Generalized Sum Graphs

Noga Alon*
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University
Ramat Aviv , Tel Aviv 69978 Israel

Edward R. Scheinerman†
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, Maryland 21218 USA

ABSTRACT

Harary [8] calls a finite, simple graph G a *sum graph* if one can assign to each $v_i \in V(G)$ a label x_i so that $\{v_i, v_j\} \in E(G)$ iff $x_i + x_j = x_k$ for some k. We generalize this notion by replacing "$x_i + x_j$" with an arbitrary symmetric polynomial $f(x_i, x_j)$. We show that for each f, not all graphs are "f-graphs". Furthermore, we prove that for every f and every graph G we can transform G into an f-graph via the addition of $|E(G)|$ isolated vertices. This result is nearly best possible in the sense that for all f and for all $m \leq \frac{1}{2}\binom{n}{2}$, there is a graph G with n vertices and m edges which, even after the addition of $m - O(n \log n)$ isolated vertices, is not an f-graph.

*Research supported in part by a U.S.A.-Israel Binational Science Foundation and by a Bergmann Memorial Grant.

†Research supported in part by the Office of Naval Research contract number N00014–85–K0622.

1 Introduction

Harary [8] makes the following definition: Let G be a (finite, simple) graph with vertex set $V = \{v_1, \ldots, v_n\}$. Such a graph is called a *sum graph* if one can assign a positive integer x_i to each vertex v_i so that $\{v_i, v_j\} \in E(G)$ if and only if $x_i + x_j = x_k$ for some k. Other authors (see, for example, [3, 6, 10]) have considered variants of this definition. We propose the following generalization:

Given a symmetric polynomial of two variables, $f : \mathbf{R}^2 \to \mathbf{R}$, we say G is an *f-graph* if one can assign real numbers $x_1, x_2, \ldots, x_n$ to its vertices $v_1, v_2, \ldots, v_n$ (respectively) so that $\{v_i, v_j\} \in E(G)$ iff $f(x_i, x_j) = x_k$ for some k.

When $f(x, y) = x+y$ this gives us (a slight generalization of) sum graphs.

Sum graphs (and their generalizations) can be very efficiently stored in a computer. An array holds the vertex labels. Adjacency can be tested by simple computations followed by a table look-up.

Ideally one would like to characterize f-graphs for each given f. In this paper we consider the following two problems. First, we show that not all graphs are f-graphs by estimating the number of f-graph on n vertices. Second, we consider the problem of how to transform a given graph with n vertices and m edges into an f-graph by the addition of isolated vertices. We show that this is always possible by the addition of m isolates and that this result is essentially best possible.

The f-graph idea provides a wide latitude for graph representations. Furthermore, our methods can be readily extended beyond the f-graph paradigm to more general representation schemes.

2 Some Graphs are not f-Graphs

Our main result in this section is an approximate count of the number of f-graphs for any symmetric polynomial f. This estimate shows that for any f, not all graphs are f-graphs. However, there is no "universal" non-f-graph: For every graph G we show there is a polynomial f such that G is an f-graph.

Theorem 1 *If f is a non-constant, symmetric polynomial and $X_f(n)$ is the* *number of labelled f-graphs on the vertex set $\{1, 2, \ldots, n\}$, then*

$$\log X_f(n) = \Theta(n \log n).$$

The key to proving this theorem is the following result due to Warren [11]. Let $p_1, p_2, \ldots, p_r$ be polynomials in ℓ variables of degree at most d. For $\mathbf{x} \in \mathbf{R}^\ell$ the *sign pattern* of $p_1(\mathbf{x}), p_2(\mathbf{x}), \ldots, p_r(\mathbf{x})$ is the vector

$$(\text{sgn } p_1(\mathbf{x}), \text{sgn } p_2(\mathbf{x}), \ldots, \text{sgn } p_r(\mathbf{x})) \in \{-1, 0, +1\}^r.$$

As $\mathbf{x}$ ranges over $\mathbf{R}^\ell$, the above vector changes values. Since each coordinate may have one of three values, a simple upper bound on the number of sign patterns is 3^r. However, since the p_i's are polynomials, Warren's theorem (see [2, 11]) gives the following sharper result:

