities are reconstructible from the ℓ-deck, so we are done in this case.

If $|V(H^+)| < k$, then the d-ball in G of a copy of H may be strictly larger than H^+ and formula (5) does not apply. This can be corrected by subtracting the number of induced copies of H in H^+ for which H^+ is not the d-neighbourhood of that induced copy of H in G. To count these, we select in turn each 'maximal' d-neighbourhood of size at least $|H^+|+1$, and subtract 1 from the relevant count for each H^+ that it contains. Any leftover H^+ that have not been accounted for must then be maximal.

Explicitly, for $H'_{\mathrm{ext}} \in \mathcal{H}_{\mathrm{ext}}$ distinct from H_{ext}, let $n(H_{\mathrm{ext}}, H'_{\mathrm{ext}})$ give the number of sub-H-extensions of H'_{ext} isomorphic to H_{ext}. We claim that

$$m_d(H_{\mathrm{ext}}, G) = n_{H^+}(G) m_d(H_{\mathrm{ext}}, H^+) - \sum_{\substack{H'_{\mathrm{ext}} \in \mathcal{H}_{\mathrm{ext}} \\ |H'_{\mathrm{ext}}| > |H_{\mathrm{ext}}|}} n(H_{\mathrm{ext}}, H'_{\mathrm{ext}}) m_d(H'_{\mathrm{ext}}, G).$$

When $|H_{\mathrm{ext}}| = k$, this formula agrees with (5). The terms $m_d(H_{\mathrm{ext}}, H^+)$, $n(H_{\mathrm{ext}}, H'_{\mathrm{ext}})$ and the domain of the summation are already known to us, and we can reconstruct $n_{H^+}(G)$ for all $H^+ \in \mathcal{H}$ using Kelly's Lemma. Moreover, we may assume that we have reconstructed the terms $m_d(H'_{\mathrm{ext}}, G)$ for $|H'_{\mathrm{ext}}| > |H_{\mathrm{ext}}|$ by induction with base case $|H_{\mathrm{ext}}| = k$, so verifying the formula will complete the proof.

The term $n_{H^+}(G) m_d(H_{\mathrm{ext}}, H^+)$ at the start of the formula counts the number of pairs $(A, B) \subseteq V(G) \times V(G)$ such that

- $G[B]$ is an induced copy of H^+,
- $A \subseteq B$ and $G[A]$ is an induced copy of H (that is, $(G[B], A)$ is an H-extension),
- B is a subset of the d-ball around A (i.e., $B \subseteq B_d(G[A], G)$).

Informally, each fixed B has exactly $m_d(H_{\mathrm{ext}}, H^+)$ sets A with which it is in a pair, and there are $n_{H^+}(G)$ sets B to count.

Compared to $m_d(H_{\mathrm{ext}}, G)$, the term $n_{H^+}(G) m_d(H_{\mathrm{ext}}, H^+)$ overcounts by 1 for each pair (A, B) with $B \subsetneq B_d(G[A], G)$. Thus, it just remains to verify that the number of pairs with $B \neq B_d(G[A], G)$ is given by

$$\sum_{|H'_{\mathrm{ext}}| > |H_{\mathrm{ext}}|} n(H_{\mathrm{ext}}, H'_{\mathrm{ext}}) m_d(H'_{\mathrm{ext}}, G).$$

To see that this is true, note that by definition the correction term counts triples (A, B, C) with $A \subseteq B \subsetneq C \subseteq V(G)$ such that

- $G[A]$ is an induced copy of H,
- $G[B]$ is an induced copy of H^+,
- $G[C] \cong B_d(G[A], G)$.

Each pair (A, B) with $B \neq B_d(G[A], G)$ is in a unique such triple, namely with $C = V(B_d(G[A], G))$; if $B = B_d(G[A], G)$, then no suitable C with $B \subsetneq C$ can be found. ∎

As an aside, we mention that by setting $d = 1$ and considering the H-extension $(H, V(H))$ in Lemma 14, one can count the number of components isomorphic to H.

Corollary 15: *Let H and G be graphs with $|V(H)| \leq \ell - 1$ and $n = |V(G)|$. If there is no induced copy of H in G for which $|B_1(H, G)| \geq \ell$, then we can reconstruct the number of components of G isomorphic to H from $\mathcal{D}_\ell(G)$.*

5.2. Recognising trees. This section contains the proof of Theorem 6, which is an application of the extension-counting result established of Section 5.1.

Theorem 6: *For $\ell \geq (2n+4)/3$, the class of trees on n vertices is recognisable from the ℓ-deck.*

Proof. Let G be a graph and suppose we are given $\mathcal{D}_\ell(G)$. By Kelly's Lemma (Lemma 8), we can reconstruct the number m of edges provided $\ell \geq 2$. Hence, we may suppose that $m = n - 1$, otherwise we can already conclude that G is not a tree. It suffices to show that we can determine whether G contains a cycle, or equivalently to determine whether G is connected.

If G has a cycle of length at most ℓ, then the entire cycle will appear on a card and we can conclude that G is not a tree. We may therefore assume that every cycle in G has length greater than ℓ. If the graph does not contain a connected card, then the graph cannot be a tree, and so we may assume that there is a connected card and the largest components in G have at least ℓ vertices each. Since $\ell \geq (2n+4)/3$, there is only one component A with at least ℓ vertices and the other components have at most $\ell - 1$ vertices.

Let $d = \lceil \ell - n/2 - 1 \rceil$. For a vertex $x \in V(G)$, denote the d-ball around x in G by $B_d(x)$. Using Lemma 14 with H being the graph consisting of a single vertex, we find that either there is an $x \in V(G)$ with d-ball of order at least ℓ or we can reconstruct the collection of d-balls (with 'distinguished' centres).

Suppose firstly that there exists $x \in V(G)$ such that $|B_d(x)| \geq \ell$. We claim that then G is a tree. Assume towards a contradiction that there is a cycle in G. Since this must have more than ℓ vertices, any cycle in G must be contained in the largest component A (the smaller components have order at most $\ell - 1$). Let C be a shortest cycle in A. Similarly, note that $x \in A$ since otherwise the d-ball around x cannot have ℓ vertices. If $|B_d(x) \cap V(C)| \leq 2d + 1$, then

$$|B_d(x)| \leq n - |V(C) \setminus B_d(x)| \leq n - (\ell + 1) + (2d + 1) \leq \ell - 1$$

by our choice of d. Thus, $B_d(x) \cap V(C)$ contains at least $2d+2$ vertices. Choose two vertices $c_1, c_2 \in B_d(x) \cap V(C)$ joined by a subpath C' of C (possibly C' is a single edge) such that C' does not contain any other vertex of $B_d(x)$. Let C'' be the other path from c_1 to c_2 in C. This must contain at least $2d$ other vertices of $B_d(x) \cap C$, so C'' is a path of length at least $2d + 1$. However, there is also a path P from c_1 to c_2 in the d-ball around x of length at most $2d$, and this intersects C' only at the endpoints c_1 and c_2. Replacing the path C'' with the path P forms a cycle which is strictly shorter than C, giving a contradiction. Hence, G cannot have any cycles and must be a tree.

