Graphs with Large Variance

Yair Caro and Raphael Yuster
Department of Mathematics
University of Haifa-ORANIM
Tivon 36006, Israel
e-mail: zeac603@uvm.haifa.ac.il
e-mail: raphy@math.tau.ac.il

ABSTRACT. For a graph G, let $Var(G)$ denote the variance of the degree sequence of G, let $sq(G)$ denote the sum of the squares of the degrees of G, and let $t(G)$ denote the number of triangles in G and in its complement. The parameters are related by: $Var(G) = sq(G)/n - d^2$ where d is the average degree of G, and $t(G) = \binom{n}{3} + sq(G)/2 - m(n-1)$. Let $Var(n)$ denote the maximum possible value of $Var(G)$ where G has n vertices, and let $sq(n,m)$ and $t(n,m)$ denote the maximum possible values of $sq(G)$ and $t(G)$, respectively, where G has n vertices and m edges. We present a polynomial time algorithm which generates all the graphs with n vertices and m edges having $sq(G) = sq(n,m)$ and $t(G) = t(n,m)$. This extends a result of Olpp which determined $t(n,m)$. We also determine $Var(n)$ precisely for every n, and show that

$$Var(n) = \frac{q(q-1)^2}{n}(1-\frac{q}{n}) = \frac{27}{256}n^2 - O(n),$$

where $q = [3n/4]$, (if $n \equiv 2 \bmod 4$ the rounding is up) thereby improving upon previous results.

1 Introduction

All graphs considered here are finite, undirected, and have no loops or multiple edges. For the standard graph-theoretic notations the reader is referred to [2]. For a graph $G = (V, E)$, let $sq(G)$ denote the sum of squares of the degrees of G, namely $sq(G) = \sum_{v \in V} {d_v}^2$, where d_v denotes the degree of vertex v. The *variance* of G, denoted $Var(G)$, is the second central moment, or variance, of the degrees of G, namely $Var(G) = \frac{1}{n}\sum_{v \in V}(d_v -$

$d)^2$ where $d = 2|E|/n$ is the average degree. It is a routine exercise to show that:

$$Var(G) = \frac{sq(G)}{n} - d^2. \tag{1}$$

Thus, if G_1 and G_2 are two graphs with the *same* number of vertices and the *same* number of edges, then $Var(G_1) > Var(G_2)$ if and only if $sq(G_1) > sq(G_2)$.

The definitions of $Var(G)$ and $sq(G)$ raise the following two extremal combinatorial problems:

1. Determine $Var(n)$, the maximum possible variance of a graph with n vertices.

2. Determine $sq(n, m)$ the maximum possible value of $sq(G)$ where G is a graph with n vertices and m edges. (By (1) this is also the maximum possible variance of a graph with n vertices and m edges. Also note that, trivially, $sq(n) = \max_m sq(n, m) = n(n-1)^2$ by considering K_n).

Note that, by (1) we have that

$$Var(n) = \max_{0 \le m \le \binom{n}{2}} \frac{sq(n, m)}{n} - 4\frac{m^2}{n^2}. \tag{2}$$

The function $sq(n, m)$ is related to the Ramsey-type problem of determining $t(n, m)$, the maximum number of monochromatic triangles in an m edge-coloring of K_n (an m edge-coloring of K_n is a red-blue coloring in which precisely m edges are colored red). It is not difficult to show (see Lemma 2.2 in the next section) that

$$t(n, m) = \binom{n}{3} + \frac{sq(n, m)}{2} - m(n-1). \tag{3}$$

The functions $sq(G)$, $sq(n, m)$ and $t(n, m)$ have been considered by several researchers. Some results on $sq(G)$ consider G to be restricted to some family of graphs, such as planar and outerplanar graphs [9, 4]. In fact, in [9] it is proved that $sq(G) \le 2n^2 + 12n - 44$ when G is a planar graph with n vertices, and that this result is sharp. Thus, by (1), $Var(G) \le 2n - 24 + 100/n - 144/n^2$ whenever G is a maximal planar graph with n vertices, and this is sharp. Several results give upper bounds on $sq(G)$ for general graphs. Székely, Clark and Entringer [8] have proved that if the degree sequence of G is $d_1, \ldots, d_n$ then