Theorem 2 (Warren) *Let $p_1, \ldots, p_r : \mathbf{R}^\ell \to \mathbf{R}$ be polynomials of degree at most d. If $r \geq \ell$ then the number of sign patterns of the p_i's is at most*

$$\left(\frac{8edr}{\ell}\right)^\ell . \square$$

Proof of Theorem 1. Let f be a symmetric polynomial in two variables of degree d. Consider the polynomials $p_{ijk} : \mathbf{R}^n \to \mathbf{R}$ defined by

$$p_{ijk}(\mathbf{x}) = f(x_i, x_j) - x_k$$

where $1 \leq i, j, k \leq n$ and $i < j$. There are $\binom{n}{2} n < \frac{1}{2} n^3$ such polynomials. Observe that the number of different f-graphs on n vertices is bounded above by the number of sign patterns of the p_{ijk}. [The vector $\mathbf{x} = (x_1, \ldots, x_n)$ indicates a labeling of the vertices. Two labelings which result in the same sign patterns for the p_{ijk}'s necessarily give the same f-graph. Thus different f-graphs must give different sign patterns.] Thus by Warren's theorem,

$$X_f(n) \leq \left(\frac{8ed\frac{1}{2}n^3}{n}\right)^n \leq (Kn)^{2n} \tag{1}$$

for some constant K.

To establish the lower bound, let $m = \lfloor n / \log n \rfloor$ and choose values $x_1, \ldots, x_m$ so that the $\binom{m}{2}$ values $f(x_i, x_j)$ are all distinct and none are equal to x_k for any $1 \leq k \leq m$.

[We can construct such a sequence inductively. We can always choose $x_1, \ldots, x_s$ such that there are only finitely many x's such that either $f(x_i, x) = f(x_j, x)$ or $f(x_i, x_j) = x$. Thus we have infinitely many choices for x_{s+1} satisfying the specified conditions. Indeed, we can do this even if we were to restrict the x's to be integers. Furthermore, the choice of m is not crucial; we just need a sublinear function which grows sufficiently quickly.]

For $x_{m+1}, \ldots, x_n$ we choose distinct values among $f(x_i, x_j)$ with $1 \le i, j \le m$. Note that by varying the choices of $(x_{m+1}, \ldots, x_n)$ we can form any $(n-m)$-edge graph on vertices $1, \ldots, m$ we please. Thus,

$$X_f(n) \ge \binom{\binom{m}{2}}{n-m} = n^{[1-o(1)]n} \tag{2}$$

Hence, by (1) and (2) we have $\log X_f(n) = \Theta(n \log n)$.□

Since there are $2^{\binom{n}{2}}$ graphs on n vertices, it is an immediate corollary that for any symmetric polynomial f, there exists a graph G which is not an f-graph. Likewise, we can conclude that there are bipartite graphs or regular graphs which are not f-graphs, etc.

One may wonder if there is some graph G which is not an f-graph for any polynomial f. This is not the case, as shown in the following simple proposition.

Theorem 3 *Let G be a graph. There exists a symmetric polynomial f so that G is an f-graph.*

Proof. Suppose the vertex set of the given graph is $V(G) = \{0, 1, \ldots, n-1\}$. Let

$$f(x, y) = \prod_{\{i,j\} \in E(G)} \left[(x+i)^2 + (y+j)^2\right] \left[(y+i)^2 + (x+j)^2\right].$$

Note that $f(x, y)$ is symmetric and is zero exactly when $x = -i$, $y = -j$ (or vice versa) and $\{i, j\} \in E(G)$; f is positive otherwise. To show that G is an f-graph we must assign a labeling to its vertices. We do this by letting the label at vertex i be $x_i = -i$. Note that if $\{i, j\} \in E(G)$ then $f(x_i, x_j) = 0 = x_0$, but if $\{i, j\} \notin E(G)$ then $f(x_i, x_j) > 0$ which cannot be the value of any x_k since they are all non-positive. Hence G is an f-graph.□

3 Every Graph is an f-Graph...

..once one adds to it sufficiently many isolated vertices.

Given an arbitrary graph G, one can transform it into a sum graph by the addition of sufficiently many isolated vertices [4, 5, 8]. (For example, label each vertex of G with distinct powers of 3. Add $m = |E(G)|$ vertices corresponding to the edges of G and label each with the sum of the labels of its corresponding endpoints. One checks that the resulting graph is $G + mK_1$.) Harary [8] asked: What is the minimum number of isolated vertices which one must add to G to make it into a sum graph?

