THE SEQUENTIAL GENERATION OF RANDOM EDGE MAXIMAL *f*-GRAPHS AS A FUNCTION OF *f*

Krystyna T. BALIŃSKA
Computer Science Center, The Technical University of Poznań, Pl. M. Skłodowskiej-Curie 5, 60-965 Poznań, Poland

and

Louis V. QUINTAS
Mathematics Department, Pace University, Pace Plaza, New York, NY 10038, USA

Abstract

A graph with no vertex of degree greater than f is called an *f-graph*. An f-graph to which no edge can be added without introducing a vertex of degree greater than f is called an *edge maximal f-graph*. We consider the following procedure. Starting with n labeled vertices and no edges, sequentially add edges one at a time so as to obtain at each step a labeled f-graph. At each step, the edge to be added is selected with equal probability from among those edges whose addition would not violate the f-degree restriction. A terminal graph of this procedure is a *sequentially generated* random edge maximal f-graph. Let $P(m; n; f)$ denote the probability that a sequentially generated random edge maximal f-graph of order n has m vertices of degree less than f. The determination of the distribution $P(m; n; f)$ is an open problem posed by P. Erdős. We have obtained various insights concerning $P(m; n; f)$. In particular, we conjecture the form of $P(m; n; f)$ as a function of f.

1. Introduction

A graph (no loops, no multiple edges) on n vertices with no vertex of degree greater than some nonnegative integer f is called an *f-graph*. Although we do not develop this here, we note that classes of random f-graphs with $f < n - 1$ have many applications in the study of physical systems. For example, see [1,2] and references therein.

The following procedure for generating a random f-graph models a non-reversible process having a form that is of interest in chemistry [3,4]. This type of model is sometimes referred to as a kinetic model, in contrast to an equilibrium model in which each of the edges of the random graph are present with some given probability. The procedure we present here has considerable mathematical interest in that it satisfies the criterion of being simple to describe but nevertheless possesses many aspects that are difficult to analyze theoretically.

A random f-graph generation procedure

Let n, N and f be positive integers such that $2 \leq f \leq n-1$ and $N \leq fn/2$.

(1) Start with n labeled vertices.

(2) From the set of all possible edges on these vertices, sequentially select edges so as to obtain at each step a labeled graph G on n vertices such that

(a) no vertex of G has degree greater than f

and

(b) each edge selected in the sequential generation of G is chosen with equal probability from the set of edges that will satisfy condition (a).

(3) Continue this procedure until either G is an f-graph with N edges or G is a graph to which no edge can be added without producing a vertex having degree greater than f. The latter graph is called an *edge maximal f-graph* [2], which we henceforth denote by *EM f-graph*.

We shall refer to the above as *procedure S*.

Observe that the Markov chain defined by procedure S with $f = n-1$ is analogous to the random graph process described in [5, p. 38]. However, the bounded degree restriction with $f < n-1$ leads to significantly different types of problems.

There are many unanswered questions about f-graphs which are generated by processes modeled by procedure S. One such question, posed by P. Erdős, is the following:

> If sequentially generated random EM f-graphs of order n are partitioned in accordance with their number of vertices of degree less than f, what is the probability distribution of these equivalence classes?

In [6] we set the stage for our investigation of sequentially generated f-graphs. There we presented an algorithm that realizes procedure S. Using data obtained from this algorithm we derived, for $f = 2, 3$, and 4, approximation functions for the probability distribution of equivalence classes of EM f-graphs, the latter being partitioned as stated in the Erdős question. These approximations were studied as functions of n ($n < 250$) for a given f. A summary of the results is given in [7]. In [8] a similar study was carried out for $f > 4$. In particular, it was found that the probability of the f-regular class (fn even) is monotonically increasing with respect to n.

In [9] we obtained approximations for these probability distributions as functions of f for a given n. In what follows, we present various insights concerning these distributions.

2. The number of orexic vertices

We use the word *orexic*, introduced in [2], to designate a vertex of degree less that f in an EM f-graph.

THEOREM 2.1

Let G be an n vertex EM $(n-t)$-graph with m orexic vertices and $2 \le t \le n-2$. Then,

$$m \le \min\{n(t-1)/(2t-1),\, n-t\}.$$

Proof

Since the m orexic vertices in an EM $(n-t)$-graph form an induced K_m, we have

$$m \le n-t. \tag{2.1}$$

The vertices in G^c, the complement of G, can be partitioned into three sets M, A, and N, where M consists of the vertices orexic in G, A the vertices that are adjacent in G^c to the vertices in M, and $N = V(G^c) - (M \cup A)$.

