Graduate School of Natural Sciences

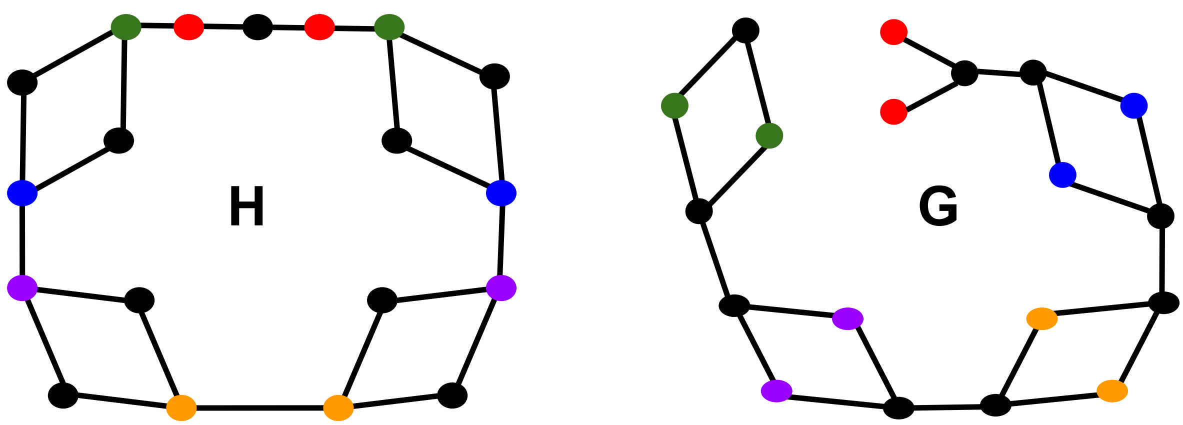
Recognizing cycles from partial decks

Determining properties from less one-vertex removed subgraphs

Master Thesis

Gabriëlle Zwaneveld

Mathematical Sciences



Supervisors:

Dr. Carla Groenland
Utrecht University

Prof. Dr. Gunther Cornelissen
Utrecht University

June 27, 2023

Abstract

The unlabeled subgraphs $G - v$ are called the cards of a graph G and the deck of the graph G is the multiset $\{G - v : v \in V(G)\}$. In 1942 Kelly conjectured that any finite, simple, undirected graph on at least 3 vertices is uniquely determined by its deck. Although this conjecture is still open, it is known that many properties can be determined from (sub)decks of G. For example, I proved that every graph G on n vertices has a subdeck of $\lfloor \frac{n}{2} \rfloor + 2$ cards from which we can determine the number of edges of G.
Wendy Myrvold [Ars Combinatoria, 1989] showed that a disconnected graph and a connected graph both on n vertices have at most $\lfloor \frac{n}{2} \rfloor + 1$ cards in common. Moreover, she found infinite families of trees and disconnected forests for which this upper bound is attained. Hence, we need at least $\lfloor \frac{n}{2} \rfloor + 2$ cards to determine whether a graph is a tree. Bowler, Brown, and Fenner [Journal of Graph Theory, 2010] conjectured that this bound is tight. In this thesis, I prove this conjecture for sufficiently large n: I show that a tree T and a unicyclic graph G on n vertices have at most $\lfloor \frac{n}{2} \rfloor + 1$ common cards.
Based on this theorem, I show that any forest and non-forest also have at most $\lfloor \frac{n}{2} \rfloor + 1$ common cards. Moreover, I have classified all pairs (except finitely many) for which this bound is strict. Furthermore, I adapted the main ideas of the proof for trees to show that the girth of a graph on n vertices can be determined based on any $\frac{2n}{3} + 1$ of its cards. Lastly, I showed that any $\frac{5n}{6} + 2$ cards determine whether a graph is bipartite.

Acknowledgements

I would like to start by thanking my master thesis supervisor Carla Groenland. First of all, she suggested this very interesting topic on which I have worked with much joy. Moreover, I really enjoyed our regular meetings in which she would give me advice about how to write everything down in a structured way and also gave some suggestions about what to do next. During these meetings she also functioned as a mentor by giving me a lot of advice about the next steps in my career. Lastly, her valuable suggestions on the first version heavily improved the quality of this thesis.
Secondly, I should thank Siebe for always being there for me, especially when I had a bad day. It was nice to have someone to which I could talk about both the mathematical aspects of my thesis as well as the process itself. Furthermore, I would like to thank you for reading my work and giving me valuable suggestions.
I also want to thank my family, friends and roommates. In particular, I want to thank my almost-roommate Anouk for all our fun evenings. My fellow trainers Dirk, Kevin, Nils, Ward and Wietze for the nice trainings activities from the Dutch Math Olympiad and Wietze also for his valuable feedback on my thesis. My parents, Naomi and Deborah for reminding me that even if things do not go as planned, that it will always work out in the end.
Lastly, I would also like to thank the second reader Gunther Cornelissen for taking the time to read and evaluate this thesis.

Contents

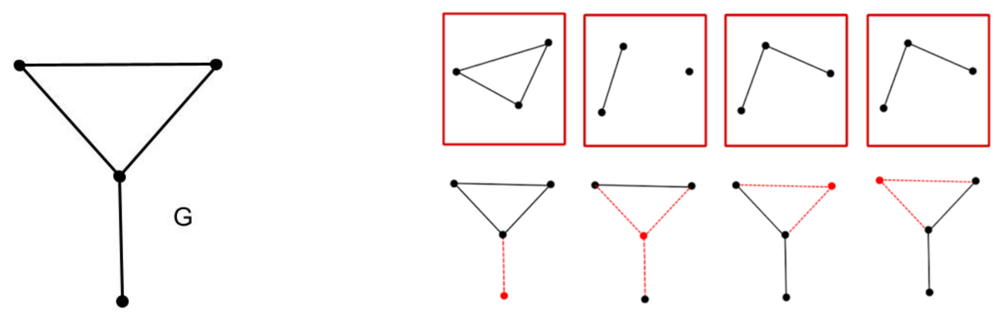
1 Introduction

Graph theory is a subfield in mathematics that was first mentioned as a field of interest by Leibniz in 1679 as geometria situs (the geometry of the location), since it is a nice way to convert location to mathematics. However, most people view the paper about the bridges of Köningsbergen written by Leonard Euler as the real start of graph theory [4].
Since then, especially in the last few years, the field has grown enormously. Nowadays, you can find graphs in various aspects of our life. For example, you can find them as maps of public transportation, friends/connections on social media, and in my favorite board game Ticket to Ride. Moreover, the theory has also many applications for example in search engines, scheduling, error correcting codes, and evolution theory.
In this thesis, we will only use finite, undirected, simple graphs. Hence, a graph G is an ordered pair (V, E) where V is the set of vertices and E is the set of edges $e = \{u, v\}$ where $u, v \in V$. We will denote $v(G) = |V| = n$ for the number of vertices and $e(G) = |E|$ for the number of edges. Figure 1a contains an example of a graph G on 4 vertices. For more information about graph theory, one can consult one of many textbooks about this topic, for instance, 'Graph Theory' by Reinhard Diestel.

1.1 Vertex reconstruction

The topic of this thesis is: what can we say about a graph if we only see the induced subgraphs obtained from deleting one vertex? To be more precise: for a graph G, we only see the graphs $G - v$ with $v \in V$. The graph $G - v$ has vertex set $V - \{v\}$ and contains exactly those edges of G that do not contain v as one of its endpoints. To make it easier to talk about these graphs, we will give the following definition.

Definition 1.1. *The* ***card*** *of G* ***corresponding to*** *the vertex v is the subgraph $G - v$. The* ***deck*** *of G is the multiset of all graphs (up to isomorphism) $G - v$ for $v \in V$.*



(a) The original graph G. (b) The deck of G consisting of all subgraphs where one vertex is deleted [6].

Figure 1: A graph G with its deck of cards. Every vertex corresponds to the unique card where that vertex is deleted. The deleted vertex is shown in red in the graph below.

As indicated by the example above, every graph on n vertices has exactly n cards. Some of these n cards can be isomorphic, for instance, the cards corresponding to the left and right vertex in the example above are isomorphic. This is the reason that we define the deck to be the multiset of all cards instead of the normal set.
A natural and important question that arises is whether the deck of cards determines the graph G uniquely. Therefore, we will give the following definition.

Definition 1.2. *A graph H is called a* ***reconstruction*** *of G if the decks of G and H are the same. A graph G is* ***reconstructible*** *if any reconstruction H of G is isomorphic to G.*

In 1942 Paul Kelly posed the following conjecture in his PhD thesis [8] (page 73).

Conjecture 1.3 (Reconstruction conjecture)**.** *Any graph having at least 3 vertices is reconstructible.*

Remark 1.4. *The condition that the graph has at least 3 vertices is necessary. Both graphs $G = K_2$ (two vertices connected with an edge) and $H = 2K_1$ (two isolated vertices) have a deck that consists of two cards with one point. Since G and H are not isomorphic, we conclude that neither graph is reconstructible.*

Remark 1.5. *Because both the graphs on 2 vertices are not reconstructible, we will assume for the rest of the thesis that every graph G has $n \geq 3$ vertices.*

Although the main reconstruction conjecture is still open, we do know that every reconstruction H of G must look somewhat similar to G. To be more precise, we can deduce many properties from the deck of G.

Definition 1.6. *A graph property $\mathcal{P}$ is called* ***recognizable*** *if every reconstruction of G also satisfies the graph property $\mathcal{P}$.*

For example, the degree sequence, all strict subgraphs of a graph G, and connectivity are recognizable from the deck. Proofs that these properties are indeed recognizable are given in Chapter 2. Moreover, it is also known that all graphs in certain classes are reconstructible.

Definition 1.7. *A class $\mathcal{C}$ of graphs is* ***reconstructible*** *if all members are reconstructible from their deck.*

For example, the classes of regular graphs, trees, and disconnected graphs are reconstructible. The proofs of the fact that regular graphs and disconnected graphs are reconstructible can be found in Chapter 2.

1.2 Ally and adversary reconstruction

Now, that we have seen that we can deduce many properties from the deck of a graph G, we can ask ourselves the question 'How many cards of G do we need to determine whether a graph G has a certain property $\mathcal{P}$?' Or to be more specific, if we get $x \leq n$ cards (without extra information on how they are picked) of G, can we determine whether the G satisfied property $\mathcal{P}$? To answer this question Wendy Myrvold [12] introduced the following two concepts:

- **Adversary reconstruction:** An opponent chooses the subdeck of x cards.
  $\Rightarrow$ We need to show that we can determine whether a graph G has property $\mathcal{P}$ based on any subdeck of G consisting of x cards.
- **Ally reconstruction:** An ally chooses the subdeck x cards.
  $\Rightarrow$ We need to show that we can determine whether a graph G has property $\mathcal{P}$ based on one specific subdeck of G consisting of x cards.

The number of cards one needs in the ally case is called the *ally recognition number* and in the adversary case the *adversary recognition number*. By definition, the ally recognition number of a certain property $\mathcal{P}$ is lower than or equal to the adversary recognition number. In the table below, the best-known upper bounds for the ally and adversary reconstruction numbers of some properties are given below.

	# Edges	Connectivity	k-Regularity	Trees
Ally	$\lfloor \frac{n}{2} \rfloor + 2$	3 [12]	$k + 3$ [12]	3 [14]
Adversary	$n - \frac{1}{20}\sqrt{n}$ [7]	$\lfloor \frac{n}{2} \rfloor + 2$ [12]	$k + 3$ [12]	?

Table 1: Best-known upper bounds (for n large enough) for the ally and adversary recognition numbers of certain properties.

Remark 1.8. *If the expression x for the number of common cards is not an integer, one should always interpret it as $\lceil x \rceil$ common cards.*

As we can see in this table, we need in most cases way fewer cards in the ally case than in the adversary case. However, for the property k-regularity we need the same number of cards. The reason why this is true is that for a vertex-transitive graph, all cards are isomorphic. And if all cards are isomorphic it does not matter whether an ally or an opponent picks the cards, because they will always choose the same cards.
Most of the bounds in this table are strict. This means that there exists (infinite) families of graphs $(G_i, H_i)_{i \in \mathbb{N}}$ such that G_i and H_i have lots of common cards and G_i satisfies some property while H_i does not. To illustrate this idea will give a well-known example:

Example 1.9. *Consider the graphs $G = K_{\frac{n}{2}} \cup K_{\frac{n}{2}}$ and $H = K_{\frac{n}{2}+1} \cup K_{\frac{n}{2}-1}$, see Figure 2 for $n = 8$. Then every card of G_n is isomorphic to $K_{\frac{n}{2}} \cup K_{\frac{n}{2}-1}$. Moreover, if we delete a vertex from $K_{\frac{n}{2}+1}$ in H, we also get a card isomorphic to $K_{\frac{n}{2}} \cup K_{\frac{n}{2}-1}$. This means that every subdeck of $\frac{n}{2}+1$ cards of G is also a subdeck of H.*

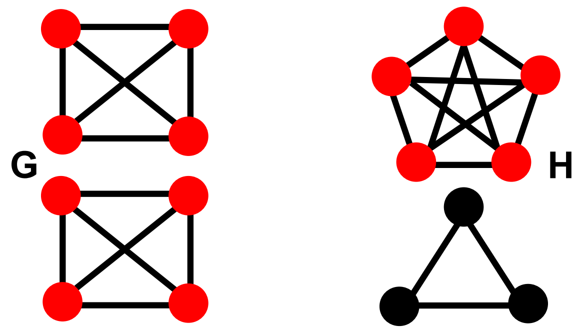


Figure 2: The graphs G and H of Example 1.9 where $n = 8$. If we delete any vertex of G and any red vertex of H, we get the same card. Since H has $\frac{n}{2}+1$ red vertices, the graphs have $\frac{n}{2}+1$ common cards.

For every $n \in \mathbb{N}$, the graph G is $(\frac{n}{2}-1)$-regular, while H is not regular. Therefore, the example above shows that the ally recognition number for k-regularity is at least $k+3$ as any deck of $\frac{n}{2}+1$ cards of a $\frac{n}{2}-1$-regular graph are also in the deck of a non-regular graph. In the example above $e(G) = \frac{n}{2}(\frac{n}{2}-1) = \frac{n^2}{4} - \frac{n}{2}$ and $e(H) = \frac{\frac{n}{2}(\frac{n}{2}+1)}{2} + \frac{(\frac{n}{2}-2)(\frac{n}{2}-1)}{2} = \frac{n^2}{4} - \frac{n}{2} - 1$. Therefore, this example also shows that the ally recognition number of the property edges is also at least $\lfloor \frac{n}{2} \rfloor + 2$.
Proving a (strict) upper bound is often the harder part. This is because many nice counting arguments, like Kelly's lemma 2.1, do not work when you do not get all cards. This means that you have to use other information, such as the (non-)existence of certain subgraphs of a card, in order to prove a property $\mathcal{P}$. However, it is often difficult to extract this kind of information in an organized way, such that you get this information for all possible graphs. This is the reason why the best-known upper bound for the number of edges is currently $n - \frac{1}{20}\sqrt{n}$, while the following is conjectured by Bowler, Brown, and Fenner [1]:

Conjecture 1.10 (Bowler, Brown, and Fenner [1])**.** *For large enough n, every graph is determined, up to isomorphism, by any $2\lfloor \frac{n-1}{3} \rfloor + 1$ of its cards.*

Since many mathematicians looked at the ally recognition number of properties, it is to be expected that someone tried proving an upper bound for the recognition number of the number of edges. However, as I did not find any proof or reference to this result, I added my proof of the following theorem in Chapter 3:

Theorem 1.11. *For every graph G there exists $\lfloor \frac{n}{2} \rfloor + 2$ cards in the deck of G such that these cards are not in the deck of any graph H that satisfies $e(G) \neq e(H)$.*

1.3 Recognizing trees in the adversary case

In this section, I will give an overview of the main results that I found while working on my thesis. I will introduce these results by giving some literature background and explaining some of the central ideas.
The example in Figure 3 is due to Wendy Myrvold [13]. This example is part of an infinite family of examples of trees G and disconnected graphs H such that the decks of these 2 graphs contain a common subset $\lfloor \frac{n}{2} \rfloor + 1$ cards. We can find the other members of this family by adding extra branches consisting of two points to the green vertices of G and H. In particular, this example indicates that we need at least $\lfloor \frac{n}{2} \rfloor + 2$ cards in the adversary case to determine whether G is connected.
As one can see in Table 1 it is yet unknown how many cards one needs in the adversary case to determine whether a certain graph is a tree. Since the graph G of Figure 3 is a tree, we need at least $\lfloor \frac{n}{2} \rfloor + 2$ cards to determine whether a graph is a tree in the adversary case. It has been conjectured that for n large enough, this number of cards also suffices.

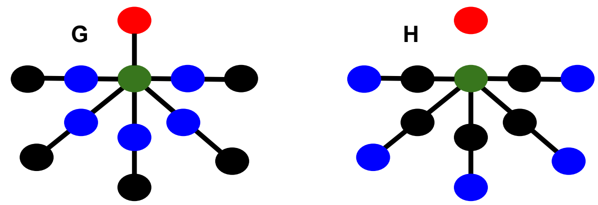


Figure 3: A tree G and a disconnected forest H that have $\frac{n}{2}+1$ common cards. Deleting non-black vertices of the same color in G and H results in the same common card.

Conjecture 1.12 (Bowler, Brown, and Fenner [1]). *For $n \geq 44$, it can be determined whether a graph is a tree from any $\lfloor \frac{n}{2} \rfloor + 2$ of its cards.*

To prove this conjecture, I looked at the properties that determine whether a graph is a tree. A graph is a tree if it satisfies at least two of the following three conditions: 'the graph is connected', 'the graph has $n-1$ edges,' and 'the graph does not contain a cycle.' In Table 2 one can find the best-known lower and upper bounds for determining these properties.

	# edges	connectivity	contains cycle	if connected contains cycle	is tree
lower bound	$2\lfloor \frac{n-1}{3} \rfloor + 1$ [1]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	$\lfloor \frac{2(n+1)}{5} \rfloor + 1$ [1]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]
upper bound	$n - \frac{1}{20}\sqrt{n}$ [7]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	?	?	?

Table 2: The best-known lower and upper bounds (for n large enough) for defining properties of a tree.

