Size reconstructibility of graphs

Carla Groenland*, Hannah Guggiari*, Alex Scott*†

July 27, 2018

Abstract

The deck of a graph G is given by the multiset of (unlabelled) subgraphs $\{G - v : v \in V(G)\}$. The subgraphs $G - v$ are referred to as the cards of G. Brown and Fenner recently showed that, for $n \geq 29$, the number of edges of a graph G can be computed from any deck missing 2 cards. We show that, for sufficiently large n, the number of edges can be computed from any deck missing at most $\frac{1}{20}\sqrt{n}$ cards.

1 Introduction

Throughout this paper, all graphs are finite and undirected with no loops or multiple edges. The *order* of a graph is the number of vertices in the graph; the *size* of a graph refers to the number of 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 and all edges incident to v. The multiset of all unlabelled cards of G is called the *deck*, $\mathcal{D}(G)$, and has size n.

It is natural to ask whether it is possible for two non-isomorphic graphs to have the same deck. Kelly and Ulam [8, 9, 15] proposed the following *Reconstruction Conjecture*.

Conjecture 1.1. *For $n > 2$, two graphs G and H of order n are isomorphic if and only if $\mathcal{D}(G) = \mathcal{D}(H)$.*

The Reconstruction Conjecture is still open, although it is known to be true for certain classes of graphs (for example trees [9]). Moreover, almost every graph can be reconstructed [2, 11, 12]. For more background, see [1, 3, 4, 10, 14].

*Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom.
†Supported by a Leverhulme Trust Research Fellowship.

A potentially easier problem is to determine which parameters of a graph can be calculated from its deck. Such parameters are said to be *reconstructible*. Given a full deck of cards, it is easy to reconstruct the number of edges m: summing over the edges present in all of the cards gives $m(n-2)$ where n is the number of vertices. It is also well-known that connectedness and the degree sequence are reconstructible.

In fact, some parameters are reconstructible even if there is not a full deck of cards. For example, Bowler, Brown, Fenner and Myrvold [6] showed that we only need $\left\lfloor \frac{n}{2} \right\rfloor + 2$ cards to determine whether the graph is connected. Myrvold [13] also found that the degree sequence is reconstructible from $n-1$ cards.

In this paper, we are concerned with reconstructing the number of edges in a graph. In a recent paper, Brown and Fenner [7] showed that, for $n \geq 29$, only $n-2$ cards are required to determine the size of a graph. Woodall [16] found that, for any $p \geq 3$ and n sufficiently large, if two graphs on n vertices have $n-p$ common cards, then the number of edges in these two graphs differs by at most $p-2$.

In Section 2, we will improve on both results by showing that the size of a graph is reconstructible with as many as $c\sqrt{n}$ missing cards. In particular, we will prove the following theorem.

Theorem 1.2. *For n sufficiently large and $k \leq \frac{1}{20}\sqrt{n}$, the number of edges m of a graph G on n vertices is reconstructible from any $n-k$ cards.*

We will also consider the following game played against an adversary. The adversary chooses a graph G of order n and gives us a collection of n cards, each showing a graph on $n-1$ vertices. We are told that there are $n-k$ *true cards*, which come from the deck $\mathcal{D}(G)$. The other k cards are *false cards*, which can depict any graph of order $n-1$. We win if we are able to reconstruct the size of G; otherwise the adversary wins. When are we guaranteed to win regardless of the graph G and the cards given by the adversary? This turns out to be a corollary of Theorem 1.2.

Corollary 1.3. *Let n be sufficiently large and $k \leq \frac{1}{40}\sqrt{n}$. The number of edges m of a graph G on n vertices is reconstructible from any collection $\mathcal{C}$ of cards where $n-k$ are true and k are false.*

2 Size Reconstruction from $n-k$ Cards

Throughout this section, G will be a graph of order n and size $m = e(G)$, where m is unknown. The vertex set of G will be $V(G) = \{v_1, \ldots, v_n\}$ and we will write G_i for the card $G - v_i$. We will assume that we are given the cards $G_1, \ldots, G_{n-k}$. For any graph G', let the number of vertices of degree t be $d_t(G') = |\{v \in V(G') : d_{G'}(v) = t\}|$. For convenience, we will write $d_t = d_t(G)$. Note that d_t is unknown for every t.