We may now assume that we can reconstruct the collection of d-balls and will show how to recognise whether the graph is connected in this case. In any component of order at most $n - \ell$, there must be some vertex x such that the distance from x to any vertex in the same component is at most $(n - \ell)/2$. By our choice of ℓ and d,

$$\frac{n-\ell}{2} \leq \ell - \frac{n}{2} - 2 \leq d - 1.$$

Thus, if there is a component of order at most $n - \ell$ (which happens if and only if G is not a tree), then there must be a d-ball with radius at most $d - 1$. Conversely, if we discover such a d-ball, then we know that the graph is disconnected since the d-ball must form a component due to its radius, yet has at most $\ell - 1$ vertices. Hence, G is a tree if and only if all d-balls have radius d. This shows that we can recognise connectedness and completes the proof. ∎

5.3. High diameter. The main result in this section is Lemma 12, which states that trees containing sufficiently long paths are reconstructible amongst connected graphs from their ℓ-decks. Our approach is based on the key property that trees with high diameter have small 1-balls around induced copies of subgraphs obtained by deleting a well-chosen edge. This is made precise within the conditions of Lemma 16, which essentially gives a reconstruction algorithm for graphs (not just trees) when this property is assumed.

Let us first develop the intuition behind our strategy using trees. Removing an edge from a tree T splits T into two components, and our goal will be to recognise a pair of graphs (R, R^c) which are the components left after removing an edge from T. However, it is not enough to know that T is formed by connecting R and R^c with an edge: we also need to know which vertices the edge is connected to, and we will actually look for pairs for which we can also deduce this.

We are specifically interested in induced subgraphs that are connected to the rest of the graph by a single edge, which leads us to consider induced copies of R (and R^c) with this property. For a graph H, let a **leaf H-extension** be a pair $H_{\text{ext}} = (H^+, A)$ where

- H^+ is obtained by adding a single vertex connected by a single edge to a vertex of H, and
- $A \subset V(H^+)$ is such that $H^+[A] \cong H$.

This is a special case of the extensions defined in Section 5.1. We will refer to the additional edge added to H to form H^+ as the **extending edge**. Note that if R is a component of $T - e$, then the 1-ball of $T[V(R)]$ in T is a leaf R-extension, but there may be multiple (non-isomorphic) leaf R-extensions in T.

The extra edge in a leaf extension indicates where to glue, so we would be done if we could identify two leaf extensions $C_{\text{ext}} = (C^+, V_C)$ and $D_{\text{ext}} = (D^+, V_D)$ for which the vertex set of G is the disjoint union of V_C and V_D. We demonstrate in Lemma 16 a case where this can be done from $\mathcal{D}_\ell(G)$ using the counts of the relevant leaf extensions obtained by Lemma 14. Lemma 16 is not specialised to trees (we still require connectedness but the R and R^c that we are looking for do not need to be acyclic), so the final step to proving Lemma 12 is to show that trees with high diameter satisfy the conditions of Lemma 16.

We say an edge e in a connected graph G is a **bridge** if the graph $G - e$ obtained from G by removing the edge e is disconnected.

LEMMA 16: *Let G be a connected graph with a bridge e, and let R and R^c be the components of $G-e$. If G has no induced subgraph H isomorphic to R or R^c with $|V(B_1(H,G))| \geq \ell$, then G is the only connected graph up to isomorphism with the deck $\mathcal{D}_\ell(G)$.*

Proof. We prove the lemma by describing an algorithm that takes in the deck $\mathcal{D}_\ell(G)$ of a connected graph G, and either returns a connected graph or a failure. We will show that if the algorithm returns a graph G', then G' must be isomorphic to G. The condition in the hypothesis that G has a suitable bridge e and corresponding R and R^c (which are all initially unknown) is only used to show that the algorithm will definitely output a graph.

The idea of the procedure is to create a finite list of candidate graphs guaranteed to contain both components of $G-e$, and then test all pairs of such graphs glued together in every way that could feasibly reconstruct G. This latter step is refined by using leaf extensions to indicate how these gluings occur. The key point is to show that we can identify when such a construction actually produces G and then terminate.

Given any connected graph H on at most $\ell-1$ vertices and a deck $\mathcal{D}_\ell(G)$, we can check directly from the cards whether there is an induced copy H' of H in G for which $|V(B_1(H',G))| \geq \ell$. Say that a graph H is **confined** if no such copy of it exists. For every confined connected graph H and every leaf H-extension H_{ext} of H, we can apply Lemma 14 to reconstruct $m_1(H_{\text{ext}},G)$. Recall that this is the number of induced copies of H in G whose 1-ball in G is obtained by adding a pendant vertex connected at a specified vertex, so a positive value would signal an extension that might correspond to a component of $G-e$ (with the extending edge corresponding to the bridge). To form our collection of candidates, let $\mathcal{H}_{\text{ext}}$ denote the set of isomorphism classes of all leaf H-extensions H_{ext} for which $m_1(H_{\text{ext}},G)>0$ and H is a confined connected graph.

We now consider all pairs $(C_{\text{ext}},D_{\text{ext}})$ of elements from $\mathcal{H}_{\text{ext}}$ for which

$$|C_{\text{ext}}|+|D_{\text{ext}}|=n+2 \quad \text{and} \quad |C_{\text{ext}}|\leq|D_{\text{ext}}|.$$

Let

$$C_{\text{ext}}=(C^+,V_C) \quad \text{and} \quad D_{\text{ext}}=(D^+,V_D),$$

where $C=C^+[V_C]$ and $D=D^+[V_D]$ denote the corresponding labelled subgraphs. Let $N(C_{\text{ext}},D_{\text{ext}})$ be the number of induced copies of C in D whose 1-ball in D^+ is an induced copy of C^+. That is, we count the induced copies of C^+ in D^+ where the extending edge of D^+ is either unused or is the extending edge of C^+. If $m_1(C_{\text{ext}},G)>N(C_{\text{ext}},D_{\text{ext}})$, then the algorithm terminates and outputs the graph G' formed by taking disjoint copies of C^+ and D^+ and identifying their extending edges as given by the extensions. If $m_1(C_{\text{ext}},G)\leq N(C_{\text{ext}},D_{\text{ext}})$, we continue on to the next pair of elements of $\mathcal{H}_{\text{ext}}$. If we have checked every suitable pair of elements from $\mathcal{H}_{\text{ext}}$ without outputting a graph, then we terminate with a failure.

Let us first verify that if the algorithm returns a graph, it must be isomorphic to G. We will later use our assumptions on G to argue that the algorithm does output a graph when the input is $\mathcal{D}_\ell(G)$, which shows that G is reconstructible.

It is useful to highlight that every leaf extension $D_{\text{ext}} = (D^+, V_D)$ with $m_1(D_{\text{ext}}, G) > 0$ has a unique partner that we will denote by $\overline{D_{\text{ext}}}$, which is the leaf extension that produces a graph isomorphic to G when joined with D_{ext} as described above: the condition $m_1(D_{\text{ext}}, G) > 0$ together with the connectedness of G imply that D is a component of the graph obtained by deleting the extending edge e_D, and its partner comes from extending the other component (with vertex set $V(G) \setminus V_D$) by the same edge e_D. Note that $|D_{\text{ext}}| + |\overline{D_{\text{ext}}}| = n+2$, where the vertices of e_D are counted twice.

