Souvik Dhara
Department of Mathematics
Massachusetts Institute of Technology
sdhara@mit.edu

Research Stieltjes Prize 2018

Universality for critical behavior of percolation processes on complex networks

In 2018 Souvik Dhara has been awarded the Stieltjes Prize for one of the two best PhD theses in mathematics in the Netherlands. The prize was awarded for his thesis entitled *Critical Percolation on Random Networks with Prescribed Degrees*, which he completed at Eindhoven University of Technology. He now is a postdoc at MIT in Cambridge, USA. In this article he discusses recent developments in percolation theory.

Complex networks are well-known tools to model complex architectures arising from applications in social, physical, and life sciences. The structural properties of complex networks have been intensively studied since the end of the nineties, and the current focus has shifted to dynamic processes taking place on networks, such as the spread of epidemics. Interestingly, these processes are often known to exhibit a *phase transition* phenomenon. In the context of epidemics, whether there will be an outbreak or not depends crucially on whether the infection probability is above or below the epidemic threshold. The behavior near the point of phase transition is called the *critical behavior*, which in a sense captures the mechanism of this phase transition. In the late nineties, physicists started studying critical behavior in complex networks due to their applications in condensed matter theory. The overall wisdom from this vast physics literature is that the intrinsic nature of the critical behavior is *universal* in the sense that the behavior does not depend on the exact description of the underlying complex network, but instead depends only on macroscopic properties such as the degree distribution. In this article, we will discuss recent developments in the mathematical literature in the context of an elementary yet important process called percolation.

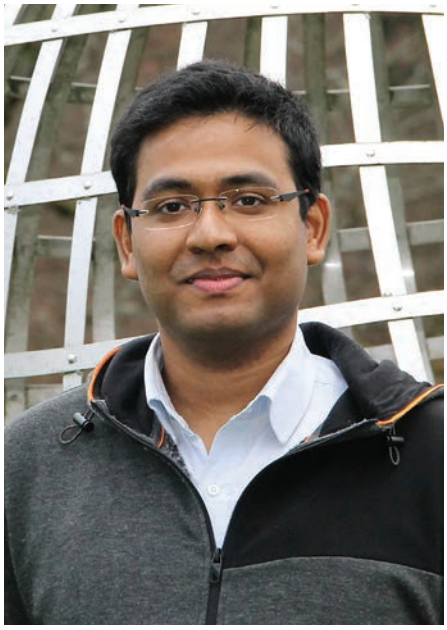

Photo: Mathematisches Forschungsinstitut Oberwolfach / Petra Lein

Souvik Dhara

Percolation phase transition

Consider a finite network G_n on n vertices, where each edge e is associated with an independent Uniform$[0,1]$ edge-weight U_e. The percolation process refers to the graph-valued stochastic process $(G_n(p))_{p\in[0,1]}$, where, at time p, we only keep edges satisfying $U_e \le p$. The phase transition can be informally understood as follows: As p increases, and more edges keep getting added to the graph, there is a point when adding a just small fraction of edges (i.e., increasing p only slightly) suddenly makes a macroscopic connected component emerge. To define the phase transition more formally, let $C_{(k)}(p)$ denote the k-th largest connected component of $G_n(p)$. We say that *phase transition*} occurs around the *critical value* $p_c = p_c(n)$, if for any $\varepsilon, \delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|C_{(1)}(p)| > \delta n) = \begin{cases} 0 \text{ for } p < p_c(1-\varepsilon) & \text{(subcritical phase)}, \\ 1 \text{ for } p > p_c(1+\varepsilon) & \text{(supercritical phase)}. \end{cases} \tag{1}$$

The critical regime lies on the boundary between the subcritical and supercritical phases, where the system exhibits an intermediate behavior. To observe the critical behavior, one must take $p = p_c(1 \mp \varepsilon_n)$ in (1), for some $\varepsilon_n \to 0$ as $n \to \infty$. Interestingly, the critical behavior is not observed for any ε_n; there is a range of ε_n where the

graph shows qualitatively similar features as the sub/supercritical phases and the critical behavior is observed only when ε_n is chosen appropriately. In most situations, this means that $\varepsilon_n = \Theta(n^{-\eta})$, where $\eta > 0$ is a constant depending on G_n. To build a more concrete understanding, let us dive deeply into a classical case where $G_n = K_n$, the complete graph.

