Perfectly balanced partitions of smoothed graphs

Ido Ben-Eliezer* Michael Krivelevich†

Submitted: Nov 27, 2008; Accepted: Apr 13, 2009; Published: May 11, 2009
Mathematics Subject Classification: 05C07,05C80

Abstract

For a graph $G = (V, E)$ of even order, a partition (V_1, V_2) of the vertices is said to be perfectly balanced if $|V_1| = |V_2|$ and the numbers of edges in the subgraphs induced by V_1 and V_2 are equal. For a base graph H define a random graph $G(H, p)$ by turning every non-edge of H into an edge and every edge of H into a non-edge independently with probability p. We show that for any constant ϵ there is a constant α, such that for any even n and a graph H on n vertices that satisfies $\Delta(H)-\delta(H) \leq \alpha n$, a graph G distributed according to $G(H, p)$, with $\frac{\epsilon}{n} \leq p \leq 1-\frac{\epsilon}{n}$, admits a perfectly balanced partition with probability exponentially close to 1. As a direct consequence we get that for every p, a random graph from $G(n, p)$ admits a perfectly balanced partition with probability tending to 1.

1 Introduction

Given a graph G of even order, an equipartition (V_1, V_2) is *perfectly balanced* if the number of edges spanned by $G[V_1]$ equals the number of edges spanned by $G[V_2]$. This definition can be extended to graphs of odd order, and in this case we require that the number of vertices and edges in each part differ by at most 1.

It is not difficult to verify that there are graphs that do not admit perfectly balanced partitions. The simplest example is the star $K_{1,n-1}$. Caro and Yuster [5] proved that for every graph $G = (V, E)$ with $n^{2-\epsilon}$ edges for some $\epsilon > 0$ there are two disjoint sets $S_1, S_2 \subseteq V$, $|S_1| = |S_2| = n/2 - o(n)$ such that the subgraphs induced by S_1 and S_2 have exactly the same number of edges, and this is asymptotically tight.

It is natural to seek for families of graphs that admit perfectly balanced partitions. In this note we show that smoothed graphs admit such a partition almost surely. Namely,

---

*School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: idobene@post.tau.ac.il.

†School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2006322, by grant 526/05 from the Israel Science Foundation, and by a Pazy memorial award.

for a given $\frac{\epsilon}{n} \leq p \leq 1 - \frac{\epsilon}{n}$, let H be a graph on n vertices that satisfies $\Delta(H) - \delta(H) \leq \alpha n$ for some small constant $\alpha(\epsilon)$ (here $\Delta(H)$ is the maximal degree and $\delta(H)$ is the minimal degree of H). If we modify every edge and non-edge independently with probability p, almost surely we get a graph that admits a perfectly balanced partition. As a corollary, we get also that for *every* $p = p(n)$, a graph distributed according to $G(n,p)$ admits a perfectly balanced partition almost surely.

Smoothed Graphs. Smoothed analysis studies the behavior of objects after adding a small amount of randomness. This concept was introduced in the context of algorithms by Spielman and Teng [8], and was studied also in the context of graphs (see, for example, [2, 3, 7]) and hypergraphs [9]. Given a fixed graph H on n vertices, define the probability distribution $G(H,p)$ as follows. Every edge of H is deleted with probability p, and every pair of non-adjacent vertices of H is connected by an edge with probability p, all these events are mutually independent. We will usually be interested in the case that p is relatively small, thus one may consider graphs from this distribution as noisy (or smoothed) variations of H.

We can now state the main result of this note.

Theorem 1 *For every $\epsilon > 0$ there are $\alpha(\epsilon) > 0$ and $c(\epsilon) > 0$ such that the following holds for every even n. For every graph H on n vertices that satisfies $\Delta(H) - \delta(H) \leq \alpha n$, a graph G distributed according to $G(H,p)$ for some $\frac{\epsilon}{n} \leq p \leq 1 - \frac{\epsilon}{n}$ admits a perfectly balanced partition with probability at least $1 - e^{-cn}$.*

Here we state the result and provide a proof for graphs of even size. It is not difficult to extend it to graphs of odd size. We note that the theorem is essentially tight: A graph distributed according to $G(K_{1,n-1}, \frac{1}{2n})$ admits a perfectly balanced partition with probability $o(1)$.

