# Analyzing ER Networks: $P_s$ Estimation, Giant Components, and Phase Transition

**Chenkai Wang**

**Email: wangck2022@mail.sustech.edu.cn**

**Abstract**

This essay investigates the structural properties and critical phenomena in Erdős–Rényi (ER) networks. The distribution $P_s$, representing the probability that a randomly selected node belongs to a component of size $s$, is computed using both simulation and numerical integration methods. The emergence and behavior of the giant component are further examined under Poisson and power-law degree distributions. Additionally, the phase transition phenomenon is analyzed as the connection probability in ER networks increases, highlighting the sudden formation of large-scale connectivity near the critical threshold.

**Keywords: ER Network, $P_S$, Giant Component, Phase Transition**

# Contents

# 1 Description of Homework 4

## Part 1

- Generate an ER network with average degree 0.5 and obtain $P_s$ by simulation.
- Calculate $P_s$ numerically using the formula for $H_0$ and compare the results.

## Part 2

- For a Poisson degree distribution with average degree $\langle k \rangle$, plot the giant component size $S$ as a function of $T$.
- For a power-law degree distribution $p(k) \propto k^{-2.5}$ (with minimum degree 2 and maximum degree $N^{0.5}$), plot the giant component size $S$ as a function of $N$.

# 2 Part I: ER Network

## 2.1 A Brief History from Seven Bridges of Königsberg to ER Network

The representation of problems using graphs can be traced back to Euler's study of the Seven Bridges of Königsberg in 1736. The city of Königsberg, formerly in East Prussia, was divided by a river and connected by seven bridges. The central question concerned the possibility of traversing all bridges exactly once. Euler established both the necessary and sufficient conditions for such a traversal, ultimately providing a negative answer. This foundational work marked the inception of graph theory.

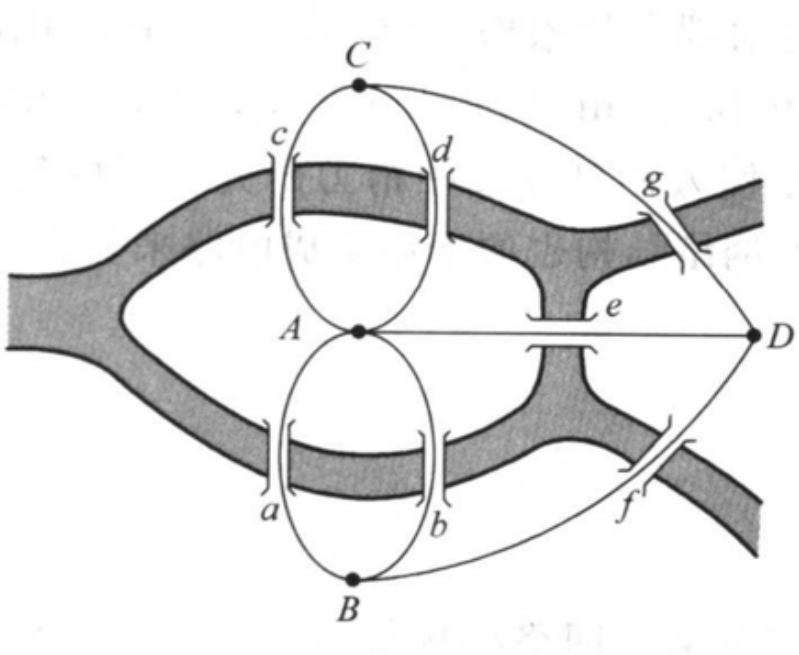


Figure 1: An illustration of the Seven Bridges of Königsberg. The figure shows the historic problem that initiated graph theory.

Graph theory advanced significantly following the publication of the seminal work by Paul Erdős and Alfréd Rényi in the 1960s[3]. The random network model introduced in their research became a fundamental tool for the analysis of real-world networks. During the same period, social experiments such as Milgram's small-world survey[6] and the Bacon game[1] were conducted, revealing structural properties of social networks.

## 2.2 Basic Structure of ER Network

The ER (Erdős–Rényi) network, also referred to as a random network, is typically characterized by two models: $G(n, l)$ and $G(n, p)$.