Table 2 shows that we need at most $\lfloor \frac{n}{2} \rfloor + 2$ cards to determine whether the graph is connected. This means that we only have to prove that can determine either 'the graph has $n-1$ edges' or 'the graph does not contain a cycle' based on any subdeck of $\lfloor \frac{n}{2} \rfloor + 2$ cards.
However, this table also shows that one needs very many cards from the deck of G to determine its number of edges. The upper bound found in 2021 by Groenland, Guggiari, and Scott [7] tells us that for n large enough one needs at most $n - \frac{1}{20}\sqrt{n}$ cards. This is not close to $\lfloor \frac{n}{2} \rfloor + 2$ especially if you consider the fact that the number of missing cards is not even linear in n. Furthermore, Bowler, Brown, and Fenner [1] found examples of graphs with a different number of edges that have $2\lfloor \frac{1}{3}(n-1) \rfloor$ cards in common. Namely, the graphs $G = K_{t+1} \cup K_t \cup K_t$ and $H = K_{t+1} \cup K_{t+1} \cup K_{t-1}$. These results both indicate that to find the minimum number of cards needed to determine whether a graph is a tree, one can not simply try to reconstruct $e(G)$. Therefore, in my thesis, I looked at the number of overlapping cards between a tree T and a connected non-tree G.
First, we show that if a tree T and a connected non-tree G have lots of common cards, the graph G can have at most one cycle.

Lemma 1.13. *A forest F and a graph G, containing at least two cycles, have at most two common cards.*

Proof. As the forest F does not contain any cycles, neither do any common cards of F and G. Next, we will now prove that G has at most two cards without any cycles. We use proof by contradiction. Suppose $G - v_i$ for $i = 1, 2, 3$ does not contain any cycles and let C and C' be two different cycles in G. Because none of the cards $G - v_i$ contain any cycles, we immediately see that all v_i lie both in C and C' as otherwise one of these cycles is visible on $G - v_i$. Since $C \neq C'$, at least one of the paths between two of the v_i (not containing the third vertex) is not the same in C as in C'. Without loss of generality, we may assume this is the path $v_1 v_2$ without v_3. Then the card $G - v_3$ contains two different paths between v_1 and v_2, implying that $G - v_3$ contains a closed walk and therefore $G - v_3$ contains a cycle. A contradiction. □

This lemma implies that the connected non-tree G has exactly one cycle. We, therefore, introduce the following concept.

Definition 1.14. *A graph is* ***unicyclic*** *if it contains exactly one cycle.*

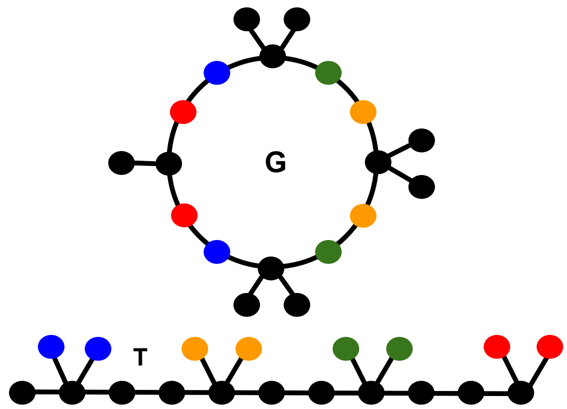


Figure 4: A member of the family found by Bowler, Brown, and Fenner [1] of a tree T and a unicyclic graph G with $\lfloor \frac{2}{5}(n+1) \rfloor$ common cards. Deleting non-black vertices of the same color in G and T results in the same common card.

This means that we will restrict ourselves to trees T and unicyclic connected graphs G with lots of common cards. Bowler, Brown, and Fenner [1] gave an example of infinitely many pairs of n-vertex trees and n-vertex connected non-trees that have $\lfloor \frac{2}{5}(n+1) \rfloor$ cards in common of which one member is shown in Figure 4. Based on their example, they stated the following conjecture:

Conjecture 1.15 (Bowler, Brown, and Fenner [1])**.** *For $n \geq 44$, the only pair of graphs on n vertices that have at least $\lfloor \frac{2}{5}(n+1) \rfloor$ common cards, where one is a tree and the other a connected non-tree, is the family corresponding to Figure 4.*

Conjecture 1.15 together with the fact that any subdeck of $\lfloor \frac{n}{2} \rfloor + 2$ cards determines whether a graph is connected implies Conjecture 1.12. However, to prove Conjecture 1.12 it also suffices to prove a (slightly) weaker version of Conjecture 1.15 namely that a tree T and a unicyclic graph G have at most $\lfloor \frac{n}{2} \rfloor + 2$ common cards.
In the example of Bowler, Brown, and Fenner (shown in Figure 4) the tree T is a caterpillar graph, which is a path plus some leaves, and the connected non-tree G is a sunshine graph, a cycle with some extra leaves. This led to further investigation into the overlapping cards between these special subclasses of trees and connected non-trees. In his PhD thesis [2], Paul Brown showed that the maximum number of overlapping cards between a sunshine and a caterpillar graph is at most $\lfloor \frac{2}{5}(n+1) \rfloor$ and that bound is only attained by the family that he earlier found together with Bowler and Fenner.
This result strengthens Conjecture 1.15, whose correctness implies there is a gap of around $\frac{n}{10}$ cards between the 'true bound' for the maximum number of common cards between a tree and a connected non-tree and the upper bound that is needed for proving Conjecture 1.12. The presumable existence of this gap was one of the main reasons why I decided to try proving Conjecture 1.12. This led to the following result whose proof is in Chapter 4.

Theorem 1.16. *For all $n \geq 5000$, it can be determined whether a graph with n vertices is a tree based on any $\lfloor \frac{n}{2} \rfloor + 2$ of its cards.*

Hence, I managed to prove Conjecture 1.12 for $n \geq 5000$. I did this by using colorings, that mark good and bad vertices on the cycle of the unicyclic graphs. These colorings are designed in such a way that good vertices lie far away from large branches and that at least one common card corresponds to a good vertex in G. These common cards give upper bounds for the length of the longest path in T in terms of the size of the cycle of G, which again gives an upper bound of $\lfloor \frac{n}{2} \rfloor + C$, where C is a constant, on the size of the cycle of G.
Thereafter, I showed that vertices in the cycle of G whose branches have 'maximal size' can not correspond to very many common cards of G and T. Hence, the cycle of G does not contain a lot of these vertices. Thus, the graph G has not a lot of symmetry, which will lead to a contradiction.

Due to the method that I used, I only proved the result for $n \geq 5000$ instead of the conjectured $n \geq 44$. Principally, Conjecture 1.12 holds except for maybe some finite number of graphs. If in the future computers get *a lot* faster, one can maybe prove the whole result by checking all trees and unicyclic graphs that satisfy the required constraints from the lemmas I used in the proof of this theorem on at most $n = 4999$ vertices.

1.4 More results about recognizing cycles

One can wonder, is it possible to adapt the proof to prove Conjecture 1.15? The answer is sadly no because in the example found by Bowler, Brown, and Fenner there is a large gap between the number of common cards of G and T, which is $\lfloor \frac{2(n+1)}{5} \rfloor$ and the number of vertices in the cycle of G, which is $\lfloor \frac{3(n+1)}{5} \rfloor$. This means that G can still have many vertices whose branches are of 'maximal size.' However, by adapting the proof I was able to prove the following slightly stronger statement:

Theorem 1.17. *For n large enough, a connected unicyclic graph G and a tree T have at most $\frac{n}{2} - \epsilon\sqrt{n}$ cards in common, where $\epsilon > 0$ is some fixed constant.*

The main importance of the theorem above is that it shows that the answer to the question 'How many cards do we need to determine whether a graph is a tree?' is determined by the number of cards one needs to determine whether a graph is connected and not by the number of cards one needs to determine whether a graph contains a cycle.
Thereafter, I looked at the generalization: How many cards does one need to determine whether a (not necessarily connected) graph contains a cycle? First, I found three well-known infinite families of pairs F_n and G_n such that F_n is always a forest and G is unicyclic both on n vertices, such that F_n and G_n have $\lfloor \frac{n}{2} \rfloor + 1$ common cards [1]. Thereafter, I tried to prove that there are no forests and non-forests with more common cards. In this proof, I used a case distinction with six (sub)cases. For case 2b, I used Theorem 1.16. Since this theorem only holds for $n \geq 5000$, the final result also only holds for $n \geq 5000$.

Theorem 1.18. *For $n \geq 5000$, a forest F and a non-forest G both on n vertices have at most $\lfloor \frac{n}{2} \rfloor + 1$ common cards.*

We can now write down all these results into Table 2 to get Table 3 with the new best-known bounds.

	# edges	connectivity	contains cycle	if connected contains cycle	is tree
lower bound	$2\lfloor \frac{n-1}{3} \rfloor + 1$ [1]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	$\lfloor \frac{2(n+1)}{5} \rfloor + 1$ [1]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]
upper bound	$n - \frac{1}{20}\sqrt{n}$ [7]	$\lfloor \frac{n}{2} \rfloor + 2$ [13]	$\lfloor \frac{n}{2} \rfloor + 2$	$\frac{n}{2} - \frac{1}{100}\sqrt{n}$	$\lfloor \frac{n}{2} \rfloor + 2$

Table 3: The best-known lower and upperbounds (for n large enough) for defining properties of a tree. The results shown in red are new.

Subsequently, I looked at pairs of graphs for which the upper bound of $\lfloor \frac{n}{2} \rfloor + 1$ is strict. I did this by looking at every case in the proof of Theorem 1.18 separately. I managed for all cases, except $2b$ that the only graphs, except for some very small graphs on 7 vertices, that attain this bound fall in one of the three families that I found before. However, I needed the proof of Theorem 1.17 to prove that there are no large pairs of graphs that fall into case $2b$. Since we can only apply this theorem for $n \geq 24000$ vertices, we have the following result

Theorem 1.19. *For $n \geq 24000$, the only forests F and non-forests G on n vertices that have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards, are part of the families generated by examples in Figure 7 and 8.*

I also used similar colorings to prove other results that are determined by the (non-)existence of certain cycles. After I proved some elementary graph-theoretical lemmas, I was quite easily able to show that both the girth, the length of the shortest cycle, and bipartiteness, the existence of cycles of odd length, can be determined even if there is a linear amount of missing cards.

Theorem 1.20. *For any graph on n vertices, we can determine its girth based on any $\frac{2n}{3} + 1$ of its cards.*

Theorem 1.21. *For any graph G on n vertices, it can be determined whether G is bipartite based on any $\frac{5n}{6} + 2$ of its cards.*

The proofs of these last two results indicate that my method of colorings is not only a nice trick that worked well for my proof of the upper bounds for the number of cards one needs to determine whether a graph is a tree, but that it also is a tool that can be used for other properties that involve cycle recognition.

2 Preliminaries

In this section, I will state some classical results in graph reconstruction and use some well-known techniques to prove some of them.
One of the first discovered results was Kelly's lemma, which although it is quite simple, is a very useful tool for proving that a certain property is recognizable. It namely tells us that based on the deck we can determine the number of subgraphs of G isomorphic to F if $v(F) < v(G)$.

Lemma 2.1 (Kelly's Lemma)**.** *For any two graph F and G with $v(F) < v(G)$ we can reconstruct the number of subgraphs of G isomorphic to F.*

Proof. Let G_F be a subgraph of G isomorphic to F. Then this subgraph occurs on a card $G - v$ if and only if $v \notin G_F$. Since there are $v(G) - v(G_F) = v(G) - v(F)$ such vertices, we find that the number of subgraphs of G isomorphic to F is equal to

$$\frac{1}{v(G) - v(F)} \sum_{v \in G} s(F, G - v),$$

where $s(F, G - v)$ is the number of subgraphs isomorphic to F in $G - v$ □

We will now give some nice corollaries of Kelly's Lemma.

Corollary 2.2. *The number of edges and the degree sequence are recognizable.*

Proof. Use Kelly's lemma with $F = K_2$ to find the number of edges $e(G)$. For the degree sequence, notice that the degree of a vertex v is equal to $\deg(v) = e(G) - e(G - v)$ as $G - v$ contains exactly those edges of G that are not incident to v. Since we can determine $e(G)$, we can also determine the degree of the missing vertex v. □

With this corollary, we can find our first class of reconstructible graphs.

Theorem 2.3. *Regular graphs are reconstructible.*

Proof. The previous corollary implies that we can determine whether a graph is k-regular. Moreover, if G is k-regular the card $G - v$ contains exactly k vertices of degree $k - 1$, which are exactly the neighbors of v. So, if G is k-regular, then we can place a vertex v back by connecting it to all vertices with degree $k - 1$. □

More generally, Kelly's lemma tells us that every graph property $\mathcal{P}$ that is determined by the subgraphs of a graph is recognizable from the deck.

Corollary 2.4. *The properties being a forest and bipartite are recognizable.*

Proof. Kelly's Lemma tells us that the deck of G determines whether G has an (odd) cycle of size at most $n - 1$. If the only (odd) cycle of G has size $n = |V|$, the graph G must be isomorphic to C_n. But C_n is a 2-regular graph and thus the previous lemma tells us that we can see from the deck whether G is C_n. □

Corollary 2.5. *The property being a tree is recognizable.*

Proof. This follows from the fact that the deck tells us both the number of edges and whether a graph is a forest (i.e. does not contain any cycles). □

Paul Kelly proved in 1957 something even stronger, namely that we can reconstruct every tree back from its deck [9]. The main idea of the proof is to look at the branches around the center for all cards that have a maximal diameter G. However, you have to be a bit careful as one or both of these branches from the center can be paths. A clear proof of this theorem can be found in the lecture notes of Alexandr Kostochka [10].

Theorem 2.6 (Kelly [9])**.** *All trees are reconstructible.*

Next, we prove that connectivity is recognizable.

Lemma 2.7. *The graph property connectivity is recognizable.*

Proof. We will prove the following claim: A graph G is connected if and only if at least 2 cards of G are connected. If G is connected, G has a spanning tree $T \subseteq G$. Deleting a leaf v from T gives a connected card $T - v \subseteq G - v$. Since every tree has at least 2 leaves, G has at least 2 connected cards. Now, suppose that G is disconnected and suppose that the card $G - v$ is connected. Then v is an isolated vertex as it can not be a neighbor of any vertex in $G - v$. Hence, if G is disconnected and has 2 connected cards, G has at least 2 isolated vertices. Then all cards of G contain at least one isolated vertex and thus none of the cards of G is connected. □

Another important class of reconstructible graphs is the class of disconnected graphs.

Lemma 2.8. *Disconnected graphs are reconstructible.*

Proof. First, Lemma 2.7 tells us that the deck determines the graph is disconnected. So the largest connected component of G has at most $n-1$ vertices. This means that there is a card on which this whole component is visible. Hence, the largest connected component of G, called M, is equal to the largest connected component over all cards.
Next, fix some (connected) subgraph L of M with $v(L) = v(M) - 1$. Then pick a card among the ones with the fewest copies of M that has the most copies of L-components. Then we can find G back by replacing one copy of L on that card with a copy of M. □

Another important observation is that the complements of the cards of a graph G are exactly the cards of the complement of G.

Observation 2.9 (The complementation principle)**.** *The cards of $\overline{G}$ are the complements of the card G.*

Proof. For every graph G and every vertex $v \in V(G)$ we have $\overline{G - v} = \overline{G} - v$. □

This principle gives us immediately the following corollaries.

Corollary 2.10. *The graph G is reconstructible if and only if $\overline{G}$ is.*

Corollary 2.11. *A graph property $\overline{\mathcal{P}}$ is recognizable if and only if the property $\overline{\mathcal{P}}$ is (i.e. what the property $\mathcal{P}$ becomes in $\overline{G}$).*

Of course, there are many more recognizable properties.

Theorem 2.12. *The following are all recognizable:*

(i) k-vertex-connectivity.

(ii) The Tutte polynomial, the chromatic polynomial, and the characteristic polynomial [15].

(iii) Planarity [5].

(iv) The number of spanning trees, Hamilton cycles, and perfect matchings [15].

Moreover, Brendan McKay showed in 2022 that all graphs on at least 3 and at most 13 vertices are reconstructible [11].

3 Ally recognition for number of edges

Example 1.9 tells us that the ally recognition number of the number of edges is at least $\lfloor \frac{n}{2} \rfloor + 2$ cards. The goal of this chapter is to prove that it is equal to $\lfloor \frac{n}{2} \rfloor + 2$. Therefore, we need to pick a clever subdeck of size $\lfloor \frac{n}{2} \rfloor + 2$ based on which we can determine the number of edges. We start with a simple observation about maximum and minimum degrees.

Definition 3.1. *The* ***maximum degree*** *in a graph G is denoted by $\Delta(G)$ and the* ***minimum degree*** *is denoted by $\delta(G)$.*

Observation 3.2. *If two graphs G and H have at least two common cards, then $\Delta(G) \leq \Delta(H) + 1$ and $\delta(G) \geq \delta(H) - 1$.*

Proof. Since G and H have at least two common cards, there is a common card $G - v = H - w$ such that a node of maximum degree v' of G is visible. However, as v can be a neighbor of v', we obtain $\Delta(H) \geq \deg_{G-v}(v') \geq \deg_G(v') - 1$. Similarly, there exists a common card $G - v = H - w$ such that a node of minimum degree v'' is visible. Because v can be a neighbor of v'', we obtain $\delta(H) - 1 \leq \delta_{G-v}(v'') \leq \delta(G)$. □

Next, we prove a lemma that generalizes the previous observation. Namely, that we can determine the entire degree sequence with a maximum deviation of one for every vertex based on any subdeck of size at least $\lfloor \frac{n}{2} \rfloor + 1$.

Lemma 3.3. *Let G and H be two graphs on n vertices such that both their decks contain the same subdeck D of size at least $\frac{n}{2} + 1$. Let $\{d_i\}_{i=1}^n$ and $\{e_i\}_{i=1}^n$ be the ordered degree sequences of G and H from low to high. Then $|d_i - e_i| \leq 1$ for all i.*

Proof. Suppose that $|d_i - e_i| > 1$ for some i. Because the statement is symmetric in G and H, we may assume $d_i - e_i > 1$. Moreover, we can use the complementation principle to assume without loss of generality $i \leq \frac{n}{2}$. If $i \leq \frac{n}{2}$, there exists a card in the deck corresponding to some e_j with $j > \frac{n}{2}$. On this card of H, there are at least i vertices with degree $\leq e_i$. Since this card is also a card of G, the graph G must have at least i vertices with degree $\leq e_i + 1 < d_i$, a contradiction. □

With this lemma, we can show the ally recognition number of the number of edges is at most $\frac{n}{2} + 2$ for regular graphs. This is an important special case as the graph from Example 1.9 is a regular graph implying that even for regular graphs this bound is strict.

