THE NUMBER OF RELATION GRAPHS

by

F. R. K. Chung
F. K. Hwang
D. H. Krantz
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

A relation graph is a graph where each pair of vertices is connected by an edge of one of the three types, "$\rightarrow$", "$-$" and "$\cdots$" satisfying the three conditions:

(i) $A \rightarrow B$ and $B \rightarrow C$ imply $A \rightarrow C$,

(ii) $A \rightarrow B$ and $A - C$ imply $B - C$,

(iii) $A \rightarrow B$ and $B \cdots C$ imply $A \cdots C$,

where A, B and C are any three vertices in the graph. In this paper we prove that the number of labeled relation graphs is asymptotically less than $(2.783)^{\binom{n}{2}}$. Relation graphs are associated with the concept of focal propositions in assessing the patterns of a body of evidence.

THE NUMBER OF RELATION GRAPHS

by

F. R. K. Chung
F. K. Hwang
D. H. Krantz
Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

We consider four possible relations between two given sets A and B: A contains B, B contains A, noncontainment but nonempty intersection, and empty intersection. We can represent the pairwise relations between n given sets by a graph, called a relation graph with each vertex representing a set, a directed edge $A \rightarrow B$ representing the relation A is contained in B, an undirected edge $A - B$ representing the relation of noncontainment but nonempty intersection, and a broken edge $A \cdots B$ representing the relation of empty intersection. Note that a relation graph must satisfy the following three conditions

(i) $A \rightarrow B$ and $B \rightarrow C$ imply $A \rightarrow C$.

(ii) $A \rightarrow B$ and $A - C$ imply $B - C$.

(iii) $A \rightarrow B$ and $B \cdots C$ imply $A \cdots C$.

Furthermore, these three conditions are also sufficient for a graph, where any two vertices are connected by an edge, either directed, undirected, or broken, to be a relation graph.

A graph is <u>labeled</u> if each vertex is labeled by a distinct number. Let $F(n)$ and $f(n)$ denote the numbers of labeled and unlabeled relation graphs on n vertices, respectively. Clearly,

$$F(n) \geq f(n).$$

We also have

$$4^{\binom{n}{2}} > F(n) \geq 2^{\binom{n}{2}}$$

since each pair of nodes can assume only one of four possible relations while a graph with only undirected edges is always a relation graph. Finally, since each unlabeled graph can induce at most $n!$ labeled graphs,

$$f(n) > F(n)/n! \geq 2^{\binom{n}{2} - n \log_2 n}.$$

In this paper we prove a new asymptotic upperbound on $F(n)$, i.e.,

$$2.783^{n^2/2} \geq F(n).$$

2. <u>The General Strategy</u>

In this section we will indicate the general strategy of an inductive (on n) proof of

$$F(n) \leq (2.783)^{n^2/2}. \tag{0}$$

We consider an arbitrary relation graph G on n vertices. Let v^* be a fixed vertex in G. We define

$$X_1 = \{v \in V(G) : v \rightarrow v^*\}$$
$$X_2 = \{v \in V(G) : v \leftarrow v^*\}$$
$$X_3 = \{v \in V(G) : v - v^*\}$$
$$X_4 = \{v \in V(G) : v \cdots v^*\}$$

Let v_i denote an arbitrary vertex in X_i. Then from the three conditions of a relation graph, we must have (i) $v_1 \rightarrow v_2$, (ii) either $v_2 \leftarrow v_3$ or $v_2 - v_3$, and (iii) no edge between v_1 and v_4. Figure 1 illustrates all possible relations between vertices in the X_i's.