Suppose that algorithm outputs the graph G' and that we terminated with $(C_{\text{ext}}, D_{\text{ext}})$, so this is a pair which produces G'. We know that $m_1(D_{\text{ext}}, G) > 0$ because $D_{\text{ext}} \in \mathcal{H}_{\text{ext}}$, so D_{ext} has a unique partner $\overline{D_{\text{ext}}}$ (unknown to us) and it suffices to show that $\overline{D_{\text{ext}}} \cong C_{\text{ext}}$ as leaf extensions. We first claim that if an induced copy of C contributes to $m_1(C_{\text{ext}}, G)$ (by definition of m_1, this means that its 1-ball in G is isomorphic as a C-extension to C_{ext}), then it cannot contain the extending edge e_D of D_{ext}. Note that $|C_{\text{ext}}| \leq |D_{\text{ext}}|$ by assumption in our algorithm, and $|C_{\text{ext}}| + |D_{\text{ext}}| = n + 2 = |\overline{D_{\text{ext}}}| + |D_{\text{ext}}|$ which implies that $|C_{\text{ext}}| = |\overline{D_{\text{ext}}}|$. Thus, an induced copy of C (which has size $|C_{\text{ext}}| - 1$) that contains e_D cannot fully contain either V_D or $V(G) \setminus V_D$. Since G is connected, the 1-ball of such an induced copy must then add at least one vertex from each of V_D and $V(G) \setminus V_D$, so it does not contribute to $m_1(C_{\text{ext}}, G)$, as claimed. It follows that

$$m_1(C_{\text{ext}}, G) = N(C_{\text{ext}}, D_{\text{ext}}) + N(C_{\text{ext}}, \overline{D_{\text{ext}}}). \tag{6}$$

In order for the algorithm to have terminated with $(C_{\text{ext}}, D_{\text{ext}})$ we must have $m_1(C_{\text{ext}}, G) > N(C_{\text{ext}}, D_{\text{ext}})$, so from (6) we see that $N(C_{\text{ext}}, \overline{D_{\text{ext}}}) \geq 1$. This, together with the fact that $|C_{\text{ext}}| = |\overline{D_{\text{ext}}}|$, implies that $C_{\text{ext}} \cong \overline{D_{\text{ext}}}$ as leaf extensions.

Finally, let us argue that the algorithm does terminate when the input is the ℓ-deck of a graph G satisfying the assumptions of the lemma. Let e be a bridge as in the hypothesis of the lemma, and let the components of $G - e$ be R and R^c. From (6) we see that $m_1(R_{\text{ext}}, G) = N(R_{\text{ext}}, \overline{R_{\text{ext}}}) + 1$, where $R_{\text{ext}}, \overline{R_{\text{ext}}}$ are the (unique) leaf-extensions of R and R^c. By the assumption that G has no induced subgraph H isomorphic to R or R^c with $|V(B_1(H,G))| \geq \ell$, both R_{ext} and $\overline{R_{\text{ext}}}$ are in $\mathcal{H}_{\text{ext}}$. We are therefore guaranteed to be able to find at least one pair, namely $(R_{\text{ext}}, \overline{R_{\text{ext}}})$, amongst our candidates that will lead to termination. ∎

We remark that the only place where we used the existence of an edge e that splits G into "nice" components R and R^c was to ensure that the algorithm output a graph. One can try to use the algorithm to reconstruct graphs whenever the deck is known to correspond to a connected graph, and the algorithm will either output the graph, or a failure (in which case one needs a different approach).

With the preceding lemma in hand, the proof of Lemma 12 boils down to showing that any tree with large enough diameter (depending on both n and ℓ) does have a bridge which splits the tree into "nice" components, and so satisfies the conditions of Lemma 16.

Lemma 12: *Let $\ell, k \in \{1,\dots,n\}$ with $k > 4\sqrt{\ell} + 2(n-\ell)$. If T is an n-vertex tree with diameter $k-1$, then T can be reconstructed amongst connected graphs from its ℓ-deck provided $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n+7} + \frac{11}{9}$.*

Proof. Fix $k, \ell \in [n]$ with $k > 4\sqrt{\ell} + 2(n-\ell)$ and $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n+7} + \frac{11}{9}$. Let T be a tree and suppose that a longest path in T contains exactly k vertices. We wish to show that T has a suitable bridge that satisfies the assumptions of Lemma 16 so that we can conclude it is reconstructible amongst connected graphs from its ℓ-deck.

Fix a longest path in T with k vertices. Let R and S be the rooted subtrees obtained from T by removing the central edge of the path if k is even, or one of the two central edges if k is odd (and rooting the subtrees at the vertex which had an incident edge removed). This S plays the role of R^c in Lemma 16, and since T is unknown, both R and S are also initially unknown. By Lemma 16, if T has no induced subgraph H isomorphic to R or S with $|V(B_1(H,T))| \geq \ell$, then we can recognise that this is the case and reconstruct T amongst connected graphs from $\mathcal{D}_\ell(T)$. We assume, in order to derive a contradiction, that T contains an induced copy S' of S with $|V(B_1(S',T))| \geq \ell$. Note that, since R contains at least $n-\ell+2\sqrt{\ell}-1$ vertices, S contains at most $\ell-2$ vertices and the 1-ball of S contains at most $\ell-1$ vertices.

Set $r = n-\ell$ and fix an isomorphism $\varphi : S \to S'$. Let P_0 be a path in R containing at least $(k-1)/2$ vertices and starting at the root of R, so $k \leq 2|P_0|+1$. We will proceed by iteratively building a sequence $(P_i)_{i=1}^{j}$ of vertex-disjoint paths in S to obtain a lower bound on $|S|$. Since $|R|+|S| = n$ and $|P_0| \leq |R|$, this leads to an upper bound on $|P_0|$ and hence on k that will contradict our initial assumption that $k > 4\sqrt{\ell} + 2(n-\ell)$.

To build the first path in the sequence, consider the intersection of S' with the path P_0. Since $V(S') \neq V(S)$, this intersection must be non-empty, and it must be connected since both T and S are trees, so S' and P_0 intersect on a subpath Q_0. We are assuming that $B_1(S', T)$ has at least ℓ vertices (that is, T has at most r vertices outside $B_1(S', T)$), and $B_1(S', T)$ has at most 2 vertices on P_0 outside Q_0, so altogether $|V(Q_0)| \geq |V(P_0)| - r - 2$.

Now let P_1 be the path $\varphi^{-1}(V(Q_0))$ in S and note that P_1 is vertex-disjoint from P_0, since P_0 is contained in R. Define Q_1 to be the intersection of S' with P_1, which is again a path. As before, T has at most r vertices outside of $B_1(S', T)$, and $B_1(S', T)$ has at most two vertices on P_i but outside Q_i for each $i = 1, 2$. Thus, we have $|V(Q_0)| + |V(Q_1)| \geq |V(P_0)| + |V(P_1)| - r - 4$. Since $|V(Q_0)| = |V(P_1)|$, we conclude that $|V(Q_1)| \geq |V(P_0)| - r - 4$.

