Network transitivity and matrix models

Z. Burda,[1] J. Jurkiewicz,[1] and A. Krzywicki[2]
[1]*M. Smoluchowski Institute of Physics, Jagellonian University, Reymonta 4, 30-059 Krakow, Poland*
[2]*Laboratoire de Physique Théorique, Bâtiment 210, Université Paris-Sud, 91405 Orsay, France*
(Received 13 October 2003; published 20 February 2004)

This paper is a step towards a systematic theory of the transitivity (clustering) phenomenon in random networks. A static framework is used, with adjacency matrix playing the role of the dynamical variable. Hence, our model is a matrix model, where matrices are random, but their elements take values 0 and 1 only. Confusion present in some papers where earlier attempts to incorporate transitivity in a similar framework have been made is hopefully dissipated. Inspired by more conventional matrix models, analytic techniques to develop a static model with nontrivial clustering are introduced. Computer simulations complete the analytic discussion.

DOI: 10.1103/PhysRevE.69.026106 PACS number(s): 05.50.+q, 05.40.−a, 02.50.Cw, 87.18.Sn

I. INTRODUCTION

Network model builders are currently adopting one of the two complementary approaches. Either a network is constructed step by step, by adding successive nodes and links, or else, what is constructed is a static statistical ensemble of networks. Each of these two approaches has its merits and shortcomings. Evolving network models shed light on the growth dynamics, while static ensembles are more appropriate for the study of structural traits. The classical model of Erdös and Rényi [1] has been generalized so as to incorporate arbitrary degree distributions and even some correlations, but a serious shortcoming of the static models proposed so far is that they do not capture the common feature of most real networks: neighbors of a randomly chosen node are directly linked to each other much more frequently than by chance, so that many short loops appear. The networks tend to have locally a tree structure (see Refs. [2–4]). And, as pointed out in Ref. [4], ''for general networks we currently have no idea how to incorporate transitivity into random graph models.'' In this paper we fill this gap, at least partially. The attention of the reader should be called to the very recent Refs. [5–7], where the clustering problem is also addressed, but following very different avenues.

Graphs are a mathematical representation of networks. For definiteness we consider in this paper undirected graphs only. Let us denote by N the number of nodes in a graph and by $M=\{M_{ij}\}, i,j=1,2,\ldots,N$ the symmetric incidence matrix, with $M_{ii}=0$ on the diagonal and $M_{ij}=1$ or 0 depending on whether the nodes labeled i and j are connected or not. Whole information about a graph is encoded in its adjacency matrix. A general random graph model can be defined by introducing the partition function (see, for example, Ref. [8]):

$$Z=\sum_{M} e^{S(M)}\delta(\mathrm{Tr}(M^2)-2L), \tag{1}$$

where L is the number of links and $S(M)$ is a function which we will call the action. The sum is over all possible adjacency matrices M. The simplest choice is $S(M)=0$. The corresponding graphs are those of the classical theory of Erdös and Rényi [1]: the value of the ratio L/N determines a variant of the model.

Probably the simplest extension of the classical theory consists in setting $S(M)=g\mathrm{Tr}(M^3)$, directly proportional to the number of triangles. This has already been attempted many years ago by Strauss [9]. His results are summarized in the recent review [4]: ''There is however, one unfortunate pathology .... If, for example, we include a term in the Hamiltonian that is linear in the number of triangles in the graph, with an accompanying positive temperature favoring these triangles, then the model has the tendency to condense, forming regions of the graph that are essentially complete cliques—subset of vertices within which every possible link exists ... Networks in the real world however do not seem to have this sort of clumpy transitivity.''

It appears to us that this negative conclusion, which faithfully reflects the content of Ref. [9], is not quite right. There is nothing wrong in Strauss's work. However, it is very incomplete and due to this incompleteness involuntarily misleading. One of the aims of our paper is to give a fresh and comprehensive discussion of Strauss's model.

The essence of Strauss's argument is as follows: assuming that the ratio L/N is kept constant, one can easily convince oneself that there exist pathological configurations for which $\mathrm{Tr}(M^3)\propto N^{3/2}$. The contribution of such a configuration to the partition function is explosive in the large N limit, since it cannot be tamed by the entropy factor falling roughly speaking like the inverse of the number of graphs, i.e., like $\exp(-\mathrm{const}\times N\ln N)$. Thus, however small the coupling g is, the only stable states of the system are the pathological ones, provided the system is large enough.

As we will show later on, the pathological crumpled states—the Strauss phase—are separated from a smooth phase by a barrier that grows with increasing N. If the system is prepared in the smooth phase, it has a very tiny probability to roll out over the barrier to the Strauss phase. This probability tends rapidly to zero in the thermodynamic limit. Strauss has missed this point, because the systems he simulated were too small to signal the relative stability of the smooth phase. Now, for all practical purposes one can work in the smooth phase, ignoring the instability. This is what one does on many occasions in physics, in particular in the context of matrix models, where the instability also goes away when the matrix size tends to infinity.

We have mentioned matrix models on purpose. The theory of random matrices is an important branch of statistical physics, with applications ranging from nuclear physics to string theory. Some of the techniques developed in this theory can be adapted to a study of the model defined by Eq. (1). This is also a matrix model, albeit dealing with rather special matrices: in standard matrix models the matrix elements are continuous random variables.

The form (1) of the partition function turns out to be very convenient for numerical simulations. In analytical calculations it will be convenient to use a slightly different formulation of the model, getting rid of the δ function and allowing small fluctuations of the number of links L. The partition function Z will be, up to a factor, the average of $\exp(S)$ in Erdös-Rényi theory. We first assume that a link is occupied with probability p. Hence, for given N and L the Erdös-Rényi weight is

$$p^L(1-p)^{N(N-1)/2-L}=(1-p)^{N(N-1)/2}\left(\frac{1}{p}-1\right)^{-L}. \quad (2)$$

This primary weight is further multiplied by $\exp(S)$. Inserting Eq. (2), integrating over L, and neglecting an irrelevant factor, we obtain the modified partition function

$$\tilde{Z}=\sum_M \exp\left[-\frac{1}{2}\ln\left(\frac{1}{p}-1\right)\mathrm{Tr}(M^2)+S(M)\right]. \quad (3)$$

In short, we have traded the δ function for a Gaussian.