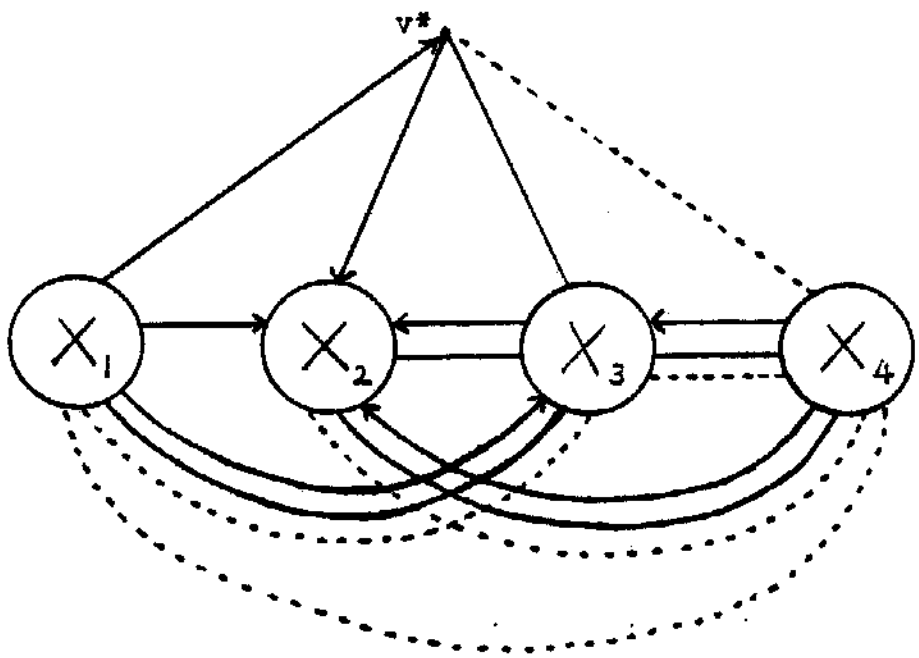

Figure 1. Possible relations between X_i's.

Let $P = \{\{v^*\}, X_1, X_2, X_3, X_4\}$ be a partition of the n points into five subsets satisfying the relations as specified in Fig. 1. Let $g(x_1, x_2, x_3, x_4)$ be the number of relation graphs given of type $P = (x_1, x_2, x_3, x_3)$ where X_i contains x_i points. Then we have

$$F(n) \leq \sum_{P} \frac{(n-1)!}{\prod_{i=1}^{4} (x_i!)} g(x_1, x_2, x_3, x_4). \tag{1}$$

By the definition of the F function, we have

$$g(x_1, x_2, x_3, x_4) \leq F(n-1). \tag{2}$$

On the other hand, we also have

$$g(x_1, x_2, x_3, x_4) \leq F(x_1)F(x_2)F(x_3)F(x_4)2^{x_2 x_3 {}_3 x_1 x_3 + x_2 x_4 + x_3 x_4} \tag{3}$$

since there are only two possible relations between X_2 and X_3 and so on. We will use both (2) and (3) to establish $F(n) \leq c^{n^2/2}$ where $c = 2.783$. Since there are at most $\binom{n+2}{3}$ choices for x_i, $i = 1,2,3,4$, with $\Sigma x_i = n-1$, it is sufficient to prove

$$\frac{(n-1)!}{\prod_{i=1}^{4} (x_i!)} g(x_1, x_2, x_3, x_4) \leq \frac{c^{n^2/2}}{n^3} \tag{4}$$

for any partition P.

Let $\alpha_i = \frac{x_i}{n-1}$. For n sufficiently large, (4) is equivalent to

$$2^{-(n-1) \sum_{i=1}^{4} \alpha_i \log \alpha_i} g(x_1, x_2, x_3, x_4) \leq \frac{c^{n^2/2}}{n^3} \tag{5}$$

for any α_i, $0 \le \alpha_i \le 1$, $\sum_{i=1}^{4} \alpha_i = 1$ and $x_i = \alpha_i(n-1)$.

(We note that all logarithms are to base 2.)

In order to prove (5), we consider the following two cases:

Case 1: $\quad -\sum \alpha_i \log \alpha_i \le \log c - \frac{3 \log n}{n-1}$.

Then

$$2^{-(n-1)\sum_{i=1}^{n} \alpha_i \log \alpha_i} g(x_1,x_2,x_3,x_4)$$

$$\le 2^{(n-1) \log c - 3 \log n} F(n-1)$$

$$\le 2^{(n-1) \log c}\, c^{(n-1)^2/2}/n^3$$

$$\le c^{n^2/2}/n^3.$$

Thus (5) holds in this case.

Case 2: $\quad -\sum \alpha_i \log \alpha_i > \log c - \frac{3 \log n}{n-1}$.

We note that from (3) we have

$$2^{-(n-1)\sum_{i=1}^{4} \alpha_i \log \alpha_i} g(x_1,x_2,x_3,x_4)$$

$$\le 2^{-(n-1)\sum_{i=1}^{4} \alpha_i \log \alpha_i} F(x_1)F(x_2)F(x_3)F(x_4)\cdot 2^{x_2x_3}3^{x_1x_3+x_2x_4+x_3x_4}$$

$$\le 2^{-(n-1)\sum_{i=1}^{4} \alpha_i \log \alpha_i} c^{(x_1^2+x_2^2+x_3^2+x_4^2)/2}\, 2^{x_2x_3}3^{x_1x_3+x_2x_4+x_3x_4},$$

by the inductive assumption.

