Circuits in Random Graphs

Enzo Marinari

(Roma *La Sapienza*, Italy)

Work in collaboration with Remi Monasson (`cond-mat 0409238`, in print on JSTAT).

$N = \infty$ (analytic) and finite N (from exact enumeration) results about number of circuits (and loops) on random graphs.

ICTP, Trieste, September 2004

Summary

- Random graphs, circuits and cavity. The meaning of an approximation.
- $N \to \infty$: mean field approach.
- Finite N: results from exact enumeration. With an algorithm by D. B. Johnson, SIAM J. Comput. **4**, 77 (1975).
- Recent interest on circuits and loops.
- Conclusions.

Random graphs: large interest in probability theory and in statistical mechanics.

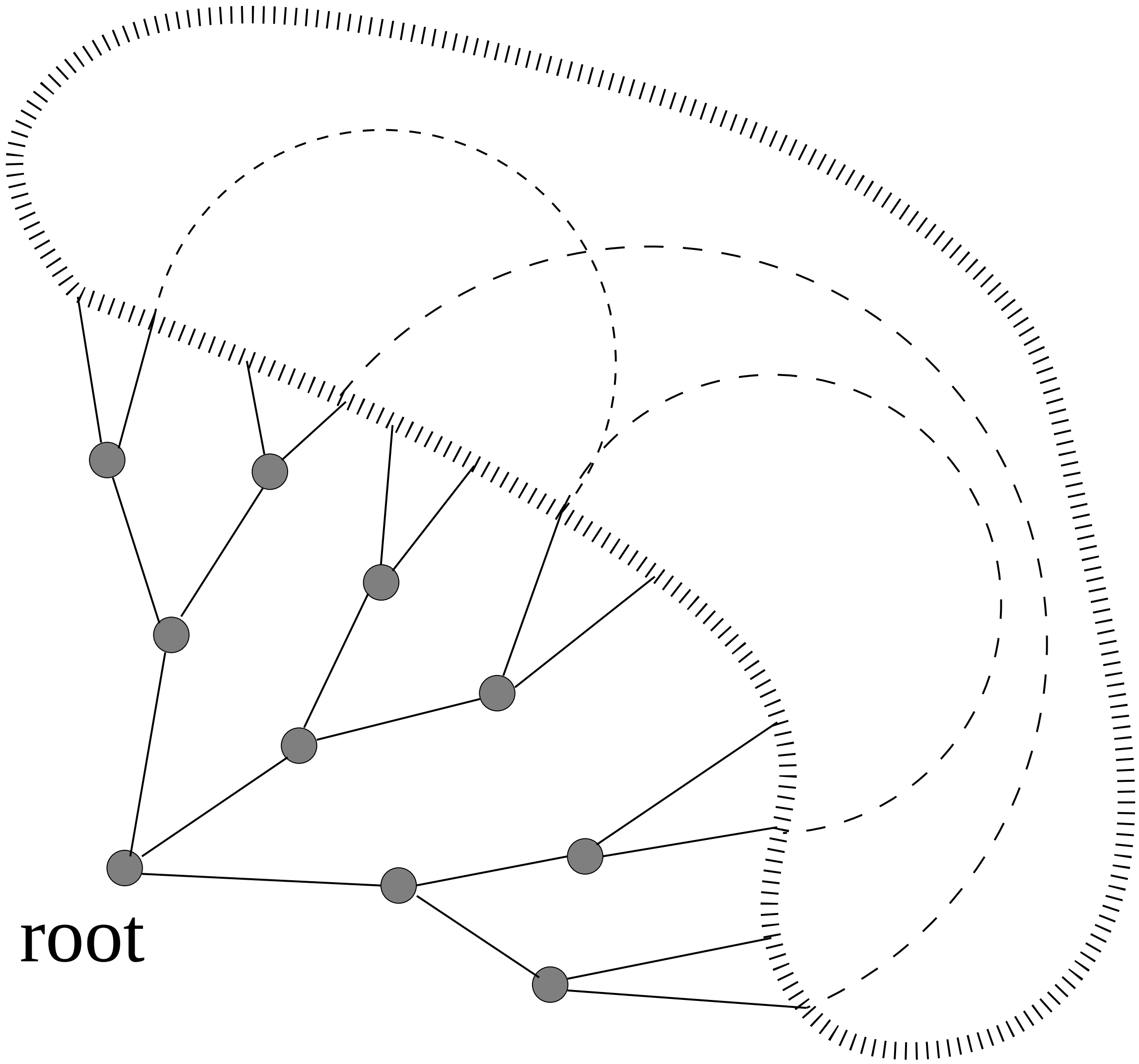
Many different (interesting...) types of random graphs exist:

- Erdös-Renyi. Here edges are chosen independently between pairs of a set of N vertices with fixed probability $O(\frac{1}{N})$
- Random c-regular graphs. Uniformly drawn from the set of all graphs with N vertices, each obliged to have degree c. Easy to generate for large N.
- Scale free graphs (modern...) with power law connectivity degree distribution.

