Reconstruction from smaller cards

Carla Groenland* Tom Johnston† Alex Scott†‡ Jane Tan†

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.

1 Introduction

Throughout this paper, all graphs are finite and undirected with no loops or multiple edges. Given a graph G 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 of graphs such as planar graphs and graphs of bounded maximum degree. 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).

*Delft Institute of Applied Mathematics, Technische Universiteit Delft, 2628 CD Delft, Netherlands, `c.e.groenland@tudelft.nl`.

†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
`tom.johnston@bristol.ac.uk, scott@maths.ox.ac.uk, jane.tan@maths.ox.ac.uk`

‡Supported by EPSRC grant EP/V007327/1.

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 all induced subgraphs of G on ℓ vertices (in this notation $\mathcal{D}(G) = \mathcal{D}_{n-1}(G)$). A graph or graph parameter is *reconstructible from the ℓ-deck* if it is determined by 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, as can be shown by a simple counting argument (see Lemma 8). Thus, if a graph is reconstructible from its ℓ'-deck then it is reconstructible from its ℓ-deck for all $\ell \geq \ell'$. The main question is then to determine the threshold; that is, to determine the smallest ℓ for which a given class of graphs or a property is reconstructible from the ℓ-deck.

Reconstruction from small cards is generally attributed to Kelly, although the strengthening of the Reconstruction Conjecture that follows seems to be formulated by Manvel (who calls it "Kelly's Conjecture").

Conjecture 1 ([19, 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 $N_1 = 3$. This stronger conjecture did not receive much attention until 1974 when it was studied by Manvel [27], who 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 (that is, if G and H are graphs with n vertices and the same $(n-2)$-deck, then either both graphs or neither belong to the class). 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 [20]) and that almost all graphs are reconstructible from $\binom{r+2}{2}$ specially chosen cards 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 such as the class of trees. In fact, Nýdl conjectured in 1990 that no two non-isomorphic trees have the same ℓ-deck when ℓ is slightly larger than $n/2$.

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

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

The conjectured bound 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 in [31]. Indeed, the best previous result is an earlier bound 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. 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 can be reconstructed from their $(n-2)$-deck for all $n \geq 6$.

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 values $4 \leq n \leq 25$, and it remains open for large n.

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, Harary and Palmer [16] showed how to recover a tree using only the cards that are subtrees, Bondy [4] showed that only the cards where peripheral vertices have been removed are needed and Manvel [26] subsequently showed that the set (as opposed to the multiset) of cards that are trees suffices (except in four cases). Lauri [25] also showed that trees with at least three cutvertices can be reconstructed (amongst all graphs) from the cards corresponding to removing a cutvertex. 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 [17, 6, 35, 1, 30, 2]), 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 [21] 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

$\mathcal{D}_\ell(G_1) = \mathcal{D}_\ell(G_2)$ for all $\ell \le \lfloor n/2 \rfloor$. However, G_1 is connected and G_2 is not. In light of this construction, Spinoza and West believe that for $n \ge 6$ and $\ell \ge \lfloor n/2 \rfloor + 1$, 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

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

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 $\ell \ge 9n/10$.*

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

Theorem 6. *For $\ell \ge (2n+4)/3$, 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 $\ell \ge \lfloor n/2 \rfloor + 1$, 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 purpose of reconstructing trees, we have retained it to keep the proof of our main result self-contained.

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 for connectedness. Chernyak [12] showed that the degree sequence of an n-vertex graph can be reconstructed from its $(n-2)$-deck for $n \ge 6$, and this was later extended by Kostochka, Nahvi, West, and Zirlin [21] to the $(n-3)$-deck for $n \ge 7$. 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 reconstructed from $\mathcal{D}_\ell(G)$ for any $\ell \ge \sqrt{2n\log(2n)}$.*

In Section 2, we give ℓ-deck versions of both Kelly's Lemma [19] 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). Our main result on reconstructing trees, Theorem 4, is proved in Section 5 which also contains a new counting tool for reconstruction that may be of independent interest as well as the proof of Theorem 6. 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

Let $\widetilde{n}_H(G)$ and $n_H(G)$ denote the number of subgraphs and induced subgraphs of G isomorphic to H respectively. We will reserve the word *copy* of H for an induced subgraph isomorphic to H, and say an *instance* of H to mean not necessarily induced.

In the classical graph reconstruction problem, Kelly's Lemma states that we can reconstruct $n_H(G)$ and $\widetilde{n}_H(G)$ provided $|V(H)| < |V(G)|$, and there are many variants of the lemma for other reconstruction problems (see [5]). We use the following variant.

