Fault Tolerance of Random Graphs with respect to Connectivity: Phase Transition in Logarithmic Average Degree

Satoshi Takabe, Takafumi Nakano, and Tadashi Wadayama
Department of Computer Science, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555, Japan
(Dated: July 6, 2018)

The fault tolerance of random graphs with unbounded degrees with respect to connectivity is investigated. It is related to the reliability of wireless sensor networks with unreliable relay nodes. The model evaluates the network breakdown probability that a graph is disconnected after stochastic node removal. To establish a mean-field approximation for the model, the cavity method for finite systems is proposed. Then the asymptotic analysis is applied. As a result, the former enables us to obtain an approximation formula for any number of nodes and an arbitrary degree distribution. In addition, the latter reveals that the phase transition occurs on random graphs with logarithmic average degrees. Those results, which are supported by numerical simulations, coincide with the mathematical results, indicating successful predictions by mean-field approximation for unbounded but not dense random graphs.

Introduction. Connectivity is a simple but fundamental property of graphs, as apparent from the fact that the first work on Erdős–Rényi (ER) random graphs examined the connectivity probability and its critical phenomenon [1]. Connectivity also has been investigated as a measure of the reliability of wireless communication networks first introduced by Moore and Shannon [2]. In the model, one considers the connectivity of networks after communication links, edges in a graph, are removed stochastically. Combined with the notion of network graphs, studies of reliability defined by connectivity [1] have provided theoretical insight into various technologies such as software fault tolerance [5] and the random key predistribution scheme [6].

After the emergence of network science dealing with complex networks [7, 8], a fraction of a giant component (GC) has also been studied as a measure of network reliability against accidental fault of (or vulnerable attacks to) nodes or edges [9–11]. This model has been applied to real networks such as power grids [12, 13]. Theoretically, the mean-field approximation in statistical mechanics contributes to those studies, revealing that network topology strongly affects fault-tolerance properties, e.g., a percolation threshold.

From the viewpoint of statistical mechanics, some mean-field approximations for random graphs depend on the order of average degrees. For dense random graphs whose average degree is proportional to the number of nodes, the replica method, which is known as a mean-field approximation, is often used, contributing to successful establishment of the physical pictures of the spin-glass theory [14]. Additionally, it has been applied to various fields such as neural networks [14–16], wireless communications [17], and inference problems including compressed sensing [18, 19].

Numerous interesting systems are mapped to spin systems on sparse graphs with bounded, i.e., $O(1)$, degree: error-correcting codes [20, 21], optimization problems [22, 23], community detection [24, 25], and so on. As mean-field approximations for sparse random graphs, the cavity method [26, 27], and the generating function method [28, 29] have been used along with the Monasson’s replica method [30]. These methods provide good evaluations of properties and physical pictures of those systems (e.g., [31, 32]). Moreover, they offer conjectures which are sometimes generalized to mathematically rigorous results. For example, percolation thresholds in the node fault model were predicted using mean-field analysis [11], as proved generally later [33].

Considering the intermediate regime, where the average degree is scaled as $O(\ln n)$, the mean-field analysis is a challenging task. Because the average degree is unbounded but not dense, the mean-field approximations for sparse or dense random graphs are naively inapplicable. In addition, the absence of the analysis might result from the fact that almost all known models exhibit phase transitions related to either sparse or dense graphs. Recently, however, probabilistic analyses have shown that some phase transitions in compressed sensing [37, 38] and community detection [39, 40] occur in the intermediate regime. Providing an example of mean-field analyses for disordered systems in the regime is expected to be useful not only to extend the scope of applications of the approximation but also to predict mathematical statements.

The fault tolerance of random graphs with respect to connectivity has attracted a great deal of interest along with development of IoT technologies [42]. In the model, disconnectivity probability of random graphs after stochastic node removals is evaluated, which describes reliability of the wireless sensor network with unreliable relay nodes [43, 44]. Nodes of this kind simplify, e.g., the energy harvesting relay node [45, 46], which accidentally fails activation. As for critical phenomena, the threshold is not estimated, although it is expected to lie in the $O(\ln n)$ regime [47], which suggests that the model is a
good example to examine the phase transition in the intermediate regime.

The motivation of this Letter is to establish a mean-field analysis for the node fault model of random graphs with logarithmic average degrees and investigate its phase transition. In the analysis, we first obtain a self-consistent cavity equation for disordered finite systems. It enables us to obtain the approximation formula with finite $n$, which coincides asymptotically with the numerical results. The asymptotic analysis then reveals that the model exhibits a discontinuous phase transition with a threshold lying in the $O(\ln n)$ regime. Finally, we discuss the validity of the mean-field analysis and some future avenues of research.

