

Statistical Physics of the Random Graph Model

Rémi Monasson

Laboratoire de Physique Théorique, École Normale Supérieure

April 6, 1998

[summary by Philippe Flajolet]

Abstract

This talk addresses the random graph model originally introduced by Erdős et Rényi in 1959. This model gives rise to a large number of threshold phenomena that are evocative of phase transitions in statistical physics. The talk illustrates the way several results on random graphs can be reexamined in a new perspective provided by a simple model of statistical physics, the Potts model. The problem addressed is principally that of the size of the giant component for which quantitative estimates are derived.

More generally, the talk is motivated by a desire to understand what statistical physics models may bring to the realm of threshold problems, not only in random graphs but also in the satisfiability of random boolean formulæ.

1. The Random Graph Models

The most natural random graph models have been introduced by Erdős and Rényi in a series of eventually famous papers that starts with [5, 6]. They are denoted by $G_{n,p}$ and $\widehat{G}_{n,e}$ and are defined as follow:

- $G_{n,p}$ considers graphs with n vertices in which each of the $N = \binom{n}{2}$ possible edges is present with probability p ;
- $\widehat{G}_{n,e}$ considers all graphs with n vertices and e edges as equally likely.

The first model is of the Bernoulli type (there are N trials, each with independent probability p of success), the second one is more “combinatorial”. Given the fact that the Bernoulli distribution $B(N, p)$ is narrowly centered around its mean Np , we expect the following fact.

The characteristics of $G_{n,p}$ resemble those of $\widehat{G}_{n,e}$ provided $e \approx Np$.

We refer globally to Bollobás’s book [4] for a discussion of these rich models and for precise conditions that make the assertion above into a valid mathematical statement. (The transfer from $\widehat{G}_{n,e}$ to $G_{n,p}$ is an Abelian one, whereas the converse transfer has a Tauberian flavour.)

Imagine a graph as evolving in time from totally disconnected to complete, through successive additions of edges that are reflected by increasing values of p from 0 to 1. What is characteristic of $G_{n,p}$ (and thus, of the companion $\widehat{G}_{n,e}$ model) is the presence of sharp thresholds. A threshold phenomenon for a property P means that there is a function $p_0(n)$ such that, with (very) high probability (as $n \rightarrow \infty$), P does not hold when $p \ll p_0(n)$ while for $p \gg p_0(n)$, P holds. (Of course, one may look for all sorts of detailed informations near the threshold $p_0(n)$.)

Here is a simplified picture of what goes on in $G_{n,p}$, expressed in terms of the mean number of edges, $m = Np$. Only isolated vertices and edges will be present when $m \ll n^{1/2}$; but trees of size 3

will start appearing at $m \approx n^{1/2}$, trees of size 4 at $m \approx n^{2/3}$, etc. There is (almost surely) no cycle when $m \ll n$. Later when $m = \lambda n/2$ and $\lambda < 1$ there is at most one cycle in each component and the largest component almost surely has size $\Theta(\log n)$. A dramatic phase transition occurs near $m = n/2$ when one or several large components of size $n^{2/3}$ appear. Still later, when $m = \lambda n/2$ and $\lambda > 1$, we find a single “giant” component of size $\Theta(n)$. However, we’ll have to wait a little longer, namely till $m \approx \frac{1}{2}n \log n$, to attain full connectedness, at which point the graph ceases to be interesting for the problems under discussion here.

There are various approaches to these problems. Most of them, following Erdős and Rényi’s original papers [5, 6], are probabilistic and well explained in [4]. Roughly, one has to cope in this framework with random variables satisfying intricate dependencies; moment methods, tail inequalities, or probabilistic inequalities are then essential. The literature in this direction is immense and Bollobás’s book already includes more than 700 references. The best results relative to connectedness that are available at this time (formulated in terms of $\widehat{G}_{n,e}$) are probably those of Bender, Canfield, and McKay; see [1, 2, 3]. In contrast, only a handful of papers starting with Knuth, Pittel, and collaborators resort to analytic methods¹; see [7, 9]. Even fewer papers rely on methods from statistical physics. The work under discussion here is a pioneering attack on this range of hard problems; see [10, 11] for applications of related ideas to the random k -satisfiability problem.