Theorem 1 implies the following corollary:

Corollary 2 *Let n be an even number. A graph distributed according to $G(n,p)$ admits a perfectly balanced partition asymptotically almost surely.*

Remark. The proof actually implies that if $\Omega(\frac{1}{n}) \leq p \leq 1 - O(\frac{1}{n})$, then the probability that a graph distributed as $G(n,p)$ admits a perfectly balanced partition is exponentially close to 1.

In the next section we give a proof of the main theorem. We do not try to optimize the constants throughout this work. Also, we omit floor and ceiling signs whenever these are not crucial.

2 Proofs

For two disjoint sets of vertices $S_1, S_2 \subseteq V(G)$, denote by $e(S_1)$ the number of edges in the induced subgraph $G[S_1]$ and by $e(S_1, S_2)$ the number of edges of G with one endpoint in S_1 and the other in S_2. We start by noting that given a partition (S_1, S_2) we have:

$$e(S_1) = \sum_{v \in S_1} d(v) - e(S_1, S_2),$$

and similarly:

$$e(S_2) = \sum_{v \in S_2} d(v) - e(S_1, S_2).$$

Therefore, in order to have a perfectly balanced partition we need to partition the vertices to two equally sized sets such that the sum of their degrees is exactly the same.

A variant of the following simple claim is proved in [5].

Claim 3 *If $a_1 \geq \ldots \geq a_{2k}$, then there exists an equipartition $[2k] = T_1 \cup T_2$ such that $|\sum_{i \in T_1} a_i - \sum_{i \in T_2} a_i| \leq a_1 - a_{2k}$.*

Proof: For a subset $R \subseteq [2k]$, denote $s(R) = \sum_{i \in R} a_i$. Consider an equipartition (T_1, T_2) of $[2k]$ minimizing $|s(T_1) - s(T_2)|$, assume w.l.o.g that $s(T_1) \geq s(T_2)$. If $s(T_1) = s(T_2)$ we are done. Otherwise find elements $a_i \in T_1$ and $a_j \in T_2$ such that $a_i > a_j$, and define a new equipartition (T_1', T_2') by swapping them: $T_1' = T_1 - a_i + a_j$; $T_2' = T_2 - a_j + a_i$. Clearly, $s(T_1') - s(T_2') = s(T_1) - s(T_2) - 2(a_i - a_j) < s(T_1) - s(T_2)$. Due to the optimality of (T_1, T_2), we get $s(T_1') - s(T_2') < 0$ and therefore $s(T_1') - s(T_2') \leq -(s(T_1) - s(T_2))$. This implies: $s(T_1) - s(T_2) \leq a_i - a_j \leq a_1 - a_{2k}$. ■

Given a graph G, a *degree matching* is a set M of disjoint pairs of vertices, such that the degrees within every pair differ by exactly 1. The following claim gives a sufficient condition for a graph to have a perfectly balanced partition in terms of the size of a largest degree matching:

Claim 4 *Suppose that a graph G contains a degree matching of size at least $\Delta(G) - \delta(G)$. Then G admits a perfectly balanced partition.*

Proof: Given such a degree matching M, we can partition the vertices in $V - \bigcup M$ according to Claim 3, and get an equipartition such that the difference between the sums of degrees in each part is bounded by $\Delta(G) - \delta(G)$. Now we add the vertices from M to the partition, pair by pair, where for each pair we add the vertex with higher degree to the set of vertices with smaller total degree. Since the initial difference is at most the number of pairs and in every round we change the difference by exactly 1, after the last stage the difference is at most 1. The sum of the degrees of all the vertices in the graph is even, hence we conclude that the two sets have exactly the same number of edges, and the claim follows. ■

Therefore, our main result will follow from the next lemma.

Lemma 5 *For every $\epsilon > 0$ there are $\alpha(\epsilon) > 0$ and $c(\epsilon) > 0$ such that for every graph H on n vertices that satisfies $\Delta(H) - \delta(H) \leq \alpha n$ and $\frac{\epsilon}{n} \leq p \leq 1 - \frac{\epsilon}{n}$, a graph distributed according to $G(H, p)$ has the following properties with probability at least $1 - e^{-cn}$:*

- $\Delta(G) - \delta(G) \leq 2\alpha n$.
- *G has a degree matching of size at least $2\alpha n$.*

In the course of the proof of Lemma 5, we need the following variation of the standard bounds of binomial variables (see, e.g., [4] Chapter 1, Theorem 6).