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.*

In particular, Kelly's Lemma means that $\mathcal{D}_{\ell'}(G)$ can be reconstructed from $\mathcal{D}_\ell(G)$ for all $\ell' \leq \ell$.

Despite its great usefulness, the proof of Kelly's Lemma requires nothing more than elementary counting. Suppose we count the number of copies of H in each of the ℓ-cards of G, and take the sum over all cards. Each copy of H in G will be counted exactly $\binom{n-|V(H)|}{\ell-|V(H)|}$ times toward this total. Hence, we can reconstruct the number $n_H(G)$ of 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 instances rather than copies. Foreshadowing later usage, we remark that Kelly's Lemma only requires the subset of the deck consisting of the cards which contain at least one copy of the fixed graph H.

2.2 Counting maximal 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 $V(F') \subsetneq V(F'')$.

Let $m(F,G)$ denote the number of maximal $\mathcal{F}$-subgraphs in G which are isomorphic to F. 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 $m(F,G)$ from the ℓ-deck.

Lemma 9. *Let $n \in \mathbb{N}$, let $\ell \in [n-1]$ and 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 $m(F, G)$ from the collection of cards in the ℓ-deck that contain an $\mathcal{F}$-subgraph.

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.

Proof. Define an (F, G)-*chain* of length k to be a sequence $(X_0, \ldots, X_k)$ 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

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

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

$$\widetilde{n}_F(G) = \sum_X \widetilde{n}_F(X) m(X, G),$$

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

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

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 $m(X, G)$-term into a double sum. The resulting triple sum can be simplified to obtain (1).

It now suffices to show that the RHS 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. □

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. We use log to mean the natural logarithm here.

Theorem 10 (Theorem A in [8])**.** *Suppose that the complex polynomial*

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

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

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

This theorem is attributed to Schmidt, but the first published proof is due to Schur and a series of simplifications have followed (see [8]). The specific application that we require was 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 two 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 $i, j \in [m]$,

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

is a polynomial 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 $j \in \{0, \ldots, \ell\}$, 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. $\square$

Condition (2) is equivalent to the condition that the first ℓ moments of α and β agree. To see this, observe that $\{x^i : i \in \{0, \ldots, \ell\}\}$ and $\{\binom{x}{i} : i \in \{0, \ldots, \ell\}\}$ 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 $\{0, \ldots, n\}$) 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 reconstructed from $\mathcal{D}_\ell(G)$ for any $\ell \geq \sqrt{2n\log(2n)}$.*

Proof. Let G have vertices $v_1, \ldots, v_n$, 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 $j \in \{2, \ldots, \ell-1\}$. 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

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

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 different degree sequence $\beta_1 \leq \cdots \leq \beta_n$ gives the same counts. Then, for $j \in \{0, \ldots, \ell-1\}$,

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

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

4 Recognising connectedness

In this section, we prove our theorem on reconstructing connectedness from the ℓ-deck. Recall that throughout this paper, a 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 on strictly less vertices than G, 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'. So we are done if G has a small component. But if G has no small components, then it is either connected or only has medium-sized components (in which case we recognize that it has no large connected subgraphs and we are done).

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 [21] 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 count 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 G is either 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 from 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, reconstructing the sequence $(\alpha_1, \ldots, \alpha_m) \in \{0, \ldots, n-h\}^m$ determines the number of components isomorphic to H.

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

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

where we also have 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

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

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

Let P 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 $(A, B) \in P$ 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 $j \geq 0$, 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 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 $G[B] \cong H'$. Since $\mathcal{H}_j$ and H are known to us, for each $H' \in \mathcal{H}_j$ we can calculate the number $n(H, H')$ of dominating copies of H in H'. Since

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

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 $|H'| = h+j \leq \ell$. For $j \leq N$ and $n \geq 39$, we find that

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

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

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

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. $\square$

We remark that the constant $9/10$ can be improved slightly in the proof above 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 address the recognition problem and prove Theorem 6.

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 cases), we will split the tree into two parts by removing a central edge, and then recognising these parts and how to glue them back together.

Having recognised that every reconstruction of 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 $\ell, k \in [n]$ 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 following lemma which we prove in Section 5.4.