We continue to iteratively build our sequence of paths P_i, together with the sequence of subpaths Q_i restricted to S', as follows: given P_i and Q_i, let $P_{i+1} := \varphi^{-1}(V(Q_i))$ and set $Q_{i+1} = P_{i+1} \cap S'$ (see Figure 2). We first note that P_{i+1} is disjoint from $P_0, \ldots, P_i$. Indeed, since P_0 is contained in R, it is clear that P_{i+1} cannot intersect P_0. If P_{i+1} intersects an earlier path P_j with $j \geq 1$, then a vertex in $P_{i+1} \cap P_j$ would be mapped by φ into $Q_i \cap Q_{j-1}$, which is contained in $P_i \cap P_{j-1}$. Hence, the paths are disjoint by induction. By the finiteness of T, we must eventually reach a j such that

$$|V(Q_{j-1})| = |V(P_j)| = 0.$$

At this point, we have disjoint paths $P_1, \ldots, P_j$ in S that satisfy

$$|V(P_i)| = |V(Q_{i-1})| \geq |V(P_0)| - r - 2i \quad \text{for all } i = 1, \ldots, j.$$

In particular, setting $i = j$ to use the fact that $|V(P_j)| = 0$ shows that $j \geq (|V(P_0)| - r)/2$. We may then calculate

$$\begin{aligned} |V(S)| &\geq |V(P_1)| + \cdots + |V(P_j)| \\ &\geq \sum_{i=1}^{\lfloor (|V(P_0)| - r)/2 \rfloor} (|V(P_0)| - r - 2i) \\ &= \left\lfloor \frac{|V(P_0)| - r}{2} \right\rfloor \left\lceil \frac{|V(P_0)| - r - 2}{2} \right\rceil \\ &\geq \frac{(|V(P_0)| - r)(||V(P_0)| - r - 2)}{4}. \end{aligned}$$

Since $|V(S)| \leq n - |V(P_0)|$, we must have $|V(P_0)| \leq \sqrt{4n-4r+1} + r - 1$ and

$$k \leq 2|V(P_0)| + 1 \leq 2\sqrt{4n-4r+1} + 2r - 1.$$

Finally, note that $2\sqrt{x+1} - 1 \leq 2\sqrt{x}$ for all $x \geq 1$ to find $k \leq 4\sqrt{\ell} + 2r$, a contradiction.

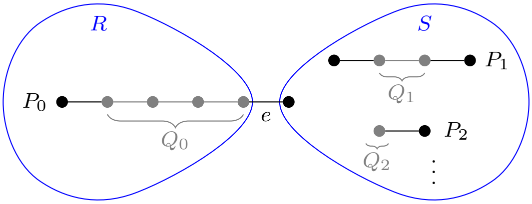

Figure 2. The start of a sequence of paths formed by the iterative process in the proof of Lemma 12.

The same argument shows that T has no induced copy R' of R for which $|V(B_1(R',T))| \geq \ell$. Hence, by Lemma 16 we can reconstruct T from $\mathcal{D}_\ell(T)$.

5.4. Low diameter. The purpose of this section is to prove Lemma 13. Since this section only considers trees, the 'number of copies' is always the same as the 'number of induced copies'. For readability, we will count copies instead of induced copies, but the reader can insert the word 'induced' everywhere if desired.

We will show that any tree T with diameter $k-1$ can be reconstructed from its ℓ-deck for any $\ell \in [n]$ such that

$$n - \ell < \frac{n-3k+1}{3} \quad \text{if } k \text{ is odd}$$

or

$$n - \ell < \frac{n-3k-1}{3} \quad \text{if } k \text{ is even,}$$

which together imply the statement directly. These conditions are equivalent to $k < \ell - \frac{2n-1}{3}$ when k is odd and $k < \ell - \frac{2n+1}{3}$ when k is even. The reason for the dependence on the parity is that, broadly, our strategy for reconstruction is to separately reconstruct branches of the tree emanating from its centre: if k is odd, the **centre** of T is the vertex in the middle of each longest path, and if k is even, the **centre** consists of the two middle vertices. The former case is easier to work with so when k is even, we subdivide the central edge and reuse the argument from when k is odd, making the even case slightly weaker. This reduction is explained in the final proof of this section.

Our motivation for using the centre of a tree is that it is unique and it does not depend on the choice of longest path. When the diameter of T is small enough to identify whether individual cards contain a longest path, we can pinpoint the centre of T on each of these cards and use this as an anchor for reconstruction.

Let us assume for the majority of this section that T is a tree with n vertices, the number k of vertices in a longest path in T is odd, and $k < \ell - \frac{2n-1}{3}$. This means that $k + 1 \leq \ell$ so we can reconstruct k from the ℓ-deck, which we shall use freely, and that T has a unique central vertex.

Given a vertex $u \in T$ with neighbours $v_1, v_2, \ldots, v_a$, let the **branches** at u be the rooted subtrees $B_1, B_2, \ldots, B_a$ where B_i is the component of $T - u$ that contains v_i, rooted at v_i. An **end-rooted path** is a path rooted at an endvertex of the path. In this section, all longest paths P_k will be rooted at the central vertex c, and are hence not end-rooted, whilst all of the shorter paths mentioned will be end-rooted. Given two rooted trees T_1 and T_2 with roots u and v respectively, let $T_1 \frown T_2$ denote the (unrooted) tree given by adding an edge between u and v (see Figure 3).

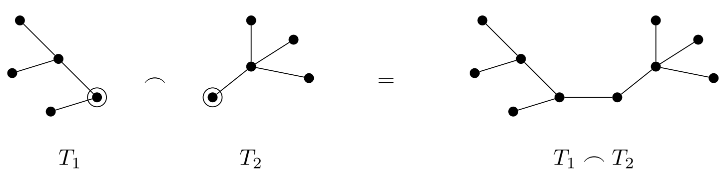

Figure 3. An example of the operation $T_1 \frown T_2$.

By restricting our attention to the cards that have diameter $k - 1$, we may assume that we can always identify the centre of the graph. Our basic strategy is to reconstruct the branches at the centre separately, knowing that we can later join them together using the centre as a common point of reference. This can be done via a counting argument when all branches at the centre have at most $\ell - k$ vertices, but when one branch is 'heavy' and contains many (at least $\ell - k$) of the vertices, a slightly more finicky version of the argument is required. This is because such a branch cannot be seen on a single card containing a longest path that is disjoint to it. It is possible to recognise these cases from the ℓ-deck.

We first address the simpler situation without heavy branches to illustrate the method.

LEMMA 17: *If T is a tree with even diameter $k-1$ for which every branch at the centre has fewer than $\ell - k$ vertices, then T is reconstructible from the subset of the ℓ-deck consisting only of cards which contain a copy of P_k.*

Proof. Let c be the central vertex of T, and let $\mathcal{B} = \{B_1, \ldots, B_a\}$ be the branches at c that we wish to reconstruct. If one of the branches at c has at least $\ell - k$ vertices, then there must be a card containing a longest path with a branch of at least $\ell - k$ vertices (the branch and the path need not be disjoint, but their union contains at most ℓ vertices). Thus we can recognise from the ℓ-deck that all branches in $\mathcal{B}$ have fewer than $\ell - k$ vertices.

