On Irregularity In Graphs

Julie Haviland*
Exeter College, Hele Road, Exeter EX4 4JS, U.K.

Abstract

Two parameters for measuring irregularity in graphs are the degree variance and the discrepancy. We establish best possible upper bounds for the discrepancy in terms of the order and average degree of the graph, and describe some extremal graphs, thereby providing analogues of results of [1], [4] and [5] for the degree variance.

1. Introduction

Let $G = (V, E)$ be a simple graph of order $|V| = n$. The *open neighbourhood* of a vertex $v \in V$ is the set $\Gamma(v) = \{w : vw \in E\}$, and the *degree* of the vertex v is $d(v) = |\Gamma(v)|$. The *average degree* of G is $d(G) = (1/n)\sum_{v\in V} d(v) = 2|E|/n$, where clearly $0 \leq d(G) \leq n-1$.

A number of parameters have been proposed as measures of graph irregularity, particularly in the context of analysing graphs derived from empirical observations. For example, the *degree variance* of G is defined to be

$$\text{var}(G) = \frac{1}{n}\sum_{v\in V} (d(v) - d(G))^2.$$

Bell [1] and Snijders [4],[5] produced tight upper bounds for var(G) in terms of the order of G, and cited the extremal graphs. Moreover the latter author used his results to derive certain 'heterogeneity indices' for social networks.

In [3], the *discrepancy* of G was defined to be

$$\text{disc}(G) = \frac{1}{n}\sum_{v\in V} |d(v) - d(G)|.$$

Thus the parameter disc(G) measures the average deviation of any vertex degree from the average degree of the graph. Note from the definitions that G has the same degree variance and discrepancy as its complement $\overline{G}$.

* Address for correspondence: 3 Church Mead, Gittisham, Honiton, EX14 3AH, U.K. Email address: juliehaviland@exe-coll.ac.uk

The motivation behind [3] was to find a convenient gauge of graph irregularity for use in modelling the spread of disease in population networks, although disc(G) was not explored in any detail in this earlier paper. In order for the discrepancy to be effective as such a measure it is necessary to find its range of possible values. Therefore in this note we establish sharp upper bounds for the discrepancy of a graph in terms of the order and average degree, and describe some graphs which attain these bounds.

In the ensuing work, all notation is standard, as given in [2].

2. Bounding The Discrepancy Of A Graph

From the definition, clearly disc(G) ≥ 0 with equality if and only if G is regular. We now establish the following best possible upper bounds for the discrepancy of a graph as functions of the order and average degree.

Theorem. *Let G be a simple graph of order n and average degree d; then*

$$\mathrm{disc}(G) \leq \frac{\psi}{n}\left(2n - 1 - \sqrt{4n\psi + 1}\right),$$

where $\psi = \min\{d, n - d - 1\}$.

Proof. Let $G = (V, E)$ be a graph of order n and average degree d with maximum discrepancy. Write $S = \{v \in V : d(v) > d\}$ and $s = |S|$; then

$$n\mathrm{disc}(G) = \sum_{v \in S} (d(v) - d) + \sum_{v \in V - S} (d - d(v)), \qquad (1)$$

where $\sum_{v \in S} (d(v) - d) = \sum_{v \in V - S} (d - d(v))$. Let $e(S)$ and $e(V - S)$ denote the number of edges with both endvertices in S and $V - S$ respectively, and let $e(S, V - S)$ denote the number of edges with one endvertex in each of S and $V - S$.

