Maximum Acyclic and Fragmented Sets in Regular Graphs

Penny Haxell*
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Oleg Pikhurko†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
United States
Web: http://www.math.cmu.edu/~pikhurko

Andrew Thomason
Centre for Mathematical Sciences
Cambridge University
Cambridge CB3 0WB
United Kingdom

October 4, 2006

Abstract

We show that a typical d-regular graph G of order n does not contain an induced forest with around $\frac{2\ln d}{d}n$ vertices, when $n \gg d \gg 1$, this bound being best possible because of a result of Frieze and Łuczak [6]. We then deduce an affirmative answer to an open question of Edwards and Farr [5] about fragmentability, which concerns large subgraphs with components of bounded size. An alternative, direct answer to the question is also given.

1 Introduction

This note is concerned with subgraphs of regular graphs that are either acyclic or have only small components. We begin by stating some appropriate graph parameters that have appeared in different places in the literature, and showing how they are related.

*Partially supported by NSERC.
†Partially supported by NSF Grant DMS-0457512.

First, with regard to acyclic subgraphs, the *decycling number* $\phi(G)$ of a graph G is the smallest number of vertices that can be removed from G to make it acyclic. Thus $v(G) - \phi(G)$ is the largest order of an induced forest in G. The problem of deciding whether $\phi(G) \leq k$ was shown to be NP-complete by Karp [8].

Second, with regard to small components, given an integer f, we say that a graph is *f-fragmented* if each of its components has at most f vertices. More generally, given a real number $\alpha > 0$, we say that G is *(α, f)-fragmentable* if it has a set of at most $\alpha v(G)$ vertices whose removal results in an f-fragmented graph.

Our interest in these parameters was prompted by a study of graph fragmentability, and in particular the parameter α_d defined by Edwards and Farr [5] as follows. For an integer d, α_d is the infimum of α for which there is an f such that any graph with maximum degree at most d is (α, f)-fragmentable. In other words, if $\alpha > \alpha_d$ then there is some finite f, such that every graph G of maximum degree d has a set of at most $\alpha v(G)$ vertices, whose removal leaves components with at most f vertices. It is easy to see that the value of α_d would not change if we consider d-regular graphs only.

Note that, trivially, $\alpha_1 = \alpha_2 = 0$. Edwards and Farr [5] proved that

$$\frac{d-2}{2d-2} \leq \alpha_d \leq \frac{d-2}{d+1} = 1 - \frac{3}{d+1}, \quad \text{for any } d \geq 3. \tag{1}$$

In particular, $\alpha_3 = \frac{1}{4}$. More recently, Edwards and Farr (personal communication) reported an improvement in the upper bound in (1). It is not clear at present what the best result given by their method is, though they can definitely show that

$$\alpha_4 \leq \frac{3}{8}. \tag{2}$$

On the other hand, Edwards and Farr [4, Problem 408] asked whether

$$\lim_{d\to\infty} \alpha_d = 1. \tag{3}$$

In this note we give an affirmative answer to this question, by two different methods. Observe that, to answer the question, we must exhibit graphs for which it is necessary to delete nearly all the vertices before the remainder is fragmented — that is, we must find graphs without any large fragmented subgraph.

In the first place, it is natural to consider what bounds are given by random regular graphs. This is where the decycling number comes in, because decycling and fragmenting are more or less the same for these graphs (see the simple Lemma 2 below). The decycling number of random d-regular graphs with d small has been studied by Bau, Wormald, and Zhou [2].

For $0 \leq d \leq n-1$ with nd even let $G_{n,d}$ be a graph selected uniformly at random from all labelled d-regular graphs of order n. Let $\phi_d(n)$ be the smallest ϕ such that with probability at least 1/2 we have $\phi(G_{n,d}) \le \phi$. (Here, as we shall see later, the value 1/2 can be changed to any constant without changing the subsequent results.)