We first reconstruct all branches that are not end-rooted paths. For any fixed B (of size at most $\ell - k$) which is a rooted tree but not an end-rooted path, we will use Lemma 9 to count each branch at c isomorphic to B once for every P_k in T. Dividing this number, denoted N_B, by the number $n_{P_k}(T)$ of copies of P_k in T then tells us the multiplicity of B in T (which may be zero). Note that $n_{P_k}(T)$ can be determined using the proof of Kelly's Lemma, the fact that $k < \ell$ and the observation that $n_H(C) = 0$ whenever C does not contain a copy of P_k.

Our main goal now is to reconstruct N_B. We will determine N_B in two parts. Let π_B be the number of pairs consisting of one copy B' of B that is a branch at c, and one copy P_k' (rooted at its centre as usual) of a longest path that is disjoint from B'. Similarly, let τ_B count pairs (B', P_k') where the copy P_k' of P_k intersects B'. It is clear that $N_B = \pi_B + \tau_B$.

We begin with π_B. Let $\mathcal{G}$ be the family of all n-vertex trees with diameter $k-1$ and where all branches from the centre have fewer than $\ell-k$ vertices. Let $\mathcal{F}$ be the family of graphs of the form $P_k \frown S$, where S is a non-empty rooted tree with less than $\ell-k$ vertices that is not an end-rooted path and P_k is rooted at its central vertex (see Figure 4). Fix $G \in \mathcal{G}$ and consider some $F \in \mathcal{F}$. If $P_k' \frown S'$ is a copy of F in G, then it is contained in a unique maximal $\mathcal{F}$-subgraph, namely P_k' together with the unique branch B' containing S'. Note that this would not be true if end-rooted paths were allowed since $P_k' \frown S'$ might then also be contained in a different maximal $\mathcal{F}$-subgraph $P_k'' \frown B''$, where S' is contained in P_k'' and B'' is a branch that contains half of the original P_k'. Also, since B' has fewer than $\ell-k$ vertices, these maximal elements have fewer than ℓ

vertices and are therefore in $\mathcal{F}$. Note also that our deck contains all cards which contain an $\mathcal{F}$-subgraph, since each $\mathcal{F}$-subgraph contains a copy of P_k. Thus, by Lemma 9, we can reconstruct the number of $\mathcal{F}$-maximal copies of F in G for each $F \in \mathcal{F}$. This is non-zero for $F = P_k \frown S$ if and only if $\pi_S \neq 0$.

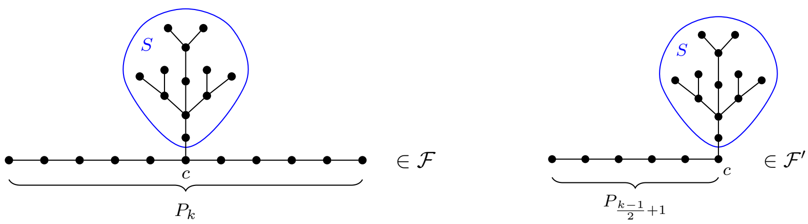

Figure 4. Elements of $\mathcal{F}$ and $\mathcal{F}'$.

In fact, the number of $\mathcal{F}$-maximal copies of $F = P_k \frown B$ is exactly π_B. To see this, consider a particular copy B' of B that occurs as a branch and observe that F occurs as a maximal $\mathcal{F}$-subgraph with this B' as the copy of B once for every longest path in the tree which avoids B'.

There is a similar argument to determine τ_B. Keeping $\mathcal{G}$ as before, let $\mathcal{F}'$ be the family of graphs of the form $P_{(k+1)/2} \frown S$ where S is a rooted tree that contains an end-rooted $P_{(k-1)/2}$ but is not itself an end-rooted path. Again, an element $F = P_{(k+1)/2} \frown S$ is $\mathcal{F}'$-maximal when S is an entire branch, and for any $G \in \mathcal{G}$ and $F \in \mathcal{F}'$, we can reconstruct the number of $\mathcal{F}'$-maximal copies of F in G by Lemma 9. This time there is at least one $\mathcal{F}'$-maximal copy of $F = P_{(k+1)/2} \frown S$ if and only if G has a branch isomorphic to S (although we do not need to use both directions explicitly).

Let $m_{F'}$ be the number of $\mathcal{F}'$-maximal copies of F' formed as $P_{(k+1)/2} \frown B$ in T, which we can reconstruct as argued above. A particular copy B' of B that occurs as a branch contributes 1 to $m_{F'}$ for each copy of $P_{(k+1)/2}$ that starts at the central vertex c and is disjoint from B'. Thus, letting $n_{P^\bullet}(B)$ be the number of end-rooted copies of $P_{(k+1)/2}$ in B' with root at the root of B (this is the same for any copy of B and does not depend on the deck), one can construct all of the copies of longest paths that intersect B' by gluing together one $P_{(k+1)/2}$ from inside B' and one that is disjoint from it. Doing so for every copy of B shows that we can reconstruct $\tau_B = m_{F'} \cdot n_{P^\bullet}(B)$. The number of copies of B that occur as a branch at c can then be reconstructed as

$$\frac{N_B}{n_{P_k}(T)} = \frac{\pi_B + \tau_B}{n_{P_k}(T)}.$$

It remains to determine the number of branches isomorphic to an end-rooted path P_j, which we will do using the fact that we know all of the other branches not of this form. Initially, let $\widetilde{T}$ be the tree obtained by gluing all of the branches that we have found so far at a single vertex c. In case $\widetilde{T}$ does not yet have a copy of P_k, update $\widetilde{T}$ by attaching up to two end-rooted paths of length $(k-1)/2$ at c (add the smallest number necessary for $\widetilde{T}$ to have at least one P_k). We will identify and glue in the remaining path branches in T that are missing from $\widetilde{T}$ in decreasing order of length, so start by setting j to be $(k-1)/2$: the maximum possible length of a path branch. Let S_j denote the graph obtained from P_k by adding a path of length j to its central vertex.

We can count the number of copies of S_j in T using the proof of Kelly's Lemma, as they only appear on cards that contain a longest path. We can also count the number of copies of S_j in $\widetilde{T}$ directly as, although $\widetilde{T}$ is growing and can have more than ℓ vertices, we have really constructed $\widetilde{T}$ and do not need to refer to the cards to look at it. If there are more copies of S_j in T than in the current $\widetilde{T}$, then there must be at least one more end-rooted P_j as a branch. We then update $\widetilde{T}$ by gluing in this new path branch at c, and repeat this step with the same j. If the counts match, then reduce j by 1 and continue iteratively. By handling the different path lengths in decreasing order, we avoid overcounting shorter paths that are contained in unknown longer paths. Once $j = 0$, we terminate and output $\widetilde{T}$. At this point, we have reconstructed all branches and the final $\widetilde{T}$ is exactly T. ∎

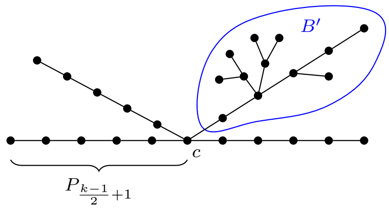

Figure 5. A tree containing three longest paths that avoid B' (so $\pi_B = 3$), and three longest paths that use B' consisting of a $P_{(k+1)/2}$ outside B' and a $P_{(k-1)/2}$ inside (so $\tau_B = 3$).