Claim 6 *Let $X \sim Bin(k,p)$ where $n/4 \leq k \leq 3n/4$ is large enough and $p \leq 1/2$.*

- *There is an absolute constant $\zeta > 0$ such that if $p \geq \frac{100}{n}$ then for every $z \in [kp - 2\sqrt{np}, kp + 2\sqrt{np}]$, $\Pr[X = z] \geq \frac{\zeta}{\sqrt{np}}$.*
- *For every constants $\epsilon > 0$ and $B > 0$ there is a constant $\xi(\epsilon, B) > 0$ such that if $\frac{\epsilon}{n} \leq p \leq \frac{100}{n}$ then for every $0 \leq z \leq B$, $\Pr[X = z] \geq \xi(\epsilon, B)$.*

Proof of Lemma 5. We will prove this theorem for $p \leq \frac{1}{2}$; the case $p > \frac{1}{2}$ follows by taking the complement. The first item follows using Chernoff-type bounds; For any pair of vertices u and v, we have $|d_H(u) - d_H(v)| \leq \alpha n$, so for every p, the probability that the difference after modifying edges is at least $2\alpha n$ is exponentially small. Using the union bound we conclude that the first property holds with probability exponentially close to 1, as claimed.

For the second item, let $V_1 = \{v : d_H(v) \leq n/2\}$ and let $V_2 = V - V_1$. We can assume that $|V_1| \geq n/2$, the other case can be treated similarly. Let $U_1 \subseteq V_1$ be an arbitrarily chosen set of size $n/4$. Clearly, for every $u \in U_1$, there are at least $n/4$ vertices outside U_1 which are non-adjacent to u.

We expose the neighbors of U_1 in G in three stages. At the first stage, we expose the edges inside U_1. For every vertex $u \in H$, denote by N_u the set of vertices outside U_1 that are adjacent to u in H, and by S_u the set of non-adjacent vertices to u outside U_1. At the second stage, for every $u \in U_1$ and $v \in N_u$, we delete the edge uv independently with probability p. We denote the degree of u after this stage by $d^*(u)$. Let $b = \max\{10, \lceil\sqrt{np}\rceil\}$. We group the vertices of U_1 to bins of width b, where the bin B_i contains all the vertices that satisfy $d^*(u) + |S_u|p \in [bi, b(i+1))$. That is, the vertices are grouped according to their expected final degree. We say that B_i is *heavy* if $|B_i| \geq \frac{b}{8}$. Recalling that $|U| = n/4$ we deduce that there are at least $\frac{n}{8}$ vertices inside heavy bins.

At the third stage, for every $u \in U_1$ and $w \in S_u$, we add the edge uw independently with probability p. The degree of u after this stage (which is also the degree of u in G) is a random variable $X_u = d^*(u) + Y_u$, where $Y_u \sim Bin(|S_u|, p)$ and all the variables $\{Y_u\}$ are independent.

A set M of disjoint pairs of random variables is a *matching* if every two variables within a pair differ by exactly 1. We next claim that the set $\{X_u\}$ contains a matching of size at least $2\alpha n$ with probability at least $1 - e^{-c(\epsilon)n}$. This will finish the proof as it proves that there is a large degree matching of the vertices of U_1 (and therefore, of the vertices of V).

We first prove that the expected number of matched variables is linear in n. Let $\{bj, bj+1, ..., b(j+2)-1\}$ be an interval that corresponds to a heavy bin B_j, and for $bj \le k < b(j+2)$ denote by T_k the number of vertices from B_j with degree exactly k after the third stage. A degree matching can be constructed by matching vertices of degree $bj+2r$ to vertices of degree $bj+2r+1$, and therefore by linearity of expectation the expected size of a largest matching in B_j is at least

$$\sum_{r=0}^{b-1} E[\min\{T_{bj+2r}, T_{bj+2r+1}\}]. \tag{1}$$

We note that the intervals that correspond to the bins are not disjoint, though the bins are; we define it so to avoid a problem where most of vertices in some bin B_j have degree $bj+b-1$ and p is very small. Let A_j be $\frac{|B_j|\zeta}{b}$, if $p \ge \frac{100}{n}$, and $|B_j| \cdot \xi(\epsilon, 100)$ otherwise, where ζ and ξ are defined in Claim 6. In both cases it is easy to verify that $A_j = \Theta\left(\frac{|B_j|}{b}\right)$. Since the final degree of every vertex u in B_j is distributed as $d^*(u) + Bin(|S_u|, p)$ and $n/4 \le |S_u| \le 3n/4$, the expected value of each T_{bj+r}, $r \in \{0, \ldots, 2b-1\}$, is at least A_j by Claim 6. We consider the following two cases:

$A_j \le 16$. In this case we have $|B_j| = \Theta(b)$. The probability that a certain vertex v will be of a certain degree in the corresponding interval is $\Theta(\frac{1}{b})$, where this probability may depend either on the constants ζ or ξ. Also, the degrees of every two vertices from B_j are independent. It is fairly easy to show that with some positive constant probability (that depends on either ξ or ζ) both T_{bj+2r} and $T_{bj+2r+1}$ are non-zero. Therefore, the expected size of a largest matching is linear in b (and hence in $|B_j|$), as desired.