2.1 Proof Overview

Using the cards we have been given, we first obtain a upper bound $\widetilde{m}$ on the number of edges m. We then show that the upper bound is close to m: writing $\alpha = \widetilde{m} - m$, we show that $0 \leq \alpha < 2k$. We then use this to estimate d_t, the number of vertices of degree t in G. If we knew the number of edges m, then we could calculate the degree of vertex v_i from its card G_i by $d(v_i) = m - e(G_i)$. Instead, we estimate the degree of the vertex corresponding to each card by $\widetilde{d}(v_i) = \widetilde{m} - e(G_i)$ and count the number of vertices

$$\widetilde{d}_t = |\{i \in \{1, \ldots, n-k\} : \widetilde{d}(v_i) = t\}|$$

with estimated degree t from the cards we have been given. Since $m \leq \widetilde{m}$, our estimate $\widetilde{d}(v_i)$ may be larger than the actual degree of vertex v_i. This means that the actual sequence (d_t) has been shifted to the right by α. Moreover, some of the cards are missing so, after applying the shift, we have $\sum_{t=0}^{n-1} |d_t - \widetilde{d}_{t+\alpha}| = k$.

Our goal is to discover the shift $\alpha = \widetilde{m} - m$, since together with $\widetilde{m}$ this allows us to calculate m. In order to do this, we construct d_t exactly for many t and match these known values to the flawed sequence $\widetilde{d}_t$ in order to discover the shift. This is done in Lemma 2.5 and Claim 1.

2.2 Preliminary Results

Let $\widetilde{m} = \left\lfloor \frac{1}{n-2-k} \sum_{i=1}^{n-k} e(G_i) \right\rfloor$ be an estimate for the number of edges in G. We can calculate $\widetilde{m}$ from the cards $G_1, \ldots, G_{n-k}$.

Lemma 2.1. *$\widetilde{m} = m + \alpha$ where $0 \leq \alpha \leq \left\lfloor \frac{k(n-1)}{n-2-k} \right\rfloor$.*

Note that if $k = o(n)$ then the error $\alpha \leq (1 + o(1))k$.

Proof of Lemma 2.1. Suppose that we have the entire deck of G. Every edge of G is on exactly $n-2$ cards and therefore $\sum_{i=1}^{n} e(G_i) = (n-2)m$. Furthermore, for every $v_i \in V(G)$, we have that $d(v_i) = m - e(G_i)$. It follows that $\sum_{i=1}^{n-k} e(G_i) = (n-2-k)m + \sum_{j=1}^{k} d(v_{n-k+j})$. Thus

$$\widetilde{m} = \left\lfloor \frac{1}{n-2-k} \sum_{i=1}^{n-k} e(G_i) \right\rfloor = m + \left\lfloor \frac{1}{n-2-k} \sum_{j=1}^{k} d(v_{n-k+j}) \right\rfloor.$$

We know $0 \leq d(v_j) \leq n-1$ if $v_j \in V(G)$. Since $\alpha = \left\lfloor \frac{1}{n-2-k} \sum_{j=1}^{k} d(v_{n-k+j}) \right\rfloor$, the result follows. $\square$

For $0 \leq t \leq n-1$, we define s_t to be the total number of vertices of degree t seen in the cards $G_1, \ldots, G_{n-k}$. Note that s_t can be calculated from the given cards.

Lemma 2.2. *Let* $\varepsilon_t = (n-1-t)d_t + (t+1)d_{t+1} - s_t$. *Then* $0 \leq \varepsilon_t \leq k(d_t + d_{t+1})$.

