2000

Paths in Graphs

Béla Bollobás
University of Memphis

Amites Sarkar
Western Washington University, amites.sarkar@wwu.edu

Paths in Graphs

BÉLA BOLLOBÁS and AMITES SARKAR

Institute for Advanced Study
Olden Lane, Princeton NJ 08540

Department of Mathematical Sciences
University of Memphis, Memphis TN 38152

Department of Pure Mathematics and Mathematical Statistics
16, Mill Lane, Cambridge CB2 1SB, England

in REVISED FORM 4 April 2001

We prove that if $10 \leq \binom{k}{2} \leq m < \binom{k+1}{2}$ then the number of paths of length three in a graph G of size m is at most $2m(m-k)(k-2)/k$. Equality is attained iff G is the union of K_k and isolated vertices. We also give asymptotically best possible bounds for the maximum number of paths of length s, for arbitrary s, in graphs of size m. Lastly, we discuss the more general problem of maximizing the number of subgraphs isomorphic to a given graph H in graphs of size m.

1. Introduction

Given a graph G and an integer $s \geq 2$, write $p_s(G)$ for the number of paths of length s in G. In this paper we study the behaviour of the function

$$p_s(m) = \max\{p_s(G) : e(G) = m\},$$

the maximum number of paths of length s in a graph of size m. To simplify the presentation, we shall assume that all our graphs G contain no isolated vertices.

For $s = 2$ and $m \geq 4$ the function $p_s(m)$ is rather trivial: $p_2(m) = \binom{m}{2}$ and the star $K_{1,m}$ is the only extremal graph. In other words, if $e(G) = m \geq 4$ then $p_2(G) \leq \binom{m}{2}$, with equality iff $G \cong K_{1,m}$. Indeed, suppose that G has order n, size m, and degree sequence $d_1 \geq d_2 \geq \ldots \geq d_n > 0$. Then

$$p_2(G) = \frac{1}{2}\sum_{i=1}^{n} d_i(d_i - 1) \leq \frac{1}{2}\sum_{i=1}^{n}(n-1)(d_i - 1) = \frac{1}{2}(n-1)(2m-n) \leq \binom{m}{2},$$

with equality in both places iff $G \cong K_{1,m}$, proving our claim. (The second inequality follows by the concavity of the function $f(x) = (x-1)(2m-x)$.) When $m = 3$ there are two extremal graphs, $K_{1,3}$ and K_3.

Also, if we fix the order and not the size of G, the problem becomes trivial for every path length: if $|G| = n$ then $p_s(G) \leq \frac{1}{2}(n)_{s+1}$ with equality iff G is complete. Ahlswede and Katona [1], continuing work of Katz [9], determined the maximum number of paths of length 2 in a graph of order n and size m.

In the third section, we consider the case when s is odd and greater than 2. Here, we have $p_s(m) \sim 2^{\frac{s-1}{2}} m^{\frac{s+1}{2}}$, and, when $m = \binom{k}{2}$, the complete graph K_k has $2^{\frac{s-1}{2}} m^{\frac{s+1}{2}} + O(m^{\frac{s}{2}})$ paths of length s. For $s = 3$, we have a much more precise result as described in our first theorem. In the following two sections, we take s to be even and greater than 2. This time, we have $p_s(m) \sim C_{s/2} m^{\frac{s}{2}+1}$ for some constant $C_{s/2}$, which we determine explicitly, and for each $s = 2l \geq 4$ and m of a suitable form (divisible by a certain integer depending on l) we give a complete bipartite graph of size m with $C_{s/2} m^{\frac{s}{2}+1} + O(m^{\frac{s}{2}})$ paths of length s. Both the results and proofs for the odd and even cases are very different in character.

After a preliminary version of this paper was written, we discovered two papers of Alon [2], [3] and one of Füredi [6] also concerned with maximizing the number $N(m, H)$ of subgraphs isomorphic to a fixed graph H in a graph of size m. We discuss these in the next section. Roughly speaking, Alon [2] obtains asymptotically best possible results when H has a spanning subgraph which is a disjoint union of cycles and isolated edges, and thus his results match ours for paths of odd length and also deal with cycles of arbitrary length. However, when H is not of this form, he only determines the order of magnitude of $N(m, H)$, and thus our results for paths of even length superseed his. In [3] and [6], the authors take H to be a disjoint union of stars, a case not considered in this paper. The methods used in [6] to establish a conjecture from [3] somewhat parallel those we use to prove our Theorem 10, although the arguments are quite different. However, we include a sketch proof of the main result of [6] since it puts §4 and §5 in perspective.

2. The work of Alon and Füredi

As mentioned in the introduction, Alon [2], [3] and Füredi [6] consider the more general problem of maximizing the number of copies of a fixed graph H in graphs of size m. They denoted this maximum by $N(m, H)$ (actually Füredi's definition differs from ours and Alon's by a factor depending on H, since he labels the edges of both graphs). Thus $p_s(m) = N(m, P_s)$. Further, we write $N(G, H)$ for the number of copies of H in G, so that for example $N(K_n, K_t) = \binom{n}{t}$ and $p_s(G) = N(G, P_s)$.

Alon makes the following definition. A graph H is *asymptotically extremally complete* (a.e.c. for short) if for all m we have

$$N(m, H) = (1 + O(m^{-\frac{1}{2}}))N(K_n, H),$$

where K_n is the largest complete graph of size at most m. Of course, it is the large graph G which is "asymptotically complete". He then proves that if H' is a spanning subgraph of H and H' is a.e.c., then H is a.e.c., and that a disjoint union of a.e.c. graphs is a.e.c.. Therefore, if H has a perfect matching, then H is a.e.c.. This last remark determines the asymptotic value of $N(m, P_{2r+1})$ and also shows that

$$N(m, C_{2r}) \sim (1 + O(m^{-\frac{1}{2}}))\frac{2^{r-2}m^r}{r},$$

although it is easy to see independently that

$$N(m, C_{2r}) \leq \frac{N(m, P_{2r-1})}{2r}.$$

It also follows that complete graphs are a.e.c., although by the Kruskal-Katona theorem [8], [10] (see also [4]) we have a much stronger result, namely that if $m = \binom{t}{2} + t'$ then

$$N(m, K_r) = \binom{t}{r} + \binom{t'}{r-1}.$$

Alon then shows that odd length cycles are a.e.c.. We sketch a proof of this for C_5s. One uses induction on m to show that

$$N(m, C_5) \leq \frac{2\sqrt{2}}{5}m^{\frac{5}{2}}.$$

Suppose that $m = \binom{k}{2}$ for notational simplicity. Let G be a graph of size m with no isolated vertices, and let v be a vertex of degree $d < k - 1$ (if there is no such vertex then the result follows immediately). We estimate the number of C_5s through v. There are $\binom{d}{2}$ choices for the edges incident to v, at most $m - d$ choices for the opposite edge, and at most two ways to join them together. By induction, there are at most

$$\frac{2\sqrt{2}}{5}(m-d)^{\frac{5}{2}}$$

C_5s whose vertex sets are disjoint from $\{v\}$. Therefore we only need

$$\frac{2\sqrt{2}}{5}(m-d)^{\frac{5}{2}} + d(d-1)(m-d) \leq \frac{2\sqrt{2}}{5}m^{\frac{5}{2}},$$

which is readily established, since by convexity

$$\frac{2\sqrt{2}}{5}\{m^{\frac{5}{2}} - (m-d)^{\frac{5}{2}}\} \geq \frac{2\sqrt{2}}{5}\frac{5}{2}(m-d)^{\frac{3}{2}}.$$