In most of this paper we set $S(M)=g\mathrm{Tr}(M^3)$, as in Ref. [9]. Thus, formally and up to a rescaling of the dynamical variable the model looks like the much studied matrix model

$$Z_{\mathrm{matrix}}=\int dM \exp\left(-\frac{1}{2}\mathrm{Tr}(M^2)+g\mathrm{Tr}(M^3)\right), \quad (4)$$

where one integrates over all possible symmetric $N\times N$ matrices. The difference is in the integration measure, which is discrete in Eq. (3) and continuous,

$$dM=\prod_{i\leq j} dM_{ij}, \quad (5)$$

in Eq. (4). This difference is crucial, of course, but we would rather like to insist on the similarities between the two models. In any case, the example of the matrix model is for us a guide in our study.

The plan of the paper is as follows. In Sec. II we develop $\exp(S)$ in powers of S and discuss the features of the perturbation series obtained by integrating term by term. In Sec. II A we recall how the behavior of the perturbation series reflects the existence of an instability of the theory, by considering two examples. In Sec. II B we introduce a helpful diagrammatic representation of the perturbative contributions to the partition function. These diagrams are counted in Sec. II C. It is argued that at finite N the perturbation series is pathological, indicating that nonperturbative phenomena are in action. However, keeping only the terms that are nonvanishing in the limit $N\to\infty$ one gets, like in the matrix model (4), a convergent series. This series is summed in Sec. II D. We obtain a simple analytic formula for the average number of triangles. We also show that the introduction of the interaction $g\mathrm{Tr}(M^3)$ leaves the degree distribution unmodified. The nonperturbative dynamics is studied in detail in Sec. III, using the Monte Carlo technique of numerical simulation. In a range of model parameters we find a remarkable agreement between the data and the perturbative predictions, showing that the nonperturbative phenomena are negligible in this range. However, at large enough coupling strength the perturbation theory breaks down, as expected. The transition point has an interesting scaling with N. This enables us to define the model so as to get a nontrivial behavior of the clustering coefficient. In Sec. IV we discuss possible generalizations. This section contains also a summary of this work and a conclusion.

II. PERTURBATION SERIES

A. An analogy

Before entering into the main discussion of our problem let us consider an elementary example, to help those readers who are not conversant with field theoretic arguments:

Consider the following integral:

$$I=\sqrt{\frac{\beta}{2\pi}}\int dx e^{\beta(-x^2/2+gx^3/3-\epsilon x^4)}, \quad (6)$$

where ϵ is infinitesimal and has been introduced only in order to satisfy purists: I can be regarded as the partition function of a particle subject to the combined action of a potential and of a heat bath. Formally, the integrand resembles the summand in Eq. (3), except that the integration variable is here just a number.

Consider a random walk in the potential given by the exponent in Eq. (6). Assume that in some initial moment the particle is located at $x=0$. This is a metastable state. The particle eventually rolls over the barrier and reaches the deep minimum of the free energy at $x\approx g/4\epsilon$. As is well known, the lifetime τ of the metastable state is given by the Arrhenius formula [10],

$$\tau\approx e^{\beta/6g^2}. \quad (7)$$

The decay of the metastable configuration is a nonperturbative phenomenon. The escape time has an essential singularity as a function of the coupling g. Of course, this nonperturbative phenomenon only occurs at nonzero temperature. When $\beta=\infty$ the particle stays forever in its initial position. Notice that the transition is more a crossover than a genuine phase transition. It occurs when the exponent in Eq. (7) is of order unity, but the value of g where the transition occurs may slightly depend on how the random walk is performed.

In more complicated models the Arrhenius formula is not so readily derived. But nonperturbative dynamics shows up, if present, in the structure of the perturbation series in the coupling constant. Let us expand the exponential in Eq. (6) with respect to terms other than the quadratic one:

FIG. 1. Diagrams representing $O(g^2)$ contributions to the partition function.

$$I=\frac{1}{\sqrt{\pi}}\sum_k \frac{\Gamma\left(3k+\frac{1}{2}\right)}{\Gamma(2k+1)}\left(\frac{8g^2}{9\beta}\right)^k[1+O(\epsilon)]. \tag{8}$$

It is evident that the series coefficients grow factorially and that the series has zero radius of convergence. This is a characteristic signal. We will find a similar behavior, at finite N, in the model defined by Eq. (3).

To conclude this section let us mention what happens when the partition function instead of being a Riemann integral, like in Eq. (6), is a matrix integral, like in Eq. (4): A perturbation expansion can again be defined and the terms in the expansion can be given a diagrammatic representation. A clever manner of cataloging these diagrams has been devised by 't Hooft [11]. One first rescales $M\to\sqrt{N}M, g\to g/\sqrt{N}, \epsilon \to \epsilon/N$. One then observes that these diagrams can be drawn on a two-dimensional surface. Such a surface is always a sphere with a number of handles. Classes of diagrams are characterized by the number h of these handles and the contributions of all diagrams belonging to the same class have the same N dependence: N^{2-2h}. Summing over all h one gets a badly divergent series. However, in the limit $N\to\infty$ the spherical topology ($h=0$) dominates and the corresponding series has a nonzero radius of convergence. We will seek a similar behavior in our model. The hint is that one should carefully examine the $N\to\infty$ limit and that it may be wise to rescale the coupling constant in order to get a physically meaningful theory.

B. Diagrammatic representation

In this section we will introduce diagrams representing terms in the perturbative expansion of Eq. (3). To avoid misunderstanding let us stress from the outset that the diagrams introduced in this section are *not* to be identified with the graphs belonging to the statistical ensemble we are working with. These diagrams are just a tool helping to catalogue contributions to the partition function. We start by setting $S(M)=g\mathrm{Tr}(M^3)$.

Let us expand in Eq. (3) the factor e^S,

$$\tilde{Z}=Z_0\sum_n \frac{g^n}{n!}\langle[\mathrm{Tr}(M^3)]^n\rangle_{ER}. \tag{9}$$

Here Z_0 is the partition function in the Erdös-Rényi ensemble of random graphs and the subscript ER in $\langle\cdots\rangle_{ER}$ indicates that the average is calculated in this ensemble. Since Z_0 does not depend on our dynamical coupling g it is for us an irrelevant normalization constant.