$$sq(G) \le (\sqrt{d_1} + \ldots + \sqrt{d_n})^2. \tag{4}$$

By applying Cauchy-Schwarz, it follows from their result that $sq(n,m) \leq 2mn$ and, consequently, $Var(n) \leq n^2/4$. An improvement upon the value of $sq(n,m)$ was obtained by de Caen [3], who proved that for $n \geq 2$,

$$sq(n,m) \leq m(\frac{2m}{n-1} + n - 2). \qquad (5)$$

This result yields $Var(n) \leq (n-1)(n-2)/8$. The inequalities (4) and (5) are incomparable since it is easy to construct graphs for which (4) is better than (5), and vice versa. Note however, that (5) is always less than $2mn$ (unless $m = 0$) and thus, (5) is a better bound for $sq(n,m)$. Another relevant paper, which generalizes the results of de Caen is [5]. Goodman [6] has made a conjecture about $t(n,m)$. This conjecture has been proved by Olpp [7], which determined $t(n,m)$, and also determines at least one m edge-coloring of the edges of K_n which has $t(n,m)$ monochromatic triangles. It follows from Olpp's result, and from (3) that $sq(n,m)$ is also determined and that one can always generate a graph G with n vertices and m edges having $sq(G) = sq(n,m)$. Olpp's result, however, does not determine the set of *all* non-isomorphic m edge-colorings of K_n with $t(n,m)$ monochromatic triangles (in fact, it determines at most two non-isomorphic colorings). Our first result in this paper is a polynomial time algorithm which generates *all* non-isomorphic m edge-colorings with $t(n,m)$ monochromatic triangles. In fact, we determine all graphs G with n vertices and m edges having $sq(G) = sq(n,m)$:

Theorem 1.1. *Let $F(n,m)$ denote the set of all graphs G with n vertices and m edges having $sq(G) = sq(n,m)$. The set of all graphs in $F(n,m)$ can be computed in $O(|F|n^2)$ time.*

Note that by coloring the edges of $G \in F(n,m)$ red, and the edges of $\overline{G}$ (the complement of G) blue, we get by (3) an m edge-coloring of K_n with $t(n,m)$ monochromatic triangles. Thus, $F(n,m)$ also determines all non-isomorphic m edge-colorings of K_n with $t(n,m)$ monochromatic triangles. It is interesting to note that the set $F(n,m)$ may contain several non-isomorphic graphs. For example, running the algorithm for $n = 9$ and $m = 18$ we obtain $sq(9,18) = 192$ and $F(9,18)$ consisting of exactly six graphs whose degree sequences are: $(0,0,3,5,5,5,6,6,6)$, $(2,2,2,3,3,3,5,8,8)$, $(0,3,3,3,3,3,7,7,7)$, $(1,1,1,5,5,5,5,5,8)$, $(0,0,4,4,4,6,6,6,6)$ and $(2,2,2,2,4,4,4,8,8)$ (it is trivial to reconstruct each of these graphs from the degree sequences). However, as can be seen from Olpp's Theorem (stated in Lemma 2 in the next section) his proof only yields the first two constructions.

Our motivation to further introduce $Var(G)$ and $Var(n)$ is the recent paper of Albertson [1] who dealt with certain ways to measure how far a graph is from being a regular graph. He introduced (with relation to a certain Ramsey type problem) the parameter $irr(G)$, called the *irregularity* of G, defined by $\sum |d_x - d_y|$ where the sum is taken over all edges (x, y) of G. It is proved in [1] that for every graph with n vertices, $irr(G) \leq 4n^2/27$. Clearly, if G is regular then $irr(G) = 0$. However, the converse is not true, as can be seen from graphs with regular connected components of different degrees. Also, it is not true that $irr(G) = irr(\overline{G})$ where $\overline{G}$ is the complement of G. With these facts in mind, $Var(G)$ seems a natural parameter to measure the irregularity of graphs and it is easy to see that:

1. $Var(G) = Var(\overline{G})$.
2. $Var(G) = 0$ if and only if G is regular.