Therefore, it is sufficient to prove that

$$-\frac{(n-1)}{\log c}\sum_{i=1}^{4}\alpha_i \log \alpha_i + \frac{(x_1^2+x_2^2+x_3^2+x_4^2)}{2} + \frac{x_2x_3}{\log c} + \frac{\log 3}{\log c}(x_1x_3+x_2x_4+x_3x_4)$$

$$< n^2/2. \tag{6}$$

Set $N = n-1$. Since $x_i = \alpha_i N$ and $\sum_{i=1}^{4} \alpha_i = 1$, (6) is equivalent to the following:

$$(\log 3 - \log c)(\alpha_1\alpha_3+\alpha_2\alpha_4+\alpha_3\alpha_4)N^2 + \left(-\sum_{i=1}^{4} \alpha_i \log \alpha_i - \log c\right)N$$

$$\leq ((\alpha_1\alpha_2+\alpha_1\alpha_4)(\log c + (\log c - 1)\alpha_2\alpha_3)N^2 + (\log c)/2. \tag{7}$$

Since $-\sum_{i=1}^{4} \alpha_i \log \alpha_i \leq 2 \leq 1 + \log c$, it is enough to prove the following in order to establish (5):

$$(\log 3 - \log c)(\alpha_1\alpha_3+\alpha_2\alpha_4+\alpha_3\alpha_4) + 1/N$$

$$\leq (\alpha_1\alpha_2+\alpha_1\alpha_4) \log c + (\log c - 1)\alpha_2\alpha_3. \tag{8}$$

Furthermore, since we are looking for asymptotic results, we can drop the $1/N$ term. Define

$$h(\alpha_1,\alpha_2,\alpha_3,\alpha_4) = (\alpha_1\alpha_2+\alpha_1\alpha_4) \log c$$

$$+ (\log c - 1)\alpha_2\alpha_3 - (\log 3 - \log c)(\alpha_1\alpha_3+\alpha_2\alpha_4+\alpha_3\alpha_4).$$

We will show that $h(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ is positive for all α_1, α_2, α_3, α_4 satisfying

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$$

$$0 \le \alpha_i, \text{ for } i = 1,2,3,4 \tag{9}$$

$$-\sum \alpha_i \log \alpha_i \ge \log c$$

We accomplish this by minimizing h over the variables one at a time.

3. Details of the Proof

First we will show the following.

Claim: The minimum value of h on the set of points satisfying (9) is achieved when $\alpha_1 = 0$.

Proof: Define

$$h_1(\alpha_1,\alpha_2,\alpha_4) = h(\alpha_1,\alpha_2,1-\alpha_1-\alpha_2-\alpha_4,\alpha_4).$$

Clearly, if h_1 is monotone increasing in α_1, the Claim follows immediately. We have

$$\frac{\partial h_1}{\partial \alpha_1}(\alpha_1,\alpha_2,\alpha_4) = -(\log 3 - \log c)(1-2\alpha_1 - \frac{1 + \log 3 - \log c}{\log 3 - \log c}\alpha_2 - \frac{2 \log 3 - \log c}{\log 3 - \log c}\alpha_4).$$

Suppose to the contrary that h_1 is not monotone increasing in α_1, say, $\frac{\partial h_1}{\partial \alpha_1} \le 0$ at the point $(\bar{\alpha}_1,\bar{\alpha}_2,\bar{\alpha}_4)$, i.e.,

$$1 - 2\bar{\alpha}_1 - \frac{1 + \log 3 - \log c}{\log 3 - \log c}\bar{\alpha}_2 - \frac{2 \log 3 - \log c}{\log 3 - \log c}\bar{\alpha}_4 \ge 0. \tag{10}$$

Let $\bar{\alpha}_3 = 1 - \bar{\alpha}_1 - \bar{\alpha}_2 - \bar{\alpha}_4$. By noting

$$\frac{1}{\log 3 - \log c} > 9$$

and

$$\frac{\log 3}{\log 3 - \log c} > 14,$$

we obtain from (10):

$$\bar{\alpha}_3 - \bar{\alpha}_1 - 9\bar{\alpha}_2 - 14\bar{\alpha}_4 > 0.$$

Define $\hat{\alpha} = \bar{\alpha}_2 + \bar{\alpha}_4$, then the above equation implies $\hat{\alpha} < .1$. Therefore $\bar{\alpha}_1 + \bar{\alpha}_3 = 1-\hat{\alpha} > .9$.