Lemma 3.4. *Let G be an arbitrary k-regular graph. Let D be any subdeck of G consisting of at least $\lfloor \frac{n}{2} \rfloor + 2$ cards. Then D is not a subdeck of any graph H with $e(H) \neq e(G)$.*

Proof. Suppose that this is not true. As the complement of a regular graph is again a regular graph, we can assume without loss of generality using the complement principle that $e(H) > e(G)$.
Now, H has at least $\lfloor \frac{n}{2} \rfloor + 2$ vertices of degree $k + e(H) - e(G)$. Lemma 3.3 tells us $k + e(H) - e(G) \leq k + 1$ implying $e(H) = e(G) + 1$ and that all other vertices have degree at least $k - 1$. Hence,

$$\begin{aligned} 2e(H) &\geq (k+1)\left(\left\lfloor \frac{n}{2} \right\rfloor + 2\right) + (k-1)\left(n - \left\lfloor \frac{n}{2} \right\rfloor - 2\right) \\ &= kn + 2\left\lfloor \frac{n}{2} \right\rfloor - n + 4 \\ &\geq kn + 3 > 2e(G) + 2 = 2e(H). \end{aligned}$$ □

We will now do the same for non-regular graphs. For a non-regular graph G, we define M to be the number of vertices of maximum degree in G and N to be the number of vertices of minimum degree. As G is non-regular, we have $M + N \leq n$.

Lemma 3.5. *There exists integers A and B satisfying $M < 2A - 1$, $N < 2B - 1$ and $A + B \leq \frac{n}{2} + 3$.*

Proof. Rewriting the equations, the numbers A and B must satisfy $A > \frac{M+1}{2}$ and $B > \frac{N+1}{2}$. Take $A = \lfloor \frac{M+1}{2} \rfloor + 1$ and $B = \lfloor \frac{N+1}{2} \rfloor + 1$, then $A + B \leq \frac{M+N+2}{2} + 2 \leq \frac{n}{2} + 3$. □

Lemma 3.6. *Let G be a non-regular graph on n vertices. Let D' be a subdeck of G consisting of at least $\lfloor \frac{n}{2} \rfloor + 2$ cards such that A cards correspond to vertices of the highest possible degrees; if two vertices have the same degree we pick the one with the most neighbors of maximum degree. If the subdeck D' is also part of the deck of a graph H satisfying $e(H) > e(G)$, then G has at least $2A - 1$ vertices of maximum degree.*

Proof. If $A = 0$ or $A = 1$ the statement is trivial, so suppose that $A \geq 2$. If $A \geq \lfloor \frac{n}{2} \rfloor + 2$ then $2A - 1 \geq n$ implying that G must be regular, which is not allowed by our assumptions.
By construction, D' contains a card corresponding to a vertex of degree $\Delta(G)$. Hence, the graph H must have a vertex of degree $\Delta(G) + e(H) - e(G)$. Because none of the cards of G contain a vertex of degree at least $\Delta(G)+1$, the cards in subdeck D' also do not contain such vertices. Since D' is a subdeck of H with at least two cards, the observation tells us $\Delta(H) \leq \Delta(G)+1$ implying $\Delta(H) = \Delta(G)+1$ and $e(H)-e(G) = 1$. Because none of the cards in D' contain a vertex of degree at least $\Delta(G) + 1$, we find that all vertices in H corresponding to a card in D' must be adjacent to all other vertices of degree $\Delta(G)+1$. Let $G-v = H-w$, then $\Delta(G)+1 = \Delta(H) \geq \deg(w) \geq |D' - \{H-w\}| = \lfloor \frac{n}{2} \rfloor + 1$ implying $\Delta(G) \geq \lfloor \frac{n}{2} \rfloor$.
Next, we will now show that $M \neq 1$ by using proof by contradiction. Suppose there is a unique vertex of maximum degree $\Delta(G)$. Since D' contains at least $\lfloor \frac{n}{2} \rfloor + 2$ cards of G, this implies that D' contains at least one card corresponding to a neighbor of this vertex. By assumption, we also selected the card corresponding to the vertex of maximum degree. This means that D' contains at least two cards that have maximum degree $\Delta(G)-1$. But as H has a vertex of degree $\Delta(G)+1$, the graph H has at most one card that does not contain a vertex of degree $\Delta(G)$ (or higher). Thus, D' is not a subdeck of H', a contradiction.
The subdeck D' contains $\min(A, M) \geq 2$ cards corresponding to a vertex of degree $\Delta(G)$ in G. These cards correspond to vertices of degree $\Delta(G)+1$ in H. Therefore, the graph H contains a clique of size $\min(A, M)$ of vertices of degree $\Delta(G) + 1$. Thus, every card in D', that is not one of the A special cards, contains a subgraph $K_{\min(A,M)}$ consisting of vertices of degree $\Delta(G)$. Similarly, the A special cards contain a subgraph $K_{\min(A,M)-1}$ consisting of vertices of degree $\Delta(G)$.
The existence of a subgraph $K_{\min(A,M)}$ of vertices of degree $\Delta(G)$ implies that there are vertices of maximum degree that have at least $\min(A, M) - 1$ neighbors of degree $\Delta(G)$. By construction of D, this subdeck contains a card $G-v$ corresponding to such a vertex v. This card contains at least $\min(A, M)-1$ vertices of degree $\Delta(G)$. Moreover, to obtain $G-v$ one vertex of maximum degree is deleted and the degree of at least $\min(A, M)-1$ vertices is dropped from $\Delta(G)$ to $\Delta(G-1)$. Hence, G has at least $2\min(A, M)-1$ vertices of maximum degree. Thus, $M \geq 2\min(A, M)-1$. However, $2M-1 > M$ for $M \geq 2$ implying $\min(A, M) = A$. Therefore, G contains at least $2A - 1$ vertices of maximum degree. □

If we apply the complementation principle to the previous lemma, we immediately get the following lemma:

Lemma 3.7. *Let G be an arbitrary non-regular graph on n vertices. Let D'' be a subdeck of G consisting of at least $\lfloor \frac{n}{2} \rfloor + 2$ cards such that B cards correspond to vertices of the lowest possible degrees, if two vertices have the same degree we pick the one with the least neighbors of minimum degree. If the subdeck D'' is also part of the deck of a graph H satisfying $e(H) < e(G)$, then G has at least $2B - 1$ vertices of minimum degree.*

Corollary 3.8. *For every graph G there exist $\frac{n}{2} + 3$ cards in the deck of G such that these cards are not in the deck of any graph H that satisfies $e(G) \neq e(H)$.*

Proof. If G is regular, we use Lemma 3.4.
If G is not regular, we let A and B be the numbers of Lemma 3.5 and let D' be a subdeck of $\frac{n}{2} + 3$ cards consisting of at least the following cards:

- A cards corresponding to the vertices of the highest degrees, where in case of a tie we select the cards the vertices with the most neighbors of maximum degree.
- B cards corresponding to the vertices of the lowest degrees, where in case of a tie we select the cards the vertices with the least neighbors of minimum degree.

Let H be an arbitrary graph such that $e(H) \neq e(G)$ and the deck of H also contains the subdeck D'.
If $e(H) > e(G)$, we apply Lemma 3.6 with A to obtain that there are at least $2A - 1$ vertices of maximum degree, which leads to a contradiction because $M < 2A - 1$.
If $e(H) < e(G)$, we apply Lemma 3.7 with B to obtain that there are at least $2B - 1$ vertices of maximum degree, which leads to a contradiction as $N < 2B - 1$. □

Although the previous corollary gives us an upper bound of $\frac{n}{2}+3$ for the ally recognition number for the number of edges, we, of course, want to show that it is at most $\lfloor \frac{n}{2} \rfloor + 2$. This is a little bit more work, but we will prove this in the next theorem.

Theorem 1.11. *For every graph G there exists $\lfloor \frac{n}{2} \rfloor + 2$ cards in the deck of G such that these cards are not in the deck of any graph H that satisfies $e(G) \neq e(H)$.*

Proof. Note by looking at the proof of the previous corollary, we only need a deck of $\lfloor \frac{n}{2} \rfloor + 3$ cards if $A + B > \lfloor \frac{n}{2} \rfloor + 2$. This only happens if $\lfloor \frac{M+1}{2} \rfloor + \lfloor \frac{N+1}{1} \rfloor > \lfloor \frac{n}{2} \rfloor$. Since all numbers are integers this is only true $\lfloor \frac{M+1}{2} \rfloor + \lfloor \frac{N+1}{2} \rfloor \geq \lfloor \frac{n}{2} \rfloor + 1 = \lfloor \frac{n+2}{2} \rfloor$. Hence, this only happens if $M + N = n$ and if at least one of M and N is odd or if $M + N = n - 1$ and both M and N are odd. By the complement principle, we can assume without loss of generality that M is odd. We take A such that $2A - 1 = M$ (so $1 \leq A \leq M$) and $B = \lfloor \frac{n}{2} \rfloor + 2 - A$. By construction,

$$2B - 1 \geq n - 1 + 4 - 2A - 1 = n - (2A - 1) + 1 \geq n - M + 1 = N + 1.$$

We will now define our deck D. The subdeck D will consist of A cards of maximum degree, where in case of a tie we select the cards corresponding to the vertices with the most neighbors of maximum degree and B cards of minimum degree, where in case of a tie we select the card corresponding to the vertex with the least neighbors of minimum degree. This gives us a subdeck of exactly $\lfloor \frac{n}{2} \rfloor + 2$ cards.
As $2B - 1 > N$, Lemma 3.7 implies that there is no graph H with $e(H) < e(G)$ such that D is a subdeck of H. Thus, we will spend the rest of the proof on the case $e(H) > e(G)$.
As $e(H) > e(G)$, we can copy the proof of Lemma 3.6 to prove that $e(H) = e(G) + 1$ and H has a subgraph K_A of vertices of degree $\Delta(G) + 1$. Moreover, $\Delta(G) \geq \lfloor \frac{n}{2} \rfloor$. This again implies that we have selected at least one neighbor of a vertex of maximum degree. Therefore, we can again prove that $M \neq 1$ and thus $A \geq 2$.

We will now show that the induced subgraph of G consisting of vertices of degree $\Delta(G)$ is ismorphic to $K_A \cup K_{A-1}$. Similar to before, every card in D not corresponding to a vertex of degree $\Delta(G)+1$ in H in the subgraph K_A has a subgraph K_A of vertices of degree $\Delta(G)$, while all other cards of H have a subgraph K_{A-1} of vertices of degree $\Delta(G)$. Thus, every card in D has a subgraph K_{A-1} of vertices of degree $\Delta(G)$. Hence, every vertex of maximum degree corresponding to a card is not adjacent to at least $A - 1$ other vertices of maximum degree. As there are exactly $2A - 1$ vertices of maximal degree, these vertices of maximal degree are adjacent to at most $A - 1$ other vertices of maximum degree. Because we picked the cards corresponding to the vertices of maximum degree with the most neighbors of maximum degree, we find that every vertex of maximum degree in G is adjacent to at most $A - 1$ other vertices of maximum degree.
On the other hand, G contains a subgraph K_A of vertices of maximum degree, so there are at least A vertices of maximum degree that are adjacent to one another and thus have $A - 1$ neighbors of maximum degree. So the other $A - 1$ vertices of maximum degree in G are not adjacent to these vertices and thus these are adjacent to at most $A - 2$ vertices of maximum degree. By the construction of our deck, every vertex in the subgraph K_A of vertices of maximum degree corresponds to a card. However, we already know that such cards have a subgraph K_{A-1} of vertices of maximum degree. Therefore, the induced subgraph of G only consisting of the vertices of maximal degree is isomorphic to $K_A \cup K_{A-1}$.

Let v be a vertex of non-maximal degree corresponding to a card $G - v = H - w$ in D. Then $H - w$ has a subgraph K_A of vertices of degree $\Delta(G)$. Therefore, the graph $G - v$ has that too. Therefore, v is not adjacent to any of the vertices in the subgraph K_A of maximal degree vertices. So there are at least B vertices of non-maximal degree that are not adjacent to any of the vertices in the subgraph K_A.
On the other hand, the graph H contains at least B vertices of degree not equal to $\Delta(G)+1$ that are incident to all vertices in the subgraph K_A of vertices of degree $\Delta(G) + 1$. On a card where we deleted one of the B vertices in H that lie outside the subgraph K_A of vertices of degree $\Delta(G)$+1, there are $B - 1$ vertices outside the clique K_A of vertices of degree $\Delta(G)$ that are connected to all vertices in this clique. Consequently, G has at least $B - 1$ extra vertices outside the component of K_A that are adjacent to all vertices in the component K_A of maximum degree vertices.
In conclusion, G has at least $2A - 1 + 2B - 1 = 2(A + B) - 2 = 2(\lfloor \frac{n}{2} \rfloor + 2) - 2 \geq 2n + 3 - 2 = n + 1$ vertices, a contradiction. □

4 Adversary recognition for trees

In this chapter, we will prove one of the main theorems of my thesis:

Theorem 1.16. *For all $n \geq 5000$, it can be determined whether a graph with n vertices is a tree based on any $\lfloor \frac{n}{2} \rfloor + 2$ of its cards.*

First, we will give a proof overview in which the key ideas are explained. The detailed proof is written down in the second section. Lastly, Section 4.3 cards contains the proof of the slightly stronger upper bound of at most $\frac{n}{2} - \epsilon\sqrt{n}$ common cards.

4.1 Proof overview

We will look at a (fixed) tree T and a connected non-tree G both on n vertices that have $\lfloor \frac{n}{2} \rfloor + 2$ common cards. First of all, every subgraph of T is a forest and therefore every common card of G and T is also a forest. If G has at least 2 cycles, then Lemma 1.13 tells us that T and G have at most 2 common cards. Hence, G has exactly one cycle, so G is *unicyclic* and every card corresponds to a vertex in the cycle of G. So the length L of the unique cycle of G is at least $\lfloor \frac{n}{2} \rfloor + 2$. Moreover, as $e(G) = n = e(T) + 1$, a common card $T - v = G - w$ satisfies $\deg(w) = \deg(v) + 1$. Because trees have a lot of vertices of degrees 1 and 2, we can use a counting argument to find a lower bound on the number of leaves of G. This then results in an upper bound for the number of vertices of degree at least 2 in the branches of G.
Next, we will use the coloring in Figure 5 to make a distinction between vertices on the cycle that lie close to large branches (the red vertices) and the vertices that lie far enough from large branches (the white vertices). Then, we use a counting argument to show that at least one common card corresponds to a white vertex w in G. Since the large branches lie 'far enough' away from w, the longest path on this card will only be a little longer than the length of the cycle itself. As $G - w = T - v$ consists of one large component and some isolated vertices, this also gives an upper bound for the longest path in T.
Thereafter, we will use another coloring, see Figure 6, to find a vertex $w \in V(G)$ such that $G - w = T - v$ consists of one large component and some isolated vertices and such that none of the vertices v' in the 'middle' of the longest path of $G - w = T - v$ can correspond to a common card of T and G. We prove this by showing that every path in $T - v'$ that uses a branch connected to the 'middle' of the longest path is at least 8 shorter than the longest path itself, which implies that the longest path in $T - v - v'$ will be somewhat shorter than $L - 1$. Adding the vertex v back increases the longest path by at most 2. Thus, we can conclude that $T - v'$ is not a common card.
As a lot of vertices of the longest path can not correspond to a common card, we will obtain an upper bound for L. This upper bound implies that almost all points on the unique cycle of G must correspond to a common card. We combine this with the fact that T is a tree to show that there are very few vertices on the cycle of G such that the sum of the sizes of the branches incident to it, is maximal. So the gaps between these branches will heavily increase as n goes to infinity. Hence, these branches will be in different places on different cards of G. On the other hand, these branches will be in (somewhat) the same places on different cards of T as T is a tree. This will yield a contradiction.

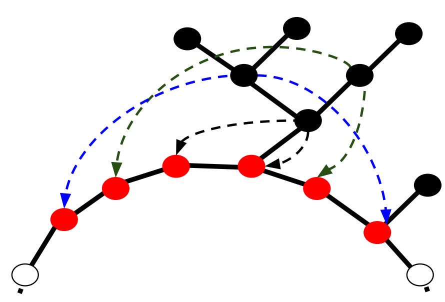


Figure 5: The coloring used in Lemma 4.6. Every vertex of degree at least 2 in a branch colors two vertices in the cycle red with the red vertices chosen to lie as close as possible to the branch.

4.2 Proof of main theorem trees

Throughout this section T will always be a tree and G will always be a non-tree such that T and G have at least $\lfloor \frac{n}{2} \rfloor + 2$ common cards. We will start by showing that G is a connected unicyclic graph. Moreover, we use the convention that all vertices 'v' lie in T, while all vertices 'w' lie in G.

4.2.1 First observations

We will start by making some important observations and using counting arguments to prove some lemmas. According to a theorem by Myrvold [12], the number of overlapping cards in the deck of a connected and a disconnected graph is at most $\lfloor \frac{n}{2} \rfloor + 1$. So we may assume that G is connected. Moreover, if a connected graph G has more than one cycle, Lemma 1.13 tells us that G and T have at most two common cards. Therefore, we are only left with the case in which a tree T and a connected unicyclic graph G have at least $\lfloor \frac{n}{2} \rfloor + 2$ common cards.
As T is a tree, it has $e(T) = n-1$ edges. Moreover, $e(G) = n = e(T)+1$ as G is connected and contains exactly one cycle. Since T does not contain a cycle, neither do common cards. Hence, every vertex corresponding to a common card must lie in the cycle of G. This immediately yields the following result:

Observation 4.1. *Let L be the length of the unique cycle of G. Then $L \geq \lfloor \frac{n}{2} \rfloor + 2$.*

Moreover, because every card contains at least $L - 1$ vertices of this cycle and these vertices lie all in one path, we obtain the following:

Observation 4.2. *Every card of G contains a path of length $L - 1$.*

Next, we will use a counting argument to show that there is a common card corresponding to a leaf in T.

Lemma 4.3. *There is a common card $C = T - v = G - w$ such that v has degree 1.*

Proof. Every common card $C_i = T - v_i = G - w_i$ satisfies $\deg(w_i) = \deg(v_i) + 1$ as $e(G) = e(T) + 1$. So if $\deg(v_i) \neq 1$ for all vertices v_i of T corresponding to a common card, $\deg(w_i) \geq 3$ for all i. Hence, G has at least $\lfloor \frac{n}{2} \rfloor + 2$ vertices of degree at least 3. Since G is connected all other vertices of G have degree at least 1. Hence,

$$2e(G) \geq 3\left(\left\lfloor \frac{n}{2} \right\rfloor + 2\right) + \left\lceil \frac{n}{2} \right\rceil - 2 \geq 3\left(\frac{n}{2} + \frac{3}{2}\right) + \frac{n}{2} - \frac{3}{2} = 2n + 3 > 2n = 2e(G).$$

A contradiction. $\square$

Next, we prove a lower bound for the number of leaves in G, which gives an upper bound for the number of vertices of degree at least 2 outside the cycle of G. We define $\text{Leaf}(X)$ to be the number of leaves in X.

