

Reconstructing trees from small 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. In 1976, Giles proved that any tree on $n \geq 5$ vertices can be reconstructed from its ℓ -deck for $\ell \geq n - 2$. Our main theorem states that it is enough to have $\ell > (8/9 + o(1))n$, making substantial progress towards a conjecture of Nýdl from 1990. In addition, we can recognise connectivity from the ℓ -deck if $\ell \geq 9n/10$, and the degree sequence from the ℓ -deck if $\ell \geq \sqrt{2n \log(2n)}$. 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 [9, 10, 19] states that all graphs on at least three vertices are reconstructible. While this classical conjecture was verified for trees by Kelly in [10], 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 connectivity, are known to be *reconstructible* for general graphs in the sense that they are determined by the deck.

A significant body of research has looked at the problem of reconstructing graphs and graph parameters from cards of smaller size. In the standard reconstruction conjecture, each card is an induced subgraph on $n - 1$ vertices, but one can instead look at cards that are induced subgraphs on ℓ vertices,

*Utrecht University, Utrecht, The Netherlands, c.e.groenland@uu.nl. Partially supported by the project CRACKNP that has received funding from the European Research Council under the EU Horizon 2020 research and innovation programme (grant agreement no. 853234).

†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
{[thomas.johnston](mailto:thomas.johnston@maths.ox.ac.uk), [scott](mailto:scott@maths.ox.ac.uk), [jane.tan](mailto:jane.tan@maths.ox.ac.uk)}@maths.ox.ac.uk

‡Supported by EPSRC grant EP/V007327/1.

where ℓ could be much smaller than $n - 1$. The ℓ -deck of G , denoted $\mathcal{D}_\ell(G)$, is the multiset of all induced subgraphs of G on ℓ vertices. In this notation, $\mathcal{D}(G) = \mathcal{D}_{n-1}(G)$.

Reconstruction from small cards was first introduced by Kelly in [10], although it did not receive much attention until it was studied by Manvel in 1974 [14]. Manvel 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, whether a given graph is a member of such a class is determined by its $(n - 2)$ -deck. Since then, the problem has been widely studied, and the reconstructibility of graphs from smaller cards is known for many classes of graphs including trees, 3-regular graphs, random graphs and graphs with maximum degree 2 [7, 11, 17]. However, many of the bounds are far from tight.

In general, it is not possible to reconstruct a graph from the ℓ -deck unless $\ell = (1 - o(1))n$, as shown by the following theorem of Nýdl.

Theorem 1 (Nýdl [16]). *For any integer n_0 and $0 < \alpha < 1$, there exist two non-isomorphic graphs on $n > n_0$ vertices that share the same multiset of subgraphs of size at most αn .*

However, it might be possible to do much better when reconstructing 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 2 (Nýdl [15]). *For any $n \geq 4$ and $\ell \geq \lfloor n/2 \rfloor + 1$, any two trees on n vertices with the same multiset of ℓ -vertex induced subgraphs are isomorphic, and this threshold is sharp.*

The conjectured bound would be sharp: Nýdl [15] constructed trees for which $\ell \geq \lfloor n/2 \rfloor + 1$ is necessary.

There has been no progress on Nýdl's conjecture since it was made in [15]. Indeed, the best result is an earlier bound of Giles [7] from 1976, which states that all n -vertex trees can be reconstructed from the $(n - 2)$ -deck for $n \geq 5$. Our main theorem improves very substantially on the result of Giles and takes a significant step towards Conjecture 2, showing that we can take $n - \ell$ to be of linear size.

Theorem 3. *For all $n \geq 3$, any n -vertex tree T can be reconstructed from $\mathcal{D}_\ell(T)$ when $\ell > \frac{8n}{9} + \frac{4}{9}\sqrt{8n + 5} + 1$.*

We remark that Conjecture 2 is false in the case $n = 13$ as demonstrated by the two graphs in Figure 1, but it is true for all other values $4 \leq n \leq 19$ and it remains open for large n .

We have already mentioned Manvel's result in [14] that the class of connected graphs is recognisable from the $(n - 2)$ -deck for $n \geq 6$. Extending this, Kostochka, Nahvi, West, and Zirlin [12] showed that the connectivity of a graph on $n \geq 7$ vertices is determined by $\mathcal{D}_{n-3}(G)$. As shown by Spinoza and West [17], if we take $G_1 = P_n$ (the path on n vertices) and $G_2 = C_{\lfloor n/2 \rfloor + 1} \sqcup P_{\lfloor n/2 \rfloor - 1}$



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

the disjoint union of a cycle and a path, we find $\mathcal{D}_k(G_1) = \mathcal{D}_k(G_2)$ for all $k \leq \lfloor n/2 \rfloor$. However, G_1 is connected and G_2 is not. This suggests the following natural conjecture.

Conjecture 4 (Spinoza and West [17]). *For $n \geq 6$ and $\ell \geq \lfloor n/2 \rfloor + 1$, the connectivity of an n -vertex graph G is determined by $\mathcal{D}_\ell(G)$, and this threshold is sharp.*

Spinoza and West proved in [17] that connectivity 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 the bound above to allow a linear gap between n and ℓ .

Theorem 5. *For all $n \geq 3$, the connectivity 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 3, 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 [13] independently announced a similar result to Theorem 6. In fact, they proved the stronger result that one can recognise if a graph is acyclic from the ℓ -deck when $\ell \geq \lfloor n/2 \rfloor + 1$, which verifies Conjecture 4 for the special case of forests. This has the particularly nice consequence that trees can be recognised from their ℓ -deck, and so Conjecture 2 is equivalent to the reconstruction of trees amongst general graphs. Our proof of Theorem 6 is considerably shorter than [13], and our theorem is (more than) sufficient for our main result on reconstructing trees, so we have retained our proof for the sake of 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 for connectivity. Chernyak [6] 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 [12] to the $(n-3)$ -deck for $n \geq 7$. The best known asymptotic result is due to Taylor [18], 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. *For $n \geq 3$, 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 [10] and a result on counting maximal subgraphs by Greenwell and Hemminger [8], as well as an algebraic result of Borwein and Ingalls [4] 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 3, is proved in Section 5. We conclude with a number of open problems 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 $\tilde{n}_H(G)$ 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 .

