Estimating the number of graphs containing very long induced paths

Steven Butler
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA

Abstract

Let $\mathcal{P}(n,k)$ denote the number of graphs on $n+k$ vertices that contain P_n, a path on n vertices, as an induced subgraph. In this note we will find upper and lower bounds for $\mathcal{P}(n,k)$. Using these bounds we show that for k fixed, $\mathcal{P}(n,k)$ behaves roughly like an exponential function of n as n gets large.

1 Introduction

In this note we will consider a graphical enumeration problem of graphs containing very long induced paths. Our graph terminology is standard and for any undefined terms we refer the reader to [3].

Let $\mathcal{P}(n,k)$ denote the number of simple graphs on $n+k$ vertices that contain P_n, a path on n vertices, as an induced subgraph. We will give an upper and lower bound for $\mathcal{P}(n,k)$ by representing such graphs as a combination of an array and a labeled graph. From these bounds it will follow that for fixed values of k that

$$\mathcal{P}(n,k) \sim \frac{2^{(nk+\binom{k}{2})}}{2k!}.$$

Our main tool will be to use Burnside's Lemma to count equivalency classes. We state the theorem below and note that its proof can be found in several places including [2].

Theorem 1 (Burnside's Lemma). *Let C be a collection of objects acted on by a group A, let N be the number of equivalence classes under A, and for $\pi \in A$ let $|C_\pi|$ be the number of elements in C fixed under π. Then $N = \sum_{\pi\in A} |C_\pi|/|A|$.*

ARS COMBINATORIA 88(2008), pp. 321-332

This note proceeds as follows. We first introduce a representation of graphs that are counted by $\mathcal{P}(n,k)$. We then count such representations to derive upper and lower bounds and look at its asymptotic behavior. At the end we will make some remarks about the proof and related problems.

2 Representing our graphs

If $G = (V, E)$ is a graph with $H = (V', E')$ as an induced subgraph of G, then G can be decomposed into three structures. Namely, the induced subgraph H, the induced subgraph H' (the induced subgraph on the vertices $V \setminus V'$), and the edges connecting H and H'.

This third structure can be represented by an array where each column is associated with a vertex of H and each row with a vertex of H'. Each entry of the array is marked with one of two colors according to whether the two corresponding vertices are connected by an edge.

Applying this to our problem, given $G = (V, E)$, a graph on $n + k$ vertices with a vertex set $N \subseteq V$ which induces P_n, then we construct the following object to represent G. Our object consists of a $k \times n$ array and a graph on k labeled vertices where the following holds.

1. Each vertex of N corresponds to a column and two columns are adjacent if and only if the corresponding vertices are adjacent.
2. The set $V \setminus N$ is labeled by $\{1, 2, \ldots, k\}$ where the ith vertex corresponds to the ith row of the array, the corresponding labeled induced subgraph on $V \setminus N$ is the graph on k labeled vertices in our object.
3. We color the entries of the array either 'edge' or 'not edge' depending on whether the vertex that corresponds to the column is adjacent to the vertex that corresponds to the row.

When we graphically represent these objects we will use dark for 'edge' and white for 'not edge', we also suppress the labeling on the graph with k vertices by putting each vertex next to its corresponding row.

Clearly, given such an object we can reconstruct G. However, given a G which contains a long induced path there are possibly many such objects which can be constructed. This can occur for two reasons.

First, we can have P_n as an induced subgraph of G in more than one way. An example of this is shown in Figure 1.

We will address this when we find the lower bound.

The second reason that we can have multiple objects is that there is freedom in how we assign vertices to the rows and columns. When assigning vertices to the columns we must assign one of the end vertices of the path to either the first or last column and then the remaining choices are

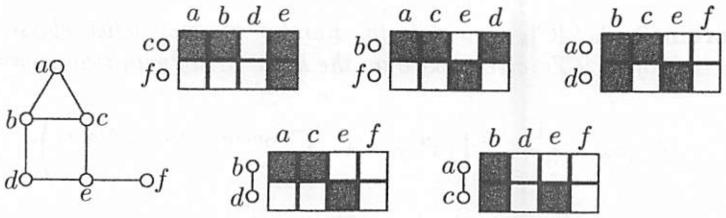

Figure 1: A graph with multiple P_4's and corresponding objects

determined. So there are two distinct ways to assign the vertices to the columns, the difference between the two a reversing of the columns in the array.

