The Giant Component: The Golden Anniversary

Joel Spencer

Paul Erdős and Alfréd Rényi

Oftentimes the beginnings of a mathematical area are obscure or disputed. The subject of *random graphs* had, however, a clear beginning, and it occurred fifty years ago. Alfréd Rényi (1921–1970) was head of the Hungarian Mathematical Institute which today bears his name. Rényi had a great love of literature and philosophy, a true Renaissance man. His life, in the words of Paul Turán, was one of intense and creative involvement in the exchange of ideas and in public affairs. His mathematics centered on probability theory. Paul Erdős (1913–1996) was a central figure in twentieth-century mathematics. This author recalls well the memorial conference organized by his long-time friend and collaborator Vera Sós in 1999, at which fifteen plenary lecturers discussed his contributions. What was so surprising to us all was the sheer breadth of his work, which spanned so many vital areas of mathematics. To this very prejudiced author, it is his work in discrete mathematics that is having the greatest lasting impact. In this area, Erdős tended toward asymptotic questions, a style very much relevant to today's world of o, O, Θ, and Ω. For both Erdős and Rényi, mathematics was a collaborative enterprise. They both had numerous coauthors, and they wrote thirty-two joint papers.

In 1960 they produced their masterwork, *On the Evolution of Random Graphs* [7]. Begin with n vertices and no edges and add edges randomly (that is, uniformly from among the potential edges) one by one. Let $G[n,e]$ be the state when there are e edges. Of course, $G[n,e]$ could be any graph with n vertices and e edges and technically is the uniform probability distribution over all such graphs. Erdős and Rényi studied the typical behavior of $G[n,e]$ as e "evolved" from 0 to $\binom{n}{2}$. When e approaches and passes $\frac{n}{2}$ the random graph undergoes a phase transition. In a typical computer run on a million vertices, the size of the largest component is only 168 at $e = 400000$ edges. By $e = 600000$ it has exploded to size 309433.

In modern language we let $G(n,p)$ be the probability space over graphs on n vertices where each pair is adjacent with independent probability p. Such graphs have very close to a proportion p of the edges. The behaviors of $G[n,e]$ and $G(n,p)$ where $e = p\binom{n}{2}$ are asymptotically the same for all the topics we discuss here, and we shall use the modern language.

We shall parameterize $p = \frac{c}{n}$. The graph with $\frac{n}{2}$ edges then corresponds to $c = 1$. In this range the graph will split into components. We let C_1, C_2 denote the largest and second largest components in the graph, with $|C_i|$ denoting their number of vertices. We define the *complexity* of a component with V vertices and E edges as $E - V + 1$. Trees and unicyclic graphs have complexity 0 and 1, respectively, and are called *simple*.

Theorem 1 (Erdős-Rényi)**.** *The behavior of $G(n,p)$ with $p = \frac{c}{n}$ can be broken into three parts.*
`Subcritical` $c < 1$*: All components are simple and very small,* $|C_1| = O(\ln n)$.
`Critical` $c = 1$*:* $|C_1| = \Theta(n^{2/3})$ *A delicate situation!*
`Supercritical` $c > 1$*:* $|C_1| \sim yn$ *where* $y = y(c)$ *is the positive solution to the equation*

$$e^{-cy} = 1 - y \quad (1)$$

Joel Spencer is professor of computer science and mathematics at the Courant Institute, New York University. His email address is `spencer@cims.nyu.edu`.

C_1 has high complexity. All other components are simple and very small, $|C_2| = O(\ln n)$.

Remark. The full statements here and below are rather unwieldy. Thus $|C_1| \sim yn$ really means that for all $c > 1$ and all $\epsilon > 0$ the limit as $n \to \infty$ of the probability that $(y-\epsilon)n < |C_1| < (y+\epsilon)n$ is 1. We allow ourselves the more informal description.

Remark. The average degree of a vertex is $p(n-1) \sim c$. The critical behavior takes place when the average degree reaches 1.

When $c > 1$, Erdős and Rényi called C_1 the *giant component*. There are two salient features of the giant component: its existence and its uniqueness. Their system does not have a Jupiter and an also huge Saturn; it is more like Jupiter and Ceres. Several books ([3], [4], [13]) give very full discussions.

