

随机图中 k -独立集的相变性质

卢友军 许道云

(贵州大学计算机科学与技术学院 贵阳 550025)
(yjlu111@126.com)

Phase Transition Properties of k -Independent Sets in Random Graphs

Lu Youjun and Xu Daoyun

(College of Computer Science and Technology, Guizhou University, Guiyang 550025)

Abstract Phase transition property is one of the most important properties of the theory of Erdős-Rényi random graphs. A subset of vertices is a k -independent set in a simple undirected graph $G=(V, E)$ if the subset is an independent set containing k vertex. In order to understand the structural property of k -independent sets in Erdős-Rényi random graphs, the phase transition properties of k -independent sets in Erdős-Rényi random graphs are investigated in this paper. It is shown that the threshold probability is $p_c=1-n^{-\frac{2}{k-1}}$ for the existence of k -independent sets in random graph $G(n, p)$ via the first moment method and the second moment method when $2 \leq k=o(\sqrt{n})$. According to this fact that random graph $G(n, p)$ is equivalent to random graph $G(n, m)$ when m is close to pC_n^2 , the threshold edge number is given by $m_c=\left[\frac{n(n-1)}{2}(1-n^{-\frac{2}{k-1}})\right]$ for the existence of k -independent sets in random graph $G(n, m)$. The simulation results show that the consistence between simulation and theoretical threshold value for the existence of k -independent sets in random graph $G(n, p)$ and $G(n, m)$ when $2 \leq k=o(\sqrt{n})$, and the threshold value is related to the total number n of vertices and the number k of vertices of independent set. However, when $k=\omega(\sqrt{n})$, the theoretical threshold value is not consistent with the simulation threshold value for the existence of k -independent sets in random graph $G(n, p)$ and $G(n, m)$.

Key words phase transition property; random graphs; k -independent set; threshold probability; threshold edge number

摘 要 相变性质是 ER(Erdős-Rényi)随机图理论具有的重要性质,一个简单无向图 $G=(V, E)$ 中的 k -独立集是一个具有 k 个顶点的独立集.为更好地理解 ER 随机图中 k -独立集的结构特性,提出并利用一阶矩和二阶矩方法严格证明了当 $2 \leq k=o(\sqrt{n})$ 时随机图 $G(n, p)$ 中 k -独立集出现相变的临界概率 $p_c=1-n^{-\frac{2}{k-1}}$.利用 $m \approx pC_n^2$ 时随机图 $G(n, p)$ 和 $G(n, m)$ 等价的性质给出了随机图 $G(n, m)$ 中 k -独立集出

收稿日期:2016-09-14;修回日期:2016-12-05

基金项目:国家自然科学基金项目(61262006,61462001,61540050,61762019);贵州省重大应用基础研究项目(JZ20142001);贵州大学研究生创新基金项目(2016047)

This work was supported by the National Natural Science Foundation of China (61262006, 61462001, 61540050, 61762019), the Major Applied Basic Research Program of Guizhou Province (JZ20142001), and the Graduate Student Innovation Foundation of Guizhou University (2016047).

通信作者:许道云(dyxu@gzu.edu.cn)

现相变的临界边数 $m_c = \left\lceil \frac{n(n-1)}{2} (1 - n^{-\frac{2}{k-1}}) \right\rceil$. 实验结果表明: 当 $2 \leq k = o(\sqrt{n})$ 时, 随机图 $G(n, p)$ 和 $G(n, m)$ 中存在 k -独立集的理论临界值和仿真得到的临界值一致且临界值与图节点总数 n 和独立集节点数 k 有关, 而当 $k = \omega(\sqrt{n})$ 时, 随机图 $G(n, p)$ 和 $G(n, m)$ 中存在 k -独立集的理论临界值和仿真临界值不一致.

关键词 相变性质; 随机图; k -独立集; 临界概率; 临界边数

中图法分类号 TP301

给定一个简单无向图 $G=(V, E)$, V 表示图 G 的顶点集合, E 表示图 G 的边集合. 一个顶点子集 $T \subseteq V$ 称为图 G 的独立集^[1] (independent set, IS), 是指对任意的顶点 $u, v \in T$ 都有 $(u, v) \notin E$. 若顶点子集 T 的元素个数为 k , 则我们称该顶点子集 T 为 k -独立集. 若独立集 T 不是其他任何一个独立集 S 的真子集, 则我们称该独立集 T 为图 G 的极大独立集. 若独立集 T 是顶点个数最大的独立集, 则我们称该独立集 T 为图 G 的最大独立集^[2] (maximum independent set, MIS). MIS 问题是指在一个给定的简单无向图 G 中找一个顶点个数最大的独立集, 该问题是组合优化领域中的著名 NP-完全问题^[3]. MIS 问题在编码理论^[4]、计算机网络以及计算机视觉等领域都有重要应用.