Here we will deal with random c-regular graphs.

We will try to reach some quantitative understanding of the behavior of closed circuits in random graphs.

Around a vertex selected at random (we call it root) the graphs looks like a regular tree.



The probability that a circuit (self-avoiding closed path) of length L passing through the root exists goes to zero when $N \to \infty$ and L is finite: it departs from zero when $log(N) = O(L)$ (there are finite length loops but not in an extensive amount).

There are two main points that motivate us.

First: reach a better understanding of equilibrium properties of models of interacting fields living on these graphs.

Frustration emerges from circuits of odd length: the number of such circuits is related to the amount of glassiness present in the system at low T.

Dynamical point of view: interacting agents, routers, internet...

Second: the fact that random graphs are locally tree like (i.e. that very are very few short circuits) is crucial to the analytic treatment of spin models on random graphs for example by cavity methods.

They are solved by assuming that f is equal to the one on a tree with self-consistent boundary conditions.

We stress that this procedure is equivalent to make assumptions on the distribution of long loops.

$N = \infty$ mean field approach: counting loops.

(Different from counting circuits...)

We define loops as closed paths going through each vertex an even number of times and through each edge at most once. (i.e. the ones that appear in the high T expansion of statistical models).

The model: Munish the graph shown before with Ising spins $S_i = \pm 1$ on vertices $i = 1, \ldots, N$ and ferromagnetic couplings $J_{ij} = 1$ on edges (i, j).

Compute f: consider the leaves of the uncovered local tree *i.e.* vertices at distance D from the root. At large β, a spontaneous magnetization m is expected to be present in the bulk.

The spins attached to the leaves will feel an external field $H > 0$. Integrating these spins out will in turn produce an external field H' acting on spins at distance $D - 1$ from the root, with

$$H' = \frac{(c-1)}{\beta} \tanh^{-1}[\tanh(\beta) \tanh(\beta H)]$$

(Bowman and Levin, PRB 1982).

After repeated iterations of this procedure the field at the root reaches a stationary value, H^*, with $m = \tanh[\beta c H^*]$ and the free-energy density

$$
\begin{aligned}
f(\beta) &= -\frac{c}{2\beta} \ln 2\left(e^{-\beta} + e^{\beta} \cosh 2\beta H^*\right) \\
&+ \frac{c-1}{\beta} \ln 2 \cosh\left(\frac{\beta c H^*}{c-1}\right) .
\end{aligned}
$$

The critical inverse temperature is the smallest value of β for which H^* is non zero *i.e.* $\beta_c = \tanh^{-1}(\frac{1}{c-1})$.

Use high T loop expansion functional form: the high temperature expansion of the partition function Z of the Ising model can be written as a sum over loops. Each loop is given a weight $(\tanh\beta)^L$ depending upon its length L:

$$
Z(\beta) = 2^N (\cosh\beta)^{\frac{cN}{2}} \sum_L M(L) \, (\tanh\beta)^L \ , \qquad (1)
$$

where $M(L)$ is the number of loops of length L that can be drawn on the graph.

Use Legendre transform: assume that loop multiplicity grows exponentially with the graph size:

$$M(L = \ell\, N) = \exp[N\,\sigma(\ell) + o(N)]\ ,$$

where ℓ is the intensive length of the loops, and σ is the entropy of loops having length ℓ. The entropy σ gives information about large-scale loops *i.e.* with lengths of the order of N.

Inserting the scaling hypothesis for $M(L)$ in the partition function Z gives

$$\begin{aligned} -\beta f(\beta) &= -\ln 2 - \frac{c}{2}\ln\cosh\beta \\ &+ \max_{\ell}\left[\sigma(\ell) + \ell\ \ln\tanh\beta\right] \end{aligned}$$

in the infinite N limit.

Results for the loop entropy: see next two figures.

Entropy departs from $\ell = \sigma = 0$ with a slope $-\ln(\tanh \beta_c) = \ln(c-1)$. Max is in $\ell_M = \frac{c}{4}$, $\sigma_M = (\frac{c}{2} - 1)\ln 2$, for loops going through half of the edges.