For a symmetric polynomial f and a graph G, let $s_f(G)$ denote the smallest integer s such that $G+sK_1$ is an f-graph. We show that $s_f(G) \leq |E(G)|$.

Theorem 4 *If G is a graph and f is a non-constant symmetric polynomial, then $s_f(G) \leq |E(G)|$.*

Proof. Suppose $V(G) = \{1, 2, \ldots, n\}$ and $m = |E(G)|$. Suppose the coefficients of the polynomial f are $c_1, \ldots, c_t$. Having chosen labels $x_1, \ldots, x_{i-1}$ for vertices 1 through $i-1$, let the label for vertex i be any real number x_i such that x_i is transcendental over the field $F_{i-1} = \mathbf{Q}(c_1, \ldots, c_t, x_1, \ldots, x_{i-1})$. (Since F_{i-1} is countable, there are only countably many real values which are algebraic over F_{i-1}.) Now we add m additional vertices u_{ij} corresponding to $\{i, j\} \in E(G)$. Let the label on u_{ij} be $x_{ij} = f(x_i, x_j)$. We must show that this labeling represents $G + mK_1$.

It is immediate that each edge $\{i, j\} \in E(G)$ is properly represented. What remains to be checked is that (1) no further edges between vertices of G are represented and (2) the additional vertices, u_{ij}, are isolated. Let

$$X = \{x_i : 1 \leq i \leq n\} \cup \{x_{ij} : \{i, j\} \in E(G)\}.$$

We must prove:

1. For all $1 \leq i < j \leq n$ with $\{i, j\} \notin E(G)$ we have $f(x_i, x_j) \notin X$, and
2. For all $\{i, j\} \in E(G)$ we have $f(x_{ij}, x_k) \notin X$ and $f(x_{ij}, x_{ab}) \notin X$ for any k and any $\{a, b\} \in E(G)$.

For (1) suppose $f(x_i, x_j) = x_k$ or $f(x_i, x_j) = x_{ab} = f(x_a, x_b)$. But in either equation we violate the transcendentality of the x with the largest subscript. Likewise for (2) we would have

- $f[f(x_i, x_j), x_k] = x_a$, or
- $f[f(x_i, x_j), x_k] = f(x_a, x_b)$, or
- $f[f(x_i, x_j), f(x_a, x_b)] = x_c$, or
- $f[f(x_i, x_j), f(x_a, x_b)] = f(x_c, x_d)$.

As before, each of these polynomial relations would violate the transcendentality of the x with the largest subscript.□

If the coefficients in f are integers then it is possible to choose the x_i's in the above proof to be integers as well (and completely avoid transcendental numbers). Put

$$\begin{aligned}
F_1 &= \prod_{i\neq j}(x_i - x_j) \\
F_2 &= \prod_{\{i,j\}\neq\{k,\ell\}} [f(x_i, x_j) - f(x_k, x_\ell)] \\
F_3 &= \prod_{i\neq j} [f(x_i, x_j) - x_k] \\
F_4 &= \prod_{i\neq j} [f(f(x_i, x_j), x_k) - x_a] \\
F_5 &= \prod_{i\neq j,\ a\neq b} [f(f(x_i, x_j), x_k) - f(x_a, x_b)] \\
F_6 &= \prod_{i\neq j,\ k\neq\ell} [f\left(f(x_i, x_j), f(x_k, x_\ell)\right) - x_a] \\
F_7 &= \prod_{i\neq j,\ k\neq\ell,\ a\neq b} [f\left(f(x_i, x_j), f(x_k, x_\ell)\right) - f(x_a, x_b)] \\
F &= F_1 \cdot F_2 \cdot F_3 \cdot F_4 \cdot F_5 \cdot F_6 \cdot F_7
\end{aligned}$$

where indicies run between 1 and n inclusive subject to the conditions shown. Now one checks that since F is not identically zero one can choose integer values for the x_i's so that $F \neq 0$. Having done this we let $\{x_1, \ldots, x_n\}$ be the labels for the original vertices in G and $\{f(x_i, x_j) : ij \in E(G)\}$ be the labels for the added isolated vertices.