- The $G(n, l)$ model consists of $n$ nodes connected by $l$ randomly placed edges[3].
- The $G(n, p)$ model consists of $n$ nodes, where each pair of nodes is connected with probability $p$[4].

Figure 2: Example of an ER network with $n = 50$ and $p = 0.1$. The figure illustrates the typical structure of a random network.

The $G(n, p)$ model is adopted in this study due to its analytical tractability and its relevance to real networks, where the number of links is typically not fixed.

For an undirected ER network, the average degree $\langle k \rangle$ is given by

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = p(n-1)$$

As $n \to \infty$, $\langle k \rangle \approx np$. Thus, appropriate values of $n$ and $p$ can be chosen to generate a random network with a desired average degree.

## 2.3 Comparison of Two Different Calculation Methods of $P_S$

## 2.4 Definition of $P_s$

Within the context of network science, $P_s$ represents the probability that a randomly selected node belongs to a connected component (cluster) of size exactly $s$. This metric provides insight into the fragmentation or cohesion of a network under varying conditions. A high value of $P_1$ indicates a prevalence of isolated nodes, while non-negligible values for larger $s$ correspond to the presence of larger connected structures.

### 2.4.1 $P_s$ from Simulation

We generate an ER network from $G(100, \frac{0.5}{99})$ to illustrate the computation of $P_s$. The ER network is generated 1000 times, and the average $P_s$ is obtained as the simulation result, as shown in the first column of Table 1.

Figure 3 displays a sample realization of the network generated during simulation. Connected components are visualized in different colors to indicate their sizes.

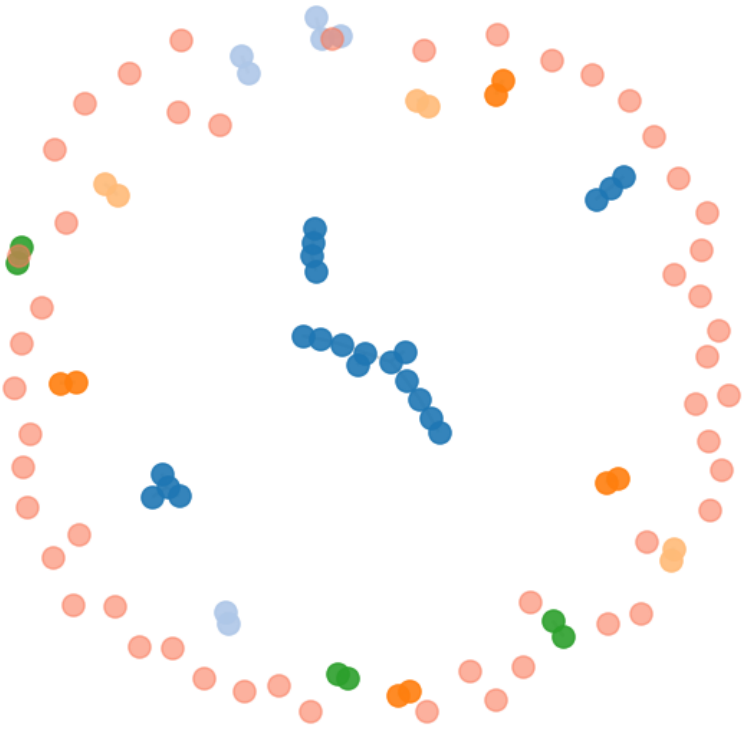

Figure 3: Visualization of a generated ER network during simulation. Connected components are visualized using different colors to indicate their sizes.

### 2.4.2 $P_s$ from Numerical Calculation

According to the slides, $P_s$ can be calculated by the following integral:

$$P_s = \frac{1}{2\pi i} \oint \frac{H_0(z)}{z^{s+1}} \, \mathrm{d}z$$

where $H_0(x)$ is the generating function, introduced below.