Clearly, $\phi_d(n) = o(n)$ if $d \le 2$. However, we do not know how to prove that $\phi_d(n)/n$ tends to a limit for $d \ge 3$ as $n \to \infty$, so we have to define

$$\begin{aligned} b(d) &= \liminf_{n\to\infty} \frac{\phi_d(n)}{n}, \\ B(d) &= \limsup_{n\to\infty} \frac{\phi_d(n)}{n}. \end{aligned}$$

Bau, Wormald, and Zhou [2] proved that *whp* (i.e., with probability $1 - o(1)$ as $n \to \infty$) we have $\phi(G_{n,3}) = \lceil \frac{n}{4} + \frac{1}{2} \rceil$ (and thus $b(3) = B(3) = \frac{1}{4}$) and they presented methods for obtaining bounds on $b(d)$ and $B(d)$ for any given d. Their techniques give the following bounds (for some small d):

$$\begin{array}{ccccccc} 1/3 & \le & b(4) & \le & B(4) & \le & 0.3787, \\ 0.3786 & \le & b(5) & \le & B(5) & \le & 0.4512, \\ 0.4232 & \le & b(6) & \le & B(6) & \le & 0.5043. \end{array} \tag{4}$$

Here we are interested in the asymptotic behavior for large d, and we establish the following result.

Theorem 1 *For any $\varepsilon > 0$ there is a d_0 such that for any $d \ge d_0$ we have*

$$b(d) \ge 1 - \frac{2\ln d}{d} + \frac{-4 + 2\ln 2 - \varepsilon}{d}. \tag{5}$$

The term $\frac{2\ln d}{d}$ is matched by the following estimate of the independence number $i(G)$ by Frieze and Łuczak [6]: for every $\varepsilon > 0$ there is a d_1 such that for any $d \ge d_1$ we have whp

$$\left| i(G_{n,d}) - \frac{2n}{d}\left(\ln d - \ln\ln d + 1 - \ln 2\right) \right| \le \frac{\varepsilon n}{d}. \tag{6}$$

Since, $\phi(G) \le v(G) - i(G)$, Theorem 1 and the estimate (6) imply that, for $n \gg d \gg 1$, the maximum size of an induced forest in $G_{n,d}$ is whp $\frac{n}{d}\left(2\ln d + O(\ln\ln d)\right)$. It would be interesting to estimate the error term more precisely.

There is a simple relationship between $b(d)$ and α_d.

Lemma 2 *For any $d \ge 2$ we have $b(d) \le B(d) \le \alpha_d$.*

Proof. Observe that any f-fragmented set $Y \subset V(G_{n,d})$ can be made acyclic by removing a vertex from each cycle of $G_{n,d}[Y]$ of length at most f. But, for $n \gg f$, there are $o(n)$ such cycles in the whole graph $G_{n,d}$ whp. ∎

Note that, in view of this lemma, the inequality (2) improves the upper bound (4) on $B(4)$.

Theorem 1 together with Lemma 2 answer the question of Graham and Farr. But we can give another proof of (3), which is much shorter and, in fact, gives a sharper estimate.

Theorem 3 *For every $k \geq 2$ we have*

$$\begin{aligned}\alpha_{2k} &\geq 1 - \frac{2}{k+1},\\ \alpha_{2k+1} &\geq 1 - \frac{2k+3}{(k+1)(k+2)}.\end{aligned}$$

In the next section we give the proof of Theorem 1. Then, in the final section, we prove Theorem 3 and give some further observations about the parameter α_d.

2 Random regular graphs

The random regular graph $G_{n,d}$ can be generated by taking a random 1-factor on dn points, these points being grouped into n *groups* of d apiece. Each group is then identified to a single vertex to produce a d-regular multigraph G'. We obtain $G_{n,d}$ by conditioning on the event that G' is *simple* (i.e. it has no loops or multiple edges), see Bollobás [3, Chapter 2.4].

Recall that $\phi_d(n)$ is the smallest ϕ such that with probability at least 1/2 we have $\phi(G_{n,d}) \leq \phi$. For fixed d and $n \to \infty$ the choice of the constant 1/2 has little effect, as $\phi(G_{n,d})$ is concentrated on an interval of length $O(n^{1/2})$. Indeed, a random pairing of dn points can be obtained by taking a random permutation and joining the points in positions $2i-1$ and $2i$. A single transposition affects at most 4 vertices of the resulting d-regular multigraph G' so it can change $\phi(G')$ by at most 4 (in fact, by at most 1). The version of Talagrand's inequality [9] for functions determined by a random permutation implies the claimed concentration for G'. The claim for $G_{n,d}$ follows by observing that, for fixed d, the probability of G' being simple is bounded from 0.

