Combinatorics, Probability and Computing (1993) **11**, 1–10

On the edge-expansion of graphs

NOGA ALON

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University, Tel Aviv, Israel. Email: noga@math.tau.ac.il.
Received

It is shown that if $n > n_0(d)$ then any d-regular graph $G = (V, E)$ on n vertices contains a set of $u = \lfloor n/2 \rfloor$ vertices which is joined by at most $(\frac{d}{2} - c\sqrt{d})u$ edges to the rest of the graph, where $c > 0$ is some absolute constant. This is tight, up to the value of c.

1. Introduction

For a graph $G = (V, E)$ and for two disjoint subsets U, W of V, let $e(U, W)$ denote the number of edges between U and W. The edge expansion coefficient $i(G)$ of G is defined by

$$i(G) = Min \frac{e(U, V - U)}{|U|},$$

where the minimum is taken over all nonempty subsets U of V of cardinality at most $|V|/2$. For any integer d define, following Bollobás [3],

$$i(d) = sup\{\gamma : i(G) > \gamma \quad \text{for infinitely many } d\text{-regular graphs } G \quad \}.$$

It is well known that there is a tight correspondence between the second largest eigenvalue of a regular graph and the above coefficient. In particular, if G is d-regular, and the second largest eigenvalue of its adjacency matrix is λ, then, by the simple remarks following Lemma 2.1 in [1],

$$i(G) \geq \frac{d - \lambda}{2}.$$

Since it is easy to deduce from the results in [4] or from the constructions in [5], [6], that for any d there are infinitely many d-regular graphs whose second largest eigenvalue is bounded by $C\sqrt{d}$, for some absolute positive constant C, this implies that there is some $C' > 0$ so that

$$i(d) \geq \frac{d}{2} - C'\sqrt{d}$$

† Research supported in part by a USA Israeli BSF grant and by the Fund for Basic Research administered by the Israel Academy of Sciences.

for any integer d.

A more direct proof of this last inequality is given by Bollobás in [3], using probabilistic arguments.

In this note we show that the above inequality is tight, up to the best possible value of C', namely, the correct value of $i(d)$ satisfies

$$\frac{d}{2} - C'\sqrt{d} \leq i(d) \leq \frac{d}{2} - c'\sqrt{d},$$

for every d, where $C', c' > 0$ are absolute constants. This follows from the following result.

Theorem 1.1. *There exists an absolute constant $c > 0$ so that if $n > 40d^9$ and $G = (V, E)$ is a d-regular graph on n vertices, then there is a set U of $u = \lfloor n/2 \rfloor$ vertices of G such that $e(U, V - U) \leq (\frac{d}{2} - c\sqrt{d})u$.*

The proof is probabilistic and is based on the idea of Shearer in [8] together with some additional combinatorial arguments and the FKG-Inequality.

2. The proof

In this section we prove Theorem 1.1. Let $G = (V, E)$ be a d-regular graph on n vertices. We must show that there is a partition $V = V_0 \cup V_1$, where $|V_0| = \lfloor n/2 \rfloor$, $|V_1| = \lceil n/2 \rceil$ and $e(V_0, V_1) \leq \frac{nd}{4}(1 - \Omega(\frac{1}{\sqrt{d}}))$. For convenience we assume that d is odd and hence n is even; the case of even d and odd or even n can be treated similarly. The basic idea is very simple: we partition the vertices randomly into two classes, and if a vertex has more neighbors in the other class than in its own, then we randomly decide whether to shift it to the other class or leave it where it is. It is then shown that the expected number of edges between the two classes is at most $\frac{nd}{4}(1 - \Omega(\frac{1}{\sqrt{d}}))$. The main problem in the process of obtaining a rigorous proof along these lines is that we have to keep the two classes of equal size. This causes several difficulties, and we overcome them by combining the FKG-Inequality with some combinatorial ideas.

We need the following lemma.

Lemma 2.1. *Let H be a graph on $n = 2m$ vertices, with maximum degree Δ, and suppose $n > 40\Delta^3$. Then there is a perfect matching $M = \{(u_i, v_i) : 1 \leq i \leq m\}$ of all vertices of H satisfying the following properties.*
(i) Each edge of M is **not** *an edge of H.*
(ii) There is no alternating cycle of length 4 or 6 consisting of edges of H and M alternately.

