Enumeration of Full Graphs: Onset of the Asymptotic Region

L. J. Cowen * D. J. Kleitman † F. Lasaga D. E. Sussman

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

A *full graph* on n vertices, as defined by Fulkerson, is a representation of the intersection and containment relations among a system of n sets. It has an undirected edge between vertices representing intersecting sets, and a directed edge from a to b if the corresponding set A contains B. Kleitman, Lasaga and Cowen gave a unified argument to show that asymptotically, almost all full graphs can be obtained by taking an arbitrary undirected graph in the n vertices, distinguishing a clique in this graph which need not be maximal, and then adding directed edges going out from each vertex in the clique to all vertices to which there is not already an existing undirected edge. Call graphs of this type members of the dominant class. This paper obtains the first upper and lower bounds on how large n has to be, so that the asymptotic behavior is indeed observed. It is shown that when $n > 170$, the dominant class predominates, while when $n < 17$, the full graphs in the dominant class comprise less than half of the total number of realizable full graphs on n vertices.

1 Introduction

There have been two kinds of questions extensively addressed for a wide range of counting problems: exact counts for small parameter values, and asymptotic results which hold in the limit as the parameters grow large. Surprisingly, the question, when are these limits achieved, or more specifically, for what parameters are the asymptotic results reasonably accurate? are rarely addressed.

This paper describes an attempt to answer this question for the number of "Full Graphs". In a previous paper, some of the authors obtained an asymptotic formula for the number of such graphs, which holds for sufficiently large number of vertices. However, that result gave no indication at all of what number was sufficiently large for the formula to be anywhere near the correct answer.

*Supported by an NSF postdoctoral fellowship.

†Supported in part by NSF grant 9108403-DMS and US NSA grant MDA904-92-H-3029.

The methods appropriate for such analysis are similar but somewhat different from those that we used in the paper establishing asymptotic results. In that paper we had to establish that certain kinds of full graphs grew with the number of vertices at a rate (r) lower than a threshold that represents the rate of growth of the class of full graphs we would use to establish the asymptotic formula. (In other words, we constructed a subclass of full graphs whose rate of growth was very close to r, and showed that this was the *dominant class* of full graphs, by demonstrating a partition of the class of all other full graphs into subclasses, each of which grew at a rate asymptotically less than r). Thus we sought the simplest arguments that established that r was less than that threshold. In this paper, we seek instead the best possible upper bounds on r that can be achieved by our methods.

This question, the finite implications of asymptotic analysis, is neglected because it appeals neither to those who seek exactitude, nor those who want to look only at limiting behavior. Yet, if asymptotics is to have any concrete meaning, this is an important question.

Below we show that the range above which our enumerated class of full graphs dominates the rest starts at somewhere between 17 and 170 vertices' or at 54 vertices within a factor of 3.2. We raise as a question of possible interest: what is the analogous statement for partial orders?

2 Background

A *full graph* (as introduced by Fulkerson and Gross [2]) represents both the containment and intersection properties of a collection of sets. Each set is represented by a vertex, and the vertex which represents set A has directed arcs which point at the vertices corresponding to sets that A contains. Undirected arcs link vertices whose sets have non-empty intersections.

$FG(n)$ denotes the number of full graphs on n vertices, then Lynch [5] conjectured that

$$\lim_{n\to\infty} \frac{\log_2 FG(n)}{n^2} = 1/2.$$

This was proved by Kleitman, Lasaga, and Cowen [3] who proved a stronger result that characterized what most full graphs look like for large n. In fact, [3] showed that the class of graphs that predominates among full graphs, for sufficiently large n, corresponds to pairs of graphs (U, C), where U is an undirected graph on n vertices, and C is a clique on some k of these vertices which is a subgraph of U. One obtains the corresponding full graph by adding directed arcs from each vertex in the clique to all vertices outside it that it is not already adjacent to. A full graph of this type is said to be in the *dominant* class.

Since this paper is concerned with getting bounds on the asymptotic region, we ask: how large must n be before the dominant class predominates?