By substituting $z = e^{i\theta}$, the integral becomes

$$P_s = \frac{1}{2\pi} \int_0^{2\pi} \frac{H_0\left(\mathrm{e}^{i\theta}\right)}{\mathrm{e}^{i\theta s}} \, \mathrm{d}\theta.$$

Dividing the interval into $N$ equal sub-intervals enables numerical approximation:

$$P_s \approx \frac{1}{2\pi} \sum_{k=0}^{N-1} \frac{H_0\left(\mathrm{e}^{i\frac{2k\pi}{N}}\right)}{\mathrm{e}^{i\frac{2k\pi s}{N}}} \times \frac{2\pi}{N}$$

The following key formulas are used. Let $G_0$ denote the generating function of the network:

$$G_0(x) = \sum_k p_k x_k = (1 - p + px)^{N-1}.$$

Additionally,

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = (1 - p + px)^{N-2}.$$

The iterative relations are

$$\begin{cases} H_1(x) = xG_1(H_1(x)) \\ H_0(x) = xG_0(H_1(x)) \end{cases}$$

These expressions permit computation of $H_0(\cdot)$ and, consequently, $P_s$. The results are presented in Section 2.4.3.

### 2.4.3 Results and Comparison

| Size $s$ | $P_s$ from simulation | $P_s$ from numerical calculation | Error |
|---|---|---|---|
| 1 | 0.60569 | 0.60576 | -0.00007 |
| 2 | 0.18488 | 0.18534 | -0.00046 |
| 3 | 0.08539 | 0.08420 | 0.00119 |
| 4 | 0.04580 | 0.04522 | 0.00057 |
| 5 | 0.02676 | 0.02666 | 0.00010 |
| 6 | 0.01745 | 0.01667 | 0.00078 |
| 7 | 0.01122 | 0.01087 | 0.00035 |
| 8 | 0.00705 | 0.00730 | -0.00025 |
| 9 | 0.00459 | 0.00502 | -0.00043 |
| 10 | 0.00343 | 0.00351 | -0.00008 |
| 11 | 0.00231 | 0.00250 | -0.00019 |
| 12 | 0.00162 | 0.00180 | -0.00018 |
| 13 | 0.00116 | 0.00131 | -0.00015 |
| 14 | 0.00084 | 0.00096 | -0.00012 |
| 15 | 0.00063 | 0.00071 | -0.00008 |
| 16 | 0.00043 | 0.00053 | -0.00009 |
| 17 | 0.00018 | 0.00039 | -0.00021 |
| 18 | 0.00017 | 0.00030 | -0.00013 |
| 19 | 0.00016 | 0.00022 | -0.00006 |
| 20 | 0.00013 | 0.00017 | -0.00004 |

Table 1: The result and comparison table of two calculation methods

From the table, the following observations can be made:

- The results obtained from both simulation and numerical calculation are consistent overall, with discrepancies remaining within a small margin.
- The simulation calculation tends to yield slightly smaller values for $P_s$ compared to the numerical results.

# 3 Part II: Giant Component

## 3.1 Definition of the Giant Component

The giant component is defined as the largest connected component in a graph. Its relative size is denoted by $S(T)$.

By definition,

$$S(T) = 1 - H_0(1, T)$$

To calculate $H_0(1, T)$, the following iterative formulas are used:

$$\begin{cases} G_0(x;T) = G_0(1 + (x-1)T) \\ G_1(x;T) = xG_1(1 + (x-1)T) \\ H_0(x;T) = xG_0(H_1(x;T);T) \end{cases}$$

By setting $x = 1$ and iterating, $H_0(1, T)$ and thus $S(T)$ can be obtained.

Here, $T$ denotes the *transmissibility*, representing the probability that a given edge successfully transmits information or infection in a percolation or epidemic process. This parameter serves as a control variable for the emergence of the giant component and typically lies within the interval $[0, 1]$.

## 3.2 Giant Component Dynamics for Poisson Distribution

Note that $T$ is theoretically a probability and thus defined on the interval $[0, 1]$. However, in the plots below, we extend the range of $T$ beyond 1 for visualization purposes. This helps to better illustrate the saturation behavior of the giant component after the phase transition.

The Poisson degree distribution is given by

$$p_k = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$$

Using the above formulas, the giant component size $S$ can be plotted as a function of $T$. For a Poisson distribution, the critical threshold is $T_c = \frac{1}{\langle k \rangle}$, where $\langle k \rangle = G_0'(1)$ is the average degree. The results are consistent with theoretical predictions.