Proof. Since the minimum degree of a vertex in the complement of H is at least $n - 1 - \Delta > n/2$, there is, by Dirac's Theorem, a Hamilton cycle in this complement and hence, by taking every second edge of this cycle we conclude that there is a matching M satisfying property (i). To complete the proof we show that as long as the matching M violates property (ii) we can modify it so that (i) still holds and the number of alternating cycles of length 4 or 6 strictly decreases. To do so, suppose there is an alternating cycle C of length 4 or 6 containing the edge (u_i, v_i) of M. Let W denote the set of all vertices of H that can be reached from either u_i or v_i by an alternating path of length at most 5, (starting with an edge of H). Clearly, $|W| \leq 2(1+2\Delta+2\Delta^2+\Delta^3) < n/2$, and hence there is an edge of M, say, (u_j, v_j), no end of which lies in W. Let M' be the matching obtained from M by replacing the two edges (u_i, v_i) and (u_j, v_j) by the edges (u_i, v_j) and (u_j, v_i). It is easy to see that

M' satisfies (i), C is not an alternating cycle after replacing M by M', and no new alternating cycle of length 4 or 6 has been formed, as the newly added edges cannot be contained in such a cycle. Thus, the number of alternating cycles of length 4 and 6 has been strictly reduced, and by repeating the procedure we must end with a matching satisfying the assertion of the lemma. □

Returning to the proof of Theorem 1.1, consider the following randomized procedure for constructing a partition of the set of vertices of $G = (V, E)$ into two equal parts V_0 and V_1. First, let H be the graph on V in which two vertices are adjacent if their distance in G is at most 3. By assumption, the maximum degree Δ in H satisfies $n > 40\Delta^3$, and hence, by Lemma 2.1 there is a matching $M = \{(u_i, v_i) : 1 \leq i \leq m\}$ satisfying the assertion of the lemma. Let $h : V \mapsto \{0, 1\}$ be a random function obtained by choosing, for each i, $1 \leq i \leq m$, randomly and independently, one of the two possibilities ($h(u_i) = 0$ and $h(v_i) = 1$) or ($h(u_i) = 1$ and $h(v_i) = 0$), both choices being equally probable. Call a vertex $v \in V$ *stable* if it has more neighbors u satisfying $h(u) = h(v)$ than neighbors w satisfying $h(w) \neq h(v)$, otherwise call it *active*. Call a pair of vertices (u_i, v_i) matched under M an *active pair* if both u_i and v_i are active, otherwise, call it a *stable pair*. Let $h' : V \mapsto \{0, 1\}$ be the random function obtained from h by randomly modifying the values of the vertices in active pairs as follows. If (u_i, v_i) is an active pair then choose randomly either ($h'(u_i) = 0$ and $h'(v_i) = 1$) or ($h'(u_i) = 1$ and $h'(v_i) = 0$), both choices being equally probable. Otherwise, define $h'(u_i) = h(u_i)$ and $h'(v_i) = h(v_i)$. Finally, define $V_0 = h'^{-1}(0)$ and $V_1 = h'^{-1}(1)$.

It is obvious that $|V_0| = |V_1| = m$ ($= n/2$). To complete the proof we show that the expected value of $e(V_0, V_1)$ is only $\frac{nd}{4}(1 - \Omega(\frac{1}{\sqrt{d}}))$. Fix an edge of G; by renaming the vertices if needed, we may assume, without loss of generality, that its two vertices are u_1 and u_2, which are matched under M to v_1 and v_2 respectively. Our objective is to estimate the probability that $h'(u_1) \neq h'(u_2)$. This is done by estimating the conditional probability of this event assuming that $h(u_1) = h(u_2)$ and the conditional probability assuming that $h(u_1) \neq h(u_2)$. Before starting to estimate these probabilities, note that by the choice of M the sets $\{v_1\} \cup N(v_1)$ and $\{v_2\} \cup N(v_2)$ of the closed neighborhoods of v_1 and v_2, respectively, are disjoint and both of them do not intersect the set $\{u_1, u_2\} \cup N(u_1) \cup N(u_2)$. Moreover, the only edges of M whose two ends lie in the set