20 世纪 50 年代末, Erdős 等人^[5]首次引入了经典的随机图模型 $G(n, p)$, 模型中的参数 n 表示随机图的顶点数; p 表示 n 个顶点中任意 2 个不同顶点之间存在边的概率, 通常情况下, 它不是一个固定的数, 而是关于顶点数 n 的函数. 该模型的显著特点是任意 2 个不同顶点之间是否存在边的概率统计独立, 容易采用解析方法对其研究. Erdős 等人^[6]系统地研究了当顶点数 $n \rightarrow \infty$ 时随机图 $G(n, p)$ 的图性质 Q 与连边概率 p 之间的关系: 随机图 $G(n, p)$ 的许多图性质 Q 存在突然涌现现象, 即对任一给定的连边概率 p , 要么几乎每一个图 $G \in G(n, p)$ 都存在图性质 Q , 要么几乎每一个图 $G \in G(n, p)$ 都不存在图性质 Q . 人们把随机图 $G(n, p)$ 中图性质 Q 的这种突然涌现现象称为图性质 Q 的相变现象, 把发生相变的临界点称为相变点. 如文献[7]中分析的随机图 $G(n, p)$ 中连通性性质的相变, 其临界值为 $\frac{\ln n}{n}$; 直径为 2 的性质的相变, 其严格临界值为 $\sqrt{\frac{2 \ln n}{n}}$; 孤立节点消失的性质的相变, 其严格临界为 $\frac{\ln n}{n}$; 环性质的相变, 其临界值为 $\frac{1}{n}$, 以及文献[8]中研究的

随机图着色的相变. 相变现象在其他问题中同样存在, 比如: 随机约束满足性问题中的相变^[9-11] 以及谣言传播的相变^[12-13] 等. 在文献[14]中, Culberson 等人证明了随机图 $G(n, p)$ 中存在控制团的临界概率为 $p_c = \frac{3-\sqrt{5}}{2}$. 2012 年, Nehéz 等人^[15]证明了随机图 $G(n, p)$ 中存在 r -控制团临界值的上下界, 即当概率 $p \leq \frac{3-\sqrt{5}}{2}$ 时, 随机图 $G(n, p)$ 中高概率不存在 r -控制团; 当概率 $p > \frac{1}{2}$ 时, 随机图 $G(n, p)$ 中高概率存在 r -控制团, 其中, $\log_{1/p} n \leq r \leq 2 \log_{1/p} n$.

本文的主要工作是研究随机图 $G(n, p)$ 和随机图 $G(n, m)$ 中 k -独立集出现相变的临界概率和临界边数, 其目的是让我们对随机图 $G(n, p)$ 和随机图 $G(n, m)$ 中 k -独立集的结构有更加深入的理解.

1 基础知识及记号

本节首先给出经典 ER 随机图模型的定义^[16], 然后给出基本概念以及与 ER 随机图有关的一个重要结果.

随机图模型 $G(n, p)$: 给定 n 个顶点, 任意 2 个不同顶点之间以概率 p 连接 1 条边所形成的随机图. 生成一个顶点数为 n 、边数为 m 的图 $G_0 \in G(n, p)$ 的概率为 $Pr[G_0 \in G(n, p)] = p^m (1-p)^{\frac{n(n-1)}{2}-m}$. 模型中的边数是一个随机变量, 其期望值为 pC_n^2 .

随机图模型 $G(n, m)$: 给定 n 个顶点, 在总共 C_n^2 条可能存在的边中, 随机地选取 m 条边所形成的随机图. 由 n 个顶点和 m 条边构成的随机图实例总共有 C_n^m 种, 全体构成一个概率空间, 每一个网络在其中出现的概率是等可能的, 其中, $N = C_n^2, 0 \leq m \leq N$.

一个图性质 Q 是指当 $G \in Q, H \in G(n, p)$ 且 $G \cong H$ (G 同构于 H) 时有 $H \in Q$, 其中, $G \in Q$ 表示图 G 中有图性质 Q . 若该图性质满足条件: $G \in Q$

并且 $G \subseteq H$ 时蕴含 $H \in Q$, 则称该图性质 Q 是单调递增的.

如果 $F \subseteq G \subseteq H$ 且 F, H 都满足性质 Q , 则 G 也满足性质 Q , 称图 G 的性质 Q 是凸的.

引理 1^[16]. 如果 Q 是凸的且 $p(1-p)C_n^2 \rightarrow \infty$, 则几乎每一个随机图 $G(n, p)$ 都满足 Q 当且仅当对任意固定的 x , 几乎每一个随机图 $G(n, m)$ 都满足 Q , 其中 $m = \lceil pC_n^2 \rceil + x \lceil p(1-p)(C_n^2)^{\frac{1}{2}} \rceil$.