By utilizing Olpp's result, together with (3), some calculus, and a symmetry argument, we are able to determine $Var(n)$, and two n-vertex graphs G with $Var(G) = Var(n)$:

Theorem 1.2.

$$Var(n) = \frac{q(q-1)^2}{n}(1 - \frac{q}{n})$$

Where $q = [3n/4]$, (in case $n = 2 \bmod 4$ the rounding is up). The n-vertex graph G consisting of a K_q and $n - q$ isolated vertices, has $Var(G) = Var(n)$, and thus $Var(\overline{G}) = Var(n)$ as well (For $n > 1$ G and $\overline{G}$ differ).

The rest of this paper contains the proofs of Theorems 1.1 and 1.2 in the next section, as well as the proof of (3) and some concluding remarks and open problems in the final section.

2 Proof of the main results

We begin this section by describing Olpp's result concerning the number of Monochromatic triangles in an m edge-coloring of K_n. For a red-blue edge-coloring c of K_n, denote by $t(c)$ the number of monochromatic triangles. Before we state Olpp's result we need to define two graphs. Let u and v be two integers which satisfy $m = \binom{v}{2} + u$ where $0 \leq u \leq v - 1$. Note that for every $m \geq 0$, v and u are uniquely defined. Let $H_1(n, m)$ be the n-vertex graph which is composed of a clique on v vertices and, if $u > 0$, there is a unique vertex outside the clique which joins exactly u vertices of the clique. (The remaining vertices, if there are any, are isolated). Note that $H_1(n, m)$ has exactly m edges. Let q and p be two integers which satisfy $m = \binom{q}{2} + q(n - q) + p$ where $1 \leq p \leq n - q - 1$. Note that for every $m > 0$, p and q are uniquely defined. Let $H_2(n, m)$ be the unique n-vertex graph composed of q vertices of degree $n - 1$, and the subgraph induced on the remaining $n - q$ vertices has exactly p edges which all share a common endpoint. Note that $H_2(n, m)$ has exactly m edges. Olpp has proved the following:

Lemma 2.1. *[Olpp [7]] Let c_1 be an m edge-coloring of K_n where the edges colored red are defined by $H_1(n,m)$. Let c_2 be an m edge-coloring of K_n where the edges colored red are defined by $H_2(n,m)$. Then $t(n,m) = \max\{t(c_1)\ ,\ t(c_2)\}$.*

Note that Lemma 2.1 also supplies a formula for $t(n,m)$ since $t(c_1)$ and $t(c_2)$ can be explicitly computed.

We now prove (3), which shows that $sq(n,m)$ and $t(n,m)$ are linearly correlated.

Lemma 2.2.

$$t(n,m) = \binom{n}{3} + \frac{sq(n,m)}{2} - m(n-1).$$

Proof: Let G be any graph with n vertices and m edges. Let c_G be the m edge-coloring of K_n where the edges of G are the ones colored red. It suffices to show that $t(c_G) = \binom{n}{3} + sq(G)/2 - m(n-1)$. Let y be the number of monochromatic copies of P_3 (the path with two edges) in c_G. Every monochromatic triangle contains three monochromatic copies of P_3, and every non-monochromatic triangle contains only one monochromatic P_3, thus $y = 3t(c_G) + (\binom{n}{3} - t(c_G)) = 2t(c_G) + \binom{n}{3}$. It is easy to compute y in terms of the degrees of G, since, clearly: $y = \sum_{x\in V}\binom{d_x}{2} + \binom{n-1-d_x}{2}$. Using the fact that $\sum_{x\in V} d_x^2 = sq(G)$ and $\sum_{x\in V} d_x = 2m$ we obtain that $y = sq(G) + 3\binom{n}{3} - 2m(n-1)$. Therefore,

$$2t(c_G) + \binom{n}{3} = y = sq(G) + 3\binom{n}{3} - 2m(n-1)$$

The result now follows. □

By Lemma 2.2 we see that if $G \in F(n,m)$ (recall that $G \in F(n,m)$ if it has n vertices, m edges and $sq(G) = sq(n,m)$) then $t(c_G) = t(n,m)$, and vice-versa. Now, since $sq(H_1(n,m)) = 2m(v-1) + u(u+1)$ and $sq(H_2(n,m)) = q(n-1)^2 + (4q+p+1)p + q^2(n-q)$ we obtain the following corollary:

Corollary 2.3. *Let n be a positive integer and let $1 \le m \le \binom{n}{2}$. Then*

$$sq(n,m) = \max\{2m(v-1)+u(u+1)\ ,\ q(n-1)^2+(4q+p+1)p+q^2(n-q)\}.$$

where q,p,u,v are defined as in the above definition of $H_1(n,m)$ and $H_2(n,m)$. If the maximum is obtained by the first expression then $sq(H_1(n,m)) = sq(n,m)$. If the maximum is obtained by the second expression then $sq(H_2(n,m)) = sq(n,m)$.

Corollary 2.3 supplies an accurate formula for computing $sq(n,m)$, and, furthermore, it establishes at least one graph (namely, $H_1(n,m)$ or $H_2(n,m)$) which belongs to $F(n,m)$. However, Olpp's proof does not characterize all graphs in $F(n,m)$.

Example: Consider the case $n = 9$ and $m = 18$. We have $v = 6$, $u = 3$, $q = 2$, $p = 3$. Thus, $sq(9,18) = \max\{192\ ,\ 192\} = 192$. We have also that $F(9,18) \supset \{H_1(9,18), H_2(9,18)\}$. The degree sequence of $H_1(9,18)$ is $(0,0,3,5,5,5,6,6,6)$ and the degree sequence of $H_2(9,18)$ is $(2,2,2,3,3,3,5,8,8)$. However, as shown in the introduction, there are at least four other graphs in $F(9,18)$. In fact, as we shall see, $|F(9,18)| = 6$.

Our goal in Theorem 1.1 is to present an *efficient* procedure which establishes all graphs in $F(n,m)$. In fact, we shall give a recursive formula for obtaining $sq(n,m)$ and $F(n,m)$ from the values of $sq(n-1,z)$ and $F(n-1,z)$ for some z, thus, by using a routine dynamic programming approach, we can solve the problem.

Before proving Theorem 1.1 we need two lemmas. The first one determines $F(n,m)$ in case m is relatively small w.r.t. n. We use the notation E_k to denote a set of k isolated vertices.

Lemma 2.4. *If $0 \le m \le n-1$, then $sq(n,m) = m(m+1)$. Furthermore, if $m \ne 3$ then $F(n,m) = \{K_{1,m} + E_{n-1-m}\}$, and if $m = 3$ then $F(n,m) = \{K_{1,3} + E_{n-4}\ ,\ K_3 + E_{n-3}\}$.*

Proof: We shall prove the lemma by induction on m. For $m \le 4$ the lemma holds by direct verification, noting, in particular, that when $m = 3$, $sq(K_3 + E_{n-3}) = sq(K_{1,3} + E_{n-4}) = 12 = 3 \cdot 4$. Assume now that $m \ge 5$. Let (x,y) be an edge. Clearly $d_x + d_y - 1 \le m$. Let $G' = G \setminus \{e\}$. By the induction hypothesis, $sq(G') \le m(m-1)$. Thus, clearly,

$$sq(G) = sq(G') + d_x^2 + d_y^2 - (d_x - 1)^2 - (d_y - 1)^2 \le m(m-1)$$
$$+ 2(d_x + d_y - 1) \le m(m-1) + 2m = m(m+1)$$

where equality is achieved if and only if $d_x + d_y - 1 = m$ and $sq(G') = m(m-1)$. Thus, we must have $G' = K_{1,m-1} + E_{n-m}$, and the requirement $d_x + d_y - 1 = m$ now forces that $G = K_{1,m} + E_{n-1-m}$. □