Let the m orexic vertices have degree sequence $s_1 \le s_2 \le \ldots \le s_m$ in G. Then, in G^c, the vertices in M have degree sequence $d_1 \le d_2 \le \ldots \le d_m$, where $d_j = n-1-s_{m+1-j}$. Since $s_{m+1-j} \le n-t-1$, it follows that $d_j \ge t$. In G^c, the vertices in A each have degree $t-1$. Furthermore, since the vertices of M are independent and each vertex of A is adjacent to a vertex of M, we have

$$(t-1)|A| \ge \sum_{j=1}^{m} d_j \ge mt. \tag{2.2}$$

Clearly, $n \ge m + |A|$, which combined with (2.2) yields

$$n \ge m + mt/(t-1). \tag{2.3}$$

Inequality (2.3) can be written as

$$m \le n(t-1)/(2t-1). \tag{2.4}$$

Combining (2.1) and (2.4) proves the theorem. □

COROLLARY 2.1.1

For $2 \le t \le (n+1)/2$, an n vertex EM $(n-t)$-graph has at most $n(t-1)/(2t-1)$ orexic vertices.

Proof

By theorem 2.1, m is bounded by $\min\{n(t-1)/(2t-1), n-t\}$. Thus, $m \le n(t-1)/(2t-1)$ whenever $n(t-1)/(2t-1) \le n-t$. The latter inequality holds if and only if $2t^2-(n+1)t \le 0$, that is, if and only if $0 \le t \le (n+1)/2$. Therefore, if $2 \le t \le (n+1)/2$, then $m \le n(t-1)/(2t-1)$. □

COROLLARY 2.1.2

An n vertex EM $(n-2)$-graph can have at most $n/3$ orexic vertices.

Proof

Let $t = 2$ in corollary 2.1.1 □

Let m denote the number of orexic vertices in a random EM f-graph of order n. Then, m is a random variable whose range is a function of n and f (see theorem 2.1). Denote by $P(m;\, n;\, f)$ the probability that a sequentially generated EM f-graph on n vertices has m orexic vertices.

Consider the three functions $X_{n,f}$, $Y_{m,f}$, and $Z_{m,n}$ defined as follows. For fixed n and f, let

$$X_{n,f}(m) = P(m;\, n;\, f) \qquad m = 0, 1, \dots .$$

That is, $X_{n,f}$ is the probability distribution function for the random variable m. For fixed m and f, let

$$Y_{m,f}(n) = P(m;\, n;\, f).$$

Then, $Y_{m,f}$ describes how $X_{n,f}$ (m) behaves as a function of n for a given f. For fixed m and n, let

$$Z_{m,n}(f) = P(m;\, n;\, f).$$

Then, $Z_{m,n}$ describes how $X_{n,f}(m)$ behaves as a function of f for a given n. By definition, for fixed m, n, and f,

$$X_{n,f}(m) = Y_{m,f}(n) = Z_{m,n}(f).$$

As noted in section 1, $Y_{m,f}(n)$ was studied in [6–8]. In this paper, our focus of attention is $Z_{m,n}(f)$.

PROBLEM 1

For fixed $n \ge f+1$ and all m, determine $Z_{m,n}(f) = P(m;\, n;\, f)$. Do the same for n going to infinity.

The cases $f = 0, 1$, and $n - 1$ are of no probabilistic interest. Thus, we need only comment on $Z_{m,n}(f)$ in the domain $f = 2, 3, \ldots, n-2$.

3. A detailed case study

It is clear that the smaller f is in relation to n, the greater will be the difference between the properties of a sequentially generated random f-graph and those of a graph generated by the random graph process of [5, p. 38]. On the other hand, if f is close to $n - 1$, then the difference in properties of such f-graphs and graph process graphs might be expected to be minimal. However, we point out that even in the extreme case $f = n - 2$ there are properties for which this difference is nontrivial. For example, an EM $(n-1)$-graph is a K_n, whereas the EM $(n-2)$-graphs, if partitioned into classes in accordance with their number of orexic vertices, fall into at most $\lfloor n/3 \rfloor + 1$ classes (see corollary 2.1.2). In particular, if $n = 241$ and f goes from 239 to 240, then the number of classes drops from 80 (there are no 239-regular graphs) to 1. Consequently, the Erdős problem is of interest for all $f < n - 1$.