Lemma 4.4. *G has at least $\frac{n}{4} - 1$ leaves.*

Proof. We use proof by contradiction. If the statement is not true, then G has at most $\frac{n}{4} - \frac{5}{4}$ vertices of degree 1, as the number of vertices of degree 1 is always an integer. Moreover, since G is connected G does not contain any isolated vertices.
By Lemma 4.3 there is a common card $C = G - w = T - v$ such that v is a leaf in T implying that C is connected. Depending on whether v's unique neighbor has degree 2, C has either $\text{Leaf}(T)$ or $\text{Leaf}(T) - 1$ leaves. Adding an extra vertex of degree 2 to that card does not increase the number of leaves, but it can decrease the number of leaves by at most 2. Hence, $\text{Leaf}(T) - 3 \leq \text{Leaf}(G)$ which implies that $\text{Leaf}(T) \leq \frac{n}{4} + \frac{7}{4}$. So at least $\frac{n}{4} - \frac{1}{4}$ of the common cards of T correspond to a vertex of degree 2 or more. This indicates that G has at least $\frac{n}{4} - \frac{1}{4}$ vertices of degree at least 3 and all vertices which are not leaves have degree at least 2. Hence,

$$2e(G) \geq 3\left(\frac{n}{4} - \frac{1}{4}\right) + \left(\frac{n}{4} - \frac{5}{4}\right) + 2\left(\frac{n}{2} + \frac{3}{2}\right) = 2n + 1 > 2e(G).$$

A contradiction. $\square$

Observation 4.5. *G has at most $\frac{3n}{4} - L + 1$ vertices of degree at least 2 outside its cycle.*

Proof. There are exactly $n - L$ vertices outside the cycle, of which at least $\frac{n}{4} - 1$ are leaves according to Lemma 4.4. Thus, there are at most $n - L - \frac{n}{4} + 1 = \frac{3n}{4} - L + 1$ vertices of degree at least 2 outside the cycle. $\square$

4.2.2 Bounding the length of the cycle

In this subsection, we will bound the length of the cycle. As said in the proof overview, we start by bounding the length of the longest path of T in terms of the length L of the cycle in G.

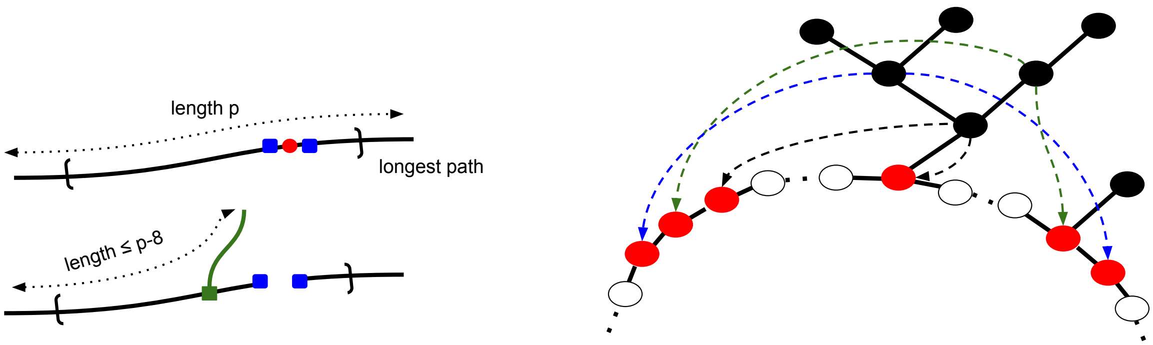
Lemma 4.6. *The longest path of T has length at most $L + 4$.*

Proof. We fix a certain drawing of G. For each branch of G connected to a vertex w in the cycle with exactly k vertices of degree at least 2, we color the vertices on the cycle which lie within distance $k - 1$ of w red; additionally, we color the vertex to the left with distance k also red. We do this for every vertex w. See Figure 5 for a small sketch of this coloring. The advantage of this coloring is that it makes a distinction between vertices that lie close to and far enough away from large branches.
Because the branches contain at most $\frac{3n}{4}+1-L$ vertices of degree at least 2 and every such vertex corresponds to at most 2 red vertices, there are at most $\frac{3n}{2} + 2 - 2L$ red vertices in the cycle of G. Hence, there are $L-(\frac{3n}{2}+2-2L) \geq 3(L-\lfloor\frac{n}{2}\rfloor-2)+\frac{5}{2}$ white vertices on the cycle. Since there are exactly $L-\lfloor\frac{n}{2}\rfloor-2 \geq 0$ vertices on the cycle that do not correspond to a common card, there is at least one white vertex w corresponding to a common card $C = G - w = T - v$.
Now, we demonstrate that the card C can only correspond to a tree with a longest path of length at most $L + 4$. Because no branch connected to w contains a vertex of degree at least 2, the card C consists of one large component and possibly some isolated vertices. We color all vertices in C that lay in the cycle of G blue. Because every branch with a height of k contains at least $k - 1$ vertices of degree at least 2, every such branch has a distance of at least $k - 2$ from the right end of the blue path and of at least $k - 1$ from the left end. Moreover, branches of height 1 extend the blue path by at most one at both ends. So the longest path is at most two longer than the blue path at the left end and at most one longer at the right end. Hence, the maximum path length on this card is at most $L - 1 + 1 + 2 = L + 2$.
As T is a tree, adding back v to T only lets the largest component grow with v and possibly leaves. Therefore, the longest path will increase by at most two. Thus, the longest path in T has a maximum length of $L + 2 + 2 = L + 4$. □

Since the longest path in T is at most $L + 4$, every path on a common card also has length at most $L + 4$. We will use this to prove that there is some special common card C, such that the only vertices on a longest path P corresponding to common cards are the vertices lying at one of the ends of P (see Figure 6a). Since at most $\lfloor\frac{n}{2}\rfloor - 2$ vertices of G do not correspond to a common card, this observation will give us an upper bound for L.

Lemma 4.7. *The unique cycle in G contains at most $\frac{n}{2} + 19.5$ vertices.*

Proof. Let w be an arbitrary vertex on the cycle of G. If the branch connected to w contains exactly k vertices of degree at least 2, we color the vertex w itself red, as well as all vertices on the cycle that lie within distance $[11, k + 9]$ red. So if $k = 1$ we only color w red. Moreover, we also color the vertex at distance 10 to the left red. Hence, for every vertex of degree at least 2 in a branch, we color at most two vertices on the cycle red. Similar to the proof of the previous lemma, there is at least one common card $G - w' = T - v'$ corresponding to a white vertex w' in G. Since w' is white, the common card $G - w' - T - v'$ consists of one large component and maybe some isolated vertices.
Moreover, we color all points in $G - w'$ corresponding to a vertex on the cycle blue. For our argument, we do not need to know which vertices are colored blue, only that these vertices exist. Besides, let P be a longest path of $G - w' = T - v'$.
First, we show that all except for maybe the 10 vertices on both ends of the blue path also lie in P. Suppose that this is not true. Because we have deleted a white vertex, every branch that lies at distance $d \geq 10$ of both ends of the blue path, has a height of at most $d - 8$, as all but one point in the path that determines the height has degree at least 2. So by taking this branch instead of following the blue path, the path becomes only smaller. Hence, these kinds of branches are not part of the longest path. Thus, P contains all vertices on the blue path that have distance at least 11 from both endpoints.
We will now show that all the vertices in the blue path, which are not in the last 10 vertices on either end, do not correspond to a common card. Let v be a vertex on the blue path such that v has a distance of at least 10 from both ends of the blue path. Previous observations tell us that v lies in the longest path P.



(a) Find a vertex $w' \in V(G)$ such that in $G - w' = T - v'$ not a lot of vertices can correspond to a common card. More specifically, we will show that the longest path in $T - v''$ will be too small for all vertices v'' in the middle of a longest path of $G - w' = T - v'$.

(b) The coloring used in Lemma 4.7. Every vertex of degree least 2 corresponds to two red vertices on the cycle, but this time we have gaps of 9 and 10 white vertices between the vertex connected to the branch and the other red vertices.

Figure 6: The goal and a picture of the coloring in Lemma 4.7.

First, we show that in $T - v' - v$, a path Q that follows the blue path until it reaches the neighbor of v is too small. We can extend the path Q by adding back v. Since both parts of P looking from the vertex v are as long as possible, the path Q is smaller or equal than the longest one of these parts of P. Moreover, both ends of P from v are at least as long as the respective ends of the blue path. Therefore, the path P becomes at least 11 shorter when we end the path P at the vertex before v. Thus, $Q \leq P - 11 \leq L + 4 - 11 = L - 7$. By adding back the vertex v' to $T - v' - v$, we can extend Q only by v' and some leaves (as T is a tree). So, the extension of Q in $T - v$ has length at most $L - 5$.
Next, we show that every path Q' in $T - v - v'$ that takes branch incident to a vertex x with such that there are at least 10 vertices before/after x on the blue path in $T - v'$ is too small. By our choice of w', the height of this branch is at most $d - 8$, because otherwise, it would have at least $d - 7 - 1 = d - 8$ vertices of degree at least 2 implying that w' was red. This means that taking such a branch results in a path that is at least 8 vertices smaller than the respective end of the blue path. Moreover, as the longest path is at most 5 vertices longer than the blue path, we can only add at most 5 vertices in $G - w'$ to a path containing this branch by taking a branch near the other end of the blue path. This means that the length of Q' is at most $L - 1 - 8 + 5 = L - 4$. Adding back the vertex v' makes the path at most 2 longer. So the extension of Q has length at most $L - 2$ in $T - v$.
Combining these two steps, we find that the longest $T - v$ path is at most $L - 2$. Since every card of G must contain a path of at least $L - 1$, the card $T - v$ is not in the deck of G. This indicates that there are at most 20 vertices on the blue path in $T - v'$ that can correspond to a common card. So there are at least $L - 1 - 20 = L - 21$ vertices of T which do not correspond to a common card. Hence, $L - 21 \leq \frac{n}{2} - 1.5$ implying that $L \leq \frac{n}{2} + 19.5$. □

In particular, this implies that there are at most 18 vertices on the cycle of G that do not correspond to a common card of G and T.

Corollary 4.8. *The maximum height of a branch of G is 14.*

Proof. We use proof by contradiction. Suppose there is a vertex w that has a branch with height at least 15. Given that the longest path on each common card is at most $L + 4$, we can conclude that none of the vertices on the G cycle with distance $d \in [1, 10]$ of w can correspond to a common card. For such vertices, we can namely use that this branch and the largest part of the blue path to obtain a path of length at least $L - 1 + 15 - (d - 1) \geq L + 5$, which should be impossible. This indicates that at least 20 vertices on the cycle do not correspond to a common card, which forms a contradiction with the remark above this corollary. □

Corollary 4.9. *A path outside the largest component on a common card of $G - w = T - v$ has length at most 27.*

Proof. Small components in common cards only consist of vertices that are in one of the branches of G. From the previous corollary, it follows that every branch has a height of at most 14. Therefore, every path consisting only of vertices in branches has length at most $2 \cdot 14 - 1 = 27$ as we only consider paths that do not contain the root, which lies in the cycle of G. □

4.2.3 Proof of the main theorem

Theorem 1.16. *For all $n \geq 5000$, it can be determined whether a graph with n vertices is a tree based on any $\lfloor \frac{n}{2} \rfloor + 2$ of its cards.*

Proof. First, we can determine whether a graph is connected from any $\lfloor \frac{n}{2} \rfloor + 2$ of its cards. So we only need to examine the possibility that some tree T and a connected unicyclic graph G have at least $\lfloor \frac{n}{2} \rfloor + 2$ cards in common.

Let K be the maximum sum of the sizes of the branches incident to a vertex in the cycle of G. Moreover, we color all vertices w' for which this maximum is attained green. So a vertex w is green if and only if the largest component in $G - w$ has exactly $n - 1 - K$ vertices.

Lemma 4.3 gives us a common card $T - v = G - w$ where v is a leaf and w is a vertex on the cycle of degree 2. Hence, this common card C is connected. We again color all vertices in $T - v = G - w$ that lie in the cycle of G blue. If we add the leaf v back to $T - v$, exactly one branch becomes larger. So there is at most one blue vertex for which the sum of the branches is at least $K + 1$. This implies that there is at most one non-blue vertex in T such that the largest component has exactly $n - 1 - K$ vertices.

If we delete vertices from the blue path, we observe that the largest component of T becomes strictly smaller for every vertex we move closer to the center (until we reach the middle). Because we can approach the center from (only) two sides, both the endpoints of the blue path, there are at most 2 blue vertices such that the largest component has exactly $n - 1 - K$ vertices. In conclusion, there are at most 3 cards of T such that the largest component has exactly $n - 1 - K$ vertices.

This also implies that there are at most 3 such common cards. So at most 3 green vertices of G can correspond to a common card. Since there are at most 18 vertices in the cycle of G that do not correspond to a card, there are at most 21 green vertices in the cycle of G.

Let $n \geq 5000$, then $L \geq \frac{n}{2} + 2 > 2500$. Because there are at most 21 green vertices, the pigeonhole principle tells us that there is a 'gap' between two green vertices consisting of at least $\frac{2500-21}{21} > 118$ adjacent non-green vertices on the cycle. Of these 118 vertices, at most 18 do not correspond to a common card.

Let w' be one of the vertices in this gap of 118 vertices such that there are at least 29 vertices before/after w' in this gap. There are at least $118 - 2 \cdot 29 = 60$ such vertices w' and at least 42 of them correspond to a common card. We will show that the green vertices in $G - w' = T - v'$ are exactly those vertices in the large component that split the large component into two components that have a path of length (at least) 29 and exactly K other vertices. Since w' lies on the blue path and the distance between w' and any green vertex is at least 29, the statement holds for all green vertices. Let w'' be a vertex that satisfies this condition, we will show that w'' is green. By Corollary 4.9 all paths consisting only of vertices in the branches have a length of at most 27. This means that both components with a path of length 29 must contain some blue vertices. Hence, w'' lies on the blue path and thus on the cycle of G. Moreover, all the extra vertices in components without a path of length at least 29, were part of a branch of w''. As the green vertices are the only ones with (at least) K vertices in the adjacent branches we conclude that w'' must have been green. Therefore, we can identify the green vertices in $G - w' = T - v'$.

Because the maximum path in a branch has size at most 27, we do know that T has no 'fake green' vertices that do not lie on the main path i.e. T has no vertices in a branch such that the largest component splits into two components with size at least 29 and K extra vertices. Hence, for T there is a path through all green vertices. Let D_1 and D_2 (with $D_1 \leq D_2$) be the maximum number of vertices before the first green vertex and after the last green vertex for a path through all green vertices.

For every vertex w' in the gap of distance at least 30, we define $e_1^{w'}$ and $e_2^{w'}$ (such that $e_1^{w'} \leq e_w^{w'}$) to be the number of vertices between w' and two green vertices at the boundary of the gap. By definition, $e_1^{w'} + e_2^{w'}$ equals the size of the gap minus one. Since all the green vertices lie on the cycle of G, there is a longest path P in $G - w' = T - v'$ through all the green vertices. Then we define $d_1^{w'}$ and $d_2^{w'}$ (with $d_1^{w'} \leq d_2^{w'}$) to be the maximum number of vertices before the first green vertex or after the last green vertices for a path through all green vertices in $G - w' = T - v'$.

As the blue path is one possible path and the longest path in T is at most 5 longer than the blue path, we see $d_1^{w'1} + d_2^{w'} \leq e_1^{w'} + e_2^{w'} + 5$ and $e_i^{w'} \leq d_i^{w'} \leq e_i^{w'} + 5$ for $i = 1, 2$. We can define these numbers for all 60 possible w'.
If we add the missing vertex v' in $T - v'$ back at most one of $d_1^{w'}$ and $d_2^{w'}$ can change. Since the updated values of $d_1^{w'}$ and $d_2^{w'}$ must be equal to the 'true' values D_1, D_2 in T, we see that $d_i^{w'} = D_j$ for some $i, j \in \{1, 2\}$. Hence,

$$e_i^{w'} \leq D_j \leq e_i^{w'} + 5.$$

Moreover, $e_1^{w'} + e_2^{w'} \leq D_1 + D_2 \leq e_1^{w'} + e_2^{w'} + 5$. Combining these results gives $D_{3-j} - 5 \leq e_{3-i}^{w'} \leq D_{3-j} + 5$. Because $D_1 \leq D_2$ and $e_1 \leq e_2$, we get that $e_1^{w'} \in [D_1 - 5, D_1 + 5]$. Hence, the minimum distance to a green vertex differs by at most 10 over all w'.
However, for every integer n in $[30, 83]$ there are exactly two vertices w' in the gap with minimal distance n to a green vertex. Since there are at most 18 vertices that do not correspond to a card, there is such a vertex w' corresponding to a common card with a minimal distance to a green vertex in the interval $[30, 39]$ and one vertex w'' in the interval $[50, 59]$. These two vertices have a difference in minimal distance of at least $50 - 39 = 11 > 10$, a contradiction. □

4.3 Slightly less cards for recognizing trees

In the previous subsection, we saw that for n large enough the maximum overlapping cards of a tree T and a connected unicyclic graph G is at most $\lfloor \frac{n}{2} \rfloor + 1$ for n large enough. In this subsection, we use the same approach to find slightly better bounds. This stronger upper bound shows that for large enough n the answer to the question 'How many cards do we need to determine whether a graph G is a tree?' is dominated by the case that determines whether G is connected.

Theorem 1.17. *For n large enough, a connected unicyclic graph G and a tree T have at most $\frac{n}{2} - \epsilon\sqrt{n}$ cards in common, where $\epsilon > 0$ is some fixed constant.*

To prove this result, we will look at a tree T and a unicyclic graph G that have $\frac{n}{2} - k$ cards in common, where $k \geq 0$ is allowed to be dependent on n. We will use (roughly) the same steps as in Section 4.2 to derive a contradiction. The only thing that changes is that for every branch consisting of ℓ nodes of degree at least 2, we only color nodes on the cycle red for $\ell - (2k + 2)$ nodes in that branch. This will lead to slightly weaker upper bounds for the longest path and the size of the cycle. Therefore, the final results will only hold for n large enough, where n is allowed to be dependent on k.
We start again with an observation about the length of the cycle of G and the existence of a 'long' path on common cards.

Observation 4.10. *Let L be the length of the unique cycle, then $L \geq \frac{n}{2} - k$. Moreover, every card of G contains a path of size $L - 1$.*

Now, we show that as long as k is not too large, Lemma 3.3 still holds.