Lemma 2.5. *Let $G \in F(n,m)$. Then, either G has an isolated vertex, or G has a vertex with degree $n-1$.*

Proof: Let G be a graph with n vertices and m edges having no isolated vertex, and no vertex with degree $n-1$. It suffices to show that there exists a graph G' with the same number of vertices and edges having $sq(G') > sq(G)$. Indeed, let x have maximum degree in G. Since $d_x < n-1$ there exists a vertex y such that (y,x) is not an edge of G. Since y is not isolated, there exists a vertex z such that (y,z) is an edge of G. Let G' be obtained from G by replacing the edge (y,z) with the edge (y,x).

$$sq(G') - sq(G) = ((d_x+1)^2 + (d_z-1)^2) - (d_x^2 + d_z^2) = 2(d_x - d_z + 1) \geq 2.$$

□

Given a graph G, denote by G^+ the graph obtained by adding to G an isolated vertex, and let G^* be the graph obtained by adding to G a new vertex which is connected to every vertex of G. The next lemma supplies a recursive formula for computing $sq(n,m)$ and $F(n,m)$ given the values of $sq(n-1,m)$ and $sq(n-1,m-n+1)$ and given $F(n-1,m)$ and $F(n-1,m-n+1)$.

Lemma 2.6. *Let $n \geq 1$ and m be integers satisfying $0 \leq m \leq \binom{n}{2}$. Then:*

1. *If $m \leq n-1$ then $sq(n,m) = m(m+1)$ and $F(n,m) = \{K_{1,m} + E_{n-1-m}\}$, unless $m = 3$ in which case $F(n,3) = \{K_{1,3} + E_{n-4}\ ,\ K_3 + E_{n-3}\}$.*
2. *If $m > \binom{n-1}{2}$ then*
$$sq(n,m) = sq(n-1, m-n+1) + (n-1)^2 + 4m - 3(n-1)$$
and
$$F(n,m) = \{G^* \mid G \in F(n-1, m-n+1)\}.$$
3. *If $n-1 < m \leq \binom{n-1}{2}$ then*
$$sq(n,m) = \max\{sq(n-1,m)\ ,\ sq(n-1,m-n+1) + (n-1)^2 + 4m - 3(n-1)\}.$$
If the maximum is obtained by the first number, then $F(n,m) = \{G^+ \mid G \in F(n-1,m)\}$. If the maximum is obtained by the second number then $F(n,m) = \{G^* \mid G \in F(n-1,m-n+1)\}$. If both numbers obtain the maximum, then $F(n,m)$ is the union of both of these sets.

Proof: The first case is identical to Lemma 2.4. Consider the second case where $m > \binom{n-1}{2}$. Let $H \in F(n,m)$. H cannot have an isolated vertex. Thus, by Lemma 2.5, H has a vertex x with degree $n-1$. Consider the graph G obtained from H by deleting x. Clearly, $H = G^*$, and $sq(H) = sq(G) + (n-1)^2 + 4m - 3(n-1)$. Now, if $K \in F(n-1, m-n+1)$ then $sq(K^*) = sq(n-1, m-n+1) + (n-1)^2 + 4m - 3(n-1)$. Since K^* has n vertices and m edges, we must have $sq(K^*) \leq sq(H)$. It follows that $G \in F(n-1, m-n+1)$ and $sq(H) = sq(n-1, m-n+1) + (n-1)^2 + 4m - 3(n-1)$, and thus, $sq(n,m) = sq(n-1, m-n+1) + (n-1)^2 + 4m - 3(n-1)$. Furthermore, for any $K \in F(n-1, m-n+1)$ we have that $K^* \in F(n,m)$. Now consider the third case, where $n-1 < m \leq \binom{n-1}{2}$. Let $H \in F(n,m)$.