Let $T = T(n, f)$ be such that $P(T; n; f) \geq P(m; n; f)$ for all m. An EM f-graph with T orexic vertices is called *predominant of type T*. The problem of determining T for a given n and f was initially investigated in [8]. Subsequently, we have found that our realizations of procedure S suggest

$$\lfloor E(m) \rfloor \leq T(n,f) \leq \lceil E(m) \rceil \quad \text{for all } n \text{ and } f$$

and

$$E(m) \leq 2 + (n - 12)(f - 2)/6(n - 4) \quad \text{for } n > 50,$$

where $E(m) = \sum_{m \geq 0} mP(m; n; f)$ is the expected value of m.

In fig. 1, we display the combination of these observations with theorem 2.1, our result on the number of orexic vertices in an EM f-graph of order n. Namely, the function **I** shows the upper bound for values of m for any distribution, that is, this upper bound depends only on the graph structure of an EM f-graph, not on the stochastic process in which such a graph is being considered. The line segment **II** from (2, 2) to $(n - 2, n/6)$ is our conjectured upper bound for both T and $E(m)$ when $n > 50$ suggested by our data for sequentially generated EM f-graphs.

Using the SG-Algorithm described in [6], we generated data to obtain the relative frequency $RF(m; n; f)$ for various values of m, n, and f. The same values of n were chosen for both fn even and fn odd: $n = 31$, 51, 101, and 241. The data were generated using samples of size 10,000 graphs for $n < 241$ and of size 5000 graphs for $n = 241$; thus, the error at $p = 0.5$ is at most 0.0098 and 0.0139, respectively.