Similarly, when we label the vertices for the induced subgraph on $V \setminus N$ we have complete freedom on assigning the labels and so there are $k!$ ways of assigning these k vertices to the rows, the difference between any two such assignments a permutation of the rows.

Different choices in our assignment can be represented by an action of $S_k \times Z_2$ on the array. This action is composed of two parts, the element of S_k determines how to permute the rows while the element of Z_2 determines whether to reverse the columns. This gives us the following lemma.

Lemma 1. For a fixed set N that induces P_n in the graph G, the set of arrays which can be generated by the above construction is an equivalence class under the action of $S_k \times Z_2$ on all of the colorings of the array by two colors.

Let C be a collection of representative colorings of the $k \times n$ array with two colors, i.e., every coloring of our $k \times n$ array is equivalent with exactly one coloring of C under the action by $S_k \times Z_2$. By Lemma 1 for each graph G and $N \subseteq V$ which induces a P_n there is exactly one element of C that will appear in a constructed object for G.

Note that restricting our colorings to C does not completely overcome the arbitrariness of how we choose to associate vertices with the rows and the columns. Ambiguity can still arise when there is a non-identity element in $S_k \times Z_2$ for which the coloring is invariant. We will address this when we find the lower bound.

2.1 Counting our inequivalent colorings

The elements of C will form the basis for our upper and lower bounds. Our next step is to use Burnside's Lemma to find the cardinality of C. Notationally, for $\sigma \in S_k$ let $e(\sigma)$ and $o(\sigma)$ denote the number of even and odd cycles respectively in the cycle decomposition of σ.

Theorem 2. *Let $|C|_{k\times n}$ denote the number of equivalence classes under the action of $S_k \times Z_2$ given above on the $k \times n$ array with two colors. Then*

$$|C|_{k\times n} = \frac{1}{2k!}\left(\prod_{i=0}^{k-1}(2^n + i) + \sum_{\sigma\in S_k} 2^{ne(\sigma)+\lfloor(n+1)/2\rfloor o(\sigma)}\right),$$

moreover, $|C|_{k\times n}/2^{nk} \to 1/2k!$ as $n \to \infty$.

Proof. To apply Burnside's Lemma we find the number of colorings of the array that are fixed under the action of $\sigma \times a \in S_k \times Z_2$. We consider three cases.

First case: a cycle in σ with columns fixed. We choose an arbitrary coloring for the row that corresponds to the first element of our cycle. Then in order for our coloring to remain fixed every other row corresponding to the elements of the cycle must have the same coloring. So for every cycle of σ we get n choices.

Second case: an even cycle in σ with columns reversed. We choose an arbitrary coloring of the row that corresponds to the first element of the cycle. Then as we go through the elements of the cycle we reverse the order and fill in the rows as we go. When we return to the row that corresponds to the first element of the cycle we will have made an even number of reversals and so we will match up with what we started with. So for every even cycle of σ we get n choices.

Third case: an odd cycle in σ with columns reversed. We start as in the previous case, now though when we return to the row that corresponds to the first element of the cycle we will have made an odd number of reversals and we will have the reverse of what we started with. In order to match up, the coloring of the first row has to be symmetric. So for every odd cycle of σ we get $\lfloor(n+1)/2\rfloor$ choices.

We now apply Burnside's Lemma and get that

$$|C|_{k\times n} = \frac{1}{2k!}\Bigg(\underbrace{\sum_{\sigma\in S_k} 2^{n(e(\sigma)+o(\sigma))}}_{\text{(i)}} + \underbrace{\sum_{\sigma\in S_k} 2^{ne(\sigma)+\lfloor(n+1)/2\rfloor o(\sigma)}}_{\text{(ii)}}\Bigg)$$

where (i) comes from elements with columns fixed (case 1) while (ii) comes from elements with columns reversed (cases 2 and 3).

Examining (i) note that $e(\sigma)+o(\sigma)$ is the number of cycles in σ. We can group the permutations of S_k according to how many cycles the permutations have into k groups (for $1, 2, \ldots, k$) each with $s(k,i)$ elements where i is the number of cycles. Here $s(k,i)$ denotes the (unsigned) Stirling numbers of the first kind. Using a property of the (unsigned) Stirling numbers of the first kind we have that

$$\sum_{\sigma\in S_k} 2^{n(e(\sigma)+o(\sigma))} = \sum_{i=1}^{k} s(k,i)2^{ni} = \sum_{i=1}^{k} s(k,i)(2^n)^i = \prod_{i=0}^{k-1}(2^n+i).$$

Substituting this in for (i) gives us our first result for $|C|_{k\times n}$.