In this paper, we achieve upper and lower bounds on n, namely 170 and 17. The upper bound is obtained by refining the method of [3]. Like the previous paper, our construction partitions full graphs into a fixed number of classes, where it is shown that the number of full graphs in each class grows asymptotically slower than the dominant class. We use a more complicated set of classes to get better bounds on the asymptotic region: we remark that our methods could also be applied directly to the Kleitman et al. construction, in which case the upper bound obtained would be greater than 750, stretching our bound on the onset of the asymptotic region by more than a factor of 3.

3 The construction

The next three sections are devoted to proving the following theorem.

Theorem 3.1 *Let $FG(n)$ be the class of all full graphs on n vertices. Let $D(n)$ be the class of full graphs obtained by taking an arbitrary undirected on n vertices, distinguishing a clique in this graph that need not be maximal, and then adding directed edges going out from each vertex in the clique to all vertices to which there is not already an existing undirected edge. Then if $n < 17$, $D(n) < FG(n)/2$, and if $n > 170$, $D(n) > FG(n)/2$.*

We will derrive an improved partition of full graphs not in $D(n)$ into a set of classes, so that the upper bound obtained for full graphs in the union of all these classes is not too large. In attempting to derive a tight upper bound for the union, there will be a tradeoff between the upper bounds obtained for the number of full graphs in a particular class, and the number of classes we specify to comprise a canonical partition into classes. We did some careful, experimental balancing to achieve a near optimal tradeoff in the construction below.

The classes of our partition are defined in terms of collection of subgraphs (of six vertices or less), together with an order on these subgraphs. A full graph not in $D(n)$ is said to be in the class indexed by particular subgraphs if it contains these subgraphs, and it has not already appeared in a previous class. Typically, the subgraphs will consist of one specified subgraph with between three and seven vertices, union any number of directed arcs, with a fixed number of singletons left over. The subgraphs will be ordered first according to the number of vertices in the longest path in multi-vertex subgraph, then among those graphs with paths of length i, but none of length $i+1$, they will be classed according to which can be partitioned into an allowed subgraph, singletons, and directed arcs, with the fewest number of singletons. Finally, for graphs with the same length for their longest directed path, for which the minimal partition into singletons and directed arcs yields the same smallest number of singletons, an order on the allowed subgraphs is also specified.

We now list the allowed subgraphs for our construction, in order, together with the number of ways to connect either a singleton or a directed arc to the subgraph, and still remain in the class; i.e. so that the resulting subgraph is a legal subgraph of a full graph, and the additional connections do not automatically imply that the full graph has already been counted as part of a previous class. We checked all the larger numbers in the below table on the computer. In the next section, we will show how to use these classes, and these calculations to get a recursive upper bound on the number of full graphs not in the dominant class.

3.1 Notation

We introduce the following notation. For a path of vertex length k, label the vertices 1 through k, where $k \supset \ldots \supset 1$. We can correspond to each possible connection pattern between a vertex path of length k and a set-vertex v, an ordered pair (i, j), or $(i, -j)$; where i is the smallest index for which v is contained in i (or $k+1$, if v is not contained in k); and j is the largest index for which v contains j, or if v does not contain 1, 0 if v intersects 1, and $-j$ if j is the largest index of a vertex disjoint from v.