Model Definition Let $G = (V, E)$ be an undirected simple graph, i.e., a graph without self-loops and multiple edges, where $V = \{1, \ldots, N\}$ and $E \subset V^2$ respectively represent a vertex set and an edge set. Connectivity is defined as a graph property by which there exists a path between any pair of nodes in the graph. We consider the node fault model [13], also known as the random node breakdown [11], in which each node is removed independently with constant probability $\epsilon$, called the node breakdown probability. The resultant graph is named a survival graph. The network breakdown probability $P^{\text{node}}(G)$ is defined as the probability that a survival graph is disconnected. Our goal is to evaluate the network breakdown probability averaged over random graphs with unbounded degrees and to investigate its phase transition.

A random graph ensemble $\Omega_n$ is defined as a probabilistic space $(G^n, 2^{G^n}, Q)$ where $G^n$ is a set of simple graphs with $n$ nodes, $2^{G^n}$ represents the power set of $G^n$, and $Q$ is a probability measure of $2^{G^n}$. As a representative model, we restrict ourselves to the case in which $\Omega_n$ is characterized solely by the number of nodes $n$ and degree distribution $p_n(k)$ ($0 \leq k \leq n - 1$) [30], which indicates that graphs sampled from $\Omega_n$ have no correlation such as degree–degree correlation [50], on average. The network breakdown probability $P_{\Omega_n}(\epsilon)$ is then defined by the average of probability $P^{\text{node}}(G)$ over a random graph ensemble $\Omega_n$. Similar to the network breakdown probability, let us denote the average fraction of GC by $\rho_{\Omega_n}$. It is noteworthy that the phase transition of $\rho_{\Omega_n}$ in the node fault model is known as network resilience [11].

Mean-field Analysis Because connectivity is realized when all nodes in survival graphs belong to the GC, we first evaluate the fraction of the GC $\rho_{\Omega_n}$. The mean-field analysis of $\rho_{\Omega_n}$ has been investigated in [51] using the generating function method, in [52–54] using the cavity method, and in [55] using nonbacktracking expansions. Herein, we briefly describe mean-field analysis based on the cavity method.

We introduce a classical spin model on a given graph $G = (V, E)$ as in [53]. We define a binary variable $s_i$ for the $i$-th node by $s_i = 1$ if the node is active and $s_i = 0$ otherwise. The other binary variable $\sigma_i$ is introduced to describe whether the $i$-th node is contained to the GC: $\sigma_i = 1$ if the node does not belong to the GC, and $\sigma_i = 0$ otherwise. After some nodes have broken down, i.e., the sequence $\{s_i\}$ is fixed, a constraint for the $i$-th node to belong to the GC is given as

$$\sigma_i = \prod_{j \in \partial_i} (1 - s_j + s_j \sigma_j),$$  \hspace{1cm} (1)

where $\partial_i$ represents a set of neighbors of the $i$-th node. To find the giant component satisfying Eq. (1), we use the sum-product algorithm. Given that $G$ is a tree, $\sigma_i$ can be obtained exactly as

$$\sigma_i = \prod_{j \in \partial_i} (1 - s_j + s_j m_{j \rightarrow i}),$$  \hspace{1cm} (2)

where $m_{j \rightarrow i}$ is a message from the $j$-th node to the $i$-th node, which takes a value of one if $\sigma_j = 1$ holds on the cavity graph $G \backslash i$, and zero otherwise. These messages satisfy cavity equations for the GC [53], expressed as

$$m_{i \rightarrow j} = \prod_{k \in \partial(i \cup j)} (1 - s_k + s_k m_{k \rightarrow i}).$$  \hspace{1cm} (3)

Although the sum-product algorithm is no longer exact for graphs with cycles, it is regarded as a kind of Bethe–Peierls approximation [56, 57].

To calculate the random graph average, the replica symmetric (RS) cavity method is applied to the system. Although earlier works directly take the large-$n$ limit, as with the case of the conventional RS cavity method, the method for random graphs with finite $n$ is proposed here to consider random graphs with unbounded average degrees.

