Cutting a Graph into Two Dissimilar Halves

Paul Erdős
MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCES
1364 BUDAPEST, HUNGARY

Mark Goldberg
DEPARTMENT OF COMPUTER SCIENCE
RENSSELAER POLYTECHNIC INSTITUTE
TROY, NEW YORK

János Pach
MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCE
1364 BUDAPEST, HUNGARY
AND DEPARTMENT OF MATHEMATICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK

Joel Spencer
DEPARTMENT OF MATHEMATICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK

ABSTRACT

Given a graph G and a subset S of the vertex set of G, the discrepancy of S is defined as the difference between the actual and expected numbers of the edges in the subgraph induced on S. We show that for every graph with n vertices and e edges, $n < e < n(n-1)/4$, there is an $n/2$-element subset with the discrepancy of the order of magnitude of $\sqrt{ne}$. For graphs with fewer than n edges, we calculate the asymptotics for the maximum guaranteed discrepancy of an $n/2$-element subset. We also introduce a new notion called "bipartite discrepancy" and discuss related results and open problems.

1. INTRODUCTION

Let G be an arbitrary graph with $v(G) = n$ vertices and $e(G) = e$ edges. For any subset S of the vertex set of G, let the *discrepancy* of S be defined as the

Journal of Graph Theory, Vol. 12, No. 1, (1988) 121–131
 CCC 0364-9024/88/010121-11$04.00

difference between the actual and expected numbers of edges in $G[S]$, i.e., in the subgraph of G induced by S. That is, let

$$\text{dis}\,(S) = e(S) - e\frac{\binom{|S|}{2}}{\binom{n}{2}} = e(S) - e\frac{|S|\,(|S| - 1)}{n(n - 1)},$$

where $e(S)$ is the shorthand form of $e(G[S])$. The average behavior of dis (S) is studied in [2].

In the problem session of the last Southeastern Conference on Combinatorics in Boca Raton (1986) the senior author raised the following question: Is it true that for every $c > 0$ there exists a constant $\hat{c} > 0$ with the property that any graph G with n vertices and $cn < e < \binom{n}{2}\text{-}cn$ edges contains two sets of vertices S and T such that $|S| = |T| = n/2$ and $|e(S) - e(T)| > \hat{c}n$? Our following result answers this question in the affirmative.

Theorem 1. Let G be a graph with n vertices and e edges, $n < e < n(n - 1)/4$, and assume that n is even. Then one can find two subsets $S, T \subset V(G)$ such that $|S| = |T| = n/2$ and

$$|e(S) - e(T)| > \alpha\sqrt{en}\,,$$

where α is an absolute constant.

At first glance, one might naively conjecture (as we did) that in the above theorem S and T can be chosen to be disjoint. However, if G is any regular graph and $S \cup T$ is any partition of its vertex set into two equal halves, then $e(S)$ and $e(T)$ are always equal.

The following, slightly weaker assertion is still true:

Theorem 2. For every μ, $0 < \mu < \frac{1}{2}$, there exists a $\upsilon > 0$ such that in any graph with n vertices and e edges, $n < e < n(n - 1)/4$, one can find two disjoint subsets S and T such that $|S| = |T| = |\mu n|$ and

$$|e(S) - e(T)| > \upsilon\sqrt{en}\,.$$

The proofs of the above theorems rely heavily on a generalization of an old quasi-Ramsey-type result of the first- and the last-named authors [5,6,1] (see section 2) and on the following *Expansion-Retraction Theorem:*

Theorem 3. Let G be a graph with n vertices and assume that $|\text{dis}\,(R)| = D$ for some subset $R \subset V(G)$. Then there exists a subset $S \subset V(G)$ with $|S| = \lfloor n/2 \rfloor$ such that

$$|\text{dis}\,(S)| > \left(\frac{1}{4} + o(1)\right)D\,,$$

where the $o(1)$ term goes to 0 as D tends to infinity.

