An $\Omega(n^{\frac{4}{3}})$ lower bound on the randomized complexity of graph properties

Péter Hajnal *
Department of Computer Science, Princeton University
and
Bolyai Institute, University of Szeged, Hungary

* The paper was written while the author was a graduate student at the University of Chicago and was completed at M.I.T. The work was supported in part by NSF under GRANT number NSF 5-27561, the Air Force under Contract OSR-86-0076 and by DIMACS (Center for Discret Mathematics and Theoretical Computer Science), a National Science Foundation Science and Technology Center – NSF-STC88-09648.

Abstract.

We improve King's $\Omega(n^{\frac{5}{4}})$ lower bound on the randomized decision tree complexity of monotone graph properties to $\Omega(n^{\frac{4}{3}})$. The proof follows Yao's approach and improves it in a different direction from King's. At the heart of the proof are a duality argument combined with a new packing lemma for bipartite graphs.

0. Introduction.

Our universe is the set of graphs on the vertex set $V = \{1, 2, ..., n\}$ i.e. the set of labeled graphs $\mathcal{G}_n$. A *graph property* P is a subset of $\mathcal{G}_n$, closed under isomorphisms. We use $\bar{P}$ to denote $\mathcal{G}_n - P$, the complement of P. The dual, P^* is defined as follows: $G \in P^*$ iff $\bar{G} \notin P$ where $\bar{G}$ is the complement of the graph G. P is monotone (increasing) if $H \subset G, H \in P$ implies $G \in P$. P is non-trivial if it differs from $\emptyset$ and $\mathcal{G}_n$. Let $\mathcal{P}_n$ be the set of non-trivial, monotone, graph properties over the universe $\mathcal{G}_n$.

Definition 0.1. A *decision tree algorithm* A computes a graph property P for any input G by asking questions of the form "Is edge {i,j} in the graph?". The choice of question may depend only on the information gained so far. Let $\mathcal{A}_P$ be the set of all decision tree algorithms for P. Let $cost(A, G)$ be the number of queries asked by A when G is the input. The *deterministic complexity* of P is $\mathcal{C}(P) = \min_{A\in\mathcal{A}_P}\max_{G\in\mathcal{G}_n} cost(A, G)$.
The *randomized decision tree algorithm* is a probability distribution α over $\mathcal{A}_P$. The expected number of queries asked by α for input G is $cost^R(\alpha, G) = \sum_{A\in\mathcal{A}}\alpha(A)cost(A, G)$. The *randomized complexity* of P is $\mathcal{C}^R(P) = \min_\alpha\max_{G\in\mathcal{G}_n} cost^R(\alpha, G)$.

It has been conjectured by Aanderaa and Rosenberg [R] that the deterministic complexity of any non-trivial monotone graph property is $\mathcal{C}(P) = \Omega(n^2)$. The conjecture was proved by Rivest and Vuillemin [RV]. Further progress was made by Kahn, Saks and Sturtevant [KSS]. Subsequently it has been conjectured that the randomized complexity of any non-trivial monotone graph property is $\mathcal{C}^R(P) = \Omega(n^2)$. The first nonlinear lower bound was obtained by Yao [Y2] who proved $\mathcal{C}(P) = \Omega(n\log^{\frac{1}{12}} n)$. More important, Yao introduced methods which form the basis of further progress. A significant improvement is due to Valerie King: $\mathcal{C}(P) = \Omega(n^{\frac{5}{4}})$. We further improve this to $\Omega(n^{\frac{4}{3}})$.

Main Theorem. (Theorem 7.1.) *The randomized decision tree complexity of any non-trivial monotone graph property $P \in \mathcal{P}_n$ is*

$$\mathcal{C}^R(P) = \Omega(n^{\frac{4}{3}}).$$

Our lower bound (as King's) will be based on a method described by Yao in [Y2]. See chapter 2 for a formal description. In the proof, we shall utilize our ability to choose between the properties $P, \bar{P}, P^*$, each having the same complexity. This liberty can be exploited through the following new packing lemma which may be of independent interest.

Packing Lemma. (Theorem 5.5.) *Let G and H be two bipartite graphs with the same two color classes, V and W, both of them of size n. Assume that*
(a) $d_{average}(G)d_{max,V}(H) \leq \frac{n}{3000}$,
(b) $d_{average}(H)d_{max,V}(G) \leq \frac{n}{3000}$,
(c) $d_{max,W}(G), d_{max,W}(H) \leq \frac{n}{10\log n}$.
Then G and H can be packed.

1.Preliminaries, notations.

Let $\mathcal{C}^R(n) = \min_{P \in \mathcal{P}_n} \mathcal{C}^R(P)$.