We now consider the case where one of the branches at the centre of T has at least $\ell - k$ vertices. This is so many that we can find a card showing all the other branches at the centre in their entirety, which then reduces the problem to reconstructing the large branch. In order to do this, we will move the "centre" one step inside the branch and continue doing this until no branch at the new centre is too big. This collection of branches can be reconstructed by essentially applying the proof of the previous lemma with minor modifications. Importantly, the condition that T has small diameter ensures that we do not have to take too many steps away from the true centre.

The following lemma sets up for this process. We shall call a branch **i-heavy** if it contains at least $\ell - k - i$ vertices (a **heavy** branch is 0-heavy), and say it is **outward** if it does not contain the centre of the tree. When we wish to talk about a branch at a vertex within a specific card, we will call it a **partial** branch to emphasise that it need not be a branch of T. Recall that $r := n - \ell$.

LEMMA 18: *Let T be a tree with even diameter $k-1$ (where $k < \ell - \frac{2n-1}{3}$) and central vertex c, and suppose we are given exactly the cards in $\mathcal{D}_\ell(T)$ that contain a copy of P_k. For any $0 \le i \le (k-1)/2$,*

- (i) *each vertex can have at most one i-heavy branch;*
- (ii) *there is at most one vertex c_i at distance i from c with an i-heavy outward branch;*
- (iii) *we can recognise whether there is a vertex c_i at distance i from c with an i-heavy outward branch;*
- (iv) *if there is such a c_i, then we can find a card among those we are given on which we can identify c_i and the root of its i-heavy branch, and all smaller branches at c_i are present in their entirety. In particular, we can completely determine the isomorphism classes of all of these smaller branches.*

Proof. Since $i \le \frac{k-1}{2}$ and $k < \ell - \frac{2n-1}{3}$ by assumption, we first deduce that

$$\ell - k - i \ge \ell - \frac{3k-1}{2} > \ell - \frac{3(\ell - \frac{(2n-1)}{3}) - 1}{2} = \frac{2n-\ell}{2} > \frac{n}{2}.$$

This proves (i), as the branches at a vertex are pairwise disjoint. Similarly, if two distinct vertices c_i and c_i' are both at distance i from c, then the only branch at c_i that can share a vertex with a branch at c_i' is that containing c. Thus, the previous calculation also proves (ii). For (iii), suppose that T does have a vertex c_i at distance i from c with an i-heavy branch B not containing c. The subtree formed by taking a copy of P_k together with the path of length i from c to c_i and any $(\ell - k - i)$-vertex subtree of B containing the root has at most $k + i + (\ell - k - i) = \ell$ vertices. Note that there may be some overlap between the vertices in the copy of P_k, the vertices in the path from c to c', and the vertices in the subtree of B, and this subgraph may have less than ℓ vertices. However, there will still be a card C with ℓ vertices which has a subtree of this form. It follows that T has a vertex c_i at distance i from c with an i-heavy branch if and only if it has a card containing a subtree of the aforementioned form.

Assuming that there exist c_i and B as above, we claim that the desired card in (iv) can be found as follows: from among the cards we have (all with a copy of P_k so we can identify c), take a connected card C in which the maximum number of vertices in any partial outward branch at any vertex with distance i from c is as small as possible. There are only $r + k + i$ vertices not in B, so C must still see at least $\ell - r - k - i$ vertices of B. On the other hand, every other partial branch at c_i has at most $r+k+i$ vertices, which is less than $\ell - k - i - r$ since

$$r + k + i \leq n - \ell + k + \frac{k-1}{2} < \frac{2n - 2\ell + 3(\ell - \frac{2n-1}{3}) - 1}{2} = \frac{\ell}{2}.$$

This means that we can identify the vertex c_i as the unique (by (i) and (ii)) vertex at the correct distance from c with a partial outward branch of size at least $\ell - k - i - r$, and the root of this partial branch is the root of the i-heavy branch in T. Moreover, by the minimality of the count used to select C, all other partial branches at c_i must actually be present in their entirety; that is, they are isomorphic to the non-i-heavy branches at c_i in T. ∎

LEMMA 19: *If T is a tree with even diameter $k - 1$ (where $k < \ell - \frac{2n-1}{3}$) for which the centre c has a branch containing at least $\ell - k$ vertices, then T is reconstructible from the subset of the ℓ-deck consisting only of cards which contain a copy of P_k.*

Proof. With $i = 0$ in Lemma 18, we can recognise whether there is a branch at c with at least $\ell - k$ vertices, and there is at most one such branch.

Starting with $c_0 = c$, we construct a sequence $c_0, c_1, c_2, \ldots$ of vertices to act as new "centres" until we reach a vertex c_j whose branches are all small enough for us to apply Lemma 9.

For the first step, let c_1 be the root of the 0-heavy branch, which is adjacent to c. Applying Lemma 18 with $i = 1$, we can recognise whether any neighbour of c has a 1-heavy outward branch. If not, then the branches at c_1 all have less than $\ell - k - 1$ vertices and we terminate with $j = 1$. Else if there is such a 1-heavy outward branch, then it follows from statement (ii) of the lemma that it must be at c_1. In addition, statement (iv) allows us to determine all but the 1-heavy branch at c_1.

Now set c_2 to be the vertex in the 1-heavy branch that is adjacent to c_1 and iterate as follows. In the ith step we check if there is a vertex at distance i from c with an i-heavy outward branch. If there is not, then all the outward branches at c_i have less than $\ell - k - i$ vertices and we terminate with $j = i$. Otherwise, there is only one such vertex and this must be c_i. Set c_{i+1} to be the root of the unique i-heavy outward branch at c_i and completely determine all of the smaller branches at c_i. The fact that we can do this is guaranteed by Lemma 18 provided $i \leq (k-1)/2$. To see that this condition holds, we note that our procedure builds a path in T with one endvertex at c. Since each step increases the length of this path by 1 and the longest path in T contains k vertices, we can take at most $(k-1)/2$ steps before terminating.

Suppose the process terminates at the jth step, where $j \leq (k-1)/2$. The remainder of the argument closely follows the proof of Lemma 17. Let $\mathcal{G}$ be the family of n-vertex trees with diameter $k-1$.

Let B be a rooted tree which is a potential branch for c_j (so of size at most $\ell - k - j$). We wish to determine the number of outward branches at c_j isomorphic to B, which will reconstruct T. We first consider the case in which B is not a path.

Again, we start by computing "branches hanging off a central path". To be precise, π_B is the number of pairs (S, P) where $S \subseteq V(T)$ induces an outward branch isomorphic to B at a vertex at distance j from c, $P \subseteq V(T)$ induces a path of length k and $P \cap S = \emptyset$.