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

The proof of Theorem 4 then amounts to verifying that the assumptions are sufficient for recognition, 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. Let k be the number of vertices in the longest path in T. The conditions on ℓ and n imply that $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n+7} + \frac{11}{9}$. This allows us to recognise that T is a tree by Theorem 6, and moreover that T is reconstructible by Lemma 12 when $k > 4\sqrt{\ell} + 2(n-\ell)$. We show that the remaining k satisfy the condition in Lemma 13. It suffices to verify that $n - \ell < \frac{n-3k-1}{3}$. The right hand side is decreasing in k, and now $k \le 4\sqrt{\ell}+2(n-\ell)$, so Lemma 13 applies provided

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

which is equivalent to our assumed condition

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

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 copies of H as induced subgraphs, so we are picking out one in particular. 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

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

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 $\varphi(A_1) = A_2$. Let $m_d(H_{\text{ext}}, G)$ be the number of copies of H in G whose d-ball is isomorphic (as an H-extension) to H_{ext}. 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 $A = B$.

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 copies of H are small enough to appear on the cards.

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, and if this is the case, the quantity $m_d(H_{ext}, G)$ is determined by the ℓ-deck for any H-extension H_{ext}.*

Proof. Let $\mathcal{H}$ denote the set of graphs H^+ such that $|V(H^+)| \le \ell$ and there is a copy H' of H in H^+ in which 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 it is not necessary (nor guaranteed) that all copies of H in H^+ satisfy the above distance condition, rather only that there is at least one such 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 the $n_{H^+}(G) = 0$ for every $H^+ \in \mathcal{H}$ with $|H^+| = \ell$, and we can tell if this is the case. Suppose that indeed the d-balls around every induced copy of H 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 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 copies of H whose d-balls are isomorphic to H_{ext} is the number of copies of H^+ in G times the number of copies of H in H^+ whose d-ball in H^+ is isomorphic to H_{ext} (as H-extensions). 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 of H may be strictly larger than H^+ and the formula (5) does not apply. This can be corrected by subtracting the number of $H \subseteq H^+$ for which H^+ is not the d-neighbourhood of that copy of H in G. To count these, we select each 'maximal' d-neighbourhood in turn, and subtract one from the relevant count for each strictly smaller H^+ that it contains. Any leftover H^+ that have not been accounted for must then be maximal.

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

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

Note that when $|H_{\text{ext}}| = k$, the formula above agrees with (5). The terms $m_d(H_{\text{ext}}, H^+)$ $n(H_{\text{ext}}, H'_{\text{ext}})$ 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 $m_d(H'_{\text{ext}}, H^+)$ for $|H'_{\text{ext}}| > |H_{\text{ext}}|$ by induction with base case $|H_{\text{ext}}| = k$, so verifying the formula will complete the proof.

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

- $G[B]$ is a copy of H^+ (contributing 1 to n_{H^+}),
- $A \subseteq B$, and $G[A]$ is a copy of H and is counted by $m_d(H_{\text{ext}}, H^+)$ for a fixed copy of H^+,
- B is a subset of the d-ball around A (i.e. $B \subseteq B_d(G[A], G)$).

Compared to $m_d(H_{\text{ext}}, G)$, the above term overcounts by 1 whenever $B \subsetneq B_d(G[A], G)$. Thus, it just remains to verify that the number of pairs with $B \neq B_d(G[A], G)$ is given by

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

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

- $G[A]$ is a copy of H,
- $G[B]$ is a 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. $\square$

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 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 in the previous section.

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 $\ell \geq 2$. 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 a tree T is reconstructible from its ℓ-deck provided it contains a sufficiently long path.

Removing an edge e 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 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 $C = (C^+, V_C)$ and $D = (D^+, V_D)$ 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 counts of the relevant leaf extensions obtained by Lemma 14. Lemma 16 is not specialised to trees with high diameter and the final step in proving Lemma 12 is showing that the lemma applies to trees with high diameter.