Proof. Consider the entire deck of G and let $v \in V(G)$ be a vertex of degree t. Then v appears as a vertex of degree $t-1$ on exactly t cards and as a vertex of degree t on $n-t-1$ cards (and does not appear on its own card). Hence

$$s_t + \sum_{j=1}^{k} d_t(G_{n-k+j}) = \sum_{i=1}^{n} d_t(G_i) = (n-1-t)d_t + (t+1)d_{t+1}.$$

Set $\varepsilon_t = \sum_{j=1}^{k} d_t(G_{n-k+j})$. For any $i \in [n]$, a vertex of degree t in G_i actually has degree t or $t+1$ in G and so $0 \leq d_t(G_i) \leq d_t + d_{t+1}$. The result now follows. $\square$

As noted by Brown and Fenner [7] and others, any result for a graph G implies a corresponding result for its complement $\overline{G}$.

Observation 2.3. *If* $\mathcal{D}(G) = \{G_1, \ldots, G_n\}$, *then* $\mathcal{D}(\overline{G}) = \{\overline{G}_1, \ldots, \overline{G}_n\}$. *Moreover, we have that* $d_t(\overline{G}) = d_{n-1-t}(G)$ *for any* $t \in \{0, \ldots, n-1\}$.

Lemma 2.4. *Let* $k \leq \frac{n}{3}$ *and* $t \in \{0, \ldots, n-1\}$. *There exists a value* d_t^*, *which can be calculated from the cards, that satisfies* $\frac{1}{4}d_t - 1 \leq d_t^* \leq d_{t-1} + d_t + d_{t+1}$.

Proof. We first consider the case where $t < \frac{n}{2}$. For these values of t, define $d_t^* = d_t^*(G) = \max\{d_t(G_i) : 1 \leq i \leq n-k\}$. Note that d_t^* can be calculated from the given cards.

Fix $t < \frac{n}{2}$ and choose $j \in [n-k]$ such that $d_t(G_j) = d_t^*$. Every vertex of degree t on G_j corresponds to a vertex in G with degree t or $t+1$. Hence, $d_t^* \leq d_t + d_{t+1}$.

Let N be the number of times a vertex of degree t in G is seen as a vertex of degree $t-1$ in the cards $G_1, \ldots, G_{n-k}$. We will find upper and lower bounds for N. For the upper bound, note that a vertex of degree t appears as a vertex of degree $t-1$ on the card $G_i = G - v_i$ if and only if v_i is one of its neighbours. Therefore, $N \leq t d_t$.

Now consider the card G_i for some $i \in [n-k]$. The card G_i has $d_t(G_i)$ vertices of degree t and each of these vertices has degree t or $t+1$ in G. The vertex v_i is not shown but may also have degree t. Therefore, there are at least $d_t - d_t(G_i) - 1$ vertices that have degree $t-1$ in G_i but degree t in G. It follows that $N \geq \sum_{i=1}^{n-k}(d_t - d_t(G_i) - 1)$. We combine these bounds on N to get

$$td_t \geq N \geq \sum_{i=1}^{n-k}(d_t - d_t(G_i) - 1) \geq (n-k)(d_t - d_t^* - 1).$$

Rearranging and using the assumptions that $t < \frac{n}{2}$ and $n - k \geq \frac{2n}{3}$, we find $\frac{2}{3}d_t^* \geq \frac{1}{6}d_t - \frac{2}{3}$. It follows that $d_t^* \geq \frac{1}{4}d_t - 1$.

We now consider the case where $t \geq \frac{n}{2}$. For these values of t, define $d_t^* = d_{n-1-t}^*(\overline{G})$. As $n-1-t < \frac{n}{2}$, this is well-defined. From the argument above, we have

$$\frac{1}{4}d_{n-1-t}(\overline{G}) - 1 \leq d_{n-1-t}^*(\overline{G}) \leq d_{n-1-t}(\overline{G}) + d_{n-t}(\overline{G}).$$

By Observation 2.3, we see that

$$\frac{1}{4}d_t(G) - 1 \leq d_{n-1-t}^*(\overline{G}) = d_t^* \leq d_t(G) + d_{t-1}(G).$$

As d_{t-1} and d_{t+1} are both non-negative for every value of t, the result follows. □

In the proof of Theorem 1.2, we will compare the unknown sequence (d_t) to a sequence $(\widetilde{d_t})$ that can be calculated from the cards. In order to do this, we will need to know some values of d_t exactly.