$$\{u_1, u_2, v_1, v_2\} \cup N(u_1) \cup N(u_2) \cup N(v_1) \cup N(v_2)$$

are the two edges $\{u_1, v_1\}$ and $\{u_2, v_2\}$. These facts, illustrated in Figure 1, will be useful as they imply that various events are independent. Thus, for example, the event (v_1 *is active and* $h(v_1) = 0$) is independent of the event (v_2 *is active and* $h(v_2) = 1$), as those are determined by disjoint sets of random choices. (Note that for this to hold it is not enough that the closed neighborhoods of v_1 and v_2 are disjoint; one also needs the fact that there are no edges of M joining these two neighborhoods.)

In order to estimate the conditional probability $Prob[h'(u_1) \neq h'(u_2) | h(u_1) = h(u_2)]$ note, first, that in case $h(u_1) = h(u_2)$ then if at least one of the pairs (u_1, v_1) or (u_2, v_2) is active, then this probability is precisely a half. On the other hand, if they are both stable, it is zero. Therefore

$$Prob[h'(u_1) \neq h'(u_2) | h(u_1) = h(u_2)] = \frac{1}{2} - \frac{1}{2} Prob[(u_1, v_1), (u_2, v_2) \text{ stable} \mid h(u_1) = h(u_2)] \tag{1}$$

Clearly

$$Prob[(u_1, v_1), (u_2, v_2) \text{ stable} \mid h(u_1) = h(u_2)]$$

u_1 v_1

u_2 v_2

Figure 1 A typical edge u_1u_2

$$= Prob[(u_1, v_1) \ \text{ stable} \mid h(u_1) = h(u_2)] \cdot Prob[(u_2, v_2) \ \text{ stable} \mid h(u_1) = h(u_2), (u_1, v_1) \ \text{ stable}] \qquad (2)$$

However,

$$Prob[(u_1, v_1) \ \text{ stable} \mid h(u_1) = h(u_2)]$$

$$= Prob[v_1 \ \text{ stable} \mid h(u_1) = h(u_2)]$$

$$+Prob[v_1 \ \text{ active} \mid h(u_1) = h(u_2)] \cdot Prob[u_1 \ \text{ stable} \mid h(u_1) = h(u_2), v_1 \ \text{ active}].$$

Since, by the choice of M, the set $\{u_1, u_2\}$ does not intersect $N(v_1)$ and none of its members is matched under M to a member of $N(v_1)$, it follows that

$$Prob[v_1 \ \text{ stable} \mid h(u_1) = h(u_2)] = Prob[v_1 \ \text{ stable}] = 1/2.$$

Define

$$\epsilon_d = \binom{d-1}{\frac{d-1}{2}} / 2^d.$$

We claim that

$$Prob[u_1 \ \text{ stable} \mid h(u_1) = h(u_2), v_1 \ \text{ active}]$$

$$= Prob[u_1 \ \text{ stable} \mid h(u_1) = h(u_2)] = \sum_{i=0}^{(d-1)/2} \binom{d-1}{i} / 2^{d-1} = \frac{1}{2} + \epsilon_d.$$

To see this, note, first, that by the choice of M the event (v_1 *active*) is determined only by the values of $|h(w) - h(v_1)|$ for $w \in N(v_1)$ and hence does not influence the conditional probability $Prob[u_1 \ \text{ stable} \mid h(u_1) = h(u_2)]$. The above expression for the last conditional probability follows from the fact that if $h(u_1) = h(u_2)$ then u_1 is stable if and only if it has at most $(d-1)/2$ neighbors $w \in N(u_1) - \{u_2\}$ satisfying $h(w) \neq h(u_1)$.