The upshot is that if H has a spanning subgraph which is a disjoint union of cycles and isolated edges, then H is a.e.c.. The converse also holds. Alon first shows that H having such a spanning subgraph is equivalent to $\theta(H) = 0$, where for $S \subset V(H)$ we set

$$N(S) = \{x \in V(H) : xy \in E(H) \text{ for some } y \in S\},$$

and

$$\theta(H) = \max\{|S| - |N(S)| : S \in V(H)\}.$$

Then, given an arbitrary graph H he constructs a graph of size m containing at least $c_1 m^{\frac{1}{2}(|H|+\theta(H))}$ copies of H. (For a path of length four his construction reduces to a complete graph with about $\frac{m}{2}$ edges together with about $\frac{m}{4}$ independent vertices joined to two vertices of the complete graph. So, owing to the generality of his construction and the fact that he is only interested in an order of magnitude estimate, his "extremal" example is quite different from ours.) Finally, he shows that any graph of size m cannot contain more than $c_2 m^{\frac{1}{2}(|H|+\theta(H))}$ copies of H. In doing all this he makes extensive use of Hall's theorem.

Füredi in [6] discusses the above question when $H = H(\mathbf{a}) = H(a_1, \dots, a_t)$ is the vertex disjoint union of stars of $a_1, \dots, a_t$ edges. Note that

$$\theta(H(\mathbf{a})) = \sum_{i=1}^{t}(a_i - 1),$$

and so $H(\mathbf{a})$ is about as far as possible from being a.e.c.. Füredi calls a graph G *maximal* for H if $N(G, H) = N(e(G), H)$. In this terminology, Alon [3] conjectured that a maximal graph for a forest of stars $H(\mathbf{a})$ is necessarily a forest of stars. (Alon proved the case $t \leq 2$.)

To avoid complications, we will assume that the a_i are distinct.

In order to investigate $N(m, H(\mathbf{a}))$ further, Füredi defines the polynomial

$$p(\mathbf{a}, \mathbf{x}) = p(a_1, \dots, a_t; x_1, \dots, x_n) = \sum_{1 \leq i_1, \dots, i_t \leq n, \text{ all } i_j \text{ distinct}} x_{i_1}^{a_1} x_{i_2}^{a_2} \dots x_{i_t}^{a_t},$$

and the following two quantities:

$$p(\mathbf{a}, n) = \max\{p(\mathbf{a}, (x_1, \dots, x_n)) : x_1 \geq 0, \dots, x_n \geq 0, \sum_{i=1}^{n} x_i = 1\},$$

$$p(\mathbf{a}) = \sup_{n \geq t} p(\mathbf{a}, n).$$

In the notation of §5 and [7], $p(\mathbf{a}, \mathbf{x})$ is a scaled Muirhead mean of degree $\sum_{i=1}^{t} a_i$. (In the following argument, $p(\mathbf{a}, \mathbf{x})$, $p(\mathbf{a}, n)$, and $p(\mathbf{a})$ will play similar roles to those of $F_{t,n}(\mathbf{x})$, $\theta_{t,n}$ and θ_t from §5. One difference is that here Füredi has $t!$ times as many terms in his sum and he does not require that the x_i are decreasing.) From now on, we will suppose that $a_i \geq 2$ for all i and that $t \geq 2$.

Füredi's main result is the following.

Theorem 1. *Suppose that $a_i \geq 2$ for all i, $\sum_{i=1}^{t} a_i = A$ and*

$$B = \frac{1}{a_1! a_2! \dots a_t!}.$$

Then

$$N(m, H(\mathbf{a})) = Bp(\mathbf{a})m^A + O(m^{A-1}).$$

Proof. The first stage of the proof, which we omit, consists of showing that for some $n_0 = n_0(\mathbf{a})$ one has $p(\mathbf{a}, n) = p(\mathbf{a}, n_0)$ for $n > n_0$ - in doing this one must assume that all $a_i \geq 2$.

As in the case of paths, that

$$N(m, H(\mathbf{a})) \geq Bp(\mathbf{a})m^A + O(m^{A-1})$$

is essentially instant. For if

$$p(\mathbf{a}) = p(\mathbf{a}, (x_1, \dots, x_n))$$

then

$$N(m, H(\mathbf{a})) \geq N(H(\lfloor x_1 m \rfloor, \ldots, \lfloor x_n m \rfloor), H(\mathbf{a})) = Bp(\mathbf{a})m^A + O(m^{A-1}).$$

The difficulty lies in establishing the upper bound. (At this stage in our proof for paths, we will use the observation that a vertex of degree at most $m^{\frac{2}{3}}$ is contained in very few paths of fixed even length. This will enable us to focus our attention on a few vertices of high degree.) Füredi shows, by an edge-switching argument, that in any graph G maximal for $H(\mathbf{a})$, any edge must be incident with a vertex of degree at least Cm.

Let G be a graph of size m maximal for $H(\mathbf{a})$, where $\mathbf{a}$ is as in the statement of the theorem. For an edge $e \in E(G)$, write $M(e)$ for the number of copies of $H(\mathbf{a})$ in G which contain e. Set $M_{\max} = \max_{e \in E(G)} M(e)$ and let uv be an edge with $M(uv) = M_{\max}$. As $p(\mathbf{a}) \geq t!t^{-A}$ we certainly have

$$M_{\max} \geq C' m^{A-1}$$

if m is sufficiently large.

Let $pq \in E(G)$ be an arbitrary edge which may or may not be incident with uv, and suppose that $M(pq) < \frac{1}{3} M_{\max}$. At least $\frac{2}{3} M_{\max}$ copies of $H(\mathbf{a})$ contain uv but not pq, of which at least $\frac{1}{3} M_{\max}$ have u (say) as the centre of a star. Write G' for G with the edge pq removed and replaced by uw, where w is a new vertex. Then $N(G', H(\mathbf{a})) > N(G, H(\mathbf{a}))$, a contradiction. As a result, for all $pq \in E(G)$,

$$M(pq) \geq \frac{1}{3} C' m^{A-1}.$$

However, if an edge is contained in many copies of $H(\mathbf{a})$, then one of its endvertices must be of high degree. For we have

$$\begin{aligned} M(pq) &\leq \sum_{i=1}^{t} \binom{d_p - 1}{a_i - 1} m^{A-a_i} + \sum_{i=1}^{t} \binom{d_q - 1}{a_i - 1} m^{A-a_i} \\ &< 2t \max\{d_p, d_q\} m^{A-1} \left(\frac{\max\{d_p, d_q\}}{m} \right)^{\min a_i - 1} \\ &\leq 2t \max\{d_p, d_q\} m^{A-2}. \end{aligned}$$