Francis Galton and Henry Watson

Let's switch gears. A *Galton-Watson* process begins with a single node "Eve".[1] Eve has Z children, where Z has a Poisson distribution with mean c. (That is, Eve has k children with probability $e^{-c}c^k/k!$. The full study of Galton-Watson processes considers other distributions as well.) These children then have children independently with the same distribution, and the process continues through the generations. Let $T = T_c$ be the total population generated. There is a precise formula

$$\Pr[T_c = k] = \frac{e^{-ck}(ck)^{k-1}}{k!}. \tag{2}$$

However, it is also possible the T is infinite.

Theorem 2. *The Galton-Watson process has three regions.*
`Subcritical` $c < 1$*: T is finite with probability* 1*, and* $E[T] = \sum_{k=0}^{\infty} c^k = \frac{1}{1-c}$.
`Critical` $c = 1$*: T is finite with probability* 1 *but has infinite expectation.*[2]
`Supercritical` $c > 1$*. T is infinite with probability* $y = y(c)$ *given by (1).*

With $c > 1$ let $z = 1 - y$ be the probability Eve generates a finite tree. If Eve has k children the full tree will be finite if and only if all of the children generate finite trees, which has probability z^k. Thus

$$z = \sum_{k=0}^{\infty} e^{-c}\frac{c^k}{k!}z^k, \tag{3}$$

and some manipulation gives that $y = 1 - z$ satisfies (1). One needs further argument to show that $z = 1, y = 0$ is a spurious solution.

[1] *In the original work the nodes were all male!*

[2] *This author recalls in undergraduate days first seeing a finite random variable with infinite expectation and thinking it was a very funny and totally anomalous creation. Wrong! Such variables occur frequently at critical points in percolation processes.*

Erdős Meets Galton

Fix a vertex v in $G(n,p)$ with $p = \frac{c}{n}$ and perform a breadth first search (BFS) to find its components. It has binomial distribution $B[n-1,p]$ neighbors, asymptotically Poisson with mean c. Its neighbors then have Poisson c new neighbors, and so on. The component of v is approximated by the Galton-Watson process. For $c < 1$ this approximation works well. But for $c > 1$ the Galton-Watson process may go on forever while the component of v can have at most n vertices. An *ecological limitation* causes the processes to converge. BFS requires new vertices. After δn vertices have been found, the new distribution is binomial $B[n(1-\delta),p]$, which is Poisson with mean $c(1-\delta)$. The success of BFS causes δ to rise, which makes it harder to find new vertices, leading the process to eventually die. The effect of the ecological limitation is only felt after a positive proportion δn of vertices have been found. Consider BFS from each vertex v. With probability $1 - y$ the process will die early, giving a small component. But for $\sim yn$ the process will not die early. All of these vertices have their components merge into the giant component.

Jupiter Without Saturn

Why can we not have Jupiter and Saturn, two components both of size bigger than δn? This would be highly unstable. Each additional edge would have probability at least $(\delta n)^2/\binom{n}{2} \sim 2\delta^2$ of merging them. High instability and nonexistence are not the same. Indeed, while there are many proofs of the uniqueness of the giant component, we do not know one that is both simple and rigorous.

The Critical Window

Erdős and Rényi normally repressed their enthusiasm in their formal writings. But not now!

> This double "jump" in the size of the largest component when $\frac{e}{n}$ passes the value $\frac{1}{2}$ is one of the most striking facts concerning random graphs. [7]

In the 1980s, spearheaded by the work of Béla Bollobás and Tomasz Łuczak, the value $c = 1$ was stretched out, and a critical window was found. The stretching was done by adding a second-order term. The correct parameterization is

$$p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}. \tag{4}$$

Now there are three regions:
`Barely Subcritical` $pn \sim 1$ and $\lambda \to -\infty$: All components are simple. $|C_1| \sim |C_2|$, their sizes increasing with λ.
`The Critical Window` $pn \sim 1$ and λ a constant: Here (and only here!) we have chaotic behavior,

distributions instead of almost sure behavior. Parameterizing $|C_i| = X_i n^{2/3}$, the X_i ($i = 1, 2$ and beyond) have a nontrivial distribution and a nontrivial joint distribution. The complexity Y_i of C_i also has a nontrivial distribution.