引理 1 的直观理解是当随机图 $G(n, m)$ 中的边数 m 与随机图 $G(n, p)$ 中的概率 p 满足关系 $m \approx pC_n^2$ 时, 二者等价, 也即是说满足随机图 $G(n, p)$ 的性质可以容易推广到随机图 $G(n, m)$ 上. 故本文的研究仅针对随机图 $G(n, p)$ 展开. 为描述随机图 $G(n, p)$ 高概率存在某图性质 Q , 引入临界函数、严格临界函数的概念^[17].

设 Q 是一个图性质, 如果:

$$\lim_{n \rightarrow \infty} Pr[G(n, p) \in Q] = \begin{cases} 1, & p(n) \ll h(n) \\ 0, & p(n) \gg h(n) \end{cases}$$

成立, 则称函数 $h(n)$ 是性质 Q 的一个临界函数. 如果对于任意的 $\epsilon > 0$, 有:

$$\lim_{n \rightarrow \infty} Pr[G(n, p) \in Q] = \begin{cases} 1, & p(n) < (1-\epsilon)h(n) \\ 0, & p(n) > (1+\epsilon)h(n) \end{cases}$$

成立, 则称函数 $h(n)$ 是性质 Q 的一个严格临界函数.

2 k -独立集存在的相变分析

本节的工作是利用一阶矩和二阶矩方法分析随机图 $G(n, p)$ 中存在 k -独立集的临界概率 p_c , 给出随机图 $G(n, m)$ 中存在 k -独立集的临界边数 m_c . 在引理 2 中先给出图 $G \in G(n, p)$ 中 k -独立集个数 X_n^k 的数学期望.

引理 2. 在随机图 $G(n, p)$ 中, 对任意的 $2 \leq k \leq n$, 非负整值随机变量 X_n^k 的数学期望为 $E(X_n^k) = C_n^k(1-p)^{C_k^2}$.

证明. 假设 Γ 是顶点个数为 k 的所有顶点子集构成的集合, 其元素包含在 $V(G)$ 中, 其中, $V(G)$ 表示图 $G \in G(n, p)$ 的顶点集, 其顶点总数为 n . 为描述方便, 用 $\alpha, \beta, \gamma, \dots$ 表示 Γ 中的元素. 记 J_α 是关于随机图 $G(n, p)$ 的示性随机变量.

如果顶点集 α 是图 $G \in G(n, p)$ 的一个 k -独立集, 则 $J_\alpha = 1$; 否则, $J_\alpha = 0$.

从顶点总数为 n 的图 $G \in G(n, p)$ 中选择一个

顶点个数为 k 的顶点集 α 有 C_n^k 种选择方式, 同时选择的顶点集 α 恰好是独立集的概率为 $(1-p)^{C_k^2}$. 故顶点集 α 是 k -独立集的期望为 $E(J_\alpha) = (1-p)^{C_k^2}$.

设 X_n^k 是一个取值为非负整数的随机变量, 表示图 $G \in G(n, p)$ 中顶点个数为 k 的独立集个数, 即 $X_n^k = \sum_{\alpha \in \Gamma} J_\alpha$. 由期望的线性性质可知, 对于给定的 $2 \leq k \leq n$, 随机变量 X_n^k 的数学期望 $E(X_n^k)$ 为

$$E(X_n^k) = \sum_{\alpha \in \Gamma} E(J_\alpha) = C_n^k(1-p)^{C_k^2}.$$

证毕.

定理 1. 对任意给定的 $2 \leq k = o(\sqrt{n})$, 随机图 $G(n, p)$ 中存在 k -独立集的临界值为 $p_c = 1 - n^{-\frac{2}{k-1}}$.

$$\lim_{n \rightarrow \infty} Pr[G(n, p) \text{ 中存在 } k\text{-独立集}] =$$

$$\begin{cases} 0, & p = 1 - o(n^{-\frac{2}{k-1}}) \\ 1, & p = 1 - \omega(n^{-\frac{2}{k-1}}) \end{cases}.$$

证明. 设 X_n^k 是一个取值为非负整数的随机变量, 表示图 $G \in G(n, p)$ 中顶点个数为 k 的独立集个数, 即 $X_n^k = \sum_{\alpha \in \Gamma} J_\alpha$. 由引理 2 有 $E(X_n^k) = C_n^k(1-p)^{C_k^2}$.