It is easy to verify by elementary calculus that for a fixed, $0 < a \leq 1$ and $0 \leq b \leq a$, the function $-b \log b-(a-b) \log (a-b)$ is concave and attains its maximum at $b = a/2$. Consequently

$$- \bar{\alpha}_2 \log \bar{\alpha}_2 - \bar{\alpha}_4 \log \bar{\alpha}_4 \leq - \hat{\alpha} \log \hat{\alpha} + \hat{\alpha},$$

$$- \bar{\alpha}_1 \log \bar{\alpha}_1 - \bar{\alpha}_3 \log \bar{\alpha}_3 \leq - (1-\hat{\alpha}) \log (1-\hat{\alpha}) + (1-\hat{\alpha}).$$

It follows that

$$\begin{aligned} - \sum_{i=1}^{4} \bar{\alpha}_i \log \bar{\alpha}_i &\leq 1 + [-\hat{\alpha} \log \hat{\alpha} - (1-\hat{\alpha}) \log (1-\hat{\alpha})] \\ &< 1 + [-(.1) \log (.1) - (.9) \log (.9)] \\ &= 1.468 \\ &< \log c, \end{aligned}$$

a contradiction to our assumption for case (ii). Therefore h_1 is monotone increasing in α_1 and the proof of the claim is complete.

Define

$$h_2(\alpha_2,\alpha_3,\alpha_4) = h(0,\alpha_2,\alpha_3,\alpha_4)$$

$$= (\log c - 1)\alpha_2\alpha_3 - (\log 3 - \log c)(\alpha_2+\alpha_3)\alpha_4,$$

where α_2, α_3, α_4 satisfy

$$0 \leq \alpha_i, \ \alpha_2 + \alpha_3 + \alpha_4 = 1, \tag{11}$$

and

$$-\sum_{i=2}^{4} \alpha_i \log \alpha_i \geq \log c.$$

From the Claim, the positiveness of h_2 will imply the positiveness of h_1. Furthermore, since α_2 and α_3 are symmetric in h_2, we may assume $\alpha_2 \leq \alpha_3$. Define $H(x)$ to be the entropy function, i.e.,

$$H(x) = -x \log x-(1-x) \log (1-x).$$

Then it is straightforward to show that

$$-\sum_{i=2}^{4} \alpha_i \log \alpha_i = (1-\alpha_4) H\left(\frac{\alpha_2}{1-\alpha_4}\right) + H(\alpha_4). \tag{12}$$

Thus the condition of Case (ii) implies

$$(1-\alpha_4)H\left(\frac{\alpha_2}{1-\alpha_4}\right) + H(\alpha_4) \geq \log c. \tag{13}$$

Define

$$h_3(\alpha_2,\alpha_4) = h_2(\alpha_2,1-\alpha_2-\alpha_4,\alpha_4) = (\log c - 1)\alpha_2(1-\alpha_2-\alpha_4)$$
$$- (\log 3 - \log c)\alpha_4(1-\alpha_4).$$

Then

$$\frac{\partial h_3}{\partial \alpha_2} = (\log c - 1)(1 - 2\alpha_2 - \alpha_4) \geq 0,$$
$$\frac{\partial^2 h_3}{\partial \alpha_2^2} < 0.$$

Hence h_2 is monotone decreasing in α_2. Note that $H\left(\frac{\alpha_2}{1-\alpha_4}\right)$ is also monotone decreasing in α_2 for $0 \leq \frac{\alpha_2}{1-\alpha_4} \leq \frac{1}{2}$. Furthermore, the condition of Case (ii) implies that for n large

$$(1-\alpha_4) + H(\alpha_4) \geq \log c \text{ (hence } \alpha_4 \leq \alpha_* \cong .522). \tag{14}$$

On the other hand

$$H(\alpha_4) \leq 1 < \log c.$$

Therefore there exists a unique α_2^0 such that

$$(1-\alpha_4)\ H\left(\frac{\alpha_2^0}{1-\alpha_4}\right) + H(\alpha_4) = \log c, \tag{15}$$

and $h_3(\alpha_2,\alpha_4)$ is minimized at $\alpha_2 = \alpha_2^0$. Define $\alpha = \alpha_4$ and $\beta = \frac{\alpha_2^0}{1-\alpha_4}$. From (15), we have

$$\frac{d\beta}{d\alpha} = \frac{H(\beta) - \log \frac{1-\alpha}{\alpha}}{(1-\alpha) \log \frac{1-\beta}{\beta}} .$$

Furthermore, define

$$h_4(\alpha) = h_3(\beta,\alpha).$$

Then it suffices to prove $h_4(\alpha) > 0$ for all α satisfying (14).

