The Average Distance and the Independence Number

F. R. K. Chung
BELL COMMUNICATIONS RESEARCH
MORRISTOWN, NEW JERSEY

ABSTRACT

We prove that in every connected graph the independence number is at least as large as the average distance between vertices.

1. INTRODUCTION

In a graph G with vertex set $V(G)$ and edge set $E(G)$, we denote by $\alpha(G)$ the independence number of G (which is the maximum number of vertices in G that are pairwise nonadjacent). The distance between two vertices u and v in G, denoted by $d_G(u, v)$ [or $d(u, v)$ for short] is the length of a shortest path joining u and v in G. The average distance of G, denoted by $\overline{D}(G)$, is the average value of the distances between all pairs of vertices in G, i.e., $(\sum_{u,v} d(u, v))/\binom{n}{2}$. The diameter $D(G)$ of G is the maximum distance $d(u, v)$ over all pairs of vertices u and v.

Although D and $\overline{D}$ are interesting graph-theoretical invariants in their own right, they play significant roles in analyzing communication networks. In a network model, the time delay or signal disgradation for sending a message from one point to another is often proportional to the number of edges a message must travel. The average distance $\overline{D}(G)$ can be used to indicate the average performance of a network whereas the diameter $D(G)$ is related to the worst-case performance.

The independence number of $\alpha(G)$ is a much-studied graph-invariant (see [2–4]). We will establish a new inequality involving $\alpha(G)$ and the average distance $\overline{D}(G)$. This inequality was first conjectured by a computer program called GRAFFITI (see [5]). A weaker inequality, namely, $\alpha(G) \geq \overline{D}(G) - 1$, was proved by Fajtlowicz and Waller [5].

Theorem. For every connected graph G, we have $\alpha(G) \geq \overline{D}(G)$, with equality if and only if G is a complete graph.

Journal of Graph Theory, Vol. 12, No. 2, 229–235 (1988)

CCC 0364-9024/88/020229-07$04.00

In the next section, we will give the proof of the theorem. In section 3, several related extremal and algorithmic problems will be discussed.

2. THE PROOF OF THE THEOREM

We will prove the inequality in the theorem by induction on the number of vertices in a connected graph vertices of G. It is easy to see that $\alpha(G) \geq \overline{D}(G)$ for a connected graph G with two vertices. Suppose for $n \geq 3$ the inequality holds for all connected graphs G' with fewer than n vertices. Let G be a graph on n vertices. Then

$$\overline{D}(G) = \left(\sum_{u \neq v} d(u,v)\right) \Big/ \binom{n}{2}$$
$$= \frac{1}{n} \sum_{u} \frac{1}{n-1} \left(\sum_{v \neq u} d(u,v)\right).$$

For a vertex v, we define the average distance from v to be

$$f(v) = \frac{1}{n-1} \sum_{u \neq v} d(u,v),$$

and therefore

$$\overline{D}(G) = \frac{1}{n} \sum_{v} f(v).$$

Let w denote a vertex with $f(w) \leq \overline{D}(G)$. We will prove a sequence of claims from which the theorem follows:

Claim 1. If $G - \{w\}$ is connected, then $\alpha(G) \geq \overline{D}(G)$.

Proof. Set $G' = G - \{w\}$. It is easy to see that

$$\begin{aligned}
\alpha(G) &\geq \alpha(G') \\
&\geq \overline{D}(G') \qquad \text{(by induction)} \\
&= \frac{2}{(n-1)(n-2)} \sum_{u,v} d_{G'}(u,v) \\
&> \frac{2}{(n-1)(n-2)} \sum_{u,v \neq w} d_G(u,v) \\
&= \frac{2}{(n-1)(n-2)} \left(\frac{n(n-1)}{2}\overline{D}(G) - (n-1)f(w)\right)
\end{aligned}$$

$$\geq \frac{n}{n-2}\overline{D}(G) - \frac{2}{(n-2)}\overline{D}(G)$$
$$= \overline{D}(G). \quad \blacksquare$$

We can therefore assume that every vertex w for which $f(w) \leq \overline{D}(G)$ is a cut vertex.