In the classical graph reconstruction problem, Kelly's Lemma states that we can reconstruct $n_H(G)$ and $\tilde{n}_H(G)$ provided $|V(H)| < |V(G)|$, and there are many variants of the lemma for other reconstruction problems (see [1]). 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 considerable usefulness, the proof of Kelly's Lemma is remarkably simple. Count the number of (possibly induced) 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

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

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 \mathcal{F} -maximal subgraphs in G which are isomorphic to F . We give a slight variation of a classical ‘‘Counting Theorem’’ due to Bondy and Hemminger [2] (see also [8]) 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 ,*

$$(i) \quad |V(F)| \leq \ell;$$

(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.

This result can be proved using following argument of Bondy and Hemminger [2] together with some additional observations. For completeness, we sketch the proof below.

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

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

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. Bondy and Hemminger prove that

$$m(F, G) = \sum_{k=0}^{\text{rank}_G(F)} \sum_{C^k} (-1)^k \tilde{n}_{X_1}(F) \tilde{n}_{X_2}(X_1) \cdots \tilde{n}_{X_k}(X_{k-1}) \tilde{n}_G(X_k), \quad (2.1)$$

where the inner sum is taken over the set C^k of all non-isomorphic (F, G) -chains (X_0, X_1, \dots, X_k) of length k . Any such chain must satisfy $|V(X_k)| \leq \ell$ by assumption (i). So, in order to prove Lemma 9, it is enough to show that it is possible to reconstruct (2.1) from the ℓ -deck.

For any $X_k \in \mathcal{F}$ on at most ℓ elements, we can compute $\tilde{n}_G(X_k)$ from the collection \mathcal{C} of cards in the ℓ -deck that contain an \mathcal{F} -subgraph. Indeed, we find

$$\tilde{n}_G(X_k) = \binom{n - |V(X_k)|}{\ell - |V(X_k)|}^{-1} \sum_{C \in \mathcal{C}} \tilde{n}_{X_k}(C)$$

since $\tilde{n}_{X_k}(C) = 0$ if $C \in \mathcal{D}_\ell(G)$ does not contain X_k .

When computing (2.1) we need to enumerate the (F, G) -chains, however, in (2.1) the inner summand will be zero if we consider a sequence

$$F \cong X_0 \subsetneq X_1 \subsetneq \dots \subsetneq X_k$$

of graphs in \mathcal{F} for which one of the terms (and hence X_k) is not a subgraph of G . We can therefore reconstruct the right-hand side of (2.1) from \mathcal{C} by instead summing over all chains of graphs from \mathcal{F} where the final graph has at most ℓ elements (which is something we can easily determine). This proves the lemma.

We note that the result above assumes that properties (i) and (ii) have been guaranteed; in particular we do not assume or guarantee that \mathcal{G} can be recognised from the given collection of cards, in contrast to Bondy and Hemminger [2].

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 a theorem of Borwein, Erdélyi and Kós [3] on the number of positive real roots of a polynomial.

Theorem 10 (Theorem A in [3]). *Suppose that the complex polynomial*

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

has k positive real roots. Then

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

The result that we require was proved by Borwein and Ingalls [4, Proposition 1]. We give 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*

$$\binom{\alpha_1}{j} + \dots + \binom{\alpha_m}{j} = \binom{\beta_1}{j} + \dots + \binom{\beta_m}{j} \quad \text{for all } j \in \{0, \dots, \ell\}, \quad (2.2)$$

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

Proof. Since $\alpha_i, \beta_j \in \{0, \dots, n\}$ for all $i, j \in [m]$,

$$p_{\alpha, \beta}(x) := \sum_{i=1}^m x^{\alpha_i} - \sum_{i=1}^m x^{\beta_i} \quad (2.3)$$

is a polynomial of degree at most n . For $c \in \mathbb{C}$, let $\text{mult}_c(p_{\alpha, \beta})$ denote the multiplicity of the root at c , or 0 if c is not a root of $p_{\alpha, \beta}$. We will show that $\ell + 1 \leq \text{mult}_1(p_{\alpha, \beta}) \leq \sqrt{2n \log(2m)}$.

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

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

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 (2.3), we have $\sum_{i=0}^{n'} |a_i| \leq 2m$.

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

$$\sqrt{2n \log \left(\frac{|a_0| + |a_1| + \dots + |a_n|}{\sqrt{|a_0 a_n|}} \right)} \leq \sqrt{2n \log(2m)}$$

and in particular $\text{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n \log(2m)}$. On the other hand, for all $j \in \{0, \dots, \ell\}$ equation (2.2) 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, $\ell + 1 \leq \text{mult}_1(p_{\alpha,\beta})$, and $\ell + 1 \leq \sqrt{2n \log(2m)}$ as desired. \square

Condition (2.2) 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 $\left\{ \binom{x}{i} : i \in \{0, \dots, \ell\} \right\}$ both form a basis for the polynomials of degree at most ℓ .

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 identical to the proof given by Taylor [18], except for the use of the stronger bounds provided by Lemma 11.