Critical window on complete graph

Erdős and Rényi [6] were the first to identify the phase transition (1) on the complete graph K_n with $p_c = \frac{1}{n}$ and $\eta = \frac{1}{3}$. An array of subsequent research has given us a more complete understanding of the critical behavior around $p = (1 \mp \varepsilon_n)p_c$, which we now describe. The proofs of these results can be found in [8]. For a component C, let $\mathrm{SP}(C)$ denote the number of surplus edges, i.e., the number of edges to be deleted to make C a tree. In a sense, $\mathrm{SP}(C)$ measures the complexity of the structure of C.

Barely subcritical phase: $p = p_c(1 - \varepsilon_n)$ with $\varepsilon_n n^{1/3} \to \infty$. Then for each fixed $i \geq 1$, as $n \to \infty$,

$$\frac{|\,C_{(i)}(p)\,|}{2\varepsilon_n^{-2}\log(n\varepsilon_n^3)} \xrightarrow{\mathbb{P}} 1,$$

and

$$\mathbb{P}(\exists i\colon \mathrm{SP}(C_{(i)}(p)) > 1) \to 0.$$

Thus, even if $p \approx p_c$, $K_n(p)$ shows the two characteristic features of the subcritical regime in this phase: first, the largest component is not distinctively larger than the second-largest component, and second, $K_n(p)$ is essentially a collection of trees.

Barely supercritical phase: $p = p_c(1 + \varepsilon_n)$ *with* $\varepsilon_n n^{1/3} \to \infty$. *Then, as* $n \to \infty$,

$$\frac{|\,C_{(1)}(p)\,|}{2n\varepsilon_n} \xrightarrow{\mathbb{P}} 1, \quad \mathrm{SP}(C_{(1)}(p)) \xrightarrow{\mathbb{P}} \infty,$$
$$\frac{|\,C_{(2)}(p)\,|}{|\,C_{(1)}(p)\,|} \xrightarrow{\mathbb{P}} 0.$$

Thus, in this regime, a unique giant component is born which outnumber all other components, and the giant is complex in the sense that there is a growing number of surplus edges.

Critical window: The phase transition takes place between the barely subcritical and supercritical regimes when $\varepsilon_n \sim n^{-\frac{1}{3}}$. This regime is known as the *critical window* for the phase transition. More precisely, the critical window is defined to be the values of p given by

$$p_c(\lambda) = p_c(1 + \lambda n^{-\eta}), \quad -\infty < \lambda < \infty, \quad (2)$$

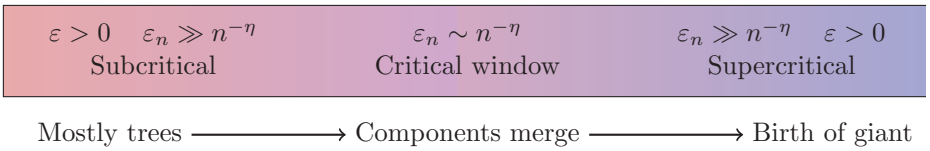

Figure 1

where $\eta = \frac{1}{3}$ for $G_n = K_n$. In this regime, the pioneering works of Janson, Knuth, Luczak and Pittel [7] and Aldous [2] identified some unique and fascinating features (described in more detail later). In fact, it was shown that $n^{-2/3}(|C_{(i)}(p)|, \mathrm{SP}(C_{(i)}(p)))_{i\geq 1}$ converges to a non-degenerate random vector, in contrast with the sub/supercritical regimes described above. Non-degenerate scaling limits is the characteristic feature of the critical window. Further, for any $i \geq 1$, $\mathrm{SP}(C_{(i)}(p))$ is of constant order, so that the surplus-edge count for large components starts to grow in the critical window. The above two properties hold for all values of λ in (2); in this sense, there is not a single critical value, but a whole 'window' of critical values over which the phase transition happens.