In the case when G has fewer than n edges we have much sharper results. To formulate them we introduce some further notations. For any graph G with n vertices, let

$$d^+(G) = \max\,\text{dis}\,(S)\,,$$
$$d^-(G) = -\min\,\text{dis}\,(S)\,,$$

and

$$d(G) = \max(d^+(G), d^-(G)) = \max|\text{dis}\,(G)|\,,$$

where the *max* and *min* are taken over all $\lfloor n/2 \rfloor$-element subsets $S \subset V(G)$. Further, for any $c > 0$, let

$$d^+(n,c) = \min\{d^+(G) : e = \lfloor cn \rfloor\}\,,$$
$$d^-(n,c) = \min\{d^-(G) : e = \lfloor cn \rfloor\}\,,$$
$$d(n,c) = \min\{d(G) : e = \lfloor cn \rfloor\}\,.$$

Theorem 4.

(*) $$\lim_{n\to\infty} \frac{d^-(n,c)}{n} = \begin{cases} c/4 & \text{if } 0 < c \le 1/2 \\ (2-c)/4 & \text{if } 1/2 < c \le 1\,. \end{cases}$$

(**) $$\lim_{n\to\infty} \frac{d^+(n,c)}{n} = \begin{cases} 3c/4 & \text{if } 0 < c \le 1/4\,, \\ (1-c)/4 & \text{if } 1/4 < c \le 1/2\,, \\ c/4 & \text{if } 1/2 < c \le 1\,. \end{cases}$$

(***) $$\lim_{n\to\infty} \frac{d(n,c)}{n} = \lim_{n\to\infty} \frac{d^+(n,c)}{n} \qquad \text{if } 0 < c \le 1\,.$$

Note that, in general, $d^+(G)$ and $d^-(G)$ can be essentially different from each other. For example, if G consists of two disjoint cliques of size $n/2$, then $d^+(G) \approx (n^2/16)$ and $d^-(G) \approx (n/16)$.

The proofs of Theorems 1–3 and Theorem 4 can be found in sections 2 and 3, respectively. The last section contains some generalizations, related results, and open problems. In particular, we will introduce and discuss a new parameter of a graph called the "bipartite discrepancy," which depends on the deviance of the most irregular bipartitions.

2. DISCREPANCY OF GRAPHS

Let G be a graph with n vertices and e edges, and let A and B be two disjoint subsets of $V(G)$. Set

$$\text{dis}\,(A,B) = e(A,B) - e\frac{|A|\,|B|}{\binom{n}{2}},$$

where $e(A,B)$ denotes the number of edges in G running between A and B.

The following theorem is a straightforward generalization of a result in [5] and [3].

Theorem 5. For every $\varepsilon > 0$ there exists $\hat{\varepsilon} > 0$ such that any graph G with n vertices and $e > n$ edges contains two disjoint subsets A and B with the property that $|A|, |B| < \varepsilon n$ and

$$|\text{dis}\,(\text{A},B)| > \hat{\varepsilon}\sqrt{en}\,.$$

Proof. Assume for simplicity that n is even, $\varepsilon < (1/16)$, and decompose $V(G)$ into disjoint parts U and V, $|U| = |V|$. Let $\mathbf{A}$ be a randomly chosen $\lfloor \varepsilon n \rfloor$-element subset of U and set

$$V(\mathbf{A}) = \left\{ v \in V\colon |\text{dis}\,(v,A)| > 10^{-2}\sqrt{\frac{\varepsilon e}{n}} \right\}.$$

Then

$$\mathbf{Pr}\left[|\text{dis}\,(v,\mathbf{A})| > 10^{-2}\sqrt{\frac{\varepsilon e}{n}} \right] > \frac{1}{2}.$$

Hence, the expected size of $V(\mathbf{A})$ equals

$$\sum_{v\in V} \mathbf{Pr}\left[|\text{dis}\,(v,\mathbf{A})| > 10^{-2}\sqrt{\frac{\varepsilon e}{n}} \right] > \frac{n}{4}.$$