A property $P \in \mathcal{P}_n$ can be characterized by the collection of minimal graphs having that property. We refer to them as minimal or critical graphs. Let $min(P)$ be the list of critical graphs for P. In many of cases we consider P as it is given by $min(P)$.

We use standard graph theoretical notions. We refer the reader to [L]. If $G \in \mathcal{G}_n$ then $V(G)$ is the set of vertices of G. $d_{max}(G)$ and $d_{average}(G)$ are the maximal and average degree of G. $K_n, K_{n,m}, E_n$ are the complete graph on n vertices, the complete bipartite class with color classes of size m and n and the empty graph on n vertices, respectively. $G[A]$ is the subgraph of G induced by the set A.

We need to build up the same system of notions for another universe, namely the universe of labeled bipartite graphs. We have bipartite graphs with the two color classes: $V = \{1, 2, ..., n\}$ and $W = \{\bar{1}, \bar{2}, ..., \bar{m}\}$. The set of all these labeled bipartite graphs is denoted by $\mathcal{G}_{n,m}$. It is straightforward to extend the previous notions, like non-trivial, monotone bipartite graph property P to this class. (When we refer to a bipartite property, we automatically mean that the property is non-trivial, monotone and that it does not depend on the labelling.) The set of these bipartite properties will be denoted by $\mathcal{P}_{n,m}$. We also use the other corresponding notions $\mathcal{C}^R(P)$, $\mathcal{C}^R(n,m)$, $min(P)$, P^* (it will be clear from the context whether the universe we are working with is graphs or bipartite graphs).

If $G \in \mathcal{G}_{n,m}$ and U is a subset of the vertices then $d_{max,U}(G)$ and $d_{average,U}(G)$ are the maximal and average degree in the set U. For example if $|V| = |W|$ then $d_{average,V}(G) = d_{average,W}(G)$. (Let us remember that if G comes from $\mathcal{G}_{n,m}$ then V and W are the two color classes.) Some obvious modifications of the graph theoretical definitions are made when dealing with bipartite graphs, for example the complement of a bipartite graph will come from $\mathcal{G}_{n,m}$.

2. Previous techniques.

In previous papers several methods giving lower bounds for the random complexity of properties were presented. The lower bounds on $\mathcal{C}^R(P)$ given by these methods depended explicitly on $min(P)$. None of these bounds could give a general lower bound for all properties, but the combination of several methods led to superlinear lower bounds.

Theorem 2.1. *(Basic Method [Y1]).*
(i) Let $P \in \mathcal{P}_n$ and $G \in min(P)$ be any minimal graph for P. Then

$$\mathcal{C}^R(P) = \Omega(|V(G)| d_{average}(G)).$$

(ii) Let $P \in \mathcal{P}_{n,m}$ and $G \in min(P)$ be any minimal graph for P. Then

$$\mathcal{C}^R(P) = \Omega(|V| d_{average,V}(G)).$$

Definition 2.2. Let $\mathcal{L}$ be a list of graphs from $\mathcal{G}_{m,n}$. For each $G \in \mathcal{L}$ let us consider the sequence of degrees in color class V. Let $\mathcal{S}(G) = d_1 \geq d_2 \geq ... \geq d_n$ be the ordered list of degrees. If $(d_1, d_2, ..., d_n)$ is the lexicographically first sequence considering all the ordered lists are gotten from elements of $\mathcal{L}$ then we refer to G as the V-lexicographically first element of $\mathcal{L}$.

Theorem 2.3. *(Yao's Method [Y2]).*
Let $P \in \mathcal{P}_{n,m}$ and $G \in min(P)$ be the V-lexicographically first graph. Then

$$\mathcal{C}^R(P) = \Omega(\frac{d_{max,V}(G)}{d_{average,V}(G)}|V|).$$

Remark. In the case of bipartite universe the roles of V and W are exchangeable.

All these lower bounds are proven using a basic lemma of Yao [Y1].

Yao's method [Y2] is very powerful. The only problem with it is that we can apply it only for a very specific graph of the list $min(P)$. We need a slight extension of this method.