Proof of Theorem 7. 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

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

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 $d(v_1) \leq \dots \leq d(v_n)$. Suppose, for a contradiction, that a different degree sequence $\beta_1 \leq \dots \leq \beta_n$ gives the same counts. Thus, 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. \square

4 Recognising connectivity

We prove that the connectivity of an n -vertex graph G can be reconstructed from ℓ -deck of G whenever $\ell \geq 9n/10$. 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 .

Proof of Theorem 5. Let G be an n -vertex graph and let $\varepsilon = 1/10$. Suppose ℓ is an integer such that $\ell \geq 9n/10 = (1 - \varepsilon)n$. We wish to recognise if G is connected from the ℓ -deck. It was shown by Kostochka, Nahvi, West, and Zirlin [12] that the connectivity of a graph can be recognised from the $(n-3)$ -deck for $n \geq 7$ so we can assume that $n \geq 39$.

Suppose that G is disconnected and let H be the largest component. If $|V(H)| \leq \ell - 1$, then we can easily recognise that G is disconnected. Indeed, let $n_H(G)$ denote the number of copies of H in G , which we can compute from the ℓ -deck by Lemma 8. We can also compute $n_{H'}(G)$ for all connected graphs H' on $|V(H)| + 1 \leq \ell$ vertices that contain H ; but $n_{H'}(G) = 0$ for all such H' as H is the largest component. This allows us to identify that G has $n_H(G)$ components isomorphic to H , and that G is disconnected. 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 of order at most $n - \ell \leq \varepsilon n$ from the ℓ -deck. Let H be a connected graph on $1 \leq h \leq \varepsilon n$ vertices. 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, \dots, H_m for the induced copies of H in G , and define the *neighbourhood* of H_i by

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

We define the *degree* of H_i to be $|\Gamma(H_i)|$, and we denote it by α_i . Note that G has a component isomorphic to H if and only if $\alpha_i = 0$ for some $i \in [m]$. We now show that we can reconstruct the sequence $(\alpha_1, \dots, \alpha_m) \in \{0, \dots, n-h\}^m$ up to a permutation.

Since $1 \leq h \leq \varepsilon n$ and $m \leq \binom{n}{h} \leq \left(\frac{en}{h}\right)^h$ we have

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

Hence by Lemma 11, it suffices to show that we can reconstruct

$$\sum_{i=1}^m \binom{\alpha_i}{j} \text{ for all integers } 0 \leq j \leq N, \quad (4.1)$$

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 at most 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 $|H'| = h + j \leq \ell$. We find that

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

for $j \leq N$ and $n \geq 31$, where the middle inequality follows from the fact that

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

for $\varepsilon \leq 1/10$.

This shows that we can reconstruct (4.1), and hence the number of components isomorphic to H . \square

We remark that the constant $9/10$ can be improved slightly in the proof above provided n is large enough. Indeed, the proof works 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

The proof of Theorem 3 is split into three parts. First, we address the recognition problem in Section 5.2 which contains the proof of Theorem 6.

Next, reconstruction is split into two cases depending on whether T contains a path that is long relative to the size of the graph and the number of

cards removed. Let the *length* of a path P be the number of edges in P , or equivalently $|P| - 1$. The *diameter* of a graph G is the maximum distance between two vertices in G , and for a tree T this is just the same as the length of a longest path. In particular, we will refer to the aforementioned cases as the high diameter and low diameter cases.

Assuming we have determined that T is a tree, the high diameter case is handled by the following lemma which we prove in Section 5.3.

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

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

Lemma 13. *Suppose that T is a tree on $n \geq 3$ vertices with diameter $k - 1$. Then T 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.*

For all three results, we crucially keep track of copies of a graph in G that have a specified distinguished subgraph. These are broadly viewed as extensions of copies of the subgraph. In the spirit of Bondy and Hemminger's Counting Theorem discussed in Section 2.2, we show that it is possible to reconstruct the number of such extensions from the ℓ -deck under certain conditions. This strategy may be of independent interest, and is introduced in Section 5.1.

The proof of Theorem 3 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.

Proof of Theorem 3. Let k be the number of vertices in the longest path in T . The assumptions 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 T is reconstructible by Lemma 12 when $k > 4\sqrt{\ell} + 2(n - \ell)$. For the remaining k , we show that the conditions of Lemma 13 are then satisfied. It suffices to verify that $n - \ell < \frac{n-3k-1}{3}$. The right hand side is decreasing in k , and $k \leq 4\sqrt{\ell} + 2(n - \ell)$, so Lemma 13 applies provided

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

which is equivalent to our assumed condition

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

□

5.1 Counting extensions

For $d \in \mathbb{N}$, the (*closed*) d -ball of an induced subgraph H of a graph G is

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

the subgraph induced by the set of vertices of distance at most d from H including the vertices of H itself. Note that the graph $B_d(H, G)$ may contain multiple copies of H . As it is useful to keep track of one specific copy, we define an H -extension to be a pair $H_e = (H^+, A)$ where H^+ is a graph and $A \subseteq V(H^+)$ is a subset of vertices with $H^+[A] \cong H$. This definition applies to any graph H . The *size* of $H_e = (H^+, A)$ is $|H_e| = |V(H^+)|$. It is natural to view the d -ball of H as the H -extension $(B_d(H, G), V(H))$.

We say that two H -extensions (G_1, A_1) and (G_2, A_2) are *isomorphic* if there is a graph isomorphism $f : G_1 \rightarrow G_2$ which maps $f(A_1) = A_2$. Then let $m_d(H_e, G)$ be the number of copies of H in G whose d -ball is isomorphic (as an H -extension) to H_e . Lastly, a H -extension (H^+, A) is a *sub- H -extension* of (H^{++}, B) if H^+ is an induced subgraph of H^{++} and $A = B$.

It is possible to reconstruct $m_d(H_e, G)$ from the ℓ -deck provided the d -balls of all copies of H are small enough to appear on a card.

