Paths of Length Four

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 FINAL FORM 30 October 2002

For each sufficiently large m, we determine the unique graph of size m with the maximum number of paths of length four. If m is even, this is the complete bipartite graph $K(\frac{m}{2}, 2)$.

Given a graph G and an integer $s \geq 2$, write $p_s(G)$ for the number of paths of length s in G. The asymptotic behaviour of the function

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

was determined in [5], where it was also shown that if $10 \leq \binom{k}{2} \leq m < \binom{k+1}{2}$ then

$$p_3(m) \leq \frac{2m(m-k)(k-2)}{k},$$

with equality if and only if $m = \binom{k}{2}$. The study of $p_3(m)$ was motivated by the results on weights of graphs in [4], and the problem of maximizing $p_2(G)$ over graphs of fixed order and size was considered in [1] and implicitly in [7]. Moreover, after the first version of this paper was written, we discovered three papers [2], [3], [6] also concerned with maximizing the number of subgraphs isomorphic to a fixed graph H in graphs of size m. Although the results in [5] (and, to a lesser extent, those in [2]) give us some information about the extremal graphs themselves, they do not tell us what they are. In particular, for large values of s we know only large families of graphs which are close to being extremal. However, in this paper we determine the unique extremal graph for $s = 4$ and every sufficiently large m.

The first stage of the proof is essentially contained in [5]: we reproduce it here for the sake of completeness.

For m even and at least two, let G_m be the complete bipartite graph $K(\frac{m}{2}, 2)$. If m is odd and at least three, we take G_m to be the complete bipartite graph $K(\frac{m-1}{2}, 2)$ with an additional edge connecting one of the vertices of degree $\frac{m-1}{2}$ to a new vertex. It turns out that G_m has many more paths of length four than a complete graph of approximately the same size. In the proof below, as throughout the paper, G will be a graph of size m with no isolated vertices.

Theorem 1. *If m is sufficiently large then*

$$p_4(m) = p_4(G_m) = \begin{cases} \frac{m^3}{8} - \frac{3m^2}{4} + m, & \text{if } m \text{ is even;} \\ \frac{m^3}{8} - \frac{7m^2}{8} + \frac{15m}{8} - \frac{9}{8}, & \text{if } m \text{ is odd,} \end{cases}$$

and G_m is the unique extremal graph.

Proof. Let G be a graph of size m with $V(G) = \{v_1, v_2, \dots, v_n\}$ and suppose that $d_1 \geq d_2 \geq \dots d_n > 0$, where $d_i = d(v_i)$. For a pair of vertices $\{v_i, v_j\}$, set

$$d_{ij} = |\Gamma(v_i) - \{v_j\}|,$$

and

$$f_{ij} = |\Gamma(v_i) \cap \Gamma(v_j)|.$$

Then, indexing a path of length four in G by the two vertices v_i, v_j adjacent to its middle vertex, we have

$$\begin{aligned} p_4(G) &= \sum_{i<j} \{(d_{ij} - f_{ij}) f_{ij} (d_{ji} - 1) + f_{ij}(f_{ij} - 1)(d_{ji} - 2)\} \\ &= \sum_{i<j} f_{ij} \{(d_{ij} - 1)(d_{ji} - 1) - (f_{ij} - 1)\}. \end{aligned}$$

Therefore certainly

$$p_4(G) \leq \sum_{i<j} d_i d_j^2,$$

and this is the first approximation we shall use.