Lemma 2.4. *Let $P \in \mathcal{P}_{n,m}$ and let us assume that there is a graph in $min(P)$ which has at least $\frac{n}{2}$ isolated nodes in V. Let G be the V-lexicographically first graph among the graphs having at least $\frac{n}{2}$ isolated nodes in V. Then*

$$\mathcal{C}^R(P) = \Omega(\frac{d_{max,V}(G)}{d_{average,V}(G)}|V|).$$

Proof. (Sketch) We assume familiarity with the proof of Yao's method [Y2, Lemma 5]. In that proof we chose the node of maximal degree from V and some other points of degree at most twice the average degree. Now we do the same but we carefully leave $\frac{n}{2}$ isolated nodes out of the consideration. The crucial point of the proof is that we construct many graphs none of which has property P. This is shown by the fact that these graphs don't have any subgraph from $min(P)$. This is still true for the new collection: For graphs having $\frac{n}{2}$ isolated nodes this follows from the fact that G was the V-lexicographically first. For the other graphs it is true because they have less than $\frac{n}{2}$ isolated nodes in V. ∎

None of these methods works for every graph property, but for any property one of them gives a good lower bound. Combining different methods one might get a superlinear lower bound. Unfortunately Yao's method doesn't seem to work on general graph properties. This is the reason that the known lower bounds handle the bipartite properties first, then they give a reduction of the general problem to the bipartite case.

Finally we mention V. King's method, that led to her improved bound. Our proof won't use this technique.

Theorem 2.5. *(King's Method [K]).*
(i) Let $P \in \mathcal{P}_n$ and $G \in min(P)$ be any minimal graph for P. Then

$$\mathcal{C}^R(P) = \Omega\left(\frac{|V(G')|^2}{d_{average}(G')d_{max}(G)}\right),$$

where G' is the subgraph of G induced by the non-isolated nodes.
(ii) Let $P \in \mathcal{P}_{n,m}$ and $G \in min(P)$ be any minimal graph for P. Then

$$\mathcal{C}^R(P) = \Omega\left(\frac{|V'|^2}{d_{average,V'}(G)d_{max,W}(G)}\right),$$

where V' is the subset of V of nodes with positive degree.

3. Using duality.

Yao's original lower bound technique had a problem with graphs from $min(P)$ with small number of edges. He solved this problem considering the dual property P^*. $min(P)$ and $min(P^*)$ determine each other. This dependence gives a possibility to get information about one of the lists knowing something about the other. This kind of information is very useful because it gives guidance in choosing the right list to work with. The very basic fact about these two lists is the following.

Definition 3.1. (a) Let $G, H \in \mathcal{G}_n$. Let us assume that G has the vertex set V and H has the vertex set V'. A *packing* is an identification between V and V' such that no edges of G is identified with any edges of H.
(b) Let $G, H \in \mathcal{G}_{n,n}$. Let us assume that G has color classes $V \cup W$ and H has color classes $V' \cup W'$. A *bipartite packing* is two identifications, one is between V and V' and another between W and W' such that no edge of G is identified with any edge of H.

Lemma 3.2. *(Fact [Y2]).*
(i) If $P \in \mathcal{P}_n$, $G \in min(P)$ and $H \in min(P^*)$ then G and H can't be packed.
(ii) If $P \in \mathcal{P}_{n,m}$, $G \in min(P)$ and $H \in min(P^*)$ then G and H can't be packed as a bipartite graphs.

Packing graphs is a heavily studied subject in graph theory. A good survey of this research can be found in [B]. For us the following theorem is specially important.

Theorem 3.3. *(Conditions on the maximal degree [SS],[C]).*
(i) If $G, H \in \mathcal{G}_n$ and $d_{max}(G)d_{max}(H) < \frac{n}{2}$ then G and H can be packed.
(ii) If $A, B \in \mathcal{G}_{n,m}$ and $d_{max,V}(A)d_{max,W}(B) + d_{max,W}(A)d_{max,V}(B) < n$, then A and B can be packed as bipartite graphs.

¿From the packing property one can get the following useful information for the bipartite case: at least one of $min(P)$ and $min(P^*)$ contains a graph where the class V has at least $\frac{n}{2}$ nodes of positive degree.

The packing property doesn't give all the dependence between the two critical lists. Actually $min(P^*)$ is the set of minimal graphs which can't be packed with any element of $min(P)$. The maximality gives us further information about the lists.

Lemma 3.4. *Let $P \in \mathcal{P}_{n,m}$. Then $min(P)$ or $min(P^*)$ has a graph G such that it has at least $\frac{n}{2}$ isolated nodes in V.*

Proof. $G = K_{\frac{n}{2},m} + E_{\frac{n}{2}} \in \mathcal{G}_{n,m}$ is a graph with $\frac{n}{2}$ isolated nodes in V and with all the possible edges. If $G \in P$ then the statement is clearly true. Otherwise $\bar{G} = G \in P^*$ and again the statement is clearly true. ∎

4. Surgery on the maximal degree.

From now on for simplicity we restrict ourselves to the bipartite universe where the two color classes have the same cardinality. So now we can use the following notation. $d_{average}(G) = d_{average,V}(G) = d_{average,W}(G)$.