For (ii) note that for σ in S_k that $2e(\sigma)+o(\sigma)\le k$ and so

$$ne(\sigma)+\lfloor (n+1)/2\rfloor\, o(\sigma) \le \lfloor (n+1)/2\rfloor\,(2e(\sigma)+o(\sigma)) \le \lfloor (n+1)/2\rfloor\, k.$$

As an immediate consequence we have

$$\sum_{\sigma\in S_k} 2^{ne(\sigma)+\lfloor (n+1)/2\rfloor o(\sigma)} \le \sum_{\sigma\in S_k} 2^{\lfloor (n+1)/2\rfloor k} = k!2^{\lfloor (n+1)/2\rfloor k}$$

Putting this in for (ii), along with what we have already done for (i), we have

$$\frac{1}{2k!}\left(\prod_{i=0}^{k-1}(2^n+i)\right) \le |C|_{k\times n} \le \frac{1}{2k!}\left(\prod_{i=0}^{k-1}(2^n+i)+k!2^{\lfloor (n+1)/2\rfloor k}\right).$$

Dividing through by 2^{nk} we see that

$$\frac{1}{2k!}\prod_{i=0}^{k-1}(1+i2^{-n}) \le \frac{|C|_{k\times n}}{2^{nk}} \le \frac{1}{2k!}\left(\prod_{i=0}^{k-1}(1+i2^{-n})+k!2^{(\lfloor (n+1)/2\rfloor -n)k}\right).$$

Letting $n\to\infty$, the first and last terms go to $1/2k!$ giving the second result. □

3 Upper and lower bounds

3.1 Upper bound

Recall that any graph which contains P_n as an induced subgraph has a presentation as some array from C and a labeled graph. Since there are $|C|_{k\times n}$ and $2^{\binom{k}{2}}$ such arrays and graphs respectively, we immediately get the following.

Theorem 3. *We have $\mathcal{P}(n,k)\le |C|_{k\times n}2^{\binom{k}{2}}$.*

3.2 Overcounting

To get our lower bound we will create a combination of colorings and labeled graphs which will correspond to non-isomorphic graphs. This will be achieved by making restrictions on the colorings of our arrays that will overcome the two problems of overcounting that are inherent in our upper bound.

The first problem we will address is having a long induced path in more than one way. In Figure 1 we saw that one graph can have P_n as an induced subgraph in multiple ways. When looking at all combinations of colorings of C and labeled graphs this caused some graphs to be counted multiple times. We can get around this problem by use of the following lemmas.

Lemma 2. Let $G = (V, E)$ and let $N \subsetneq V$ be a collection of n vertices which will induce a graph with maximum degree q. Then for all $v \in N$ we have $\deg(v) \leq |V| - n + q$.

Proof. Since the maximum degree of a vertex in the induced graph is q, if $v \in N$ then v can be adjacent to at most q other vertices of N. In particular v is not adjacent to $n - q$ of the vertices of V lying in N. Thus, the maximum degree that v can have is $|V| - (n - q)$. □

Lemma 3. If the $k \times n$ array in a representation of G has at least $k+3$ entries in each row colored 'edge' then G contains P_n as an induced subgraph in only one way.

Proof. Each row corresponds to a vertex in the graph, and by our assumption each vertex corresponding to a row has degree at least $k + 3$. By Lemma 2, with $|V| = n + k$ and the maximal degree of P_n as 2, it follows that none of the k vertices that correspond to the rows can lie in an induced subgraph which is P_n. Thus only n of the $n + k$ vertices can lie in an induced subgraph which is P_n and so we have only one way to have P_n as an induced subgraph. □

So by Lemma 3, adding the restriction that our colorings have at least $k + 3$ or more entries in each row colored 'edge' eliminates the problem of having P_n as an induced subgraph in more than one way. We now derive an upper bound for the number of graphs not satisfying this property.