Considering that the messages $\{m_{i \rightarrow j}\}$ are updated recursively by the cavity equations [3], let $I_n^{(t)}$ be the probability that the message $m_{i \rightarrow j}$ takes one for randomly chosen graph $G$ and its edge $(i, j)$ from $\Omega_n$ at the $t$-th iteration step. If the correlation between any pair of spins is negligible, then $I_n^{(t)}$ satisfies the following relations.

$$I_n^{(t+1)} = \sum_{k=1}^{n} k p_n(k) \langle k \rangle (\epsilon + (1 - \epsilon) I_n^{(t)})^{k-1},$$  \hspace{1cm} (4)

where $\langle k \rangle \equiv \sum_{k=1}^{n} k p_n(k)$ represents the average degree of random graphs. A fixed-point solution $I_n$ is obtained in the large-$t$ limit, which is given as

$$I_n = \sum_{k=1}^{n} k p_n(k) \langle k \rangle (\epsilon + (1 - \epsilon) I_n)^{k-1}.$$  \hspace{1cm} (5)

The average fraction of the GC under the mean-field approximation is then obtained as

$$\rho_{\Omega_n}^{\text{MF}} = \sum_{k=1}^{n} p_n(k) (\epsilon + (1 - \epsilon) I_n)^k.$$  \hspace{1cm} (6)
The breakdown probability $P_{\Omega_n}(\epsilon)$ is then approximated using the evaluation presented above. As a graph is connected when each active node belongs to the GC of the survival graph, the approximation formula of the network breakdown probability is given as

$$P_{\Omega_n}^{\text{MF}}(\epsilon) = 1 - \left(1 - (1 - \epsilon) \sum_{k=0}^{\infty} p_n(k)[\epsilon + (1 - \epsilon)I_n]^k\right)^n.$$  

Equation (7) has a trivial solution $I_n = 1$ corresponding to a “broken phase” in which $P_{\Omega_n}^{\text{MF}} = 0$ holds. It also possibly has another solution $\tilde{I}_n$ in $[0, 1)$. Because $\epsilon$ is assumed to be constant, $P_{\Omega_n}^{\text{MF}}(\epsilon)$ converges to one if $\tilde{I}_n$ remains positive in the large-$n$ limit. We therefore examine the case in which the average degree and resultant $I_n$ depend on $n$ using the RS cavity method for finite-size systems.

**Asymptotic Analysis.** To investigate the phase transition of connectivity, here we execute an asymptotic analysis of the approximation formula. Unfortunately, it is difficult to obtain an explicit form of the critical threshold in general. Alternatively, we describe the analytical results for some well-known random graph ensembles.

(i) **ER random graphs.** The ER random graph with $n$ nodes is characterized by the binomial degree distribution $p_n(k) = \binom{n-1}{k}(1-p)^{n-k-1}p^k$, where $p \equiv c/(n-1)$ and $c$ is the average degree. Then Eq. (5) reads as

$$I_n = 1 - p(1 - \epsilon)(1 - I_n)^{n-2}.$$  

(8)

The approximate network breakdown probability $P_{n,c}^{\text{MF}}(\epsilon)$ is given as

$$P_{n,c}^{\text{MF}}(\epsilon) = 1 - \left[1 - (1 - \epsilon)\left(1 - p(1 - \epsilon)(1 - I_n)^{n-1}\right)\right]^n.$$  

(9)

If we set $c = d \ln n$, then the asymptotic analysis of Eq. (9) shows that the leading term of $I_n$ is given as $n^{-d/(1 - \epsilon)}$. This asymptotic form is then substituted to $I_n$ in Eq. (9). Using the fact that, in the large-$n$ limit, $(1 - an^{-\delta})^n$ converges to zero when $b < 1$, to $e^{-a}$ when $b = 1$ and to one otherwise, the threshold $\epsilon^*$ of node breaking probability is obtained by $\epsilon^* = 1 - 1/d$. In other words, $P_{n,c}^{\text{MF}}(\epsilon)$ jumps from zero to one at the threshold $\epsilon^*$. This result suggests that ER random graphs lose connectivity even when $\epsilon = 0$ if $c < \ln n$ holds, which is proved in (ii).

(ii) **Regular random graphs.** The regular random (RR) graph is characterized by $p_n(k) = \delta_{k,\lambda}$, where $\delta_{n,m}$ represents Kronecker’s delta. Although RR graphs themselves have connectivity with high probability if $\lambda \geq 3$, results show that connectivity is lost by node break downs for any $\epsilon > 0$. As in the case of ER random graphs, we scale the degree as $\lambda = d \ln n$. The asymptotic analyses yield that the leading term of $\tilde{I}_n$ is given as $n^{-d\ln(1/\epsilon)}$. The threshold is given as $\epsilon^* = e^{-1/d}$.