The basic idea of this chapter is the following. Let us fix a bipartite property P. Let us consider $min(P)$. If we have a graph of high average degree in $min(P)$ then we get a good lower bound by our basic method. If we take a graph to which we can apply Yao's method, the average degree is low and the maximal degree in the corresponding color class is high then Yao's technique gives us a good bound. In some sense we can interpret these statements as if we can't get a good lower bound on $\mathcal{C}(P)$ then we have an upper bound on the maximal degree of a special graph from $min(P)$ in one of the color classes. Let us assume that the second case occurs (so we could not get a good lower bound in the way we described).If we do the same thing for $min(P^*)$ then we are left with two graphs. We have bounds on the maximal degree in some color classes. That gives us the possibility to force a contradiction using packing theorems. The first step in this program is to choose graphs to which we can apply Yao's method.

Definition 4.1. Let $P \in \mathcal{P}_{n,n}$.
(a) Remember that the two color classes in this universe are V and W. We assume that $min(P)$ has some graphs with at least $\frac{n}{2}$ isolated nodes in V. Let G be the V-lexicographically first among them.
(b) For clarity we denote the two color classes of H by V' and W'. Let H be the W'-lexicographically first element of $min(P^*)$.

Our previous remark suggests that we can handle G and H as we have an upper bound on the maximal degrees in their color classes. In order to apply for example Catlin's packing theorem, Theorem 3.3.(ii) we need bounds on the maximal degree in all other color classes. We proceeds as follows. We start to build a packing between G and H. This partial packing leaves us with some leftover, unpacked nodes. These define a new packing problem which, in some sense, is independent from the original one. The same time we will have a more complete knowledge about the maximal degrees in the color classes.

Definition 4.2. (Prepacking)
We are going to define sets $V_0 \subset V$, $W_0 \subset W$, $V_0' \subset V'$ and $W_0' \subset W'$ ($|V_0| = |V_0'|, |W_0| = |W_0'|$) and a packing between the corresponding induced subgraphs of G and H.
(a) Let V_0 be the set of $\frac{n}{2}$ isolated nodes in G.
(b) Let W_0' be the set of $\min\{\frac{n}{8d_{average}(H)}, \frac{n}{2}\}$ nodes of lowest degree in H.
(c) Let V_0' be the neighborhood of W_0' and plus as many of the highest degree nodes from V' as needed in order to get a set of size $\frac{n}{2}$. (We will see that the size of the neighborhood V_0 is at most $\frac{n}{4}$.)
(d) Let W_0 be the set of $\min\{\frac{n}{8d_{average}(H)}, \frac{n}{2}\}$ nodes in W of highest degree in G.
Let G_0 and H_0 be the subgraphs of G and H induced by the sets defined above. Note that G_0 is empty. Let V_1, W_1, V_1' and W_1' the leftover parts of the corresponding vertex sets. Let G_1 and H_1 be the subgraphs induced by these leftover sets.
Any pair of identifications between V_0 and V_0' plus between W_0 and W_0' gives a packing between G_0 and H_0. Choose arbitrary one of these, and call it prepacking.

Lemma 4.3. *(i) If G_1 and H_1 can be packed then the packing and the prepacking together yield a total packing.*
(ii) $d_{max,V_1}(G_1) \leq d_{max,V}(G)$.
(iii) $d_{max,W_1'}(H_1) \leq d_{max,W'}(H)$.
(iv) $d_{max,V_1'}(H_1) \leq 4d_{average}(H)$.
(v) $d_{max,W_1}(G_1) \leq 8d_{average}(G)d_{average}(H)$.

Proof (i) We don't have any edges in G between V_0 and W_1 and in H between W_0' and V_1'. Thus we won't have any conflict putting together the two packings.

(ii) and (iii) are obvious because G_1 is a subgraph of G and H_1 is a subgraph of H.

(iv) $V_1' = V' - V_0'$. We are going to show that the neighborhood of W_0' has at most $\frac{n}{4}$ nodes. This implies that V_0' has the $\frac{n}{4}$ highest degrees in V'. The claim about the size of the neighborhood of W_0' is clear because all degrees in W_0' are not greater then $2d_{average}(H)$.

(v) If the statement were not true then the contribution of the edges having an end-point in W_0 to the total number of edges in G would be greater than $|W|d_{average}(G)$. ∎

Theorem 4.4. *The randomized decision tree complexity of any non-trivial monotone bipartite graph property $P \in \mathcal{P}_{n,n}$ is $\Omega(n^{\frac{5}{4}})$, i.e.,*

$$\mathcal{C}^R(n,n) = \Omega(n^{\frac{5}{4}}).$$

Proof Let us fix an arbitrary $P \in \mathcal{P}_{n,n}$. Let $G \in min(P)$ and $H \in min(P^*)$ be the two graphs defined in Definition 4.1.