`Barely Supercritical` $pn \sim 1$ and $\lambda \to +\infty$: $|C_1| \gg n^{2/3} \gg |C_2|$. C_1 is the *dominant component*, much bigger than C_2 but still small. C_1 has high complexity, but all other components are simple.

Stirling's formula applied to (2) with $c = 1$ gives $\Pr[T_1 = k] \sim (2\pi)^{-1/2}k^{-3/2}$ and $\Pr[T_1 \geq k] \sim 2(2\pi)^{-1/2}k^{-1/2}$. Now consider $G(n, p)$ with $pn = 1$ and let $C(v)$ be the component containing v. Call $C(v)$ large if its size is at least $Kn^{2/3}$ and let Z be the number of v with $C(v)$ large. Estimating $|C(v)|$ by T_1 would give that $C(v)$ is large with probability $2(2\pi)^{-1/2}K^{-1/2}n^{-1/3}$, and so Z would have expectation $2(2\pi)^{-1/2}K^{-1/2}n^{2/3}$. The actual value of $E[Z]$ is somewhat smaller due to the ecological limitation, but let us assume it as a heuristic. If any large component exists, every vertex of it would be in a large component, so that Z would be at least $Kn^{2/3}$. When K is large, $E[Z]$ is much lower than $Kn^{2/3}$, so that with high probability there would be no large component. Conversely, when ϵ is a small positive constant, we expect many components of size bigger than $\epsilon n^{2/3}$.

A Strange Physics

Let $c_i n^{2/3}$ be the size of the ith largest component at $p = n^{-1} + \lambda n^{-4/3}$. Let $\Delta\lambda$ be an "infinitesimal" and increase λ by $\Delta\lambda$. There are $(c_i n^{2/3})(c_j n^{2/3})$ potential edges that would merge C_i, C_j, and each is added with probability $(\Delta\lambda)n^{-4/3}$. The n factors cancel: C_i, C_j merge to form a component of size $(c_i + c_j)n^{2/3}$ with probability $c_i c_j(\Delta\lambda)$. This gravitational attraction merges the large components and forms the dominant component. We can include the complexity in this model. When C_i, C_j with complexities r_i, r_j merge, the new component has complexity $r_i + r_j$. Further, each C_i has $\sim \frac{1}{2}c_i^2 n^{4/3}$ potential internal edges. In the infinitesimal time $\Delta\lambda$ with probability $c_i^2(\Delta\lambda)/2$, such an edge is added, and the complexity of C_i is incremented by 1. Over time, the complexities get larger and larger. The limiting process, called the multiplicative coalescent process, has interesting connections to Brownian motion [2].

A Computer Exercise

Computer experimentation vividly shows the rapid development in the critical window. In the run[3] on the following page, we begin with $n = 10^5$ vertices and no edges. At each step a random edge is added, and a Union-Find algorithm is used to keep track of component sizes. We parameterize the number of edges as $e = \binom{n}{2}(n^{-1} + \lambda n^{-4/3})$ and take "snapshots" at $\lambda = -4, -3, \ldots, +4$. The ten largest component sizes (listed $0, \ldots, 9$ here, and divided by $n^{2/3}$) are given for each λ. At $\lambda = +2$ there is a 1.16 Jupiter and 0.86 Saturn. The next digit, under N, gives the new ranking (– if not in the top ten) for that component for the next λ. Components $0, 1, 2, 3, 4$ have $N = 0$, meaning they have all merged by $\lambda = +3$. At $\lambda = 3$ Jupiter has blown up to 4.21. (Smaller components have also joined Jupiter, explaining the discrepancy in the sum.) The size of the second largest component has *decreased* (it is the component formerly ranked 5) to a 0.22 Ceres.

Inside the Critical Window

At $\lambda = -4$ there is a "jostling for position" among the top components, while by $\lambda = +4$ a dominant component has emerged. The last time the largest component loses that distinction occurs during the critical window [8]. At $\lambda = -4$ all components are simple, while by $\lambda = +4$ the dominant component has high complexity. Complexity at least 4 is necessary for nonplanarity. Planarity is lost in the critical window [16]. In a masterful work [12], the development of complex components is studied. One exceptionally striking result: the probability that the evolution ever simultaneously has two complex components is, asymptotically, $\frac{5}{18}\pi$.