For the proof of Theorem 1.2, we will only need the following lemma in the case when $\beta = \frac{1}{2}$ and in the interval $[\frac{n}{3}, \frac{2n}{3}]$. However, the result may be of independent interest and so we state it in a more general form.

Lemma 2.5. *Suppose $0 \leq \beta < 1$ and write $\gamma = \frac{3}{4} + \frac{1}{4}\beta < 1$. Let n be sufficiently large and $k = O(n^\beta)$. Then, for any graph G of order n and any deck of $n-k$ cards, the value of d_t can be calculated exactly for all but $O(n^\gamma)$ values of t.*

Proof. We will assume that n is sufficiently large to ensure $k \leq \frac{n}{3}$. Let $I = \{0, \dots, n-1\}$ and $A = \{t \in I : d_t^* + 1 \geq \frac{1}{4}K\}$ where $K = n^{1-\gamma}$. By Lemma 2.4, we have $\sum_{t\in A}(d_t^* + 1) \leq \sum_{t \in A}(d_{t-1} + d_t + d_{t+1} + 1) \leq 4n$. Hence, $|A| \leq \frac{16n}{K} = 16n^\gamma$. Note that, if $t \notin A$, then $d_t \leq K$ by Lemma 2.4 and so, for n sufficiently large, d_t is small in comparison to n. Let $I' = I \setminus (A \cup (A-1))$. We will show that we can calculate d_t exactly for most $t \in I'$.

If $t \in I'$, then $t, t+1 \notin A$ and hence $d_t, d_{t+1} \leq K$. By Lemma 2.2, we have $s_t = (n-1-t)d_t + (t+1)d_{t+1} - \varepsilon_t$ where $0 \leq \varepsilon_t \leq k(d_t + d_{t+1}) \leq 2kK$.

For convenience, we will write $t+1 = qn$ where $q = \frac{t+1}{n} \in [0,1] \cap \mathbb{Q}$. Then $\frac{s_t}{n} = (1-q)d_t + qd_{t+1} - \frac{\varepsilon_t}{n}$. Note that we are able to find the value $\frac{s_t}{n}$ from the cards $G_1, \ldots, G_{n-k}$.

Consider the set $X = \{(1-q)a + qb : a, b \in \{0, \ldots K-1\}$. Choose a, b such that $(1-q)a+qb$ is the closest element of this set to $\frac{s_t}{n}$. We would like to conclude that $d_t = a$ and $d_{t+1} = b$. This is true if every pair of elements of X takes values that are at least $\left|\frac{2\varepsilon_t}{n}\right|$ apart.

Suppose that, for some $\delta < \left|\frac{2\varepsilon_t}{n}\right|$, we are able to find $a > a'$ and $b < b'$ satisfying $a(1-q) + bq = a'(1-q) + b'q + \delta$. Rearranging, we get

$$a - a' = (b' - b + a - a')q + \delta.$$

In particular, $(b'-b+a-a')q+\delta$ is an integer. As $|b'-b+a-a'| \leq 2K-2$, it suffices to ensure that q is at distance at least δ from every element of

$$R = \left\{\frac{x}{y} : y \in \{1, \ldots, 2K-2\}, x \in \{0, \ldots, y\}\right\}.$$

The set R has size strictly less that $2K^2$. For $i \in \{0, \ldots, n-1\}$ and $\frac{x}{y} \in R$, suppose that $\left|\frac{x}{y} - \frac{i}{n}\right| < \left|\frac{2\varepsilon_t}{n}\right|$. Then $\left|\frac{xn}{y} - i\right| < 2|\varepsilon_t|$. It is straightforward to see that, for each choice of $\frac{x}{y} \in R$, there are at most $4|\varepsilon_t|$ choices for i. Define

$$S = \left\{t : \exists r \in R \text{ such that } \left|\frac{t+1}{n} - r\right| < \left|\frac{2\varepsilon_t}{n}\right|\right\}.$$

We see that $|S| \leq 4|\varepsilon_t||R| < 16kK^3$. Let $J = I' \setminus S$. Then we have that $|J| \geq n - \frac{32n}{K} - 16kK^3 > n - 48n^\gamma$ for n sufficiently large. For every $t \in J$, we can calculate d_t exactly. $\square$

2.3 Main Result

We are now ready to prove that the size of a graph of order n is reconstructible from $n-k$ cards.