Case 1. $d_{average}(G)$ or $d_{average}(H)$ is at least $\frac{1}{10}n^{\frac{1}{4}}$.
In this case our basic method gives the lower bound.

Case 2. $d_{max,V}(G)$ or $d_{max,W'}(H)$ is at least $\frac{1}{10}n^{\frac{1}{2}}$ and case 1 doesn't hold.

Without loss of generality we can assume that $d_{max,V}(G)$ is at least $\frac{1}{10}n^{\frac{1}{4}}$. Because of the choice of G we can apply Yao's method and we get the lower bound

$$\Omega(\frac{d_{max,V}(G)}{d_{average}(G)}n).$$

We know that $d_{average}(G)$ is at most $\frac{1}{10}n^{\frac{1}{4}}$. Thus we get the desired lower bound.

Case 3. None of the previous cases holds.

Let G_1 and H_1 be the graphs defined in Definition 4.2. It is easy to check that the condition of Theorem 3.3.(ii) is satisfied. So G_1 and H_1 can be packed. Using Lemma 4.3.(i) we get that G and H can also be packed, which is a contradiction. ∎

5. The improved packing theorem for bipartite graphs.

In the previous section we heavily used Catlin's packing theorem for bipartite graphs. The proof of that theorem is very simple. In this section we improve his result and in this way we get an improved lower bound on our problem too. We summarize Catlin's idea. Given $G, H \in \mathcal{G}_{n,n}$ with color classes V, W, V' and W'. We want to find a sufficient condition for existence a packing. We take an arbitrary identification of V and V'. Define a bipartite graph between W and W' based on whether two nodes can be identified or not. Now the problem is simply finding a matching in this auxiliary graph.

Definition 5.1. Let $G, H \in \mathcal{G}_{n,n}$. Let be V, W, V' and W' the corresponding color classes. Given f, an identification of V and V'. B_f is a bipartite graph between W and W'. $x \in W$ and $y \in W'$ are adjacent iff x and y can be identified, i.e., the neighborhoods of x and y are disjoint in the identified V and V'.

Now it is easy to show that if G and H satisfy the condition of Theorem 3.3.(ii) then B_f satisfies the condition of König's theorem [see L]. Our improvement comes from the idea that we don't pick an arbitrary f but rather choose a random one. Then with positive probability B_f has the property that all vertices in it have degree at least $\frac{n}{2}$. We need a probabilistic lemma. For that we use a well-known inequality for Bernoulli random variable.

Theorem 5.2. *Let $X_1, X_2, ..., X_N$ be independent 0-1 random variables such that $Prob(X_i = 1) = p$.*
(i) (Chernoff's inequality [Ch])
If $m \geq Np$ is an integer then

$$Prob(\sum_{i=1}^{N} X_i \geq m) \leq \left(\frac{Np}{m}\right)^m \exp(m - Np).$$

(ii) [AV,Ch]
For every $0 < \beta < 1$,

$$Prob(\sum_{i=1}^{N} X_i \leq \lfloor(1-\beta)Np\rfloor) \leq \exp\left(-\frac{\beta^2 Np}{2}\right).$$

Lemma 5.3. *Let $0 \le d_1, d_2, ..., d_n \le L = \frac{n}{10 \log n}$ integers and let us define $d_{average}$ by $\sum_{i=1}^{n} d_i = d_{average} n$. Let $X_1, X_2, ..., X_n$ be independent $0-1$ random variables such that $Prob(X_i = 1) = p = \frac{1}{1000 d_{average}}$. Then $Prob(\sum_{i=1}^{n} X_i d_i > \frac{n}{10}) \le \frac{1}{n^2}$.*

Proof. Let $X_{i,j} = X_i$ for $1 \le i \le j \le d_i$. Using this notation

$$\sum\nolimits_{i=1}^{n} X_i d_i = \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{d_i} X_{i,j}$$

Let $\mathcal{X}$ be the set of random variables defined above. The size of $\mathcal{X}$ is $\sum_{i=1}^{n} d_i = n d_{average}$. The elements of $\mathcal{X}$ are arranged in a matrix such a way that each column consists of independent random variables. It is easy to see that $\mathcal{X}$ can be partitioned into L subsets $(\mathcal{X} = \mathcal{X}_1 \cup ... \cup \mathcal{X}_L)$ such a way that each $\mathcal{X}_i$ has independent random variables in it and their sizes are the same (for all i,j $\Big||\mathcal{X}_i - \mathcal{X}_j|\Big| \le 1$).