Classical Bond Percolation

Mathematical physicists examine Z^d as a lattice; the pairs $\vec{v}, \vec{w}$ that are one apart are called *bonds*. They imagine that each bond is *occupied* with independent probability p. (For graph theorists, the occupied bonds form a random subgraph of Z^d.) The occupied components then form clusters, or components. There is a critical probability, denoted p_c (dependent on d), so that:

`Subcritical` $p < p_c$: All components are finite.

`Critical` $p = p_c$: A delicate situation!

`Supercritical` $p > p_c$: There is precisely one infinite component.

There are natural analogies between this infinite model and the asymptotic Erdős-Rényi model. Infinite size corresponds to $\Omega(n)$, while finite size corresponds to $O(\ln n)$. There is particular interest in p being very close to p_c. Let $f(p)$ be the probability that $\vec{0}$ (or, by symmetry, any particular $\vec{v}$) lies in an infinite component. The *critical exponent* β is that[4] real number such that $f(p_c + x) \sim x^{\beta + o(1)}$ as $x \to 0^+$. The probability $p_c + x$ corresponds to $pn = 1 + x$ and f to the probability that a given vertex v lies in the giant component or, equivalently, the proportion $y(1 + x)$ of vertices in

[3] *Thanks to Juliana Freire.*

[4] *β might not exist, but all mathematical physicists assume it does.*

·	-4	N	-3	N	-2	N	-1	N	0	N	+1	N	+2	N	+3	N	+4
0	0.14	1	0.18	0	0.24	1	0.28	0	0.37	0	0.82	0	1.16	0	4.21	0	5.88
1	0.10	2	0.16	1	0.19	2	0.26	1	0.36	1	0.39	1	0.86	0	0.22	0	0.24
2	0.10	3	0.13	3	0.13	4	0.19	3	0.26	0	0.28	3	0.49	0	0.13	0	0.12
3	0.09	4	0.12	4	0.13	0	0.16	2	0.21	2	0.23	4	0.46	0	0.12	0	0.10
4	0.07	0	0.09	5	0.13	3	0.14	5	0.16	0	0.20	2	0.32	0	0.11	2	0.10
5	0.07	0	0.08	7	0.09	5	0.12	4	0.15	4	0.14	5	0.16	1	0.10	1	0.10
6	0.07	5	0.06	6	0.09	7	0.10	8	0.12	5	0.12	2	0.12	3	0.09	1	0.10
7	0.06	8	0.06	2	0.08	6	0.09	-	0.12	3	0.10	1	0.11	4	0.09	4	0.10
8	0.05	-	0.06	8	0.06	9	0.09	-	0.10	3	0.10	7	0.09	2	0.09	0	0.08
9	0.05	9	0.05	-	0.06	0	0.07	-	0.10	9	0.10	0	0.08	5	0.08	6	0.07

the giant component. As $y(1+x) \sim 2x = x^{1+o(1)}$ the β value for the Erdős-Rényi model is considered 1. This is known to be the β value in Z^d for all sufficiently high dimensions d. Grimmett [10] gives many other critical exponents, and in all cases the analogous value for the Erdős-Rényi model matches the known value in high-dimensional space. Mathematical physicists loosely use the term *mean field behavior* to describe percolation phenomena in high dimensions, and the Erdős-Rényi model has this mean field behavior.

Recent Results

Today it is recognized that percolation and the critical window appear in many guises. Here is a highly subjective description of recent work. We generally give simplified versions.

Random 2-SAT

We generate m random clauses $C_1, \ldots, C_m$ on Boolean variables $x_1, \ldots, x_n$. That is, each clause $C = y \vee z$ with y, z drawn randomly from $\{x_1, \overline{x_1}, \ldots, x_n, \overline{x_n}\}$. We ask if all C_i can be simultaneously satisfied. The answer changes from yes to no in the critical window $m = n + \lambda n^{2/3}$ [5].

d-regular Graph

Let G_n be a sequence of transitive d-regular graphs. Under reasonable conditions $p_c = \frac{1}{d-1}$ acts as the critical probability for a random subgraph of G_n. For $p < p_c$ the components are small, while for $p > p_c$ there is a giant component. More delicately, at $p = p_c$ the largest component has size $\Theta(n^{2/3})$. The scaling $pd = 1 + \lambda n^{-1/3}$ acts as the critical window [18].