From straightforward calculations, we have

$$\frac{d\ h_4}{d\alpha}(\alpha) = (\log c - 1)(-\beta(1-\beta) + (1-\alpha)(1-2\beta)\frac{d\beta}{d\alpha}) - \log 3 + \log c.$$

We note that

$$-\beta(1-\beta) + (1-\alpha)(1-2\beta)\frac{d\beta}{d\alpha} = \frac{\beta^2 \log \beta - (1-\beta)^2 \log (1-\beta) - (\log \frac{1-\alpha}{\alpha})(1-2\beta)}{\log \frac{1-\beta}{\beta}}$$

and $\beta^2 \log \beta - (1-\beta)^2 \log (1-\beta) \le 0$ for all $\beta \le \frac{1}{2}$. Therefore $\frac{d\ h_4}{d\alpha} < 0$ for $\alpha \le \frac{1}{2}$. For $.5 < \alpha < .522$, by (15) we have $\beta > .37$ and $1 - 2\beta \le \frac{1}{2} \log \frac{1-\beta}{\beta}$

$$\begin{aligned} \frac{d\ h_4}{d\alpha}(\alpha) &\le -\log\left(\frac{1-0.522}{0.522}\right) \cdot \frac{1}{2} - \log 3 + \log c \\ &\le -0.04 < 0. \end{aligned}$$

The minimum value of h_4 is achieved at α_* where α_* is as large as possible, i.e., α_* satisfies

$$(1-\alpha_*) + H(\alpha_*) = \log c \text{ (hence } \beta = \tfrac{1}{2}). \tag{17}$$

It is easy to see that $h_4(\alpha) > 0$ for all $0 \leq \alpha < \alpha_*$. (Note that α_* and c are joint solutions of (17) and $h_4(\alpha) = 0$.)

We have proved that $H(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ is positive for any $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ satisfying (9). Thus (6) and (0) are proved for sufficiently large n.

4. Some Concluding Remarks

Shafer [1] introduced the concept of a proposition which is focal with respect to an item of evidence: the proposition is exactly what the evidence tends to support. For example, suppose that a man has been murdered. It is learned that, a few days earlier, he had an angry meeting with three of the seven people that he supervised at work, during which he berated and threatened them. In isolation, this fact supports (though of course it does not prove) a proposition something akin to the following: "One or more of those three people was involved in the murder." Propositions which are more general (such as "One of the seven people he supervised at work was involved in the murder") or more specific propositions (such as "One of the three people he met with was solely responsible for the murder") are not focal propositions for that particular item of evidence; whereas the first mentioned proposition is the focal proposition.

A more complex body of evidence may not have any single focal proposition, yet it may be possible to subdivide the evidence into conceptually independent items, each of which does have a single focus. When this is possible, the degree of belief in any particular proposition A will depend positively on the strengths of those items which <u>support</u> A -- i.e., whose focal propositions logically imply A -- and will depend negatively on the strengths of those items which <u>contradict</u> A -- i.e., whose focal propositions logically imply not-A.

We are thus led to the idea of different abstract patterns of argumentation, depending on the logical relations among various focal propositions. If we have focal propositions A,B,C,..., then the abstract pattern of argumentation will depend on the logical relations among these propositions.

A partial characterization of the above abstract pattern, which may be useful, is obtained by graphing the <u>pairwise</u> logical relations among the propositions. Any pair of nonequivalent propositions, A and B, stands in one of four possible relations : A contains B; B contains A; noncontainment with nonempty intersection; and empty intersection. For three propositions, A, B, C, each of 3 pairs could give rise to 4 possibilities; but not all of the resulting 4^3 patterns are possible, since obviously, two of the pairwise relations may dictate or constrain the third one. Enumeration shows that there are in fact only 41 possible patterns; for example, there are 6 different fully nested patterns $A \supset B \supset C$, $B \supset A \supset C$,...,1 fully disjoint pattern, etc.

The pairwise relations do not fully characterize the logical pattern of argumentation (triple intersections, etc. are also relevant), but they do provide a lower bound on the number of distinct patterns; and in view of limited human capacity for grasping complex structures quickly, they may correspond well to subjectively distinct patterns.

The purpose of this note is to provide upper and lower bounds for the number of such distinct pairwise patterns.

REFERENCE

1. Shafer, G. A mathematical theory of evidence. Princeton: Princeton Univ. Press (1976).