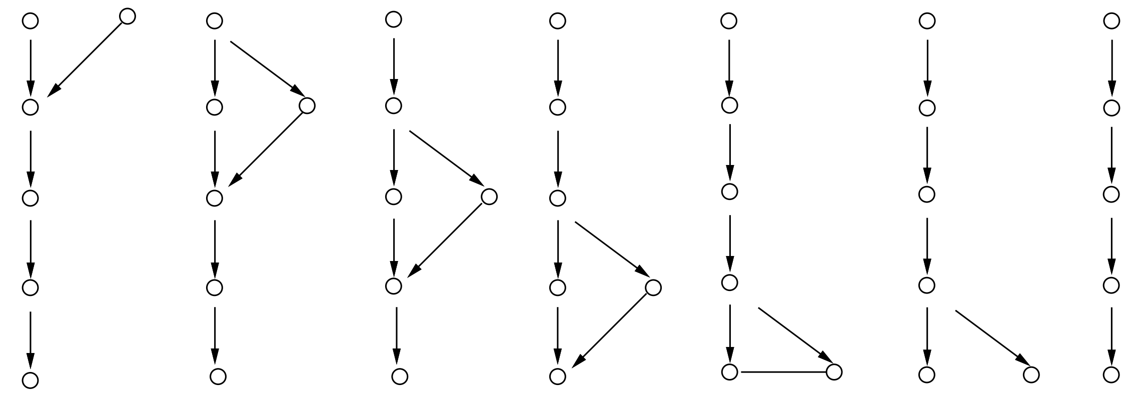

Figure 1: Specified configurations for a path of length five.

3.2 The Construction

Our first class (as in [3]) is full graphs containing a directed path of six vertices or more. There are 49 legal ways to connect a single vertex to a directed path of six vertices.

We next consider all partitions of a full graph that does not contain a directed path of six vertices or more into one of the configurations listed below, with the remainder of the vertices partitioned into singletons and arcs, with the smallest possible number of single vertices.

To handle full graphs which have a directed path of vertex length five (but no directed path of length greater), we consider, in order, the following configurations, all of which contain a directed five vertex path, and all which contain additional vertices are indexed by the name of the connection that an additional singleton makes to the path: (6,4), (5,3), (4,2), (3,1), (2,0), (2,-1), and also consider the directed path of vertex length five.

For full graphs with paths of vertex length four, we use the same canonical form with a configuration (in this case on 5 or 4 vertices), singletons and directed arcs, with the minimum possible number of single vertices. our 7 5-vertex configurations are, in order, (5,3) (4,2) (3,1) (2,0) (2,-1) (3,0) and (3,-1), plus a directed path of vertex length four.

For full graphs with directed paths of vertex-length three, we would like to do the same thing, but this time, we also keep track of the *level* of a vertex in order to hold down the number of connections of two specified configurations: the checkmark and the 3-path + s (see figure 3.2) to a directed arc. A vertex at the bottom of a 3-path is at level 1, the middle vertex is at level 2, and the top vertex is at level 3. For vertices which are not ncontained in any directed 3-path, we say their level is indeterminate. Now, for a single subgraph H, including a vertex of indeterminate with respect to H, we can specify several different classes, indexed by the level of this vertex in the entire graph, plus a class for when this vertex remains of indeterminate level.

The configurations, in order, will be the superbell, houses, the Vs, the int.bell, the open bell, the three-path, the Y +s, the intersecting lambda +s, the disjoint lambda +s, the diamond +s, the Y, the intersecting lambda, the diamond, the disjoint lambda, the checkmark,

6-vertex configurations:[a]		
	singleton[b]	directed arc
(6,4)	12	707
(5,3)	21	755
(4,2)	27	795
(3,1)	30	801
(2,0)	30	740
(2,-1)	27	930
5-vertex configuration:[c]		
	singleton[d]	directed arc
path	24	441

[a]Removing those that contain a 6-path.
[b]Removing those that contain a 5-path plus a directed arc, or a previous 6-configuration plus a singleton.
[c]Removing those that contain a 6-path.
[d]removing those that make a 6-configuration.

Figure 2: Paths of Length Five: Summary.

5-vertex configurations:[a]		
	singleton[b]	directed arc
(5,3)	10	368
(4,2)	17	404
(3,1)	21	412
(2,0)	22	386
(2,-1)	22	488
(3,0)	24[c]	514
(3,-2)	24[d]	571
4-vertex configuration:[e]		
	singleton[f]	directed arc
path	13	144

[a]Removing those that contain 5-paths
[b]Removing those that contain a 4-path plus a directed arc
[c]Also removing previous 5-configurations.
[d]Also removing previous 5-configurations.
[e]Removing those that contain a 5-path.
[f]Removing those that form a 5-configuration

Figure 3: Paths of Length Four: Summary.