An Improving Walk

Consider an infinite walk starting at $W_0 = 1$ with $W_t = W_{t-1} + X_t - 1$ where X_t is Poisson with mean $\frac{t}{n}$. When $W_t = 0$ (crashes) it is reset to $W_t = 1$. When $t = \frac{1-\epsilon}{n}$ the walk has negative drift and crashes repeatedly. When $t = \frac{1+\epsilon}{n}$ the walk has positive drift and goes to infinity. The walk will crash for the last time in the critical window $t = n + \lambda n^{2/3}$ [9].

A First-Order Phase Transition

Modify the Erdős-Rényi evolution as follows. In each round an edge is added to G, initially empty. *Two* random pairs $\{u, v\}$, $\{w, x\}$ are given. Add that pair for which the product of the component sizes of the two vertices is smaller. This provides a powerful antigravity that deters large components from joining. Parameterizing $e = t\frac{n}{2}$ edges chosen (so that $t = 1$ is the critical value in the Erdős-Rényi evolution) the giant component occurs at $t \sim 1.77$. More interesting, extensive computer simulation (but no mathematical proof!) indicates strongly that when the critical value is reached, there is a first-order phase transition. That is, let $t = t_c$ be the critical probability and let $f(t)$ be the proportion of vertices in the largest component at "time" t. Then the limit of $f(t)$ as t approaches t_c from above appears not to be zero but rather something like 0.6 [1].

General Critical Points

Let G be a graph on n vertices. Set $d^* = (\sum_v d_v^2)(\sum_v d_v)$, noting d_v^* denotes the average degree of a vertex if you first select an edge uniformly and then one of its vertices uniformly. Let G_p denote the random subgraph of G, accepting each edge with independent probability p. Then, under certain mild conditions on G, $p = \frac{1}{d^*}$ is the critical point in the evolution of G_p. When $p = \frac{1-\epsilon}{d^*}$, G_p contains no giant component, while when $p = \frac{1+\epsilon}{d^*}$, G_p does contain a unique giant component [6].

Degree Sequence

For given $d_1, \ldots, d_n$ we consider the graph on n vertices chosen uniformly among all with that degree sequence, that is, v having precisely d_v neighbors. Suppose for each n we have a degree sequence, $\lambda_i(n) \sim \lambda_i n$ vertices having degree i. Then (with d^* from above), $d^* \sim [\sum i^2 \lambda_i]/[\sum i\lambda_i]$. Set $Q := \sum_i i(i-2)\lambda_i$ so that $Q > 0$ if and only if $d^* > 2$. In analyzing BFS one thinks of an edge going to a vertex with the above distribution, which then is on an expected number $d^* - 1$ of new edges. With $d^* < 2$ the process will die, while for $d^* > 2$ it might continue. When $Q < 0$ the random graph with this degree sequence has no giant component, while when $Q > 0$ it does [17].

Set $Q_n := \sum_i i(i-2)\lambda_i(n)$ and assume $Q_n \to 0$. Under moderate assumptions, $Q_n = \lambda n^{-1/3}$ provides a critical window. For $\lambda \to -\infty$ the random graph is subcritical, and all components have size $o(n^{2/3})$. When $\lambda \to +\infty$ the random graph is supercritical, there is a dominant component of size $\gg n^{2/3}$ and all other components have size $o(n^{2/3})$. In the power law random graphs, thought by many to model the Web graph and other phenomena, it is assumed that $\lambda_i \sim i^{-\gamma}$ for a constant γ. For certain γ the above critical window does not work, and work in progress indicates that there is a critical window whose exponent depends on γ [14] [11].