Substituting the expressions above we conlude that

$$Prob[(u_1,v_1) \text{ stable} \mid h(u_1)=h(u_2)] = \frac{1}{2}+\frac{1}{2}(\frac{1}{2}+\epsilon_d) = \frac{3}{4}+\frac{1}{2}\epsilon_d. \quad (3)$$

We can now apply a similar reasoning to estimate the conditional probability

$$Prob[(u_2,v_2) \text{ stable} \mid h(u_1)=h(u_2),(u_1,v_1) \text{ stable}].$$

The crucial point is that when $h(u_1)=h(u_2)$, the event ((u_2,v_2) *stable*) and the event ((u_1,v_1) *stable*) behave monotonely with respect to the h-values on the intersection $N(u_1)\cap N(u_2)$, in case this intersection is nonempty. That is, if one of these events occurs, then by changing the value of some $h(w)$ for w in this intersection from $1-h(u_1)=1-h(u_2)$ to $h(u_1)$, this event still occurs. It thus follows from the FKG Inequality (cf. e.g., [2], Chapter 6) that

$$Prob[(u_2,v_2) \text{ stable} \mid h(u_1)=h(u_2),(u_1,v_1) \text{ stable}] \geq \frac{3}{4}+\frac{1}{2}\epsilon_d. \quad (4)$$

Combining (2),(3) and (4),

$$Prob[(u_1,v_1),(u_2,v_2) \text{ stable} \mid h(u_1)=h(u_2)] \geq \frac{9}{16}+\frac{3}{4}\epsilon_d+\frac{1}{4}\epsilon_d^2,$$

and therefore, by (1)

$$Prob[h'(u_1)\neq h'(u_2)|h(u_1)=h(u_2)] \leq \frac{1}{2}(\frac{7}{16}-\frac{3}{4}\epsilon_d-\frac{1}{4}\epsilon_d^2). \quad (5)$$

Similar arguments can be used to estimate the conditional probability

$$Prob[h'(u_1)\neq h'(u_2)|h(u_1)\neq h(u_2)].$$

Here are the details. Note, first, that

$$Prob[h'(u_1)\neq h'(u_2)|h(u_1)\neq h(u_2)] = \frac{1}{2}+\frac{1}{2}Prob[(u_1,v_1),(u_2,v_2) \text{ stable} \mid h(u_1)\neq h(u_2)] \quad (6)$$

Next, observe that

$$Prob[(u_1,v_1),(u_2,v_2) \text{ stable} \mid h(u_1)\neq h(u_2)]$$

$$= Prob[(u_1,v_1) \text{ stable} \mid h(u_1)\neq h(u_2)]\cdot Prob[(u_2,v_2) \text{ stable} \mid h(u_1)\neq h(u_2),(u_1,v_1) \text{ stable}] \quad (7)$$

However,

$$Prob[(u_1,v_1) \text{ stable} \mid h(u_1)\neq h(u_2)]$$

$$= Prob[v_1 \text{ stable} \mid h(u_1)\neq h(u_2)]$$

$$+Prob[v_1 \text{ active} \mid h(u_1)\neq h(u_2)]\cdot Prob[u_1 \text{ stable} \mid h(u_1)\neq h(u_2), v_1 \text{ active}].$$

As before, by the choice of M,

$$Prob[v_1 \text{ stable} \mid h(u_1)\neq h(u_2)] = Prob[v_1 \text{ stable}] = 1/2,$$

and

$$Prob[u_1 \text{ stable} \mid h(u_1)\neq h(u_2), v_1 \text{ active}]$$

$$= Prob[u_1 \text{ stable} \mid h(u_1)\neq h(u_2)] = \sum_{i=0}^{(d-3)/2}\binom{d-1}{i}/2^{d-1} = \frac{1}{2}-\epsilon_d,$$

since if $h(u_1) \neq h(u_2)$ then u_1 is stable if and only if it has at most $(d-3)/2$ neighbors $w \in N(u_1) - \{u_2\}$ satisfying $h(w) \neq h(u_1)$.

Substituting, we conlude that

$$Prob[(u_1, v_1) \text{ stable} \mid h(u_1) \neq h(u_2)] = \frac{1}{2} + \frac{1}{2}(\frac{1}{2} - \epsilon_d) = \frac{3}{4} - \frac{1}{2}\epsilon_d. \tag{8}$$

By a similar computation, and using the FKG-Inequality it follows, next, that

$$Prob[(u_2, v_2) \text{ stable} \mid h(u_1) \neq h(u_2), (u_1, v_1) \text{ stable}] \leq \frac{3}{4} - \frac{1}{2}\epsilon_d, \tag{9}$$

since when $h(u_1) \neq h(u_2)$ then the event $((u_1, v_1)$ *stable*$)$ is monotone increasing with respect to changing the values of some $h(w)$ for $w \in N(u_1) \cap N(u_2)$ from $h(u_2)$ to $h(u_1)$, whereas the event $((u_2, v_2)$ *stable*$)$ is monotone decreasing with respect to such a change.