1) 由文献[18]可知, 当 $k = o(\sqrt{n})$ 时, 有:

$$A_n^k = \prod_{j=1}^k (n+1-j) = n^k \left(1 - C_k^2 \frac{1}{n} + O\left(\frac{k^4}{n^2}\right) \right) \sim n^k, \quad (1)$$

由式(1)以及 Stirling's 公式 $k! \approx \left(\frac{k}{e}\right)^k \sqrt{2\pi k}$, 有:

$$C_n^k = \frac{A_n^k}{k!} \sim \frac{1}{\sqrt{2\pi k}} \left(\frac{ne}{k}\right)^k. \quad (2)$$

以 $p = 1 - n^{-\frac{2}{k-1}}$ 代入随机变量 X_n^k 的期望表达式 $E(X_n^k)$, 结合式(2)有:

$$\begin{aligned} E(X_n^k) &= C_n^k(1-p)^{C_k^2} \sim \frac{1}{\sqrt{2\pi k}} \left(\frac{en}{k}\right)^k (1-p)^{\frac{k(k-1)}{2}} = \\ &= \frac{1}{\sqrt{2\pi k}} \left(\frac{e}{k}\right)^k (n(1-p)^{\frac{k-1}{2}})^k = \\ &= \frac{1}{\sqrt{2\pi k}} \left(\frac{e}{k}\right)^k n^{(1-c)k}. \end{aligned} \quad (3)$$

当 $c > 1$ 时, 有

$$\lim_{n \rightarrow \infty} E(X_n^k) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi k}} \left(\frac{e}{k}\right)^k n^{(1-c)k} = 0.$$

由一阶矩方法可知, 当 $p = 1 - o(n^{-\frac{2}{k-1}})$ 时, 随机图 $G(n, p)$ 中高概率不存在 k -独立集.

2) 当 $c < 1$ 时, 有 $\lim_{n \rightarrow \infty} E(X_n^k) = \infty$. 因为:

$$\begin{aligned} E((X_n^k)^2) &= E\left(\left(\sum_{a \in \Gamma} J_a\right)^2\right) = \\ &= E(X_n^k) + \sum_{\alpha \neq \beta \in \Gamma} E(J_\alpha J_\beta) = \\ &= E(X_n^k) + C_n^k \sum_{s=0}^{k-1} C_k^s C_{n-k}^{k-s} (1-p)^{2C_k^2 - C_s^2}, \quad (4) \end{aligned}$$

其中, $C_n^k C_k^s C_{n-k}^{k-s}$ 表示从顶点个数为 n 的顶点集中选择 2 个顶点个数为 k 的顶点子集的组合数, 选择的 2 个顶点子集的公共顶点个数为 s ; $(1-p)^{2C_k^2 - C_s^2}$ 表示选择的 2 个顶点子集同时为 k -独立集的概率.

又因为 $E^2(X_n^k) = (C_n^k)^2 (1-p)^{2C_k^2}$, 结合式(4)有:

$$\begin{aligned} \frac{Var(X_n^k)}{E^2(X_n^k)} &= \frac{1}{E(X_n^k)} + \\ &+ \sum_{s=0}^{k-1} \frac{C_k^s C_k^s C_{n-k}^{k-s} (1-p)^{2C_k^2 - C_s^2}}{(C_n^k)^2 (1-p)^{2C_k^2}} - 1 = \\ &= \frac{1}{E(X_n^k)} + A_n + B_n - 1. \quad (5) \end{aligned}$$

其中:

$$\begin{aligned} A_n &= \frac{C_{n-k}^k + k C_{n-k}^{k-1}}{C_n^k}, \\ B_n &= \sum_{s=2}^{k-1} \frac{C_k^s C_{n-k}^{k-s} (1-p)^{-C_s^2}}{C_n^k}. \end{aligned}$$

由 $n-k < n$, $C_n^k \leq \frac{n^k}{k!}$ 可知: $\frac{C_{n-k}^k}{C_n^k} \rightarrow 1, (n \rightarrow \infty)$,

$\frac{k C_{n-k}^{k-1}}{C_n^k} \leq \frac{k^2}{n} e^{-\frac{k(k-1)}{n}} \rightarrow 0, (n \rightarrow \infty)$. 于是有:

$$\lim_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \frac{C_{n-k}^k + k C_{n-k}^{k-1}}{C_n^k} = 1. \quad (6)$$

由 $C_n^k = \frac{A_n^k}{k!}, A_k^s = \frac{k!}{(k-s)!}$ 以及 $A_n^k = A_n^s A_{n-s}^{k-s}$ 可知:

$$\begin{aligned} \frac{C_k^s C_{n-k}^{k-s}}{C_n^k} &= \frac{k!}{A_n^k} \times \frac{A_k^s}{s!} \times \frac{A_{n-k}^{k-s}}{(k-s)!} = \frac{A_k^s (k-s)!}{(k-s)!} \times \frac{A_k^s}{s!} \times \\ &\times \frac{A_{n-k}^{k-s}}{A_n^s A_{n-s}^{k-s}} \leq \frac{A_k^s A_k^s}{s! A_n^s} \leq \frac{k^{2s}}{s! A_n^s} \sim \frac{k^{2s}}{s! n^s}. \quad (7) \end{aligned}$$

