The difference between a graph and its square

Donald Aingworth* Rajeev Motwani†
Stanford University

Frank Harary
New Mexico State University

Abstract

Frank will write the abstract at the end.

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 (u, v) is present if and only if vertices u and v are at distance at most 2 in G. This notion was defined and first studied in the article by Harary and Ross [3]. A related notion introduced by Exoo and Harary [1] is that of the *2-step graph*, $G_{[2]}$, which is defined to be $G^2 - G$. 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 the following lower bound on the number of edges added to G when G^2 is formed.

*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.

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 graph P_n containing $n-1$ edges. It is easy to verify that the number of edges in 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 $r+s$ vertices such that:

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

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

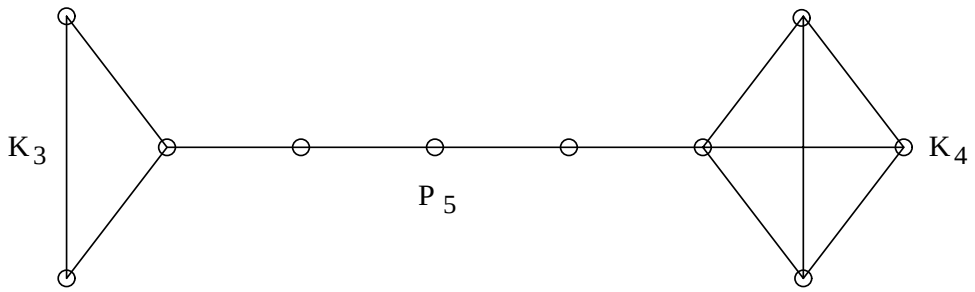

Figure 1: The dumbbell graph $D_{2,3,5}$

We define this graph for all $r, s \geq 0$ and $t \geq 1$ 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 total number of

vertices in $D_{r,s,t}$ 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 gives us 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 far from a path as to force the existance of additional edges in $G_{[2]}$. Paths have the nice property that every node has only two other nodes at disance 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 when $m - n = 2$, 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.

3 The Proof

To establish Theorem 1, it will be convenient to work with the following reformulation. Suppose the complement graph $\overline{G}$ is connected, then the number of edges in $G_{[2]}$ 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 vertices), and hence its square will be complete. □

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, 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 assumption, G has a triangle that crosses 3 connected components of $G_{[2]}$. Call the components X, Y, and Z, and the respective triangle nodes x, y, and z. Given any other node w, we proceed by induction on the length of a shortest path in G from w to the triangle.

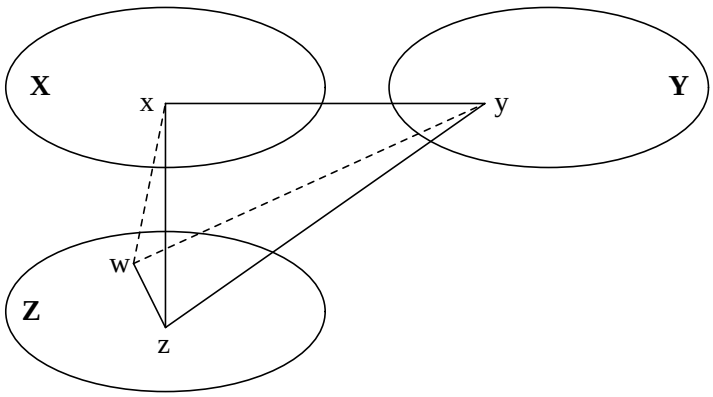

Figure 2: w is in Z

In the base case, w is adjacent to a triangle node, say z. There are four possibilities: w is in Z, w is in Y, w is in X, or w is in a fourth component. If w is in Z, then the existence of edges (x,z) and (z,w) implies that if (x,w) did not exist in G then it would exist in $G_{[2]}$. Since that edge would link two separate connected components, it must exist in G. A similar