By (7),(8) and (9),

$$Prob[(u_1, v_1), (u_2, v_2) \text{ stable} \mid h(u_1) \neq h(u_2)] \leq \frac{9}{16} - \frac{3}{4}\epsilon_d + \frac{1}{4}\epsilon_d^2,$$

and therefore, by (6)

$$Prob[h'(u_1) \neq h'(u_2) | h(u_1) \neq h(u_2)] \leq \frac{1}{2}(\frac{25}{16} - \frac{3}{4}\epsilon_d + \frac{1}{4}\epsilon_d^2). \tag{10}$$

Combining (5) and (10) we finally conclude that

$$Prob[h'(u_1) \neq h'(u_2)]$$

$$= Prob[h(u_1) = h(u_2)] \cdot Prob[h'(u_1) \neq h'(u_2) | h(u_1) = h(u_2)]$$

$$+ Prob[h(u_1) \neq h(u_2)] \cdot Prob[h'(u_1) \neq h'(u_2) | h(u_1) \neq h(u_2)]$$

$$\leq \frac{1}{4}(\frac{32}{16} - \frac{3}{2}\epsilon_d) = \frac{1}{2} - \frac{3}{8}\epsilon_d.$$

Since (u_1, u_2) was a typical edge, by linearity of expectation, the expected value of $e(V_0, V_1)$ is at most

$$\frac{nd}{2}(\frac{1}{2} - \frac{3}{8}\epsilon_d).$$

We have thus proved the following explicit form of Theorem 1.1.

Proposition 2.2. *If $n > 40d^9$ and d is odd, then any d-regular graph $G = (V, E)$ on n vertices contains a partition $V = V_0 \cup V_1$ where $|V_0| = |V_1| = n/2$ such that*

$$e(V_0, V_1) \leq \frac{nd}{4}(1 - \frac{3}{4}\epsilon_d).$$

In [8], Lemma 1, it is observed that for every $d \geq 1$, $\epsilon_d \geq \frac{1}{2\sqrt{2}\sqrt{d}}$, and since it is not difficult to prove a version of the last proposition for even values of d as well, it follows that

$$i(d) \leq \frac{d}{2}(1 - \frac{3}{8\sqrt{2}\sqrt{d}}). \tag{11}$$

3. Concluding remarks

- The assumption $n > 40d^9$ in Theorem 1.1 can be easily relaxed, and we make no attempt to optimize it here. Note, however, that some assumption of the form $n > n_0(d)$ is essential, since the assertion of the theorem fails for the complete graph K_{d+1} or for the complete bipartite graph $K_{d,d}$.
- As noted in the end of the previous section our proof actually shows that $i(d) \le \frac{d}{2}(1 - \frac{3}{8\sqrt{2}\sqrt{d}})$ for all $d \ge 1$. It seems possible to improve the constant $\frac{3}{8\sqrt{2}}$ by modifying slightly the update probability in the definition of h', but the problem of finding the best possible value of this constant seems very difficult. In [3] it is shown that as d tends to infinity

$$i(d) \ge \frac{d}{2}(1 - \frac{(2+o(1))(ln2)^{1/2}}{\sqrt{d}}),$$

  Thus the correct value of $i(d)$ is $\frac{d}{2} - \Theta(\sqrt{d})$, but the problem of determining this function precisely remains open.
- Our proof clearly works for non-regular graphs as well, as long as the maximum degree is small with respect to the number of vertices. Repeating the arguments in Section 2 one can prove the following.