Moreover, applying the method in [9], random substitutions into the variables x_i (with, say, integers between 1 and kn^7 for some constant k) will yield a non-zero value for F with high probability. This can be converted into a deterministic polynomial time labelling algorithm (or even a deterministic efficient parallel algorithm) using the methods in [1].

Note that for graphs with n vertices and m edges, the above theorem implies that $s_f(G) \leq m = O(n^2)$. One may suspect that for some f there is a sub-quadratic upper bound, however, we can apply Theorem 1 to show that a much better bound cannot be obtained. If one compares the number of graphs on n vertices with the number of f-graphs on $n + x$ vertices, one gets the existence of a graph for which $s_f(G) \geq cn^2/\lg n$ for any $c < \frac{1}{4}$. (Here and below, lg denotes the base-2 logarithm.) However, a more careful use of Warren's theorem gives us the following result.

Theorem 5 *Let f be a non-constant, symmetric polynomial. There exists a constant $C > 0$ (depending only on f) such that for all n and all $m \leq \frac{1}{2}\binom{n}{2}$ there is a graph G with n vertices and m edges such that*

$$s_f(G) > m - 3n \lg n - Cn.$$

Proof. For convenience we put $N = \binom{n}{2}$. Suppose $m \leq \frac{1}{2}N$ and suppose that for all graphs G with n vertices and m edges we have $s_f(G) \leq m - a$ where $a = 3n \lg n + Cn$. Clearly, we may assume that $m \geq a$. Suppose $V(G) = \{1, 2, \ldots, n\}$ and let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the corresponding labels we give to these vertices. Consider the (fewer than) n^4 polynomials

$$q_{ijkl}(\mathbf{x}) = f(x_i, x_j) - f(x_k, x_l).$$

Note that if the labels on vertices of G are given in $\mathbf{x}$ and $q_{ijkl}(\mathbf{x}) = 0$, then $\{i, j\} \in E(G) \iff \{k, l\} \in E(G)$. Now these (fewer than) n^4 polynomials in n variables have degree $d = \deg(f)$, and by Warren's theorem, the number of sign patterns for the q_{ijkl} is at most

$$\left(\frac{8edn^4}{n}\right)^n \leq (Kn)^{3n}$$

for some constant K.

Now suppose the sign pattern for the q's is fixed. We claim there are at most $\sum_{j=0}^{m+n-a} \binom{N+n-a}{j}$ graphs with n vertices, m edges that are f-graphs upon the addition of $m - a$ (isolated) vertices:

To prove this claim observe that the sign pattern of the q's induced by a given $\mathbf{x}$ creates a partition on the set of pairs of vertices so that $\{i, j\}$ and $\{k, l\}$ are in the same block of the partition iff $q_{ijkl}(\mathbf{x}) = 0$. Suppose there are t blocks in this partition and the block sizes are $r_1, \ldots, r_t$ where $r_1 + \cdots + r_t = N$. Note that the pairs in one block are either all edges or all non-edges in G. Moreover, at most $m+n-a$ of the blocks can be used. Thus there are at most $\sum_{j=0}^{m+n-a} \binom{t}{j}$ graphs G we can form given the sign pattern of the q's.

Now if no subset of the r's sums to m, it is impossible to make an f-graph with m edges on the vertices $1, \ldots, n$ regardless of how many additional vertices we add. Thus, without loss of generality, suppose $r_1 + \cdots + r_j = m$ with $1 \leq j < t$, where j is the minimum number of r's whose sum is m. If $j > n + m - a$ we cannot make G into an f-graph where the total number of vertices is at most $n + m - a$. Thus, $j \leq n + m - a$. Since each $r_i \geq 1$, $t - j \leq r_{j+1} + \cdots + r_t = N - m$ and therefore

$$t \leq (N - m) + j \leq (N - m) + (n + m - a) = N + n - a.$$

Thus there are at most $\sum_{j=0}^{m+n-a} \binom{N+n-a}{j}$ graphs G we can represent for each sign pattern of the q's. This completes the justification of the claim.