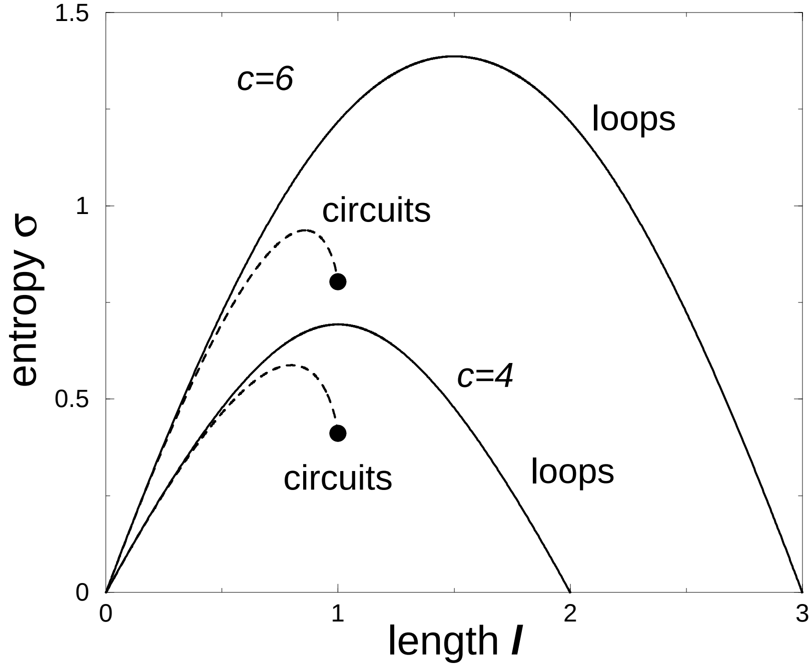
The left part of the curve ($\ell \leq \ell_M$) is parametrized by β going from β_c ($\ell = 0$: there is a para/ferro transition when extensive loops start contributing to Z) to ∞ (top of the curve: at $T = 0$, Z is dominated by the most numerous paths).

The right part of the curve ($\ell \geq \ell_M$) is for $\tanh \beta > 1$, that is, for inverse temperatures with an imaginary part equal to $\frac{\pi}{2}$.

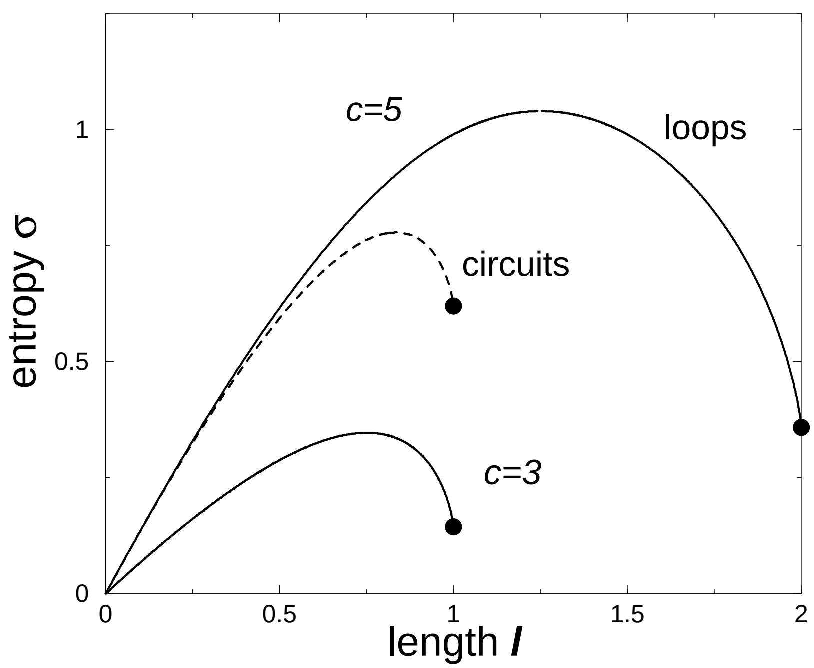
Even c. σ is unchanged under the transformation $\ell \to \frac{c}{2} - \ell$. The right part of the curve is the mirror symmetric of the left part, from a duality between long and short extensive loops. The largest loop has length $\frac{c}{2}$.

Odd c. Duality does not hold. The maximal length ℓ_+ is reached with an infinite slope (finite entropy σ_+). For odd degrees c loops cannot occupy all edges: the longest loops have one free edge per vertex, acting as defects, the positions of which can be chosen with some freedom, giving rise to a finite entropy. The frustration coming from the parity of c is less important as c increases. G. Parisi, private communication: RS solution is correct in this range of (complex-valued) temperature.

even c:


1.5
1
0.5
0
entropy σ
0
1
2
3
length *l*
c=6
loops
circuits
c=4
circuits
loops

odd c:


1
0.5
0
entropy σ
0
0.5
1
1.5
2
length *l*
c=5
loops
circuits
c=3

$N = \infty$ mean field approach: counting circuits.