以 $p = 1 - n^{-\frac{2}{c-1}}$ 带入表达式 B_n , 结合式(7)有:

$$\begin{aligned} B_n &= \sum_{s=2}^{k-1} \frac{C_k^s C_{n-k}^{k-s} (1-p)^{-C_s^2}}{C_n^k} \leq \\ &\leq \sum_{s=2}^{k-1} \frac{k^{2s}}{s! n^s} (1-p)^{-C_s^2} = O(1) \sum_{s=2}^{k-1} \left(\frac{k^2}{n} (1-p)^{-\frac{s-1}{2}}\right)^s = \\ &= O(1) \sum_{s=2}^{k-1} \left(\frac{k^2}{n} (1-p)^{-\frac{k-1}{2}}\right)^k = O(1) k^{2k+1} n^{-(c-1)k}. \quad (8) \end{aligned}$$

当 $c < 1$ 时, 有 $\lim_{n \rightarrow \infty} k^{2k+1} n^{-(c-1)k} = 0$, 结合式(8)可以得到 $\lim_{n \rightarrow \infty} B_n = 0$. 结合式(5)、式(6)有 $Var(X_n^k) =$

$o(E^2(X_n^k))$. 由二阶矩方法可知, 当 $p = 1 - \omega(n^{-\frac{2}{k-1}})$ 时, 随机图 $G(n, p)$ 中高概率存在 k -独立集.

综合证明 1) 2) 的分析可知, 当独立集节点数 $2 \leq k = o(\sqrt{n})$ 时随机图 $G(n, p)$ 中存在 k -独立集的临界值为 $p_c = 1 - n^{-\frac{2}{k-1}}$. 证毕.

注: $Var(X_n^k)$ 表示随机变量 X_n^k 的方差, $E(X_n^k)$ 表示随机变量 X_n^k 的期望, $Pr(X_n^k)$ 表示随机变量 X_n^k 的概率, A_n^k 表示从 n 个元素中取 k 个元素的排列数. C_n^k 表示从 n 个元素中取出 k 个元素的组合数.

由引理 1 可知, 当随机图 $G(n, m)$ 中的边数 m 与随机图 $G(n, p)$ 中的边概率 p 满足关系 $m \approx p C_n^2$ 时二者等价, 所以容易得到下面的定理 2.

定理 2. 对任意给定的 $2 \leq k = o(\sqrt{n})$, 随机图 $G(n, m)$ 中存在 k -独立集的临界值为

$$m_c = \left\lceil \frac{n(n-1)}{2} (1 - n^{-\frac{2}{k-1}}) \right\rceil,$$

其中, $[x]$ 为取整函数.

$$\begin{aligned} \lim_{n \rightarrow \infty} Pr[G(n, m) \text{ 中存在 } k\text{-独立集}] = \\ \begin{cases} 0, & m = \omega(m_c), \\ 1, & m = o(m_c). \end{cases} \end{aligned}$$

3 独立集算法

给定一个图 $G=(V, E)$, V 表示图 G 的顶点集合, E 表示图 G 的边集合. 若顶点集合 V 的顶点子集 T 中任意 2 个节点 u, v 之间都没有边, 则称该顶点子集 T 为图 G 的独立集. 若顶点子集 T 的元素个数为 k , 则称该顶点子集 T 为 k -独立集. 下面给出独立集的求解算法 ^[19]:

算法 1. 独立集求解算法.

输入: 简单无向图 $G=(V, E)$, 整数 $k \geq 2$, 其中, $|V|=n$;

输出: 独立集 $|S| \geq k$, 其中 $|S|$ 表示独立集 S 中的元素个数.

Step1. 对图 G 中的每个顶点用数字 $i=1, 2, \dots, n$ 标记, 即 $V=\{1, 2, \dots, n\}$.

Step2. 当 $i \leq n$ 时, 初始化独立集 $S_i=\{i\}$, 同时执行 $I=V-i \cup N(i)$; 如果 $I=\emptyset$, 此时 $|S_i|=1$, 那么执行 $i=i+1$, 然后转回 Step2; 否则, 转到 Step3.

Step3. 当 $|S_i| < k$ 时, 对集合 I 中的所有 j , 分别执行 $H_j=V-N[S_i] \cup j \cup N(j)$, 转到 Step4.

Step4. 找到节点 $j_{\max} = \arg \max_{j \in I} (|H_j|)$, 执行 $I=H_{j_{\max}}$, 若 $I=\emptyset$, 更新 $S_i=S_i \cup \{j_{\max}\}$, 转到 Step5; 否则, 更新 $S_i=S_i \cup \{j_{\max}\}$, 转到 Step6.