According to Lemma 2.5, either H has a vertex of degree $n-1$, or H has an isolated vertex. If H has a vertex of degree $n-1$ then, as in the previous case, $sq(n,m) = sq(n-1, m-n+1) + (n-1)^2 + 4m - 3(n-1)$, and $F(n,m) \supset \{K^* \mid K \in F(n-1, m-n+1)\}$. If H has an isolated vertex x, then, denoting by G the graph obtained from H by deleting x we have $H = G^+$ and $sq(G) = sq(H)$. Now, if $K \in F(n-1,m)$ then $sq(K^+) = sq(n-1,m)$, and since K^+ has n vertices and m edges we must have $sq(K^+) \leq sq(H)$. It follows that $G \in F(n-1,m)$ and $sq(H) = sq(n,m) = sq(n-1,m)$. Furthermore, for any $K \in F(n-1,m)$ we have that $K^+ \in F(n,m)$. Obviously, if $F(n,m)$ contains both a graph with an isolated vertex and another graph with a vertex of degree $n-1$ then we must have $sq(n-1,m) = sq(n-1,m-n+1)+(n-1)^2+4m-3(n-1)$ and $F(n,m)$ is a union of both $\{K^* | K \in F(n-1, m-n+1)\}$ and $\{K^+ \mid K \in F(n-1,m)\}$. Note also that this union is *disjoint* since $m-n+1 \neq m$ (if $n=1$ the third case cannot occur). □

We can now turn Lemma 2.6 into an algorithm:

Proof of Theorem 1.1: We wish to generate all graphs in $F(n,m)$. We shall use the dynamic programming approach based on the recursive equation shown in Lemma 2.6. We begin by computing $sq(n,m)$ according to the lemma. We maintain a two dimensional array, denoted by SQ, whose dimensions are n by $m+1$. For $i = 1, \dots, n$ and for $j = 0, \dots, m$ $SQ(i,j)$ will hold $sq(i,j)$ at the end of the algorithm. We initialize by $SQ(1,0) = 0$. Now, for every $i = 1, \dots, n$ and for every $j = 0, \dots, m$, Lemma 2.6 shows us how to compute $SQ(i,j)$ in constant time. Namely, if $j \leq i-1$ we put $SQ(i,j) = j(j+1)$. Otherwise, if $j > \binom{i-1}{2}$ we put $SQ(i,j) = SQ(i-1, j-i+1) + (i-1)^2 + 4j - 3(i-1)$. Otherwise, we put $SQ(i,j) = \max\{SQ(i-1,j)\,,\, SQ(i-1,j-i+1)+(i-1)^2+4j-3(i-1)\}$. Clearly, all the elements in the array SQ can be filled in $O(nm)$ time. After completing the array, we are now ready to determine the graphs in $F(n,m)$. We use recursion, and assume the graphs are represented via their n by n adjacency matrices. We shall also use the fact that whenever we need to compute some $sq(x,y)$ we can do this in constant time (naturally, under the uniform cost model) using Corollary 2.3, since p, q, u and v can be computed by solving quadratic equations in x and y. We begin by computing $sq(n,m)$. If $m \leq n-1$ then we know by Lemma 2.6 that $F(n,m) = \{K_{1,m} + E_{n-m-1}\}$, unless $m = 3$ in which case we also have the graph $K_3 + E_{n-3}$, and we are done. Otherwise, if $sq(n,m) = sq(n-1,m)$ (this should be checked only if $m \leq \binom{n-1}{2}$) we recursively obtain the set of all graphs in $F(n-1,m)$, and to each graph in this set we add the n'th line and n'th column to its adjacency matrix, where all the entries are zero, since the n'th vertex represents an isolated vertex. This is $F(n,m)$. Now we check whether $sq(n,m) = sq(n-1,m-n+1) + (n-1)^2 + 4m - 3(n-1)$ (recall that it is possible that both cases occur). If so, we recursively obtain the set of all graphs in $F(n-1, m-n+1)$, and to each graph in this set we add the n'th line and n'th column to its adjacency matrix, where all the entries are 1 (except for the entry (n,n) which is 0), since the n'th vertex represents a vertex of degree $n-1$. The obtained graph is added to $F(n,m)$. Note that, as mentioned in the end of the proof of Lemma 2.6, each element added to $F(n,m)$ is unique. The amount of work invested in each graph going from stage $n-1$ to stage n is $O(n)$. Thus, we can compute $F(n,m)$ in $O(n^2|F(n,m)|)$ time. □