Therefore,

$$\max\{d_p, d_q\} > \frac{C' m}{6t}.$$

Consequently, there is a set $W = \{w_1, \ldots, w_n\} \subset V(G)$ of less than $C'' = 12tC'$ vertices, each of degree more than $\frac{C'm}{6t}$, which together intersect every edge of G. Let G'' be the bipartite graph obtained from G by deleting all edges inside W. Then

$$\begin{aligned} N(G, H(\mathbf{a})) &\leq N(G'', H(\mathbf{a})) + \binom{C''}{2} m^{A-1} \\ &\leq Bp\left(\mathbf{a}, \frac{d_{w_1}}{m}, \ldots, \frac{d_{w_n}}{m}\right) m^A + O(m^{A-1}) \\ &\leq Bp(\mathbf{a}) + O(m^{A-1}). \end{aligned}$$

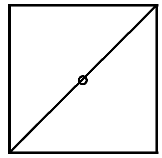

Figure 1 G_1

$\square$

Füredi's follow-up argument is that if for all i, $a_i > \log_2(t+1)$, then $p(\mathbf{a}) = p(\mathbf{a}, t)$, and so an $H(\mathbf{a})$-maximal graph is in this case a disjoint union of t stars.

Much remains to be done. In particular, for most graphs H with $\theta(H) \neq 0$, the asymptotic behaviour of $N(m, H)$ has yet to be determined. Once this has been done, there is still the problem of obtaining the H-maximal graphs themselves.

3. Paths of odd length

First let us consider the case $s = 3$. Here we restrict attention to graphs G of size $m \geq 6$ and we show that if $m = \binom{k}{2}$ (for some $k \geq 4$) then $p_3(G)$ is largest when G is complete, although for $k = 4$ there are two extremal graphs, each with 12 P_3's: one is K_4 and the other one, G_1, is drawn in Figure 1. The method of proof is similar to that of Theorem 4 in [5].

Theorem 2. *Let G be a graph of size m containing no isolated vertices, with $6 \leq \binom{k}{2} \leq m < \binom{k+1}{2}$. Then $p_3(G) \leq 2m(m-k)(k-2)/k$. Equality holds if and only if either $m = \binom{k}{2}$ and $G \cong K_k$ or $m = 6$ and $G \cong G_1$.*

Proof. Let $V(G) = \{x_1, \ldots, x_n\}$, and for each i, set

$$\begin{aligned} d_i &= d(x_i), \\ c_i &= e(G[\Gamma(x_i)]), \\ F(x_i) &= V(G) - \Gamma(x_i) \cup \{x_i\}, \\ f_i &= e(G[F(x_i)]), \\ e_i &= m - c_i - d_i - f_i. \end{aligned}$$

Note that $F(x_i)$ is the set of vertices at distance at least 2 from x_i, and that e_i is just the number of edges from $\Gamma(x_i)$ to $F(x_i)$. Further, we have $c_i \leq \binom{d_i}{2}$ for each i. Summing over paths whose middle edges are incident with x_i, we have

$$p_3(G) = \frac{1}{2} \sum_{d_i \geq 2} \{(d_i - 1)e_i + 2(d_i - 2)c_i\}$$

$$\leq \frac{1}{2}\sum_{d_i\geq 2}\left\{(d_i-1)e_i+2(d_i-2)\min\left\{\binom{d_i}{2}, m-e_i-d_i\right\}\right\}.$$

Hence, $2p_3(G)$ is at most

$$\begin{aligned}
&\sum_{d_i\geq 2} d_i \min\left\{(d_i-1)(d_i-2)+\frac{(d_i-1)e_i}{d_i}, 2m-\frac{4m}{d_i}-\frac{(d_i-3)e_i}{d_i}-2d_i+4\right\}\\
= &\sum_{d_i\geq 2} d_i \min\left\{(d_i-1)(d_i-2)+\frac{(d_i-1)e_i}{d_i}, 2m-\frac{4m+2d_i^2}{d_i}+4-\frac{(d_i-3)e_i}{d_i}\right\}\\
\leq &\sum_{d_i\geq 2} d_i \min\left\{(d_i-1)(d_i-2)+\frac{(d_i-1)e_i}{d_i}, 2m-\frac{4m+2k^2}{k}+4-\frac{(d_i-3)e_i}{d_i}\right\}\\
= &\sum_{d_i\geq 2} d_i \min\left\{(d_i-1)(d_i-2)+\frac{(d_i-1)e_i}{d_i}, \frac{2}{k}(m-k)(k-2)-\frac{(d_i-3)e_i}{d_i}\right\},
\end{aligned}$$

where we have used the convexity of $\frac{2m}{x}+x$ to obtain the inequality. When $d_i \geq 3$,

$$\frac{2}{k}(m-k)(k-2)-\left(1-\frac{3}{d_i}\right)e_i \leq \frac{2}{k}(m-k)(k-2),$$

and when $d_i = 2$,

$$(d_i-1)(d_i-2)+\left(1-\frac{1}{d_i}\right)e_i = \frac{e_i}{2} \leq \frac{m-2}{2} \leq \frac{2}{k}(m-k)(k-2),$$

provided $k \geq 4$, with the last inequality being strict for $m \geq 7$. Therefore,

$$\begin{aligned}
p_3(G) &\leq \frac{1}{2}\sum_{1\leq i\leq n, d_i\geq 2}\frac{2d_i}{k}(m-k)(k-2)\\
&\leq \sum_{i=1}^{n}\frac{d_i}{k}(m-k)(k-2)\\
&= \frac{2m(m-k)(k-2)}{k},
\end{aligned}$$

as claimed.

If $m \geq 7$, then for equality we need $\frac{2m}{d_i}+d_i = \frac{2m}{k}+k$ for every i. When $\binom{k}{2} < m < \binom{k+1}{2}$, this means that each vertex must have degree k, an impossibility. When $m = \binom{k}{2}$ we require that each vertex has degree k or $k-1$. However, since in this case all degrees must sum to $k(k-1)$, G is necessarily K_k.

Finally, suppose $G \not\cong K_4$ has $\delta(G) \geq 1$, $e(G) = 6$ and $p_3(G) = 12$. From the above proof we need $\delta(G) = 2$ ($\delta(G) = 3$ would give $G \cong K_4$, while $\delta(G) = 1$ would give strict inequality in the final step). Moreover, supposing $d_1 = 2$, we require $e_1 = m - 2 = 4$ so that $c_1 = f_1 = 0$. This forces $|G| = 5$ and $G \cong G_1$. □

It is an almost trivial matter to get fairly good bounds on $p_s(m)$ when s is odd, say $s = 2r+1$. For suppose G is a graph of size m. Let A_r be the set of all paths of length

$2r+1$ in G with a distinguished "initial" vertex. For simplicity, we call a path with a distinguished initial vertex a *directed* path. As a path can be directed in two different ways, $|A_r| = 2p_{2r+1}(G)$. Let B_r be the set of all ordered $(r+1)$-tuples of vertex disjoint directed edges (that is, paths of length 1) in G. Then $|B_r| \leq 2^{r+1}(m)_{r+1}$. There is an injection from A_r into B_r, mapping a path $v_1v_2\ldots v_{2r+2} \in A_r$ to an $(r+1)$-tuple $(v_1v_2, v_3v_4, \ldots, v_{2r+1}v_{2r+2}) \in B_r$. Therefore

$$p_{2r+1}(G) = \frac{1}{2}|A_r| \leq \frac{1}{2}|B_r| \leq 2^r(m)_{r+1} \sim 2^r m^{r+1}.$$

When $m = \binom{k}{2}$, we have $e(K_k) = m$ and

$$p_{2r+1}(K_k) = \frac{1}{2}(k)_{2r+2} = m(k-2)_{2r} \sim 2^r m^{r+1},$$

so that our bound gives the correct highest order term. Instead of glueing together directed edges, we can use $(r+1)$ paths of length 3 or 1 edge and r paths of length 3 to obtain the following slightly improved result.