Proposition 3.1. *There exists an absolute positive constant c so that the following holds. Let $G = (V, E)$ be a graph on n vertices with degree sequence $d_1, \ldots, d_n$, and suppose $n > 40d_i^9$ for all i. Then there is a partition $V = V_0 \cup V_1$, where $|V_0| = \lfloor n/2 \rfloor$, $|V_1| = \lceil n/2 \rceil$, such that*

$$e(V_0, V_1) \le \frac{1}{2}\sum_{i=1}^{n} d_i(\frac{1}{2} - \frac{c}{\sqrt{d_i}}).$$

- Theorem 1.1 together with the fact mentioned in the introduction that for any d-regular graph G on n vertices with second largest eigenvlaue λ, $i(G) \ge \frac{d-\lambda}{2}$, imply that if n is sufficiently large with respect to d then $\lambda \ge \Omega(\sqrt{d})$. A tighter inequality is known, however. In this case $\lambda \ge (1-o(1))2\sqrt{d-1}$, where the $o(1)$ term tends to 0 as n tends to infinity. This is tight, and the proof is, in fact, much easier than the proof of Theorem 1.1 here; see, e.g., [7]. We note that the fact that for **any** d there are infinitely many d-regular graphs satisfying $\lambda \le 2\sqrt{d} + o(d)$ can be deduced from the results in [4] or from the constructions in [5], [6] by packing together various graphs, so as to obtain d-regular ones. We omit the details.
- The *bisection width* of a graph on n vertices is the minimum possible value of $e(V_0, V_1)$, as V_0, V_1 range over all partitions of V into two parts satisfying $|V_0| = \lfloor n/2 \rfloor$, $|V_1| = \lceil n/2 \rceil$. This notion arises naturally in the study of embedding problems motivated by questions in the design of VLSI circuits. Thus, our result here shows that the maximum possible bisection width of a d-regular graph on n vertices is $\frac{nd}{4}(1 - \Omega(1/\sqrt{d}))$. Moreover, the proof, as presented in the previous section, supplies an efficient randomized algorithm for finding a partition V_0, V_1 with $e(V_0, V_1)$ bounded as ensured by the theorem, and it is not too difficult to apply the method of conditional expectations (see, e.g., [2], Chapter 15) and convert this algorithm into an efficient deterministic one.
- For small values of d it is possible to improve the upper bound provided for $i(d)$ in (11). Trivially, $i(1) = i(2) = 0$ and in [3] it is proved that $i(3) \le 1$. In fact, this bound can be improved to $i(3) \le 1/2$ as follows. Let G be a cubic graph on n vertices, where n is large. It is easy to see we may assume G is connected.

Take a spanning tree in G and observe that since its maximum degree is at most 3, it can be broken into connected pieces each containing, say, between $\sqrt{n}$ and $4\sqrt{n}$ vertices. Now split these pieces randomly into two classes. By standard Chernoff estimates, with high probability the total number of vertices in each class will be $(\frac{1}{2}+o(1))n$. The expected number of edges whose ends lie in the two classes is only half of the number of edges that do not lie in the pieces, namely, half of $\frac{3}{2}n-(1-o(1))n$. This implies that $i(3) \leq 1/2$ and a similar argument shows that $i(4) \leq 1$. We omit the details. The known lower bounds for these quantities, proved in [3], are $i(3) > 2/11$ and $i(4) > 11/25$.

Acknowledgment I would like to thank M. Krivelevich and D. Spielman for helpful comments.

References

[1] N. Alon and V. D. Milman, λ_1*, isoperimetric inequalities for graphs and superconcentrators*, J. Combinatorial Theory, Ser. B 38(1985), 73-88.

[2] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons Inc., New York, 1992.

[3] B. Bollobás, *The isoperimetric number of random regular graphs*, Europ. J. Combinatorics 9 (1988), 241-244.

[4] J. Friedman, J. Kahn and E. Szemerédi, *On the second eigenvalue in random regular graphs*, Proc. 21^{st} ACM STOC (1989), 587-598.

[5] A. Lubotzky, R. Phillips and P. Sarnak, *Explicit expanders and the Ramanujan conjectures*, Proc. 18^{th} ACM STOC (1986), 240-246. See also: A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan graphs*, Combinatorica 8 (1988), 261-277.

[6] G. A. Margulis, *Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and superconcentrators* , Problemy Peredachi Informatsii 24(1988), 51-60 (in Russian). English translation in Problems of Information Transmission 24(1988), 39-46.

[7] A. Nilli, *On the second eigenvalue of a graph*, Discrete Mathematics 91 (1991), 207-210.

[8] J. B. Shearer, *A note on bipartite subgraphs of triangle-free graphs*, Random Structures and Algorithms 3 (1992), 223-226.