First suppose $s > d$. If $e(V - S) > 0$, then there must exist $x, y \in V - S$ with $xy \in E$. Since $d(x)$ and s are integers with $d(x) \leq d < s$, then we can find a non-neighbour of x in S, z say. Deleting the edge xy and adding xz produces a graph G' of average degree d with $n\mathrm{disc}(G') \geq n\mathrm{disc}(G) + 2$, which contradicts the maximality of G. We conclude that if $s > d$ then $e(V - S) = 0$. Therefore from (1) and the fact that $\sum_{v \in S} (d(v) - d) = \sum_{v \in V - S} (d - d(v))$ we have

$$\begin{aligned} n\mathrm{disc}(G) &= 2 \sum_{v \in V - S} (d - d(v)) \\ &\leq 2\left[(n - s)d - e(S, V - S)\right]. \qquad (2) \end{aligned}$$

A lower bound for $e(S, V-S)$ is given by

$$e(S, V-S) = |E| - e(S) \geq \frac{nd}{2} - \binom{s}{2}.$$

The right-hand side of the above expression is non-negative provided $s \leq \left(1+\sqrt{4nd+1}\right)/2$; if this condition holds then substituting for $e(S, V-S)$ in (2) yields

$$n\mathrm{disc}(G) \leq nd - 2sd + s^2 - s = f(s).$$

Now $f(s)$ is a quadratic function of s with positive coefficient of s^2 and $\partial f/\partial s = -2d + 2s - 1 = 0$ for $s = (d+1)/2$. Thus $f(s)$ achieves its maximum value over the range $d < s \leq \left(1+\sqrt{4nd+1}\right)/2$ at one endpoint. When $s = \left(1+\sqrt{4nd+1}\right)/2$ we have

$$n\mathrm{disc}(G) \leq f\left(\left(1+\sqrt{4nd+1}\right)/2\right) = d\left(2n-1-\sqrt{4nd+1}\right),$$

whilst if $s = d$ then

$$n\mathrm{disc}(G) \leq f(d) = d(n-d-1) \leq d\left(2n-1-\sqrt{4nd+1}\right)$$

for $d \leq n-1$. On the other hand, if $s > \left(1+\sqrt{4nd+1}\right)/2$ then using the lower bound $e(S, V-S) \geq 0$ in (2) yields

$$n\mathrm{disc}(G) \leq 2d(n-s) \leq d\left(2n-1-\sqrt{4nd+1}\right).$$

Otherwise $s \leq d$, and as each vertex of S has at most $s-1$ neighbours in S, then

$$\begin{aligned} e(S, V-S) &\geq \sum_{v\in S} (d(v) - s + 1) \\ &\geq \sum_{v\in S} (d(v) - d) + s(d-s+1). \end{aligned} \tag{3}$$

In addition, counting degrees of vertices in $V-S$ we have

$$\sum_{v\in V-S} (d - d(v)) \leq (n-s)d - e(S, V-S) - 2e(V-S). \tag{4}$$

Therefore eliminating $e(S, V-S)$ between (3) and (4) implies

$$\sum_{v\in S} (d(v) - d) + \sum_{v\in V-S} (d - d(v)) \leq nd - 2sd + s^2 - s - 2e(V-S),$$

and applying (1) gives

$$n\text{disc}(G) \leq nd - 2sd + s^2 - s - 2e(V - S). \tag{5}$$

A lower bound for $e(V - S)$ is given by

$$e(V - S) = |E| - e(S) - e(S, V - S) \geq \frac{nd}{2} - \binom{s}{2} - s(n - s).$$

Now if $s \leq \left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right)/2$ then the right-hand side of this last expression is positive, in which case substituting for $e(V - S)$ in (5) yields

$$\begin{aligned} n\text{disc}(G) &\leq 2(n - d - 1)s \\ &\leq (n - d - 1)\left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right). \end{aligned}$$

We are left to dispose of the case $s > \left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right)/2$. Using the bound $e(V - S) \geq 0$, the right-hand side of (5) becomes $f(s)$, with $\partial f/\partial s = -2d + 2s - 1 < 0$ for $s \leq d$. Hence we may assume that s is minimum possible, i.e.

$$\begin{aligned} n\text{disc}(G) &\leq f\left(\left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right)/2\right) \\ &= (n - d - 1)\left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right). \end{aligned}$$