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

Lemma 16. *Let G be a connected graph with a bridge e, and $R, R^c \subseteq G$ 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', it must be isomorphic to G. This shows that such a G is reconstructible since if $\mathcal{D}_\ell(G_1) = \mathcal{D}_\ell(G_2)$, then applying the algorithm to this shared deck will produce a single graph G' for which $G' \cong G_1 \cong G_2$. The condition in the hypothesis that G has a suitable bridge e 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 feasible way that could 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 a 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. Then 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 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 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 loop over all pairs $(C_{\text{ext}}, D_{\text{ext}})$ of elements from $\mathcal{H}_{\text{ext}}$ for which $|C_{\text{ext}}| + |D_{\text{ext}}| = n + 2$ and $|C_{\text{ext}}| \leq |D_{\text{ext}}|$. Let $C_{\text{ext}} = (C^+, V_C)$ and $D_{\text{ext}} = (D^+, V_D)$ 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 copies of C in D whose 1-ball in D^+ is a copy of C^+. That is, we count the copies of C^+ in D^+ where the extending edge of D^+ is either unused or is the extending edge of C^+. If $m_1(C_{\text{ext}}, G) \geq N(C_{\text{ext}}, D_{\text{ext}}) + 1$, then the algorithm outputs the graph G' formed by taking disjoint copies of C^+ and D^+ and identifying their extending edges as given by these extensions. If $m_1(C_{\text{ext}}, G) < N(C_{\text{ext}}, D_{\text{ext}}) + 1$, 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 output a failure.

Let us first verify that if the algorithm returns a graph, it must be isomorphic to G. In fact, this is true for any connected graph. 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* leaf extension, which we will denote by $D^c_{\text{ext}} = ((D^c)^+, V_{D^c})$, that produces a graph isomorphic to G when joined with D_{ext} as described above. Explicitly, if e_D is the extending edge in D_{ext}, then D^c_{ext} is given by taking $V_{D^c} = (V_D)^c$ and setting $(D^c)^+$ to be the complement of D in G together with e_D as an additional edge. Indeed, this is true when G is any connected graph.

We now argue that any output graph G' is isomorphic to G. Let $(C_{\text{ext}}, D_{\text{ext}})$ be a pair that produces G'. If $C_{\text{ext}} \cong D^c_{\text{ext}}$ as leaf extensions then $G \cong G'$ by definition of D^c_{ext}, so suppose this is not the case. It is enough to show that $m_1(C_{\text{ext}}, G) < N(C_{\text{ext}}, D_{\text{ext}}) + 1$, giving a contradiction to the fact that we terminated when considering $(C_{\text{ext}}, D_{\text{ext}})$. We first claim that if a copy of C contributes to $m_1(C_{\text{ext}}, G)$ (in the sense that $G[V_C]$ coincides with this copy for a leaf extension counted by $m_1(C_{\text{ext}}, G)$), then it cannot use the edge between D and D^c. To see this, note that since G is connected, both D and D^c are connected. We have assumed that $|V(C)| = |V(D^c)| \leq |V(D)|$ and $C_{\text{ext}} \ncong D^c_{\text{ext}}$, so no copy of C can completely contain either D or D^c. This means that if a copy of C were to use this edge, then its 1-ball would contain at least one vertex from D and one from D^c meaning it does not contribute to $m_1(C_{\text{ext}}, G)$.

It now follows that

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

If $D^c_{\text{ext}} \ncong C_{\text{ext}}$ as we have assumed, then $N(C_{\text{ext}}, D^c_{\text{ext}}) = 0$. Indeed, since no copy of C uses the extending edge of $(D^c)^+$, this would have to be the extending edge of C^+ and we would have $C_{\text{ext}} \cong D_{\text{ext}}$. This leaves $m_1(C_{\text{ext}}, G) = N(C_{\text{ext}}, D_{\text{ext}})$ which gives the desired contradiction to the supposition that we terminated when considering $(C_{\text{ext}}, D_{\text{ext}})$, so G' must be isomorphic to G.

Finally, let us argue that the algorithm does terminate when the input is $\mathcal{D}_\ell(G)$. From (6) we easily see that $m_1(R_{\text{ext}}, G) = N(R_{\text{ext}}, R^c_{\text{ext}})+1$, where R_{ext} and R^c_{ext} are as defined earlier. Our assumption on G guarantees that both of these are in$\mathcal{H}_{\text{ext}}$, and we are guaranteed to have at least one pair among our candidates that will lead to termination. □

We remark that the only place where we used that the existence of the edge e which 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). We now show 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 required condition to be reconstructible amongst connected graphs.

Proof of Lemma 12. Let $k, \ell \in [n]$ 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 it has a suitable bridge satisfying the assumptions of Lemma 16 so that we can conclude that it is reconstructible from its ℓ-deck amongst connected graphs.

Fix a longest path in T with k vertices. Create two rooted subtrees R and $S = R^c$ 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). By Lemma 16, if T has no induced subgraph H isomorphic to R or S with $|V(B_1(H,T))| \geq \ell$, then T is reconstructible from $\mathcal{D}_\ell(T)$. We assume, in order to derive a contradiction, that T contains a 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.