Step5. 若 $|S_i| < k$, 则执行 $i = i + 1$, 转到 Step2; 否则, 输出 S_i 程序终止.

Step6. 若 $|S_i| \geq k$, 则输出 S_i 程序终止; 否则, 转到 Step3.

注: 在 Step2 中, $I = V - i \cup N(i)$ 表示从图 G 的节点集 V 中去除节点 i 以及节点 i 的邻居节点之后的剩余节点集, 其中, $N(i) = \{j \in V: (i, j) \in E\}$. 在 Step3 中, $N[S_i]$ 表示集合 S_i 中节点和相应节点的邻居节点所构成的节点集合, $|H_j|$ 表示去除 S_i 中节点和相应节点的邻居节点, 以及节点 j 和节点 j 的邻居节点之后的剩余节点集的个数, $j_{\max} = \arg \max_{j \in I} (|H_j|)$ 是使得 $|H_j|$ 取最大值的节点.

4 数值实验分析

本节的主要工作是进行数值仿真实验分析, 该实验主要基于 3 个算法: 1) 随机图 $G(n, p)$ 实例产生

算法^[6,20]; 2) 随机图 $G(n, m)$ 实例产生算法^[6,20]; 3) 独立集算法^[19]. 数值实验的具体操作方法是首先利用随机图 $G(n, p)$ (或 $G(n, m)$) 算法产生一个随机图实例, 然后调用独立集算法判断图 $G \in G(n, p)$ (或 $G \in G(n, m)$) 中是否存在 k -独立集. 判断方法主要是看独立集算法在图 G 中是否能找到一个节点数不小于 k 的独立集, 若独立集算法找到的独立集的节点数不小于 k , 则确定在 $G \in G(n, p)$ (或 $G \in G(n, m)$) 中存在一个 k -独立集, 否则, 我们确定在 $G \in G(n, p)$ (或 $G \in G(n, m)$) 中不存在 k -独立集.

从定理 1 可知, 随机图 $G(n, p)$ 中 k -独立集性质出现相变的临界概率 p_c 与节点总数 n 和独立集节点数 k 有关. 所以, 在模拟随机图 $G(n, p)$ 中 k -独立集性质的相变现象时, 分别取节点总数 n 值为 100, 200, 300, 400, 500, 独立集节点数 k 值为 5, 10, 15, 20, 25, 如表 1 所示:

Table 1 Comparison Between the Theoretical Threshold and the Numerical Simulations for the Existence of k -Independent Set in Random Graph $G(n, p)$

表 1 随机图 $G(n, p)$ 中存在 k -独立集的理论临界和仿真临界的对比

k	$n=100$		$n=200$		$n=300$		$n=400$		$n=500$	
	Theory	Simulation	Theory	Simulation	Theory	Simulation	Theory	Simulation	Theory	Simulation
5	0.90	0.80	0.93	0.86	0.94	0.88	0.95	0.90	0.96	0.91
10	0.64	0.68	0.69	0.79	0.72	0.83	0.74	0.84	0.75	0.85
15	0.48	0.65	0.53	0.77	0.56	0.80	0.58	0.83	0.59	0.84
20	0.38	0.63	0.43	0.76	0.45	0.79	0.47	0.82	0.48	0.84
25	0.32	0.59	0.36	0.74	0.38	0.78	0.39	0.81	0.40	0.83

表 1 中的每一个理论值是取对应的 k 和 n 时利用定理 1 的临界概率 p_c 计算得到的值, 实验值是仿真得到的图 $G \in G(n, p)$ 中存在 k -独立集的临界概率. 从表 1 可以看到, 当独立集的节点数 $k \ll \sqrt{n}$ 时, 随机图 $G(n, p)$ 中存在 k -独立集的理论临界值与仿真得到的实验值误差很小, 产生这一误差的原因主要是由于我们的 n 值取得较小, 因为定理 1 中的临界概率 p_c 是在 $n \rightarrow \infty$ 时的极限情况. 从表 1 还可以看到, 当 $k \gg \sqrt{n}$ 时, 随机图 $G(n, p)$ 中存在 k -独立集的理论临界值与仿真得到的实验值之间的误差很大, 而且随着 k 的增加, 理论值与实验值之间的误差也在增大, 说明定理 1 的理论临界概率对 $k \gg \sqrt{n}$ 失效.

图 1 给出了当独立集节点数 k 固定时随机图 $G(n, p)$ 中存在 k -独立集的临界概率 p_c 随节点总数 n 的变化情况, 这里取独立集节点数 $k=10$, 节点总数 n 分别取值为 100, 200, 300, 400, 500. 图 1 中的

每个数据点是对每个给定的概率 p 和节点总数 n 利用随机图 $G(n, p)$ 生成算法生成的 100 个图 $G \in G(n, p)$ 实例中存在 10-独立集的统计概率.