Let E be the event that

$$\sum\nolimits_{i=1}^{n} X_i d_i = \sum\nolimits_{X_{i,j} \in \mathcal{X}} X_{i,j} > \frac{n}{10},$$

and E_k be the event that

$$\sum\nolimits_{X_{i,j} \in \mathcal{X}_k} X_{i,j} > |\mathcal{X}_k| \frac{1}{10 d_{average}} = 100p|\mathcal{X}_k|.$$

Then

$$E \subset \cup_{k=1}^{L} E_k.$$

Using Chernoff's inequality we get

$$\begin{aligned} Prob(E_k) &\le \left(\frac{p|\mathcal{X}_k|}{100p|\mathcal{X}_k|}\right)^{100p|\mathcal{X}_k|} \exp(100p|\mathcal{X}_k| - p|\mathcal{X}_k|) \\ &\le \left(\frac{e}{100}\right)^{\log n} \le \frac{1}{n^3} \end{aligned}.$$

Thus

$$Prob(E) \le \frac{d_{average} n}{L} Prob(E) \le \frac{1}{n^2}.$$

∎

We need the same probabilistic lemma but in a different probabilistic model.

Lemma 5.4. *Let $0 \le d_1, d_2, ..., d_n \le \frac{n}{10 \log n}$ integer numbers such that $\sum_{i=1}^{n} d_i = d_{average} n$. Choose D such that $D d_{average} \le \frac{n}{3000}$. Let Δ be a random subset of $\{1, 2, ..., n\}$ of size D. Then $Prob(\sum_{i \in \Delta} d_i \ge \frac{n}{2}) < \frac{2}{n^2}$.*

Proof. Let Δ_i be a random subset of $\{1, 2, ..., n\}$ of size i, all i-subsets of $\{1, 2, ..., n\}$ being equally likely. Let $P_i = Prob(\sum_{j \in \Delta_i} d_j > \frac{n}{2})$. Then $P_0 \le P_1 \le ... \le P_n$.

Let $X_1, X_2, ..., X_n$ be independent 0-1 random variables such that $Prob(X_i = 1) = p = \frac{1}{1000 d_{average}}$. Then

$$Prob(\sum_{i=0}^{n} X_i d_i > \frac{n}{2}) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} P_k$$
$$\geq P_{\lfloor \frac{1}{2} np \rfloor} \sum_{\lfloor \frac{1}{2} np \rfloor \leq k \leq \lfloor \frac{3}{2} np \rfloor} \binom{n}{k} p^k (1-p)^{n-k} .$$
$$\geq \frac{1}{2} P_{\lfloor \frac{1}{2} np \rfloor} \geq \frac{1}{2} P_D$$

Using our bound from Lemma 5.3 we get the desired inequality. ∎

As sketched above this lemma leads to an improved packing theorem.

Theorem 5.5. *Let $G, H \in \mathcal{G}_{n,n}$. Assume that*
(a) $d_{average}(G) d_{max,W'}(H) \leq \frac{n}{3000}$,
(b) $d_{average}(H) d_{max,W}(G) \leq \frac{n}{3000}$,
(c) $d_{max,V}(G), d_{max,V'}(H) \leq \frac{n}{10 \log n}$.
Then G and H can be packed.

Proof For the sake of clarity let us assume that the color classes of G and H are V, W, V' and W'. Let $f : V \to V'$ be a random $1-1$ map, all maps being equally likely. Then B_f is a random bipartite graph between the sets W and W'.

We are interested in the event

$$E = \text{ Each node of } B_f \text{ has degree at least } \frac{n}{2}.$$

One elementary bad event is

$$E_x = x \text{ has degree in } B_f \text{ less then } \frac{n}{2} \text{ (for } x \in W \cup W').$$

Using this notation

$$E = \Omega - \cup_{x \in W \cup W'} E_x.$$

Thus

$$Prob(E) \geq 1 - \sum_{x \in W \cup W'} Prob(E_x).$$

For $x \in W$, E_x is exactly the event that the image $f(N(x))$ of $N(x)$ ($f(N(x)) \subset V'$) has a neighborhood in W' of size more than $\frac{n}{2}$. The event that the sum of the degrees in $f(N(x))$ is at least $\frac{n}{2}$ is a superset of E_x. $f(N(x))$ is a random set of size $|N(x)|$ and its size is at most $d_{max,W}(G)$. Applying Lemma 5.4 we get that $Prob(E_x) < \frac{1}{2n}$. So $Prob(E)$ is positive. This proves that there exists an f such that for the corresponding B_f each node has degree at least $\frac{n}{2}$. Thus there is a perfect matching in B_f. This perfect matching is an identification of W and W', which together with f gives us a packing. ∎

We are going to use this improved packing theorem in order to get the following improved lower bound.