The same procedure can be applied to derive the entropies of other spin models. The ferromagnetic $O(n)$ model with $n \to 0$ gives information on circuits.

A spin $\vec{S}$ is submitted to two fields H_1, H_2 conjugated to the magnetization and its squared value. One gets

$$f = -\frac{c-2}{2\beta} \ln \left[\frac{c(c-1)\beta - 2}{c-2} \right] + \frac{c}{2\beta} \ln \left[(c-1)\beta \right]$$

(for an alternative derivation see Mézard, Montanari and Müller, PRL 2004). The entropy of circuits is

$$\begin{aligned} \sigma(\ell) &= -(1-\ell)\ln(1-\ell) \\ &+ \left(\frac{c}{2} - \ell\right) \ln\left(1 - \frac{2\ell}{c}\right) + \ell \ln(c-1) \,. \end{aligned}$$

Results are in the former figure. The rightmost point corresponds to Hamiltonian cycles.

Our result coincides with the output of rigorous calculations (Garmo 1999; Janson, Luczak and Rucinski 2000); the replica symmetric hypothesis is exact for the $O(n \to 0)$ model (see again Mézard et al. 2004).

From now on we only refer to circuits (and not to loops).

For finite L, $M(L)$ is asymptotically Poisson-distributed when $N \to \infty$,

$$\mathrm{P}[M(L) = M] = \frac{1}{M!}\left[\frac{(c-1)^L}{2L}\right]^M e^{-\frac{(c-1)^L}{2L}}$$

that holds for circuit-length $L \ll \log N$. The expected number of circuits of intensive length $\ell = L/N$ is for $\ell < \frac{\log N}{N}$,

$$\langle M(\ell) \rangle = \frac{(c-1)^L}{2L} = e^{N\sigma(\ell) - \log(N) + \tilde{\sigma}(\ell)} ,$$

with $\sigma(\ell) = \ell \ \log(c-1)$ and $\tilde{\sigma}(\ell) = \log(2\ell)$.

Finite size corrections: exact enumeration

Finite N: we have implemented a fast algorithm for finding all circuits in a given graph. We *find*, not only count, all the circuits: our method we can in principle give all interesting characterizations.

We first generate a random graph and then count the circuits: we average over a number of samples.

To generate a fixed connectivity random graph we start by assuming that each site has c connections that connect it to c different sites. Self-connections and double edges are not allowed.

We start with all connections free: pairs of connections are extracted and matched together. We continue filling them up (we use a table which is resized after each step to keep the process effective) till all connections are set, or till we are stuck (if for example there are only two free connections belonging to the same site): in this case we discard the full graph and restart the procedure from scratch.

The algorithm for circuit enumeration

To enumerate circuits we have implemented an algorithm by Johnson SIAM J. Comput. **4**, 77 (1975) that finds all elementary circuits of a graph.

Computer time is bounded by $O((N+E)(M+1))$, where N is the number of vertices of the graph, E is the number of edges, and M is the total number of circuits in the graph. The time used between the output of two consecutive circuits is bound by $O(N+E)$ (this is true also for the time elapsed before the output of the first circuit and after the output of the last one). The memory space is bounded by $O(N+E)$.

One first orders the vertices in some lexicographic sequence, and labels them with integers. The search starts from a *root* vertex r, in the subgraph induced by r and by vertices after r. The input to the procedure is the adjacency list $A(v)$ for each vertex v: A contains u if and only if $(v,u) \in \mathcal{E}$, where $\mathcal{E}$ is the set of edges of the graph.

We *block* a vertex v when it is added to a path beginning in r.

We build elementary paths starting from r.

The vertices of the current trial paths are loaded onto a stack.

A procedure adds the vertex to the path, if appropriate, and appends the vertex to the stack: the vertex is deleted from the stack when exiting from this procedure.

Ingenious part: keep a vertex *blocked* as long as possible. This has to be done while maintaining the procedure correctness: the basic rule that has to be satisfied to guarantee that all circuits are found (only once) is that if a path exists from the vertex v to r that does not intersect the path loaded on the stack, then v has to be free (i.e. it cannot be in a *blocked* state).

On an Intel Xeon 2.8 GHz processor our implementation takes of the order of 0.07 seconds for finding all circuits of a $N = 30$ graph (they are $O(50000)$), $2.4s$ for $N = 40$ ($O(1.5\ 10^6)$ circuits) and $80s$ for $N = 50$ ($O(4\ 10^7)$ circuits).