Proof of Theorem 1.2. As noted earlier, we can obtain an estimate $\widetilde{m}$ for the number of edges in G from the cards $G_1, \ldots, G_{n-k}$. By Lemma 2.1, we have $\widetilde{m} = m + \alpha$ where $0 \leq \alpha \leq \left\lfloor \frac{k(n-1)}{n-2-k} \right\rfloor$ and so it suffices to find α. For n sufficiently large, we have $n-1 < 2(n-2-k)$ and hence $\alpha < 2k$.

Throughout the remainder of this proof, we will say that d_t is *large* if $d_t > \sqrt{n}$ and *small* if $d_t \leq \frac{3}{4}\sqrt{n}$.

Claim 1. *Suppose that, for some $t \leq \frac{2n}{3} - 1$, the value of d_{t+1} is known exactly and is not large. Then either d_t can be calculated exactly or d_t can be identified as being large.*

Proof. We know, from Lemma 2.2, that $s_t = (n-1-t)d_t + (t+1)d_{t+1} - \varepsilon_t$ where $0 \leq \varepsilon_t \leq k(d_t + d_{t+1})$. Define $d'_t = \frac{1}{n-1-t}(s_t - (t+1)d_{t+1})$. Then d'_t can be calculated exactly and, furthermore, $d_t = d'_t + \frac{\varepsilon_t}{n-1-t}$ so $d_t \geq d'_t$.

Suppose that $d_t \leq 2\sqrt{n}$. By assumption, $d_{t+1} \leq \sqrt{n}$ and so we find that $d_t - d'_t = \frac{\varepsilon_t}{n-t-1} \leq \frac{3}{n}k(d_t + d_{t+1}) \leq \frac{9}{20} < \frac{1}{2}$. Hence, if $d_t \leq 2\sqrt{n}$, the closest integer to d'_t is precisely d_t. In particular, this holds if d_t is not large.

Now suppose that $d_t > 2\sqrt{n}$. It suffices to show that the closest integer to d'_t must be large. We know $d'_t \geq d_t - \frac{\varepsilon_t}{n-1-t}$, $t+1 \leq \frac{2n}{3}$ and $d_{t+1} \leq \sqrt{n}$. Together, these give us that $\frac{\varepsilon_t}{n-1-t} \leq \frac{3}{n}k(d_t + d_{t+1}) \leq \frac{3d_t}{20\sqrt{n}} + \frac{3}{20} \leq \frac{1}{2}d_t$ for n sufficiently large and hence $d'_t \geq \frac{1}{2}d_t > \sqrt{n}$. We therefore correctly identify that d_t is large in this case. $\Diamond$

Claim 2. *Suppose that, for some $t \geq \frac{n}{3} + 1$, the value of d_{t-1} is known exactly and is not large. Then either d_t can be calculated exactly or d_t can be identified as being large.*

Proof. If $t \geq \frac{n}{3} + 1$, then $n - t - 1 \leq \frac{2n}{3} - 1$. By Observation 2.3, we have $d_{n-t}(\overline{G}) = d_{t-1}(G)$. Apply Claim 1 to $\overline{G}$ to see that either $d_t(G) = d_{n-t-1}(\overline{G})$ can be calculated exactly or it can be identified as being large. $\Diamond$

Claim 3. *The interval $[\frac{n}{3}, \frac{2n}{3}]$ contains $2k$ consecutive values of t such that every d_t can be calculated exactly and they are all small.*

Proof. Let $I = [\frac{n}{3}, \frac{2n}{3}] \cap \mathbb{N}$. Lemma 2.5 with $\beta = \frac{1}{2}$ gives a set $J \subseteq I$ and a constant c such that $|I \setminus J| \leq cn^{\frac{7}{8}}$ and we can calculate d_t exactly if $t \in J$. Hence there are at most $cn^{\frac{7}{8}}$ values of $t \in I$ for which we cannot calculate d_t exactly.