Theorem 3. *Let r be a positive integer, and let $m = \binom{k}{2}$ for some $k \geq 4r+4$. Then*

$$\frac{1}{2}(k)_{4r+2} \leq p_{4r+1}(m) \leq 2^r m(p_3(m))_r = \frac{1}{2}k(k-1)((k)_4)_r$$

and

$$\frac{1}{2}(k)_{4r+4} \leq p_{4r+3}(m) \leq 2^r(p_3(m))_{r+1} = \frac{1}{2}((k)_4)_{r+1}.$$

Proof. Let G be a graph of size $m = \binom{k}{2}$. Let C_r be the set of all directed paths of length $4r+1$ in G and let D_r be the set of all directed paths of length $4r+3$ in G. Let E_r be the set of all ordered pairs (x,y), where x is an ordered r-tuple of vertex disjoint directed paths of length 3 in G and y is a directed edge, disjoint from all the paths in x. Finally, let F_r be the set of all $(r+1)$-tuples of vertex disjoint directed paths of length 3 in G. We have injections

$$i_1 : C_r \longrightarrow E_r$$

$$i_2 : D_r \longrightarrow F_r$$

given by

$$i_1(v_1v_2\ldots v_{4r+2}) = ((v_1v_2v_3v_4, v_5v_6v_7v_8, \ldots, v_{4r-3}v_{4r-2}v_{4r-1}v_{4r}), v_{4r+1}v_{4r+2})$$

and

$$i_2(v_1v_2\ldots v_{4r+4}) = (v_1v_2v_3v_4, v_5v_6v_7v_8, \ldots, v_{4r+1}v_{4r+2}v_{4r+3}v_{4r+4}).$$

Therefore

$$p_{4r+1}(G) = \frac{1}{2}|C_r| \leq \frac{1}{2}|E_r| \leq 2^r m(p_3(G))_r \leq 2^r m(p_3(m))_r$$

and

$$p_{4r+3}(G) = \frac{1}{2}|D_r| \leq \frac{1}{2}|F_r| \leq 2^r(p_3(G))_{r+1} \leq 2^r(p_3(m))_{r+1}.$$

The lower bounds arise from the graph K_k. □

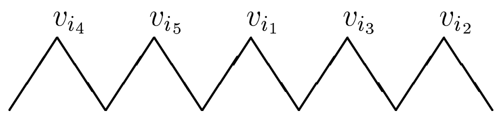

Figure 2

Any exact result for paths of length $2r+1$, where $r \geq 2$, would lead to an improvement in the upper bounds for p_{2t+1}, for all $t \geq r$, following the method used in the proof of Theorem 3. Alon [2] uses a variant of the above argument in maximizing the number of copies of H in a graph of size m when H has a perfect matching.

4. Paths of length four

The situation for paths of length four, and indeed paths of even length, is somewhat different. Suppose we are trying to maximize $p_4(G)$ over graphs G of size m. For simplicity, we assume that $m = \binom{k}{2}$ (for some k) and also that m is even (so that $k \equiv 0, 1(4)$). While K_k has $\frac{k}{2}(k-1)(k-2)(k-3)(k-4) \sim 2\sqrt{2}m^{\frac{5}{2}}$ paths of length four, $K(\frac{m}{2}, 2)$, the complete bipartite graph with class sizes $\frac{m}{2}$ and 2, has $\frac{m}{2}(\frac{m}{2}-1)(\frac{m}{2}-2) \sim \frac{m^3}{8}$ such paths, since any ordered triple (w_1, w_2, w_3) of vertices from the large class determines a unique path $w_1 v_1 w_2 v_2 w_3$ in the graph. Similarly, when $m = \binom{k}{2}$ and l divides m, we have

$$e(K(\frac{m}{l}, l)) = e(K_k) = m,$$

$$p_{2l}(K(\frac{m}{l}, l)) = \frac{l!}{2}\left(\frac{m}{l}\right)_{l+1} \sim \sqrt{\frac{\pi}{2l}}e^{-l}m^{l+1}$$

and

$$p_{2l}(K_k) = \frac{1}{2}(k)_{2l+1} \sim 2^{l-1}\sqrt{2}m^{l+\frac{1}{2}}.$$

For paths of length four, Theorem 5 will show that $K(\frac{m}{2}, 2)$ is essentially best possible. The proof of this result is completely different from that of Theorem 2. We require a preliminary lemma which gives a bound in terms of the degree sequence.

Lemma 4. *Let G be a graph with degree sequence $d_1 \geq d_2 \geq \ldots \geq d_n > 0$. Then*

$$p_{2l}(G) \leq \frac{l!}{2} \sum_{i_1 < i_2 < \ldots < i_l} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2.$$

Proof. Let $V(G) = \{v_1, v_2, \ldots, v_n\}$, where $d(v_i) = d_i$. Any path in G of length $2l$ specifies a unique l-tuple $(i_1, i_2, \ldots, i_l)$, with $i_1 < i_2 < \ldots < i_l$, where $v_{i_1}, v_{i_2}, \ldots, v_{i_l}$ are, in some order, the vertices at an odd distance from either end of the path (see Figure 2).