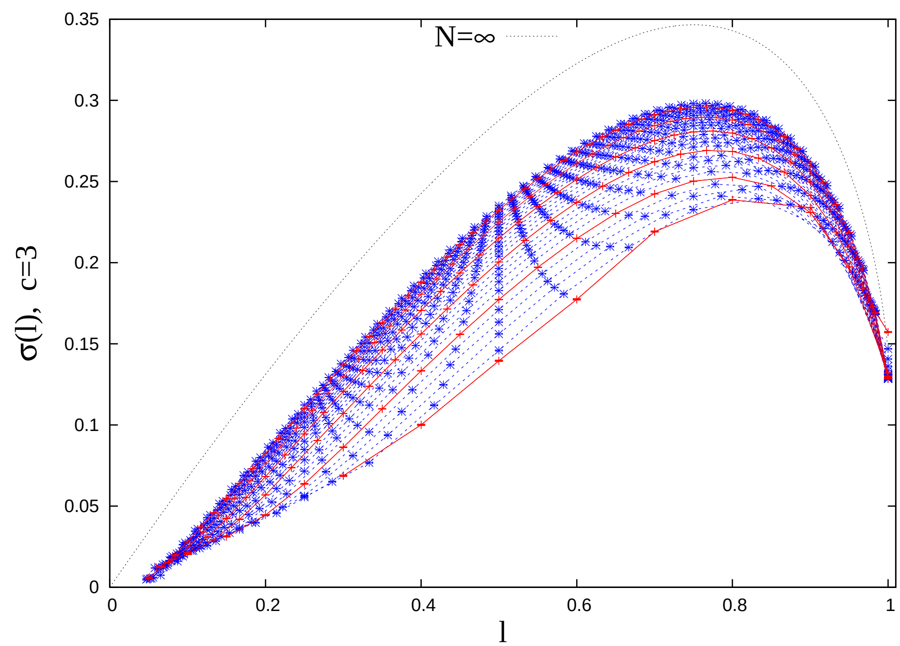
Thanks to our algorithm and implementation we have been able to enumerate of the order of 10^{14} circuits (a large number).

For small c values we can study larger graphs (we have analyzed graphs with up to 64 vertices in the $c = 3$ case and up to 22 vertices for $c = 6$, and averaged our results over samples ranging from 1000 to 10000 random graphs).

Typically we find of the order of 300 million circuits for a $N = 56$, $c = 3$ graph, one billion circuits on a $N = 26$, $c = 5$ graph and 1.5 billion circuits on a $N = 22$, $c = 6$ graph.

For each value of N, we average over of the order of 10000 samples for all the $c = 3$ enumerations, and 1000 graphs for $c > 3$.

$\sigma_N(\ell)$ as a function of ℓ for $c = 3$, and for graph sizes ranging from $N = 10$ to 64 (from bottom to top). The full curve is for the analytical calculation. Data for sizes multiple of 10 use a different drawing style.



Finite N data approach very slowly the $N = \infty$ limit.

Verify that finite N and $N \longrightarrow \infty$ results are compatible.

$\log\langle M(\ell)\rangle \sim \ell \log(c-1)$ as $\ell \to 0$ where $\langle M(\ell)\rangle$ is the average number of circuits of length $L = \ell \cdot N$.

Numerical investigations do not allow us to be very close to $\ell = 0$ since the minimal intensive circuit length, ℓ_{min} is of the order of $1/N$.

Flattening of $\log\langle M(\ell)\rangle$ when ℓ approaches the minimum allowed value: finite size effects.