The problem now is to calculate the averages appearing in the sum in Eq. (9). We use a method largely inspired by Ref. [12], adopting also some of their notations, like

$$N^{\underline{k}}=N!/(N-k)!, \tag{10}$$

which is the number of ways to choose k among N indices, the different permutations of the selected indices being considered as distinct. We have $\mathrm{Tr}(M^3)=\Sigma_{abc}M_{ab}M_{bc}M_{ca}$, which is up to the factor 3! the number T of triangles in the graph. We represent a matrix element M_{ab} by a line segment. Indices a,b are then associated with the ends of the segment. The product $M_{ab}M_{bc}M_{ca}$ is represented by a triangle. Notice that $M_{ab}M_{bc}M_{ca}$ is a random variable which can only take value 0 or 1. Since the diagonal elements of M are by definition 0 the probability that this product equals 1 is p^3. There are $N^{\underline{3}}$ configurations of indices which correspond to nonvanishing contributions to the sum. Hence

$$\langle\mathrm{Tr}M^3\rangle_{ER}=p^3N^{\underline{3}}. \tag{11}$$

The power of p is equal to the number of the sides of the triangle and the underlined power of N is the number of indices one is summing over. The contribution to the perturbation series is, of course, $gp^3N^{\underline{3}}$.

A complication arises as one goes to higher order of the perturbation theory. As one multiplies the traces one produces strings $M_{ab}M_{bc}\ldots M_{ef}$ where some pairs of indices, possibly interchanged like ab and ba, repeat themselves referring to the same element of the adjacency matrix. The corresponding probability factor is then p and not some power of p. One has to identify such pairs of indices. It is also necessary to identify the independent summation indices and to count their number. One has also to count in how many distinct combinations the independent indices can appear in the string. All this may seem a bit confusing and is best explained with an example.

Consider the second order term in Eq. (9), $n=2$. One deals with the strings that have the following structure

$$M_{ab}M_{bc}M_{ca}M_{de}M_{ef}M_{fd}. \tag{12}$$

When one sums over indices there appear $N^{\underline{6}}$ terms where the indices $abcdef$ are all distinct. Then, all the six index pairs are also distinct and the probability of individual strings is p^6. We illustrate this situation by drawing two triangles (because there are two traces) that are nonoverlapping, like in Fig. 1(d). The corresponding contribution to the perturbation series is

$$\text{fig 1d}\to\frac{g^2}{2!}p^6N^{\underline{6}}.$$

Another possibility is that the indices def are identical to abc but appear in a different order. For each choice of abc there are six such possibilities and there are $N^{\underline{3}}$ such choices. The probability of the string taking value 1 is p^3. We illustrate this situation by drawing two overlapping triangles, as in Fig. 1(a). Remember, that there are six manners of putting one triangle on top of another. Hence

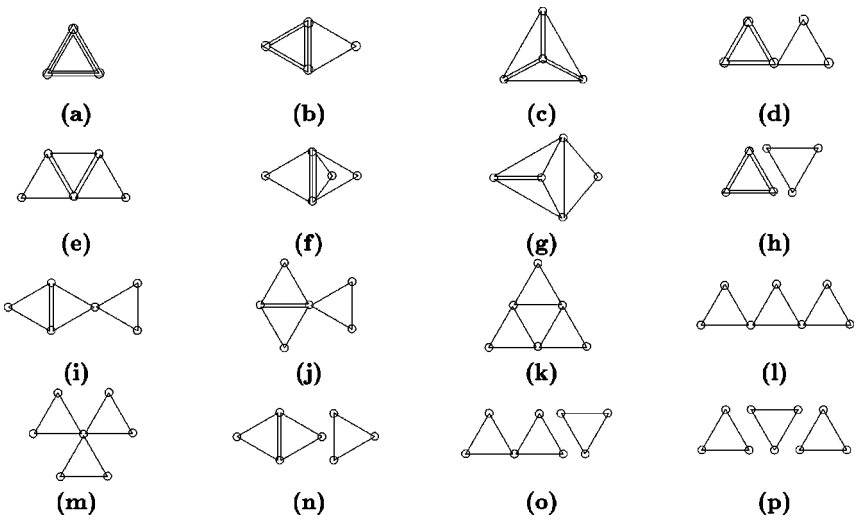

FIG. 2. Diagrams representing $O(g^3)$ contributions to the partition function.

$$\text{fig 1a} \rightarrow \frac{g^2}{2!} 6 p^3 N^{\underline{3}}.$$

Still another possibility is that two and only two indices are equal. They necessarily belong to two distinct traces. Thus one of the indices def equals either a, or b, or c: there are three possible choices. Let us take one of them, say a. Then there are three possible structures for the second trace in the product: $M_{ae}M_{ef}M_{fa}$, $M_{ea}M_{af}M_{fe}$, and $M_{ef}M_{fa}M_{ae}$. Notice that e and f are dummy indices. We illustrate this situation by drawing two triangles with one common vertex. Remember that there are three manners of attaching a triangle to a vertex of another triangle, like in Fig. 1(c). On the whole there are $3\times 3=9$ arrangements of the five independent indices and six distinct index pairs. Thus

$$\text{fig 1c} \rightarrow \frac{g^2}{2!} 9 p^6 N^{\underline{5}}.$$

Finally, two indices among def can be identical to two indices among abc. There are three choices, let us take ab. Then there are six possible structures for the second trace in the product: $M_{ab}M_{be}M_{ea}$, $M_{ba}M_{ae}M_{eb}$, $M_{ae}M_{eb}M_{ba}$, $M_{ea}M_{ab}M_{be}$, $M_{be}M_{ea}M_{ab}$, and $M_{eb}M_{ba}M_{ae}$. We illustrate this situation by drawing two triangles with one common edge, like in Fig. 1(b). Remember that attaching a triangle to two vertices can be done is six manners. There are $3\times 6=18$ arrangements of the four independent indices and five distinct pairs:

$$\text{fig 1b} \rightarrow \frac{g^2}{2!} 18 p^5 N^{\underline{4}}.$$

The game can be extended to higher order n, although the number of diagrams increases very fast. The third order diagrams are listed in Fig. 2. The general rule is that the power of p equals the number of triangle edges, the number of triangle vertices appears as the underlined power of N and, in the nth order, there is a factor $g^n/n!$. The determination of the number of independent index arrangements is rather tedious. The best way is to proceed recursively.

Using the diagrams one can actually forget about indices. It is sufficient to construct the diagrams of order n by adding one triangle to the diagrams of order $n-1$ in all possible manners. One has to multiply the number of arrangements factor in the target diagram of order $n-1$ by the number of ways the new triangle can be attached to it. These numerical factors should be added when a given diagram of order n can be constructed from several diagrams of order $n-1$. We repeat again the rules: free triangle, factor 1; triangle attached to one vertex, factor 3; triangle attached to a pair of vertices, factor 6; triangle attached to three vertices, factor 6.