On the other hand,

$$\frac{n}{4} < \mathbf{E}[|V(\mathbf{A})|] \le \frac{n}{2}\mathbf{Pr}\left[|V(\mathbf{A})| > \frac{n}{8} \right] + \frac{n}{8}\left(1 - \mathbf{Pr}\left[|V(\mathbf{A})| > \frac{n}{8} \right] \right),$$

implying

$$\mathbf{E}\left[|V(\mathbf{A})| > \frac{n}{8}\right] > \frac{1}{3}.$$

Thus, one can choose a specific A and an $[\varepsilon n]$-element subset $B \subset V(A)$ such that dis $(v, A) > 10^{-2}\sqrt{\varepsilon e/n}$, or dis $(v, A) < -10^{-2}\sqrt{\varepsilon e/n}$, hold for all $v \in B$. In both cases, A and B meet the requirements of the theorem with $\hat{\varepsilon} = 10^{-2}\varepsilon^{3/2}$. ∎

Corollary. For every $\varepsilon > 0$ there exists a $\delta > 0$ with the property that any graph G with n vertices and $e > n$ edges contains an at most $2\varepsilon n$-element subset $R \subseteq V(G)$ such that

$$|\text{dis}\ (R)| > \delta\sqrt{en}\,.$$

Proof. It is sufficient to note that

$$\text{dis}\ (A \cup B) = \text{dis}\ (A) + \text{dis}\ (B) + \text{dis}\ (A, B)\,.$$

Hence, if A and B satisfy the conditions in Theorem 5, then the absolute value of the discrepancy of at least one of the sets A, B, or $A \cup B$ exceeds $\hat{\varepsilon}(\sqrt{en}/3)$. ∎

Next we prove the *Expansion-Retraction Theorem* stated in the introduction.

Proof of Theorem 3. Let $|R| = m$ and suppose for convenience that n is even. If $m \geq n/2$, then let $\mathbf{S}$ be a randomly chosen $\lfloor n/2 \rfloor$-element subset of R. The expected number of edges in $G[\mathbf{S}]$ is

$$\mathbf{E}[e(\mathbf{S})] = e(R)\frac{\binom{n/2}{2}}{\binom{m}{2}} \approx \frac{1}{4}e(R)\left(\frac{n}{m}\right)^2,$$

implying

$$\mathbf{E}[\text{dis}\ (\mathbf{S})] \approx \text{dis}\ (R)\left(\frac{n}{2m}\right)^2.$$

Thus there exists a specific S with $|\text{dis}\ (S)| \geq |\text{dis}\ (R)|/4$.

Now assume $m < n/2$ and denote $\overline{R}$ the complement of R. Let $\mathbf{P}$ be a randomly chosen $(n/2)$-element subset of $\overline{R}$ and let $\mathbf{Q}$ be a random set consisting of R and $n/2$-m randomly chosen vertices of $\overline{R}$. Denote $E_1 = \mathbf{E}[e(\mathbf{P})]$ and $E_2 = \mathbf{E}[e(\mathbf{Q})]$. We will establish an upper bound for $\min(E_1, E_2)$ in the case of $D \geq 0$ and a lower bound for $\max(E_1, E_2)$ in the opposite case.

Clearly,

$$E_1 \approx \frac{1}{4} e(\overline{R}) \frac{n^2}{(n-m)^2} = F_1,$$

$$E_2 \approx e(R) + e(R,\overline{R}) \frac{(n/2)-m}{n-m} + e(\overline{R}) \frac{((n/2)-m)^2}{(n-m)^2} = F_2.$$

Since $e(R,\overline{R}) = e - e(R) - e(\overline{R})$, for fixed e and $e(R)$, F_1 and F_2 are linear functions of $x = e(\overline{R})$. Therefore, $\min(\max(F_1,F_2))$ as well as $\max(\min(E_1,E_2))$ is achieved if $F_1 = F_2$. Thus,

$$\frac{1}{4} x_0 \left(\frac{n}{n-m}\right)^2 = e(R) + \frac{1}{2}(e - e(R) - x_0)\frac{n-2m}{n-m} + \frac{1}{4} x_0 \left(\frac{n-2m}{n-m}\right)^2,$$

$$x_0 = e(R) + e\frac{n-2m}{n}.$$

Substituting $e(R)$ for $e(m/n)^2 + D$ we get

$$F_1(x_0) = F_2(x_0) = \frac{1}{4}e + \frac{1}{4}D\left(\frac{n}{n-m}\right)^2.$$