Paul Erdős described the percolation process as the race between the components to become the giant. The mental picture is that the collection of trees in the barely subcritical regime are the participants of this race and the component that outnumbers the other components in terms of the number of vertices wins the race. As the percolation parameter transitions through the critical window with λ increasing, components merge with each other and grow in size and complexity. Due to this merging dynamics, the largest component can consist of completely disjoint sets of vertices at two different times $\lambda_1 \neq \lambda_2$, and the race is on. However, at the end of the critical window, when λ becomes sufficiently large, the leader $C_{(1)}(p_c(\lambda))$ stops changing and this leader becomes the young giant component at the end of the critical window. At the barely supercritical phase, the race ends and the largest component stays the largest throughout the future of the percolation process. See Figure 1.

Scaling limits of critical components

We now describe the key scaling limit results about the percolation process in the critical window.

(1) Evolution of component functionals. For each fixed $-\infty < \lambda < \infty$, consider $p_c(\lambda)$ defined in (2), and define

$$\mathbf{Z}_n(\lambda) := (n^{-\rho}\,|C_{(i)}(p_c(\lambda))|, \mathrm{SP}(C_{(i)}(p_c(\lambda))))_{i\geq 1}.$$

Then $(\mathbf{Z}_n(\lambda))_{-\infty<\lambda<\infty}$ can be viewed as a stochastic process as the percolation parameter transitions through the critical window, i.e. λ increases from $-\infty$ to $+\infty$. In the context of the 'race to become a giant', $(\mathbf{Z}_n(\lambda))_{-\infty<\lambda<\infty}$ describes the movie of this race. Therefore it is desirable to study the limit of $(\mathbf{Z}_n(\lambda))_{-\infty<\lambda<\infty}$.

For the complete graph, $\rho = \frac{2}{3}$, and Aldous [2] showed that $(\mathbf{Z}_n(\lambda))_{-\infty<\lambda<\infty}$ converges to a process called an augmented multiplicative coalescent. To intuitively understand this evolution, note that, after increasing λ slightly, a new edge might appear in the graph, and due to the homogeneity in the connectivity structure of K_n, this edge selects two end-points uniformly at random. For this reason, two components $C_{(i)}(p_c(\lambda))$ and $C_{(j)}(p_c(\lambda))$ merge if the end-points are selected from $C_{(i)}(p_c(\lambda))$ and $C_{(j)}(p_c(\lambda))$ respectively, which occurs at rate proportional to $|C_{(i)}(p_c(\lambda))| \times |C_{(j)}(p_c(\lambda))|$ and create a component of size $|C_{(i)}(p_c(\lambda))| + |C_{(j)}(p_c(\lambda))|$. This merging dynamics of a collection of particles according to the product of their sizes is known as the *multiplicative coalescent*. Moreover, following the same logic, a surplus edge is created in $C_{(i)}(p_c(\lambda))$ at rate proportional to $|C_{(i)}(p_c(\lambda))|^2$. The creation of surplus edges can also be augmented in the evolution of the component sizes (the augmented version was later observed in [3]).

(2) Global metric structure. One can also ask, not only about component functionals but about the *global structure* of these critical components. Of course, the term *global structure* is a bit vague; however, this can be formalized. Each component C can be viewed as a metric space, equipped with a measure on the associated Borel sigma-algebra. The metric on C is the graph-distance where (i) each edge has length one, (ii) the measure being proportional to the counting measure, i.e., for any $A \subset C$, the measure of A is given by $\mu_{ct}(A) = |\,A\,|/|\,C\,|$. Then, $C_{(i)}(p_c(\lambda))$ can be viewed as a ran-