Example: We shall construct the graphs in $F(6,7)$. According to the algorithm, we recursively consider $F(5,7)$ and $F(5,2)$. Clearly $F(5,2) = \{K_{1,2} + E_2\}$,. For $F(5,7)$ we need to recursively consider $F(4,3)$. Now, $F(4,3) = \{K_3 + E_1, K_{1,3}\}$ and $sq(4,3) = 12$. It follows that $sq(5,7) = 12 + 4^2 + 4 \cdot 7 - 3(5-1) = 44$ and $F(5,7) = \{(K_3 + E_1)^* \ , \ K^*_{1,3}\}$. Now, $sq(6,7) = \max\{sq(5,7)\ , = \max\{44\ , \ 44\} = 44$. Thus, $|F(6,7) = 3|$ and

$$F(6,7) = \{((K_3 + E_1)^*)^+ \ , \ (K^*_{1,3})^+ \ , \ (K_{1,2} + E_2)^*\}.$$

We now turn to prove Theorem 1.2. Recall the graphs $H_1(n,m)$ and $H_2(n,m)$ defined at the beginning of this section:

Lemma 2.7. *There exists* $0 \le m \le \binom{n}{2}$ *such that* $Var(H_1(n,m)) = Var(n)$.

Proof: Let G be a graph with n vertices having $Var(G) = Var(n)$. Let $m = e(G)$. By (1) $sq(G) = sq(n,m)$. By corollary 2, either $sq(H_1(n,m)) = sq(n,m)$ or $sq(H_2(n,m)) = sq(n,m)$. In the first case, we have, again by (1) that $Var(H_1(n,m)) = Var(G) = Var(n)$ and we are done. In the second case, notice that $H_2(n,m)$ is the complement graph of $H_1(n, \binom{n}{2} - m)$. Since the variance of any graph is the same as the variance of its complement, we have $Var(H_1(n, \binom{n}{2} - m)) = Var(H_2(n,m)) = Var(G) = Var(n)$ and we are done. □

Proof of Theorem 1.2: According to Lemma 2.7, Corollary 2.3 and (1) we know that

$$Var(n) = \max_{0 \le m \le \binom{n}{2}} \frac{2m(v-1) + u(u+1)}{n} - 4\frac{m^2}{n^2} \tag{6}$$

where v and u are uniquely defined by $m = \binom{v}{2} + u$ and $0 \le u \le v-1$. Note that we cannot just compute (6) by derivation, since, by definition, u and v are always integers, and so the function cannot be replaced by a continuous one. We will first show that the maximum in (6) is obtained when $u = 0$. Since $Var(1) = Var(2) = 0$, we can assume $n \ge 3$ and $1 \le m \le \binom{n}{2} - 1$. Thus, $2 \le v \le n-1$. In order to show that (6) is obtained when $u = 0$ it suffices to show that:

1. When $v = n-1$ and $u = 0$, the r.h.s. of (6) is larger than when $v = n-1$ and $1 \leq u \leq n-2$.

2. For every $2 \leq v \leq n-1$, when $m = \binom{v}{2}$ the value of the r.h.s. of (6) is larger than when $m = \binom{v}{2} - \lambda$ for every $1 \leq \lambda \leq v-2$.

We verify that both conditions hold:

1. Since $m = \binom{n-1}{2} + u$ then, putting $k = \binom{n}{2} - m$, the r.h.s. of (6) is $k(k+1)/n - 4k^2/n^2$, and, since $1 \leq k \leq n-1$, the maximum of this expression is obtained when $k = n-1$, which, in turn, implies that $m = \binom{n-1}{2}$, hence $v = n-1$ and $u = 0$.