A Potts Model

In the Potts model, the distribution of graphs is biased toward having more components. There are three parameters, $p \in [0,1]$, $q \geq 1$, and the number of vertices n. A graph G with e edges, $s := \binom{n}{2} - e$ nonedges, and c components has probability $p^e(1-p)^s q^c/Z$, where Z is a normalizing constant chosen so that the sum of the probabilities is 1. For $q = 2$ this is called the Ising model, for $q \geq 3$ and integral, this is the Potts model. For $2 < q < 3$ the critical value is $pn = c_q := 2\frac{q-1}{q-2}\ln(q-1)$. At $pn = c_q + \epsilon$ there is a giant component, while at $pn = c_q - \epsilon$ the largest component has logarithmic size. The critical window has parameterization $pn = c_q + \frac{\lambda}{n}$. There the graph has two different personalities. Either it has a giant component *or* the largest component has logarithmic size. Both occur with positive limiting probability, and these limiting probabilities sum to 1. At no p is there a middle ground with the size of the largest component being bigger than logarithmic but sublinear [15].

In Conclusion

The mathematical landscape at the time of Paul Erdős's birth, nearly one hundred years ago, was very different from what it is today. Discrete mathematics was disparaged as "the slums of topology". Probability was useful for gambling, but not proper work for a serious mathematician. Today both areas are thriving. It is the fecund intersection of discrete mathematics and probability that has seen the most spectacular growth. A wide variety of random processes on large discrete structures are studied. These processes, to use Erdős and Rényi's well-chosen word, undergo an *evolution*. At a critical moment they undergo a phase transition, from water to ice, from satisfiable to not satisfiable, from freeflow to gridlock, from small components to a giant component. To understand a process we need to understand these critical moments. The Erdős-Rényi process provides a bedrock to which all other processes may be compared.

References

[1] D. Achlioptas, R. M. D'Souza and J. Spencer, Explosive percolation in random networks, *Science* **323** (5920), 1453–1455, 2009.

[2] David Aldous, Brownian excursions, critical random graphs, and the multiplicative coalescent, *Annals of Probability* **25** (1997), 812–854.

[3] N. Alon and J. Spencer, *The Probabilistic Method*, 3rd ed., John Wiley, 2009.

[4] Béla Bollobás, *Random Graphs*, 2nd ed., Cambridge University Press. 2001.

[5] Béla Bollobás, Christian Borgs, Jennifer Chayes, Jeong Han Kim, and David B. Wilson, The scaling window of the 2-SAT transition, *Random Structures and Algorithms* **21** (2002), 182–195.

[6] Fan Chung, Paul Horn, and Lincoln Lu, The giant component in a random subgraph of a given graph, *Proceedings of the 6th Workshop on Algorithms and Models for the Web-Graph (WAW2009)*, Lecture Notes in Computer Science 5427, 38–49.

[7] Paul Erdős and Alfréd Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

[8] Paul Erdős and Tomasz Łuczak, Changes of leadership in a random graph process, *Random Structures and Algorithms* **5** (1994), 243–252.

[9] Juliana Freire, Analysis techniques for nudged random graphs, Ph.D. Thesis, Courant Institute, 2009.

[10] G. Grimmett, *Percolation*, 2nd ed., Springer-Verlag, 1999.

[11] S. Janson and M. Luczak, A new approach to the giant component problem, *Random Structures and Algorithms* **34** (2008), 197–216.

[12] Svante Janson, Tomasz Łuczak, Donald Knuth, and Boris Pittel, The birth of the giant component, *Random Structures and Algorithms* **3** (1993), 233–358.

[13] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random Graphs*, John Wiley, 2000.

[14] Milyun Kang and T. Seierstad, The critical phase for random graphs with a given degree sequence, *Combin. Probab. Comput.* **17** (2008), 67–86.

[15] Malwina Luczak and Tomasz Łuczak, The phase transition in the cluster-scaled model of a random graph, *Random Structures and Algorithms* **28** (2006), 215–246.

[16] Tomasz Łuczak and John Wierman, The chromatic number of random graphs at the double jump threshold, *Combinatorica* **9** (1989), 39–49.

[17] M. Molloy and B. Reed, The size of the largest component of a random graph on a fixed degree sequence, *Combin. Probab. Comput.* **7** (1998), 295–306.

[18] Asaf Nachmias, Mean-field conditions for percolation on finite graphs, GAFA (to appear).