Lemma 14. *Let $\ell, d \in \mathbb{N}$ and let G be a graph on at least $\ell + 1$ vertices. For any graph H on at most $\ell - 1$ vertices, at least one of the following conditions must hold:*

1. *There is a copy of H in G whose d -ball in G has at least ℓ vertices.*
2. *For any H -extension H_e , we can reconstruct $m_d(H_e, G)$ from the ℓ -deck of G .*

Proof. Let \mathcal{H} denote the set of graphs H^+ such that $|V(H^+)| \leq \ell$, and there is a copy H' of H in H^+ such that all vertices of H^+ are at distance at most d from H' in H^+ . These represent all possible d -neighbourhoods, and in particular, \mathcal{H} contains all actual neighbourhoods of copies of H in G that we wish to examine. Note that H^+ may contain several copies of H , and these could have vertices outside of H^+ in their d -ball in G .

We may reconstruct $n_{H^+}(G)$ from the ℓ -deck for any $H^+ \in \mathcal{H}$ using Lemma 8, and we can recognise (from the ℓ -deck) whether Condition 1 of Lemma 14 holds. Suppose that it does not hold, so no copy H has a d -ball containing more than $\ell - 1$ elements. Then 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}_e denote the set of isomorphism classes of H -extensions (H^+, A) with $H^+ \in \mathcal{H}$. By the preceding observation, if $H_e = (H^+, A) \in \mathcal{H}_e$ with $|H^+| = k$, then

$$m_d(H_e, G) = n_{H^+}(G)m_d(H_e, H^+), \tag{5.1}$$

the number of copies of H^+ in G times the number of copies of H in H^+ whose d -ball (within H^+) is isomorphic to H_e (as H -extensions). Both of these quantities are reconstructible from the ℓ -deck, so we are done in this case.

If $|H^+| < k$, the d -ball of H may be strictly larger than H^+ and the formula (5.1) 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.

For $H'_e \in \mathcal{H}_e$ distinct from H_e , let $n(H_e, H'_e)$ give the number of sub- H -extensions of H'_e isomorphic to H_e . We claim that

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

Note that when $|H_e| = k$, this formula agrees with (5.1). The terms $m_d(H_e, H^+)$, $n(H_e, H'_e)$ 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 $m_d(H'_e, H^+)$ for $|H'_e| > |H_e|$ by induction with base case $|H_e| = k$, so the formula (if correct) implies the lemma.

We now verify the formula. The first term $n_{H^+}(G)m_d(H_e, H^+)$ counts the number of pairs $(A, B) \subseteq V(G) \times V(G)$ such that

- $G[B]$ is a copy of H^+ and corresponds to one object counted by n_{H^+} ,
- $A \subseteq B$,
- $G[A]$ is a copy of H and corresponds to an extension counted by $m_d(H_e, H^+)$ after B and hence the copy of H^+ is fixed,
- B is a subset of the d -ball around A (i.e. $B \subseteq B_d(G[A], G)$) due to the definition of m_d .

Compared to $m_d(H_e, G)$, we are overcounting whenever $B \subsetneq B_d(G[A], G)$. Thus, it just remains to verify that there are $\sum_{|H'_e| > |H_e|} n(H_e, H'_e)m_d(H'_e, G)$ pairs for which $B \neq B_d(G[A], G)$. To see this, we can think of the correction term as counting 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^+
- $G[C] \cong B_d(G[A], G)$.

The first two conditions follow from the definition when H_e is a sub- H -extension of H'_e , and the latter follows from the definition of $m_d(H'_e, G)$. The fact that $B \subsetneq C$ follows from the strict inequality $|H'_e| > |H_e|$. 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 appropriate 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

In this section we prove Theorem 6, which allows us to recognise whether a given ℓ -deck belongs to a tree on n vertices for $\ell \geq (2n + 4)/3$.

Proof of Theorem 6. 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 of edges m 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 . By applying Kelly's Lemma with every connected graph on ℓ vertices, we can determine whether every component of G has size at most $\ell - 1$, and if so, G cannot possibly be a tree. We may therefore assume that the largest component in G , say A , has at least $\ell \geq (2n + 4)/3$ vertices, and all other components have at most $n - \ell \leq \ell - 1$ vertices.

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

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

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

by our choice of d . Thus, $B_d(x) \cap C$ contains at least $2d + 2$ vertices. We can choose two vertices $c_1, c_2 \in B_d(x) \cap C$ such that there is a subpath C' of C between c_1 and c_2 that does not contain any other vertices of $B_d(x) \cap C$, allowing the possibility that C' is a single edge. 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 size 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 size at most $n - \ell$ (which happens if and only if G is not a tree), then we discover 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 connectivity and completes the proof. \square

5.3 High diameter case

The main result in this section is Lemma 12, which states that a tree T is reconstructible provided it has a sufficiently long path.

Consider the collection of components of $T - e$ as e varies over all edges, viewed as specific copies. All elements R in this collection have a natural counterpart $R^c := T[V(G) - V(R)]$. Our goal is to recognise a pair (R, R^c) for which we can also deduce which vertex in each subgraph was incident to e (assuming this is the pair of components in $T - e$). With this information, one can easily obtain T by gluing via one extra edge.

Instead of working just with copies of R (and R^c), we are interested in copies which connect to the rest of the graph by a single edge. For a graph H let a *leaf H -extension* be a pair $H_e = (H^+, A)$ 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. Note that the 1-ball of our special component R of T gives 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 that this is possible under assumptions that ensure that we can count leaf extensions by Lemma 14. The final step to proving Lemma 12 consists of showing that there exist suitable R and R^c in T for which Lemma 16 applies, and it is only then that we use the assumption of high diameter.