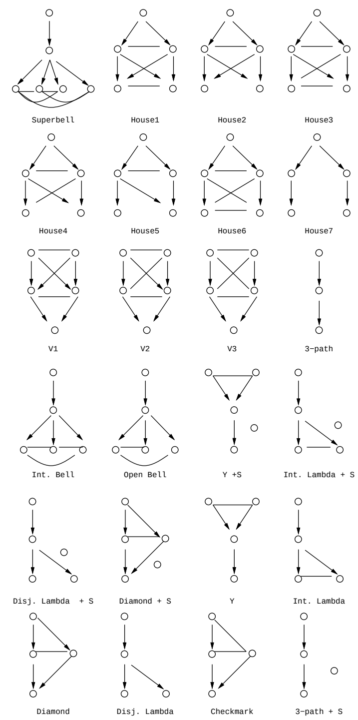

Figure 4: Specified configurations for a path of length three.

and the three-path +s. These configurations are as they appear in Figure 3.2. The connections to a singleton or directed arc are summarized below in Figure 5. In the numbers cited in the figure, we have removed connections that make four-paths from everything.

What remains, is full graphs with paths of vertex length 2, which still are not in the dominant class. These must contain either subgraphs A or B. First, we deal with anything that contains subgraph A. We partition it into A or A1, A2 and directed arcs and singletons, as usual with the smallest possible number of singletons. Here A is the graph with two directed arcs, both completely disjoint. A1 are the two configurations that add a vertex to be contained in one of the two directed arcs (these two bottom vertices can be disjoint or intersect). A2 are two configurations that add a vertex to contain the bottom vertex in one of the directed arcs (there are 4 such legal configurations, depending on the intersection pattern with the remaining directed arc; choose any two, we choose for example A2 with the new vertex entirely disjoint from the remaining directed arc, and A2 with the new vertex containing the bottom vertex in the remaining directed arc (see figure).)

Next, we deal with graphs which do not contain A, but which do contain a triangle. Here a triangle is two vertices both containing a third. We partition such graphs into a triangle, directed arcs, and singletons, minimizing singletons.

Graphs which fail to be in the dominant class and do not contain A must contain B. We now deal with those which have B but no triangle, since we have already dealt with any that contain a triangle.

We partition them into B plus arcs and singletons, and consider the number of connections

	no. vertices	singleton[a]	directed arc[a]
superbell	6	51[b]	2073
Houses and Vs		[c]	
house1	5	18	397
house2	5	18	461
house3	5	18	477
house4	5	18	546
house5	5	18	523
house6	5	18	563
house7	5	18	505
V1	5	20	342
V2	5	20	407
V3	5	20	432
int. bell	5	24[d]	507
open bell	5	25[e]	555
3path	3	6[f]	42[g]
+ ∅ configurations:		[h]	
Y + s	5	20	444
int. lambda + s	5	20	512
dis. lambda + s	5	18[i]	604
diamond + s	5	19	570
The 4-configurations:		[j]	
Y	4	7	165
int. lambda[k]	4	13	160
dis. lambda[l]	4	11	167
diamond[m]	4	12	157
3path + ∅[n]	4	12	¡ 169[o]
checkmark[p]	4	12	¡ 169[q]

[a] All counts remove those that contain a path of length 4.
[b] Removing those that contain a bell plus a directed arc.
[c] Removing those that contain a 4-configuration plus a directed arc
[d] Removing those that contain superbells, or lambda plus a directed arc.
[e] Removing those that contain superbells, lambda plus a directed arc, or int. bells plus singleton.
[f] Removing those that contain any 4-configuration.
[g] Removing those that contain a House, V, or bell.
[h] Removing those that contain a 4-configuration plus a directed arc.
[i] Also removing those that contain an int. lambda plus a directed arc.
[j] Removing those that contain a 3-path and a directed arc, bells, or a + s configuration.
[k] Removing those that contain Y.
[l] Removing those that contain Y or int. lambda.
[m] Removing those that contain Y or lambdas
[n] Removing those that contain Y or lambdas or diamond, plus 2 3-paths.
[o] When further partitioned by possible level of disjoint vertex in entire graph.
[p] Removing those that contain Y or lambdas or diamond or 3path + s
[q] When further partitioned by possible level of disjoint vertex in entire graph.