Thus we have obtained two different upper bounds for $\text{disc}(G)$, and in order to complete the proof it remains to compare them over the range $0 \leq d \leq n-1$. Straightforward calculations reveal that $d\left(2n - 1 - \sqrt{4nd + 1}\right) \geq (n - d - 1)\left(2n - 1 - \sqrt{4n(n - d - 1) + 1}\right)$ for $0 \leq d \leq (n - 1)/2$, which implies the result. □

3. Extremal Graphs & Maximum Values

Let $H(n, d)$ be the graph of order n and average degree d with vertex and edge sets

$$\begin{aligned} V &= \{v_1, \ldots, v_n\}, \\ E &= \{v_i v_j : 1 \leq i \neq j \leq q\} \cup \{v_{q+1} v_i : 1 \leq i \leq r\}, \end{aligned}$$

where q and r are integers defined by $|E| = nd/2 = \binom{q}{2} + r$, $0 \leq r < q$.

Note that when $r = 0$ from the definition we must have $H(n,d) \cong K_q \cup (n-q)K_1$, where $\binom{q}{2} = nd/2$, so $q = (1+\sqrt{4nd+1})/2$; similarly $\overline{H}(n,n-d-1) \cong K_t + (n-t)K_1$, where $\binom{t}{2} + t(n-t) = nd/2$, so $t = \left(2n-1-\sqrt{4n(n-d-1)+1}\right)/2$. It is easily checked that in this case if $0 \le d \le (n-1)/2$ then $H(n,d)$ is extremal and if $(n-1)/2 \le d \le n-1$ then $\overline{H}(n,n-d-1)$ is extremal (although we do not claim uniqueness for either of these extremal graphs).

In [1], [4] and [5] it is shown that graphs of the type $H(n,d)$ and $\overline{H}(n,n-d-1)$ are also extremal for $\mathrm{var}(G)$, although the proof techniques employed are very different to the one given here, and the ranges of d for which $H(n,d)$ and $\overline{H}(n,n-d-1)$ are extremal are also different. With respect to $\mathrm{var}(G)$, $\overline{H}(n,n-d-1)$ is extremal for $0 \le d \le (n-1)/2$ and $H(n,d)$ is extremal for $(n-1)/2 \le d \le n-1$. Furthermore, in [5] it is proved that

$$\begin{aligned}\mathrm{var}(G) &\le q(q-1)^2(n-q)/n^2\\ &\le (n-2)(3n+2)(3n-2)^2/256n^2\\ &= \mathrm{var}_{\max}(G),\end{aligned}$$

with this maximum value being attained when $r = 0$ and $\max\{d, n-d-1\} = (9n^2-4)/16n$. Clearly $\mathrm{var}_{\max}(G) = 27n^2/256 + o(n^2)$. An analogous upper bound for the discrepancy may be obtained by observing that, as a function of ψ in the range $0 \le \psi = \min\{d, n-d-1\} \le (n-1)/2$, the upper bound of the Theorem is maximised at $\psi = \left(2n^2-2n-1+(2n-1)\sqrt{n^2-n+1}\right)/9n$, whence

$$\begin{aligned}\mathrm{disc}(G) &\le 2\left((2n-1)(n+1)(n-2) + 2(n^2-n+1)^{3/2}\right)/27n^2\\ &= \mathrm{disc}_{\max}(G),\end{aligned}$$

with $\mathrm{disc}_{\max}(G) = 8n/27 + o(n)$.

Acknowledgements

The author would like to thank the anonymous referee whose helpful suggestions greatly improved the clarity of this paper, and Maggie Walker for many stimulating discussions.

References

[1] F.K. Bell, A note on the irregularity of graphs, *Linear Algebra Appl.* **161** (1992), 45–54.

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press, London (1978).
[3] C.J. Lawrence, K. Tizzard and J. Haviland, Disease-spread and stochastic graphs, *Proc. International Conference On Social Networks, London* **1** (1995), 143–150.
[4] T.A.B. Snijders, Maximum value and null moments of the degree variance, Department of Mathematics Report TW-229, University of Groningen, 1981.
[5] T.A.B. Snijders, The degree variance: an index of graph heterogeneity, *Social Networks* **3** (1981), 163–174.