The difference between a graph and its square

Donald Aingworth*   Rajeev Motwani†
Department of Computer Science
Stanford University
Stanford, CA 94305

Frank Harary
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003–8001

Dedicated to the memory of Uncle Paul Erdös, 1913–1996.

Abstract

The square of a graph G of order n nodes and size m edges is denoted G^2. It is obtained from G by adding new edges joining all pairs of nodes at distance 2 in G. The 2-step graph of G, denoted $G^{[2]}$, has the same node set as G and contains just those edges which were added to G to construct G^2. We prove that the size of $G^{[2]}$, $\left|E\left(G^{[2]}\right)\right| = \left|E\left(G^2\right)\right| - |E(G)|$, is at least $n-2$ when G^2 is not complete, and that this bound is sharp.

1 Introduction

Let $G = G(V, E)$ be a connected graph with n nodes and m edges. Then G^2, the *square* of G, is the graph on V where an edge uv is present if and only if the distance $d(u, v)$ in G is at most 2. This unary operation on G was introduced by Harary and Ross [3]. We remark that Motwani and Sudan [4] have established the NP-hardness of deciding whether a given graph H is the square of any graph G. A related unary operation, introduced by Exoo and Harary [1], is that of the *2-step graph* $G^{[2]}$. It has node set V and

*Supported by an NSF Graduate Fellowship and NSF NYI Award CCR-9357849.

†Supported by an Alfred P. Sloan Research Fellowship, an IBM Faculty Development Award, an OTL grant, and NSF Young Investigator Award CCR-9357849, with matching funds from IBM, Schlumberger Foundation, Shell Foundation, and Xerox Corporation.

edge set $E(G^{[2]}) = E(G^2) - E(G)$. Our results are concerned with the extremal properties of the 2-step graph. We follow in general the graph notation and terminology of Harary [2].

We establish a sharp lower bound on the number of edges added to G when G^2 is formed.

Theorem 1 *If G^2 is not a complete graph, then the number of edges in $G^{[2]}$ is at least $n-2$.*

Before we present the proof of the theorem in Section 3, we discuss the tightness of this result.

2 An Extremal Family of Graphs

To see that this result is sharp, consider the case where G is the path P_n containing $n-1$ edges. It is easy to verify that the size of P_n^2 is exactly $2n-3$, which realizes the lower bound in Theorem 1. This is not satisfactory in general since we would like extremal graphs with an arbitrary number m of edges. However, the example can be generalized by considering the "dumbbell graph" $D_{r,s,t}$ which is obtained by taking the path graph P_t and adding $r+s$ nodes such that:

- the first r additional nodes form a clique of order $r+1$ with the left endpoint of P_t, and
- the remaining s nodes form a clique of order $s+1$ with the right endpoint of P_t.

Note that the additional nodes are only adjacent to the other nodes in their respective cliques. Figure 1 shows the dumbbell graph $D_{2,3,5}$.

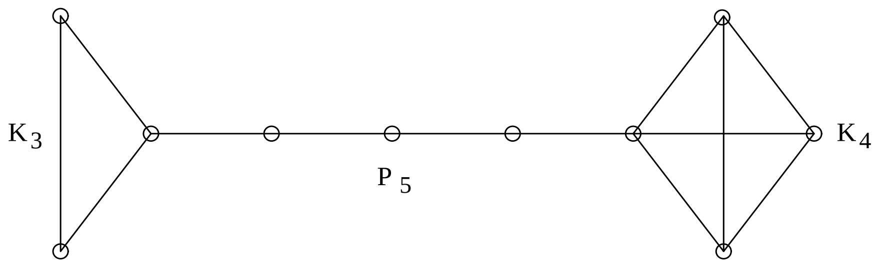

Figure 1: The dumbbell graph $D_{2,3,5}$

We define this graph for all $r, s \geq 0$ and $t > 1$ but note that $D_{1,s,t} = D_{0,s,t+1}$, $D_{r,1,t} = D_{r,0,t+1}$, and $D_{1,1,t} = D_{0,0,t+2} = P_{t+2}$. The order of

$D_{r,s,t}$ is $n = r + s + t$. It is easy to verify that the number of additional edges in its square is given by $r + (t-2) + s = n - 2$.

The point of this construction is that it provides more flexibility in the choice of the number of edges in the extremal graph. To see this, first observe that the total number of edges in $D_{r,s,t}$ is given by

$$m = \binom{r+1}{2} + (t-1) + \binom{s+1}{2},$$

which implies that for $r, s \geq 2$,

$$m - (n-1) = \binom{r}{2} + \binom{s}{2}.$$

In other words, we can obtain an extremal graph for any density of edges which has the property that the excess of the number of edges over that required for connectedness can be expressed as the sum of the number of edges in the two cliques.