Figure 5: Paths of Length Three: Summary.

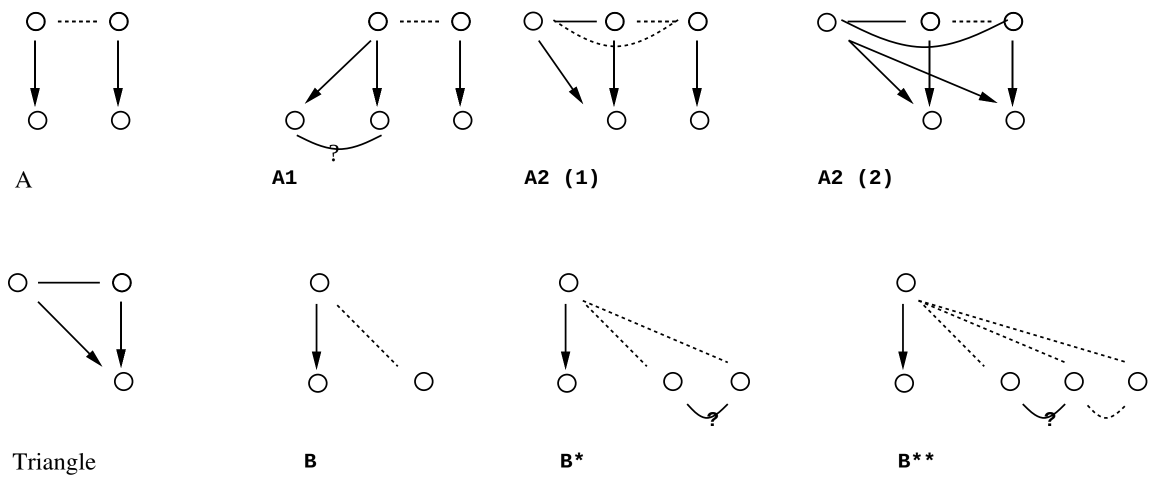

Figure 6: The A and B configurations.

of B to a singleton or an arc. We remove from the B graphs the two graphs of the form B* (corresponding to the two singletons intersecting or being disjoint), and from B*, a single graph B** (the new singleton is entirely disjoint from the rest of the configuration). For B** we again use the level idea to hold down the number of connections of a singleton. All results are summarized in Figure 7.

4 The recursion

The recursion takes as input the first 6 values for FG(n) computed exactly. There is 1 legal full graph on 1 vertex, 4 on two vertices, 41 on three vertices, 916 on four vertices, 41,099 legal full graphs on 5 vertices, and 3,528,258 full graphs on 6 vertices. We carried out the recursion by means of a program which works works with the logarithms of the numbers, rather than the numbers themselves, for computational purposes, where the log of a sum of k terms is upperbounded by k times the maximum term. We balance the recursion as follows. Consider the following 8 terms, grouped by size, which account for all the configurations in our construction, where m here is the maximum of (maximum number of connections to a singleton, the square root of the number of connections to a directed arc:

1. the 6-vertex configuration with r=51
2. 7 6-vertex configurations with $r \leq 49$
3. 2 5-vertex configurations with $r = 26$
4. 1 5-vertex configurations with $r \leq 25$
5. 27 5-vertex configurations with $r \leq 24$
6. 11 4-vertex configurations with $r \leq 13$

	singleton	directed arc
A configurations:		
A1 (disjoint)	26	465
A1 (intersect)	26	385
A2 (disjoint)	24	293
A2 (contain bot.)	24	328
A	13	121
triangle	6	42
B configurations:		
B (3 vertices)	6	25
B* (4 vertices)	13	69
B** (5 vertices)	< 24[a]	256

[a] Partitioning by the level of the singleton in the entire graph

Figure 7: Paths of Length Two: Summary.