We will proceed by building a sequence of vertex-disjoint paths in S to obtain a lower bound on the size of S, which leads to an upper bound on the maximum length of a path in R and hence also an upper bound on k. Let us sketch the main idea. Since the 1-ball around S' only extends one vertex further along paths not in S', the existence of a long path in T that is not in S' would indicate that this 1-ball misses lots of vertices and so cannot be too big. This means that S' should contain a lot of our chosen longest path, and S' should reach a long way into R. However, the long path in S' that reaches into R corresponds to a long path in S (under the isomorphism that makes it a copy), and S' must also contain many vertices from this path. These form another long path in S, and S' must reach a long way down this path as well. Continuing this argument eventually forces S to be so large that there are not enough vertices remaining to form a path of sufficient length in R for our assumption on k to be true.

Set $r = n - \ell$. Proceeding as laid out above, let $\varphi : S \to S'$ be an isomorphism, and let P_0 be a path in R containing at least $(k-1)/2$ vertices which starts at the root of R. Consider the intersection of S' with the path P_0. Since $V(S') \neq V(S)$, this intesection 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. Moreover, the intersection of $B_1(S', T)$ and P_0 must also be a path with at most $|V(Q_0)| + 2$ vertices. This means that there are at least $|V(P_0)| - |V(Q_0)| - 2$ vertices on P_0 which are not in $B_1(S', T)$. We are assuming that T has at most r vertices which are not in $B_1(S', T)$, so $|V(Q_0)| \geq |V(P_0)| - r - 2$.

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 as P_0 is contained in R. Define Q_1 to be the intersection of S' with P_1, which is again a path. Furthermore, the intersection of $B_1(S', T)$ and P_1 is also a path, this time with at most $|V(Q_1)| + 2$ vertices. The number of vertices of P_1 and P_0 which are not in $B_1(S', T)$ is at least $|V(P_0)| + |V(P_1)| - |V(Q_0)| - |V(Q_1)| - 4$, which gives the inequality $|V(Q_0)| + |V(Q_1)| \geq |V(P_0)| + |V(P_1)| - r - 4$. Since $|V(Q_0)| = |V(P_1)|$, this becomes $|V(Q_1)| \geq |V(P_0)| - r - 4$.

We now 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 $Q_{i+1} = P_{i+1} \cap S'$. We first note that P_{i+1} is disjoint from $P_0, \ldots, P_i$. Indeed, since P_0 is contained in R, P_{i+1} cannot intersect P_0. If P_{i+1} intersects a path P_j, then Q_i must intersect Q_{j-1} which in turn implies P_i intersects 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 $|V(P_i)| = |V(Q_{i-1})| \geq |V(P_0)| - r - 2i$ 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

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

where we have used $|P_0|$ as shorthand for $|V(P_0)|$.

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. The same argument shows that T has no copy R' of R with $|V(B_1(R', T))| \geq \ell$. 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. Specifically, 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 $n-\ell < \frac{n-3k+1}{3}$ if k is odd or $n-\ell < \frac{n-3k-1}{3}$ if k is even, which together imply the statement directly. These conditions are equivalent to odd $k < \ell - \frac{2n-1}{3}$ and even $k < \ell - \frac{2n+1}{3}$. The reason for 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 centre is unique, so in particular it does not depend on the choice of longest path.

The case when k is odd (so the diameter is even) is easier to work with, and the follow lemma provides a simple reduction that will allow us to proceed with this assumption.

Lemma 17. *If all trees with* $n+1$ *vertices and even diameter are reconstructible from the cards in their* $(\ell+1)$*-decks that contain a longest path, then all trees with* n *vertices and odd diameter are reconstructible from the cards in their* ℓ*-decks that contain a longest path.*

Proof. Let T be any tree with n vertices and odd diameter. This means that it has two middle vertices joined by one central edge. Let T' be the tree obtained by subdividing the central edge of T, noting that it has $n+1$ vertices and even diameter. We can obtain the cards in the $(\ell+1)$-deck of T' that contain a longest path by taking the cards in the ℓ-deck of T that contain a longest path and subdividing the central edge, and thus reconstruct T' by assumption. It is then straightforward to recover T by recognising the central vertex in T' and smoothing out the vertex created by subdivision. $\square$

Let us assume for the remainder 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 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