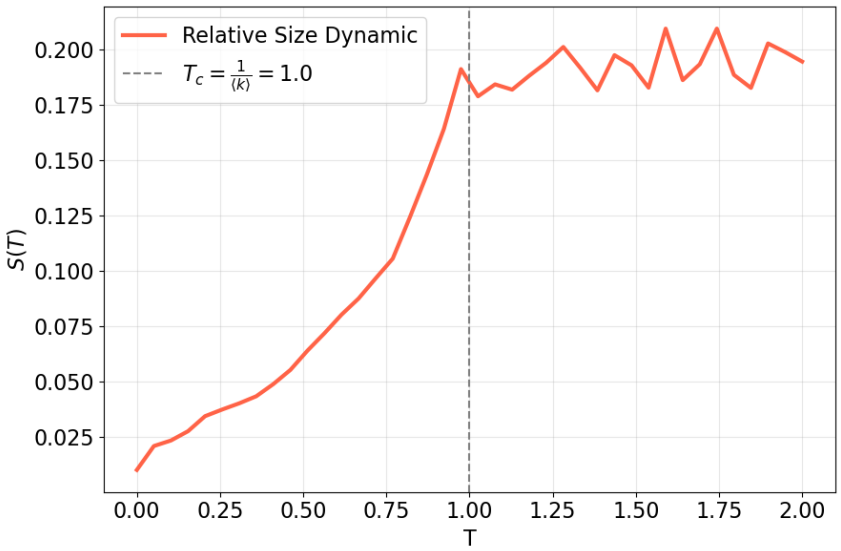


Figure 4: Relative giant component size $S$ as a function of $T$ for a Poisson degree distribution with $\langle k \rangle = 1$ and $N = 1000$. The figure demonstrates the emergence of the giant component near the critical threshold.

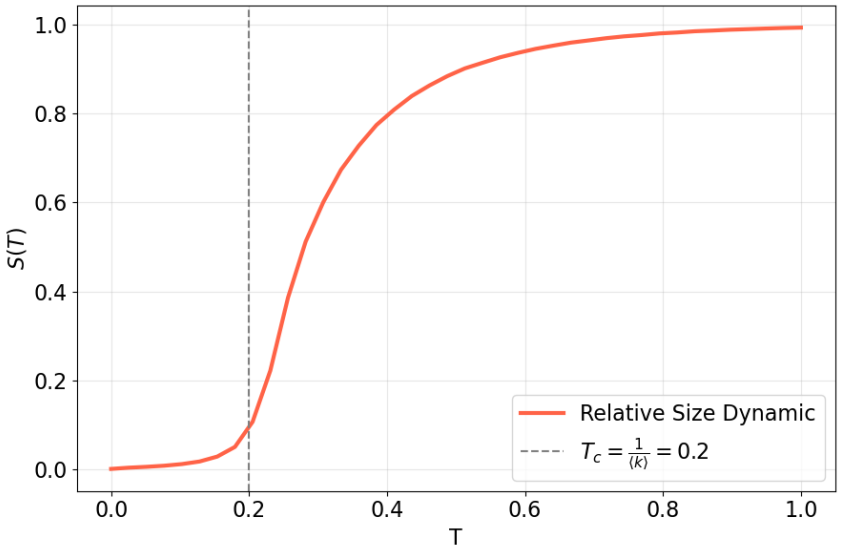


Figure 5: Relative giant component size $S$ as a function of $T$ for a Poisson degree distribution with $\langle k \rangle = 5$ and $N = 1000$. The figure illustrates the rapid growth of the giant component above the threshold.