Lemma 4. Let D be a maximal collection of inequivalent colorings with two colors of the $k \times n$ array under the action of $S_k \times Z_2$, such that each coloring of D contains at least one row with $k+2$ or fewer elements colored 'edge.' Then

$$|D| \leq \left(\sum_{i=0}^{k+2} \binom{n}{i} \right) |C|_{k-1 \times n}.$$

Proof. Consider the collection of $k \times n$ arrays where the first row contains $k+2$ or fewer elements colored 'edge', and the remaining $k-1$ rows come from $C_{k-1\times n}$. Since there are a total of $\sum_{i=0}^{k+2} \binom{n}{i}$ different ways that the first row can have $k+2$ or fewer elements colored 'edge', and there are $|C|_{k-1\times n}$ colorings, we have at most

$$\left(\sum_{i=0}^{k+2} \binom{n}{i}\right) |C|_{k-1\times n}$$

such objects.

All that remains is to show that a coloring in D is equivalent to one of them. To see this, start with any element in D then act on it to send a row with $k+2$ or fewer elements colored 'edge' to the first row. Rows 2 through k are a coloring of the $(k-1)\times n$ array and so there is an element in $S_{k-1} \times Z_2$ which acts on these $k-1$ rows which takes it to a coloring of $C_{k-1\times n}$. We can extend this action to the $k \times n$ array. Doing so the first row will still have $k+2$ or fewer elements colored 'edge' and the remaining rows are a coloring in $C_{k-1\times n}$, as desired. □

The second source for overcounting is because some colorings of C are invariant under multiple actions. This can cause the object associated with a graph to have several possible labeled graphs associated with a fixed coloring of the array. An example of this situation is shown in Figure 2.

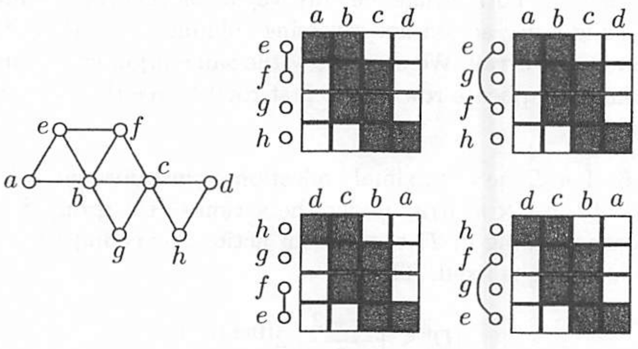

Figure 2: Objects with a fixed coloring and multiple labeled graphs

The different actions for which the coloring is invariant corresponds to different sets of choices in how we assign vertices, this can possible lead to different labeled graphs. If we add the restriction that our colorings remain invariant only under the identity automorphism then this problem is eliminated. We need to get a bound on the number of colorings which satisfy this condition, this will be done by separating into two cases and examining each case in turn as done in the following lemmas.

Lemma 5. Let D be a maximal collection of inequivalent colorings with two colors of the $k \times n$ array under the action of $S_k \times Z_2$ such that for each coloring of D there is a non-identity action, not reversing the columns, for which the coloring is fixed. Then

$$|D| \le (k-1)|C|_{k-1\times n}$$

Proof. For each coloring in $C_{k-1\times n}$ we construct $k-1$ colorings of the $(k-1)\times n$ array by first putting the coloring in rows 2 through k and then for the first row we duplicate, in turn, each of the rows 2 through k. An example of this construction is shown in Figure 3.

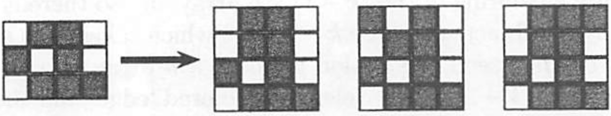

Figure 3: An example of the construction in Lemma 5

This constructs at most $(k-1)|C|_{n-1\times k}$ colorings. All that remains is to show that each coloring of D is equivalent to at least one of the colorings that we have constructed.

This follows by noting that the only way a coloring can remain invariant under a non-identity action not reversing columns is for there to be a duplicate row in the array. We now apply the same argument as in Lemma 4 (now sending a duplicate row to the first row) to see that we have all possible colorings. □

Lemma 6. Let D be a maximal collection of inequivalent colorings with two colors of the $k \times n$ array under the action of the group $S_k \times Z_2$ such that for each coloring of D there is an action, reversing the columns, in which the coloring is fixed. Then

$$|D| \le \left\lfloor \frac{k+2}{2} \right\rfloor 2^{\lfloor (n+1)/2 \rfloor k}.$$

Proof. Consider a row of a coloring in D. If the row of is not symmetric then in order for the coloring to remain invariant under an action which reverses the order of the columns there must be some row which has the coloring in reverse order. In particular, the rows of d are either symmetric or they can be placed in pairs which are the reverse of each other.