It is easy to see that there are at most

$$\frac{l!}{2} d_{i_l}(d_{i_l}-1)(d_{i_{l-1}}-1)\ldots(d_{i_1}-1) < \frac{l!}{2} d_{i_l}^2 d_{i_{l-1}} d_{i_{l-2}} \ldots d_{i_2} d_{i_1}$$

paths corresponding to the l-tuple $(i_1, i_2, \ldots, i_l)$. □

Theorem 5. *For all m,*

$$p_4(m) = \frac{m^3}{8} + O(m^{\frac{8}{3}}).$$

Proof. By virtue of the example $K(\frac{m}{2}, 2)$, we need only show that

$$p_4(m) \leq \frac{m^3}{8} + O(m^{\frac{8}{3}}).$$

Let G be a graph of size m. As in the proof of Lemma 4, let $V(G) = \{v_1, v_2, \ldots, v_n\}$, where $d(v_i) = d_i$, and suppose that $d_1 \geq d_2 \geq \ldots \geq d_n > 0$. Write

$$\begin{aligned} S &= \{i \in [n] : d_i > m^\alpha\}, \\ T &= \{i \in [n] : d_i \leq m^\alpha\}, \\ W &= \{v_i : i \in S\}, \end{aligned}$$

for some $\frac{1}{2} < \alpha < 1$, so that W is the set of vertices of large degree. Now, applying Lemma 4,

$$\begin{aligned} p_4(G) &\leq \sum_{1 \leq i < j \leq n} d_i d_j^2 \\ &= \sum_{1 \leq i < j \leq n, j \in S} d_i d_j^2 + \sum_{1 \leq i < j \leq n, j \in T} d_i d_j^2 \\ &\leq \sum_{1 \leq i < j \leq n, j \in S} d_i d_j^2 + \sum_{i=1}^{n} d_i \sum_{j \in T} d_j^2 \\ &\leq \sum_{1 \leq i < j \leq n, j \in S} d_i d_j^2 + 2m \sum_{j \in T} d_j m^\alpha \\ &\leq \sum_{1 \leq i < j \leq n, j \in S} d_i d_j^2 + 4m^{2+\alpha}. \end{aligned}$$

There are less than $2m^{1-\alpha}$ vertices in W, and so they span less than $2m^{2-2\alpha}$ edges. Therefore

$$\sum_{i \in S} d_i \leq m + 2m^{2-2\alpha}.$$

Writing $\beta = 1 + 2m^{1-2\alpha}$, we have

$$\begin{aligned} p_4(G) &\leq \sum_{1 \leq i < j \leq n, j \in S} d_i d_j^2 + 4m^{2+\alpha} \\ &\leq \frac{1}{2} \sum_{i,j \in S, i \neq j} d_i d_j^2 + 4m^{2+\alpha} \\ &= \frac{1}{2} \left(\sum_{i \in S} d_i \sum_{j \in S} d_j^2 - \sum_{j \in S} d_j^3 \right) + 4m^{2+\alpha} \end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{2}\left(\beta m \sum_{j\in S} d_j^2 - \sum_{j\in S} d_j^3\right) + 4m^{2+\alpha} \\
&= \frac{1}{2}\sum_{j\in S} d_j^2(\beta m - d_j) + 4m^{2+\alpha} \\
&= \frac{1}{2}\sum_{j\in S} d_j\,\{d_j(\beta m - d_j)\} + 4m^{2+\alpha} \\
&\le \frac{1}{2}\left(\sum_{j\in S} d_j \left(\frac{\beta m}{2}\right)^2\right) + 4m^{2+\alpha} \\
&\le \left(\frac{\beta m}{2}\right)^3 + 4m^{2+\alpha} \\
&= \frac{m^3}{8}(1 + 6m^{1-2\alpha} + 12m^{2-4\alpha} + 8m^{3-6\alpha}) + 4m^{2+\alpha}.
\end{aligned}
$$

Taking $\alpha = \frac{2}{3}$, we obtain the desired result. □

The main idea in the proof of Theorem 5 is that there are very few paths $v_1v_2v_3v_4v_5$ where either $d(v_2)$ or $d(v_4)$ are small, and so when using the bound $p_4(G) \le \sum_{1\le i<j\le n} d_i d_j^2$, we need only consider terms for which d_i and d_j are both large. Since there can only be a few vertices of large degree, the sum of their degrees is roughly m, not $2m$: this is why the truncated bound is at most approximately $\frac{m^3}{8}$. Had we not ignored the terms with d_j small, we would have obtained the bound m^3, the same as that resulting from counting the number of ways of glueing together a directed path of length two with a directed edge.

5. Paths of even length

When s is even and greater than four, essentially the same argument goes through, with one or two additional complications. For $n \ge t \ge 1$, we define the function $F_{t,n}$ and the constant $\theta_{t,n}$ by the formulae:

$$F_{t,n}(x_1,\ldots,x_n) = \sum_{1\le i_1<i_2<\ldots<i_t\le n} x_{i_1}x_{i_2}\ldots x_{i_{t-1}}x_{i_t}^2$$

$$\theta_{t,n} = \max\{F_{t,n}(x_1,\ldots,x_n) : x_1 \ge x_2 \ge \ldots \ge x_n \ge 0, \sum_{i=1}^n x_i = 1\}.$$

The $\theta_{t,n}$ increase with n for fixed t and they are all bounded above by 1. We write $\theta_t = \sup_{n\ge t}\theta_{t,n}$. Use of Lemma 4 in bounding $p_{2l}(m)$ from above leads one to consider maximizing $F_{l,n}$ over the unit simplex. The main difficulty in determining $p_{2l}(m)$ for $l \ge 3$ is that while it is easy to show that $\theta_2 = \theta_{2,n} = F_{2,n}(\frac{1}{2},\frac{1}{2},0,0,\ldots,0) = \frac{1}{8}$ for $n \ge 2$ (this was essentially done in the proof of Theorem 5), evaluation of θ_l for $l \ge 3$ is slightly harder. Set $\phi_t = \frac{1}{2}t!\theta_t$.

First, we will give an upper bound for $p_{2l}(m)$ in terms of ϕ_l. We then turn our attention to the problem of calculating ϕ_l in terms of l. However, as regards lower bounds for $p_{2l}(m)$, the nature of the point $P_{l,n}$ on the unit simplex at which the maximum of $F_{l,n}$ is attained (for n large compared to l) is of crucial significance. In particular, an inspection of the proof of Lemma 4 shows the following: we cannot construct bipartite graphs which have asymptotically the same number of paths of length $2l$ as that given by the upper bound in Lemma 6 unless, for sufficiently large n, $P_{l,n} = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}, 0, 0, \ldots, 0)$ for some N depending on l. Fortunately, as Lemma 7 will show, $P_{l,n}$ is always of the above form, allowing us to show that our upper bounds are asymptotically best possible.

Lemma 6. *For $l \geq 2$,*

$$p_{2l}(m) \leq \phi_l m^{l+1} + O(m^{l+\frac{2}{3}}).$$

Proof. Let G be a graph of size m with $V(G) = \{v_1, v_2, \ldots, v_n\}$, $d(v_i) = d_i$, $d_1 \geq d_2 \geq \ldots \geq d_n > 0$ and set

$$\begin{aligned} S &= \{i \in [n] : d_i > m^{\frac{2}{3}}\}, \\ T &= \{i \in [n] : d_i \leq m^{\frac{2}{3}}\}, \\ W &= \{v_i : i \in S\}, \\ \beta &= 1 + 2m^{-\frac{1}{3}}. \end{aligned}$$