Partition I into $\left\lfloor \frac{n}{6k} \right\rfloor$ intervals of length $2k$. At most $\left\lfloor \frac{cn^{7/8}}{2k} \right\rfloor$ of them are completely contained in $I \setminus J$. For n sufficiently large, $\left\lfloor \frac{n}{6k} \right\rfloor - \left\lfloor \frac{cn^{7/8}}{2k} \right\rfloor \geq \frac{n}{8k}$. Therefore, for these values of n, there are at least $\frac{n}{8k}$ intervals which have non-empty intersection with J. By Claim 1 and 2, we are able to calculate d_t exactly for all values of t in each of these intervals unless the interval happens to contain a value of t for which d_t is large.

We know that there are at most $\frac{4}{3}\sqrt{n}$ values of $t \in \{0, \ldots, n-1\}$ for which d_t is not small. Therefore, as $\frac{n}{8k} \geq \frac{5}{2}\sqrt{n} > \frac{4}{3}\sqrt{n}$, there exists an interval which has non-empty intersection with J and which only contains small values of d_t, each of which we can calculate exactly. $\Diamond$

By Claim 3, we can find an interval $\mathcal{I} = \{b, b+1, \ldots, b+2k-1\} \subset [\frac{n}{3}, \frac{2n}{3}]$ such that, for every $t \in \mathcal{I}$, we can calculate d_t exactly and it is small. We may then recursively apply Claim 1, starting with $t+1 = b$. We continue until either we reach d_0 or we hit a large vertex d_{t_ℓ} for some $t_\ell < b$. Similarly, we may recursively apply Claim 2, starting with $t-1 = b+2k-1$. Again, we will either calculate d_{n-1} or we will identify that d_{t_r} is large for some $t_r > b+2k-1$.

If we are able to calculate both d_0 and d_{n-1}, then we will know d_t for every $t \in \{0, \ldots, n-1\}$. This tells us the degree sequence of G and hence we can directly calculate m.

Therefore, we may assume that we have the following situation: there exists an interval $\mathcal{J} \supseteq \mathcal{I}$ with endpoints t_ℓ and t_r such that $t_\ell < t_r$. For every $t \in \mathcal{J} \setminus \{t_\ell, t_r\}$, the value d_t is known exactly and is not large. At least one of d_{t_ℓ} and d_{t_r} has been identified as being large. By Observation 2.3, we may assume that d_{t_ℓ} is large.

Recall from the proof overview that $\widetilde{d}_t = |\{i \in \{1, \ldots, n-k\} : \widetilde{d}(v_i) = t\}|$ can be calculated from the cards and also that $\sum_{t=0}^{n-1} |d_t - \widetilde{d}_{t+\alpha}| = k$. (Note that we need to calculate $\widetilde{d}_t$ for $0 \leq t \leq n+2k$ and that, for $t + \alpha \geq n$, it is possible for $\widetilde{d}_{t+\alpha}$ to take a non-zero value.) This means that the overall shape of $\widetilde{d}_0, \ldots, \widetilde{d}_{n-1}$ will be the same as the overall of shape of $d_0, \ldots, d_{n-1}$ but shifted to the right by α. We will determine α.

Although we do not know the exact value of d_{t_ℓ}, it is sufficient to redefine each d_t and $\widetilde{d}_t$ to be the minimum of their current value and $\sqrt{n}$. After doing this, we still have $\sum_{t=0}^{n-1} |d_t - \widetilde{d}_{t+\alpha}| \leq k$. It follows that $\sum_{t=t_\ell}^{t_r-1} |d_t - \widetilde{d}_{t+\alpha}| \leq k$.

Claim 4. *For $s \in \{0, \ldots, 2k-1\}$, $\sum_{t=t_\ell}^{t_r-1} |d_t - \widetilde{d}_{t+s}| \leq k$ if and only if $s = \alpha$.*