Lemma 16. *Let G be a connected graph with a bridge e , and $R, R^c \subseteq G$ be the connected 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 reconstructible from $\mathcal{D}_\ell(G)$.*

Proof. Given any connected graph H on at most $\ell - 1$ vertices and $\mathcal{D}_\ell(G)$, we can check whether there is a copy of H in G with $|V(B_1(H, G))| \geq \ell$. Suppose H is a connected graph for which no such copy exists. For every leaf H -extension H_e of H , apply Lemma 14 to reconstruct $m_1(H_e, G)$. Recall that this is the number of copies of H in G whose 1-ball in G is obtained by adding an edge at a specified vertex. Let \mathcal{H} denote the set of connected graphs H for which we have now reconstructed that $m_1(H_e, G) > 0$ for at least one leaf H -extension H_e .

We may assume that $|V(R^c)| \geq |V(R)|$. Note that \mathcal{H} is reconstructible from $\mathcal{D}_\ell(G)$, and by assumption R and R^c are elements of \mathcal{H} . Consider all pairs (C, D) of elements in \mathcal{H} for which $|V(C)| + |V(D)| = n$ and $|V(C)| \leq |V(D)|$. Given a leaf C -extension $C_e = (C^+, V_C)$ with $m_1(C_e, G) > 0$ and also a leaf D -extension $D_e = (D^+, V_D)$ with $m_1(D_e, G) > 0$, we will show that it is possible to determine whether $C \cong D^c$, and also whether gluing C_e and D_e on their additional edge gives G . This will complete the proof that G is reconstructible from its ℓ -deck since the existence of such a good pair is guaranteed by the fact that (R, R^c) is necessarily among the pairs considered, and suitable leaf extensions R_e and R_e^c exist.

Fix any pair (C, D) together with leaf extensions C_e and D_e . Since D_e is a leaf extension, we know that G is obtained by adding a single edge between D and D^c for some graph D^c with $|V(D^c)| = |V(C)|$. The fact that G is connected implies that D and D^c are connected. In particular, if $D' \subset D$ is a non-spanning subgraph, then $B_1(D')$ contains a vertex of $V(D) - V(D')$. The same holds if we replace D with D^c . This implies that any copy of C containing vertices from both D^c and D satisfies $|B_1(C)| \geq |C| + 2$, and so does not contribute to $m_1(C_e, G)$. Since $|V(C)| = |D^c|$, a copy of C cannot cover some of D^c and none of D . Hence, the only way a copy of C which contributes to $m_1(C_e, G)$ can contain vertices from D^c is if it covers all of D^c , and since $|V(C)| = |V(D^c)|$ this implies that $C \cong D^c$. There is hence at most one leaf C -extension for which the copy of C contains vertices from D^c and it exists if and only if C_e and D_e glue on the indicated edge to give G .

Any other contributing copy of C must be contained in D , and the corresponding copy of C^+ is contained in D^+ (possibly using the extra edge). Now let $N(C_e, D^+)$ be the number of leaf C -extensions $(C^{+'}, V_C')$ in D^+ isomorphic to C_e with $V_C' \subseteq V_D$, which can be calculated directly for our fixed C_e and D^+ . By the preceding discussion, either $m_1(C_e, G) = N(C_e, D^+)$ or $m_1(C_e, G) = N(C_e, D^+) + 1$. In the latter case, the additional leaf C -extension contains all vertices of D^c (and hence $C = D^c$), and this exists if and only if $C \cong D^c$ and C_e and D_e are glued on their respective edges. Since $m_1(C_e, G)$ and $N(C_e, D^+)$ are known, we can hence recognise whether C_e and D_e glue on the indicated edge to give G . \square

We are now ready to prove the main result in this section. It remains to show that Lemma 16 applies to trees with large enough diameter (depending on both n and the size of the deck). For this, we need to find subtrees R and S for which T has no copy of R or S with a large 1-ball; informally, we would like T to not be too star-like, and this is the case when T has a long path.

Proof of Lemma 12. Let $n \geq 3$ and $k, \ell \in [n]$ with $k > 4\sqrt{\ell} + 2(n - \ell)$ and $\ell \geq \frac{2n}{3} + \frac{4}{9}\sqrt{6n} + 7 + \frac{11}{9}$. We will show that the assumptions of Lemma 16 apply for any n -vertex tree T where a longest path contains at least $4\sqrt{\ell} + 2(n - \ell)$ vertices.

We assume that we have already determined that T is a tree, and note that we can recognise from the ℓ -deck if a longest path contains at least $4\sqrt{\ell} + 2(n - \ell)$ vertices by our choice of ℓ . Indeed, our choice of ℓ guarantees that $\ell \geq 4\sqrt{\ell} + 2(n - \ell) + 1 \geq \lceil 4\sqrt{\ell} + 2(n - \ell) \rceil$.

Fix a longest path in T with k vertices. Create two rooted subtrees R and S 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 $|B_1(H, T)| \geq \ell$, 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 $|B_1(S', T)| \geq \ell$.

Set $r = n - \ell$. Let $\varphi : S \rightarrow S'$ be an isomorphism, and let P_0 be a path in R of length at least $(k - 1)/2$ which starts at the root of R . Consider the intersection of S' with the path P_0 . It is non-empty since φ must be non-trivial, and 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 $|Q_0| + 2$ vertices. This means that there are at least $|P_0| - |Q_0| - 2$ vertices on P_0 which are not in $B_1(S', T)$. By assumption we have $|B_1(S', T)| \geq n - r$ meaning T has at most r vertices not in $B_1(S', T)$, so it follows that $|Q_0| \geq |P_0| - r - 2$.