We proceed as before:

$$\begin{aligned} p_{2l}(G) &\leq \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 \\ &= \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n, i_l \in S} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 \\ &+ \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n, i_l \in T} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 \\ &\leq \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n, i_l \in S} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 \\ &+ \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_{l-1} \leq n} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} \sum_{i \in T} d_i^2 \\ &< \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n, i_l \in S} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 \\ &+ \frac{l!}{2} (d_1 + d_2 + \ldots + d_n)^{l-1} \sum_{i \in T} d_i m^{\frac{2}{3}} \\ &\leq \frac{l!}{2} \sum_{1 \leq i_1 < i_2 < \ldots < i_l \leq n, i_l \in S} d_{i_1} d_{i_2} \ldots d_{i_{l-1}} d_{i_l}^2 + 2^{l-1} l! m^{l+\frac{2}{3}} \end{aligned}$$

$$\begin{aligned} &\leq \frac{l!}{2}\beta^{l+1}m^{l+1}\theta_l + 2^{l-1}l!m^{l+\frac{2}{3}} \\ &= \phi_l\beta^{l+1}m^{l+1} + 2^{l-1}l!m^{l+\frac{2}{3}} \\ &= \phi_l m^{l+1} + O(m^{l+\frac{2}{3}}). \end{aligned}$$

For the final inequality, we again used the fact that the sum of the degrees of the vertices in W is at most βm. □

Our next goal is to determine θ_t (and so ϕ_t) for all t, and to show that for sufficiently large n, depending on t, $\theta_{t,n}$, the maximum of $F_{t,n}$ over the unit simplex, is attained at the point $(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}, 0, 0, \ldots, 0)$ for some N depending on t. For $t \geq 2$, $N \geq t$, $n \geq t$, $x \geq t$ write

$$\begin{aligned} a_{t,n} &= F_{t,n}\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \binom{n}{t}\frac{1}{n^{t+1}}, \\ c_{t,N} &= \max\{a_{t,n} : t \leq n \leq N\}, \\ c_t &= \max\{a_{t,n} : n \geq t\}, \\ f(x) &= \binom{x}{t}\frac{1}{x^{t+1}} = \frac{x(x-1)\ldots(x-t+1)}{t!}\frac{1}{x^{t+1}}, \\ g(x) &= \sum_{i=0}^{t-1}\frac{1}{x-i} - \frac{t+1}{x}, \\ h(x) &= xg(x) + 1 = \sum_{i=1}^{t-1}\frac{i}{x-i}. \end{aligned}$$

Observe that $h'(x) < 0$ on $[t, \infty)$, so that $h(x)$ and $g(x) = \frac{1}{x}(h(x)-1)$ are both decreasing on $[t, \infty)$. We have $f'(x) = f(x)g(x)$ and $g(x) = 0$ precisely when $h(x) = 1$. Now $h(t) \geq t - 1$, while $h(x) \to 0$ as $x \to \infty$, and so f has a unique maximum in $[t, \infty)$. The upshot of this is that for fixed t the $a_{t,n}$ increase up to a point and then decrease: we define $N_t = \min\{n : a_{t,n} = c_t\}$ (the minimum is taken over at most two values).

Lemma 7. *For $n \geq t \geq 2$,*

$$\theta_{t,n} = c_{t,n}.$$

In particular,

$$\theta_t = c_t.$$

Proof. For fixed t, we proceed by induction on $n \geq t$. Throughout, we suppose that $x_1 \geq x_2 \geq \ldots \geq x_n \geq 0$ and that $\sum_{i=1}^n x_i = 1$. Firstly,

$$F_{t,t}(x_1, x_2, \ldots, x_t) = x_1 x_2 \ldots x_{t-1} x_t^2 \leq \left(\frac{1 - x_t}{t-1}\right)^{t-1} x_t^2.$$

As $a(x) = x^2(1-x)^{t-1}$ is increasing on $[0, \frac{2}{t+1}]$ and $x_t \leq \frac{1}{t} < \frac{2}{t+1}$, the function $F_{t,t}$ is maximized at $x_1 = x_2 = \ldots = x_t = \frac{1}{t}$. Therefore $\theta_{t,t} = c_{t,t}$.

For the induction step, some more notation will be convenient. For $\alpha \geq t^2-1$, $t \geq 2$, $x \geq 0$, write

$$\begin{aligned}
A_t(\alpha,x) &= (1-x)^{t+1} + \alpha x^2 (1-x)^{t-1},\\
B_t(\alpha,x) &= (1+\alpha)(1+t)x^2 - 2(1+\alpha+t)x + (1+t),\\
R_1(\alpha,t) &= \frac{1}{t+1} + \frac{t}{t+1}\frac{1}{1+\alpha} - \frac{1}{t+1}\frac{\alpha}{1+\alpha}\sqrt{\frac{\alpha-(t^2-1)}{\alpha}},\\
R_2(\alpha,t) &= \frac{1}{t+1} + \frac{t}{t+1}\frac{1}{1+\alpha} + \frac{1}{t+1}\frac{\alpha}{1+\alpha}\sqrt{\frac{\alpha-(t^2-1)}{\alpha}}.
\end{aligned}$$

Note that $R_1(\alpha,t)$ and $R_2(\alpha,t)$ are the zeros of $B_t(\alpha,x)$, and that $B_t(\alpha,x)$ divides the derivative of $A_t(\alpha,x)$.

Now,

$$\begin{aligned}
F_{t,n}(x_1,x_2,\ldots,x_n) &= F_{t,n-1}(x_1,x_2,\ldots,x_{n-1})\\
&+ x_n^2\left(\sum_{1\leq i_1<i_2\ldots<i_{t-1}\leq n-1} x_{i_1}x_{i_2}\ldots x_{i_{t-1}}\right)\\
&\leq \theta_{t,n-1}(1-x_n)^{t+1} + x_n^2(1-x_n)^{t-1}\frac{\binom{n-1}{t-1}}{(n-1)^{t-1}}\\
&= c_{t,n-1}(1-x_n)^{t+1} + x_n^2(1-x_n)^{t-1}\frac{\binom{n-1}{t-1}}{(n-1)^{t-1}}\\
&= c_{t,n-1}A_t\left(\frac{\binom{n-1}{t-1}}{c_{t,n-1}(n-1)^{t-1}}, x_n\right).
\end{aligned}$$

Here, the first inequality holds by the definition of θ and by the inequality between the arithmetic and geometric means. We know that $x_n \leq \frac{1}{n}$, so our objective is to maximize

$$A_t\left(\frac{\binom{n-1}{t-1}}{c_{t,n-1}(n-1)^{t-1}}, x\right)$$

over $[0,\frac{1}{n}]$. We have

$$\frac{dA_t(\alpha,x)}{dx} = -(1-x)^{t-2}B_t(\alpha,x).$$

For $\alpha \geq t^2-1$, the derivative is zero when $x=1$, $R_1(\alpha,t)$ or $R_2(\alpha,t)$, while if $\alpha < t^2-1$ then $A_t(\alpha,t)$ is decreasing on $[0,1]$. Also, by the definition of R_1 and R_2, if $\alpha \geq t^2-1$ then

$$0 < R_1(\alpha,t) \leq R_2(\alpha,t) < 1$$

and

$$R_2(\alpha,t) > \frac{1}{t+1} \geq \frac{1}{n}$$

so that $A_t(\alpha,x)$ is decreasing on $[0,R_1(\alpha,t)]$, increasing on $[R_1(\alpha,t),R_2(\alpha,t)]$ and decreasing on $[R_2(\alpha,t),1]$. Thus,

$$A_t\left(\frac{\binom{n-1}{t-1}}{c_{t,n-1}(n-1)^{t-1}},x\right)$$

is maximized over $[0,\frac{1}{n}]$ at either 0 or $\frac{1}{n}$.