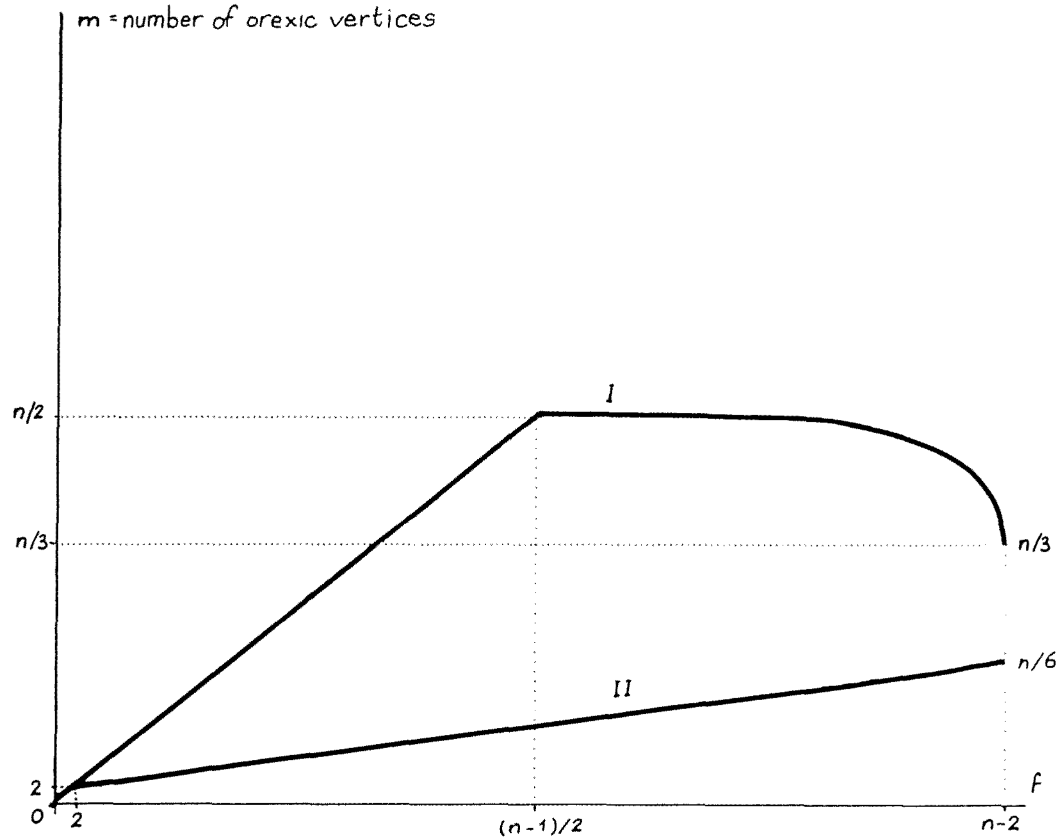


Fig. 1. EM f-graphs of order n: **I** – the upper bound for the number of orexic vertices for any distribution and **II** – an upper bound for both T and $E(m)$ ($n > 50$) for the sequential generation.

As an illustration, we describe in detail $RF(m;\, n;\, f) = RF_{m,n}(f)$ as a function of f for $n = 241$. The tables of these data are given in appendices A1 and A2 of [9]. The behavior of $RF_{m,n}(f)$ seen here is representative in that the shape of the data for various values of n is in general the same. The plot of the data for $n = 241$ is given in fig. 2 for both fn even and fn odd. The figure clearly shows the relative positions of these data for different values of m. The significant difference between these cases is seen in the data for $m = 0$, 1, and 2 and $f < 40$. For example, there are no f-regular graphs when fn is odd and the $m = 1$ data behavior is clearly seen to be quite different. In fig. 3 we show the same data as a function of m, and in fig. 4 the mean $\bar{m}$ and variance s^2 of m when fn is even.

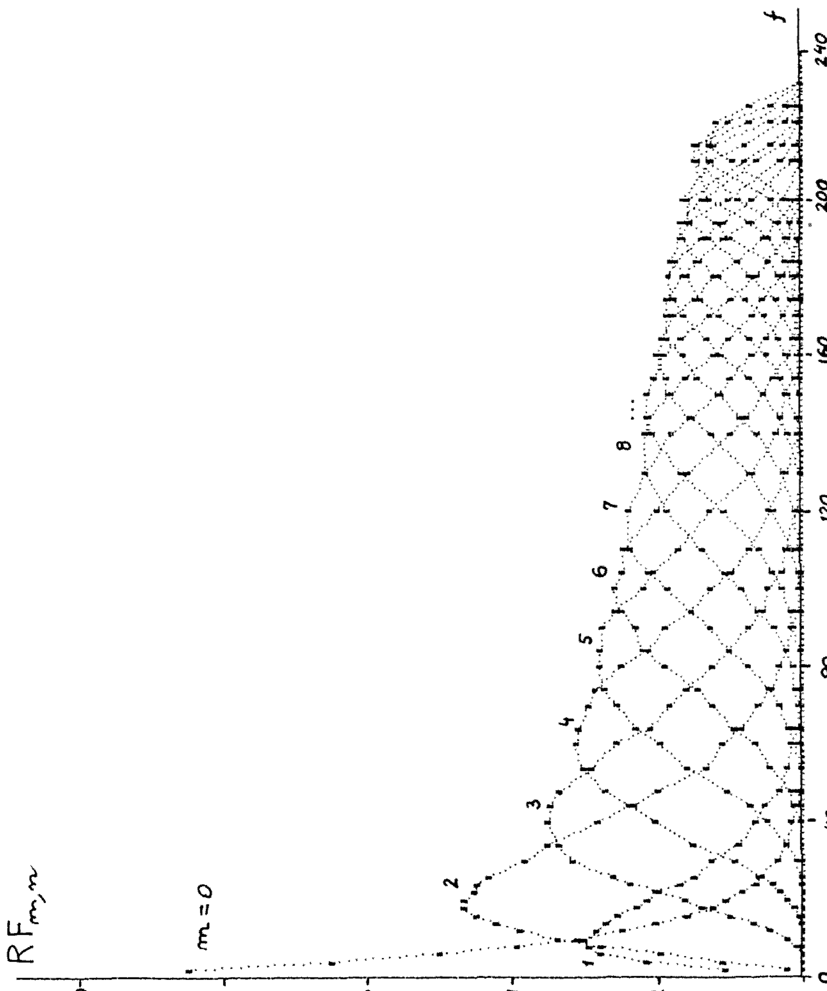


Fig. 2(a). $RF_{m,n}(f)$, $m = 0, 1, \ldots, 20$; $n = 241$, fn even.