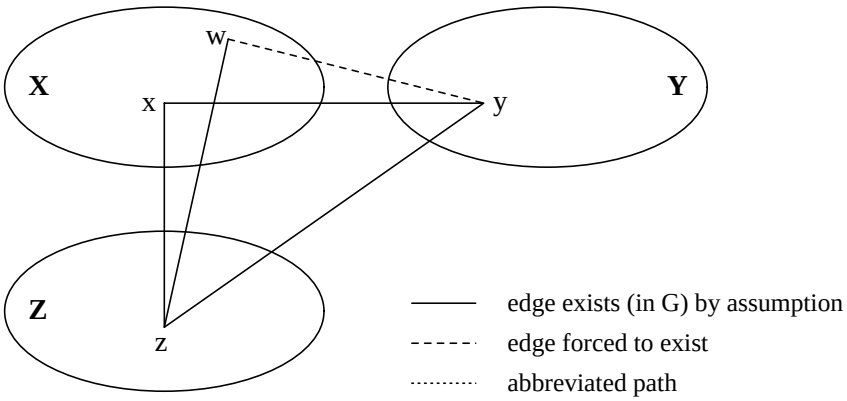

Figure 3: w is in X

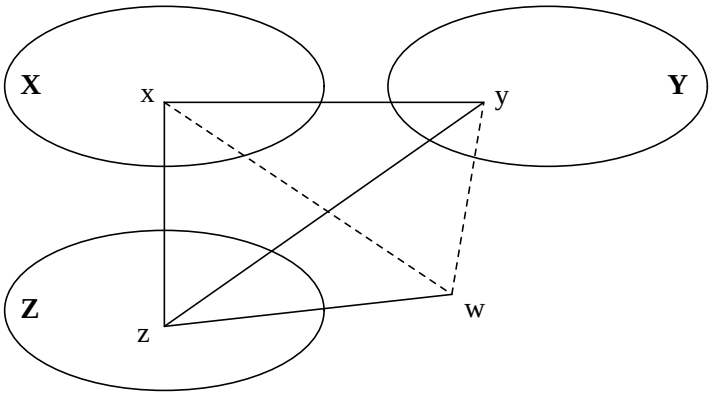

Figure 4: w is outside of X, Y, Z

argument applies to (y, w), and in 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.

We take as the inductive hypothesis that this is true for all paths of length less than k, and let v be the first node on the path from the triangle to w. By our base case, we know that there is a triangle in G with v and two of x, y, z, so by induction w is connected to at least one of x, y, z. This both contradicts the assumption that the shortest path is of length k and shows, by an analysis identical to the base case, that the inter-component edges exist. □

It is fairly easy to bootstrap the preceding lemma to establish the following.

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.

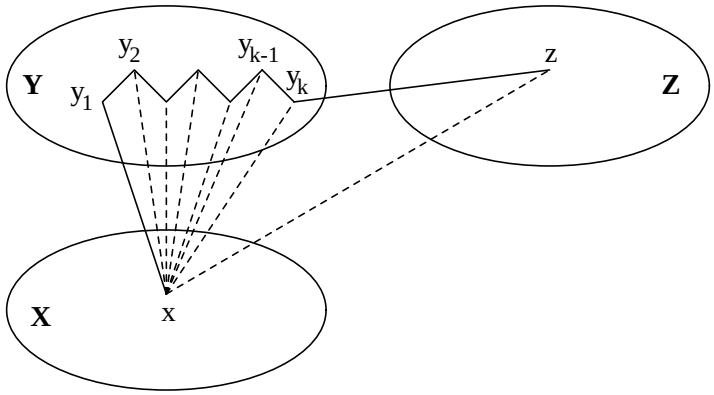

Figure 5: Paths force triangles

Lemma 4 *If $G_{[2]}$ has more than two connected components, then there must exist a triangle in G that crosses three components.*

Proof: Since G is connected, there must exist a path that crosses 3 components in $G_{[2]}$. Take the 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 (x, y_1) and (y_1, y_2) necessitate the existence of (x, y_2) (for otherwise the components would be joined in $G_{[2]}$). This continues until we find that (x, y_k) must also exist in G, and similarly that (x, z) must exist in G. So x, y_k, z is the triangle that we are looking for. This is shown in Figure 5. □

Combining these 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.