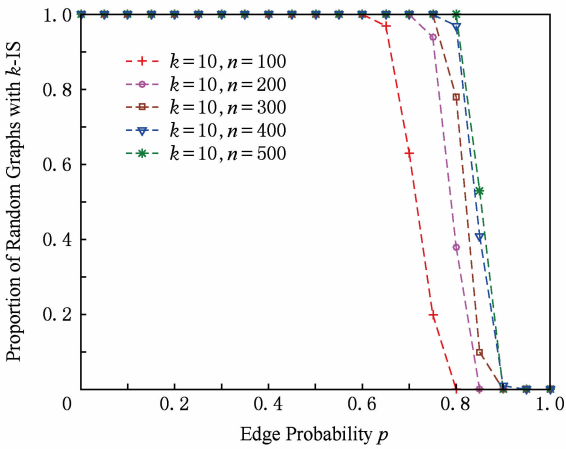


Fig. 1 Phase transition of 10-IS in random graph $G(n, p)$

图 1 随机图 $G(n, p)$ 中 10-独立集的相变

从图 1 可以看出,当独立集节点数 $k=10$ 固定时,随机图 $G(n, p)$ 中存在 10-独立集的临界概率 p_c 随节点总数 n 的增加而增加,而且临界概率 p_c 向右移动.

图 2 给出了当节点总数 n 固定时随机图 $G(n, p)$ 中 k -独立集出现相变的临界概率 p_c 随独立集节点数 k 的变化情况,这里取节点总数 $n=400$,独立集节点数 k 分别取值为 5, 10, 15, 20, 25. 图 2 中的每个数据点是对每个给定的概率 p 和节点总数 n

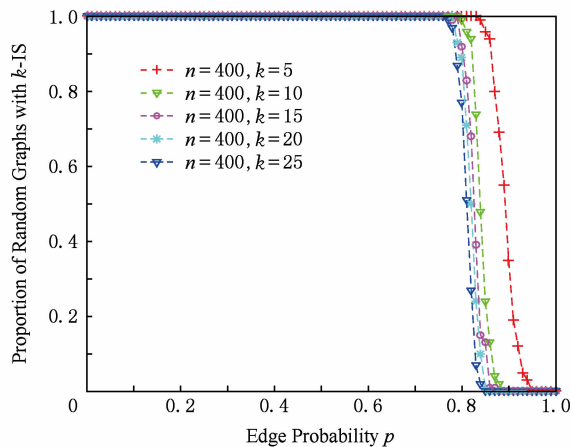


Fig. 2 Phase transition of k -IS in random graph $G(n, p)$

图 2 随机图 $G(n, p)$ 中 k -独立集的相变

利用随机图 $G(n, p)$ 生成算法生成的 100 个图 $G \in G(n, p)$ 实例中存在 k -独立集的统计概率. 从图 2 可以看出,当节点总数 $n=400$ 固定时,随机图 $G(n, p)$ 中存在 k -独立集的临界概率 p_c 随独立集节点数 k 的增加而减小,而且临界值 p_c 向左移动.

从定理 2 可知,随机图 $G(n, m)$ 中 k -独立集性质出现相变的临界边数 m_c 与节点总数 n 和独立集节点数 k 有关. 所以,在模拟随机图 $G(n, m)$ 中 k -独立集性质的相变现象时,节点总数 n 分别取值为 100, 200, 300, 400, 500, 独立集节点数 k 分别取值为 5, 10, 15, 20, 25, 如表 2 所示. 表 2 中的理论值是取对应的 k 和 n 时利用定理 2 的临界边数 m_c 计算得到的值,实验值是仿真实验得到的临界边数. 从表 2 可以看出,当独立集的节点数 $k \ll \sqrt{n}$ 时,随机图 $G(n, m)$ 中存在 k -独立集的理论临界边数 m_c 与仿真实验边数之间的误差很小,理论结果与仿真结果一致. 然而,当 $k \gg \sqrt{n}$ 时,随机图 $G(n, m)$ 中存在 k -独立集的理论临界边数与仿真得到的实验边数之间的误差很大,而且随着 k 的增加,理论值与实验值之间的误差也在增大,说明定理 2 的理论临界值对 $k \gg \sqrt{n}$ 失效.