The proof of Theorem 1 is based on a first moment calculation which in fact shows that whp any set $A \subset V(G_{n,d})$ with $|A| = (1 - f(d,\varepsilon))n$ spans at least $|A|$ edges, where $f(d,\varepsilon)$ denotes the right-hand side of (5). Although the approach is straightforward, we have to go through somewhat messy calculations. The method in [2] for proving a lower bound on $b(d)$ is also based on the expectation argument but it estimates the number of induced forests via generating functions. Our calculations indicate that this approach gives a bound comparable to ours, namely $b(d) \geq 1 - \frac{2\ln d}{d} + O(1/d)$, but the proof would be longer.

Let $n \gg d \gg 1/\varepsilon$. Choose a λ such that $m = \lambda n/d$ is an integer and

$$2\ln d + 4 - 2\ln 2 + \frac{\varepsilon}{2} \leq \lambda \leq 2\ln d + 4 - 2\ln 2 + \varepsilon. \tag{7}$$

Let $\mathcal{X}$ consist of all m-subsets of $V(G')$ that span at most $m-1$ edges in the multigraph G'. We show that the expectation of $|\mathcal{X}|$ is $o(1)$.

The expected number of m-subsets X spanning $i \le m$ edges in G' is

$$\begin{aligned} E_i &= \binom{n}{m}\binom{dm}{2i}\Phi(2i)\binom{dn-dm}{dm-2i}(dm-2i)!\,\frac{\Phi(dn-dm-(dm-2i))}{\Phi(dn)} \\ &= \frac{n!\,(dm)!\,2^{dm}\,(dn-dm)!\,(dn/2)!}{(n-m)!\,m!\,(dm-2i)!\,2^{2i}\,i!\,(dn/2-dm+i)!\,(dn)!}, \end{aligned}$$

where $\Phi(2m)=(2m)!/(2^m m!)$ is the number of perfect matchings on $2m$ points. The ratio

$$\frac{E_{i+1}}{E_i} = \frac{(dm-2i)(dm-2i-1)}{4(i+1)(dn/2-dm+i+1)}.$$

is a decreasing function of i. Thus, for any $i<m$ we have

$$\frac{E_{i+1}}{E_i} \ge \frac{E_m}{E_{m-1}} = (1+o(1))\,\frac{(d-2)^2\lambda}{d(2d-4\lambda+4\lambda/d)},$$

which is at least, for example, 2 if the constant d is large. Hence,

$$E\big(\,|\mathcal{X}|\,\big) = \sum_{i=0}^{m-1} E_i \le E_m(2^{-1}+2^{-2}+\dots) = E_m.$$

From now on it is more convenient to operate with λ rather than with m which is dependent on n. We have

$$E_m = \frac{n!\,(\lambda n)!\,2^{(\lambda-\frac{2\lambda}{d})n}\,(dn-\lambda n)!\,(dn/2)!}{(n-\frac{\lambda}{d}n)!\,((\frac{\lambda n}{d})!)^2\,(\lambda n-\frac{2\lambda}{d}n)!\,(\frac{d}{2}-\lambda+\frac{\lambda}{d})n)!\,(dn)!} = n^{O(1)}(h(d,\lambda))^n,$$

where, by Stirling's formula,

$$h(d,\lambda) = \frac{\lambda^\lambda\,2^{\lambda-\frac{2\lambda}{d}}\,(d-\lambda)^{d-\lambda}(d/2)^{d/2}}{(1-\frac{\lambda}{d})^{1-\frac{\lambda}{d}}\,(\frac{\lambda}{d})^{\frac{2\lambda}{d}}\,(\lambda-\frac{2\lambda}{d})^{\lambda-\frac{2\lambda}{d}}\,(\frac{d}{2}-\lambda+\frac{\lambda}{d})^{\frac{d}{2}-\lambda+\frac{\lambda}{d}}\,d^d}.$$