Suppose that we have j rows that are paired together. Then the remaining $k-2j$ rows must be symmetric. For every pair of rows we get to make a full choice for one row and the other row will have its coloring determined, so we get a total of nj choices. For the symmetric rows we get to color half the row and the other half must be colored in reverse order to be symmetric and so we get $(k-2j)\lfloor (n+1)/2 \rfloor$ choices.

The number of pairs that we can have is between 0 and $\lfloor k/2 \rfloor$, so we get $|D| \leq \sum_{j=0}^{\lfloor k/2 \rfloor} 2^{nj+(k-2j)\lfloor (n+1)/2 \rfloor}$. Since $n \leq 2\lfloor (n+1)/2 \rfloor$ we have

$$\begin{aligned}\sum_{j=0}^{\lfloor k/2 \rfloor} 2^{nj+(k-2j)\lfloor (n+1)/2 \rfloor} &\leq \sum_{j=0}^{\lfloor k/2 \rfloor} 2^{2j\lfloor (n+1)/2 \rfloor+(k-2j)\lfloor (n+1)/2 \rfloor} \\ &= \sum_{j=0}^{\lfloor k/2 \rfloor} 2^{\lfloor (n+1)/2 \rfloor k} = \left(\left\lfloor \frac{k}{2} \right\rfloor + 1\right) 2^{\lfloor (n+1)/2 \rfloor k} \\ &= \left\lfloor \frac{k+2}{2} \right\rfloor 2^{\lfloor (n+1)/2 \rfloor k}.\end{aligned}$$

Any coloring that is invariant under an action which reverses the columns will be equivalent to one of these, i.e., we permute the rows to put the pairs in order at the top of the array and the symmetric rows at the bottom. This concludes the proof. □

3.3 The lower bound

Theorem 4. *We have*

$$(|C|_{k\times n} - (\sum_{i=0}^{k+2} \binom{n}{i} + (k-1))|C|_{k-1\times n} - \left\lfloor \frac{k+2}{2} \right\rfloor 2^{\lfloor (n+1)/2 \rfloor k}) 2^{\binom{k}{2}} \leq \mathcal{P}(n,k).$$

Proof. Let D be a maximal collection of inequivalent colorings of the $k \times n$ array where every coloring satisfies the following two conditions.

1. Every row of the coloring has at least $k+3$ entries colored 'edge'.
2. The only action in $S_k \times Z_2$ for which the coloring is fixed is the identity.

Consider all combinations of arrays with colorings of D and graphs on k labeled vertices.

Any graph on $n+k$ vertices which contains P_n as an induced subgraph is represented at most once in this set of combinations. To see this, let G be a graph on $n+k$ vertices. Then if the array in the constructed object for G is not in D it is not one of our graphs.

So suppose that the array constructed in our object for G is in D. Then by our restrictions of our arrays and Lemma 3 we know that the graph contains P_n as an induced subgraph in exactly one way, so there is only one coloring of D which can be used in a representation of the graph.

Because of our other restriction on the array there is only one possible way of assigning the vertices to the columns and the rows so that the coloring matches with the coloring of D. (If there were two distinct ways of assigning the vertices to the columns and the rows then we could form a non-identity automorphism for which the coloring would be invariant.) It follows that for G there is only one labeled graph that we can associate with the coloring of D. In particular, G can only show up once in the combinations of the colorings of D and all labeled graphs. So we can conclude that $|D|2^{\binom{k}{2}} \leq \mathcal{P}(n,k)$. All that remains is to bound $|D|$.

To bound $|D|$ we start with $C_{k\times n}$ and then get an upper bound for the number of colorings which need to be removed in order to get D. Our upper bound for the number of colorings that we need to remove comes by combining Lemmas 4, 5 and 6. By Lemma 4, the number of graphs which do not satisfy the first condition placed on our array is bounded above by $\sum_{i=0}^{k+2}\binom{n}{i}|C|_{k-1\times n}$. By Lemmas 5 and 6, the number of graphs which do not satisfy the second condition placed on our array is bounded above by $(k-1)|C|_{k-1\times n} + \lfloor\frac{k+2}{2}\rfloor 2^{\lfloor (n+1)/2\rfloor k}$. Subtracting these terms out from $|C|_{k\times n}$ gives us our desired bound, concluding the proof. □