This implies that for some specific $n/2$-element subset S of the form **P** or **Q**,

$$|\mathrm{dis}\,(S)| \geq \left(\frac{1}{4} + o(1)\right)D.$$

Moreover, the signs of dis (S) and dis (R) are identical. Note also that the extreme value $\frac{1}{4}$ in Theorem 3 is only achieved if $|R|/n$ is nearly 0 or 1; otherwise, the constant can be improved. ∎

Proof of Theorem 1. To obtain S, apply Theorem 3 to the set R constructed in the corollary. Let **T** be a randomly chosen $n/2$-element subset of $V(G)$. Then

$$\mathbf{E}[e(S) - e(\mathbf{T})] = \mathbf{E}[\mathrm{dis}\,(S) - \mathrm{dis}\,(\mathbf{T})] = \mathrm{dis}\,(S),$$

yielding the result. ∎

For the proof of Theorem 2 we need the following slightly generalized form of the *Expansion-Retraction Theorem:*

Theorem 3′. Let G be a graph with n vertices, ε and υ positive numbers, $\varepsilon < 1 - \upsilon$, and assume that

$$|\text{dis}(R)| = D$$

for some subset $R \subset V(G)$ having at most εn elements. Then there exists a subset $S \subset V(G)$ with $|S| = \lfloor \upsilon n \rfloor$ such that

$$|\text{dis}(S)| \geq (\upsilon \min(\upsilon, 1 - \upsilon) + o(1))D\,,$$

where the $o(1)$ term goes to 0 as D tends to infinity.

Proof of Theorem 2. Divide the vertex set of G into two disjoint equal parts U and V such that $e(G[U]) \geq e/4$. Applying the corollary to the graph $G[U]$ with $\varepsilon = 1 - 2\mu$, we obtain that there exists an at most $(1 - 2\mu)n$-element subset R of U with $|\text{dis}(R)| > \delta\sqrt{(e/4)(n/2)}$. By Theorem 3′, there is $S \subset U$ with $|S| = \lfloor 2\mu \frac{1}{2} n \rfloor = \lfloor \mu n \rfloor$ and

$$|\text{dis}(S)| > (2\mu \min(2\mu, 1 - 2\mu) + o(1))\delta\sqrt{\frac{en}{8}} = D'\,,$$

so we can choose another $\lfloor \mu n \rfloor$-element subset $S' \subset U$ such that

$$|e(S) - e(S')| \geq D'.$$

Then, for any $\lfloor \mu n \rfloor$-element subset $T \subset V$, either $|e(S) - e(T)| > \frac{1}{2}D'$ or $|e(S') - e(T)| > \frac{1}{2}D'$. ∎

3. SPARSE GRAPHS

In this section, we consider graphs with n vertices and cn edges, where $c \leq 1$. The following form of Túran's theorem will be used:

Theorem 6 [7]. Every graph with n vertices and e edges contains an independent set of size $\geq n^2/(2e + n)$.

Proof of Theorem 4. If $c \leq \frac{1}{2}$, then by Túran's theorem we can find in G an independent set J of size $\geq n^2/(2e + n) \geq n/2$. Obviously, dis $(J) = -cn \times (\frac{1}{4} + o(1))$ and thus $d^-(n, c) = n[(c/4) + o(1)]$ for $0 \leq c \leq \frac{1}{2}$.

To prove the second part of (∗), we show that every graph with n vertices and e edges $((n/2) \leq e \leq n)$ contains an independent set J of size $\geq (2n - e)/3$. Indeed, this is true for $n = 2$ and, due to Túran's theorem, it follows for every graph with n vertices and $e = n$ edges. Let $n > 2$ and $e < n$. We may assume without loss of generality that G has no isolated vertices. Then G must have a vertex of degree 1. Let w be such a vertex and let z be adjacent to w. We delete z together with all edges incident to it. The remaining graph has an isolated vertex w and a subgraph H with $n - 2$ vertices and $\leq e - 1$ edges. By induction, H contains an independent set Q of size $\geq [2(n - 2) - (e - 1)]/3 = [(2n - e)/3] - 1$. Thus, the independent set $J = Q \cup w$ contains $\geq (2n - e)/3$ vertices.