Lemma 4.11. *For $k \leq \frac{n}{6} - 3$, there is at least one common card corresponding to a leaf in T.*

Proof. Suppose not. Let K be the number of common cards corresponding to a vertex of degree 2 in T, then all the other $\frac{n}{2} - k - K$ cards correspond to a vertex of degree at least 3 in T. Therefore, G has at least K vertices of degree 3 and at least $\frac{n}{2} - k - K$ vertices of degree ≥ 4. All other vertices in G have at least degree 1 as G is connected. Hence,

$$2e(G) = 2n \geq 3K + 4\left(\frac{n}{2} - k - K\right) + \frac{n}{2} + k = \frac{5n}{2} - 3k - K.$$

Because $k \leq \frac{n}{6} - 3$ this implies that $K \geq 9$, so there is a common card $T - v = G - w$ with $\deg(v) = 2$. This card has at least $K - 1 - 2 = K - 3$ vertices of degree 2. Since $\deg(w) = 3$, adding back w changes the degree of three vertices, implying that G has at least $K - 6$ vertices of degree 2. Hence,

$$2e(G) = 2n \geq 4\left(\frac{n}{2} - k - K\right) + 3K + 2(K - 6) + \frac{n}{2} + k - (K - 6) = \frac{5n}{2} - 3k - 6.$$

This yields a contradiction as $3k + 6 < \frac{n}{2}$. □

We will use this common card to find a lower bound on the number of leaves in G which gives an upper bound for the number of vertices outside the cycle of G of degree at least 2.

Lemma 4.12. *G has at least $\frac{n}{4} - \frac{k}{2} - \frac{3}{2}$ leaves and at most $\frac{3n}{4} - L + \frac{k}{2} + \frac{3}{2}$ of degree at least 2 outside the cycle.*

Proof. If the statement is not true, G has at most $\frac{n}{4} - \frac{k}{2} - \frac{7}{4}$ leaves. Similar to the proof of Lemma 4.4. we obtain that $\mathrm{Leaf}(T) - 3 \leq \mathrm{Leaf}(G)$. Therefore, T has at most $\frac{n}{4} - \frac{k}{2} + \frac{5}{4}$ leaves. Thus, there are at least $\frac{n}{4} - \frac{k}{2} - \frac{5}{4}$ cards corresponding to a vertex of degree at least 2 in T. These cards correspond to a vertex of degree at least 3 in G. So

$$2e(G) = 2n \geq 3\left(\frac{n}{4} - \frac{k}{2} - \frac{5}{4}\right) + \frac{n}{4} - \frac{k}{2} - \frac{7}{4} + 2\left(\frac{n}{2} + k + 3\right) = 2n + \frac{1}{2},$$

a contradiction. Thus, we can conclude that the number of vertices outside the cycle of G of degree at least 2 is at most

$$(n - L) - \left(\frac{n}{4} - \frac{k}{2} - \frac{3}{2}\right) = \frac{3n}{4} - L + \frac{k}{2} + \frac{3}{2}.$$

□

We will now bound the length of the path of G.

Lemma 4.13. *The longest path in T has size at most $L + 8k + 13$*

Proof. We use similar coloring as in Lemma 4.6. However, for every branch of size ℓ, we only color nodes red for $\ell - (2k + 2)$ vertices in that branch. This means that we either do not color any node or that at least $2k + 2$ vertices of degree at least 2 outside the cycle do not correspond to any red vertex. In the last case, there at most

$$2\left(\frac{3n}{4} - L + \frac{k}{2} + \frac{3}{2} - (2k + 2)\right) = \frac{3n}{2} - 2L - 3k - 1$$

red vertices on the cycle. So then there are at least

$$L - \left(\frac{3n}{2} - 2L - 3k - 1\right) = 3\left(L - \frac{n}{2} + k\right) + 1$$

white vertices on the cycle. Hence, in both cases there is a card $T - v = G - w$ corresponding to a white vertex. Every branch adds at most $2k+4$ as at most $2k+3$ vertices in a longest path from the cycle do not correspond to a red vertex (the first $2k + 2$ vertices and the leaf at the end). So the longest path on this card has size at most $L - 1 + 2(2k + 4) = L + 4k + 7$. The longest path in a branch of a white vertex (without containing the white vertex itself) is at most $2(2k + 3) - 1$. Therefore, by adding back v the longest path can increase by at most $2(2k + 3)$. Thus, the maximum path length is at most $L + 4k + 7 + 2(2k + 3) = L + 8k + 13$. □

We would now like to bound the cycle similar to in Lemma 4.7.

Lemma 4.14. *The length of the cycle L satisfies $L \leq \frac{n}{2} + 29k + 53$.*

Proof. We will use a similar coloring as in Lemma 4.7. However, for every branch of size ℓ, we only color nodes red for $\ell - (2k+2)$ vertices in that branch. We will color the following vertices red: the vertex where the branch is connected to the cycle and the vertices on the cycle whose distance till this vertex lies in the interval $[14k+26, \ell+12k+22]$. Then we color at most $2(\ell+12k+22-(14k+26)+1)+1 = 2(\ell-2k-3)+1 < 2(\ell-(2k+2))$ vertices red for this branch, as desired.
Similar to the proof of the previous lemma, there exists a card $T - v = G - w$ such that w is a white vertex. We then know that the longest path on this card has size at most $L + 8k + 13$. The vertices on the card $T - v = G - w$ that lay in the cycle of G are colored blue. Notice that the blue vertices form a path. Let v' be a vertex in the blue path such that v' has distance at least $14k + 26$ from both endpoints. Similar to Lemma 4.7 we can show that the vertex v' must lie in the longest path of T and that the distance from v' to both of the endpoints is at least $14k + 26$. This means that every path that stops a neighbor of v' (and that does not contain v') has length at most $L + 8k + 13 - (14k + 26) = L - 6k - 13$.

All other paths that do not contain v', contain a branch incident to a vertex at distance at least $14k + 26$. By our coloring, every such branch that has l vertices of degree 2, has distance of at least $l + 12k + 23$ to w, because w was white. This means that paths that contain (part of) such a branch gain at most $l + 1$ vertices from the branch, but also miss at least $l + 12k + 22$ vertices from the cycle. This means that such paths have size at most $L + 8k + 13 - (l + 12k + 22) + l + 1 = L - 4k - 8$.
Now we conclude, that the longest path in $T - v - v'$ has size at most $L - 4k - 8$. Since adding back v can only increase the path length with at most $2(2k + 3) - 1 + 1 = 4k + 6$, the longest path in $T - v'$ has size at most $L - 2$. Because every card of G has a path of size $L - 1$, this tells us that the vertex v' can not correspond to a common card. As this is true for all vertices v' on the blue path with distance at least $14k + 26$ from both endpoints, this implies that $L - 1 - 2(14k + 26) = L - 28k - 53$ vertices of T do not correspond to a common card. Therefore, $L - 28k - 53 \leq \frac{n}{2} + k$ implying that $L \leq \frac{n}{2} + 29k + 53$. □

In particular, this means that at most $30k + 53$ vertices on the cycle do not correspond to a common card. We will now prove something similar to Corollary 4.8 and 4.9.

Corollary 4.15. *The maximum height of a branch of G is $23k + 50$.*

Proof. We use proof by contradiction. Suppose there is a vertex v that has a branch with height at least $23k + 51$. Given that the longest path on each G card is at most $L + 8k + 13$, we can conclude that none of the vertices on the G cycle with distance $d \in [1, 15k + 27]$ of V can correspond to a common card. For such vertices, we can namely use this branch and the largest part of the blue path to obtain a path of size at least $L - 1 + 23k + 51 - (d - 1) \geq L + 8k + 14$, which should be impossible. This indicates that at least $2(15k + 27) = 30k + 54$ vertices on the cycle do not correspond to a common card, a contradiction. □

Corollary 4.16. *A path outside the largest component on a common card of $G - w = T - v$ has size at most $46k + 99$.*

Proof. Small components in common cards only consist of vertices that are in one of the branches of G. From the previous corollary, it follows that every branch has a height of at most $23k + 50$. Therefore, every path consisting only of vertices in branches has length at most $2(23k + 50) - 1 = 46k + 99$ as we only consider paths that do not contain the root, at it lies in the cycle of G. □

We will now prove the main theorem of this section.

Theorem 1.17. *For n large enough, a connected unicyclic graph G and a tree T have at most $\frac{n}{2} - \epsilon\sqrt{n}$ cards in common, where $\epsilon > 0$ is some fixed constant.*

Proof. Let $\epsilon = \frac{1}{100}$, so we take $k = \frac{\sqrt{n}}{100}$. Then $n > 6.000k^2 + 24.000k + 24.000$ for n large enough.
Let K be the maximal sum of the vertices in branches connected to some vertex w' in the cycle of G. Since there are at most $30k + 53$ vertices in the cycle not corresponding to a common card, we can use a similar argument as in Theorem 1.16 to show that there are at most $30k + 56$ vertices in the cycle of G such that the sum of the vertices in their branches is at least K. We again color these vertices green. By the pigeonhole principle, there exists an interval of size at least $200k + 400$ between two green vertices. Moreover, this interval contains at least $108k + 200$ vertices w that have distance at least $46k + 100$ from both green vertices. Moreover, for these vertices w, we can spot the green vertices in $G - w$ as these are exactly the points such that by deleting them the largest component splits into two components with a path of size at least $46k + 100$ and exactly K other vertices.
As all the green vertices lie on the cycle of G, there is a longest path P in $G - w' = T - v'$ through all the green vertices. Moreover, the fact that branches have a maximal path length of at most $46k + 99$ implies that all the 'green looking' vertices in T i.e. all vertices such that by deleting them T splits into two components with a path of length 30 and K extra vertices, lie again on one path. We let D_1 and D_2 (with $D_1 \leq D_2$) be the maximum number of vertices before/after the first/last green vertex on this path.
For every vertex w' in the gap, we define $e_1^{w'}$ and $e_2^{w'}$ (such that $e_1^{w'} \leq e_2^{w'}$) to be the number of vertices between w' and the green vertices at the boundary of the gap. By definition, $e_1^{w'} + e_2^{w'}$ equals the size of the gap minus one. As again the green vertices lie on a path, we can again define $d_1^{w'}$ and $d_2^{w'}$ (with $d_1^{w'} \leq d_2^{w'}$ as the maximum number of vertices before/after the first/last green vertex for a path through all green vertices. We do this for all possible w' in the gap.

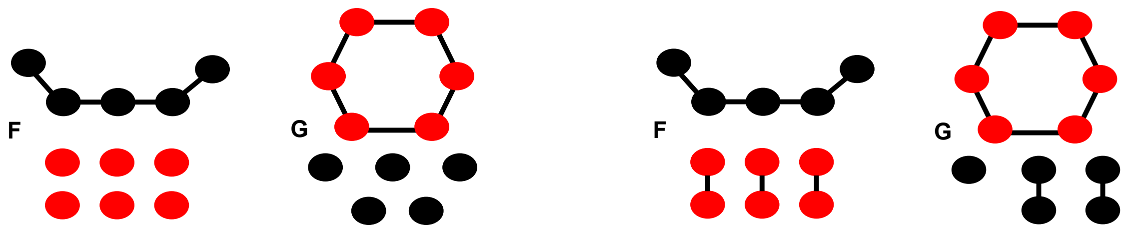
As the blue path is one such possible path and the longest path is at most $8k + 14$ longer, we obtain $e_i^{w'} \leq d_i^{w'} \leq e_i^{w'} + 8k + 14$. If we add the missing vertex v' back in $G - w' = T - v'$, at most one of $d_1^{w'}$ and $d_2^{w'}$ changes. Since the updated values must be equal to D_1 and D_2, implying $d_i^{w'} = D_j$ for some $i, j \in \{1, 2\}$. Hence,

$$e_i^{w'} \leq D_j \leq e_i^{w'} + 8k + 14.$$

Moreover, $e_1^{w'} + e_2^{w'} \leq D_1 + D_2 \leq e_1^{w'} + e_2^{w'} + 8k + 14$. If we combine these results we also get $D_{3-j} - 8k - 14 \leq e_{3-i}^{w'} \leq D_{3-j} + 8k + 14$. Because $D_1 \leq D_2$ and $e_1 \leq e_2$, we get that $e_1^{w'} \in [D_1 - 8k - 14, D_1 + 8k + 14]$. Hence, the minimum distance to a green vertex differs by at most $16k + 28$ over all w'.
On the other hand, for every number in $[46k + 100, 100k + 200]$, there are exactly two vertices w with that minimal distance. Since at most $30k + 53$ vertices on the cycle do not correspond to a common card there is at least one vertex with minimum distance in $[46k + 100, 61k + 127]$ and one vertex with minimum distance in $[85k + 173, 100k + 200]$. But the minimum distance between these vertices is at least $23k + 46 > 16k + 28$, a contradiction. $\square$

5 Adversary recognition for forests

In the previous chapter, we proved for n large enough that any $\lfloor \frac{n}{2} \rfloor + 2$ cards determine whether a graph G is a tree. Next, we want to know how many cards we need to prove that a graph is a forest i.e. whether the graph contains a cycle. After some puzzling, I found some examples of forests F that have $\lfloor \frac{n}{2} \rfloor + 2$ cards in common with some unicyclic graph G. Three of these examples are shown in the figures below.



(a) A forest F consisting of a path and some isolated vertices and a unicyclic graph G consisting of a cycle and some isolated vertices.

(b) A forest F consisting of a path and some isolated edges together with unicyclic graph G consisting of a cycle and some isolated vertices.

Figure 7: Two examples of a forest F and a non-forest G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards (where $n = 11$). All red vertices correspond to isomorphic common cards.

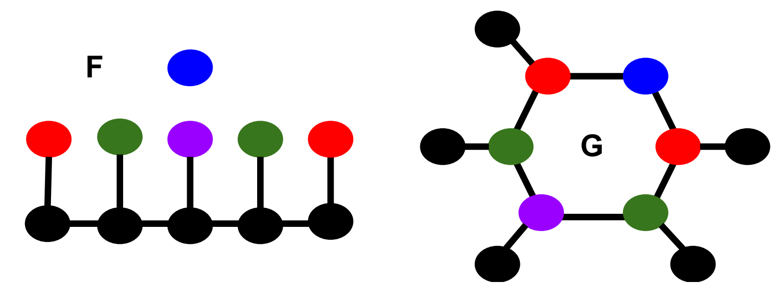


Figure 8: Another example between a forest F and a connected non-forest G with $\lfloor \frac{n}{2} \rfloor + 1$ common cards (where $n = 11$). All vertices of the same color (not black) correspond to the same cards.

Because these three examples all have a nice description, we can extend these three examples to infinite families of pairs of trees F and unicyclic graphs G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards. For example, in the pair of Figure 7a the forest F consists of some path with some isolated vertices, whereas G consists of a cycle with also some isolated vertices. Deleting an arbitrary vertex from a cycle of size $k + 1$ always results in a path of size k. So if the size of the cycle of G is one larger than the length of the path of F (and thus if F has one more isolated vertex than G), deleting an isolated vertex of F and a vertex from the cycle in G always results into the same common card. If $n = 2k + 1$, then the graphs F consisting of a path of size k and $k + 1$ isolated vertices and G consisting of a cycle of size $k + 1$ and k isolated vertices have always $\lfloor \frac{n}{2} \rfloor + 1$ common cards. As n can become arbitrarily large, we have found a lower bound for the property forest of $\lfloor \frac{n}{2} \rfloor + 2$.
In order to also extend the other two examples into infinite families, we need some notation. We use C_k for the cycle on k vertices, P_k for the path on k vertices, K_2 for an isolated edge, and K_1 for isolated vertices. Moreover, for the example in Figure 8 we need the following definition:

Definition 5.1. *Let G be any graph, then the* ***the star of*** *G, denoted $S_1[G]$, is the graph G with an extra leaf connected to each of its vertices. If G is vertex-regular, then $S_1[G]'$ is the graph $S_1[G]$ minus one leaf.*

For the second part of the definition, the graph G must be vertex-regular as otherwise the graph $S_1[G]'$ is not well-defined.
With this notation, we find the following families of pairs F and G:

Observation 5.2. *There are (at least) three infinite families of pairs consisting of a forest F and a unicyclic graph G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards.*

- *Family 1:* $F_{2n+1} = P_n \cup (n+1)\,K_1$ *and* $G_{2n+1} = C_{n+1} \cup n\,K_1$ *corresponding to Figure 7a.*
- *Family 2:* $F_{4n-1} = P_{2n-1} \cup n\,K_2$ *and* $G_{4n-1} = C_{2n} \cup (n\text{-}1)\,K_2 \cup K_1$ *corresponding to Figure 7b.*
- *Family 3:* $F_{2n+1} = S_1[P_n] \cup K_1$ *and* $G_{2n+1} = S_1[C_{n+1}]'$ *corresponding to Figure 8.*

These families were already known, for example, Family 3 is a special case of Theorem 3.4 from 'Families of Pairs of Graphs with a Large Number of Common Cards' [1].
To obtain these common cards, we always must delete a vertex from the cycle of size $\lfloor \frac{n}{2} \rfloor + 1$ of G and a point outside the middle $\lfloor \frac{n}{2} \rfloor$ vertices of the longest path from F. One can see a visualization of this process in Figures 7 and 8.

5.1 Proof upper bound

The second goal of this chapter is to prove that, for n large enough, there are no forests and non-forests that have more than $\lfloor \frac{n}{2} \rfloor + 1$ common cards. In order to prove this, we will use a case distinction based on the following properties of forest F and unicyclic graph G: the number of connected components κ_F respectively κ_G, and the size of the largest connected component M_F respectively M_G. We start by proving a lemma which gives us a relation between κ_F and κ_G.

Lemma 5.3. *If a forest F and a non-forest G have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards, then $\kappa_F \leq \kappa_G + 1$.*

Proof. As the graph G is not a forest, G contains a cycle. By Lemma 1.13 a graph with at least 2 cycles and a forest have at most 2 common cards implying that G has exactly one cycle. The length L of this cycle satisfies $L \geq \lfloor \frac{n}{2} \rfloor + 1$ and every vertex corresponding to a common card must lie in the unique cycle because this cycle is not visible on any of the common cards.
As G contains one cycle, there are at least $L - (n - L) \geq 2(L - \lfloor \frac{n}{2} \rfloor - 1) + 1$ vertices of degree 2 in the unique cycle. Since at most $L - \lfloor \frac{n}{2} \rfloor - 1$ vertices on the cycle do not correspond to a common card, there is a common card $G - w = F - v$ such that $\deg(w) = 2$.
Then $G - w$ is a forest, as subgraphs of forests are forests. Moreover, the connected components of $G - w$ are the same as G except for the one with the cycle (which is still one connected component). In particular, $\kappa_{G-w} = \kappa_G$. Because $G - w = F - v$, we immediately get that $\kappa_{F-v} = \kappa_{G-w}$. At last, $\kappa_F \leq \kappa_{F-v} + 1 = \kappa_G + 1$ as card $F - v$ only contains less components than F if v is an isolated vertex. In that case, $\kappa_{F-v} = \kappa_F - 1$. □

To prove the main theorem, we will use a case distinction based on the difference between κ_F and κ_G. For two of these cases, we will use subcases which are determined by M_F and M_G, the sizes of largest component of F and G. Families 1 and 3 from Observation 5.2 fall both into subcase 1 whereas Family 2 falls into subcase 2a.