Claim 2. If, for some $w, f(w) \leq \overline{D}(G) - 1$, then $\alpha(G) > \overline{D}(G)$.

Proof. For any neighbor w' of w, $f(w') \leq f(w) + 1 \leq \overline{D}(w)$, so w' must also be a cut vertex. For $w_1, \ldots, w_k$ denote the neighbors of w in G and let C_i be the connected component of $G - \{w\}$ containing w_i. Also, let c_i denote the order of C_i. We consider the following two cases:

Case 1. $k = 2$.

Let G_1 be the graph obtained from G' by adding the edge $\{w_1, w_2\}$ and joining each of w_1 and w_2 to all neighbors of the other. Namely, $E(G_1) = \{\{u, v\} \in E(G): u, v \in V(G_1)\} \cup \{\{u, v\}: u = w_i$ for some i and v is adjacent to w_j in G for some $j\} \cup \{\{w_1, w_2\}\}$. It is easy to see that $d_G(u, v) \leq d_{G_1}(u, v) + 2$ for $u \in C_1$ and $v \in C_2$, and $d_G(u, v) = d_{G'}(u, v)$ if $u, v \in C_i$ for some i.

We note that $\alpha(G) \geq \alpha(G_1) + 1$ since any independent set of G_1 can be extended to an independent set of G by adding one more vertex. On the other hand, we have

$$\binom{n}{2}\overline{D}(G) = (n-1)f(w) + \sum_{u, v \neq w} d_G(u, v)$$
$$< (n-1)(\overline{D}(G) - 1) + \sum_{u, v \neq w} d_{G_1}(u, v) + 2c_1c_2 .$$

This implies

$$\binom{n-1}{2}\overline{D}(G) < \binom{n-1}{2}\overline{D}(G_1) - (n-1) + \frac{(n-1)^2}{2}$$
$$\overline{D}(G) < \overline{D}(G_1) + 1$$
$$\leq \alpha(G_1) + 1$$
$$\leq \alpha(G) .$$

Case 2. $k \geq 3$.

Let $N(w)$ denote the set of neighbors of w in G and let $N^2(w)$ denote the set of vertices in $G - \{w\}$ that are adjacent to some vertex in $N(w)$.

We now construct a graph G_2 with $V(G_2) = V(G) - \{w\}$ and $E(G_2) = \{\{u, v\} \in E(G): \{u, v\} \subseteq V(G)\} \cup \{\{u, v\}: u \in N(w), v \in N(w) \cup N^2(w)\} \cup \{\{u, v\}: u \in N(w_i), v \in N(w_j)$ and $i \neq j\}\}$.

It is easily checked that $d_G(u,v) \leq d_{G_2}(u,v) + 3$ for $u \in C_i, v \in C_j$, and $i \neq j$, and $d_G(u,v) = d_{G_2}(u,v)$ for $u, v \in C_i$.

We have $\alpha(G) \geq \alpha(G_2) + k - 1$ since any independent set of G_2 can be extended to an independent set of G by adding one vertex from each of the other $k - 1$ components of G'. On the other hand, we have

$$\binom{n}{2}\overline{D}(G) = (n-1)f(w) + \sum_{u,v\neq w} d_G(u,v)$$

$$< (n-1)(\overline{D}(G) - 1) + \sum_{u,v\neq w} d_{G_2}(u,v) + 3\sum_{i\neq j} c_i c_j .$$

Since

$$\sum_{i\neq j} c_i c_j \leq \frac{(k-1)}{2}\sum_i c_i^2 \leq \frac{(k-1)}{2k}(n-1)^2 ,$$

then

$$\binom{n-1}{2}\overline{D}(G) < \binom{n-1}{2}\overline{D}(G_2) - (n-1) + \frac{3(k-1)(n-1)^2}{2k}$$

$$\overline{D}(G) < D(G_2) + \frac{3(k-1)}{k} + \left(\frac{3(k-1)}{k} - 2\right)\frac{1}{(n-2)}$$

$$\leq D(G_2) + k - 1 \qquad \text{for } k \geq 3$$

$$\leq \alpha(G_2) + k - 1$$

$$\leq \alpha(G) . \quad \blacksquare$$

Claim 3. If $D(G) \geq 2\lfloor\overline{D}\rfloor - 1$, then $\alpha(G) > \overline{D}(G)$.