It remains to argue that $h=h(d,\lambda)$ is strictly less than 1. For $d\to\infty$, we have $h\to 1$, so we have to take into account smaller order terms. We use the following estimates. (Recall that $\lambda=\Theta(\ln d)$.)

$$\begin{aligned} \frac{\lambda^\lambda}{(\lambda-\frac{2\lambda}{d})^{\lambda-\frac{2\lambda}{d}}} &= \lambda^{\frac{2\lambda}{d}}\left(1-\frac{2}{d}\right)^{-\lambda+\frac{2\lambda}{d}} = \lambda^{\frac{2\lambda}{d}}\,\mathrm{e}^{\frac{2\lambda}{d}+O(\frac{\lambda}{d^2})}, \\ \left(1-\frac{\lambda}{d}\right)^{1-\frac{\lambda}{d}} &= \mathrm{e}^{-\frac{\lambda}{d}+O(\frac{\lambda^2}{d^2})}, \\ \frac{(d-\lambda)^{d-\lambda}}{d^d} &= d^{-\lambda}\left(1-\frac{\lambda}{d}\right)^{d-\lambda} = d^{-\lambda}\left(\mathrm{e}^{-\frac{\lambda}{d}-\frac{\lambda^2}{2d^2}+O(\frac{\lambda^3}{d^3})}\right)^{d-\lambda} = d^{-\lambda}\,\mathrm{e}^{-\lambda+\frac{\lambda^2}{2d}+O(\frac{\lambda^3}{d^2})}, \\ \frac{(d/2)^{d/2}}{(\frac{d}{2}-\lambda+\frac{\lambda}{d})^{\frac{d}{2}-\lambda+\frac{\lambda}{d}}} &= (d/2)^{\lambda-\frac{\lambda}{d}}\left(\mathrm{e}^{-\frac{2\lambda}{d}+\frac{2\lambda}{d^2}-\frac{(2\lambda)^2}{2d^2}+O(\frac{\lambda^3}{d^3})}\right)^{-\frac{d}{2}+\lambda-\frac{\lambda}{d}} = (d/2)^{\lambda-\frac{\lambda}{d}}\,\mathrm{e}^{\lambda-\frac{\lambda^2}{d}-\frac{\lambda}{d}+O(\frac{\lambda^3}{d^2})}. \end{aligned}$$

We obtain

$$h = \frac{2^{\lambda - \frac{2\lambda}{d}}}{(\frac{\lambda}{d})^{\frac{2\lambda}{d}}} \lambda^{\frac{2\lambda}{d}} \mathrm{e}^{\frac{2\lambda}{d}} \mathrm{e}^{\frac{\lambda}{d}} d^{-\lambda} \mathrm{e}^{-\lambda + \frac{\lambda^2}{2d}} (d/2)^{\lambda - \frac{\lambda}{d}} \mathrm{e}^{\lambda - \frac{\lambda^2}{d} - \frac{\lambda}{d}} \mathrm{e}^{O(\frac{\lambda^3}{d^2})}$$
$$= 2^{-\frac{\lambda}{d}} d^{\frac{\lambda}{d}} \mathrm{e}^{\frac{2\lambda}{d}} \mathrm{e}^{-\frac{\lambda^2}{2d}} \mathrm{e}^{O(\frac{\lambda^3}{d^2})} = \exp\left(-\frac{\varepsilon \ln d}{2d} + O(d^{-1})\right) < 1.$$

Thus $E(|\mathcal{X}|) = o(1)$. By the Markov inequality, whp $\mathcal{X} = \emptyset$, that is, any m vertices of the multigraph G' span at least m edges. The result for $G_{n,d}$ follows because the probability that G' is simple can be bounded from 0 by a function depending on d only.

This completes the proof of Theorem 1. ▮

3 Fragmentability

We begin with the proof of Theorem 3.

For $d = 2k$ we use the construction of Alon, Ding, Oporowski, and Vertigan [1], see also Haxell, Szabó, and Tardos [7]. Let f be an arbitrary constant. Let H be a $(k+1)$-regular graph of girth larger than f. Let $n = v(H)$. Let G be the line graph of H. It is regular of degree $2k$. The order of G is $n(k+1)/2$. Note that any set of n vertices of G has a component of size at least $f+1$ because the corresponding edge set of H must span a cycle. Hence, G is not $(1 - \frac{2}{k+1}, f)$-fragmentable and the required bound on α_{2k} follows.