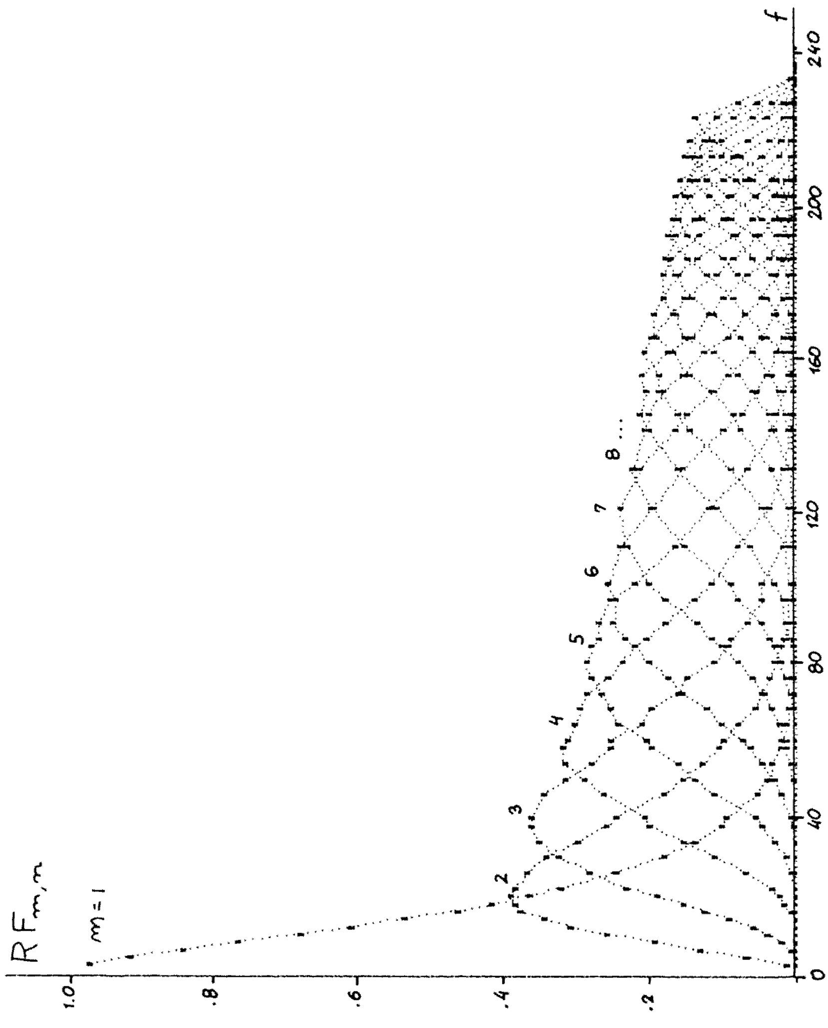


Fig. 2(b). $RF_{m,n}(f)$, $m = 1, 2, \ldots, 21$; $n = 241$, fn odd.