The general structure of the perturbation series is

$$\tilde{Z}/Z_0 = \sum_n \frac{g^n}{n!} \sum_k N^{\underline{k}} \sum_m W_{km}^{(n)} p^m. \tag{13}$$

The summation over k goes from 3 to $3n$. The power m is always $\leqslant k$ and $m=k$ corresponds to diagrams where there are one or several groups of triangles lying one on top of another. Clearly, these are the only diagrams that survive in the limit $N\rightarrow\infty$.

C. Counting diagrams

The quantity $W_{km}^{(n)}$ appearing in Eq. (13) is the number of paths leading to a given diagram topology. In a sense it is the number of diagrams of that topology. We are able to determine it recursively, step by step, but we are unable to give a general formula for it. It is relatively easy to follow the evolution of the number of triangle vertices during the recursive process, it is much more tedious to keep track of the number of triangle edges [13]. Thus, one can write a recursion equation for the sum

$$W_k^{(n)} = \sum_m W_{km}^{(n)}, \tag{14}$$

that is, for the total number of diagrams of order n, with k triangle vertices. This recursion relation, in essence, summarizes the rules listed in the preceding section:

TABLE I. Contributions to the partition function corresponding to the diagrams of Fig. 2; the common factor $g^3/3!$ is omitted.

(a)	$36N^3p^3$	(b)	$324N^4p^5$	(c)	$216N^4p^6$	(d)	$162N^5p^6$
(e)	$648N^5p^7$	(f)	$108N^5p^7$	(g)	$324N^5p^8$	(h)	$18N^6p^6$
(i)	$324N^6p^8$	(j)	$324N^6p^8$	(k)	$216N^6p^9$	(l)	$162N^7p^9$
(m)	$27N^7p^9$	(n)	$54N^7p^8$	(o)	$27N^8p^9$	(p)	$1N^9p^9$

$$W_k^{(n+1)}=k(k-1)(k-2)W_k^{(n)}+3(k-1)(k-2)W_{k-1}^{(n)}+3(k-2)W_{k-2}^{(n)}+W_{k-3}^{(n)}. \tag{15}$$

The coefficients result from elementary combinatorics. The initial condition is $W_k^{(1)}=\delta_{k3}$. The first two iterations are listed below. One can check that the numbers match those given in Figs. 1 and 2, provided the weights of diagrams with the same number of vertices are summed.

$n=2$.

$$W_3^{(2)}=6W_3^{(1)}=6,$$

$$W_4^{(2)}=18W_3^{(1)}=18,$$

$$W_5^{(2)}=9W_3^{(1)}=9,$$

$$W_6^{(2)}=W_3^{(1)}=1.$$

$n=3$.

$$W_3^{(3)}=6W_3^{(2)}=36,$$

$$W_4^{(3)}=24W_4^{(2)}+18W_3^{(2)}=540,$$

$$W_5^{(3)}=60W_5^{(2)}+36W_4^{(2)}+9W_3^{(2)}=1242,$$

$$W_6^{(3)}=120W_6^{(2)}+60W_5^{(2)}+12W_4^{(2)}+W_3^{(2)}=882,$$

$$W_7^{(3)}=90W_6^{(2)}+15W_5^{(2)}+W_4^{(2)}=243,$$

$$W_8^{(3)}=18W_6^{(2)}+W_5^{(2)}=27,$$

$$W_9^{(3)}=W_6^{(2)}=1.$$

One can easily see that the numbers given above agree with those presented in Table I. For example, the multiplicity of the diagrams (b) and (c) is, respectively, 324 and 216, which gives together $W_4^{(3)}=540$ as expected in the third order for the sum of diagrams occupying $k=4$ vertices.

We can also estimate the number of diagrams $W^{(n)}$ in the large order of the expansion, that is, for $n\to\infty$. As argued, this number is expected to grow faster than a factorial, reflecting the nonperturbative transition to Strauss's phase. As we will show this is indeed the case. Define the polynomial function

$$W^{(n)}(y)=\sum_k W_k^{(n)}y^k. \tag{16}$$

Multiplying both sides of Eq. (15) by y^{k-3} and summing over k one obtains the following differential equation:

$$W^{(n+1)}(y)=y^3(\partial_y+1)^3W^{(n)}(y). \tag{17}$$

The solution is

$$W^{(n)}(y)=[y^3(\partial_y+1)^3]^n\times 1, \tag{18}$$

which can also be rewritten as

$$W^{(n)}(y)=e^{-y}[y^3\partial_y^3]^n\times e^y. \tag{19}$$

Let us assume for a moment that $W^{(n)}(y)$ grows with n less rapidly than $(n!)^\kappa$, with some fixed κ, uniformly in y. It is then meaningful to introduce a generating function

$$W(x,y)=\sum_n W^{(n)}(y)x^n/(n!)^\kappa. \tag{20}$$

This function has a formal expansion

$$Z=e^{-y}\mathcal{W}[xy^3(\partial_y^3)^3]e^y, \tag{21}$$

where

$$\mathcal{W}(z)=\sum_n \frac{z^n}{(n!)^\kappa}. \tag{22}$$

Developing e^y in a power series we have

$$Z=e^{-y}\sum_k \frac{\mathcal{W}[xk(k-1)(k-2)]y^k}{k!}. \tag{23}$$

To check the convergence of this sum we need the asymptotic behavior of a function $\mathcal{W}(z)$. Examples of such functions for integer $\kappa=0,1,2$ are well known, being a simple exponential e^z, the Bessel function $I_0(2\sqrt{z})$, or the generalized hypergeometric function (see Ref. [14]) ${}_0F_2(z)$, respectively. For arbitrary κ one has

$$\mathcal{W}(z)\sim\frac{C_\kappa}{z^{(\kappa-1)/2\kappa}}e^{\kappa z^{1/\kappa}}(1+\cdots) \tag{24}$$

with some κ-dependent constant C_κ. It is obvious that if $\kappa<3$ the series (23) is meaningful only when $x=0$. For $\kappa\geqslant 3$ it becomes convergent for arbitrary y and x. We conclude that $W^{(n)}$ grows faster than $(n!)^{3-\epsilon}$, but slower than $(n!)^3$ for arbitrary small ϵ. Such an explosive behavior of the number of perturbation theory diagrams is a signal that nonperturbative phenomena are present. Indeed, it means that the coefficients of high powers of g and N^{-1} are increasing dra-

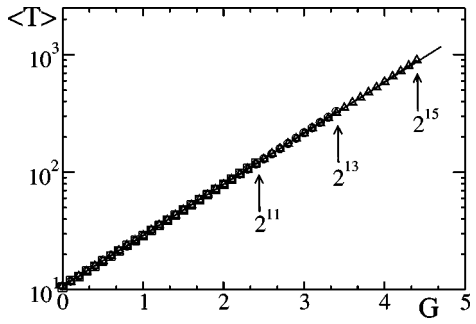

FIG. 3. The average number of triangles $\langle T\rangle$ vs the coupling constant $G\equiv 6g$. The average degree is set to $\alpha=4$ and the simulation is performed for the number of nodes $N=2^{11}$ (squares), 2^{13} (circles), and 2^{15} (triangles). The arrows indicate the position of the transition point $G=G_{out}$. The continuous line represents the analytic result $\langle T\rangle=(\alpha^3/6)\exp(G)$.

matically with the order of perturbation theory: what was assumed to be just a perturbation is in fact huge [15].