Theorem 1.18. *For $n \geq 5000$, a forest F and a non-forest G both on n vertices have at most $\lfloor \frac{n}{2} \rfloor + 1$ common cards.*

Proof. We will use proof by contradiction. Let F be a forest and G be a non-forest that have at least $\lfloor \frac{n}{2} \rfloor + 2$ common cards. Since the cycle of G contains at least $\lfloor \frac{n}{2} \rfloor + 2$ cards, the largest component of G is at least two larger than the next largest component.
By Lemma 5.3 the graphs F and G satisfy $\kappa_F \leq \kappa_G + 1$. Moreover, we do know that $e(F) = n - \kappa_F$ as F is a forest and $e(G) = n + 1 - \kappa_G$ because G contains one cycle. Lastly, the proof of Lemma 5.3 tells us that there exists a common card $G - w = F - v$ such that w has degree 2. For this card, we see $M_F \geq M_{F-v} = M_{G-w} = M_G - 1$ implying $M_F \geq M_G - 1$. For the remainder of the proof, we will distinguish four cases, where the second and third case are each split into two subcases. We will show that for every case there are no forests F and unicyclic graphs G that have $\lfloor \frac{n}{2} \rfloor + 2$ common cards.

Case 1: $\kappa_F = \kappa_G + 1$ implying $e(F) = n - \kappa_F = n - \kappa_G - 1 = e(G) - 2$.
In this case, every vertex in G of degree 2 corresponding to a common card corresponds to an isolated vertex in F. Let C be the number of common cards corresponding to a vertex of degree 2 in G. Then F has at least C isolated vertices. Since deleting a vertex of degree 2 from the cycle of G does not let the number of isolated vertices of G increase, G has at least $C - 1$ isolated vertices.

Furthermore, for every vertex of degree $d \geq 3$ in the cycle of G, there is at least one extra vertex in the largest component, sometimes also called the main component, (outside the cycle) connected to that vertex. Hence, G has at least $\lfloor \frac{n}{2} \rfloor + 2 + (\lfloor \frac{n}{2} \rfloor + 2 - C) + C - 1 \geq n + 2$ vertices, a contradiction.

Case 2: $\kappa_F = \kappa_G$ implying $e(F) = e(G) - 1$.
For a common card $G - v = F - w$ where w has degree 2, the missing vertex v has degree 1. This means $M_F = M_{F-v} + 1$ or $M_F = M_{F-v}$, implying $M_F = M_G$ or $M_F = M_G - 1$.
Case 2a: $\kappa_F = \kappa_G$ and $M_F = M_G - 1$.
In this case, every common card corresponding to a vertex of degree 2, corresponds to a vertex of degree 1 in F outside its main component. Moreover, for every vertex of degree at least 3 on the cycle of G, there is at least one vertex in the main component of G outside the cycle. Let C be the number of common cards corresponding to a vertex of degree 2. Then F has at least C vertices outside its main component. By looking at a common card $G - w = F - v$ where $\deg(w) = 2$, the graph G has at least $C - 1$ vertices outside its main component. Moreover, G has at least $\lfloor \frac{n}{2} \rfloor + 2 + (\lfloor \frac{n}{2} \rfloor + 2 - C)$ vertices in its main component. Therefore, G must have more than n vertices, a contradiction.

Case 2b: $\kappa_F = \kappa_G$ and $M_F = M_G$.
As deleting a vertex of degree 2 in G does not create any extra small components (and neither does deleting any vertex of degree 1 from the largest component of F), all small components of F and G are isomorphic. Moreover, every common card corresponds to a vertex in the largest component of G (as every such vertex lies in the cycle) and thus also to a vertex in the largest component of F. Then as all small components are the same, the largest component of F (which is a tree) and the largest component of G (which is a connected unicyclic graph) have at least $\lfloor \frac{n}{2} \rfloor + 2$ common cards, while their size is at most n. This forms a contradiction with Theorem 1.16.

Case 3: $\kappa_F = \kappa_G - 1$ implying $e(F) = e(G)$.
Case 3a: $\kappa_F = \kappa_G - 1$ and $M_F = M_G - 1$.
In this case, every common card corresponding to a vertex of degree 2 in G corresponds to a vertex of degree 2 outside the large component in F. Because there is at least one common card corresponding to a vertex of degree 2 in G, this means that F has at least one vertex of degree 2 outside its main component, but as F is a forest this also gives us that F has at least 2 extra vertices of degree 1 outside its main component.
On the other hand, for every common card corresponding to a vertex of degree 3 in G, there must be at least one new vertex of degree 1 outside the cycle connected to that vertex. With this bijection, we can show, in a similar way as 2a, that G must have more than n vertices to derive a contradiction.

Case 3b: $\kappa_F = \kappa_G - 1$ and $M_F > M_G - 1$.
As $M_F \neq M_G - 1$, we have that for every common card $G - w = F - v$ such that w lies in the largest component of G, that the vertex v also lies in the largest component of F. Therefore, there is a common card $G - w = F - v$ where w and v have degree at least 2 implying $M_F \geq M_{F-v} + 2 = M_{G-w} + 2 = M_G + 1$. Because $M_F \geq M_G + 1$, there exists a (small) connected graph H such that G contains more connected components of H than F.
As all cards correspond to a vertex in the main component of G and thus also to a vertex in the main component of F, the main component M_F of F should have at least $\lfloor \frac{n}{2} \rfloor + 2$ cards that contain a component isomorphic to H. This forms a contradiction with Lemma A.1.

Case 4: $\kappa_F < \kappa_G - 1$, so $e(F) > e(G)$.
As all cards of G correspond to a vertex of degree ≥ 2, the forest F has at least $\lfloor \frac{n}{2} \rfloor + 2$ vertices of degree ≥ 3. Since every connected component of F is a tree and a tree with k vertices of degree ≥ 3 has at least $k + 2$ leaves, F has at least $\lfloor \frac{n}{2} \rfloor + 4$ vertices of degree leaves. This means that F has more than n vertices, a contradiction. $\square$

Remark 5.4. *The requirement that $n \geq 5000$ is only necessary in case 2b where we use Theorem 1.16. This means that if Theorem 1.16 is proven for all $n \geq k$ (where k is some positive integer), then also Theorem 1.18 is proven for all $n \geq k$.*

5.2 Classification of (infinite) families for which the upper bound is strict

Next, we are going to investigate which forests on non-forests on n vertices have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards. We already have found three of these families in Observation 5.2. We are going to prove that these three are the only ones. Lemma 5.3 immediately tells us that $\kappa_F \leq \kappa_G$, so this means that we can use the same case distinction as in the previous section.

Theorem 1.19. *For $n \geq 24000$, the only forests F and non-forests G on n vertices that have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards, are part of the families generated by examples in Figure 7 and 8.*

Proof. Let F and G be a forest and a unicyclic graph that have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards. Lemma 5.3 tells us that $\kappa_F \leq \kappa_G$ and that there exists a common card $G - w = F - v$ such that w had degree 2. We will use the same six cases as in Theorem 1.18. We will do this in reverse order so that we end with the most difficult cases, which are the ones that lead to infinite families of examples.

Case 4: In this case, we can use a similar argument to show that F must have at least $\lfloor \frac{n}{2} \rfloor + 1$ vertices of degree 3 or higher and also at least $\lfloor \frac{n}{2} \rfloor + 3$ of degree 1. This is impossible as F has n vertices.
Case 3b: We can use the same arguments to show that the maximum component M_F of F should have at least $\lfloor \frac{n}{2} \rfloor + 2$ cards that contain a component isomorphic to H (for some connected graph H). However, this still forms a contradiction with Lemma A.1.
Case 3a: For this case, we need to use the fact that F has at least 2 extra vertices of degree 1 outside of its main component, so F has at least $C + 2$ vertices in small components. Then we can show that G has at least $C + 1$ vertices outside its main component and at least $\lfloor \frac{n}{2} \rfloor + 1 + (\lfloor \frac{n}{2} \rfloor + 1 - C)$ vertices in its main component. So, G must have more than n vertices, a contradiction.
Case 2b: We can again prove that the main components M_F and M_G (which are trees) must have at least $\lfloor \frac{n}{2} \rfloor + 1$ common cards. However, if we use $k = 0$ in the proof of Theorem 1.17, we get that this is impossible for $n \geq 24000$.

We now continue with the two cases that lead to our found infinite families.
Case 2a: The bijection between vertices of degree 2 corresponding to a common card in G and vertices outside the main component in F still works. Let C be again the number of common cards corresponding to a vertex of degree 2 in G, then we can show in a similar way that G has at least $C - 1$ vertices outside its main component and $\lfloor \frac{n}{2} \rfloor + 1 + (\lfloor \frac{n}{2} \rfloor + 1 - C)$ in its main component. If n is even this implies that G has at least $n + 1$ vertices, so we have again a contradiction.
However, if n is odd, then $C - 1 + \lfloor \frac{n}{2} \rfloor + 1 + (\lfloor \frac{n}{2} \rfloor + 1 - C) = n$. Hence, for our graphs G and F, there must be equality in all inequalities. Thus, $L = \lfloor \frac{n}{2} \rfloor + 1$ implying that all vertices on the cycle of G correspond to a common card. Moreover, all vertices on that cycle have degree 2 or 3 and the unique neighbor outside the cycle of a vertices of degree 3 is a leaf. Lastly, this implies that F has exactly C vertices outside its main component which all have degree 1. Hence, F has $\frac{C}{2}$ small components isomorphic to K_2.
Suppose there exists a vertex w' of degree 3 such that $G - w' = F - v'$. Then, $M_{F-v'} = M_{G-w'} = M_G - 2$. Hence, $F - v'$ has a component of size $M_G - 2 = M_F = 1$. As $\deg(v) = 3 - 1 = 2$, deleting v from M_F will cut off at least two vertices of M_F. Therefore, one of the small components should have size $M_F - 1$. But since all small components of F have size 2, we get $M_F - 1 = 2 = M_G - 2$. Thus $M_G = 4$ implying $L \leq 4$ and thus $n \leq 7$.
So for larger graphs, all vertices in the cycle of G must have degree 2, so $C = L$. Hence, F equals $P_{\lfloor \frac{n}{2} \rfloor} \cup \frac{\lfloor \frac{n}{2} \rfloor + 1}{2} K_2$ and $G = C_{\lfloor \frac{n}{2} \rfloor + 1} \cup \frac{\lfloor \frac{n}{2} \rfloor - 1}{2} K_2 \cup K_1$. So it is part of Family 2.

Case 1: In this case, $\kappa_F = \kappa_G + 1$ and thus $e(F) = e(G) - 2$. Hence, every vertex of degree 2 corresponding to a common card in G corresponds to an isolated vertex in F. Let C be the number of common cards corresponding to a vertex of degree 2. Then F has at least C isolated vertices. Because deleting a vertex of degree 2 from the cycle of G does not let the number of isolated vertices of G increase, G has at least $C - 1$ isolated vertices.
Moreover, for every vertex of degree $d \geq 3$ in the cycle of G, there are at least $d - 2 \geq 1$ extra vertices in the main component of G. If we combine these facts, we obtain that G has at least $\lfloor \frac{n}{2} \rfloor + 1 + (\lfloor \frac{n}{2} \rfloor + 1 - C) + C - 1$ vertices. If n is even, this implies that G has at least $n + 1$ vertices, a contradiction. Therefore, n is odd,

in which case this sum is n. So all inequalities are equalities. In particular, $L = \lfloor \frac{n}{2} \rfloor + 1$ and the cycle only contains vertices of degrees 2 and 3. Moreover, the neighbors of the vertices of degree 3 outside the cycle are all leaves.
Since all isolated vertices v and v' in F satisfy $F - v = F - v'$, all vertices w and w' of degree 2 in the cycle of G satisfy $G - w = G - w'$. So in particular the degrees of the neighbors of such vertices must be the same. Suppose that they as a set are equal to $\{d_1, d_2\}$.
If $d_1 = d_2 = 2$, all the vertices on the cycle must have degree 2. In this case, we get an element of Family 1. If $d_1 = d_2 = 3$, then every vertex of degree 2 has two neighbors of 3. Because $M_F = M_G - w'$ (where $\deg(w') = 2$) and all branches outside the cycle are leaves, the longest path in M_F is equal to $\lfloor \frac{n}{2} \rfloor + 2$ as the leafs adjacent to the neighbors of degree 3 each add one to the length of the path. Every graph contains at most 2 leaves such that by deleting that leaf the length of the longest path becomes shorter. Hence, all but at most 2 vertices of degree 3 of G have two neighbors of degree 3. So there is at most one vertex of degree 2. As we also knew there is at least one vertex of degree 2, this means that exactly one vertex on the cycle has degree 2. In this case, we find that G must be equal to $S_1[C_{\lfloor \frac{n}{2} \rfloor + 1}]'$. By looking at the card of G corresponding to a vertex of degree 2, the graph F is equal to $S_1[P_n] \cup K_1$. Hence, F and G must be the graphs of Family 3.
If $d_1 \neq d_2$, then every vertex of degree 2 has a neighbor of degree 2 and a neighbor of degree 3. Hence, for every node w of degree 2, one of the ends of the longest path in $M_F = M_G - w$ corresponded to the neighbor of degree 2, hence it is a leaf in M_F. This vertex also had a neighbor of degree 3, so the vertex next to it has degree 3. This vertex of degree 3 is adjacent to two leaves. Thus, all but at most 2 common cards $G - w' = T - v'$ must contain a vertex of degree 3 that is adjacent to two leaves.
However, this can only happen for a vertex w' in G if w' has a neighbor of degree 2, whose other neighbor is of degree 3. But if $\deg(w) = 3$ this implies that its neighbor of degree 2 has two neighbors of degree 3. A contradiction. Therefore, there are at most two common cards corresponding to a vertex of degree 3, so the cycle of G has at most two vertices of degree 3. On the other hand, because every vertex of degree 2 has a neighbor of degree 3 at least one-third of the vertices on the cycle has degree 3. Then $\frac{1}{3}\lfloor \frac{n}{2} \rfloor + 1 \leq 2$ implying that $n \leq 11$. □

Remark 5.5. *By looking at pairs of graphs with at most* 11 *vertices that fall into case 1 where* $d_1 \neq d_2$ *and pairs of graphs with at most* 2 *vertices that fall into case 2a, one can find three more small graphs for which the upper bound is attained.*

6 Adversary recognition for the girth

The next graph property that we will examine is the girth, which is the length of the smallest cycle in a graph. For example, the girth of C_n is n and the girth of K_n is 3 for $n \geq 3$. If the graph does not contain any cycle, the girth is defined to be infinity.
In this section, we will show that we can determine the girth of a graph G on n vertices based on any of its $\frac{2n}{3}+1$ cards. Notice that we need at least $\lfloor \frac{n}{2} \rfloor + 2$ cards as we found in Chapter 5 three infinite families of pairs of graphs such that one is unicyclic and the other a forest that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards. As the girth for forests is infinite and for unicyclic graphs finite, these graphs do not have the same girth. We start with an elementary lemma about graphs where the girth is at least $\frac{2n}{3}$.

Lemma 6.1. *Every graph G on n vertices with girth at least $\frac{2n}{3}+1$ has at most one cycle.*

Proof. We will use proof by contradiction. Let G be such a graph and let C and C' be two cycles in G. By assumption, $|C| \geq \frac{2n}{3}+1$ and $|C'| \geq \frac{2n}{3}+1$. So they have at least $\frac{2n}{3}+1-(n-\frac{2n}{3}+1) = \frac{n}{3}+2$ common vertices. In particular, the induced subgraph $C \cup C'$ of G is connected and has at most n vertices.
Let T be a spanning tree of $C \cup C'$. Since $C \cup C'$ contains at least 2 cycles, we have to add at least 2 edges to T. We call these extra edges u_1v_1 and u_2v_2. Then $d_T(u_i, v_i) \geq \frac{2n}{3} \geq \frac{2v(C\cup C')}{3}$ for $i = 1, 2$ because otherwise $C \cup C'$ (and thus also G) contains a cycle of length at most $\frac{2n}{3}$.
For $i = 1, 2$ we define P_i to be the path between u_i and v_i. Then $v(P_i) \geq \frac{2n}{3}+1$ as both cycles have length at least $\frac{2n}{3}+1$. Moreover, as P_1 and P_2 are both paths in a tree $P_1 \cap P_2$ is also a path.
Let x and y be the endpoints of the path $P_1 \cap P_2$. Since x and y lie on the cycle $P_1 \cup u_1v_1$, there exists a path from x to y in $((P_1 \cup u_1v_1)\backslash P_2) \cup \{x, y\}$. Similarly, there exists a path from x to y in $((P_2 \cup u_2v_2)\backslash P_1) \cup \{x, y\}$. We can combine these internal vertex-disjoint paths with the same endpoints into a cycle with

$$\begin{aligned} v(P_1\backslash P_2) + v(P_2\backslash P_1) + |\{x,y\}| &= v(P_1) + v(P_2) - 2v(P_1 \cap P_2) + 2 \\ &\leq v(P_1) + v(P_2) - 2[v(P_1) - (n - v(P_2))] + 2 \\ &= 2n + 2 - v(P_1) - v(P_2) \\ &\leq 2n + 2 - 2\left(\frac{2n}{3} + 1\right) \\ &= \frac{2n}{3} \end{aligned}$$

vertices. A contradiction. □

In chapter 5, we saw that for $n \geq 24000$ a forest and a non-forest on n vertices have at most $\lfloor \frac{n}{2} \rfloor + 1$ common cards. However one can prove, even for small n, using (mostly) counting arguments that the following result holds, of a proof is written down in Appendix A.

Lemma 6.2. *Forests and non-forests have at most $\frac{2n}{3}$ common cards.*

We will now combine the previous two lemmas with a version of our standard coloring to find the following theorem.