Finally, since we supposed that all graphs with n vertices and m edges have $s_f \leq m - a$, we have

$$\binom{N}{m} \leq (Kn)^{3n} \sum_{j=0}^{m+n-a} \binom{N+n-a}{j} \tag{3}$$

When C is large enough, $n \leq a$ giving, $m + n - a \leq \frac{1}{2}(N + n - a)$. Thus we have

$$\sum_{j=0}^{m+n-a} \binom{N+n-a}{j} \leq N \binom{N+n-a}{m+n-a}$$

Since $a = 3n \lg n + Cn$, if C is a sufficiently large constant then

$$\begin{aligned} \frac{\binom{N+n-a}{m+n-a}}{\binom{N}{m}} &= \frac{m(m-1)\cdots(m+n-a+1)}{N(N-1)\cdots(N+n-a+1)} \\ &\leq \left(\frac{m}{N}\right)^{a-n} \leq \left(\frac{1}{2}\right)^{a-n} < \frac{1}{N(Kn)^{3n}}, \end{aligned}$$

contradicting the inequality (3) above. This completes the proof.□

Note that we have not used the fact that the added vertices should be isolated. Regardless of the structure of the vertices we add to G, we must add at least $m - O(n \log n)$ vertices to make some graphs G into an f-graph.

In the specific case of sum graphs, the authors of [5] prove a similar result, namely: for all n, there is a graph G with n vertices, $m = cn^2$ edges (for some constant c) and $s(G) \geq m - O(n \log n)$.

4 Further Generalization

We can generalize the notion of f-graphs in several ways. For example, we can assume the labels on the vertices are complex numbers z_i. Or we may wish to assign a t-vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{it})$ to vertex v_i and have the edge $\{v_i, v_j\}$ precisely when $f(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_k$ for some k (where each coordinate function in $f : \mathbf{R}^{2t} \to \mathbf{R}^t$ is a polynomial). Finally, we can postulate that given a polynomial $f : \mathbf{R}^{3t} \to \mathbf{R}^s$, we have $\{v_i, v_j\} \in E(G)$ iff for some k we have $f(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = \mathbf{0}_s$. Note that this last generalization encompasses the previous two. In this case we can apply Warren's theorem to show that the number of f-graphs on n vertices is bounded by $(Kn)^{2tn}$ for some constant K (depending on f). Thus, even under this most liberal definition of f-graphs, almost all graphs are not f-graphs .

The results above depend heavily on the polynomial nature of f (both in our application of Warren's theorem and in our transcendental trickery in Theorem 4). If we were to let f be an arbitrary symmetric function of two variables we could easily contrive an f for which all graphs were f-graphs. Trenk [10] investigated "gcd-graphs" in which the label x_i on vertex v_i is a positive integer and $\{v_i, v_j\}$ is an edge iff $\gcd(x_i, x_j) = x_k$ for some k. The conclusion of Theorem 1 fails miserably in this case as all bipartite graphs are gcd-graphs, and therefore there are at least $2^{n^2/4}$ labelled gcd-graphs on n vertices.

References

[1] Alon, N. L. Babai and A. Itai, "A fast and simple randomized parallel algorithm for the maximal independent set problem" *J. Algorithms* **7** (1986) 567–583.

[2] Alon, N. and E.R. Scheinerman, "Degrees of freedom versus dimension for containment orders" *Order* **5** (1988) 11–16.

[3] Boland, J., G. Domke, R. Laskar and C. Turner, "Mod sum graphs", preprint.

[4] Ellingham, M., "Sum graphs from trees", preprint.

[5] Gould, R.J. and V. Rödl, "Bounds on the number of isolated vertices in sum graphs", Technical Report No. 77, Department of Mathematics and Computer Science, Emory University (1988).

[6] Grimaldi, R., "Graphs from rings", preprint.

[7] Hao, T. "On sum graphs", Mathematical Preprint No. 1988-7, Queen's University, Kingston, Ontario, Canada (1988).

[8] Harary, F., "Sum graphs and difference graphs", talk at the Nineteenth Southeastern Conference on Combinatorics, Graph Theory and Computing, Baton Rouge (1988).

[9] Schwartz, J.T., "Fast probabilistic algorithms for verification of polynomial identities" *SIAM J. Comput.* **27** (1980) 701–717.

[10] Trenk, A., personal communication.

[11] Warren, H.E., "Lower bounds for approximation by nonlinear manifolds" *Trans. AMS* **133** (1968) 167–178.