We can immediately get the correct order of magnitude for $p_4(m)$, since we may write

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

However, a little more care gives the correct constant also. Indeed, the terms $d_i d_j^2$ where d_j is small contribute very little to the above sum, and so we can restrict attention to the "large degree terms", where d_i and d_j are both large. The sum of the large degrees is approximately m, so the following two-stage calculation saves us a factor of 8. Write

$$S = \{i \in [n] : d_i > m^{\frac{2}{3}}\},$$

$$T = \{i \in [n] : d_i \leq m^{\frac{2}{3}}\},$$

$$W = \{v_i : i \in S\}.$$

There are less than $2m^{\frac{1}{3}}$ vertices in W, and so they span less than $2m^{\frac{2}{3}}$ edges. This means that

$$\sum_{i \in S} d_i \leq m + 2m^{\frac{2}{3}} = \beta m,$$

where $\beta = 1 + 2m^{-\frac{1}{3}}$. Now,

$$p_4(G) \leq \sum_{i<j} d_i d_j^2 = \sum_{i<j, j \in S} d_i d_j^2 + \sum_{i<j, j \in T} d_i d_j^2 = s(G) + t(G), \tag{1}$$

say. We shall bound $s(G)$ and $t(G)$ separately. Clearly,

$$t(G) \leq \sum_{i=1}^{n} d_i \sum_{j \in T} d_j^2 \leq 2m \sum_{j \in T} d_j m^{\frac{2}{3}} \leq 4m^{\frac{8}{3}}, \tag{2}$$

and further

$$\begin{aligned} s(G) &\leq \frac{1}{2} \sum_{i,j \in S, i \neq j} d_i d_j^2 = \frac{1}{2} \left(\sum_{i \in S} d_i \sum_{j \in S} d_j^2 - \sum_{j \in S} d_j^3 \right) \leq \frac{1}{2} \left(\beta m \sum_{j \in S} d_j^2 - \sum_{j \in S} d_j^3 \right) \\ &= \frac{1}{2} \sum_{j \in S} d_j^2 (\beta m - d_j) \leq \frac{1}{2} \sum_{j \in S} d_j \left(\frac{\beta m}{2} \right)^2 \leq \left(\frac{\beta m}{2} \right)^3 \\ &= \frac{m^3}{8} + \frac{3}{4} m^{\frac{8}{3}} + \frac{3}{2} m^{\frac{7}{3}} + m^2 \leq \frac{m^3}{8} + m^{\frac{8}{3}} \end{aligned}$$

for $m \geq 1000$. Together with (2) this gives

$$p_4(G) \leq \frac{m^3}{8} + 5m^{\frac{8}{3}}$$

for $m \geq 1000$.

The next step is to show that unless d_1 and d_2 are both very close to $\frac{m}{2}$, G will contain far fewer than $\frac{m^3}{8}$ paths of length four.

If there are less than two vertices in W, then there are no more than $4m^{\frac{8}{3}}$ paths of length four in G, and

$$4m^{\frac{8}{3}} < \frac{m^3}{8} - m^2 < p_4(G_m)$$

for $m \geq 64000$. Therefore we may suppose that $d_1 \geq d_2 > m^{\frac{2}{3}}$.

From now on, we shall assume that $m \geq 16^{21}$ and also that $p_4(G) \geq p_4(G_m)$; in particular

$$p_4(G) > \frac{m^3}{8} - m^2,$$

and, using (1) and (2),

$$s(G) > \frac{m^3}{8} - 4m^{\frac{8}{3}} - m^2. \tag{3}$$

Our next aim is to prove that

$$d_2 > \frac{\beta m}{2} - 4m^{\frac{6}{7}}. \tag{4}$$

We shall do this in three stages.

First, we require

$$\frac{\beta m}{2} - m^{\frac{6}{7}} < d_1 < \frac{\beta m}{2} + m^{\frac{6}{7}}. \tag{5}$$

If d_1 is out of this range, then so are all the other degrees in G. For if $d_1 \geq \frac{\beta m}{2} + m^{\frac{6}{7}}$ then for $2 \leq j \leq n$ we have

$$d_j \leq d_2 \leq m - d_1 + 1 < \beta m - d_1 \leq \frac{\beta m}{2} - m^{\frac{6}{7}}.$$