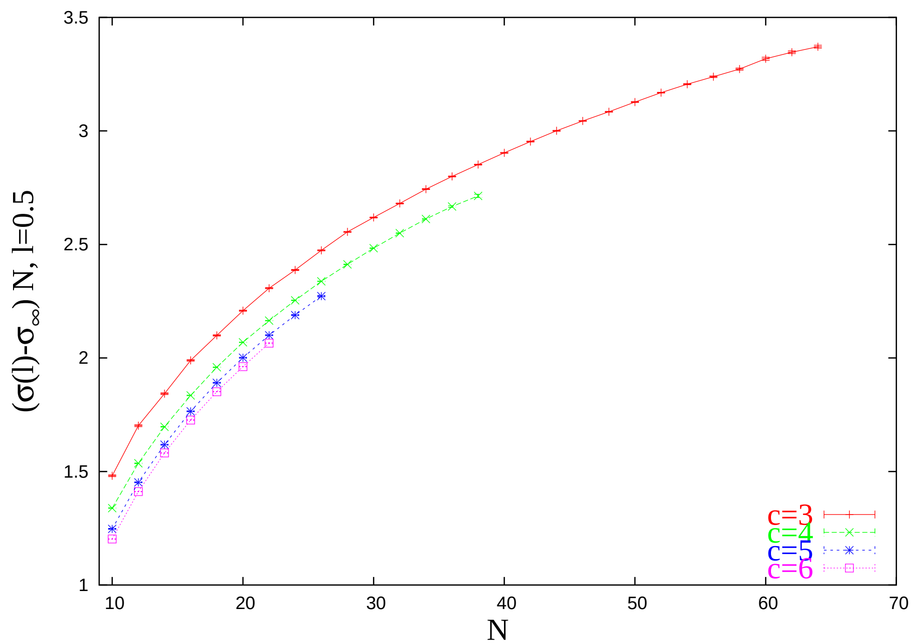
In the small ℓ region, relatively safe from finite size effects, the slope is very similar to the asymptotic slope. We have fitted a linear behavior (that is clear in the data) for example for ℓ in the range $(.13, .19)$ for $c = 3$.

Using this approach we find for c from 3 to 6 slopes about 20% smaller than the theoretical prediction (on the larger graphs we can study).

For $c = 3$ we find 0.54 versus a theoretical $\log 2 \sim 0.69$; for $c = 4$ we find 0.87 versus 1.10; for $c = 5$ we find 1.12 versus 1.39; for $c = 6$ we find 1.31 versus 1.61.

Finite size effect can be drastically reduced if we compare directly different c values. The ratio of the slopes corresponding to c and $c+1$ is 0.62 for $c = 3$ versus a theoretical 0.63, 0.78 versus 0.79 for $c = 4$ and 0.85 versus 0.86 for $c = 5$.

This remarkable agreement gives us confidence that we have a good control over finite size effects.



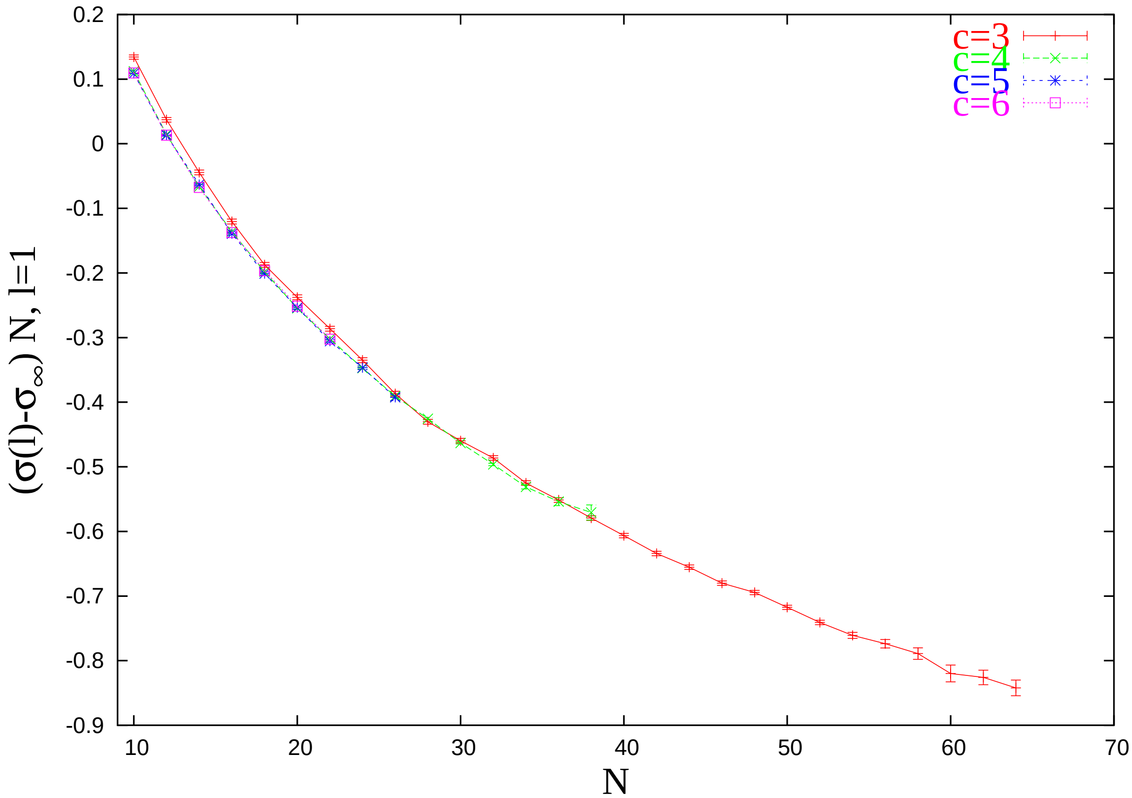
We have seen that for small values of ℓ, N times the difference between the circuit entropy $\sigma_N(\ell)$ and its asymptotic value behaves as

$$\big(\sigma_N(\ell) - \sigma_\infty(\ell)\big)N = -\log N + \tilde{\sigma}(\ell)$$

with $\tilde{\sigma}(\ell) = -\log(2\ell)$: it is independent of c, with a logarithmic dependence upon the graph size N.