2. The Potts Model

The Potts model of statistical physics considers particles or sites whose states (sometimes referred to as “colours”) may assume any of q values. In the particular case of random graphs, it instantiates as follows. Consider n sites that we may imagine as regularly spaced on a circle. Each site may be in a certain *state* that is an integer of $\{0, 1, \dots, q - 1\}$. The integer q is a parameter of the model and when $q = 2$, one can think of the states as “spins” representing the orientation of some vector, e.g., a magnetic moment. A *configuration* $\Sigma = (\sigma_1, \dots, \sigma_n)$ is an assignment of states to each site, so that there are q^n possible configurations. The energy of a configuration is defined as

$$E(\Sigma) = -\frac{\gamma}{n} \sum_{i < j} \mathbf{1}_{\sigma_i, \sigma_j},$$

where $\mathbf{1}_{x,y}$ is the indicator of $x = y$ that has value 1 if $x = y$ holds and 0 otherwise. There γ is a parameter; the fact that one takes all the $N = n(n - 1)/2$ combinations $i < j$ corresponds to a model with complete interactions, that is, the underlying graph is the complete graph. (Models of statistical physics often consider instead an underlying graph constrained to be a regular lattice in dimension 1, 2, or 3.)

An essential object of statistical physics is the *partition function* defined here as

$$Z(\gamma, n) = \sum_{\Sigma} e^{-E(\Sigma)},$$

which is thus a sum of q^n terms. There are two main points in the talk: (i) the function Z provides information on the random graph model; (ii) it is possible to estimate analytically various characteristics of Z .

¹As said by Frieze in his discussion of the paper by Janson, Knuth, Łuczak, and Pittel [9] in the *Mathematical Reviews* [MR94h:05070]: “The paper [9] and its predecessor [7] mark the entry of generating functions into the general theory of random graphs in a significant way. Previously, their use had mainly been restricted to the study of random trees and mappings. However, at the early stages of the evolution of a random graph we find that it is usually not too far from being a forest, and this allows generating functions an entry.”

3. The Partition Function and Random Graphs

First of all, the partition function is (almost) a counting generating function in disguise. One has

$$Z(\gamma, n) = \sum_{\Sigma} \prod_{i < j} \exp\left(\frac{\gamma}{n} \mathbf{1}_{\sigma_i, \sigma_j}\right) = \sum_{\Sigma} \prod_{i < j} \left(1 + \mathbf{1}_{\sigma_i, \sigma_j} (e^{\gamma/n} - 1)\right),$$

as results from the identity $\mathbf{1}^2 = \mathbf{1}$. Next,

$$\begin{aligned} (1) \quad Z(\gamma, n) &\approx \sum_{\Sigma} \prod_{i < j} \left(1 + \frac{\gamma}{n} \mathbf{1}_{\sigma_i, \sigma_j}\right) \\ (2) \quad &= \sum_G \left(\frac{\gamma}{n}\right)^{\epsilon(G)} q^{c(G)}. \end{aligned}$$

The first line (1) is a natural approximation for n large. In the second line (2), $\epsilon(G)$ is the number of edges of the graph G , $c(G)$ the number of its connected components, and the sum is over all graphs G with n vertices. The reason why (2) is true is that the general term of the sum involves a product over all possible edges, and a product like $u_{\sigma_1, \sigma_2} u_{\sigma_2, \sigma_3}$ has value 1 only if $\sigma_1, \sigma_2, \sigma_3$ are of the same colour, in which case there are altogether q degrees of freedom.

Now, a graph G with n vertices and ϵ edges has probability

$$P_{\gamma}(G) = \left(\frac{\gamma}{n}\right)^{\epsilon} \left(1 - \frac{\gamma}{n}\right)^{N - \epsilon},$$

in the model $G_{n,p}$, where $p = \gamma/n$ and, like before, $N = n(n-1)/2$. Then, equation (2) yields the approximate formula,

$$(3) \quad Z(\gamma, n) \approx e^{\gamma n/2} \sum_G P_{\gamma}(G) q^{c(G)}.$$

(These approximations are stated here without error terms but it is not hard to assign them sufficient validity conditions.)