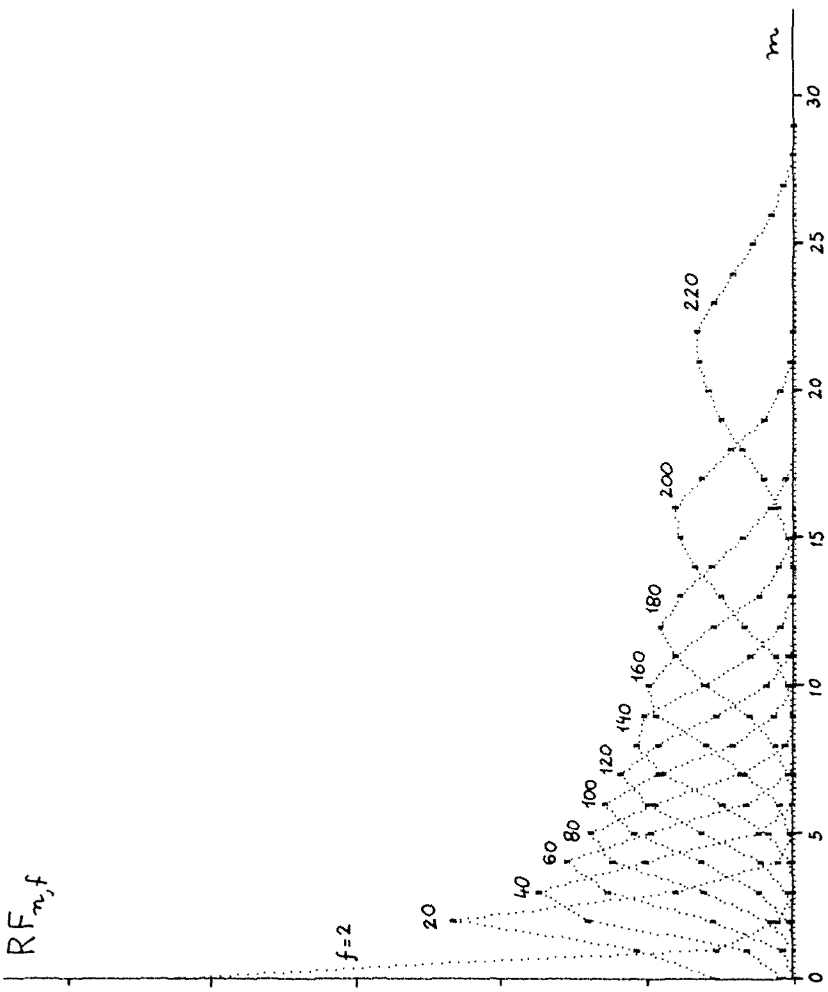


Fig. 3. $RF_{n,f}(m)$, $f = 2, 20, 40, \ldots, 220$; $n = 241$, fn even.

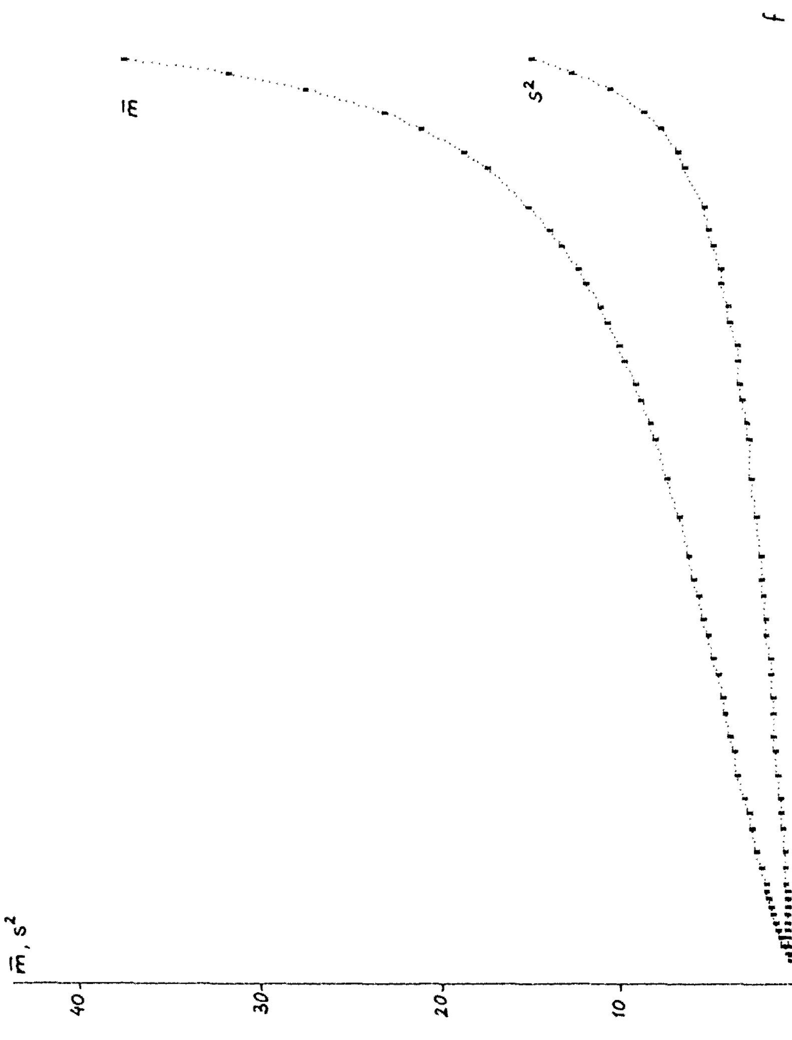


Fig. 4. The mean and variance of m; $n = 241$, fn even.