If $n-1\le N_t$, then $c_{t,n-1}=a_{t,n-1}$,

$$A_t\left(\frac{\binom{n-1}{t-1}}{c_{t,n-1}(n-1)^{t-1}},0\right)=A_t\left(\frac{\binom{n-1}{t-1}}{a_{t,n-1}(n-1)^{t-1}},0\right)=1$$

and

$$A_t\left(\frac{\binom{n-1}{t-1}}{c_{t,n-1}(n-1)^{t-1}},\frac{1}{n}\right)=A_t\left(\frac{\binom{n-1}{t-1}}{a_{t,n-1}(n-1)^{t-1}},\frac{1}{n}\right)=\frac{\binom{n}{t}\frac{1}{n^{t+1}}}{\binom{n-1}{t}\frac{1}{(n-1)^{t+1}}}.$$

Hence,

$$\theta_{t,n}=F_{t,n}\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right)=\binom{n}{t}\frac{1}{n^{t+1}}$$

if $n\le N_t$ and

$$\theta_{t,n}=F_{t,n}\left(\frac{1}{n-1},\frac{1}{n-1},\ldots,\frac{1}{n-1},0\right)=\binom{n-1}{t}\frac{1}{(n-1)^{t+1}}$$

if $n=N_t+1$.

If $n-1>N_t$, then by the induction hypothesis, $\theta_{t,n-1}=c_{t,n-1}=a_{t,N_t}=c_t$ and we claim that $\theta_{t,n}=c_t$ also holds. To see this, we have to show that

$$A_t\left(\frac{\binom{n-1}{t-1}}{c_t(n-1)^{t-1}},0\right)>A_t\left(\frac{\binom{n-1}{t-1}}{c_t(n-1)^{t-1}},\frac{1}{n}\right)$$

or that

$$1>\left(1-\frac{1}{n}\right)^{t+1}+\frac{\binom{n-1}{t-1}}{c_t(n-1)^{t-1}}\left(\frac{1}{n}\right)^2\left(1-\frac{1}{n}\right)^{t-1}$$

for $n\ge N_t+2$. We can rewrite this inequality as

$$n^{t+1}-(n-1)^{t+1}-\frac{1}{c_t}\binom{n-1}{t-1}>0,$$

that is

$$\left\{n^{t+1}-\frac{1}{c_t}\binom{n}{t}\right\}-\left\{(n-1)^{t+1}-\frac{1}{c_t}\binom{n-1}{t}\right\}>0.$$

Therefore we are done if we can show that

$$G(x)=x^{t+1}-\frac{1}{c_t}\binom{x}{t}$$

increases on $[N_t+1,\infty)$. Since

$$c_t=\frac{1}{N_t^{t+1}}\binom{N_t}{t},$$

we have

$$\begin{aligned} G(x) &= x^{t+1} - \frac{N_t^{t+1}\binom{x}{t}}{\binom{N_t}{t}} \\ &= \binom{x}{t}\left\{\frac{x^{t+1}}{\binom{x}{t}} - \frac{N_t^{t+1}}{\binom{N_t}{t}}\right\} \\ &= \binom{x}{t}\left\{\frac{1}{f(x)} - \frac{N_t^{t+1}}{\binom{N_t}{t}}\right\}, \end{aligned}$$

which is increasing on $[N_t + 1, \infty)$ because $f(x)$ is decreasing on $[N_t + 1, \infty)$. Therefore $\theta_{t,n} = c_t$ as required. □

The $F_{t,n}$ closely resemble the family $A_{t,n}$, defined by

$$A_{t,n} = \frac{1}{(t-1)!(n-t)!} \sum_{\pi \in S_n} x_{\pi(1)}^2 x_{\pi(2)} \dots x_{\pi(t)}.$$

(The factor $\frac{1}{(t-1)!(n-t)!}$ is included to ensure that each term appears with multiplicity one.) The $A_{t,n}$ are examples of (scaled) Muirhead means of degree $t+1$ (see e.g. [7]). The extreme n-variable Muirhead means of degree $t+1$ are

$$B_{t,n} = \frac{1}{(t+1)!(n-t-1)!} \sum_{\pi \in S_n} x_{\pi(1)} x_{\pi(2)} \dots x_{\pi(t+1)}$$

and

$$C_{t,n} = \sum_{i=1}^{n} x_i^{t+1};$$

A, B and C are of type $(2,1,1,\dots,1,0,0,\dots,0)$, $(1,1,\dots,1,0,0,\dots,0)$ and $(t+1,0,0,\dots,0)$ respectively. On the simplex $S(n) = \{(x_1, x_2, \dots, x_n) : x_i \geq 0, \sum_{i=1}^n x_i = 1\}$, B is maximized by setting all the x_i equal and C is maximized by setting one of the x_i equal to 1 and the rest equal to 0. For large n, the form of $A_{t,n}$ approaches that of $B_{t,n}$ with increasing t, and accordingly it is plausible that the extreme points of $A_{t,n}$ on $S(n)$ approach the centre of the simplex which is the unique extreme point of $B_{t,n}$.

Muirhead [11] gave necessary and sufficient conditions for two appropriately scaled n-variable Muirhead means of the same degree to be comparable - the type-sequence of one must be majorized by the type-sequence of the other.

Returning to the problem about paths, in view of Lemma 7, we can construct graphs that show that the bound in Lemma 6 is essentially best possible.

Theorem 8. *For $l \geq 2$,*

$$p_{2l}(m) = \frac{1}{2} l! c_l m^{l+1} + O(m^{l+\frac{2}{3}}).$$

Proof. Let $G_{l,m}$ be the vertex disjoint union of the complete bipartite graph with class sizes N_l and $\left\lfloor \frac{m}{N_l} \right\rfloor$ and $m - N_l \left\lfloor \frac{m}{N_l} \right\rfloor$ isolated edges. Then

$$\begin{aligned} p_{2l}(G_{l,m}) &= \frac{(N_l)_l}{2} \left(\left\lfloor \frac{m}{N_l} \right\rfloor \right)_{l+1} \\ &= \frac{1}{2} l! \binom{N_l}{l} (\frac{m}{N_l})^{l+1} + O(m^l) \\ &= \frac{1}{2} l! c_l m^{l+1} + O(m^l). \end{aligned}$$

Therefore

$$p_{2l}(m) \geq \frac{1}{2} l! c_l m^{l+1} + O(m^l).$$

For the other direction, apply Lemmas 6 and 7. □

It is easy to give an explicit form for N_l and c_l; however, as the calculations are cumbersome, we give a detailed proof.