To check if this behavior applies to finite values of ℓ, we look at the number of circuits with $\ell = 0.5$, for different values of N and c.

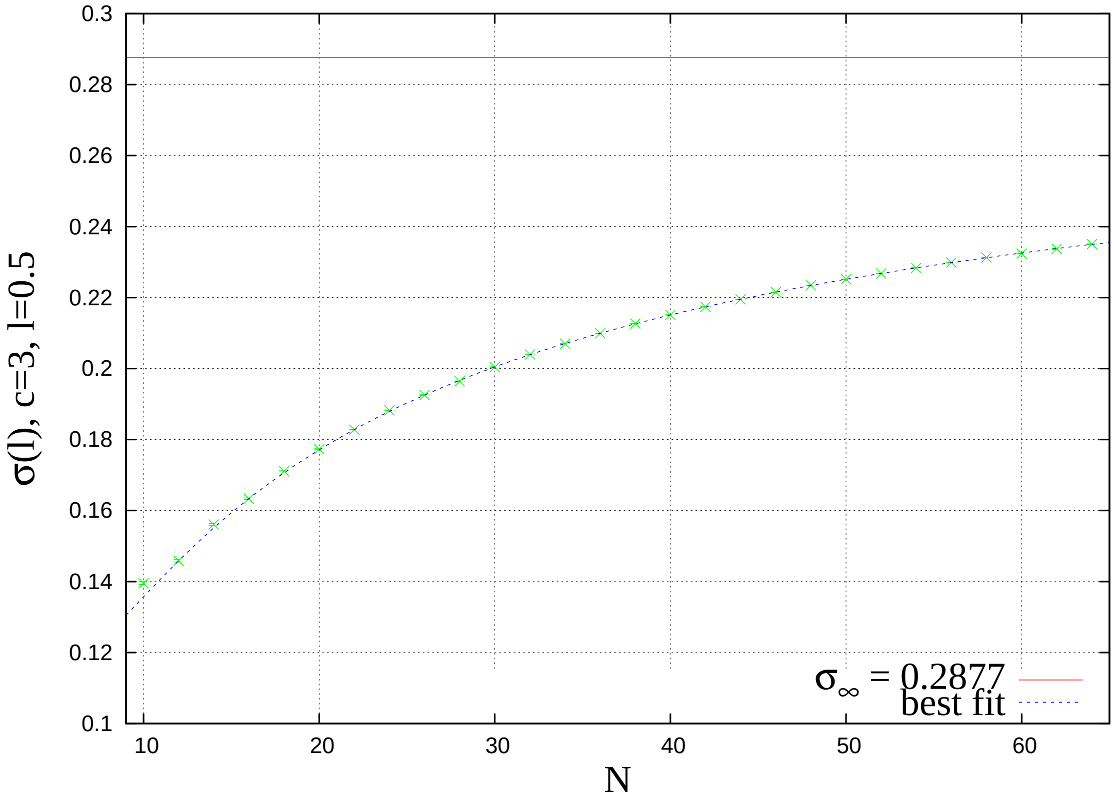
The data show only a very weak dependence upon c, that becomes weaker with increasing c. $c = 5$ data are already indistinguishable $c = 6$ data.



Here we show the same quantity for $\ell = 1$ *i.e.* for Hamiltonian circuits, that pass through all vertices of the random graph.

The scaling of Hamiltonian circuits is excellent already at $c = 3$.

We will come back later about the fact that scaling property of Hamiltonian circuits are very different from the ones of all other finite ℓ, less dense circuits.



Inspired by the case of small circuits we fit the circuit entropy for finite values of ℓ to

$$\sigma_N(l) = \sigma_\infty(l) + c_1 \frac{\log N}{N} + c_2 \frac{1}{N} \ .$$

In the figure we show our results for $c = 3$, $\ell = 0.5$. The quality of the best fit to data with sizes $N \geq 30$ only is excellent, and in very good agreement with all data with $N \geq 12$. This two parameter fit is clearly superior to power law fits.