4. Approximation functions

In [9] we were led to approximation functions for $Z_{m,n}(f)$ of the form $Af^B\exp(-Cf)$, $A > 0$, $C > 0$, and $2 \leq f \leq (n-1)/2$. For $(n-1)/2 < f \leq n-2$, we conjecture the approximation functions are of the same form but with different coefficient values for a given value of m. We believe that one reason for this is the fact that the strictly graph-theoretical upper bound for the number of orexic vertices changes abruptly from a linear to a nonlinear bound at $f = (n-1)/2$ (see fig. 1). We hold in reserve further speculation at this time.

For $n = 241$ and $m \leq 3$, the coefficients A, B, and C were computed using the least-squares method. The results are shown in table 1. In fig. 5 we display this function for both fn even and fn odd.

Table 1

The coefficients for $Z_{m,n}(f)$ for $n = 241$, $m = 0, 1, 2, 3$ with fn even, and $m = 1, 2, 3$ with fn odd

fn even	m	A	B	C
	0	1.15	−0.12	0.100
	1	0.87×10^{-1}	0.98	0.098
	2	1.72×10^{-2}	1.66	0.083
	3	1.75×10^{-5}	3.69	0.092
fn odd	**m**	**A**	**B**	**C**
	1	0.99	0.20	0.076
	2	0.52×10^{-2}	2.05	0.091
	3	2.21×10^{-5}	3.62	0.090

We can make the following general observations and conjectures about the behavior of the coefficients of the approximation function $Af^B\exp(-Cf)$ for $2 \leq f \leq (n-1)/2$.

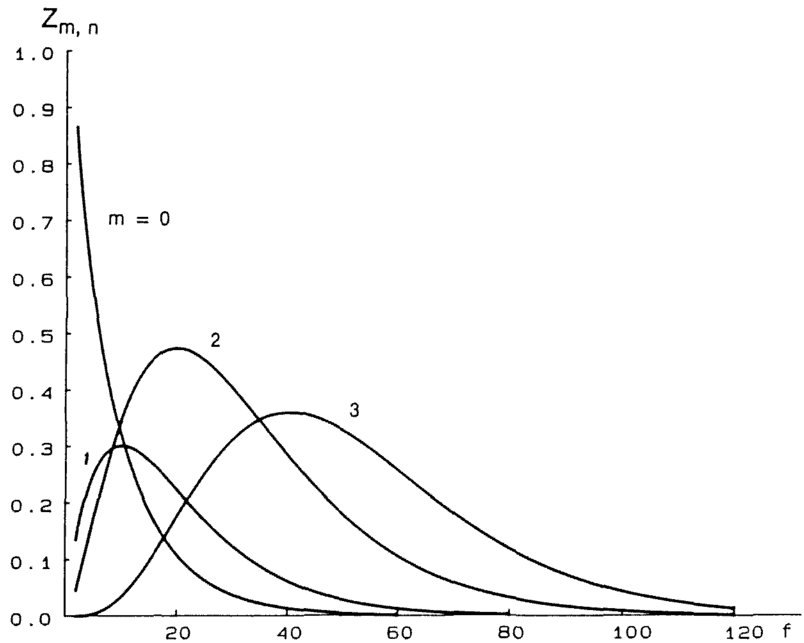
The coefficients A, B, and C are positive for all n except for the case corresponding to f-regular graphs, that is, when $m = 0$. In the latter case B is negative.

For increasing large n and $m = 0, 1, 2$, or 3, the coefficients B and C decrease and A increases.