Proof. Fix $s \in \{0, \ldots, 2k-1\}$. We noted above that $\sum_{t=t_\ell}^{t_r-1} |d_t - \widetilde{d}_{t+\alpha}| \leq k$. Now suppose $s \neq \alpha$. We have

$$\begin{aligned}\sum_{t=t_\ell}^{t_r-1} |d_t - \widetilde{d}_{t+s}| &= \sum_{t=t_\ell}^{t_r-1} |d_t - d_{t+s-\alpha} + d_{t+s-\alpha} - \widetilde{d}_{t+s}| \\ &\geq \sum_{t=t_\ell}^{t_r-1} |d_t - d_{t+s-\alpha}| - \sum_{t=t_\ell}^{t_r-1} |d_{t+s-\alpha} - \widetilde{d}_{t+s}|. \qquad (\star)\end{aligned}$$

Since $\sum_{t=0}^{n-1} |d_t - \widetilde{d}_{t+\alpha}| \leq k$, it follows that

$$\sum_{t=t_\ell}^{t_r-1} |d_{t+s-\alpha} - \widetilde{d}_{t+s}| = \sum_{t=t_\ell+s-\alpha}^{t_r+s-\alpha-1} |d_t - \widetilde{d}_{t+\alpha}| \leq k.$$

Hence, $(\star)$ will be strictly greater than k whenever $\sum_{t=t_\ell}^{t_r-1} |d_t - d_{t+s-\alpha}| > 2k$.

Recall that the interval $\mathcal{I}$ consists of $2k$ consecutive values of t such that every d_t is small. As $s \leq 2k-1$ and $s \neq \alpha$, then there exists some $\eta \in \mathbb{Z}$ such that $t_\ell + \eta(s-\alpha) \in \mathcal{I}$. We know that $\eta(s-\alpha) > 0$. First assume $\eta > 0$. Since d_{t_ℓ} is large and $d_{t_\ell+\eta(s-\alpha)}$ is small, we find

$$\begin{aligned}\sum_{t=t_\ell}^{t_r-1} |d_t - d_{t+s-\alpha}| &\geq \sum_{i=0}^{\eta-1} |d_{t_\ell+i(s-\alpha)} - d_{t_\ell+(i+1)(s-\alpha)}| \\ &\geq |d_{t_\ell} - d_{t_\ell+\eta(s-\alpha)}| \\ &\geq \sqrt{n} - \frac{3}{4}\sqrt{n} = \frac{1}{4}\sqrt{n} \\ &> 2k.\end{aligned}$$

If $\eta < 0$, then $s - \alpha < 0$ and

$$\sum_{t=t_\ell}^{t_r-1} |d_t - d_{t+s-\alpha}| \geq \sum_{i=0}^{-\eta} |d_{t_\ell+(i+1)(\alpha-s)} - d_{t_\ell+i(\alpha-s)}| \geq |d_{t_\ell-\eta(\alpha-s)} - d_{t_\ell}|.$$

The result then follows in a similar fashion. $\Diamond$

By Claim 4, we see that α is the only value $s \in \{0, \ldots, 2k-1\}$ satisfying $\sum_{t=t_\ell}^{t_r-1} |d_t - \widetilde{d}_{t+s}| \leq k$. Once we have identified α, we can then calculate m, the number of edges in G. $\square$

We now consider the situation where our deck contains $n-k$ true cards and k false cards and show that we can calculate the number of edges in our graph exactly if $k \leq \frac{1}{40}\sqrt{n}$.

Proof of Corollary 1.3. Suppose that G and H are two graphs on n vertices and each has at least $n-k$ cards in common with a deck of cards $\mathcal{C}$. Then G and H must have at least $n-2k$ cards in common. We may apply Theorem 1.2 to these $n-2k$ common cards. If n is sufficiently large and $2k \leq \frac{1}{20}\sqrt{n}$, then G and H must have the same number of edges. $\square$

3 Conclusion

We have shown that the size of a graph can be reconstructed if we are given a deck from which either at most $\frac{1}{20}\sqrt{n}$ cards are missing or at most $\frac{1}{40}\sqrt{n}$ cards are false. The constants can be improved a little, although we do not know whether the result remains true with $\sqrt{n}$ missing cards. However, we suspect that stronger results could be proved by using more information about the degree sequences on the cards.