Having constructed J, we expand it to an $(n/2)$-element subset S by adding one by one the necessary number of vertices in such a way that each addition brings at most one new edge. Such an expansion certainly exists, since otherwise we would find a subset T such that

(1) $|T| > n/2$, and
(2) every $x \in T$ is adjacent to at least two vertices in $V - T$.

This would imply that $|E| \geq 2|T| > n$, which is impossible. Thus, $S \supseteq J$ induces a subgraph with $\leq (n/2) - [(2n - e)/3] = (2e - n)/6$ edges. This proves that both $d^-(G)$ and $d^-(n, c)$ are $\geq [(2 - c)/12]n + o(n)$. To see that $d^-(n, c) \leq [(2 - c)/12]n + o(n)$, take the union of $(1 - c)n$ edges and $(2c - 1)/3$ triangles (all are disjoint).

Next we show (**). If $e \leq n/4$, then, evidently, G has a subgraph with $n/2$-vertices which contains all edges. This yields $d^+(n, c) \approx [(3c)/4]n$.

If $e > n/4$, then consider the connected components $G_1, G_2, \ldots, G_r$ of G. Let $e(G_i) = v(G_i) - 1 + \delta_i (i = 1, \ldots, r)$ and let $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r$. If k is the smallest i with $\delta_i = 0$, then we assume that $v(G_k) \geq v(G_{k+1}) \geq \cdots \geq v(G_r)$. Let, also, $H = \cup_{i=1}^{k+1} G_i$ and

$$s^* = \sum_{i=1}^{k+1} v(G_i) .$$

Obviously, $e(H) \geq s^* - 1$. Therefore, if $s^* \geq n/2$, then

$$d^+(G) \geq \frac{2 - c}{4} n + o(n) .$$

In the case $s^* \leq n/2$, we add to H some components $G_{k-2}, G_{k+3}, \ldots$ to get a graph F, with $n/2$ vertices (it is possible that the last component will be only partially included). Clearly, $e(F) \geq e/2$ and thus $d^+(n, c) \geq c/4$. In addition, $e(F) \geq n/4$, otherwise

$$e(F) = \sum_{x \in F} d_F(x) \leq \frac{n}{4} - 1$$

would imply that F contains at least two isolated vertices, therefore $e(F) = e \geq n/4$.

So, if $c \geq \frac{1}{4}$ then

$$d^+(n,c) \geq \begin{cases} \dfrac{1-c}{4}n + o(n) & \text{if } \dfrac{1}{4} \leq c \leq \dfrac{1}{2}, \\ \dfrac{c}{4}n + o(n) & \text{if } \dfrac{1}{2} \leq c \leq 1. \end{cases}$$

To show that this bound is best possible, consider a graph with n vertices and e edges, which consists of $p = n - e - 1$ disjoint paths of length $\lceil e/p \rceil$, and another component, which is a path of length $l = e - p\lceil e/p \rceil$ (in case $e > 0$).

Finally, note that (∗∗∗) follows from (∗) and (∗∗). ∎

4. BIPARTITE DISCREPANCY

For any graph G with n vertices and e edges, let the bipartite discrepancy of G be defined by

$$\text{bdis}\,(G) = \max\left(|\text{dis}\,(S,T)| : S \cup T = V(G), |S| = \left\lfloor \frac{n}{2} \right\rfloor, |T| = \left\lceil \frac{n}{2} \right\rceil\right).$$

That is, bdis (G) is the maximum deviation of the number of edges running between two complementary halves of $V(G)$ from

$$e\frac{\left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil}{\binom{n}{2}},$$

i.e., from its expected value.

Conjecture 1. For any $0 < \varepsilon < \frac{1}{2}$, there exists a δ such that

$$\text{bdis}\,(G) \geq \delta n^{3/2}$$

holds for every graph G with n vertices and $\frac{1}{2}\binom{n}{2} \leq e \leq (1-\varepsilon)\binom{n}{2}$ edges.