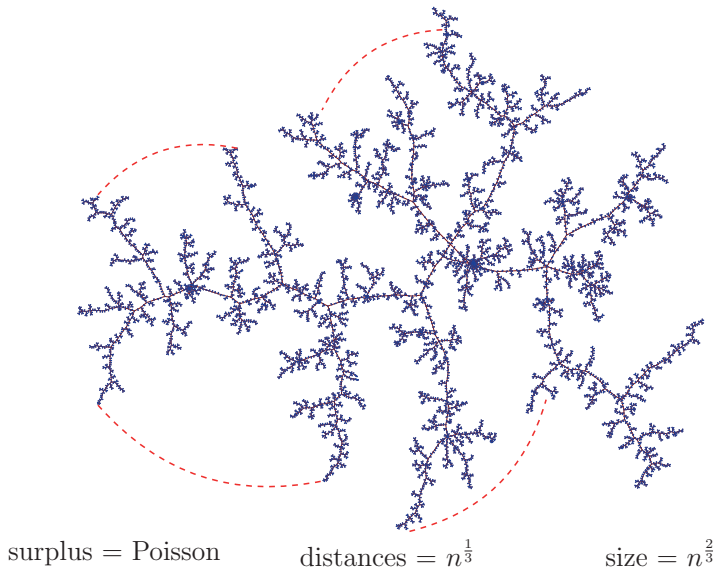

Figure 2 A visualization of the global structure of largest components in the critical window for $G_n = K_n$.

dom element from $\mathcal{M}$, the space of metric spaces with an associated probability measure. For $M = (M, d, \mu) \in \mathcal{M}$ and $a > 0$, define aM to be the measured metric space (M, ad, μ). Then the goal is to

find the distributional limit of
$\big((n^{-\delta} C_{(i)}(p_c(\lambda))\big)_{i \geq 1}$.

The limit is usually a random metric space that is compact with probability 1. Since the limit is obtained after rescaling of graph-distances by n^{δ}, the distances in $C_{(i)}(p_c(\lambda))$ scale as n^{δ}. The above quantity is an $\mathcal{M}$-valued sequence. Of course, the topology on $\mathcal{M}$ is important, and one can consider the Gromov weak-topology, or stronger the Gromov–Hausdorff–Prokhov topology.

Addario-Berry, Broutin and Goldschimdt [1] proved such a scaling limit result for $G_n = K_n$ with $\delta = \frac{1}{3}$. The description of this beautiful limiting random metric space is out of scope here, and the reader is referred to Figure 2 for a visualization of this metric space. (This is not an exact simulation. The picture is taken from Igor Kortchemski's website.) The visualization shows that the global structure of $C_{(i)}(p_c(\lambda))$ looks mostly like a tree with $n^{2/3}$ vertices with distances scaling like $n^{1/3}$, and the red lines indicate the surplus edges.

Effect of degree-inhomogeneity

Two major universality classes

While studying phase transition on complex networks, one is interested in the universality of the critical behavior with respect to the behavior on complete graphs. This has motivated extensive literature in physics and has led to a wide array of conjectures and heuristic deductions of the associated critical exponents. In a nutshell, these conjectures can be described as follows:

The intrinsic nature of the critical behavior does not depend on the exact description of the model, but only on moment conditions on the degree distribution. There are two major universality classes corresponding to the critical regime and the nature of the emergence of the giant depends on whether the degree distribution has an asymptotically finite third moment or infinite third moment.

In [5], our objective was to understand the notion of universality of this critical behavior. More precisely, suppose that we are given a degree sequence $\mathbf{d}^n = (d_i^n)_{i \in [n]}$ for each $n \geq 1$, and G_n is uniformly chosen from all graphs with degree sequence $\mathbf{d}^n$. Instead of stating the moment conditions, let us suppose that the asymptotic empirical degree distribution follows a power-law with exponent τ.

Erdős–Rényi universality class. For $\tau > 4$, the asymptotic empirical degree distribution has a finite third moment. In this case, the critical window turns out to be $p = p_c(1 + \lambda n^{-1/3})$, the maximal component sizes $|C_{(i)}(p_c(\lambda))|$, for any fixed i, are of the order $n^{2/3}$ in the critical regime, whilst typical distances in these maximal connected components scale like $n^{1/3}$. Thus $\rho = \frac{2}{3}$ and $\eta = \delta = \frac{1}{3}$. Moreover, the evolution of $(\mathbf{Z}_n(\lambda))_{-\infty < \lambda < \infty}$ over the critical window $(\mathbf{Z}_n(\lambda))_{-\infty < \lambda < \infty}$ turns out to be same as the complete graph. Thus, the critical behavior is indistinguishable compared to that in the homogeneous instance of the complete graph.