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 . 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 $|Q_1| + 2$ vertices. The number of vertices of P_1 and P_0 which are not in $B_1(S', T)$ is at least $|P_0| + |P_1| - |Q_0| - |Q_1| - 4$, which gives the inequality $|Q_0| + |Q_1| \geq |P_0| + |P_1| - r - 4$. Since $|Q_0| = |P_1|$, this becomes $|Q_1| \geq |P_0| - r - 4$.

We now continue iteratively to build two sequences of paths: given P_i , define $Q_i := S' \cap P_i$ which is a subpath of P_i and then let $P_{i+1} := \varphi^{-1}(V(Q_i))$ which is a path in S . By finiteness of T , we must eventually reach a j such that $|Q_{j-1}| = |P_j| = 0$. At this point, we have disjoint paths P_1, \dots, P_j in S that satisfy $|P_i| = |Q_{i-1}| \geq |P_0| - r - 2i$ for all $i = 1, \dots, j$. In particular, setting $i = j$ to use the fact that $|P_j| = 0$ shows that $j \geq (|P_0| - r)/2$. We

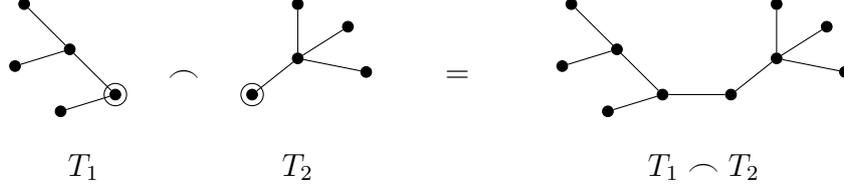


Figure 2: An example of the tree grafting operation $T_1 \frown T_2$.

may then calculate

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

Since $|S| \leq n - |P_0|$, we must have $|P_0| \leq \sqrt{4n - 4r + 1} + r - 1$ and $k \leq 2|P_0| + 1 \leq 2\sqrt{4n - 4r + 1} + 2r - 1$. 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 $|B_1(R', T)| \geq \ell$. Hence, by Lemma 16 we can reconstruct T from $\mathcal{D}_\ell(T)$. \square

5.4 Low diameter

Throughout this section, we will assume that $r = n - \ell < \frac{n - 3k + 1}{3}$, where k is the number of vertices in a longest path in T . This means that $k \leq \ell + 1$ and we can reconstruct k from the ℓ -deck.

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 vertices in the middle. The centre is unique, so in particular it does not depend on the choice of longest path. Let us assume for now that there is always a central vertex, leaving the even case to be handled later by subdividing the central edge. Given a vertex $u \in T$ with neighbours v_1, v_2, \dots, v_a , let the *branches* at u be the rooted subtrees b_1, b_2, \dots, 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. 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 least $\ell - k$) of the vertices a slightly more finicky argument is required to reconstruct the heavy branch which cannot be seen on a single card. It is possible to recognise these cases from the ℓ -deck, a statement which we prove as part of Lemma 13. As such, we simply state Lemma 17 and Lemma 18 with the assumption that we know *a priori* whether or not all branches have less than $\ell - k$ vertices.

Lemma 17. *Let T be a tree with even diameter $k - 1$. Suppose it is known that 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 b_1, \dots, b_a be the branches at c that we wish to reconstruct. Let b be a rooted tree which is not an end-rooted path. We will count each branch isomorphic to b once for every copy of P_k in T . From this we can determine the number of branches isomorphic to b by dividing by the number of copies of P_k in T , which can be determined by Kelly’s Lemma (Lemma 8) as $k < \ell$.

We start by applying Lemma 9. Let \mathcal{G} be the family of all trees with diameter $k - 1$, and let \mathcal{F} be the family of graphs of the form $P_k \frown S$, where S is a rooted tree 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' . Since every branch has fewer than $\ell - k$ vertices by assumption, these maximal elements have fewer than ℓ vertices. Thus, Lemma 9 provides the number of \mathcal{F} -maximal copies of each F . If this is non-zero, then F consists of P_k together with an entire branch.

Let $F = P_k \frown b$ and let m be the number of \mathcal{F} -maximal copies of F in T . Consider a particular copy b' of b that occurs as a branch. We 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' . It follows that m is the number of branches isomorphic to b multiplied by the number of copies of P_k which do not pass through a copy b' (and this is the same for every copy S' of S).

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 a rooted $P_{(k-1)/2}$, but is not itself an end-rooted path. The number of maximal \mathcal{F}' -subgraphs isomorphic to $P_{(k-1)/2+1} \frown b$, call this m' , is reconstructible by Lemma 9. The copy b' contributes one to m' for each copy of $P_{(k-1)/2+1}$ starting at the central vertex c and disjoint from b' . The product of m' and the number of copies of the rooted path $P_{(k-1)/2}$ inside b' represents the number of longest paths which pass through b' , say m'' . Each longest path contributes one to either m or m'' for b' , dependent on whether the longest path passes through b' or not. Hence,

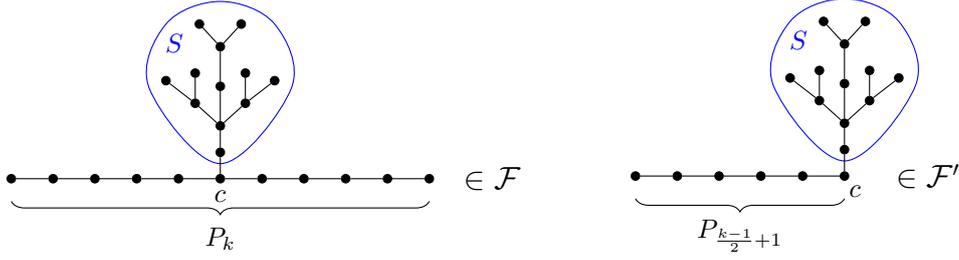


Figure 3: Elements of \mathcal{F} and \mathcal{F}' .