D. Summation of leading diagrams

We are interested in the limit $N\to\infty$ with $pN=\alpha=$ const. The structure of the perturbation expansion is given by Eq. (13). As already mentioned, in general, the number of triangle edges [denoted m in Eq. (13)] is larger than or equal to the number of triangle vertices [denoted k in Eq. (13)]. In the limit under consideration, only those diagrams contribute to the leading N-independent term for which the number of triangle edges is equal to the number of triangle vertices. One can easily see that in these diagrams the triangles can overlap, but otherwise do not touch. In the expansion in g up to the third order, the following diagrams belong to this class: a single triangle in the first order, diagrams in Figs. 1(a,d) in the second, and those in Figs. 3(a,h,p) in the third. Using the previously found results we have up to the third order,

$$Z(G,\gamma)=1+\frac{G}{1!}\gamma+\frac{G^2}{2!}(\gamma+\gamma^2)+\frac{G^3}{3!}(\gamma+3\gamma^2+\gamma^3)+\cdots, \tag{25}$$

where the convenient notation $G=6g,\gamma=\alpha^3/6$ has been introduced. In general, one can write this expansion as follows:

$$Z(G,\gamma)=1+\sum_{n=1}^{\infty}\frac{G^n}{n!}Z^{(n)}(\gamma)=1+\sum_{n=1}^{\infty}\frac{G^n}{n!}\sum_{k=1}^{n}z_k^{(n)}\gamma^k\cdots. \tag{26}$$

The coefficients $z_k^{(n)}$ can be interpreted as the number of all diagrams which consist of n triangles located at k isolated positions, with possible multioccupation of a position. Hence

$$z_k^{(n)}=\sum_P\frac{n!}{(n_1!)^{m_1}(n_2!)^{m_2}\cdots(n_k!)^{m_k}m_1!m_2!\cdots m_k!}, \tag{27}$$

where $n_1>n_2>\cdots n_k$ and the sum is over all the partitions P of n: $n_1m_1+n_2m_2+\cdots+n_km_k=n$. It turns out that the numbers $z_k^{(n)}$ can be calculated recursively:

$$z_k^{(n+1)}=kz_k^{(n)}+z_{k-1}^{(n)},\quad k=1,2,\ldots,n+1 \tag{28}$$

with the initial condition $z_1^{(1)}=1$ and $z_k^{(n)}=0$ for k outside of the closed interval $[1,n]$. The meaning of the equation is as follows. If one adds a new triangle to a configuration with n triangles, this triangle can be put at either of the k existing positions or at a new position. Hence, all configurations with $n+1$ triangles located at k positions can be obtained from configurations with n triangles at k positions, by placing a new triangle at one of the k old positions, or from configurations with n triangles at $(k-1)$ positions by placing a new triangle at a new position. The first few terms resulting from this recursion relation are

$$Z^{(2)}(\gamma)=1\gamma^1,$$

$$Z^{(2)}(\gamma)=1\gamma^1+1\gamma^2,$$

$$Z^{(3)}(\gamma)=1\gamma^1+3\gamma^2+1\gamma^3,$$

$$Z^{(4)}(\gamma)=1\gamma^1+7\gamma^2+6\gamma^3+\gamma^4,$$

$$Z^{(5)}(\gamma)=1\gamma^1+15\gamma^2+25\gamma^3+10\gamma^4+\gamma^5,$$

$$Z^{(6)}(\gamma)=1\gamma^1+31\gamma^2+90\gamma^3+65\gamma^4+15\gamma^5+\gamma^6.$$

The recursion relation can be converted into a partial differential equation for $Z(G,\gamma)$. Multiplying both sides of the equation by γ^{k-1} and summing over k one finds

$$\frac{1}{\gamma}Z^{(n+1)}=\frac{\partial}{\partial\gamma}Z^{(n)}+Z^{(n)}, \tag{29}$$

where

$$Z^{(n)}(\gamma)=\sum_{k=1}^{n}z_k^{(n)}\gamma^k. \tag{30}$$

Now, multiplying both sides by $G^n/n!$ and summing over n one obtains

$$\frac{\partial}{\partial G}Z=\gamma\frac{\partial}{\partial\gamma}Z+\gamma Z, \tag{31}$$

where Z is given by Eq. (26). An even simpler equation is satisfied by $F=\ln Z$:

$$\frac{\partial}{\partial G}F=\gamma\frac{\partial}{\partial\gamma}F+\gamma. \tag{32}$$

One easily checks that the general solution is

$$F(\gamma,G)=f(\gamma e^G)-\gamma, \tag{33}$$

where f is an arbitrary differentiable function. It results, however, from Eq. (26) that $F(\gamma,0)=0$. Hence

$$F(\gamma,G)=\ln Z(\gamma,G)=\gamma(e^G-1). \tag{34}$$

This result is not surprising for a practitioner of quantum field theory. Indeed, $\ln Z(\gamma,G)$ should equal the sum of contributions of connected diagrams. The only connected diagrams, in the large N limit, are those where triangles are all put on top of each other and according to our rules the diagram of nth order yields just $(g^n/n!)p^3 6^{n-1}N^{\underline{3}}\sim \gamma G^n/n!$.