Lemma 9. *For $l \geq 2$,*

$$N_l = \left\lfloor \frac{9l^2 + 3l + 5}{18} \right\rfloor.$$

Proof. It is easy to check that $N_2 = 2$, $N_3 = 5$, $N_4 = 8$ and $N_5 = 13$. For $n \geq 3l \geq 18$ we have

$$r_{l,n} = \frac{a_{l,n+1}}{a_{l,n}} = \frac{\binom{n+1}{l} \frac{1}{(n+1)^{l+1}}}{\binom{n}{l} \frac{1}{n^{l+1}}} = \left(1 + \frac{l}{n - l + 1}\right) \left(1 - \frac{1}{n+1}\right)^{l+1}$$

and

$$\log r_{l,n} = -\sum_{j=1}^{\infty} \frac{(-1)^j l^j}{j(n-l+1)^j} - \sum_{j=1}^{\infty} \frac{(l+1)}{j(n+1)^j}.$$

We already know that, for fixed l, the sequence $a_{l,n}$ is strictly increasing for $n \leq N_l$ and strictly decreasing for $n \geq N_{l+1}$. We defined N_l so that $a_{l,N_l} \geq a_{l,N_{l+1}}$, although it is still possible that $a_{l,N_l} = a_{l,N_{l+1}}$. This much follows from the discussion preceding Lemma 7. The following more detailed analysis will show that in fact we always have $a_{l,N_l} > a_{l,N_{l+1}}$, since the function $\log r_{l,n}$ (considered as a continuous function of n) changes sign (strictly) between $n = N_{l-1}$ and N_l. In addition, we derive a formula for N_l. In order to do this, we need to focus attention on the values of n lying in an interval of length $\frac{l}{3}$. Specifically, take $n + 1 = \frac{l^2}{2} + \frac{l}{6} + \alpha$, where $|\alpha| \leq \frac{l}{6}$, so that

$$n - l + 1 \geq \frac{l^2}{3}$$

and

$$n + 1 \geq \frac{l^2}{2} > \frac{l}{3}(l+1)$$

giving

$$\begin{aligned}\left|\log r_{l,n}+\sum_{j=1}^{4}\frac{(-1)^j l^j}{j(n-l+1)^j}+\sum_{j=1}^{2}\frac{(l+1)}{j(n+1)^j}\right| &\leq \frac{1}{5}\sum_{j=5}^{\infty}\frac{l^j}{(n-l+1)^j}+\frac{1}{3}\sum_{j=3}^{\infty}\frac{(l+1)}{(n+1)^j}\\ &< \frac{1}{5}\sum_{j=5}^{\infty}\left(\frac{3}{l}\right)^j+\frac{l}{3}\sum_{j=3}^{\infty}\left(\frac{3}{l^2}\right)^j\\ &< \frac{108}{l^5}.\end{aligned}$$

It is easily checked that, again with $n+1=\frac{l^2}{2}+\frac{l}{6}+\alpha$,

$$\left|\sum_{j=1}^{4}\frac{(-1)^j l^j}{j(n-l+1)^j}+\sum_{j=1}^{2}\frac{(l+1)}{j(n+1)^j}+\frac{2(5-18\alpha)}{9l^4}\right|<\frac{42}{l^5}.$$

Therefore,

$$\log r_{l,\frac{l^2}{2}+\frac{l}{6}+\alpha}=\frac{2(5-18\alpha)}{9l^4}+E(l,\alpha)$$

where

$$E(l,\alpha)<\frac{150}{l^5}.$$

For integral l, the expression $\frac{l^2}{2}+\frac{l}{6}+\frac{5}{18}$ is never an integer, and it differs from the nearest integer by at least $\frac{1}{18}$. So, for $l\geq 675$, we have

$$\log r_{l,\left\lfloor\frac{l^2}{2}+\frac{l}{6}+\frac{5}{18}\right\rfloor}>0$$

and

$$\log r_{l,\left\lfloor\frac{l^2}{2}+\frac{l}{6}+\frac{5}{18}\right\rfloor+1}<0,$$

which implies that $a_{l,n}$ increases from $n=\left\lfloor\frac{l^2}{2}+\frac{l}{6}+\frac{5}{18}\right\rfloor-1$ to $n=\left\lfloor\frac{l^2}{2}+\frac{l}{6}+\frac{5}{18}\right\rfloor$ and decreases thereafter. For smaller values of l, the assertion is readily checked on a computer. □

Putting the pieces together, we have the following result.

Theorem 10. *For $l\geq 2$,*

$$p_{2l}(m)=C_l m^{l+1}+O(m^{l+\frac{2}{3}}),$$

where

$$C_l=\frac{\left(\left\lfloor\frac{9l^2+3l+5}{18}\right\rfloor\right)_l}{2\left(\left\lfloor\frac{9l^2+3l+5}{18}\right\rfloor\right)^{l+1}}.$$

Proof. This is immediate from Theorem 8 and Lemma 9. □

It may be of interest to have an asymptotic formula for C_l itself. This is again easy to obtain, for we have

$$\begin{aligned} C_l &\sim \frac{1}{2}\left(\frac{2}{l^2}\right)^{l+1}\left(\frac{l^2}{2}\right)_l = \frac{1}{l^2}\prod_{i=1}^{l-1}\left(1-\frac{2i}{l^2}\right) \\ &= \frac{1}{l^2}\prod_{i=1}^{l-1} e^{-2i/l^2}(1+O(l^{-2})) = \frac{1}{el^2}e^{1/l}(1+O(l^{-1})) \sim \frac{1}{el^2}. \end{aligned}$$

We suspect that the complete graphs (in the case of odd s), and the bipartite graphs $K(N_{s/2}, \frac{m}{N_{s/2}})$ (in the case of even s) are the extremal graphs with exactly $p_s(m)$ paths of length s for m with suitable divisibility properties. However, obtaining exact rather than just asymptotic results seems to require a much more detailed analysis.

References

[1] Ahlswede, R. and Katona, G.O.H., Graphs with maximal number of adjacent pairs of edges, *Acta Math. Acad. Sci. Hungar.* **32** (1978), 97-120.

[2] Alon, N., On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math* **38** (1981), 116-130.

[3] Alon, N., On the number of certain subgraphs contained in graphs with a given number of edges, *Israel J. Math* **53** (1986), 97-120.

[4] Bollobás, B., *Combinatorics*, Cambridge University Press, 1986.

[5] Bollobás, B. and Erdős, P., Graphs of extremal weights, *Ars Combinatorica* **50** (1998), 225-233.

[6] Füredi, Z., Graphs with maximum number of star-forests, *Studia Sci. Math. Hungar.* **27** (1992), 403-407.

[7] Hardy, G.H., Littlewood, J.E. and Pólya, G., *Inequalities*, Cambridge University Press, 1951.

[8] Katona, G.O.H., A theorem on finite sets, in *Theory of Graphs* (Erdős, P. and Katona, G.O.H., eds.), Akadémiai Kiadó, Budapest, 187-207.

[9] Katz, M., Rearrangements of (0-1) matrices, *Israel Jour. Math.* **9** (1971), 53-71.

[10] Kruskal, J.B., The number of simplices in a complex, in *Mathematical Optimization Techniques*, Univ. California Press, Berkeley, 251-278.

[11] Muirhead, R.F., Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters, *Proc. Edinburgh Math. Soc.* **21** (1903), 144-157.