Corollary 5.6. *The randomized decision tree complexity of any non-trivial monotone bipartite graph property $P \in \mathcal{P}_{n,n}$ is $\Omega(n^{\frac{4}{3}})$, i.e.,*

$$\mathcal{C}^R(n,n) = \Omega(n^{\frac{4}{3}}).$$

Proof. Let $P \in \mathcal{P}_{n,n}$ be an arbitrary graph property. Let G and H be the graphs defined in Definition 4.1. We are going to consider three cases.

Case 1. $d_{average}(G)$ or $d_{average}(H)$ is at least $\frac{1}{100}n^{\frac{1}{3}}$.

Applying Lemma 2.1. we get the lower bound.

Case 2. $d_{max,V}(G)$ or $d_{max,W'}(H)$ is at least $\frac{1}{100}n^{\frac{2}{3}}$ and case 1 does not hold.

Because of the choice of G we can apply Yao's method and we get the lower bound $\Omega(\frac{d_{max,V}(G)}{d_{average}(G)}n)$. In this case we know that $d_{average}(G)$ is at most $\frac{1}{100}n^{\frac{1}{3}}$. These imply the lower bound.

Case 3. None of the previous cases holds.

Let us consider G_1 and H_1, the graphs defined in Definition 4.2. It is easy to check that the conditions of the new packing theorem, Theorem 5.5. are satisfied. So G_1 and H_1 can be packed. This leads to a contradiction that proves our theorem. ∎

6. The improved reduction.

Given a graph property P one can construct other graph properties, that are useful for proving lower bounds on the complexity of P.

Definition 6.1. *Let $P \in \mathcal{P}_n$. Let us assume that $K_{\frac{n}{2}} + E_{\frac{n}{2}} \notin P$ and $K_n - K_{\frac{n}{2}} \in P$. Let $\tilde{P} \in \mathcal{P}_{\frac{n}{2},\frac{n}{2}}$ be the following property. $G \in \tilde{P}$ iff adding all the possible edges between the nodes of V to G gives us a graph having property P.*

Definition 6.2. *Let $P \in \mathcal{P}_n$. Let us assume that $K_n - K_{\frac{n}{2}} \notin P$. Let $\hat{P} \in \mathcal{P}_{\frac{n}{2}}$ be the following property. $G \in \hat{P}$ iff adding $\frac{n}{2}$ new nodes and all the possible edges incident to a new node to G gives us a graph having property P.*

Considering these problems helps us because of the following lemma. Basically it says that it is enough to give a lower bound on the constructed property. In the case when the first definition might be applied the advantage is obvious, we get a reduction to the bipartite case.

Lemma 6.3. *Let $P \in \mathcal{P}_n$ be an arbitrary non-trivial graph property. Let us assume that $\tilde{P}$ and $\hat{P}$ are the properties defined in 6.1. and 6.2. Then the following are true.*
(i) $\tilde{P}$ is a non-trivial, monotone bipartite graph property.
(ii) $\mathcal{C}^R(P) \geq \mathcal{C}^R(\tilde{P})$.
(iii) $\hat{P}$ is a non-trivial, monotone graph property.
(iv) $\mathcal{C}^R(P) \geq \mathcal{C}^R(\hat{P})$. ∎

Another advantage is that we might have "nice" critical graphs for the constructed property. This way it is easier to handle a lower bound on that property.

Lemma 6.4. *Let $P \in \mathcal{P}_n$ be a graph property and $\hat{P}$ be the property defined in 6.2. Let $G \in min(P)$. Then there is an $H \in min(\hat{P})$ such that the following are true.*
(i) $d_{max}(H) \leq 4d_{average}(G)$.
(ii) H has at least $\frac{n}{10d_{average}(G)}$ isolated nodes.

Proof. Let V be the vertex set of G, ($|V| = n$). Let us take any subset V_0 of V of size $\frac{n}{2}$. Then the subgraph of G induced by V_0, $G[V_0]$ has the property $\hat{P}$. Thus $min(\hat{P})$ has an element that is a subgraph of $G[V_0]$. So it is enough to show that for an appropriate set V_0, $G[V_0]$ has the properties (i) and (ii).

Choose $\min\{\frac{n}{2}, \frac{n}{10d_{average}(G)}\}$ nodes by the following greedy algorithm. Choose the node of minimum degree in G. Throw away that point and its neighborhood. Choose the node of minimum degree in the remaining graph and continue this procedure. The set that we shall get will be an independent set and its neighborhood will have size less then $\frac{n}{4}$. Let us refer to this independent set as A. Let us extend $N(A)$ to a set of size $\frac{n}{2}$ by adding some nodes of largest degree. Notice that we add at least $\frac{n}{4}$ new nodes. Let B be the set obtained after this extension of $N(A)$. Let V_0 be the complement of B. Let us remark that $A \subset V_0$.