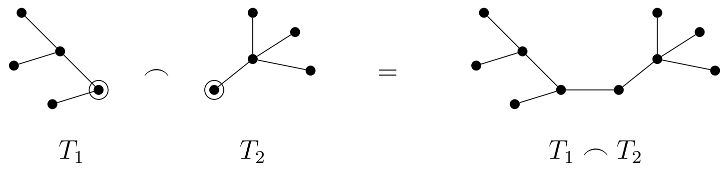

Figure 2: An example of the tree grafting operation $T_1 \frown T_2$.

least $\ell - k$) of the vertices a slightly more finicky version of the argument is required to reconstruct this branch as it cannot be seen on a single card. 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 18. *If T is a tree with even diameter $k-1$ for which every branch from the centre has fewer than $\ell - k$ vertices, then T is reconstructible from the subset of the ℓ-deck consisting only of cards that contain a copy of P_k.*

Proof. Let c be the central vertex of T, and let $\mathcal{B} = \{B_1, \dots, B_a\}$ 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 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 by Kelly's Lemma as $k < \ell$, so it suffices to reconstruct N_B.

Following the preceding outline, fix B to be any rooted tree that is not an end-rooted path. We will actually 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 of a longest path that is disjoint from B'. Similarly, let τ_B count pairs (B', P'_k) where the copy 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 3). Fix $G \in \mathcal{G}$ and consider some $F \in \mathcal{F}$. If $F' = P'_k \frown S'$ 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 the resulting F' might then also be contained in a different maximal $\mathcal{F}$-subgraph $P''_k \frown B''$ where S' is contained in the 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 $\mathcal{F}$. Thus, by Lemma 9 we can reconstruct the number of $\mathcal{F}$-maximal copies of each F in G from $\mathcal{D}_\ell(G)$. This is non-zero for $F = P_k \frown S$ if and only if $\pi_S \neq 0$.

Now let $F = P_k \frown B$. Since $T \in \mathcal{G}$ and $F \in \mathcal{F}$, we may reconstruct the number of $\mathcal{F}$-maximal copies of F in T as above. This is precisely 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 $P_{(k-1)/2+1} \frown S$ where S is a rooted tree which contains an end-rooted $P_{(k-1)/2}$, but is not itself an end-rooted path. Again, an element $F = P_{(k-1)/2+1} \frown S$ 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 each F in G by Lemma 9. This time there is at least one $\mathcal{F}'$-maximal copy of $F = P_{(k-1)/2+1} \frown S$ 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 $F' = P_{(k-1)/2+1} \frown B$ 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 $P_{(k-1)/2+1}$ 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 $P_{(k-1)/2+1}$ in B' with roots coinciding (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 $P_{(k-1)/2+1}$ 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_i, which we do using the fact that we know all of the other branches not of this form. Starting with $j = (k-1)/2$, this being the maximum possible length of a path branch, we compare the number of copies of $P_{(k-1)/2+j+1}$ in T to the number of copies in the graph $\widetilde{T}$ obtained by gluing all of the known branches at a single vertex c. The former count can be obtained by Kelly's Lemma, and the latter by directly inspecting $\widetilde{T}$. If there are more copies in T than in the current $\widetilde{T}$, then there must be at least one more end-rooted P_j as a branch so we add one copy to our list of known branches. We then repeat this step with the same j but a new $\widetilde{T}$ updated to include this new path branch. If the counts match, meaning all copies of $P_{(k-1)/2+j+1}$ in G are already present in $\widetilde{T}$, then reduce j by 1 and continue iteratively until $j = 0$. Note that it is important that we handle the different path lengths in this order. At this point, we have reconstructed all branches and the final $\widetilde{T}$ is exactly T. □

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, in fact, 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

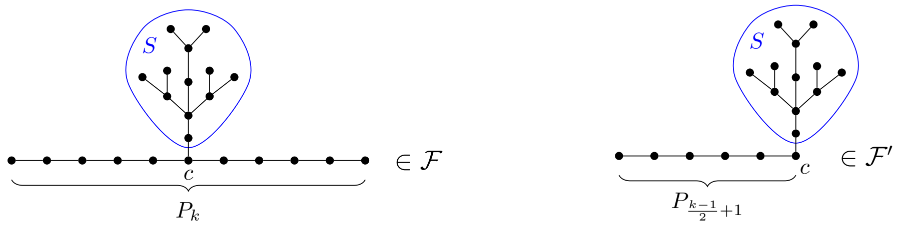

Figure 3: Elements of $\mathcal{F}$ and $\mathcal{F}'$.