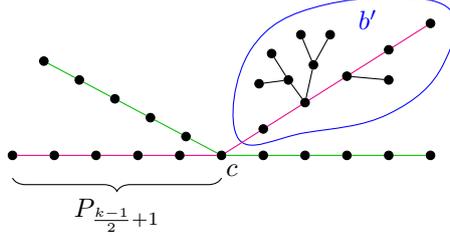


Figure 4: A longest path that avoids b' (green) contributes one to the number of times $P_k \setminus b$ occurs as a maximal subgraph. A longest path that uses b' (magenta) consists of a $P_{(k-1)/2+1}$ outside b' and a $P_{(k-1)/2}$ inside.

$m + m'$ is the number of branches isomorphic to b times the number of longest paths, so we can reconstruct the number of branches isomorphic to b .

It remains to determine the number of branches isomorphic to a path P_i (for $i \in [k]$), but we can now use 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 \tilde{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 directly from inspecting \tilde{T} . If there are more copies in T than in the current \tilde{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 j fixed but \tilde{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 \tilde{T} , then reduce j by 1 and continue iteratively until $j = 0$. At this point, we have reconstructed all branches and the final \tilde{T} is exactly T . \square

We now consider the case where one of the branches contains a lot of vertices. Since this branch contains so many vertices, we can find a card showing all the other branches in their entirety. This reduces the problem to reconstructing the large branch. We will move the ‘‘centre’’ one step inside the branch and continue doing this until no branch is too big. At this point we can apply the proof of the previous lemma (with some minor modifications) to reconstruct the branches off the new ‘‘centre’’.

Lemma 18. *Let T be a tree of even diameter $k - 1$. Suppose it is known that T has a branch with 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. Let c be the central vertex of T . First, note that $\ell - k = n - r - k > 2n/3$, so there can be at most one branch with at least $\ell - k$ vertices. Call this the *heavy branch*. The total number of vertices in the remaining branches is at most $k + r$. On any connected card in which only vertices of the heavy branch have been removed, there are still $n - 2r - k > k + r$ vertices of the heavy branch present. Taking a connected card in which the maximum number of vertices in any branch is as small as possible among those which contain a copy of P_k , we see that the heavy branch must still have at least $n - 2r - k$ vertices whilst each of the other branches have at most $k + r$ vertices. Thus, we can directly identify which is the heavy branch and the entirety of all of the smaller branches.

Set $c_0 := c$. To reconstruct the heavy branch, we construct a sequence of vertices c_0, c_1, c_2, \dots to act as new ‘centres’ until the branches at some c_j are all small enough for us to apply Lemma 9. Let c_1 be the vertex in the heavy branch adjacent to c_0 . If the branches at c_1 , which we call *1-branches*, all have less than $\ell - k - 1$ vertices, then we terminate with $j = 1$. Otherwise, take a connected card containing a copy of P_k in which the heaviest 1-branch is as small as possible. The heaviest branch has at least $\ell - k - r - 1$ vertices, which is greater than the maximum number of vertices in any other branch (now $k + r + 1$). This ensures that the heaviest 1-branch is contained in the original heavy branch, and we can identify all smaller 1-branches. Now set c_2 to be the vertex in the heaviest 1-branch adjacent to c_1 and repeat the argument. In the i th step, we terminate if every branch from c_i has weight less than $\ell - k - i$, and otherwise completely determine all but the heaviest i -branch and proceed by taking a step into this i -branch. To do this we only require $\ell - k - r - i > k + r + i$, which holds for $i \leq (k - 1)/2$ by our choice of r . Suppose the process terminates at the j th step. Since the longest path in G contains at most k vertices, the longest path in the heavy branch with one endvertex at the root contains at most $(k + 1)/2$ vertices and hence $j \leq (k - 1)/2$.

The remainder of the argument closely follows the proof of Lemma 17. Let \mathcal{G} be the family of trees of 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 which is not an end-rooted path to the vertex u_{j-i} . This condition on the tree 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 off u_{j-i} . Note that this is not necessarily a j -branch: it might be part of the graph that we already know. Applying Lemma 9 allows us to determine number of maximal \mathcal{F} -subgraphs, as we did in the proof of Lemma 17.

At this point, each branch b' has contributed one to the relevant count for each copy of P_k which does not use b' , and we again need to determine the number of P_k which use b' . This can be done using an identical argument to

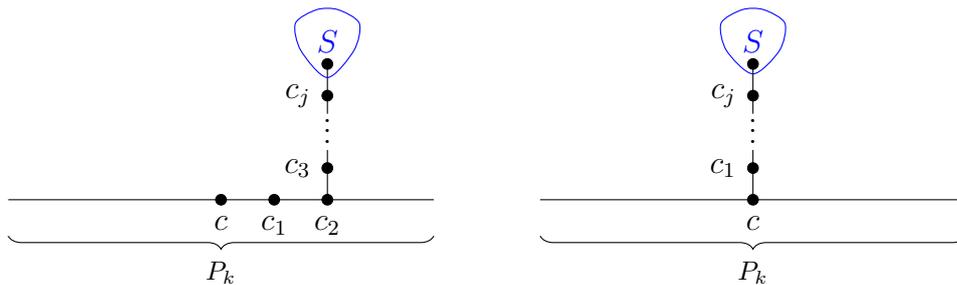


Figure 5: Potential elements of \mathcal{F} along with their ‘moving centres’.

that in Lemma 17 except for replacing c with c_j , replacing $P_{(k-1)/2+1} \cap S$ with $P_{(k-1)/2+j+1} \cap S$ and suitably adjusting S . We have so far identified the total number of branches of each isomorphism class (except those which are rooted paths), but we do not know from which vertices in the graph they branch off. However, we know all branches except the j -branches attached at c_j , so we can remove them from the counts. The remainder must be attached at c_j .