It is easy to see that V_0 defined above is a good set. (i) follows from the fact that B has the set of nodes of the greatest $\frac{n}{4}$ degrees. (ii) is true because $A \subset V_0$. ∎

After this we are ready to prove the improved reduction to the bipartite case.

Theorem 6.5. *The randomized decision tree complexity of any non-trivial monotone bipartite graph property $P \in \mathcal{P}_n$ is*

$$\mathcal{C}^R(P) = \Omega\big(\min\{\, n^{\frac{4}{3}}, \mathcal{C}^R(\frac{n}{2}, \frac{n}{2})\,\}\big).$$

Proof Let $P \in \mathcal{P}_n$ be arbitrary graph property. We consider two cases.

Case 1. $K_{\frac{n}{2}} + E_{\frac{n}{2}} \notin P$ and $K_n - K_{\frac{n}{2}} \in P$.

Then $\mathcal{C}^R(P) \geq \mathcal{C}^R(\tilde{P}) = \Omega(\mathcal{C}^R(\frac{n}{2}, \frac{n}{2}))$.

Case 2. Case 1. does not hold.

Without loss of generality we assume that $K_n - K_{\frac{n}{2}} \notin P$. (This must hold for P or P^*.)

Let us consider $\hat{P}$. For any $G \in min(P)$ construct an $H \in min(\hat{P})$ guaranteed in Lemma 6.4 to exist. Choose any $F \in min(\hat{P}^*)$. We know that F and H have no packing. Start a prepacking the following way. Pack all the nodes of the top $\frac{n}{10d_{average}(G)}$ degrees of F into isolated nodes of H. Let the unpacked nodes span the graphs F_1 and H_1.

It is easy to see that F_1 and H_1 can't have a packing. F_1 has maximal degree at most $10d_{average}(G)d_{average}(F)$. From 6.4 H_1 has maximal degree at most $4d_{average}(G)$.

We finish the proof by considering the following two subcases.

Subcase 1. $d_{average}(F) \geq \frac{1}{10}n^{\frac{1}{3}}$ or $d_{average}(G) \geq \frac{1}{10}n^{\frac{1}{3}}$.

Then our basic methods gives us that $\mathcal{C}^R(P) \geq \mathcal{C}^R(\hat{P}) = \Omega(n^{\frac{4}{3}})$.

Subcase 2. Subcase 1 is not satisfied.

Then Catlin's theorem, Theorem 3.3.(ii) gives us a contradiction. ∎

7. The improved lower bound.

Combining Corollary 5.6. and Theorem 6.5. we get the following improved lower bound on general graph properties.

Theorem 7.1. *The randomized decision tree complexity of any non-trivial monotone graph property* $P \in \mathcal{P}_n$ *is* $\Omega(n^{\frac{4}{3}})$, *i.e.,*

$$\mathcal{C}^R(n) = \Omega(n^{\frac{4}{3}}).$$

∎

Acknowledgement.

The author would like to thank M. Szegedy, J. Simon, L. Babai and H. Karloff for fruitful discussions and helpful suggestions.

References

[AV] D. Angluin and L. G. Valiant, Fast probabilistic algorithms for Hamiltonian circuits and matchings, *Journal of Computer and System Sciences*, **19**, 155-193.

[B] B. Bollobás, *Extremal Graph theory*, Chapter VIII., Academic Press, 1978.

[C] P. A. Catlin, Subgraphs of graphs I., *Discrete Math.* **10** (1974), 225-233.

[Ch] H. Chernoff, A measure of asymptotic effiency for tests of a hypothesis based on the sum of observations, *Annals of Math. Stat.*, **23** (1952), 493-509.

[K] V. King, An $\Omega(n^{\frac{5}{4}})$ lower bound on the randomized complexity of graph properties, to appear in *Combinatorica*.

[KSS] J. Kahn, M. Saks and D. Sturtevant, A topological aproach to evasiveness, *Combinatorica*, **4** (4) (1984), 297-306.

[L] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland 1979.

[R] A. L. Rosenberg, On the time required to recognize properties of graphs: A problem, *SIGACT News*, **5** (4) (1973), 15-16.

[RV] R. Rivest and J. Vuillemin, A generalization and proof of the Aanderaa-Rosenberg conjecture, *Proc. 7th SIGACT Conference*, (1975), ACM 1976.

[SS] N. Sauer and J. Spencer, Edge-disjoint replacement of graphs, *J. of Combinatorial Theory* Ser. B **25** (1978), 295-302.

[Y1] A. Yao, Probabilistic computation: towards a unified measure of complexity, *Proc. 18th IEEE FOCS*, 1977, pp. 222-227.

[Y2] A. Yao, Lower bounds to randomized algorithms for graph properties, *Proc. 28th IEEE FOCS*, 1987, pp. 393-400.