$A_j > 16$. For a fixed k, as $E[T_k] \ge A_j$ we have

$$E[T_k] - 2\sqrt{E[T_k]} \ge E[T_k]/2 \ge A_j/2.$$

Note that since T_k is a sum of independent indicator variables, the standard deviation of T_k is bounded by $\sqrt{E[T_k]}$. Therefore, by Chebyshev's Inequality, for every $0 \le r < b$, with probability at least $3/4$ we have $T_{bj+r} \ge A_j/2$. We conclude that with probability at least $1/2$, a pair of adjacent values T_{bj+2r} and $T_{bj+2r+1}$ are both at least $A_j/2$, and therefore by (1) the expected value of the largest matching in B_j is $\Omega(|B_j|)$.

To summarize, we proved that the expected size of a largest matching in every heavy bin B_j is $\Omega(|B_j|)$. Denote by Z the maximum size of a degree matching in U_1. Since there are at least $n/8$ vertices in heavy bins and the expected value of Z is at least the sum of expectations over all heavy bins, we conclude that $E[Z] \ge c_1(\epsilon)n$, where $c_1(\epsilon)$ is a constant that may depend on ϵ.

Next we will prove that the size of the largest matching is concentrated near the expectation. We use a vertex exposure martingale, where Z_i is the expected size of the largest degree matching after having exposed the first i vertices from U_1 and their neighbors, and $Z = Z_{U_1}$ is the size of a largest matching. Since exposing the edges of a certain vertex may change the expected size of a maximum degree matching by at most 1, the martingale satisfies the Lipschitz condition (see, e.g., [1], Chapter 7). By Azuma's Inequality, for $\alpha = c_1(\epsilon)/4$ and $c(\epsilon) = 8 \cdot c_1^2(\epsilon)$, the size of the matching is at least $2\alpha n$ with probability $1 - e^{-c(\epsilon)n}$, and the lemma follows. ■

Proof of Corollary 2. Again we consider only $p \leq \frac{1}{2}$, otherwise the proof will follow by taking the complement. The case $p = \Omega(\frac{1}{n})$ follows easily by taking H to be the empty graph. For smaller values of p, every graph distributed as $G(n, p)$ asymptotically almost surely has $n - o(n)$ isolated vertices and is bipartite (see, e.g., [6]). Assuming that the graph has these properties, we have two sets W_1, W_2 of vertices that are independent and contain all non-isolated vertices of G. Next we add the isolated vertices to W_1 and W_2 and get an equipartition (V_1, V_2), such that V_1 and V_2 are independent and thus the partition is perfectly balanced, as desired. ■

References

[1] N. Alon and J. Spencer, **The Probabilistic Method**, Third Edition, Wiley, New York, 2008.

[2] T. Bohman, A. Frieze, M. Krivelevich and R. Martin, *Adding random edges to dense graphs*, Random Structures and Algorithms 24 (2004), 105-117.

[3] T. Bohman, A. Frieze and R. Martin, *How many random edges make a dense graph Hamiltonian?*, Random Structures and Algorithms 22 (2003), 33-42.

[4] B. Bollobás, **Random Graphs**, Academic Press, 1985.

[5] Y. Caro and R. Yuster, *Large disjoint subgraphs with the same order and size*, European Journal of Combinatorics 30 (2009), 813-821.

[6] S. Janson, T. Łuczak and A. Ruciński, **Random Graphs**, Wiley, New York, 2000.

[7] M. Krivelevich, B. Sudakov and P. Tetali, *On smoothed analysis in dense graphs and formulas*, Random Structures and Algorithms 29 (2006), 180-193.

[8] D. A. Spielman and S. H. Teng, Why the Simplex algorithm usually takes polynomial time, Proceedings of the 33th Annual ACM Symposium, STOC 2001, 296-305.

[9] B. Sudakov and J.Vondrak, *How many random edges make a dense hypergraph non-2-colorable?*, Random Structures and Algorithms 32 (2008), 290-306.