Finally, we are left to handle the end-rooted paths attached at c_j , but this can be done in a similar fashion to the end of the proof of Lemma 17. \square

Proof of Lemma 13. Let T be an n -vertex tree with diameter $k - 1$, and hence k vertices in any longest path. First assume that k is odd, so there is a central vertex c of T . 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 whether there is a branch with at least $\ell - k$ vertices. If there is no such branch then we are done by Lemma 17, and if there is one, we are done by Lemma 18.

When k is even, we reduce to the odd case by instead considering the tree T' obtained by subdividing the central edge. Since we can recognise the central vertex in T' and hence recover T , it suffices to reconstruct T' . Moreover, 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. These are the only cards required by Lemma 17 and Lemma 18. \square

6 Open problems

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 $\ell(n) = (\frac{1}{2} + o(1))n$ such that all n -vertex trees be reconstructed from their $\ell(n)$ -deck?*

We also leave the following natural question open.

Problem 2. *Asymptotically, what is the minimal $\ell = \ell(n)$ for which the degree sequence of every n -vertex graph can be reconstructed from the ℓ -deck?*

In terms of lower bounds on $\ell(n)$, we remark that it is easy to obtain one which is polynomial in $\log n$. Indeed, each ℓ -vertex graphs 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 at least $\Omega(4^n/n)$ possible degree sequences [5], and hence we need

$$2^{\log_2(n)\ell 2^{\ell^2}} \geq 2^{2n - \log_2(n)} \implies \ell = \Omega(\sqrt{\log n}).$$

By considering restricted graph classes, this can be slightly improved, but it would be interesting to see whether the bound can be improved to n^ε for some $\varepsilon > 0$.

While Theorem 10 is tight up to constants, we do not know if the $\log(2m)$ factor is required in Lemma 11. Without it, our upper bound of $O(\sqrt{n} \log n)$ would become $O(\sqrt{n})$.

Problem 3. *Can the upper bound on ℓ in Lemma 11 be improved to $O(\sqrt{n})$?*

Lastly, we pose the following problem.

Problem 4. *Let $\ell_k(n)$ be the minimum ℓ such that, for every n -vertex graph G , whether G is k -colourable can be recognised from the ℓ -deck. What are the asymptotics of $\ell_k(n)$?*

A special case that would make an interesting starting point is to recognise whether the graph is bipartite. A lower bound of $\lfloor n/2 \rfloor$ follows from the example of Spinoza and West [17] mentioned in the introduction (consider a path and the disjoint union of an odd cycle and a path). Manvel [14] proved that the $(n - 2)$ -deck suffices, but it is possible that a non-constant or even linear number of vertices could be removed from the cards in this case as well.

References

- [1] J. A. Bondy. A graph reconstructor's manual. *Surveys in Combinatorics, London Mathematical Society Lecture Note Series*, **166**:221–252, 1991.
- [2] J. A. Bondy and R. L. Hemminger. Graph reconstruction—a survey. *J. Graph Theory*, **1**(3):227–268, 1977.
- [3] P. Borwein, T. Erdélyi and G. Kós. Littlewood-type problems on $[0, 1]$. *Proc. Lond. Math. Soc.*, **79**(1):22–46, 1999.
- [4] P. Borwein and C. Ingalls. The Prouhet-Tarry-Escott problem revisited. *Enseign. Math.*, **40**(2):3–27, 1994.
- [5] J. M. Burns. *The number of degree sequences of graphs*. PhD thesis, Massachusetts Institute of Technology, 2007.

- [6] Z. A. Chernyak. Some additions to an article by B. Manvel: “Some basic observations on Kelly’s conjecture for graphs” (Russian). *Vestsi Akad. Navuk BSSR Ser. Fz.-Mat. Navuk*, **126**:44–49, 1982.
- [7] W. B. Giles. Reconstructing trees from two point deleted subtrees. *Discrete Math.*, **15**(4):325–332, 1976.
- [8] D. L. Greenwell and R. L. Hemminger. Reconstructing the n -connected components of a graph. *Aequationes Math.*, **9**:19–22, 1973.
- [9] P. J. Kelly. *On isometric transformations*. PhD thesis, University of Wisconsin, 1942.
- [10] P. J. Kelly. A congruence theorem for trees. *Pacific J. Math.*, **7**(1):961–968, 1957.
- [11] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin. 3-regular graphs are 2-reconstructible. *arXiv:1908.01258* preprint, 2019.
- [12] 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 Combin.*, **36**:491–501, 2020.
- [13] 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.
- [14] B. Manvel. Some basic observations on Kelly’s conjecture for graphs. *Discrete Math.*, **8**(2):181–185, 1974.
- [15] V. Nýdl. A note on reconstructing of finite trees from small subtrees. *Acta Universitatis Carolinae. Mathematica et Physica*, **31**(2):71–74, 1990.
- [16] V. Nýdl. Finite undirected graphs which are not reconstructible from their large cardinality subgraphs. *Discrete Math.*, **108**(1-3):373–377, 1992.
- [17] H. Spinoza and D. B. West. Reconstruction from the deck of k -vertex induced subgraphs. *J. Graph Theory*, **90**(4):497–522, 2019.
- [18] R. Taylor. Reconstructing degree sequences from k -vertex-deleted subgraphs. *Discrete Math.*, **79**(2):207–213, 1990.
- [19] S. M. Ulam. *A Collection of Mathematical Problems*, volume 8 of *Interscience tracts in pure and applied mathematics*. Interscience Publishers, 1960.