Therefore, if

$$d_1 \notin \left(\frac{\beta m}{2} - m^{\frac{6}{7}}, \frac{\beta m}{2} + m^{\frac{6}{7}}\right)$$

then

$$\begin{aligned}
s(G) &\leq \frac{1}{2}\sum_{j\in S} d_j^2(\beta m - d_j) \\
&\leq \frac{1}{2}\sum_{j\in S} d_j \left(\frac{\beta m}{2} - m^{\frac{6}{7}}\right)\left(\frac{\beta m}{2} + m^{\frac{6}{7}}\right) \\
&\leq \frac{1}{8}(m + 2m^{\frac{2}{3}})(m + 2m^{\frac{2}{3}} - 2m^{\frac{6}{7}})(m + 2m^{\frac{2}{3}} + 2m^{\frac{6}{7}}) \\
&\leq \frac{m^3}{8} - \frac{m^{\frac{19}{7}}}{2} + m^{\frac{8}{3}} \\
&< \frac{m^3}{8} - 4m^{\frac{8}{3}} - m^2,
\end{aligned}$$

contradicting (3).

Second, we must have $d_2 > m^{\frac{6}{7}}$ since otherwise

$$\begin{aligned}
s(G) &\leq \frac{1}{2}\sum_{j\in S} d_j^2(\beta m - d_j) \\
&= \frac{1}{2}d_1^2(\beta m - d_1) + \frac{1}{2}\sum_{j\in S, j\geq 2} d_j^2(\beta m - d_j) \\
&< \frac{1}{2}\left(\frac{\beta m}{2} + m^{\frac{6}{7}}\right)^2\left(\frac{\beta m}{2} - m^{\frac{6}{7}}\right) + \frac{1}{2}\sum_{j\in S, j\geq 2} d_j^2(\beta m - d_j) \\
&< \frac{1}{2}\left(\frac{\beta m}{2} + m^{\frac{6}{7}}\right)\left(\frac{\beta m}{2}\right)^2 + \frac{\beta^2}{2}m^{\frac{20}{7}} \\
&= \frac{1}{2}\left(\frac{\beta m}{2}\right)^3 + \frac{5\beta^2}{8}m^{\frac{20}{7}}
\end{aligned}$$

$$< \frac{m^3}{15},$$

which contradicts (3).

Third, if $d_1 - d_2 \geq 3m^{\frac{6}{7}}$ then

$$\begin{aligned}
s(G) &= \sum_{i<j, j\in S} d_i d_j^2 \\
&\leq \frac{1}{2}\sum_{i,j\in S, i\neq j} d_i d_j^2 + \frac{1}{2}(d_1 d_2^2 - d_1^2 d_2) \\
&\leq \frac{m^3}{8} + m^{\frac{8}{3}} - \frac{1}{2} d_1 d_2 (d_1 - d_2) \\
&\leq \frac{m^3}{8} + m^{\frac{8}{3}} - \frac{1}{2}\frac{m}{3} m^{\frac{6}{7}}(3m^{\frac{6}{7}}) \\
&< \frac{m^3}{8} - 4m^{\frac{8}{3}} - m^2,
\end{aligned}$$

contradicting (3), as before, and so proving (4).

Define $l = l(G)$ by

$$d_{12} + d_{21} = m - l.$$

Inequalities (4) and (5) imply that $l \leq 5m^{\frac{6}{7}}$. Our aim is to show that in fact $l = 0$. Once we have established this, the remainder of the proof will be easy. Indeed, if $l = 0$ then

$$\begin{aligned}
p_4(G) &= f_{12}\{(d_{12}-1)(d_{21}-1) - (f_{12}-1)\} \\
&\leq d_{21}\{(d_{12}-1)(d_{21}-1) - (d_{21}-1)\} \\
&= d_{21}(d_{21}-1)(d_{12}-2),
\end{aligned}$$

and we maximize this last function over $d_{12} + d_{21} = m$, $d_{12} \geq d_{21}$ by making d_{12} and d_{21} as equal as possible, so that $G \cong G_m$. Suppose then that $l > 0$.