Figure 1 presents numerical results and mean-field evaluations of the network breakdown probability for RR graphs with degree $\lambda = \log_2 n$. The numerical simulation is performed by generating $10^6$ random graphs and by checking the connectivity of $10^5$ survival graphs per random graph using depth-first search. Results show that the mean-field evaluations by Eqs. (5) and (7) are asymptotically close to the numerical results, indicating that the analytical result $\epsilon^* = 1/2$ is correct.

(iii) **Scale-free networks.** We also consider scale-free (SF) networks, with the degree distribution given as a power-law distribution $p_n(k) \propto k^{-\gamma}$. To parameterize the ensemble with $\gamma$ and average degree $\langle k \rangle$, we consider random SF networks with the minimum degree $\tilde{k}$. The degree distribution is therefore written as

$$p_n(k) = C_n^{-1}k^{-\gamma} (\tilde{k} \leq k < n), C_n = \zeta(\gamma, \tilde{k}) - \zeta(\gamma, n),$$  

(10)

where $\zeta(s, a) = \sum_{k=0}^{\infty}(k + a)^{-s}$ is the Hurwitz zeta function. Then, the average degree is given as $\langle k \rangle = C_n^{-1}[\zeta(\gamma - 1, \tilde{k}) - \zeta(\gamma - 1, n)]$, which is well-defined when $\gamma > 2$. As the above cases illustrate, it is apparent that the possible threshold is trivial, i.e., $\epsilon^* = 0$ when $\langle k \rangle = O(1)$. If the average degree is again scaled as $d \ln n$, then the minimum degree also depends on $n$ because $\langle k \rangle = (\gamma - 1)/(\gamma - 2)\tilde{k} + O(1)$ holds. The self-consistent equation of $I_n$, Eq. (3), is expressed as

$$I_n = \frac{1}{C_n(k)} \left[\epsilon + (1 - \epsilon)I_n^{k-1}\Phi\left(\epsilon + (1 - \epsilon)I_n, \gamma - 1, \tilde{k}\right)\right],$$  

(11)

where $\Phi(z, s, a) = \sum_{k=0}^{\infty}z^k/(k + a)^s$ is called the Lerch
Final result is given as $O(n \ln n)$. The cavity method is recognized as the Bethe–Peierls approximation. In the case of sparse random graphs, the use of approximation is justified by the locally tree-like structure of graphs that the length of cycles are typically scaled logarithmically. In this manner, our analysis is justified because the length of cycles is evaluated as $O(n \ln n / \ln \ln n)$. From a mathematical perspective, however, the justification remains an open problem because the known rigorous results only treat random graphs with bounded degrees. In other words, our results provide not only conjectures related to analytical results, but also a challenging open problem in the mean-field theory.

The node fault model presented in this Letter has some possible extensions. The first direction is consideration of intentional attacks to nodes and correlations in graphs. These are analogous to the original node fault model with respect to percolation. Studies of these extensions reveal the network topology dependence on reliability. It is another direction to examine the node fault model on random graphs in metric space. The representative model is known as random geometric graphs, which is often used as an abstract model of wireless ad hoc networks. These extensions enable us to optimize networks while conserving reliability, as in the case of percolation.

Our results suggest that the mean-field approximation in this Letter will be applicable to other systems with critical phenomena, possibly in the intermediate regime. We hope that this work will be helpful in elucidating interesting critical phenomena related to complex networks with logarithmic average degrees and resultant extensions of statistical mechanics for disordered systems.

The authors warmly thank T. Nozaki, T. Hasegawa, K. Hukushima, and Y. Kabashima for fruitful discussions. ST also thank T. Kawamoto for carefully reading the manuscript. This research is partially supported by JSPS KAKENHI Grant Nos. 16H02878, 16K14267 (TW), and 17H06758 (ST).

The result by Erdős and Rényi is specific to the ensemble, the rigor of our predictions on other ensembles is left as an open problem. We emphasize, however, that the correctness of the specific case is strong evidence of the validity of our analyses.

Next we can discuss the validity of the approximation itself. In the view of spin-glass theory, the cavity method is known as the Bethe–Peierls approximation. In the case of sparse random graphs, the use of approximation is justified by the locally tree-like structure of graphs that the length of cycles are typically scaled logarithmically. In this manner, our analysis is justified because the length of cycles is evaluated as $O(n \ln n / \ln \ln n)$. From a mathematical perspective, however, the justification remains an open problem because the known rigorous results only treat random graphs with bounded degrees. In other words, our results provide not only conjectures related to analytical results, but also a challenging open problem in the mean-field theory.