Let $\mathcal{F}$ be the family of graphs that can be constructed as follows. Let $i \in \{0, \ldots, j-1\}$, let $v_1, \ldots, v_k$ be the vertices in a copy of P_k and let $u_1, \ldots, u_{j-i}$ be the vertices in a (disjoint) copy of P_{j-i}. A graph in $\mathcal{F}$ is formed by adding an edge from u_1 to $v_{\frac{k+1}{2}+i}$, and then attaching a rooted tree S which is not an end-rooted path to the vertex u_{j-i}. The condition that the attached tree is not a path ensures that it is easy to distinguish the copy of P_k and the added tree in any $\mathcal{F}$-graph. Two examples are given in Figure 6.

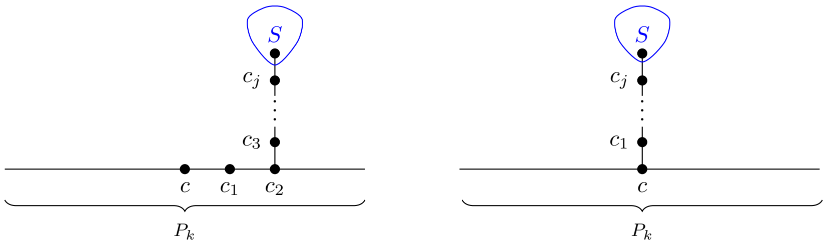

Figure 6. Potential elements of $\mathcal{F}$ along with their ‘moving centres’.

Each $\mathcal{F}$-subgraph of $G \in \mathcal{G}$ is contained in a unique maximal $\mathcal{F}$-subgraph, given by extending the tree attachment to the whole of the relevant branch at u_{j-i}. Applying Lemma 9 allows us to determine the number of occurrences of each maximal $\mathcal{F}$-subgraph, as we did in the proof of Lemma 17. Here we use the fact that there is no vertex at distance j from c with a j-heavy outward branch, and so indeed all sought-after subgraphs fit on the cards. Note that indeed we count each branch once per longest path which is disjoint from it.

Next, we count “each branch once per longest path intersecting it”. To be precise, let σ_B denote the number of pairs (S, P) where $S \subseteq V(T)$ is the vertex set of an outward branch isomorphic to B at a vertex at distance j from c, and P is the vertex set of a copy of P_k with $P \cap S \neq \emptyset$.

Let $\mathcal{F}'$ be the family of graphs of the form $P_{(k-1)/2+j+1} \frown S$ where S is a rooted tree that contains an end-rooted $P_{(k-1)/2-j}$ but is not itself an end rooted path. An element $F = P_{(k-1)/2+j+1} \frown S$ is $\mathcal{F}'$-maximal when S is the entire outward branch, and we can reconstruct the number of $\mathcal{F}'$-maximal copies of each F in G using Lemma 9 as in Lemma 17. We obtain $\tau(B)$ by multiplying the number of $\mathcal{F}'$-maximal graphs (for $S = B$) by the number of end-rooted paths contained in B.

The number $n_{P_k}(T)$ of P_k in T can again be obtained using the proof of Kelly’s Lemma. The total number of outward branches from vertices at distance j from c isomorphic to B is given by $\frac{\pi_B+\tau_B}{n_{P_k}(T)}$. This includes all the outward branches at c_j, but also outward branches at other vertices. However, we have already reconstructed all of the tree except for the outward branches at c_j, so we can subtract the counts for all the outward branches not at c_j from the total: the remainder must be attached at c_j.

Finally, we reconstruct the end-rooted paths attached at c_j using a similar argument to that at the end of the proof of Lemma 17. Build the subtree $\widetilde{T}$ of T (rooted at c_j) that includes all of the branches reconstructed thus far: the only parts of T missing from $\widetilde{T}$ are the outward branches of c_j that are end-rooted paths. We will add end-rooted paths of length i to c_j in $\widetilde{T}$, starting with $i = \frac{k-1}{2} - j$: the maximum possible path length of an outward branch at c_j as the tree has diameter $k-1$.

Let H_0 be the smallest subgraph of $\widetilde{T}$ that contains c_j and a path of length k. Then let H_i be the graph obtained from H_0 by adding an edge between a path on i vertices and the vertex c_j in H_0. We compute the number of subgraphs isomorphic to H_i in T using the proof of Kelly's lemma, and in $\widetilde{T}$ by inspection. If the count in T is the same as the count in $\widetilde{T}$, then we decrease i by 1, terminating once $i = 0$ with the current $\widetilde{T}$. If the counts are not the same, we add a path of length i to the root vertex c_j in $\widetilde{T}$. Our procedure adds path branches from the longest possible length to the shortest so that our counts are not inflated by subpaths of longer paths, meaning the discrepancy in counts can only arise from path branches of c_j in T that are missing in $\widetilde{T}$. At the end of this procedure, we have reconstructed T as $\widetilde{T}$. ∎

LEMMA 13: *Let $\ell, k \in \{1, \ldots, n\}$ with $k < \ell - \frac{2n+1}{3}$. If T is an n-vertex tree with diameter $k-1$, then T is reconstructible amongst trees from its ℓ-deck.*

Proof. If k is odd, then by Lemma 19 we can reconstruct T from its ℓ-deck provided $k < \ell - \frac{2n-1}{3}$, which is slightly better than the bound claimed in the statement.

Suppose that k is even. This means that there is a central edge instead of a central vertex, but this is only a minor inconvenience. Indeed, let T' be formed from T by subdividing the central edge, and consider a new partial deck $\mathcal{D}'$ formed by subdividing the central edge in every card in $\mathcal{D}_\ell(T)$ which contains a longest path, and discarding cards which do not contain a longest path. Then $\mathcal{D}'$ is the subdeck of the $(\ell+1)$-deck of the tree T' consisting of the cards containing a longest path in T'. Note that the tree T' has $k+1$ vertices in a longest path and that $k+1 < \ell + 1 - \frac{2(n+1)-1}{3}$ by our choice of k and ℓ.

If a branch at the centre c of T' has size at least $\ell - k$ (which we can recognise), then we are done by Lemma 19. If not, then we are done by Lemma 17. ∎

6. Conclusion

The example in Figure 1 shows that Nýdl's lower bound of $\ell \geq \lfloor n/2 \rfloor + 1$ for the ℓ such that no two non-isomorphic trees have the same ℓ-deck is not sharp for $n = 13$. However, it is still the best known lower bound for all other values of n. It may well be the case that Conjecture 3 is asymptotically true, or even true exactly for large enough n.

Problem 1: Is there a function $\ell(n) = (1/2 + o(1))n$ such that all n-vertex trees can be reconstructed from their $\ell(n)$-deck?

For the problem of reconstructing the degree sequence, let $\ell = \ell(n)$ be the smallest integer such that the degree sequence of every n-vertex graph can be reconstructed from the ℓ-deck. We have shown in Theorem 7 that $\ell(n) \leq \sqrt{2n \log(2n)} + 1$. It is easy to obtain a lower bound of the form $\ell(n) = \Omega(\sqrt{\log n})$: indeed, each ℓ-vertex graph appears at most $\binom{n}{\ell}$ times in the ℓ-deck, so there are at most $(n^\ell)^{2^{\ell^2}}$ possible ℓ-decks. There are $\Omega(4^n/n)$ possible degree sequences as determined by Burns [11], and hence we need $2^{\log_2(n)\ell 2^{\ell^2}} \geq 2^{2n - \log_2(n)}$, which implies the bound. By considering restricted graph classes, this can be improved slightly, but it would be interesting to see whether the lower bound can be improved to n^ε for some $\varepsilon > 0$.