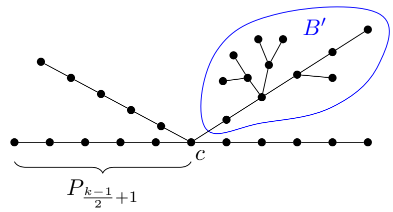

Figure 4: A tree containing three longest paths that avoid B' (so $\pi_B = 3$), and three longest paths that use B' consisting of a $P_{(k-1)/2+1}$ outside B' and a $P_{(k-1)/2}$ inside (so $\tau_B = 3$).

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 to not be a branch of T. Recall that $r := n - \ell$.

Lemma 19. *Let T be a tree with even diameter $k-1$ (where $k < \ell - \frac{2n-1}{3}$ is odd) and central vertex c, and suppose we are given only 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

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

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 P_k together with the path of length i from c to c' and any $(\ell - k - i)$-vertex subtree of B containing the root has at most $k + i + (\ell - k - i) = \ell$ vertices. On the other hand, if a card C has a subtree with a P_k (allowing us to identify c) and a vertex c_i at distance i from c with a partial outward branch that has at least $\ell - k - i$ vertices, then we would be done. It follows that T has such a vertex c_i and i-heavy branch if and only if it has a card containing a subtree of this 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

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

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 20. *If T is a tree of diameter $k - 1$ (where $k < \ell - \frac{2n-1}{3}$ is odd) and central vertex c, 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 19, we can recognise whether there is a branch at c with at least $\ell - k$ vertices. Let us suppose there is, since we are otherwise done by Lemma 18 (or equivalently by setting $j = 0$ and proceeding with the present proof).

To reconstruct the heavy branch at $c := c_0$, we construct a sequence of vertices $c_0, c_1, c_2, \ldots$ to act as new "centres" until the branches at some c_j are all small enough for us to apply Lemma 9. For the first step, let c_1 be the root

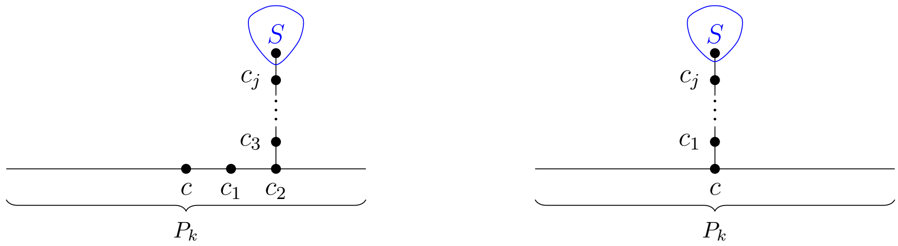

Figure 5: Potential elements of $\mathcal{F}$ along with their 'moving centres'.

of the heavy branch, which is adjacent to c_0. Applying Lemma 19 with $i = 1$, we can recognise whether any neighbour of c_0 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 repeat the argument. In the ith step, we terminate if every branch at c_i has weight less than $\ell - k - i$, and otherwise we completely determine all but the i-heavy branch and proceed by setting c_{i+1} to be the root of this branch. The case at hand is recognisable by Lemma 19 and we can also reconstruct the smaller branches at each step provided $i \leq (k-1)/2$. To see that this condition is maintained, 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 18. Let $\mathcal{G}$ be the family of n-vertex trees with diameter $k-1$, and $\mathcal{F}$ be the family of graphs that can be constructed as follows. Let $i \in \{0, \dots, j-1\}$, let $v_1, \dots, v_k$ be the vertices in a P_k and let $u_1, \dots, u_{j-i}$ be the vertices in a (disjoint) 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 identify P_k and the added tree in any $\mathcal{F}$-graph. An example is given in Figure 5.

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 18.

At this point, each branch B' has contributed one to the relevant count for each copy of P_k which does not use B', so we again need to determine the number of P_k which use B'. This can be done using an identical argument to that in Lemma 18 except replacing c with c_j, replacing $P_{(k-1)/2+1} \frown S$ with $P_{(k-1)/2+j+1} \frown S$ and suitably adjusting S.

We have now identified the total number of branches of each isomorphism class from vertices at distance j from c (except those which are end-rooted paths), although we do not know they are all branches at c_j. However, we have already reconstructed all of the tree except for the branches at c_j, so we can subtract the counts of all the appropriate branches not at c_j from the total: the remainder must be attached at c_j.