Heavy-tailed behavior. For $\tau \in (3, 4)$, the asymptotic degree distribution has an infinite third moment, but a finite second moment. Here the critical window turns out to be $p = p_c(1 + \lambda n^{-(\tau-3)/(\tau-1)})$, $|C_{(i)}(p_c(\lambda))|$ is of the order $n^{(\tau-2)/(\tau-1)}$ for each fixed $i \geq 1$, whilst distances scale like $n^{(\tau-3)/(\tau-1)}$. Thus $\rho = (\tau-2)/(\tau-1)$ and $\eta = \delta = (\tau-3)/(\tau-1)$. The scaling limits for the component sizes and its metric structure turn out to be fundamentally different. Figure 3 is a visualization of the metric structure of these critical components. (This is not an exact simulation. The picture is taken from Igor Kortchemski's website.) As we can see in this case, the extremal-degree vertices play a crucial role in the connectivity pattern of the critical components, and if we remove these extremal degree vertices, then the critical components fall apart.

Why is finite third-moment condition important?

To understand the relevance of the finite third-moment condition, let us look at the local structures of G_n, the uniform graph with given degrees. For such random graphs, neighborhoods of most vertices can be *approximated* by neighborhoods of the root of an infinite random rooted tree. The degree of a neighbor of a vertex is approximately given by the *size-biased distribution*

$$\mathbb{P}(D_n^* = k) = \frac{k \times |\{i : d_i^n = k\}|}{\sum_i d_i^n}.$$

Therefore, the finite neighborhoods can be coupled with a branching process with progeny $D_n^* - 1$. The expectation is

$$\nu_n := \frac{\sum_i d_i^n (d_i^n - 1)}{\sum_i d_i^n}.$$

When the percolation probability is $p \approx \frac{1}{\nu_n}$, then the local neighborhood structure looks like a critical branching process, and therefore $p_c = \frac{1}{\nu_n}$. Notice now that the variance of this branching process depends on the third moment of the degree distribution. This is the reason why we observe

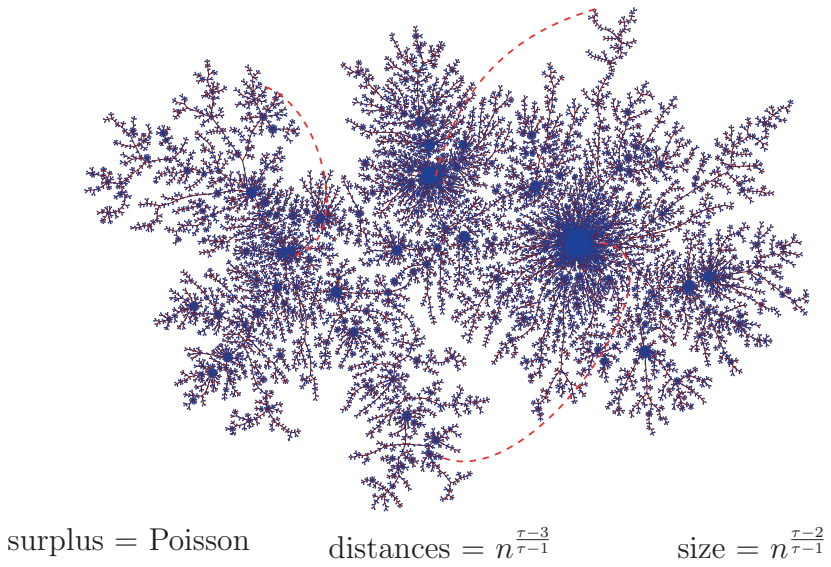

Figure 3 A visualization of the global structure of largest components in the critical window for $\tau \in (3,4)$.

universality of these local neighborhood structures compared to the homogeneous model as long as the finite third-moment condition holds. On the other hand, if this condition fails, then vertices of extremal degree start appearing leading to structures like Figure 3.