4. The Potts Model and the Process of “Analytic Continuation”

In order to approach the number of connected components, we return to the definition of the partition function and aim at transforming its expression. The energy, $E(\Sigma)$ depends on n variables, but only through their q possible values, with q the parameter of the Potts model. Indeed, for $\sigma \in \{0, \dots, q-1\}$, define the occupancy variables,

$$X_{\sigma}(\Sigma) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\sigma_i, \sigma},$$

that describe how many times each value of $\{0, \dots, q-1\}$ is used by a configuration. One finds easily that

$$E(\Sigma) = \frac{\gamma}{2} - \frac{\gamma n}{2} \sum_{\sigma=0}^{q-1} X_{\sigma}^2.$$

(The term $\gamma/2$ is ignored in subsequent computations.) Now, grouping states according to the values of their occupancy vectors $\{X_\sigma\}$ yields

$$Z(\gamma, n) \approx \sum_{\{X_\sigma\}} \binom{n}{nX_0, \dots, nX_{q-1}} \exp\left(-\frac{\gamma n}{2} \sum_{\sigma=0}^{q-1} X_\sigma^2\right),$$

where the X_σ are such that $\sum_\sigma X_\sigma = 1$ and the X_σ go by steps of $\frac{1}{n}$. (The original derivation of Monasson makes use of manipulations with indicator variables that are related to the theory of “replicas”.) Then, Stirling’s formula employed to approximate the factorials present in the multinomial coefficient produces

$$(4) \quad Z(\gamma, n) \approx \sum_{\{X_\sigma\}} \exp\left(-n \left(\sum_\sigma X_\sigma \log X_\sigma + \frac{\gamma}{2} \sum_\sigma X_\sigma^2\right)\right).$$

The form (4) is indeed a q -fold sum and the original n -fold summations and products have been eliminated.

The next step consists in evaluating what happens with the approximation (4) taken as

$$(5) \quad Z(\gamma, n) \approx \sum_{\{X_\sigma\}} \exp(-nG(\{X_\sigma\})).$$

The idea is to estimate the sum by means of the q -dimensional Laplace method, which requires locating the *global* extrema of the exponential. It is observed that *local* extrema at least are obtained by trying

$$(6) \quad X_0 = \frac{1}{q}(1 + (q-1)s), \quad X_1 = \dots = X_{q-1} = \frac{1}{q}(1-s).$$

Then, the argument of the exponential in (4) is locally maximized if one fixes s as a root of

$$(7) \quad \log\left(\frac{1 + (q-1)s}{1-s}\right) = \gamma s.$$

It is believed that all global extrema are obtained in this way, up to permutation of indices. Under this assumption, the Laplace method can then be applied to approximate $Z(\gamma, n)$ as

$$(8) \quad Z(\gamma, n) \approx e^{-nF(q, \gamma)},$$

where F is essentially $G(\{X_\sigma\})$ of (5) evaluated at the X_σ given by (6) in terms of q, s , where $s = s(\gamma)$ satisfies the condition (7).

We now dive into a more conjectural world based on a special kind of formal reasoning. The principle of the heuristic analysis consists in extrapolating the asymptotic approximation of the partition function that is defined a priori for integer values of q only and make it an analytic function of q . Then, let q tend to 1 and hope for consistency. Once this is done, various results that can be checked successfully against known ones (via analytic or probabilistic methods) are obtained.

Let us postulate the validity of (8) for all *real* q , and in particular for q near 1. Within this framework, the mean number of connected components of a random graph of $G_{n, \gamma/n}$ is for instance accessible as ($\langle X \rangle$ denotes expectation):

$$(9) \quad \langle c(G) \rangle = \left. \frac{d}{dq} \log Z(\gamma, n) \right|_{q=1},$$

where *real* values $q \rightarrow 1$ are used. For the region of interest which is thus q near 1, equation (7) becomes $1 - s = e^{\gamma s}$. This equation defines s as a function of γ , $s = s(\gamma)$. The parameter $s(\gamma)$ is in fact an indicator of the fraction of sites in the largest connected component of the random graph. There is a bifurcation at $s = 1$. The function $s(\gamma)$ is identically 0 when $\gamma < 1$, a fact consistent with known properties of the random graph before the emergence of the giant component. At $s = 1^+$, the function $s(\gamma)$ has a square-root singularity, while it becomes analytic for $s > 1$. Thus informations about the giant component and its “phase transition” become amenable to this approach (details omitted in this abstract).

5. Discussion

A summary of the methodology is as follows. For a given problem, there are a priori two “partition functions”,

$$Z_{\text{comb}} := \sum_G P(G) q^{c(G)}, \quad Z_{\text{phys}} = \sum_{\Sigma} e^{-E(\Sigma)}.$$

The process is then as follows.