Proof. Let s denote $\lfloor\overline{D}\rfloor$. Suppose for some vertices u and v we have $d(u,v) \geq 2s - 1$. Let L_i consist of all vertices x with $d(u,x) = i$. Clearly, $L_i \neq \emptyset$ for $0 \leq i \leq d(u,v)$. Let $I(L_i)$ denote a maximum independent set in L_i. We now consider two independent sets in G, namely,

$$I_1 = \{x : x \in I(L_i) \text{ for some } i \text{ even}\}$$

$$I_2 = \{x : x \in I(L_i) \text{ for some } i \text{ odd}\}$$

Clearly,

$$|I_1| + |I_2| \geq \sum_i |I(L_i)| \geq 1 + d(u,v) .$$

If $d(u,v) \geq 2s$, one of I_1 and I_2 has at least $s + 1$ vertices and therefore $\alpha(G) \geq s + 1 > \overline{D}(G)$. Suppose $d(u,v) = 2s - 1$ and $\alpha(G) \leq s$. We have $2s \geq \sum_i |I(L_i)| \geq 1 + d(u,v) = 2s$. This implies, for all i, the induced subgraph on L_i is complete. We can easily estimate $\overline{D}(G)$. Among all such G, $\overline{D}(G)$ is maximized when L_i has exactly one element except for $i = 0$ or $2s - 1$. It can be easily calculated that $\overline{D}(G) < R \leq \alpha(G)$. This completes the proof of Claim 3.

From Claim 3, we may assume $d(u,v) \leq 2\lfloor\overline{D}\rfloor - 2$ for all u and v. From Claim 2, we may assume $f(v) > \overline{D}(G) - 1$ for all v. We are now ready to check the last case.

Claim 4. If $f(w) > \overline{D}(G) - 1$, then $\alpha(G) > \overline{D}(G)$.

Proof. Without loss of generality, we may assume that w is chosen so that $t = \max\{d(w,v) : v \in G\}$ is minimized among all w with $f(w) \leq \overline{D}(G)$ and the number of v with $d(w,v) = t$ is as small as possible. Since w is a cut-vertex, let A denote a connected component in $G - \{w\}$ containing a vertex u with $d(u,w) = t \geq s = \lfloor\overline{D}\rfloor$. Let B denote the union of the remaining connected components. That is, $B = G - A - \{w\}$. From Claim 3, it follows that $d(w,x) \leq s - 2 - d(u,w) \leq s - 2$ for any x in B.

Choose a vertex z on the shortest path $p(u,w)$ joining u and w such that $d(z,u) = s - 1$. We consider the following two subcases:

Subcase (a). Suppose there is a vertex v for which $d(z,u) = s + 1$.
Clearly, v is not in B, since otherwise

$$d(u,v) = d(u,z) + d(z,w) + d(w,v)$$

$$= (s - 1) + (s + 1),$$

which contradicts the assumption that $d(u,v) \leq 2s - 2$. Let M_i denote the vertices x in A with $d(x,w) = i$. Suppose z is in M_r. We say M_i is complete if the induced subgraph on M_i is complete. Let a denote the least number such that M_{r+a} is complete. Clearly, v is not in M_j with $j \geq r + a$ since for any y in such M_j, $d(z,y) \leq 1 + d(z,u) \leq s$. So v is in some M_j with $j < r + a$. Let b denote the least number such that M_{r-b} is complete. If v is in M_j with $j \geq r - b$, we have $d(z,u) \leq b + 1 + (a + b - 1)$. We consider two independent sets I'_1 and I'_2 where

$$I'_1 = \{x : x \in I(M_i) \text{ for some even } i\}$$

$$I'_2 = \{x : x \in I(M_i) \text{ for some odd } i\} \cup I(B).$$

Then

$$|I_1'| + |I_2'| \geq 1 + \sum_i I(M_i)$$
$$\geq 1 + 2(a + b - 1) + s - a + 1$$
$$\geq s + 2a + b \geq 2s + 1\,.$$

Therefore $\alpha(G) \geq s + 1 > \overline{D}(G)$. We may assume v is in M_j with $j < r - b$. Then

$$d(u,v) \geq d(u,z) + d(z,v) - 1$$
$$\geq 2s - 1\,,$$

since $p(u,v)$ contains a vertex in M_{r-b}. From Claim 3, we get $\alpha(G) \geq \overline{D}(G)$.