With very good accuracy (surely better than one percent)

$$c_1 = -1 \ ,$$

i.e. even at finite ℓ the previous relation gives the correct leading corrections.

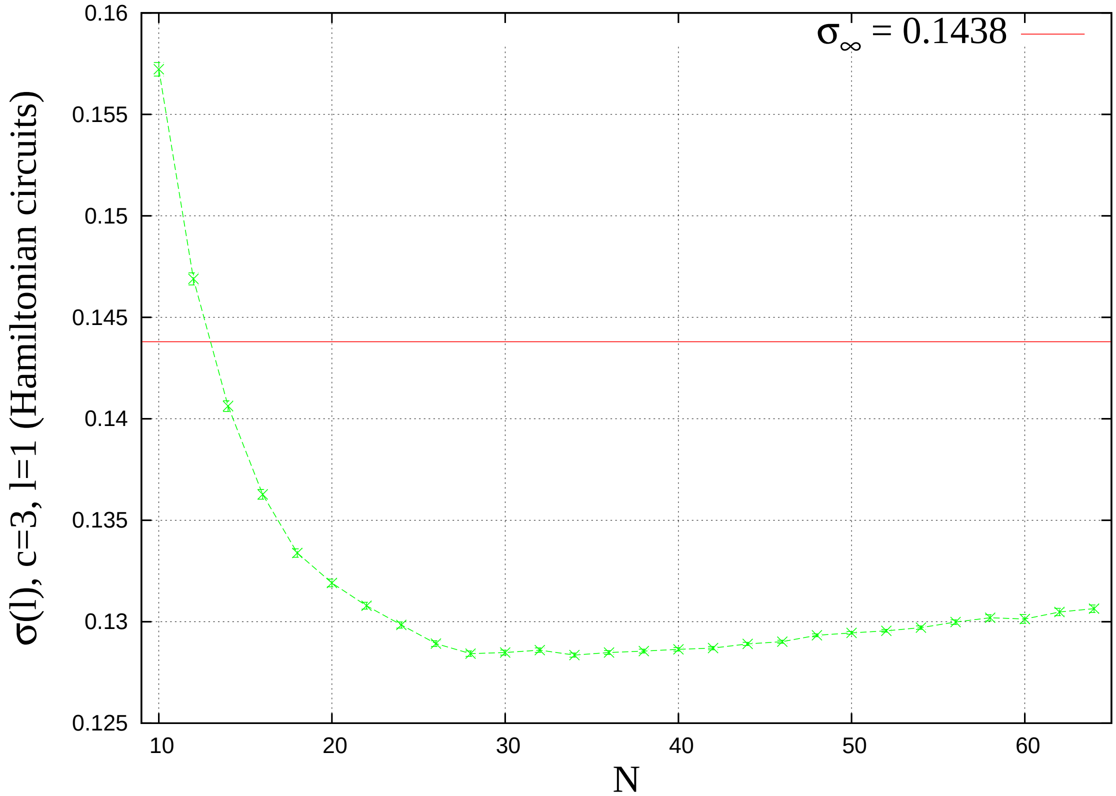
For all ℓ values (maybe excluding $\ell = 1$, see later) we find that the average number of circuits of reduced length ℓ equals

$$\langle M(\ell\, N) \rangle = \big(K(\ell) + o(1)\big)\ \frac{e^{N\,\sigma_\infty(\ell)}}{N}\ ,$$

where $K(\ell)$ is a bounded function of ℓ.

For c_2, we find values close to 1 e.g. .78 in the case of $\ell = 0.5$. Here precision is not as good since this is a sub-leading correction.

What is clear from our data is that sub-leading corrections to the circuit entropy are of the order of $1/N$.



As we have said above the case of Hamiltonian circuits ($\ell = 1$) is exceptional.

Finite size effects are very strong; this is intuitively expected since these circuits fill the graph and are deeply affected by its finite size.

It is clear that here the structure of finite size effects is completely different. On the contrary we have already explained that we find exactly the same behavior for all intermediate ℓ values: the case $\ell = 1$ appears to be isolated.

The absence of small loops in random graphs allows one to argue that the free-energy of a spin model defined on the graph is equal to the one on a regular tree with a self-consistent external field at boundary (leaves).

In turn, this free-energy fully determines the distribution of large-scale loops in the random graph.

We have added to our exact computation, valid in the $N \longrightarrow \infty$ limit, results from exact enumeration at finite N.

Thanks to them we have been able to determine precisely the behavior of the leading corrections to the thermodynamical behavior (at least for circuits with $\ell < 1$): we have found that Hamiltonian circuits have stronger finite size corrections and a peculiar finite N behavior.