1. Choose the configuration space and energy function so that $Z_{\text{comb}} \approx Z_{\text{phys}}$.
2. Evaluate Z_{phys} by: (i) identifying the order parameters (the X_σ); (ii) determining asymptotic approximations (here by the Laplace method); (iii) performing an analytic “continuation” according to the chain “(q integer) \mapsto (q real) \mapsto ($q \rightarrow 1$)”.

The major question to ask is why and to what extent does this approach provide useful quantitative result. Certainly, the approximations for fixed integer q can be justified. Also there is a possibility of matching the analysis near $q = 1$ against what we know from analytic approaches. (For instance, it is known that the emergence of the giant component is related in an essential way to the occurrence of two coalescing saddle points.) So, in a way, the most surprising fact to be explained is that estimates initially conducted for integer q only (we used a q -dimensional Laplace method!) can be “analytically continued” to the region of q near 1.

We observe that complex analysis does *sometimes* provide a framework for such analytic continuation. For instance, a theorem of Carlson asserts that when a function $\phi(s)$ is analytic (holomorphic) in a right-hand half-plane and is of moderate growth, $\phi(s) = O(e^{(\pi-\epsilon)|s|})$, then: *$\phi(s)$ vanishes identically if and only if it vanishes at the nonnegative integers*. Therefore, an identity $A(s) = B(s)$ can be inferred just from its specialization at the integers² provided it is known a priori that A, B don’t grow too fast. For instance, it suffices to establish

$$\sin^2 \frac{\pi s}{4} + \cos^2 \frac{\pi s}{4} = 1,$$

for $s = 0, 1, \dots$, in order to be sure that it holds for all complex s . Observations of this kind however fall short of providing a basis for the analytic continuation process employed here, given the intricate nature of the approximations involved.

Bibliography

- [1] Bender (Edward A.), Canfield (E. Rodney), and McKay (Brendan D.). – The asymptotic number of labeled graphs with given number of vertices and edges. *Random Structures & Algorithms*, vol. 1, n° 2, 1990, pp. 127–169.
- [2] Bender (Edward A.), Canfield (E. Rodney), and McKay (Brendan D.). – Asymptotic properties of labeled connected graphs. *Random Structures & Algorithms*, vol. 3, n° 2, 1992, pp. 183–202.

²It is well worth pointing at Hardy’s discussion in [8, Ch. 11] of “Ramanujan’s heuristic” on which Ramanujan based so much of his definite integral evaluations.

- [3] Bender (Edward A.), Canfield (E. Rodney), and McKay (Brendan D.). – The asymptotic number of labeled graphs with n vertices, q edges, and no isolated vertices. *Journal of Combinatorial Theory. Series A*, vol. 80, n° 1, 1997, pp. 124–150.
- [4] Bollobás (Béla). – *Random Graphs*. – Academic Press, 1985.
- [5] Erdős (P.) and Rényi (A.). – On random graphs. I. *Publ. Math. Debrecen*, vol. 6, 1959, pp. 290–297.
- [6] Erdős (P.) and Rényi (A.). – On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 5, 1960, pp. 17–61.
- [7] Flajolet (P.), Knuth (D. E.), and Pittel (B.). – The first cycles in an evolving graph. *Discrete Mathematics*, vol. 75, 1989, pp. 167–215.
- [8] Hardy (G. H.). – *Ramanujan: Twelve Lectures on Subjects Suggested by his Life and Work*. – Chelsea Publishing Company, New-York, 1978, third edition. Reprinted and Corrected from the First Edition, Cambridge, 1940.
- [9] Janson (Svante), Knuth (Donald E.), Łuczak (Tomasz), and Pittel (Boris). – The birth of the giant component. *Random Structures & Algorithms*, vol. 4, n° 3, 1993, pp. 233–358.
- [10] Monasson (Rémi) and Zecchina (Riccardo). – Entropy of the K -satisfiability problem. *Physical Review Letters*, vol. 76, n° 21, 1996, pp. 3881–3885.
- [11] Monasson (Rémi) and Zecchina (Riccardo). – Statistical mechanics of the random K -satisfiability model. *Physical Review E. Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics. Third Series*, vol. 56, n° 2, 1997, pp. 1357–1370.