In a different direction, it would be interesting to determine how large ℓ needs to be in order to recognise the k-colourability of a graph on n vertices from its ℓ-deck. A special case of a result of Tutte [36] from 1979 states that the chromatic number of a graph is reconstructible when $\ell = n - 1$, but nothing more is known in the direction of taking smaller cards. An interesting starting point would be to pinpoint the threshold for recognising whether a graph is bipartite (2-colourable). In this case, a lower bound of $\lfloor n/2 \rfloor$ follows from the example of Spinoza and West [33] mentioned in the introduction (consider a path and the disjoint union of an odd cycle and a path). Manvel [27] proved that the $(n-2)$-deck suffices, but it seems likely that it should be possible to determine bipartiteness when a linear number of vertices are removed. More generally, for fixed k, it may even be true that k-colourability is recognisable from the cn-deck for some $c = c(k) < 1$.

Acknowledgements. We would like to thank Doug West and the anonymous reviewer for their helpful suggestions.

References

[1] T. Andreae, *On the reconstruction of locally finite trees*, Journal of Graph Theory **5** (1981), 123–135.

[2] T. Andreae, *On reconstructing rooted trees*, Journal of Combinatorial Theory. Series B **62** (1994), 183–198.

[3] B. Bollobás, *Almost every graph has reconstruction number three*, Journal of Graph Theory **14** (1990), 1–4.

[4] J. A. Bondy, *On Kelly's congruence theorem for trees*, Mathematical Proceedings of the Cambridge Philosophical Society **65** (1969), 387–397.

[5] J. A. Bondy, *A graph reconstructor's manual*, in *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series, Vol. 166, Cambridge University Press, cambridge, 1991, pp. 221–252.

[6] J. A. Bondy and R. L. Hemminger, *Reconstructing infinite graphs*, Pacific Journal of Mathematics **52** (1974), 331–340.

[7] J. A. Bondy and R. L. Hemminger, *Graph reconstruction: a survey*, Journal of Graph Theory **1** (1977), 227–268.

[8] P. Borwein, T. Erdélyi and G. Kós, *Littlewood-type problems on* $[0, 1]$, Proceedings of the London Mathematical Society **79** (1999), 22–46.

[9] P. Borwein and C. Ingalls, *The Prouhet–Tarry–Escott problem revisited*, L'Enseignement Mathématique **40** (1994), 3–27.

[10] N. Bowler, J. Erde, P. Heinig, F. Lehner and M. Pitz, *A counterexample to the reconstruction conjecture for locally finite trees*, Bulletin of the London Mathematical Society **49** (2017), 630–648.

[11] J. M. Burns, *The number of degree sequences of graphs*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2007.

[12] Z. A. Chernyak, *Some additions to an article by B. Manvel: "Some basic observations on Kelly's conjecture for graphs"*, Vestsi Akadèmii Navuk Belarusi. Seryya Fizika-Matèmatychnykh Navuk **126** (1982), 44–49.

[13] W. B. Giles, *The reconstruction of outerplanar graphs*, Journal of Combinatorial Theory. Series B **16** (1974), 215–226.

[14] W. B. Giles, *Reconstructing trees from two point deleted subtrees*, Discrete Mathematics **15** (1976), 325–332.

[15] D. L. Greenwell and R. L. Hemminger, *Reconstructing the* n*-connected components of a graph*, Aequationes Mathematicae **9** (1973), 19–22.

[16] F. Harary and E. Palmer, *The reconstruction of a tree from its maximal subtrees*, Canadian Journal of Mathematics **18** (1966), 803–810.

[17] F. Harary, A. J. Schwenk and R. L. Scott, *On the reconstruction of countable forests*, l'Institut Mathématique. Publications **13(27)** (1972), 39–42.

[18] P. J. Kelly, *On isometric transformations*, Ph.D. thesis, University of Wisconsin, Madison, WI, 1942.

[19] P. J. Kelly, *A congruence theorem for trees.*, Pacific Journal of Mathematics **7** (1957), 961–968.

[20] A. V. Kostochka, M. Nahvi, D. B. West, and D. Zirlin, *Degree lists and connectedness are 3-reconstructible for graphs with at least seven vertices*, Graphs and Combinatorics **36** (2020), 491–501.

[21] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin, *3-regular graphs are 2-reconstructible*, European Journal of Combinatorics **91** (2021), Article no. 103216

[22] A. V. Kostochka, M. Nahvi, D. B. West, and D. Zirlin, *Acyclic graphs with at least* $2\ell+1$ *vertices are* ℓ*-recognizable*, Journal of Graph Theory **109** (2025), 184–200.

[23] A. V. Kostochka and D. B. West, *On reconstruction of graphs from the multiset of subgraphs obtained by deleting* ℓ *vertices*, IEEE Transactions on Information Theory **67** (2021), 3278–3286.

[24] J. Lauri, *The reconstruction of maximal planar graphs*, Journal of Combinatorial Theory. Series B **30** (1981), 196–214.

[25] J. Lauri, *Proof of Harary's conjecture on the reconstruction of trees*, Discrete Mathematics **43** (1983), 79–90.

[26] B. Manvel, *Reconstruction of trees*, Canadian Journal of Mathematics **22** (1970), 55–60.

[27] B. Manvel, *Some basic observations on Kelly's conjecture for graphs*, Discrete Mathematics **8** (1974), 181–185.

[28] V. Müller, *Probabilistic reconstruction from subgraphs*, Commentationes Mathematicae Universitatis Carolinae **17** (1976), 709–719.

[29] W. Myrvold, *The ally-reconstruction number of a tree with five or more vertices is three*, Journal of Graph Theory **14** (1990), 149–166.

[30] C. St. J. A. Nash-Williams, *Reconstruction of infinite graphs*, Discrete Mathematics **95** (1991), 221–229.

[31] V. Nýdl, *A note on reconstructing of finite trees from small subtrees*, Acta Universitatis Carolinae. Mathematica et Physica **31** (1990), 71–74.

[32] V. Nýdl, *Finite undirected graphs which are not reconstructible from their large cardinality subgraphs*, Discrete Mathematics **108** (1992), 373–377.

[33] H. Spinoza and D. B. West, *Reconstruction from the deck of* k*-vertex induced subgraphs*, Journal of Graph Theory **90** (2019), 497–522.

[34] R. Taylor, *Reconstructing degree sequences from* k*-vertex-deleted subgraphs*, Discrete Mathematics **79** (1990), 207–213.

[35] C. Thomassen, *Reconstructing* 1*-coherent locally finite trees*, Commentarii Mathematici Helvetici **53** (1978), 608–612.

[36] W. T. Tutte, *All the king's horses (a guide to reconstruction)*, in *Graph Theory and Related Topics*, Academic Press, New York–London, 1979. pp. 15–33.

[37] S. M. Ulam, *A Collection of Mathematical Problems*, Interscience Tracts in Pure and Applied Mathematics, Vol. 8, Interscience, New York–London, 1960.