Conjecture 1′. For any $0 < \varepsilon < \frac{1}{2}$, there exists a $\hat{\delta}$ such that, if G is any graph with n vertices and $\frac{1}{2}\binom{n}{2} \leq e \leq (1-\varepsilon)\binom{n}{2}$ edges, and $w_1, w_2, \ldots, w_n$ are any weights assigned to the vertices of G, then one can always find an $\lfloor n/2 \rfloor$-

element subset $S \subset V(G)$ satisfying

$$\left|e(S) - \sum_{i \in S} w_i\right| \geq \hat{\delta} n^{3/2}.$$

Proposition. Conjecture 1′ implies Conjecture 1.

Proof. Assume, for simplicity, that n is a multiple of 6, and let T_0 be an arbitrary set of $n/3$ vertices of G. For any $i \in V(G) - T_0$ set

$$w_i = \left|\{t \in T_0: (i,t) \in E(G)\}\right| - 3\frac{e(T_0)}{n}.$$

Applying Conjecture 1′ to the subgraph of G induced by $V(G) - T_0$, we can find an $n/3$-element subset $S \subseteq V(G)$, disjoint from T_0, with

$$\left|e(S) - \sum_{i \in S} w_i\right| = \left|e(S_0) + e(T_0) - e(S_0, T_0)\right| \geq \hat{\delta}\left(\frac{2n}{3}\right)^{3/2}.$$

Now split $V(G) - S_0 - T_0$ arbitrarily into $n/6$ pairs x_j, y_j, and let $\mathbf{S}$ be a random set that contains S_0 and exactly one vertex from each pair. Further, let $\mathbf{T} = V(G) - \mathbf{S}$. Then any edge of G with at least one endpoint not in $S_0 \cup T_0$ has probability precisely $\frac{1}{2}$ of being in $e(S,T)$, unless it is an edge of the form (x_j, y_j). Thus

$$\mathbf{E}[e(\mathbf{S}) + e(\mathbf{T}) - e(\mathbf{S}, \mathbf{T})] = e(S_0) + e(T_0, T_0) - \Delta,$$

where $0 < \Delta \leq n/12 = o(n^{3/2})$. Hence, there exist S and T with $|\mathrm{dis}\,(S,T)| = |e(S) + e(T) - e(S,T)| \geq \delta n^{3/2}$. Note that, in the special case when $w_i = e/(2n)$, the truth of Conjecture 1′ follows from [5] or from the corollary in section 2. ∎

Let c_0 denote the maximal positive c such that a random graph with n vertices and cn edges has a partition of the vertex set into two subsets of sizes $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, respectively, for which the number of edges with endpoints in different parts is $o(n)$. By [4], a random graph with n vertices and cn edges consists of a "giant" component of size $[1 - x(c)/2c]n$ and small components of sizes $O(\ln n)$, where $x(c)$ is the solution satisfying $0 < x(c) < 1$ of the equation $x(c)e^{-x(c)} = 2ce^{-2c}$. For $c = \ln 2$, the size of the "giant" component is $n/2$, implying that $c_0 \geq \ln 2$.

Conjecture 2. $c_0 = \ln 2$.

Conjecture 2 would proceed from the following:

Conjecture 3. For every $\varepsilon > 0$, there is but $o((1+\varepsilon)^n)$ partitions of the vertex set of a random tree T into two subsets of sizes $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, respectively, for which the number of edges with endpoints in different parts is $o(n)$.

ACKNOWLEDGMENTS

The work of M. Goldberg was supported in part by the National Science Foundation under grant DCR-8520872.

References

[1] B. Bóllobás, *Random Graphs*. Academic Press, London–New York (1985).
[2] D. de Caen, A note on the probabilistic approach to Túran's problem. *J. Combinat. Theory Ser. B* **34** (1983).
[3] P. Erdós, *The Art of Counting (Selected Writings)*. MIT Press, Cambridge, MA–London (1973).
[4] P. Erdós and A. Renyi, On the evolution of random graphs. *Mat. Kutato Int. Kozl.* **5** (1960) 17–60.
[5] P. Erdós and J. Spencer, Imbalances in k-colorations. *Networks* **1** (1972) 379–385.
[6] P. Erdós and J. Spencer, *Probabilistic Methods in Combinatorics*. Academic Press, New York–London and Akademiai Kiado, Budapest (1974).
[7] P. Túran, On the theory of graphs. *Colloq. Math.* **3** (1954) 19–30.