7. 3 3-vertex configurations with $r \leq 6.5$

To compute FG(i), for a t-vertex term, we take the number of configurations of size t that appear in our canonical form, and multiply by $\binom{n}{t}t!$ times the recursive upper bound for FG($i-t$) times the maximum number of ways to FG($i-t$) to H, which we upperbound by the r listed above for H, raised to the $i-t$. For an upper bound on the dominant class, we compute each of the n terms in the sum, take the maximum, and multiply by n. Then we compare all terms (the t-configuration terms, and the upper bound on the dominant class), take the maximum, and multiply by the number of terms. We note that we have heuristically balanced the terms above quite well: term 8 dominates for the first dozen values of n, then term 7 and term 8 take turns for n in the mid-twenties to thirties, then 7 dominates uncontested until the mid-fifties, when 3 becomes the dominant term. In the 150s, 3 vies with the first term for dominance, which is then dominant thereafter. We compare what we get (without adding in the upper bound to the dominant class, above) to the lower bound on the dominant class, which just consists of the maximum of the n terms. Then we add back in the upper bound to the dominant class and make this our new estimate for $FG(i)$.

We remark that we have balanced the terms uniformly, for all n less than 170, and that a slightly better upper bound could be achieved by balancing these terms differently for different ranges of n. However, since such tinkering will definitely not push our bound in any case much below 160, and since we believe the true answer is probably closer to the lower bound, we are content with this upper bound.

In the next section, we give an explicit construction for a class of legal full graphs which lie outside the dominant class, but which accounts for a larger proportion of full graphs than the dominant class, for $n < 17$.

5 The lower bound

Consider the following class of full graphs of size n. Take a member of the $D(n-2)$. An additional two vertices, joined by a directed arc, are connected to the other $n-2$ vertices as follows: vertices in the clique of size k have three choices as to how they connect to these two vertices, either they intersect both, intersect the top and contain the bottom, or contain both. The remaining vertices also have three choices as to how to connect to these two vertices: they can be disjoint from both, intersect only the top, or intersect both. For each fixed k less than $n-2$, for each member of the dominant class on $n-2$ vertices, there are $\binom{n-k}{2}$ ways to name the additional vertices, two ways to arrange them in a directed arc, times 3^{n-2} ways to connect them. On the other hand, for the dominant class there are 4^{n-2} ways to connect them. We calculate all n terms of the dominant class and our construction exactly. These graphs dominate over the dominant class for $n < 17$.

6 Conclusions and open problems

We have shown both upper and lower bounds on the value where the asymptotic region commences. The most obvious open problem is to try to narrow the gap between the upper and lower bounds presented here. We conjecture that the lower bound is closer to the exact value.

Another open problem is to look at the same question as we did for Full graphs in this paper, for partial orders. Kleitman and Rothschild [4] provide an asymptotic enumeration of partial orders, including the specification of a predominant class of partial orders for n large enough. A more recent paper of Erné and Stege [1] counts exactly the number of legal partial orders on n vertices, for $n \leq 14$. What is interesting, is that the asymptotic behavior proved by Kleitman and Rothschild is not yet in evidence for n this small. Thus stopping at $n = 14$, Erné and Stege have not yet reached the asymptotic region: is 14 anywhere near the value of n where asymptotic behavior will begin to be observed? We pose the open question of finding tight bounds on the smallest n for which we get the asymptotic behavior in the enumeration of partial orders.

Acknowledgment

Many thanks to Craig Fields, for expertly answering some of our C questions.

References

[1] M. Érne and K. Stege. Counting finite posets and topologies. *Order*, 8:247–265, 1991.

[2] D. R. Fulkerson and O. A. Gross. Incidence matrices and interval graphs. *Pacific Journal of Math*, 15:835–855, 1965.

[3] D. J. Kleitman, F. Lasaga, and L. Cowen. Asymptotic enumeration of full graphs. *Journal of Graph Theory*, 20:59–69, 1995.

[4] D. J. Kleitman and B. L. Rothschild. Asymptotic enumeration of partial orders on a finite set. *Transactions of the American Math. Society*, 205:205–220, 1975.

[5] J. F. Lynch. The visually distinct configurations of k sets. *Discrete Mathematics*, 133:281–287, 1981.