Finally, the end-rooted paths attached at c_j can be reconstructed using the argument from the end proof of Lemma 18. □

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

Proof. If k is odd, then by Lemma 20 we can reconstruct any n-vertex tree with diameter $k-1$ using only the cards in its ℓ-deck that contain a longest path provided $k < \ell - \frac{2n-1}{3}$, which is slightly better than the bound claimed in the theorem. The trees for which k is even are then also reconstructible by Lemma 17, provided $k < \ell + 1 - \frac{2(n+1)-1}{3} - 1 = \ell - \frac{2n+1}{3}$. □

6 Conclusion

The example in Figure 1 shows that the conjectured lower bound for reconstructing trees of $\lfloor n/2 \rfloor + 1$ is false for $n = 13$, but the bound is still the best known for all other values of n. It may well be the case that the conjecture 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 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 slightly improved, 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 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**(2):123–135, 1981.

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

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

[4] J. A. Bondy. On Kelly's congruence theorem for trees. *Mathematical Proceedings of the Cambridge Philosophical Society*, **65**(2):387–397, 1969.

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

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

[7] J. A. Bondy and R. L. Hemminger. Graph reconstruction: a survey. *Journal of Graph Theory*, **1**(3):227–268, 1977.

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

[9] P. Borwein and C. Ingalls. The Prouhet-Tarry-Escott problem revisited. *L'Enseignement Mathématique*, **40**(2):3–27, 1994.

[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**(4):630–648, 2017.

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

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

[13] W. B. Giles. The reconstruction of outerplanar graphs. *Journal of Combinatorial Theory Series B*, **16**(3):215–226, 1974.

[14] W. B. Giles. Reconstructing trees from two point deleted subtrees. *Discrete Mathematics*, **15**(4):325–332, 1976.

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

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

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

[18] P. J. Kelly. *On isometric transformations.* PhD thesis, University of Wisconsin, 1942.

[19] P. J. Kelly. A congruence theorem for trees. *Pacific Journal of Mathematics*, **7**(1):961–968, 1957.

[20] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin. 3-regular graphs are 2-reconstructible. *arXiv:1908.01258* preprint, 2019.

[21] 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**:491–501, 2020.

[22] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin. Acyclic graphs with at least $2\ell+1$ vertices are ℓ-recognizable. *arXiv:2103.12153* preprint, 2021.

[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(6)**:3278 – 3286, 2021.

[24] J. Lauri. The reconstruction of maximal planar graphs. *Journal of Combinatorial Theory Series B*, **30**(2):196–214, 1981.

[25] J. Lauri. Proof of Harary's conjecture on the reconstruction of trees. *Discrete Mathematics*, **43**(1):79–90, 1983.

[26] B. Manvel. Reconstruction of trees. *Canadian Journal of Mathematics*, **22**(1):55–60, 1970.

[27] B. Manvel. Some basic observations on Kelly's conjecture for graphs. *Discrete Mathematics*, **8**(2):181–185, 1974.

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

[29] W. Myrvold. The ally-reconstruction number of a tree with five or more vertices is three. *Journal of Graph Theory*, **14**(2):149–166, 1990.

[30] C. S. J. A. Nash-Williams. Reconstruction of infinite graphs. *Discrete Mathematics*, **95**(1-3):221–229, 1991. Directions in infinite graph theory and combinatorics (Cambridge, 1989).

[31] V. Nýdl. A note on reconstructing of finite trees from small subtrees. *Acta Universitatis Carolinae. Mathematica et Physica*, **31**(2):71–74, 1990.

[32] V. Nýdl. Finite undirected graphs which are not reconstructible from their large cardinality subgraphs. *Discrete Mathematics*, **108**(1-3):373–377, 1992.

[33] H. Spinoza and D. B. West. Reconstruction from the deck of k-vertex induced subgraphs. *Journal of Graph Theory*, **90**(4):497–522, 2019.

[34] R. Taylor. Reconstructing degree sequences from k-vertex-deleted subgraphs. *Discrete Mathematics*, **79**(2):207–213, 1990.

[35] C. Thomassen. Reconstructing 1-coherent locally finite trees. *Commentarii Mathematici Helvetici*, **53**(4):608–612, 1978.

[36] W. T. Tutte. All the King's Horses (A Guide to Reconstruction). In (J. A. Bondy and U. S. R. Murty, editors) *Graph Theory and Related Topics*. Academic Press, New York, 1979.

[37] S. M. Ulam. *A Collection of Mathematical Problems*, volume 8 of *Interscience tracts in pure and applied mathematics*. Interscience Publishers, 1960.