Subcase (b). For any vertex $v, d(z,v) \leq s$.

Since $f(z) < s \leq \overline{D}(G)$, we have $t = s$ and z is adjacent to w. From Claim 2, we may assume $f(z) > \overline{D}(G) - 1$. Let v denote a vertex with $d(z,v) = s$. $d(v,w)$ is either $s - 1$ or s. If $d(v,w) = s - 1$, we can consider M_i as in Case (a). Each M_i, $i \neq 0$ or s, contains at least two independent vertices one of which is in a shortest path joining w and u and the other in a shortest path joining w and v. In particular $|I(M_1)| \geq 3$. Therefore $|I_1'| + |I_2'| \geq 2s + 1$ which implies $\alpha(G) \geq s + 1 > \overline{D}(G)$. We may assume $d(v,w) = s$. This implies the set of all v with $d(z,v) = s$ has fewer elements than that of w that is a contradiction. This completes the proof of Claim 4.

We remark that $\alpha(G) = \overline{D}(G)$ can only occur when $G - \{w\}$ is connected for every w with $f(w) \leq \overline{D}(G)$. In fact, in the proof of Claim 1, $\alpha(G) = \overline{D}(G)$ only if $\alpha(G) = \alpha(G - \{w\}) = \overline{D}(G\{w\})$. By the induction assumption, $G - \{w\}$ is a complete graph. Therefore $\alpha(G) = 1$ and G is a complete graph. This completes the proof of the theorem.

3. CONCLUDING REMARKS

The study of the average distance has recently attracted the attention of many researchers. In particular, Winkler [7] proposed the following conjecture:

For any given graph with average distance $\overline{D}$, there exists a vertex whose deletion results in a graph with the average distance no more than $4/3\overline{D}$.

It is not difficult to prove the existence of a vertex whose removal results a graph with the average distance $2\overline{D}$. On the other hand, for a cycle C_n, the deletion of any vertex increases the average distance by a factor of $4/3$. Recently Bienstock and Györi [1] proved that, for $n = |V(G)|$ sufficiently large, there is a vertex in G whose removal results in a graph with average distance $(\frac{4}{3} + o(1))\overline{D}(G)$.

Another interesting direction is the algorithmic complexity of determining the average distance of a graph. Of course, the average distance can be calculated by first finding the distances of all pairs of vertices. The best known algorithm for finding all distances requires time $O(n^2 + ne)$, or for dense graphs, $O(n^3(\log \log n/\log n)^{1/3})$, due to M. L. Fredman [6] (also see [7]). Is the problem of determining the average distance easier than or just as hard as the all distances problem? The best known algorithm for determining the diameter of a graph requires time $O(ne)$ or $O(\log nf(n))$ where $f(n)$ is the complexity for matrix multiplication (the current champion for matrix multiplication [D. Coppersmith, private communication] has complexity $O(n^{2.38})$. The problem of determining the diameter and all distances as well as the problem of determining the average distance remain an open-ended challenge in this area.

REFERENCES

[1] D. Bienstock and E. Györi, Average distance with removed elements. Preprint.

[2] B. Bollobás, *Extremal Graph Theory,* Academic Press, New York (1978).

[3] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications.* North Holland, (1976).

[4] F. R. K. Chung, Diameters of communications networks. *Mathematics of Information Processing,* AMS Publications (1984) 1–18.

[5] S. Fajtlowicz and W. A. Waller, On two conjectures of graffiti. Preprint (1986).

[6] M. L. Fredman, New bounds on the complexity of the shortest path problem. *SIAM J. Comput.* **5** (1976) 83–89.

[7] R. E. Tarjan, *Data Structures and Network Algorithm.* SIAM Publications (1983).

[8] Peter M. Winkler, Mean distance and the 4/3 conjecture. *Congressus Numerantium* (to appear).