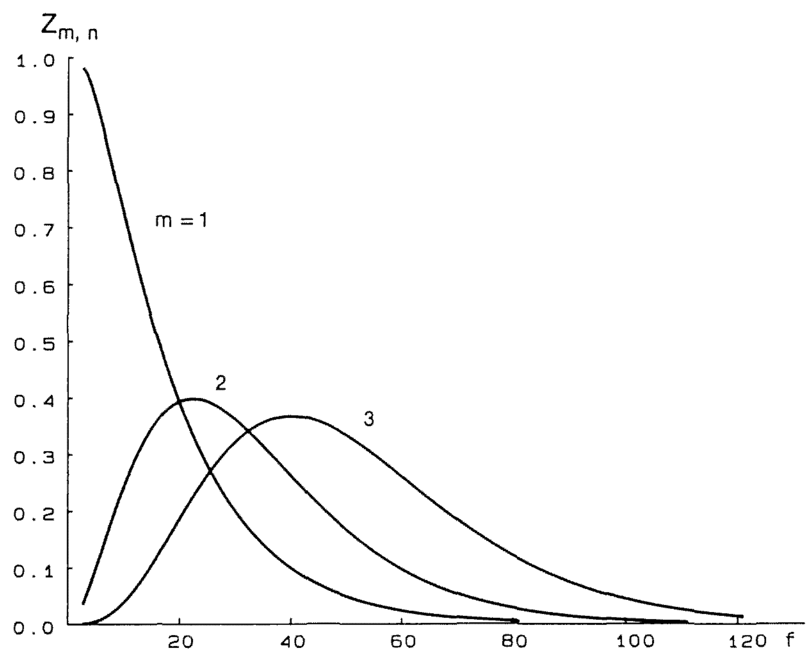
If n is fixed and large, then as m increases, C tends to a constant, B increases linearly, and A decreases rapidly.

If $m \geq 3$ is fixed, then the maximum value of $Af^B\exp(-Cf)$ and the value $f = B/C$ at which this maximum occurs both increase with increasing n.

Our results clearly suggest the conjecture that for fixed m, n, and $2 \leq f \leq (n-1)/2$, we have $Z_{m,n}(f) = Af^B\exp(-Cf)$, where $A = A(m, n)$, $B = B(m, n)$ and



(a)



(b)

Fig. 5. (a) $Z_{m,n}(f)$, $m = 0, 1, 2$, and 3; $n = 241$, fn even.
(b) $Z_{m,n}(f)$, $m = 1, 2$, and 3; $n = 241$, fn odd.

$C = C(m,n)$ are functions of m and n. Assuming the verification that $Z_{m,n}(f)$ is of this form, the additional following problem is posed.

PROBLEM 2

Determine analytic expressions for $A(m, n)$, $B(m, n)$, and $C(m, n)$ and/or determine the limits of the coefficients as n goes to infinity.

Acknowledgements

Support of this work for L.V.Q. was provided in part by Dyson College of Arts and Sciences Faculty Research and Pace University Scholarly Research Committee Grants. Both authors appreciate the support and hospitality provided by The Institute of Mathematics of the Technical University of Poznań (April 1988) and The Higher College of Engineering of Zielona Góra (January 1989).

References

[1] K.T. Balińska and L.V. Quintas, Random graph models for physical systems, *Graph Theory and Topology in Chemistry*, Studies in Physical and Theoretical Chemistry 51 (Elsevier, Amsterdam, 1987), pp. 349–361.

[2] J.W. Kennedy and L.V. Quintas, Probability models for random f-graphs, *Combinatorial Mathematics New York 1985*, Ann. N.Y. Acad. Sci. 555(1989)248–261.

[3] H. Galina and A. Szustalewicz, A kinetic theory of stepwise crosslinking polymerization with substitution effect, Macromolecules 22(1989)3124–3129.

[4] J. Mikeš and K. Dušek, Simulation of polymer network formation by the Monte Carlo method, Macromolecules 15(1982)93–99.

[5] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).

[6] K.T. Balińska and L.V. Quintas, The sequential generation of random f-graphs. Preliminaries, an algorithm, and line maximal 2-, 3-, and 4-graphs (presented in Poznań, April 1988), Mathematics Department, Pace University, New York, NY (October 7, 1988).

[7] K.T. Balińska and L.V. Quintas, The sequential generation of random f-graphs. Line maximal 2-, 3-, and 4-graphs, Comput. Chem. 14, 4(1990)323–328.

[8] K.T. Balińska and L.V. Quintas, The sequential generation of random f-graphs. Distributions and predominant types of line maximal f-graphs for $f > 4$, presented at French–Israeli Conference on Combinatorics and Algorithms, Jerusalem (Nov. 13–17, 1988).

[9] K.T. Balińska and L.V. Quintas, The sequential generation of random f-graphs. Distributions of edge maximal f-graphs as functions of f, Computer Science Center Report No. 316, The Technical University of Poznań (May 15, 1989; revised May 18, 1990).