Theorem 1.20. *For any graph on n vertices, we can determine its girth based on any $\frac{2n}{3}+1$ of its cards.*

Proof. First, Lemma 6.2 gives that we can determine whether a graph has a cycle based on any $\frac{2n}{3}+1$ of its cards. So if G does not have any cycle, we know that its girth is ∞.
Let G be an arbitrary graph with a cycle and let L be the girth of G.Any cycle of length at most $\frac{2n}{3}$ is visible on at least one of the $\frac{2n}{3}+1$ cards. This implies that if the girth L of G is $\leq \frac{2n}{3}$, we see a cycle of size L and since we do not see any smaller cycles, we know that L is the girth of G. Hence, we will only look at graphs with a cycle and a girth of $L \geq \frac{2n}{3}+1$.
By Lemma 6.1 the graph G has exactly one cycle. If this unique cycle is visible on some card, we can use this card to can determine the girth L. Hence, we assume that every card in the subdeck corresponds to a vertex in the cycle of G. Now, we notice that every card $G - v$ of G contains a path of size $L - 1$ because G contains a cycle of length L.

To determine the length of the unique cycle, we will use a coloring somewhat similar to the one in Figure 5 to distinguish good and bad vertices on the cycle. If there are k vertices in a branch connected to a vertex v on the cycle, then we color all vertices on the cycle with a distance between 1 and k of v red. In particular, we do not color the vertex v itself red. So every vertex outside the cycle corresponds to at most two red vertices. Thus, there are at least

$$L - 2(N - L) = 3L - 2N = 3\left(L - \frac{2n}{3} - 1\right) + 3$$

white vertices on the cycle. Since at most $L - \frac{2n}{3} - 1$ vertices on the cycle do not correspond to a card in the subdeck, there is at least one card in the subdeck that corresponds to a white vertex w.
Every branch with a path of size k from the cycle contains at least k vertices. Therefore, such branches have distance at least $k + 1$ from w. So these branches can not increase the length of the path corresponding to the cycle, implying that the longest path on $G - w$ has a length of exactly $L - 1$.
In conclusion, we can reconstruct $L - 1$ because it is the smallest among the longest paths over all cards. So we can determine $L - 1$ and thus also the girth L self. $\square$

7 Adversary recognition for bipartiteness

The last property that we will examine is bipartiteness. A graph is bipartite if it does not contain any cycle of odd length. The main goal of this chapter is to show that we can determine whether a graph G has such a cycle even if there is a linear amount of cards missing.

Theorem 1.21. *For any graph G on n vertices, it can be determined whether G is bipartite based on any $\frac{5n}{6}+2$ of its cards.*

We will use a similar approach as in the proof of Theorem 1.16. In this case, we look at a bipartite graph G and a non-bipartite graph H that have at least $\frac{5n}{6}+2$ common cards. We fix D to be a subdeck of $\frac{5n}{6}+2$ of these common cards.
We will call cycles of length at most $\frac{5n}{6}+1$, *small cycles* because these cycles are all at least once visible on a card of D. All cycles with at least $\frac{5n}{6}+2$ are called *large cycles.* This gives us the following observations.

Observation 7.1. *Let $k \leq \frac{5n}{6}+1$. Then G contains a cycle of size k if and only if H does.*

As G does not contain an odd cycle, neither do cards in D. Thus all small cycles of G and H have even length.

Observation 7.2. *All odd cycles in H are large cycles and contain at least $\frac{5n}{6}+2$ vertices.*

For the rest of this section, we fix a smallest odd cycle L of H. The previous observation gives $L \geq \frac{5n}{6}+2$, where abused notation by using L for the number of vertices in L. We will do this more often to make the proof easier to read. As the cycle L can not be visible on any common card, all vertices $w \in H$ corresponding to a common card lie in L. We will use this fact to prove that a lot of common cards in D correspond to vertices of degree 2. With this information, we will derive an upper bound for the largest even cycle of H.

Lemma 7.3. *At least $\frac{2n}{3}+4$ cards of D correspond to a vertex of degree 2 in H.*

Proof. First, note that there are exactly $n-L$ vertices outside the cycle L of H. Now we will prove that every vertex outside the cycle L can be connected to at most 2 vertices on the cycle L. First, notice that if v is connected to w_1 and w_2, the distance between w_1 and w_2 should be even because otherwise we get an odd cycle of size at most $L/2+1 \leq \frac{n}{2}+1$. Hence, there is at least one vertex on the cycle of L between w_1 and w_2. Moreover, we can replace the (short even) arc w_1w_2 of L with the path w_1vw_2 to get a new odd cycle. Since this cycle is as least as big as L, there is at most one vertex between w_1 and w_2. This concludes the proof of the claim.
Next, we see that if a vertex v outside the cycle L is connected to two vertices w_1 and w_2 on the cycle, that the vertex on the cycle between w_1 and w_2, called w, can not correspond to a common card, as the card $H-w$ contains a cycle of length L by consisting of the long arc w_2w_1 and the 'branch' w_1vw_2. This means that this vertex outside the cycle corresponds to a unique vertex on the cycle that can not correspond to a common card, implying that the length of L should also be one higher. Since this vertex lies directly between the two vertices whose degree increased, we see that every vertex v, which increases the degree of two vertices on the cycle of G from 2 to 3, adds a new vertex to the cycle that can not correspond to a common card.
Hence, we see that (on average) every vertex not corresponding to a card increases the degree of at most one vertex corresponding to a card from 2 to something higher. Thus, there are at least $\frac{5n}{6}+2-(n-\frac{5n}{6}-2) \geq \frac{2n}{3}+4$ cards in D corresponding to a vertex of degree 2. □

With this lemma, we can show that the number of edges of G and H are equal.

Lemma 7.4. $e(G) = e(H)$.

Proof. Suppose $e(G) \neq e(H)$ we see that there is a degree d (not 2) such that G contains $\frac{2n}{3}+4$ vertices of degree d. But if $d > 3$, the graph G also has at least $\frac{2n}{3}+3$ vertices of degree at most 3, a contradiction. So $d = 1$ or $d = 3$. If $d = 1$, then G also has at least $\frac{2n}{3}+4-1$ vertices of degree at least 2, a contradiction. And similarly, if $d = 3$ G also has at least $\frac{2n}{3}+4-1-3 = \frac{2n}{3}$ vertices of degree at most 2, a contradiction. Hence, $e(G) = e(H)$. □

We call a vertex in L **simple** if it has degree 2 and if it does not lie in any even cycle C.

Lemma 7.5. *Every vertex of degree 2 of H corresponding to a common card is simple.*

Proof. Suppose w is a vertex of degree 2 in H corresponding to a common card, but w is not simple. Then w lies in an even cycle C and in L. Let w_1 and w_2 be the neighbors of w. Since w has degree 2, the vertices w_1 and w_2 also lie in $C \cap L$. Replacing the path $w_1 w w_2$ in L with the other arc of C gives us an odd closed walk. This odd closed walk lies also in $H - w$ implying that $H - w$ also contains an odd cycle, a contradiction. □

As H has at least $\frac{2n}{3} + 4$ simple vertices that correspond to a common card, we can conclude that:

Corollary 7.6. *The largest even cycle of H has size at most $\frac{n}{3} - 4$.*

Proof. Combining Lemmas 7.3 and 7.5 results that H has at least $\frac{2n}{3} + 4$ vertices that do not lie in any even cycle. hence, every even cycle has size at most $n - \frac{2n}{3} + 4 = \frac{n}{3} - 4$. □

If we combine this lemma with Observation 7.1, we get the following corollary:

Corollary 7.7. *The largest small cycle in G has size at most $\frac{n}{3} - 4$.*

With this information, we can show that G must have lots of cut vertices.

Corollary 7.8. *Every common card $G - v = H - w$ contains $\frac{2n}{3} + 1$ cut vertices of degree 2.*

Proof. H has at least $\frac{2n}{3} + 4$ simple vertices, which by definition are of degree 2 and do not lie in any even cycle. Since common cards do not contain odd cycles, none of these vertices lie in a cycle on a common card. By deleting a vertex at most one vertex of degree 2 disappears and the degree of at most two vertices of degree 2 in the cycle L decreases. Because every vertex not in a cycle is a cut vertex, common cards have at least $\frac{2n}{3} + 4 - 1 - 2$ cut vertices of degree 2. □

We now show that G also has a large cycle.

Lemma 7.9. *G contains a cycle of size at least $\frac{5n}{6} + 2$.*

Proof. Suppose not. Then Lemmas 7.1 and 7.6 imply that the maximum length of any cycle in G is at most $\frac{n}{3} - 4$. Let $G - w = H - v$ be a common card corresponding to vertices of degree 2. By Corollary 7.8 this card has at least $\frac{2n}{3} + 1$ cut vertices of degree 2. However, by adding back w (which has degree 2) we can put at most $\frac{n}{3} - 5$ cut vertices in a cycle as every cycle of G contains at most $\frac{n}{3} - 4$ vertices. This means that G has at least $\frac{n}{3} + 5$ cut vertices of degree 2. In particular, at least one of these cut vertices v corresponds to a common card $G - v$ which is disconnected. However, the card $H - w$ is connected for every vertex w of degree 2 as w lies in L. So $G - v \neq H - w$ for all $w \in H$ and thus v does not correspond to a common card, a contradiction. □

Let K be the smallest large cycle of G, as G is bipartite the cycle K has even length. Since the even cycles in H contain at most $\frac{n}{3} - 4$ vertices, this cycle is not visible on any common card, implying that all vertices $v \in G$ corresponding to a common card lie in K. Let C be an arbitrary small cycle in G and/or H.

Lemma 7.10. *For every two vertices $u, v \in K \cap C$ (respectively $L \cap C$) the distance in K (or L) is smaller or equal to the distance in C.*

Proof. We only prove the statement for K as the proof for L is the same. If we replace the short arc of K with the short arc in C, the cycle should either become larger (implying that the short arc in C is larger than the short arc in K) or smaller. However, if it becomes smaller we must obtain a small cycle by definition of K. This means that the small arc in C is at least $\frac{5n}{6} + 2 - (\frac{n}{3} - 4) \geq \frac{n}{2} + 6$ smaller than the small arc in K. However, since the size of K is at most n, the small arc has size at most $\frac{n}{2}$, a contradiction. □

Corollary 7.11. *For every small cycle C, the vertices of $K \cap C$ (respectively $L \cap C$) lie in one interval of size at most $\frac{C}{2} + 1$ in K (respectively L).*

Proof. Corollary 7.10 gives that every two vertices in $C \cap K$ lie within distance $\frac{C}{2}$ in K. Fix a vertex $v \in K \cap C$, then every vertex lies within distance $\frac{C}{2}$ from a vertex v. Since $2C \leq 2\frac{n}{3} - 4 < L$, the shortest arc between vertices left and right from v is always via v. Hence, all the vertices in $K \cap C$ lie in an interval of size at most $\frac{C}{2} + 1$. □

Similar to the definition in H we define a simple vertex of G to be a vertex in the cycle K such that this vertex does not lie in any small cycle of G. We now prove that every vertex of degree 2 corresponding to a common card of G and H is simple.

Lemma 7.12. *Every vertex of degree 2 in G corresponding to a common card is simple.*

Proof. Suppose v is a vertex of degree 2 in G corresponding to a common card, but v is not simple. Then v lies in a small cycle C (and a large cycle). Since v has degree 2, both neighbors of v also lie in C and K. Let v_1' and v_2' be the vertices in $K \cap C$ that have the largest distance between them in K.
Because $C \leq \frac{n}{3} - 4$ and every distance in L is smaller than the distance in C, we obtain $d_K(v_1', v_2') \leq \frac{n}{6} - 2$. Moreover, there are no other vertices on the long arc $v_1'v_2'$ of K that also lie in C as those vertices will have a larger distance to either v_1' or v_2' than the distance between v_1' and v_2' in K. So v lies on the short arc of K. If we now replace the short arc $v_1'v_2'$ in K with the arc $v_1'v_2'$ of C without v, we get a cycle that does not contain v. This new arc of C is larger than the short arc in K implying that this cycle is larger than K and thus not small. Since H does not contain any large even cycles and $G - v$ contains a large even cycle, $G - v$ is not a card of H. □

We will now combine all the information from the previous lemmas to derive a contradiction.

Theorem 1.21. *For any graph G on n vertices, it can be determined whether G is bipartite based on any $\frac{5n}{6} + 2$ of its cards.*

Proof. We use proof by contradiction. Suppose there exists a bipartite graph G and a non-bipartite graph H that have a subdeck D of at least $\frac{5n}{6} + 2$ common cards.
By the proof of Lemma 7.4, at least $\frac{2n}{3} + 4$ common cards in D correspond to a simple vertex in H (and G) of degree 2. Hence, we can determine the number of edges in G and H based on D and thus we can determine the degrees of the missing vertex for all cards in D.
We will now prove that $K - 1$ and $L - 1$ are both equal to the smallest longest path of all common cards corresponding to a vertex of degree 2. We will prove this statement for K as the proof for L will be the same. We start by observing the following. The neighbors of a simple vertex v must have distance $K - 1$ in $G - v$ as otherwise, v would lie in a smaller cycle than K. Hence, the diameter on all cards corresponding to a vertex of degree 2 is at least $K - 1$.
We will use a similar coloring as in Figure 5. This means that for every branch of size k outside our fixed cycle K connected to a vertex v in K, we color all vertices on K with distance at most k to v red. Since the vertices of a path of size m outside the cycle are only connected to vertices that lie in an interval of size $m + 1$ on the cycle by Lemma 7.11, we color at most $3k + 2 \leq 5k$ vertices red for every branch of size k. This indicates that we color at most $5(n - K)$ vertices on K red. Hence, there are at least

$$K - 5(n - K) = 6K - 5n = 6\left(K - \frac{5n}{6} - 2\right) + 12$$

white vertices on the cycle. Since at most $K - \frac{5}{6}n - 2$ vertices on the cycle of G do not correspond to a common card, there exists a white vertex v corresponding to a card in D, which is of degree 2, and thus simple, by our coloring.
Next, we show that for every white vertex v the diameter of $G - v$ is exactly $K - 1$. Suppose not. Let x and y be two vertices in $G - v$. Let v_x and v_y be the vertices closest to x and y such that v_x and v_y lie on the cycle L and such that v_x lies closer to v_1 and v_y to v_2. Then as v is white, we see $d(x, v_x) < d(v, v_x)$ and $d(y, v_y) < d(v, v_y)$ implying that

$$d(x, y) \leq d(x, v_x) + d(v_x, v_y) + d(v_y, y) \leq d(v_1, v_x) + d(v_x, v_y) + d(v_y, v_2) = d(v_1, v_2).$$

Where the last equality follows from the fact that the unique path between v_1 and v_2 is the path corresponding to the large arc of the cycle.
Hence, $K - 1$ is equal to the minimal diameter of (common)cards $G - v$, where v has degree 2. Similarly, we can prove that $L - 1$ is equal to the minimal diameter of (common)cards $H - w$, where w has degree 2. For a common card $G - v = H - w$, the vertex v has degree 2 if and only if w has degree 2. Implying that we take the minimum over the same set of cards. Thus, $K - 1 = L - 1$ implying $K = L$, a contradiction. □

So we have proved an upper bound of $\frac{5n}{6}+2$ for determining whether a graph on n vertices is bipartite. Now, I would like to shortly look at possible lower bounds. Notice that if we take an element of Family 1 with $n \equiv 1 \mod 4$ vertices, then G contains an odd cycle of length $\frac{n+1}{2}$, while F does not contain such a cycle. This means that we found an infinite family of pairs of bipartite graphs F and non-bipartite graphs G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards.
Moreover, for a member of Family 3 on $n \equiv 1 \mod 4$ vertices, the graph G has an odd cycle of size $\frac{n+1}{2}$, while F does not contain any cycles. This gives our second infinite family of pairs of bipartite graphs F and non-bipartite graphs G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards.
Below, I give another infinite family of pairs of graphs G and H, G is bipartite and H is not and such that G and H have $\frac{n+1}{2}$ common cards. What is special about this example is the fact that G and H both contain cycles and that both are connected. We can extend this example to a whole family by adding extra diamonds to G and H. Since every diamond increases the large cycle of H by 3, H is not bipartite if it contains an even number of diamonds H. However, G is always bipartite implying that we have found our desired infinite family of pairs.

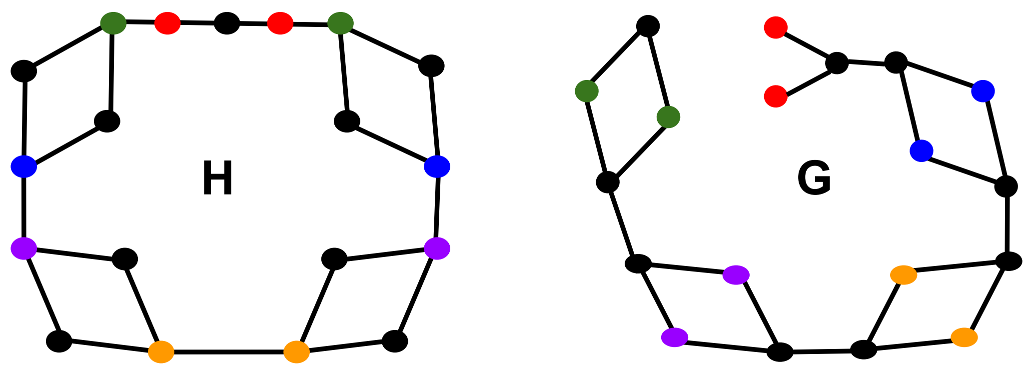
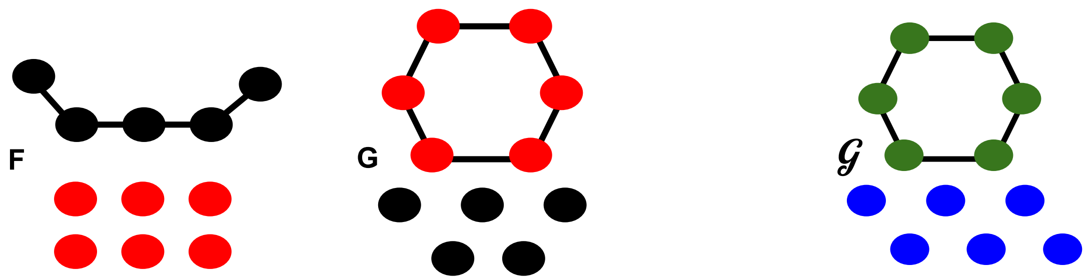


Figure 9: Another example of a bipartite graph G and an non-bipartite graph H that have $\frac{n+1}{2}$ common cards. For every two vertices $v \in G$ and $w \in H$ with the same color (not black), we have $G - v \cong H - w$.