A new universality class

Both the above universality classes assume a finite second-moment condition on the degree distribution. This condition fails if the power-law exponent of the degree distribution satisfies $\tau \in (2,3)$. Such networks are often called *scale-free* networks in the literature and are of particular interest due to their ubiquitous appearance in the World Wide Web, social networks, protein interaction networks. One of the well-known features of scale-free networks is that these networks are *robust* under random edge-deletion, i.e., for any sequence $(p_n)_{n\geq 1}$ with $\liminf_{n\to\infty} p_n > 0$, the graph obtained by applying percolation with probability p_n remains supercritical. Thus, in order to observe the percolation critical behavior, one needs to take $p_c \sim n^{-\eta}$ for some $\eta > 0$.

In this case, working with uniform graphs becomes challenging, since well-known algorithms to generate them do not work. Instead, let us consider a simpler model with independent edge connections. Suppose each vertex is equipped with weight w_i (think of it as the average degree), and the empirical distribution of w_i's is asymptotically a power-law with exponent $\tau \in (2,3)$. The graph G_n is constructed by keeping an edge between i and j with probability

$$p_{ij} := 1 - e^{-w_i w_j / \Sigma_i w_i}.$$

An investigation [4] about the critical behavior for this model shows that the critical window is given by

$$p_c(\lambda) = \lambda n^{-\frac{3-\tau}{2}} \quad \text{for } \lambda \in (0, \lambda_c),$$

for some explicitly computable λ_c. The key observation here is that the critical window is of finite length, i.e., $\lambda \in (0,\lambda_c)$, which is in stark contrast with the other behaviors above. Thus the race between components suddenly ends at a finite location of the critical window, and a tiny giant component emerges at $\lambda = \lambda_c$. This observation is rather surprising and had not been predicted by the extensive literature in the substantial physics community.

We have just started exploring the critical behavior on scale-free networks, with many interesting questions yet to be studied. For example, recalling the analogy of race between the components, currently, we know the scaling limits of the rescaled component sizes at each fixed location of this race, i.e., for each fixed $\lambda \in (0,\lambda_c)$. Can we understand the whole movie of this race as λ increases? The evolution of components is clearly quite different from the multiplicative coalescent, since all the large components merge with each other within a finite time. Also, can we understand the global structure of these critical components? The study of such questions will unravel completely new types of scaling objects, which makes the critical behavior of scale-free networks an interesting avenue of future research for all of us.

Acknowledgement
The author sincerely thanks his advisors Remco van der Hofstad and Johan van Leeuwaarden for their guidance in this project.

References

1 Louigi Addario-Berry, Nicolas Broutin and Christina Goldschmidt, The continuum limit of critical random graphs, *Probab. Theory Relat. Fields* 152(3) (2012), 367–406.

2 David Aldous, Brownian excursions, critical random graphs and the multiplicative coalescent, *Ann. Probab.* 25(2) (1997), 812–854.

3 Shankar Bhamidi, Amarjit Budhiraja and Xuan Wang, The augmented multiplicative coalescent and critical dynamic random graph models, *Probab. Theory Relat. Fields* 160(3) (2014), 733–796.

4 Shankar Bhamidi, Souvik Dhara and Remco van der Hofstad, Multiscale genesis of the tiny giant for percolation on scale-free random graphs (2021), in preparation.

5 Souvik Dhara, *Critical Percolation on Random Networks with Prescribed Degrees*, PhD Thesis, Eindhoven University of Technology, 2018, arXiv:1809.03634.

6 P. Erdős and A. Rényi, On the Evolution of Random Graphs, *Publ. Math. Inst. Hung. Acad. Sci.* 5 (1960), 17–61.

7 Svante Janson, Donald E Knuth, Tomasz Łuczak and Boris Pittel, The birth of the giant component, *Random Struct. Algor.* 4(3) (1993), 233–358.

8 Svante Janson, Tomasz Łuczak and Andrzej Rucinski. *Random Graphs*. Wiley, 2000.