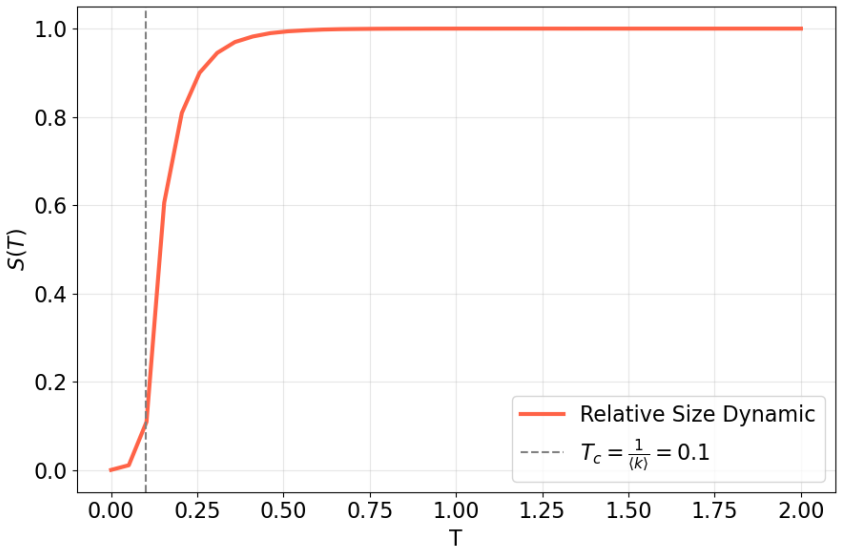


Figure 6: Relative giant component size $S$ as a function of $T$ for a Poisson degree distribution with $\langle k \rangle = 10$ and $N = 1000$. The figure shows the dominance of the giant component at high average degree.

## 3.3 Giant Component Dynamics for Power-Law Degree Distribution

Given the specified constraints, the degree distribution $p_k$ is

$$p_k = \sum_{k=2}^{\sqrt{N}} k^{-2.5}$$

This corresponds to a discrete, truncated power-law distribution. The probability mass function (PMF) is defined as:

$$P(k) = \frac{k^{-\gamma}}{\sum_{j=k_{\min}}^{k_{\max}} j^{-\gamma}}$$

where $\gamma$ is the power-law exponent, and $k_{\min}$ and $k_{\max}$ denote the minimum and maximum degrees, respectively.

The theoretical epidemic threshold $T_c$ is given by

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

with

$$\langle k \rangle = \sum_{k=k_{\min}}^{k_{\max}} k \cdot P(k) = \frac{\sum_{k=k_{\min}}^{k_{\max}} k^{1-\gamma}}{\sum_{k=k_{\min}}^{k_{\max}} k^{-\gamma}}$$

$$\langle k^2 \rangle = \sum_{k=k_{\min}}^{k_{\max}} k^2 \cdot P(k) = \frac{\sum_{k=k_{\min}}^{k_{\max}} k^{2-\gamma}}{\sum_{k=k_{\min}}^{k_{\max}} k^{-\gamma}}$$

The size of the giant component $S$ can be computed using the same iterative procedures as for the Poisson distribution, as described in Section 3.1. The results are shown in Figure 7.

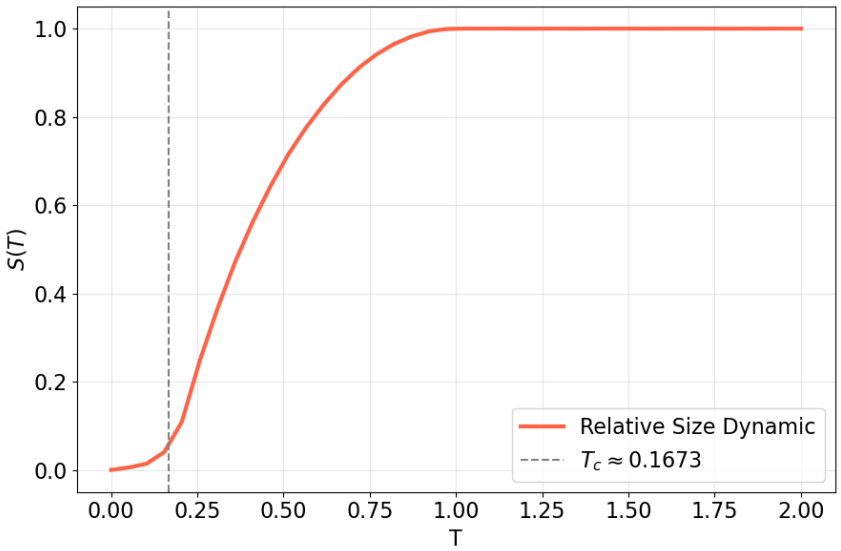


Figure 7: Relative giant component size $S$ as a function of $T$ for a power-law degree distribution with $\gamma = 2.5$ and $N = 1000$. The figure illustrates the behavior of the giant component in networks with heavy-tailed degree distributions.