Notice, that the same result (34) is obtained assuming that in the Erdös-Rényi model the number of triangles has a Poisson distribution with average γ. Indeed, the average of $\exp(GT)$ is

$$\sum_{T=0}^{\infty}\frac{\gamma^T}{T!}e^{-\gamma}e^{GT}=\exp[\,\gamma(e^G-1)\,]. \tag{35}$$

The average number of triangles is

$$\langle T\rangle=\frac{\partial}{\partial G}\ln Z=\gamma e^G. \tag{36}$$

It is important to note that in the $N\to\infty$ limit the average degree of a graph node becomes independent of G and is just equal to α, like in pure Erdös-Rényi theory. This can be easily seen in our formalism. Add to the action a source term $\eta\mathrm{Tr}(PM)$, where P is the matrix $P_{ij}=\delta_{i1}$, so that $\mathrm{Tr}(PM)$ is the degree of the graph node with label 1. In our diagrammatics $\eta\mathrm{Tr}(PM)$ produces a line instead of a triangle. But only one end of this line has a running index, the other end has index 1. The diagram of the lowest order in η is just this line and gives the contribution $\eta pN=\eta\alpha$. All corrections due to the interaction $g\mathrm{Tr}(M^3)$ yield terms proportional to some inverse powers of N, because there is no way to put a triangle on a line. For example, the diagram of order ηg, where one has one triangle and one line on top of one of its edges gives $3\eta g p^3 N^{\underline{2}}\sim 3\eta g\alpha^3/N$ (we have $N^{\underline{2}}$ and not $N^{\underline{3}}$ because one of the triangle edges has the fixed label 1). In conclusion, the only connected diagram of order $O(\eta)$ is independent of g, as is, to this order the free energy $\ln Z$. The average degree is just the derivative, at $\eta=0$, of the free energy and equals α. One can extend this argument to higher order moments of the degree distribution.

Using the results of this section we can propose a rough estimate of the expected region where nonperturbative physics sets in. With $p=\alpha/N$ and N large the summand in Eq. (3) can be rewritten as

$$\exp\left\{\ln\frac{N}{\alpha}(-L+G_0T)\right\}. \tag{37}$$

We have rescaled the coupling by $\ln(N/\alpha)$, so that this large factor multiplies now both terms in the action. We expect that the perturbation series breaks down when the fluctuations of the two terms in the action become comparable. The number of links is $\sim N$ and we expect $\langle(\delta L)^2\rangle\sim N$. In the large N limit the fluctuation of T is given by the second derivative of the free energy and equals $\langle(\delta T)^2\rangle=\gamma\exp(G)=\gamma N^{G_0}$. The two fluctuations become comparable when $G_0\approx 1$. The numerical calculations confirm the logarithmic scaling of G, but as we will see in a moment the critical G_0 seems to lie below 1.

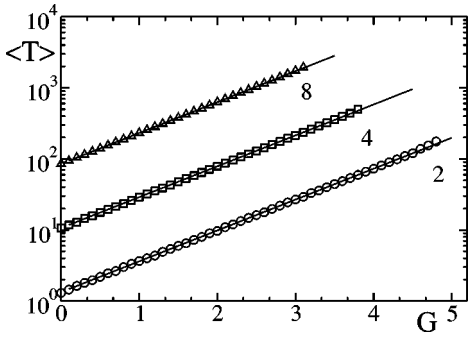

FIG. 4. As in Fig. 3, $\langle T\rangle$ vs G, at $N=2^{14}$, but for three values of the average degree $\alpha=2$ (circles), 4 (squares), and 8 (triangles).

III. NUMERICAL SIMULATIONS

We have shown in the preceding section that at finite N the perturbative series has a behavior which signals the presence of a nonperturbative phenomenon. The similarity of our problem with the example exhibited in the first subsection suggests an educated guess: there is a barrier separating the perturbative phase from a pathological but stable configuration; the nonperturbative phenomenon in question is the rolling of the system over the barrier towards this stable configuration. The barrier must become unpenetrable in the $N\to\infty$ limit, because in this limit the perturbation series becomes well behaved, actually summable, whatever is the coupling. We will confirm this guess with the help of numerical simulations.

In constructing an algorithm manipulating adjacency matrices it is most important to take into account the sparse nature of these matrices. Only the positions of $2L$ matrix elements carry a relevant information. This makes it possible to reduce the amount of computer memory, needed to store an adjacency matrix, from $O(N^2)$ in the naive coding to $O(N)$ in the linear coding. In effect, we simulate systems with the number of nodes of order 10^4, i.e., three orders of magnitude larger than those simulated by Strauss [9]. In the present work we use the algorithm introduced in Refs. [17,18] by straightforwardly upgrading it so as to include the term in the action proportional to the number of triangles.

In the first numerical experiment we set $\alpha=2L/N=4$ and we measure the average number of triangles $\langle T\rangle$ for $N=2^{11}$, 2^{13}, and 2^{15}. The coupling G is changed in small steps until the system makes a transition to Strauss's phase. The results are shown in Fig. 3. The continuous line corresponds to Eq. (36). It is remarkable that the points follow this line. The error bars are smaller than the symbol size. The transition points are indicated by an arrow. A closer examination of the data shows that near the transition the points start to deviate from the line and lie systematically above it.

In the next experiment we set $N=2^{14}$ and measure $\langle T\rangle$ for $\alpha=2$, 4, and 8. The coupling G again varies up to the transition point. The result is shown in Fig. 4. The lines correspond to $\langle T\rangle=(\alpha/6)\exp(G)$. The agreement is remarkable.

We have also measured the local clustering measure C_j as defined in Ref. [16]:

$$C_j = \frac{2T_j}{L_j(L_j-1)}, \tag{38}$$

where T_j is the number of triangles touching the vertex j, and L_j is the number of links emerging from it. We set $C_j=0$ when L_j is zero or 1. A global clustering coefficient C is obtained by averaging over vertices. In Fig. 5 we plot $N\langle C\rangle$ versus G. It is seen that $N\langle C\rangle = \sigma(\alpha)\exp(G)$, with $\sigma(\alpha) = \alpha(1-(1+\alpha)\exp(-\alpha))$, which is the value of $N\langle C\rangle$ in Erdös-Rényi model. It is interesting that for very different values of α the transition occurs at roughly the same value of the clustering coefficient. This is presumably not a numerical accident, but we have no explanation to offer.

In Fig. 6 we show the variation of the transition point $G = G_{out}$ with the system size N. In our experiments the coupling G was always changed by $\delta G = 0.1$. Thus, in the figure we have associated an error 0.1 with the data points. Here a comment is in order: After having changed G we always made 1000 thermalization sweeps, then we carried out 20 000 sweeps, performing measures every ten sweeps. It is important to remember that the number of sweeps was always the same. Indeed, at *finite N* the system will sooner or later roll over the barrier, it is sufficient to wait long enough. The transition point $G = G_{out}$ is well defined when one decides to fix the waiting time. Actually, we are more interested by the scaling of G_{out} with N than by its exact value.