3.4 Asymptotic behavior

Theorem 5. *Let k be fixed. Then*

$$\lim_{n\to\infty}\frac{\mathcal{P}(n,k)}{2^{nk}} = \frac{2^{\binom{k}{2}}}{2k!}.$$

Proof. Starting with our bound from Theorem 3 and dividing both sides by 2^{nk} we have

$$\frac{\mathcal{P}(n,k)}{2^{nk}} \leq \frac{|C|_{k\times n}}{2^{nk}}2^{\binom{k}{2}},$$

applying Theorem 2 it follows that

$$\limsup_{n\to\infty}\frac{P(n,k)}{2^{nk}} \leq \frac{2^{\binom{k}{2}}}{2k!}.$$

Starting with our bound from Theorem 4 and dividing both sides by 2^{nk} and simplifying we have

$$\frac{\mathcal{P}(n,k)}{2^{nk}} \geq$$

$$\left(\frac{|C|_{k\times n}}{2^{nk}} - \frac{\sum_{i=0}^{k+2}\binom{n}{i} + (k-1)}{2^n}\frac{|C|_{k-1\times n}}{2^{n(k-1)}} - \left\lfloor\frac{k+2}{2}\right\rfloor 2^{(\lfloor (n+1)/2\rfloor - n)k}\right)2^{\binom{k}{2}}.$$

We proceed as before. The expression $\left(\sum_{i=0}^{k+2}\binom{n}{i}\right)+(k-1)$ is a polynomial in n of degree $k+2$, this is dominated by 2^n as $n \to \infty$. All the other terms are straightforward by the use of Theorem 2. It follows that

$$\liminf_{n\to\infty} \frac{P(n,k)}{2^{nk}} \geq \frac{2^{\binom{k}{2}}}{2k!}.$$

Combining the lim inf and lim sup gives us our desired result. □

Theorem 5 is equivalent to saying that for fixed values of k that

$$\mathcal{P}(n,k) \sim \frac{2^{(nk+\binom{k}{2})}}{2k!}.$$

So for fixed values of k we have that $\mathcal{P}(n,k)$ behaves as an exponential function of n as n gets large. We note the rate of growth is much smaller than that for all graphs (which is approximately $2^{\binom{n}{2}}/n!$). Showing, unsurprisingly, that graphs with very long induced paths become rare.

4 Generalizations and open problems

The approach presented here can be used for other graphs. For instance by minor modifications to the argument we have the following. Let $\{H_i\}_{i>m}$ be an infinite family of simple graphs where each H_i is a graph on i vertices with a trivial automorphism group and there is a universal q that bounds the maximum degree of H_i for all i. If $\mathcal{H}(n,k)$ denotes the number of simple graphs on $n+k$ vertices which contains H_n as an induced subgraph, then for fixed values of k we have that

$$\mathcal{H}(n,k) \sim \frac{2^{(nk+\binom{k}{2})}}{k!}.$$

Our approach, as is, will not work for all graphs. The advantage of the path was that its nontrivial automorphism reduced the number of squares to color in the array by essentially half. The graph shown in Figure 4 also has the same automorphism group as the path but the actions on the $k \times n$ array is notably different and the proof above cannot be easily adopted.

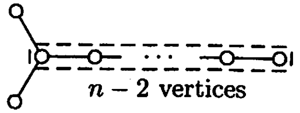

Figure 4: A more challenging graph

This is the simplest case of a whole range of problems. We might ask if there is a similar result for the infinite family in Figure 4. If the only thing that is important is the size of the automorphism group of the induced subgraph then we would expect approximately the same number of graphs on $n+k$ vertices with $n \gg k$ with long induced paths as we would for graphs which have Figure 4 as an induced subgraph. It is unknown if this holds.

Another interesting problem would be to try to find a similar, non-heuristic proof, for n-cycles. An initial conjecture would be as follows.

Conjecture 1. Let $C(n,k)$ denote the number of graphs on $n+k$ vertices which contain C_n, a cycle on n vertices as an induced subgraph. Then for $n \gg k$ we have

$$C(n,k) \approx \frac{2^{nk+\binom{k}{2}}}{2nk!}.$$

References

[1] Harary, F., Palmer, E.: Graphical Enumeration. New York: Academic Press 1973

[2] Harris, K., Hirst, J., Mossinghoff, M.: Combinatorics and Graph Theory. New York: Springer 2000

[3] D. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, 1996.