For $d = 2k+1$ we slightly modify the above construction: namely, take a bipartite graph H with partition $V(H) = V_1 \cup V_2$ such that the girth of H is larger than f and every vertex in V_1 (resp. in V_2) has degree $k+1$ (resp. $k+2$). The existence of such a graph can be established e.g. by randomly pairing the points in $(k+2)n$ groups each of size $k+1$ with the points in $(k+1)n$ groups each of size $k+2$, where $n \to \infty$. Let G be the line graph of H. It has order $\frac{(k+1)(k+2)}{2k+3} v(H)$ while any f-fragmented set in G contains at most $v(H) - 1$ vertices. The theorem is proved. ▮

It is convenient to denote $\gamma_d = 1 - \alpha_d$. Theorem 3 and (1) then imply that γ_d has magnitude $\Theta(d^{-1})$ as $d \to \infty$. We can show that this function behaves regularly.

Theorem 4 *The expression $d\gamma_d$ tends to some limit γ as $d \to \infty$. Moreover, for any $d \geq 2$ we have*

$$\gamma \geq (d+1)\gamma_d. \tag{8}$$

Proof. We prove first that for any $D \geq d \geq 2$, we have

$$(D + d + 1)\gamma_D \geq (d+1)\gamma_d. \tag{9}$$

Fix any small $\varepsilon > 0$. Choose an f such that any graph of maximum degree d is $(\alpha_d + \varepsilon, f)$-fragmentable. Let G be an arbitrary graph of maximum degree at most D. Let $n = v(G)$. Let $q = \lceil \frac{D+1}{d+1} \rceil \leq \frac{D+d+1}{d+1}$. Take a partition $V(G) = V_1 \cup \cdots \cup V_q$ which maximizes the number of edges across the parts. Since $q(d+1) > D$, it follows that each $G[V_i]$ has maximum degree at most d. Suppose without loss of generality that $|V_1| \geq n/q$. By the definition of ε, applied to $G[V_1]$, we can find an f-fragmented set $X \subset V_1$ with $|X| \geq (\gamma_d - \varepsilon)|V_1| \geq (\gamma_d - \varepsilon)n/q$. The same set $X \subset V(G)$ shows that G is (α, f)-fragmentable, where $\alpha = 1 - (\gamma_d - \varepsilon)/q$. We conclude that

$$\gamma_D \geq \frac{\gamma_d - \varepsilon}{q} \geq \frac{d+1}{D+d+1}\,(\gamma_d - \varepsilon).$$

Since ε was arbitrary, (9) follows.

All claims of the theorem routinely follow from (9). ∎

Theorem 3 implies that $\gamma \leq 4$ while (2) and (8) imply that $\gamma \geq \frac{25}{8}$.

Problem 5 *Compute γ exactly.*

Acknowledgements

The authors thank Alan Frieze and Nick Wormald for helpful discussions.

References

[1] N. Alon, G. Ding, B. Oporowski, and D. Vertigan, *Partitioning into graphs with only small components*, J. Combin. Theory (B) **87** (2003), 231–243.

[2] S. Bau, N. C. Wormald, and S. Zhou, *Decycling numbers of random regular graphs*, Random Struct. Algorithms **21** (2002), 397–413.

[3] B. Bollobás, *Random Graphs*, 2d ed., Cambridge Univ. Press, 2001.

[4] P. Cameron, *Research problems from the 18th British Combinatorial Conference*, Discrete Math. **266** (2003), 441–451.

[5] K. Edwards and G. Farr, *Fragmentability of graphs*, J. Combin. Theory (B) **82** (2001), 30–37.

[6] A. M. Frieze and T. Łuczak, *On the independence and chromatic numbers of random regular graphs*, J. Combin. Theory (B) **54** (1992), 123–132.

[7] P. Haxell, T. Szabó, and G. Tardos, *Bounded size components—partitions and transversals*, J. Combin. Theory (B) **88** (2003), 281–297.

[8] R. M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), Plenum, New York, 1972, pp. 85–103.

[9] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. **81** (1995), 73–205.