For this case, $T_c \approx 0.1673$.

# 4 Part III: Additional Analysis of ER Network

In an ER random network, increasing the connection probability $p$ induces a phase transition. When $p$ is below a critical threshold, all connected components remain small. As $p$ surpasses this threshold, a giant component emerges, containing a significant fraction of the nodes. As depicted in Figure 8, the size of the giant component does not increase linearly with $p$, but instead exhibits a pronounced jump near the transition point.

References such as [2] and [5] describe this phenomenon as a phase transition in the ER network as $p$ increases. Specifically[5]:

- For $p = \frac{1-\epsilon}{n}$, where $\epsilon > 0$ is a small constant and $G \sim G(n, p)$, with high probability (w.h.p.), all connected components are of size at most $\frac{7}{\epsilon^2} \ln n$.
- For $p = \frac{1+\epsilon}{n}$, with high probability, $G$ contains a connected component with at least $\frac{\epsilon n}{2}$ nodes.

Here, "w.h.p." denotes "with high probability"[5].

An ER network with $n = 150$ is used to illustrate the sudden emergence of the giant component as $p$ crosses the critical threshold.

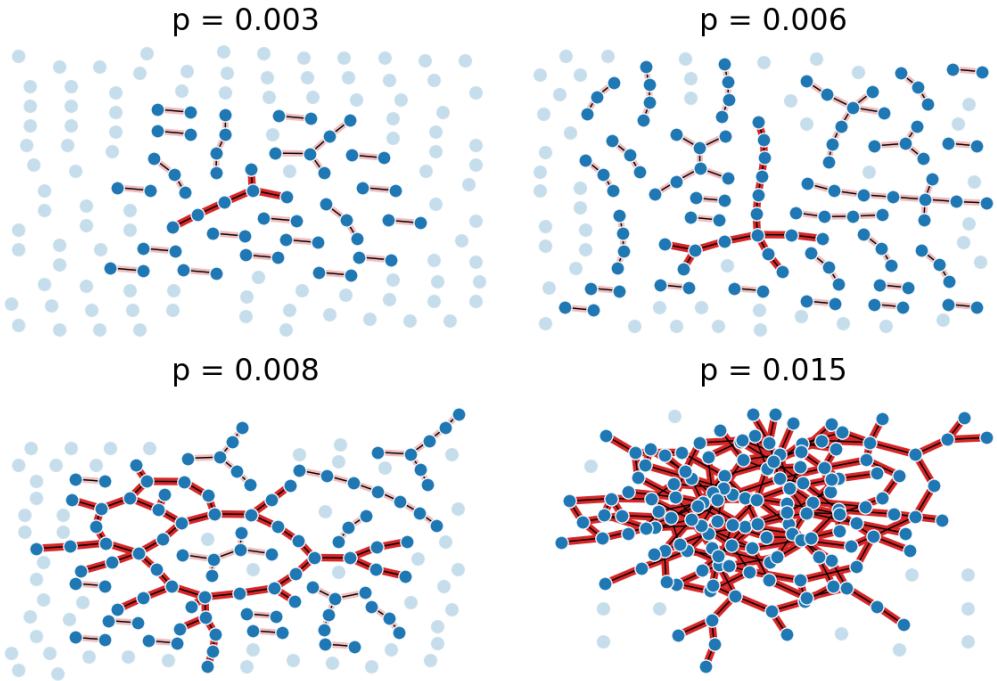


Figure 8: Phase transition in ER network with $n = 150$.

# References

[1] https://oracleofbacon.org/.

[2] http://networksciencebook.com/.

[3] Paul Erdős, Alfréd Rényi, et al. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[4] Edgar N Gilbert. "Random graphs". In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[5] Michael Krivelevich and Benny Sudakov. "The phase transition in random graphs: A simple proof". In: *Random Structures & Algorithms* 43.2 (2013), pp. 131–138.

[6] Stanley Milgram. "The small world problem". In: *Psychology today* 2.1 (1967), pp. 60–67.