8 Why do we need $\lfloor \frac{n}{2} \rfloor + 2$ cards for many properties ?

In this last chapter, I would like to give some insight into a question that I asked myself while writing this thesis. Namely, 'Why do we often need $\lfloor \frac{n}{2} \rfloor + 2$ cards to determine a certain property?' In this chapter, I will not write down any formal proof about why this is the case, but I hope to give you some intuition for answering this question.
The answer lies in the construction of the counter-examples for the lower bound. Most of these examples are of some special form. I will illustrate this with an example, namely Family 1 from the section about Forests.



(a) An example of a forest F and a non-forest G that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards (where $n = 11$). All red vertices correspond to isomorphic common cards.

(b) The supercard $\mathcal{G}$ which consists of $n+1$ vertices. Deleting a green vertex results in the graph F, whereas deleting a blue vertex results in the graph G.

Figure 10: Two graphs F and G with a supercard $\mathcal{G}$. Deleting a green vertex from $\mathcal{G}$ results in the graph F and deleting a blue vertex results in G.

For the graphs F and G of this example there exists some graph $\mathcal{G}$ on $n+1$ vertices such that both F and G are cards of $\mathcal{G}$. Such a graph $\mathcal{G}$ is also called a 'supercard' of F and G [3]. If we delete a green vertex from the cycle of $\mathcal{G}$ we get our graph F, whereas deleting a blue isolated vertex results in G. For this supercard $\mathcal{G}$, there are exactly $\frac{n+1}{2} = \lfloor \frac{n}{2} \rfloor + 1$ vertices corresponding to a card isomorphic to F and $\frac{n+1}{2} = \lfloor \frac{n}{2} \rfloor + 1$ vertices corresponding to a card isomorphic to G.
The action of the automorphism group of $\mathcal{G}$ divides the vertices of $\mathcal{G}$ into two orbits of size $\frac{n+1}{2}$. One orbit of all isolated vertices and one orbit consisting of the cycle. Moreover, deleting a vertex of the first orbit results in a card isomorphic to G and deleting a vertex of the second orbit results in a card isomorphic to F. Because these two orbits do not interact, all cards of F that are obtained from deleting an isolated vertex (that lies in the first orbit of $\mathcal{G}$) are isomorphic to cards where a vertex from the cycle of G (which laid in the second orbit of $\mathcal{G}$) is deleted as these cards are exactly the graphs obtained from deleting one vertex from both orbits of $\mathcal{G}$. Since there are exactly $\frac{n+1}{2} = \lfloor \frac{n}{2} \rfloor + 1$ vertices in the other orbit, we see that F and G have $\lfloor \frac{n}{2} \rfloor + 1$ common cards. Notice that this construction works for every pair of the infinite Family 1.
For the graphs G and H of Figure 9, we can also find a supercard $\mathcal{G}$, see Figure 11, such that its vertices are split into two sets of $\frac{n+1}{2}$ by the automorphism group. In this case, we have the orbit of vertices of degree 3 on the cycle and the vertices on the diamonds of degree 2. Also in this case, we see that common cards of $F = \mathcal{G} - v$ and $G = \mathcal{G} - w$ arise by deleting a vertex v' from F that lies in the other orbit than v (and deleting w' from G such that w' and w lie in different orbits in $\mathcal{G}$).
However, for this supercard $\mathcal{G}$ the vertices of different orbits are connected. Therefore, not every pair of v' and w' satisfies $F - v' = G - w'$. However, for every v' in the same orbit as w (so a different orbit as v), we can find an automorphism ϕ of G such that $\phi(w) = v'$. Because ϕ is an automorphism, we obtain that $G - \phi^{-1}(v) \cong F - v'$. By selecting automorphisms for every possible v' such that all $\phi_{v'}^{-1}(v)$ are distinct, we can find again $\frac{n+1}{2}$ common cards.
In conclusion, if we start with a supercard $\mathcal{G}$ such that the vertices of $\mathcal{G}$ are split into at least two orbits of roughly $\frac{n+1}{2}$ vertices, we can construct non-isomorphic graphs G and H that have (almost) $\frac{n+1}{2} = \lfloor \frac{n}{2} \rfloor + 1$ common cards by deleting vertices from different orbits. So there is a nice way to construct many pairs of graphs with $\lfloor \frac{n}{2} \rfloor + 1$ common cards. Since such a construction is not possible for $\lfloor \frac{n}{2} \rfloor + 2$ cards, we will often get a lower bound of $\lfloor \frac{n}{2} \rfloor + 2$.
In the examples that I constructed, the graph $\mathcal{G}$ was chosen such that it had exactly one (odd) cycle. Then to obtain the first graph, I deleted a vertex such that this cycle disappeared, while for the other graph, I

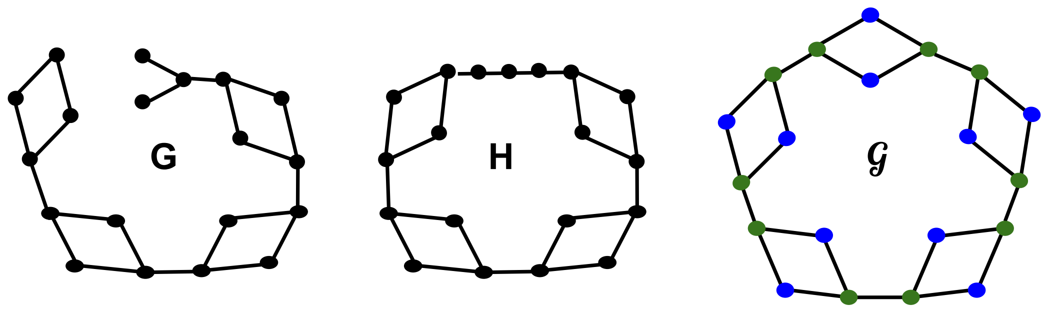
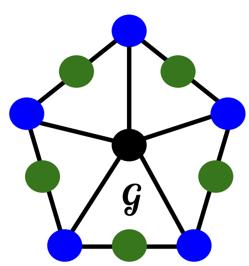
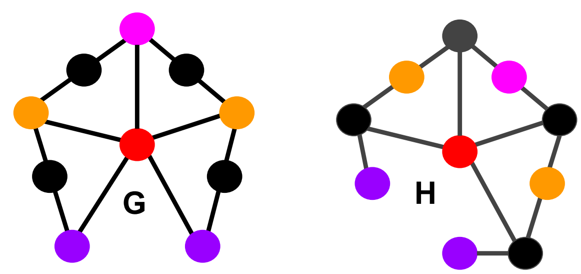


Figure 11: The graphs G and H from Figure 9 together with a super card. Deleting a green vertex from $\mathcal{G}$ results in the graph G and deleting a blue vertex results in H.

picked a vertex such that this cycle was still visible. Therefore, I constructed pairs of graphs such that one has an (odd) cycle while the other does not and such that they have $\lfloor \frac{n}{2} \rfloor + 1$ common cards. But by starting with a graph $\mathcal{G}$ that looks different, we can construct graphs G and H such that G satisfies $\mathcal{P}$ and H does not.
In the figure below, I use a graph $\mathcal{G}$ with three orbits, one of size one, and two of size $\frac{n}{2}$ such that these last two orbits are only different because of their (non-)adjacency to the vertex in the orbit of size 1, to construct (an infinite family of) graphs G and H such that G contains a Hamiltonian cycle and H does not contain a Hamiltonian path, so, in particular, H also contains no Hamiltonian cycle.



(a) The supercard $\mathcal{G}$ which consists of $n+1$ vertices. Deleting a green vertex results in the graph G, whereas deleting a blue vertex gives H.



(b) An example of a graph G with a Hamiltonian cycle and graph H without an Hamiltonian path that have $\lfloor \frac{n}{2} \rfloor + 1$ common cards (where $n = 10$). All non-black vertices of the same color correspond to the same common cards.

Figure 12: Two graphs G and H with a supercard $\mathcal{G}$.

This example generalizes to a whole family of pairs by taking $\mathcal{G}$ equal to C_{2k} plus some extra vertex u that is connected to every even vertex on the cycle. The graph G is obtained by deleting a neighbor of u, while H is obtained by deleting a non-neighbor of u. We can switch any green odd vertex in the cycle C_{2k} of $\mathcal{G}$ with the vertex u to obtain a new cycle of size $2k$ in $\mathcal{G}$. Hence, deleting a green odd vertex of C_{2k} results in a graph G with a Hamiltonian cycle. Deleting a blue even vertex of C_{2k} from $\mathcal{G}$, results in a graph H with two vertices of degree 1, implying that H does not contain a Hamiltonian cycle. Moreover, if H contains a Hamiltonian path then these two vertices must be the end vertices of this path. Furthermore, H contains $k-2$ vertices of degree 2, so H has k vertices of degree at most 2. These vertices of degree at most 2 lie on C_{2k} such that the vertices of degree at most 2 and of degree 3 alternate. Hence, the only path through all vertices of degree at most 2 is the path corresponding to the cycle C_{2k}. But this path does not contain the extra vertex u, so H does not contain a Hamiltonian path.

9 Discussion

In this thesis, we have developed some machinery to determine whether a graph has a 'big' hidden cycle if we have $n - \epsilon n$ arbitrary cards of a graph G on n vertices. The main idea was to mark vertices that lie close to big branches or other cycles as bad (in this thesis always red). This means that for good vertices v, most of the vertices that laid in the 'big hidden' cycle are lying in the longest path. This also implies that you can clearly see what most of the branches/small cycles around the big cycle in $G - v$ look like.
However in order to use this information, you have to ensure that at least one common card corresponds to a good vertex. This requirement led to different bounds for different properties. We were for example able to show that we can determine whether a connected graph on n vertices has a cycle based on any $\frac{n}{2} - \epsilon\sqrt{n}$ cards, whereas we needed at least $\frac{2n}{3} + 2$ cards to determine the length of a smallest cycle. Moreover, if we wanted to determine whether a graph had an odd cycle we needed at least $\frac{5n}{6} + 2$ cards, as we had to ensure that there was a big gap between 'large' and 'small cycles' as otherwise our coloring would not work.
I believe that these colorings can also be used to determine other properties that have to do with the (non-)existence of a few cycles based on any $n - \epsilon n$ cards. The main idea is to pick enough cards such that you can see whether there are any 'small' cycles and then you need to determine whether there is a big hidden one. For example, I believe it is possible to determine whether a graph has at most one cycle or more than one cycle based on any $n - \epsilon n$ cards. However, if one wants to prove this one has to take into account the possibility that there are two small cycles of the same length.
Although we have found upper bounds for girth and bipartiteness of the form of $n - \epsilon n$, I suspect that the true upper bounds are way lower. To be honest, I suspect that the lower bounds of $\lfloor \frac{n}{2} \rfloor + 2$ are in fact the real bounds. Therefore, I propose the following conjectures and questions:

Conjecture 9.1. *For n large enough, the number of common cards between a graph G and H, both on n vertices, with different girth is at most $\lfloor \frac{n}{2} \rfloor + 1$.*

Conjecture 9.2. *For n large enough, the number of common cards between a bipartite graph G and a non-bipartite graph H that both have n vertices, is at most $\lfloor \frac{n}{2} \rfloor + 1$.*

Question 9.3. *What are all infinite families of pairs of graphs G and H both on n vertices, such that G is bipartite and H not, that have exactly $\lfloor \frac{n}{2} \rfloor + 1$ common cards?*

References

[1] Bowler, A., Brown, P., and Fenner, T. Families of pairs of graphs with a large number of common cards. *Journal of Graph Theory 63*, 2 (2010), 146–163.

[2] Brown, P. *On the maximum number of common cards between various classes of graphs*. PhD thesis, University of London, 2008.

[3] Brown, P., and Fenner, T. A new approach to graph reconstruction using supercards. *Journal of Combinatorics 9*, 1 (2018), 95–118.

[4] Euler, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae* (1741), 128–140.

[5] Fiorini, S. A theorem on planar graphs with an application to the reconstruction problem, i. *The Quarterly Journal of Mathematics 29*, 3 (1978), 353–361.

[6] Groenland, C. Blog post: Size reconstructibility and the graph reconstruction conjecture, 2018.

[7] Groenland, C., Guggiari, H., and Scott, A. Size reconstructibility of graphs. *Journal of Graph Theory 96*, 2 (2021), 326–337.

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

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

[10] Kostochka, A. Lecture notes: Reconstruction, 2021.

[11] McKay, B. Reconstruction of small graphs and digraphs. *Preprint* (2022).

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

[13] Myrvold, W. The ally-reconstruction number of a disconnected graph. *Ars Combinatoria 28* (1989), 123–127.

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

[15] Tutte, W. *All the Kings Horses*. Academic Press, New York, 1979.

A Omitted proofs

This section of the appendix contains some results and proofs that were earlier skipped.

A.1 Components is cards of connected graphs

Lemma A.1. *Let H, G be connected graphs such that $v(H) < \frac{v(G)}{2}$. Then at most $\left\lfloor \frac{v(G)}{v(H)+1} \right\rfloor$ cards of G contain a connected component isomorphic H.*

Proof. If there is no card of G that contains a connected component isomorphic to H we are immediately do. So let v be a vertex such that H is a connected component of $G - v$. We look at the set $V(H_G)$ of vertices in G that lie in this component. As G is connected, there is a vertex $w_1 \in V(H_G)$ and $x \in G - v - V(H_G)$ such the edges (w_1, v) and (v, x) exist.
Let $w_i \in V(H_G)$, we will show that the graph $G - w_i$ does not contain a connected component H. The graph $G/V(H_G)$ is connected as G is connected and H is a connected component in $G - v$. This means that the component in $G - w_i$ containing v has at least $v(G) - v(H_G) > v(H_G)$ vertices and is thus not isomorphic to H. All other connected components are a subset of $V(H_G) - w_i$ which has strictly less vertices than H. Thus these components are also not isomorphic to H. This means that for all vertices $w_i \in V(H_G)$, the graph $G - w_i$ does not contain a connected component isomorphic to H.
Let $u \neq v$ and $G - u$ be another card with a connected component H' which is isomorphic to H. By the previous section, we see that $u \neq w_i$. We will show that $V(H) \cap V(H'_G) = \emptyset$. As H is a connected component in $G - v$ and $u \notin V(H_G)$, we find that in $G - u$ the vertex v is in the same connected component as H. This means that all vertices of H lie in a connected component with more vertices then H, so none of the vertices in $V(H_G)$ are also in $V(H'_G)$.
This means that for every vertex v such that $G - v$ contains a connected component of size k, there are uniquely defined $v(H)$ vertices for which $G - w$ does not contain such a component. Let $|C|$ be the number of cards which have H as a connected component, then we see that $|C| \cdot (v(H) + 1) \leq v(G)$ which concludes the proof. □

Remark A.2. *The bounds of the previous lemma are sharp. Draw a cycle C with $v(G) - k \left\lfloor \frac{v(G)}{k+1} \right\rfloor$ vertices and add then a copy of H to $\left\lfloor \frac{v(G)}{k+1} \right\rfloor$ points in this cycle C. Then for each of those $\left\lfloor \frac{v(G)}{k+1} \right\rfloor$ points, we have that $G - v$ contains a connected component isomorphic to H.*

Looking at the proof of Lemma A.1, we only looked at the number of vertices of the components of $G - v$ and never whether a connected component of size k was isomorphic to H. This means that we have proved the following even stronger statement.

Theorem A.3. *Let G be a connected graph and let $1 \leq k < \frac{v(G)}{2}$. Then there are at most $\left\lfloor \frac{v(G)}{k+1} \right\rfloor$ vertices $v \in G$ such that $G - v$ contains a connected component of size exactly k.*

A.2 Upper bound common cards between forests and non-forests

Lemma A.4. *Forests and non-forests have at most $\frac{2n}{3}$ common cards.*

Proof. We will use a proof by contradiction. Suppose that a forest T and a graph G with a cycle, both on n vertices, have at least $\frac{2n}{3} + 1$ common cards. Let L be the length of a cycle of G, then $L \geq \frac{2n}{3} + 1$ as the cycle is not visible on any of the common cards. Now, Lemma 6.1 tells us that G is unicyclic, so let L be the length of the unique cycle. Then as G is unicyclic, G has at least $L - (n - L) = 2L - n = 2(L - \frac{2n}{3} - 1) + \frac{n}{3} + 2$ vertices of degree 2 on its cycle. As at most $L - (\frac{2n}{3} + 1)$ vertices on the cycle of G do not correspond to a common card, at least $\frac{n}{3} + 2$ common cards do correspond to a vertex of degree 2.
Let K_G be the number of connected components in G and K_T the number of connected components of T. Then $e(G) = n + 1 - K_G$ and $e(T) = n - K_T$. For every common card corresponding to a vertex of degree 2 in the cycle of G, the number of connected components is equal to K_G.
Therefore, $K_T \leq K_G + 1$. If $K_T = K_G + 1$ then T must have at least $\frac{2n}{3} + 1$ isolated vertices, as deleting them is the only way to decrease the number of connected component. This indicates that $K_G \geq \frac{2n}{3} + 1$.

However, as the cycle of G already has $\frac{2n}{3}+1$ vertices, G also has a connected component with at least $\frac{2n}{3}+1$ vertices, which leads to a contradiction. Therefore, $K_T \leq K_G$ and $e(G) > e(T)$.
Therefore, for every common card $T - v = G - w$ with $\deg(w) = 2$, the fact $e(G) > e(T)$ implies $\deg(v) \leq 1$. This indicates that T has at least $\frac{n}{3} + 2$ vertices of degree at most 1. Then $T - v = G - w$ has at least $\frac{n}{3} + 1$ such vertices. As w has degree 2, this indicates that G has at least $\frac{n}{3} - 1$ vertices of degree at most 1 as this number only decreases if a neighbor of w had degree 1 or less. But as G already has $\frac{2n}{3} + 1$ vertices on the cycle, which all have degree at least 2, all bounds must be tight.
In particular, this means that $L = \frac{2n}{3} + 1$ and thus every vertex on the cycle of G corresponds to a common card. Moreover, this also tells us that every vertex of degree 2 in the cycle has two neighbors of degree at most 2. Hence, all vertices in the cycle have we degree 2. Therefore, T has at least $\frac{2n}{3} + 1$ vertices of degree at most 1 and G has $\frac{2n}{3} - 2$ vertices of degree at most 1. As G has n vertices, this gives us that $n \geq \frac{2n}{3} - 2 + \frac{2n}{3} + 1$, implying $3 \geq n$.
On the other hand, we also know that $\frac{2n}{3} + 1 \leq n$, so $n = 3$. In this case, G is a triangle and all 3 cards of G are connected. However, as all vertices of T have degree at most 1, we see that T is disconnected. So T does not have 3 connected cards, a contradiction. $\square$