Call a path of length four in G *regular* if it is of the form xv_1yv_2z, and *irregular* otherwise. Write $p_4(G, v_1, v_2)$ for the number of regular paths and $q_4(G, v_1, v_2)$ for the number of irregular ones, so that $p_4(G) = p_4(G, v_1, v_2) + q_4(G, v_1, v_2)$. Clearly, $p_4(G, v_1, v_2) \leq p_4(G_{m-l})$ so that

$$\begin{aligned}
p_4(G_m) - p_4(G) &= p_4(G_m) - p_4(G, v_1, v_2) - q_4(G, v_1, v_2) \\
&\geq p_4(G_m) - p_4(G_{m-l}) - q_4(G, v_1, v_2).
\end{aligned}$$

Hence, assuming $d_{12} = d_{21}$ and $l \equiv 0(2)$ for simplicity, our task amounts to showing that when we replace l edges from $G - \{v_1, v_2\}$ by $\frac{l}{2}$ edges incident with v_1 and $\frac{l}{2}$ edges incident with v_2 our gain in regular paths is more than our loss in irregular ones.

If m is even, then

$$p_4(G_m) - p_4(G_{m-1}) = \frac{(m-2)(m-4)}{2}$$

and

$$p_4(G_m) - p_4(G_{m-1}) = \frac{(m-1)(m-3)}{4}$$

if m is odd. Further, if $m \geq 16^{21}$ and $l_1 \leq l \leq 5m^{\frac{6}{7}}$ we have

$$p_4(G_{m-l_1+1}) - p_4(G_{m-l_1}) \geq \frac{m^2}{5}.$$

Summing, we obtain

$$p_4(G_m) - p_4(G_{m-l}) \geq \frac{lm^2}{5}.$$

It remains to show that if $l > 0$ then

$$q_4(G, v_1, v_2) < \frac{lm^2}{5}.$$

In what follows, x, y, z, w and v will denote vertices chosen from $V(G) - \{v_1, v_2\}$. There are at most l^2 paths of the form v_1v_2xyz, l^2 of the form v_2v_1xyz, ml of each of the types v_1xv_2yz and v_2xv_1yz, $2lm$ of either of the types v_1xyv_2z and v_2xyv_1z, l^2 of the form v_1xyzv_2, $2lm$ of either of the types xv_1v_2yz and xv_2v_1yz, $4l^2$ of each of the types v_1xyzw and v_2xyzw, l^2m of either of the types xv_1yzw and xv_2yzw, $4l^2$ of each of the types xyv_1zw and xyv_2zw and l^3 of the form $xyzwv$. For example, when counting the number of paths of type v_1v_2xyz, there are at most l choices for the edge $e = xy$, at most $l-1$ choices of a fifth vertex z (adjacent to either one of the endvertices of e) and the choice of z determines which endvertex of e is joined to v_2. Thus

$$q_4(G, v_1, v_2) \leq l^3 + 19l^2 + ml^2 + 6ml.$$

Each of the four terms on the right hand side is strictly less than $\frac{lm^2}{20}$, so finally

$$p_4(G) < p_4(G_m),$$

a contradiction. $\square$

Although we suspect that the conclusion of the theorem holds for much smaller values of m than above, it is certainly false for $m = 4, 5$ and 6: G_4, G_5 and G_6 have fewer P_4s than a path of length four, a cycle of order five and a cycle of order five with a pendant edge, respectively.

There are various obstacles to extending the theorem to paths of different lengths. When we deal with paths of odd length, complete graphs are asymptotically extremal [2], [5], and entirely different methods are used. (As it happens, the problem is easier for paths of odd length.) For paths of length six, the complete bipartite graphs with *five* vertices in one class (and so about $\frac{m}{5}$ in the other) are asymptotically extremal, and it therefore seems likely that new ideas are needed to give exact results for paths of six and higher even orders.

References

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

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

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

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

[5] B. Bollobás and A. Sarkar, Paths in graphs, Studia Sci. Math. Hungar. **38** (2001) 115-137.

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

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