It seems unlikely that extremal graphs exist for all other values of m. The problem is that the tight examples are very similar to paths, and that at many values of m and n, it seems that you must deviate sufficiently from a path as to force the existence of additional edges in $G^{[2]}$. Paths have the nice property that every node has only two other nodes at distance 2. (If there are more than two nodes at distance 2, to keep the number of new edges low, some of the nodes at distance 2 must also be neighbors. This process continues until you are left with cliques, as in the dumbell graph.) A specific example is given by $m - n = 2$, where there are only 3 edges beyond those needed to connect the graph. A case analysis here shows that $n - 2$ cannot be achieved here.

We leave open the following question.

Question 1 *For which values of n and m is the bound $n-2$ realized?*

3 The Proof

To establish Theorem 1, it will be convenient to work with the following reformulation.

Theorem 2 *If the complement graph $\overline{G}$ of G is connected, then the size of the 2-step graph $G^{[2]}$ is at least $n-2$.*

The following lemma shows why the two formulations are equivalent.

Lemma 1 *If $\overline{G}$ is not connected, then G^2 is complete.*

Proof: If $\overline{G}$ is disconnected, then we can partition its nodes into 2 sets that have no edges between them in $\overline{G}$. This implies that G contains a complete bipartite spanning subgraph between these two sets of nodes, and hence its square will be complete. $\square$

The idea behind the proof of Theorem 1 is to show that the graph of added edges, $G^{[2]}$, has at most two components, which clearly implies that at least $n-2$ edges are added. We will do this by showing that if there are more than 2 components in the 2-step graph, then $\overline{G}$ must be disconnected, contradicting our assumption.

Lemma 2 *If there exists a triangle in G which has its nodes in three separate components of $G^{[2]}$, then each node x of the triangle is adjacent in G to every other node in the graph that is not in the component of $G^{[2]}$ containing x.*

Proof: By hypothesis, G has a triangle whose nodes are in 3 connected components of $G^{[2]}$. Call the components X, Y, and Z, and the respective nodes x, y, and z. First we will show that if a node w is adjacent to the triangle in G then w is adjacent in G to all of the triangle nodes lying in separate components of $G^{[2]}$. Then we shall show that all nodes must be adjacent to any such triangle in G.

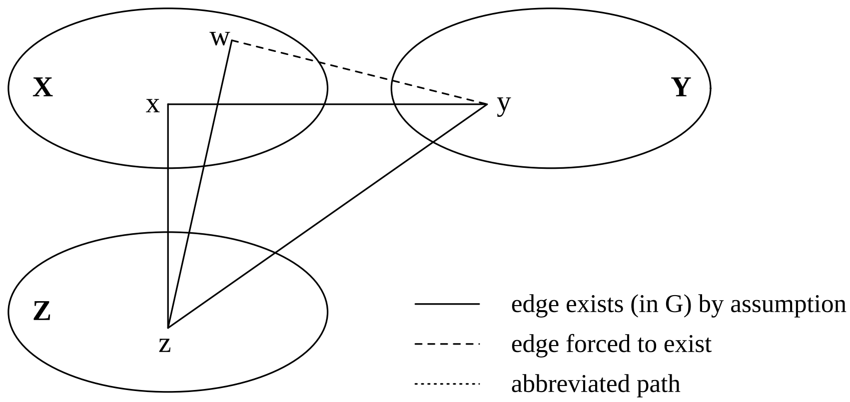

Figure 2: w is in X

Assume that w is adjacent to a triangle node, say z. There are four possibilities: $w \in Z$, $w \in Y$, $w \in X$, or w is in a fourth component. If $w \in Z$, then the existence of edges xz and zw in G implies that if edge xw is not in G then it must be in $G^{[2]}$. Since that edge would link two separate connected components, it must exist in G. A similar argument applies to