The curves in Fig. 6 are

$$G_{out} = 0.75 \ln N - 2.4 \quad \text{for} \quad \alpha = 2, \tag{39}$$

$$G_{out} = 0.70 \ln N - 2.9 \quad \text{for} \quad \alpha = 4, \tag{40}$$

$$G_{out} = 0.60 \ln N - 2.7 \quad \text{for} \quad \alpha = 8. \tag{41}$$

It is very interesting, although not really surprising (see Sec. II D), that G_{out} scales like $\ln N$. This means that setting $G = G_0 \ln N$ one obtains a model with the clustering coefficient scaling nontrivially, $C \sim N^{G_0 - 1}$.

Figure 7 illustrates the fact that the degree distribution is in the smooth phase insensitive to the value of the coupling G. We show the distribution at $N=2048$ and $\alpha=2$ for $G=0$ and for a large value of G, i.e., $G=3.0$, close to G_{out}. The distributions are almost identical and correspond to the Poisson distribution with average equal to $\alpha=2$ (the line). This has been repeated for $N=2048$ and $\alpha=4$, where we measured at $G=0$ and $G=2.3$.

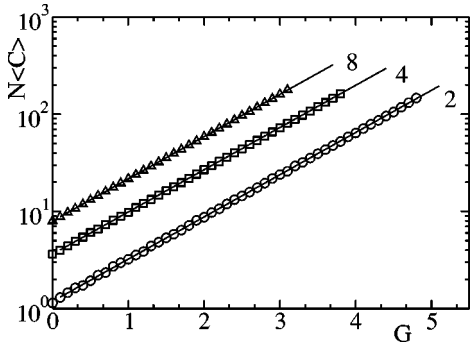

FIG. 5. The scaled clustering coefficient $N\langle C\rangle$ vs G at $N=2^{14}$ and for $\alpha=2$ (circles), 4 (squares), and 8 (triangles). The line represents the expected behavior $N\langle C\rangle \propto \exp(G)$, where the proportionality coefficient is chosen so as to get the value expected in Erdös-Rényi model at $G=0$.

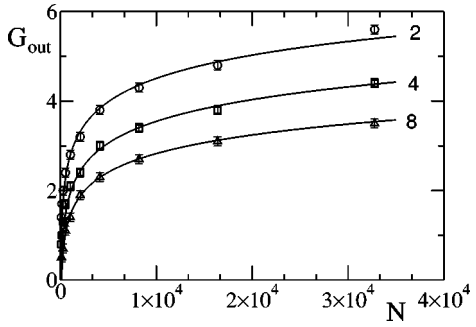

FIG. 6. The transition from the perturbative to Strauss's phase occurs at $G=G_{out}$. The figure shows how G_{out} depends on the system size N, for $\alpha=2$ (circles), 4 (squares), and 8 (triangles). Notice the logarithmic growth of the curves.

IV. DISCUSSION, SUMMARY, AND CONCLUSION

A. Possible generalizations

Up to now we assumed that the interaction has a simple form $S(M) = g\mathrm{Tr}(M^3)$. The question which immediately comes to mind is what happens to the network transitivity when the action is more complicated? Assume that $S(M)$ has the polynomial form, $S(M) = \Sigma_{n\geq 3} g_n \mathrm{Tr}(M^n)$, with $g_3 \equiv g$. Our diagrammatic rules can easily be extended to include this case. $\mathrm{Tr}(M^n)$ is represented by a polygon with n sides. However, the polygon can be folded, the same line segment being covered several times. In particular, when $n>4$, a polygon can be folded so that some of its edges form a triangle.

The calculation of leading diagrams given in Sec. II D can be generalized to interactions involving odd powers of M.

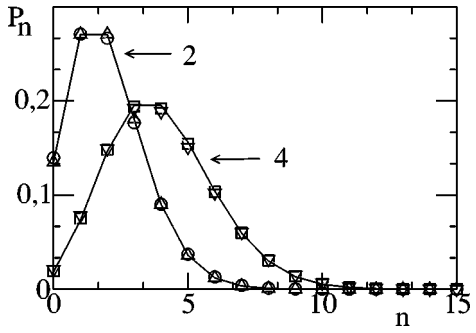

FIG. 7. We compare the degree distribution calculated at two values of the coupling constant G. The number of nodes is set to $N=2048$. At $\alpha=2$ the calculation is performed setting $G=0$ (triangles up) and 3.0 (circles). At $\alpha=4$ the calculation corresponds to $G=0$ (squares) and $G=2.3$ (triangles down). The continuous lines represent Poisson distributions with averages equal to 2 and 4, respectively.

One can limit oneself to connected diagrams, those contributing to the free energy. In the $N\to\infty$ limit the contribution of a diagram is proportional to N to a power equal to (number of vertices)$-$(number of edges) (because $p=\alpha/N$). We are interested in diagrams which overlap with a triangle. In this case (number of vertices)$-$(number of edges)$=0$. In all other cases this quantity is negative and the diagram does not contribute in the limit. It is easy to see that diagrams surviving in the $N\to\infty$ limit are those where triangles and folded polygons are put on top of each other.

The situation is more complicated for even powers of M. The leading diagrams look like branched polymers, with (number of vertices)$=$(number of edges)$+1$, and their contribution diverges like N. In order to avoid an unwanted renormalization of the quadratic term in the action one has to subtract from $\mathrm{Tr}(M^{2k})$ a counterterm $\sim\mathrm{Tr}(M^2)$ with an appropriate coefficient in front. Then the calculation is like for the odd power case.

We have not pushed this calculation very far. As far as we can see one expects a certain degree of universality: the higher powers of M should not change the qualitative picture very much, although they may be important for phenomenology, to fit the data. A comprehensive study of these more general interactions is certainly worth being done. This is, however, beyond the scope of the present paper.

It is not quite clear what is the best way of extending the theory of this paper so as to obtain an arbitrary degree distribution. The field theoretical methods extensively used in this paper usually fail when the action becomes nonanalytic. The simplest, although perhaps not the most elegant, extension consists in using instead of the Erdös-Rényi model, a general model with uncorrelated nodes [18] as the zeroth order approximation. Preliminary numerical results look encouraging, although it is clear that much has to be done in order to get a fully satisfactory phenomenology. We hope to return to this problem elsewhere.