We also note that $c\sqrt{n}$ is still very far away from the best known lower bounds, which are linear. For example, for $n = 3p + 1$, Bowler, Brown and Fenner [5] have given the following two graphs which differ in the number of edges but have $\frac{2}{3}(n-1)$ cards in common: the graphs $G = 2K_{p+1} + K_{p-1}$ and $H = K_{p+1} + 2K_p$ both have at least $2p$ cards of the form $K_{p+1} + K_p + K_{p-1}$. We suspect that the lower bound is closer to the truth and propose the following question.

Problem 3.1. *Does there exist some $\varepsilon > 0$ such that, for any graph G on n vertices, we can reconstruct the number of edges of G from any $(1 - \varepsilon)n$ cards?*

Another direction for future work is to reconstruct other graph parameters, such as the degree sequence or the number of triangles. The following problem seems very natural.

Problem 3.2. *Fix $k \in \mathbb{N}$ and a graph H and let n be sufficiently large. Can we reconstruct the number of copies of H in G given any $n - k$ cards from $\mathcal{D}(G)$ for any graph G on n vertices?*

If we are given the entire deck $\mathcal{D}(G)$ (i.e. $k = 0$), then this problem is solved by Kelly's Lemma [9].

Lemma 3.3. *For any two graphs G and H with $|G| > |H|$, the number of subgraphs of G isomorphic to H is reconstructible.*

If the number of edges is known, then the degree of a vertex can be calculated from the number of edges on its card. Therefore, by our main result, if $k \leq \frac{1}{20}\sqrt{n}$, then all but k of the degrees are known. If k is larger, then Lemma 2.5 still allows us to construct most of the degree sequence. We expect that, for a large range of k, it is possible to determine the whole degree sequence exactly. As a first step, we make the following conjecture.

Conjecture 3.4. *Fix $k \in \mathbb{N}$ and let n be sufficiently large. For any graph G on n vertices, the degree sequence of G is reconstructible from any $n - k$ cards.*

Note that a positive answer to Problem 3.2 would give a positive answer to Conjecture 3.4: for fixed k and n sufficiently large, we can find the number of edges of the graph by Theorem 1.2 and hence determine all but k elements of the degree sequence. Provided n is sufficiently large, we can reconstruct the number of copies of the star $K_{1,j}$ for $j = 1, \ldots, k + 1$; this is given by

$\sum_{v\in V(G)} \binom{d(v)}{j}$. By subtracting the terms corresponding to vertices of known degree, we obtain a sequence of polynomials in the unknown degrees. Adding constants, these form a basis for all polynomials of degree at most $k+1$. From these, it is straightforward to evaluate the remaining degrees.

References

[1] K. Asciak, M. Francalanza, J. Lauri and W. Myrvold, A survey of some open questions in reconstruction numbers, *Ars Combinatoria* **97** (2010), 443–456.

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

[3] J. Bondy, A graph reconstructor's manual, *Surveys in Combinatorics*, Cambridge University Press, (1991) 221–252.

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

[5] A. Bowler, P. Brown and T. Fenner, Families of pairs of graphs with a large number of common cards, *Journal of Graph Theory* **63** (2010), 146–163.

[6] A. Bowler, P. Brown, T. Fenner and W. Myrvold, Recognising connectedness from vertex-deleted subgraphs, *Journal of Graph Theory* **67** (2011), 285–299.

[7] P. Brown and T. Fenner, The size of a graph is reconstructible from any $n-2$ cards, *Discrete Mathematics* **341** (2018), 165–174.

[8] P. Kelly, On isometric transformations (PhD thesis), University of Wisconsin (1942).

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

[10] J. Lauri and R. Scapellato, Topics in graph automorphisms and reconstruction, Cambridge University Press, (2016).

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

[12] W. Myrvold. Ally and Adversary Reconstruction Problems, *PhD thesis*, University of Waterloo, (1988).

[13] W. Myrvold, The degree sequence is reconstructible from $n-1$ cards, *Discrete Mathematics* **102** (1992), 187–196.

[14] C. Nash-Williams, The reconstruction problem, *Selected Topics in Graph Theory*, (1978).

[15] S. Ulam, A collection of mathematical problems, Interscience, New York (1960).

[16] D. Woodhall, Towards size reconstruction from fewer cards, *Discrete Mathematics* **338** (2015), 2515–2522.