2. We need to show that

$$\frac{2\binom{v}{2}(v-1)}{n} - \frac{v^2(v-1)^2}{n^2} \geq \frac{2(\binom{v}{2} - \lambda)(v-2) + (v-\lambda-1)(v-\lambda)}{n} - \frac{4(\binom{v}{2} - \lambda)^2}{n^2}.$$

   Multiplying both sides by n^2 and rearranging the terms, the last inequality is equivalent to:

$$n(4v - 5 - \lambda) \geq 4v^2 - 4v - 4\lambda.$$

   Since $v \geq 2$ and $\lambda \leq v-2$, it suffices to show that

$$n \geq \frac{4v^2 - 4v - 4\lambda}{4v - 5 - \lambda} = v + \frac{v + \lambda v - 4\lambda}{4v - 5 - \lambda}.$$

   Since $\lambda \leq v - 2$, we have $\frac{v+\lambda v - 4\lambda}{4v-5-\lambda} \leq 1$, and so the last inequality holds since $n \geq v+1$.

We have shown that for all $n > 2$.

$$Var(n) = \max_{2 \leq v \leq n-1} \frac{2\binom{v}{2}(v-1)}{n} - \frac{v^2(v-1)^2}{n^2} \tag{7}$$

Although v in constrained to an integer, we can use derivation to determine (7). It turns out that (7) is maximized when $v = \lfloor 3n/4 \rfloor$, and when $n \equiv 2 \bmod 4$, we need to perform the rounding upwards, namely $v = (3n+2)/4$. Note that (7) is obtained by the graph $H_1(n, \binom{v}{2})$, and its complement. □

3 Concluding remarks and open problems

By running the algorithm of Theorem 1.1 for each $n \le 150$ and each $0 \le m \le \binom{n}{2}$ we have observed the following:

1. Since $Var(n)$ is obtained whenever $m = \binom{q}{2}$ or $m = \binom{n}{2} - \binom{q}{2}$, where $q = [3n/4]$ (rounding is up when $n \equiv 2 \bmod 4$), it is interesting to determine $F(n,q)$ since this tells us the number of n-vertex graphs which have variance $Var(n)$. It turns out that for $n \le 150$ we always have that $|F(n,q)| = 1$, and, in fact, $F(n,q) = \{K_q + E_{n-q}\}$, unless $n = 4$ in which case $F(4,3) = \{K_3 + E_1, K_{1,3}\}$. It is therefore safe to conjecture that:

   Conjecture 3.1 *For every $n > 1$ there are exactly two graphs with n vertices having variance $Var(n)$. They are $K_q \cup E_{n-q}$ and its complement.*

2. For every $n \le 150$, the number of graphs in $F(n,m)$ is either 1,2,3,4 or 6. Representing examples are: $|F(6,3)| = 2$, $|F(6,4)| = 1$, $|F(6,7)| = 3$, $|F(7,9)| = 4$, $|F(9,18)| = 6$. It is therefore interesting to solve the following problem:

   Problem 3.2 Is it true that for every n and m, $|F(n,m)| \le C$ where C is an absolute constant. In particular, is it true that $C = 6$? Is it true that for every n and m, $F(n,m) \ne 5$?

References

[1] M. Albertson, The irregularity of a graph, *Ars Combin.* **46** (1997), 219–225.

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press, 1978.

[3] D. de Caen, An upper bound on the sum of squares of degrees in a graph, *Discrete Math.* **185** (1998), 245–248.

[4] R.J. Cook, Vertex degrees of planar graphs, *J. Combin. Theory, Ser. B* **26** (1979), 337–345.

[5] E. van Dam, A Cauchy-Khinchin matrix inequality, *Linear Algebra and its Applications* **280** (1998), 163–172.

[6] A.W. Goodman, Triangles in a complete chromatic graph, *J. Austral. Math. Soc. Ser. A* **39** (1985), 86–93.

[7] D. Olpp, A conjecture of Goodman and the multiplicities of graphs, *Australian J. Combin.* **14** (1996), 267–282.

[8] L.A. Székely, L.H. Clark and R.C. Entringer, An inequality for degree sequences, *Discrete Math.* **103** (1992), 293–300.

[9] Miroslaw Truszczynski, Note on vertex degrees of planar graphs, *J. Graph Theory* **8** (1984), 171–176.