B. Summary

Let us now summarize what has been achieved in this work. In most of this paper we have discussed a model of random graphs where the classical Erdös-Rényi theory is generalized by the introduction into the action of an interaction term $(G/3!)\mathrm{Tr}(M^3)=GT$, M being the $N\times N$ adjacency matrix and T the number of triangles, respectively. This model is our test model. It is a matrix model, but of a special kind, because the dynamical variable is a random matrix whose elements equal either 0 or 1.

Inspired by the analogy with more conventional matrix models we develop a diagrammatic technique, enabling us to calculate the perturbation series analytically. We count the diagrams and show that, at finite N, their number grows so rapidly that the perturbation series becomes pathological. This also happens in conventional matrix models and, as is well known, indicates the presence of a nonperturbative phenomenon. The nature of this phenomenon is identified through numerical simulations. There is a ''potential barrier'' and the system can roll over it and fall into a pathological phase, where all triangles form a unique clan. This phase was first discovered long ago by Strauss [9], who did not notice, however, that it is separated from a smooth phase by a barrier which becomes impenetrable at large N. We propose to consider this smooth phase as the physical one.

We show that for large enough N and in a range of values of the coupling constant G the smooth phase can be considered, for all practical purposes, as stable. In this range of G it is meaningful to neglect the nonperturbative physics and to limit oneself to leading diagrams (those obtained setting $N=\infty$). We are able to sum all these diagrams up, obtaining simple analytic expressions for the free energy and for the average number of triangles. We also show analytically that in this regime the degree distribution is insensitive to the value of the coupling G.

A heuristic argument, confirmed by numerical simulations, indicates that the transition point $G=G_{out}$, where the system jumps to the Strauss's phase, scales with N like $\ln N$. Hence, the physical coupling is not so much G but rather G_0 defined by the equation $G=G_0\ln N$. Our simulations indicate that at the transition point $G_0=0.6$ to 0.75, depending on the average degree, but this result should not be regarded as definitive. Anyhow, the clustering coefficient scales nontrivially, like $C\sim N^{G_0-1}$ and is larger by one to two orders of magnitude than in the unperturbed model.

It appears that the analytic treatment can be extended to more complicated, but polynomial actions. In the present state of affairs the extension of our approach to more realistic, for example, scale-free models can only be done numerically.

C. Conclusion

Clustering is a rather striking trait of many observed networks. The local treelike structure characterizing most static models is clearly nonrealistic. We have argued elsewhere that static models are an important ingredient of network theory. Thus, we believe that it is important to be able to construct static models with nontrivial clustering. There was some confusion concerning the feasibility of such an enterprise. We hope to have dissipated it. For the sake of clarity we have focused our attention on a model where much can be done analytically. It is a specific matrix model, where matrices are random, but their elements take values 0 and 1 only. In the zeroth order approximation it is equivalent to the classical Erdös-Rényi model of graphs. Nontrivial clustering is generated by an appropriate interaction. A comprehensive phenomenologically oriented study is beyond the scope of this paper and remains to be carried out. After this manuscript was completed we learned about Ref. [19], which contains some overlapping material.

ACKNOWLEDGMENTS

This work was partially supported by the EC IHP Grant No. HPRN-CT-1999-000161, by Polish State Committee for Scientific Research (KBN) Grant Nos. 2P03B 08225 (2003-2006) and 2P03B 09622 (2002-2004), and by EU IST Center of Excellence ''COPIRA.'' Laboratoire de Physique Théorique is Unité Mixte du CNRS UMR 8627.

[1] B. Bollobás, *Random Graphs*, 2nd ed. (Academic Press, New York, 2001).
[2] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[3] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002); *Evolution of Networks: From Biological Nets to to the Internet and WWW* (Oxford Univerity Press, New York, 2003).
[4] M.E.J. Newman, SIAM Rev. **45**, 167 (2003).
[5] S.N. Dorogovtsev, e-print cond-mat/0308444.
[6] M. Boguñ a and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003); S.N. Dorogovtsev, e-print cond-mat/0308336.
[7] M.E.J. Newman, Phys. Rev. E **68**, 026121 (2003).
[8] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).
[9] D. Strauss, SIAM Rev. **28**, 513 (1986), and references therein; in particular, see P.W. Holland and S. Leinhardt, J. Am. Stat. Assoc. **76**, 33 (1981).
[10] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[11] G. 't Hooft, Nucl. Phys. **72**, 461 (1974).
[12] M. Bauer and O. Golinelli, J. Stat. Phys. **103**, 301 (2001).
[13] We have also a recursive Monte Carlo algorithm calculating $W_{km}^{(n)}$. The algorithm produces a string of indices, starting with 123 and adding successive triples. Suppose that at the jth step, among $3j$ indices there are i *distinct* ones. The algorithm goes over to the next step with probability $P(i \to i+1) = \binom{i+3}{3}/\binom{j+3}{3}$. In the $(j+1)$st step three indices are drawn at random from the i distinct indices supplemented with three *indistinguishable* ones. These randomly selected indices are accepted as the new triple with probability 1/6,1/6,1/3, and 1 when they contain 3, 2, 1, and 0 indistinguishable indices, respectively. After $n-1$ successful steps one has constructed a string representing a particular contribution to $\mathrm{Tr}(M^n)$. One determines k and m by counting distinct indices and links. The acceptance of this algorithm drops rapidly with n, one can hardly go beyond $n=10$.
[14] http://functions.wolfram.com/ HypergeometricFunctions/ HypergeometricPFQRegularized/06/02/04/
[15] In field theory such a behavior of the number of diagrams implies that the radius of convergence of the perturbation series is zero. This is not the case in our model. The terms of the series are the largest when $p=1$. But for this value of p all links of the graph are, by construction, occupied and the sum of all diagrams of order n is just $[N^3]^n$ (cf. Ref. [12]). Consequently the sum of the series (13) is finite and less than $\exp(gN^3)$. The series is pathological, but it is regulated as a result of a specific excluded volume effect. Indeed $N^{\underline{k}}=0$ for $k>N$, so that for fixed N the proliferation of diagrams does not continue indefinitely.
[16] S.H. Strogatz and D.J. Watts, Nature (London) **393**, 440 (1998).
[17] Z. Burda, J.D. Correia, and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
[18] Z. Burda and A. Krzywicki, Phys. Rev. E **67**, 046118 (2003). The distinctive trait of this paper, compared to other publications on the subject, is that it deals with simple graphs, without self- and multiple connections between nodes and contains a discussion of important finite N effects, absent in pseudographs.
[19] S. Coulomb and M. Bauer, e-print cond-mat/0310170.