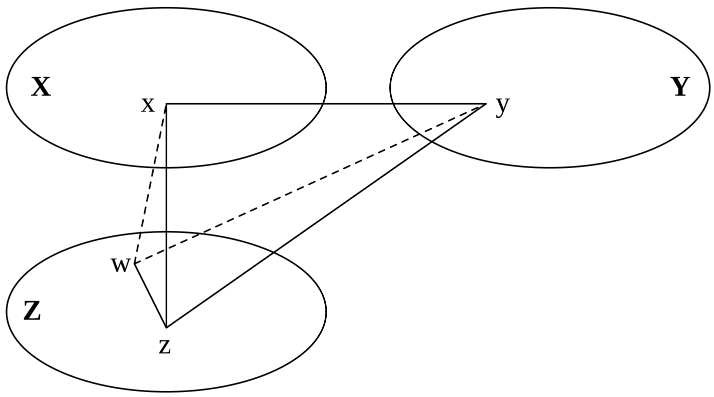

Figure 3: w is in Z

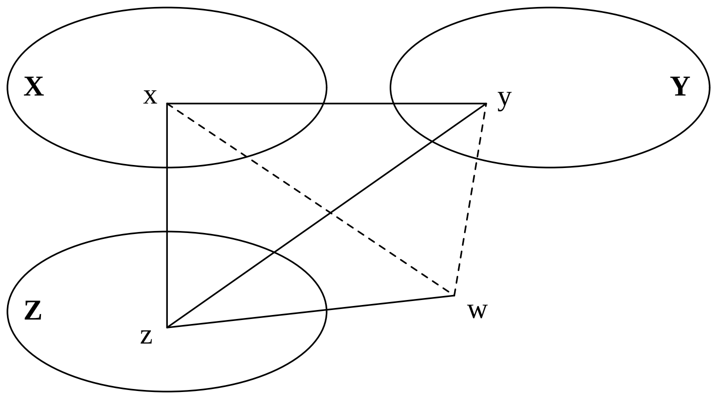

Figure 4: w is outside of X, Y, Z

yw and the other cases. We illustrate this proof in Figures 2, 3, and 4. Note that all of these cases also imply the existence of a triangle involving w and two other components.

Suppose now that w is at distance 2 from the triangle, and let v be the node between the triangle and w. By the above argument, we know that there is a triangle in G with v and two of x, y, z. An analysis identical to that above (using the triangle that we just found) shows that w is adjacent to at least one of x, y, z. But this would imply that w was actually at distance 1 from the triangle, a contradiction. Therefore there are no nodes at distance 2 from the triangle, and so the connectedness of G implies that all nodes in G must be adjacent to the triangle. Now we are in the situation that we analyzed before (w adjacent to the triangle), so an identical argument implies the existence of a triangle involving w and two other components. □

It is fairly easy to bootstrap the preceding lemma to the next preliminary result.

Lemma 3 *If there exists a triangle in G which has its nodes in three separate components of $G^{[2]}$, then every edge that between two components of $G^{[2]}$ must exist.*

Such a situation would disconnect G^2, and prove the theorem, so now it suffices to show the existence of such a triangle.

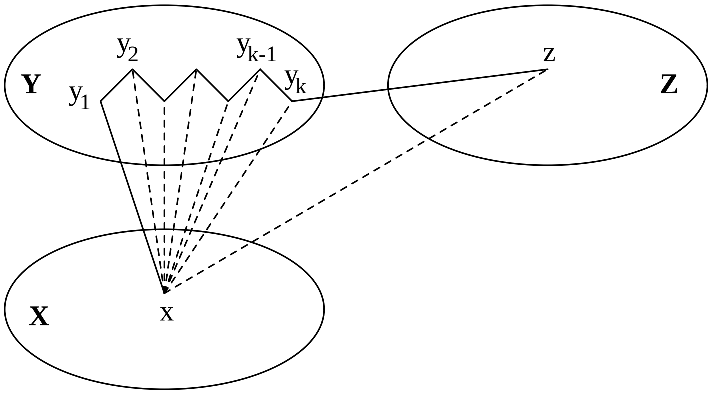

Figure 5: Paths force triangles

Lemma 4 *If $G^{[2]}$ has more than two connected components, then there must exist a triangle in G whose nodes are in three components of $G^{[2]}$.*

Proof: Since G is connected, it must contain a path on three 3 components of $G^{[2]}$. Take a shortest such path. In particular, this path must start at some node x in component X, cross to node y_1 in component Y, proceed through nodes $y_2, ..., y_k$, all in component Y, and then cross to node z in component Z. The existence of edges xy_1 and y_1y_2 necessitate the existence of edge xy_2, for otherwise the components would be joined in $G^{[2]}$. This continues until we find that xy_k must also exist in G, and similarly that xz must exist in G. So x, y_k, z is a triangle that we are seeking (see Figure 5). □

Combining the lemmas, we see that if $G^{[2]}$ has more than two components, then $\overline{G}$ must be disconnected. This establishes Theorem 1. □

References

[1] G. Exoo and F. Harary. Step graphs. *J. Combin. Inform. System Sci.* 5(1980):52–53.

[2] F. Harary. *Graph Theory.* Addison-Wesley, Reading (1969).

[3] F. Harary and I. Ross. The square of a tree. *Bell System Technical J.* 39(1960):641–647.

[4] R. Motwani, and M. Sudan. Computing Roots of Graphs is Hard. *Discrete Applied Mathematics*, 54(1994):81–88.