Table 2 Comparison Between the Theoretical Threshold and the Numerical Simulations for the Existence of k -Independent Set in Random Graph $G(n, m)$

表 2 随机图 $G(n, m)$ 中存在 k -独立集的理论临界边数和仿真临界边数对比

k	$n=100$		$n=200$		$n=300$		$n=400$		$n=500$	
	Theory	Simulation	Theory	Simulation	Theory	Simulation	Theory	Simulation	Theory	Simulation
5	4 455	4 000	18 493	17 114	42 261	39 063	75 810	71 022	119 171	113 568
10	3 171	3 550	13 769	15 721	32 223	36 818	58 725	67 830	93 398	107 328
15	2 386	3 350	10 565	15 323	24 994	36 369	45 894	66 234	73 408	106 080
20	1 902	3 250	8 506	15 124	20 246	35 920	37 328	65 436	59 896	104 832
25	1 578	3 050	7 103	14 925	16 967	35 022	31 364	64 638	50 426	103 584

图 3 给出了当独立集的节点数 k 固定时随机图 $G \in G(n, m)$ 中存在 k -独立集的临界边数 m_c 随节点总数 n 的变化情况,这里分别取独立集节点数 $k=10$,节点总数 n 分别取值为 100, 200, 300, 400, 500. 图 3 中的每个数据点是对每个给定的边数 m 和节点总数 n 利用随机图 $G(n, m)$ 生成算法生成的 100 个图 $G \in G(n, m)$ 实例中存在 k -独立集的统计概率. 从图 3 可以看出,当固定独立集节点数 $k=10$ 时,随机图 $G(n, m)$ 中存在 k -独立集的临界边数 m_c 随节点总数 n 的增加而增加,临界边数 m_c 向右移.

图 4 给出了当节点总数 n 固定时随机图 $G(n, m)$ 中 k -独立集出现相变的临界边数 m_c 随独立集的节点数 k 的变化情况,这里分别取节点总数 $n=200$,独立集节点数 k 分别取值为 5, 10, 15, 20, 25. 图 4 中的每个数据点是对每个给定的边数 m 和节点总数 n 利用随机图 $G(n, m)$ 生成算法生成的 100 个图 $G \in G(n, m)$ 实例中存在 k -独立集的统计概率. 从图 4 可以看出,当固定顶点数 $n=200$ 时,随机图 $G(n, m)$ 中存在 k -独立集的临界边数 m_c 随着独立集节点数 k 的增加而减小,临界边数 m_c 向左移.

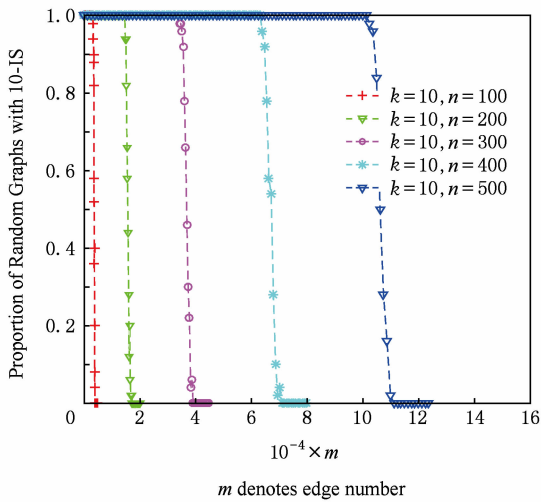


Fig. 3 Phase transition of 10-IS in random graph $G(n, m)$
图 3 随机图 $G(n, m)$ 中 10-独立集的相变

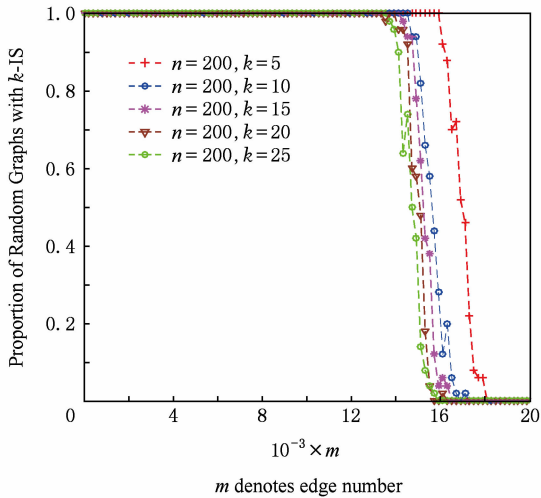


Fig. 4 Phase transition of k -IS in random graph $G(n, m)$
图 4 随机图 $G(n, m)$ 中 k -独立集的相变

5 结束语

本文得到了当 $2 \leq k = o(\sqrt{n})$ 时随机图 $G(n, p)$ 中存在 k -独立集的临界值 $p_c = 1 - n^{-\frac{2}{k-1}}$, 即当 $p = 1 - \omega(n^{-\frac{2}{k-1}})$ 时, 随机图 $G(n, p)$ 中大概率存在 k -独立集; 当 $p = 1 - o(n^{-\frac{2}{k-1}})$ 时, 随机图 $G(n, p)$ 中大概率不存在 k -独立集. 同时还给出了随机图 $G(n, m)$ 中存在 k -独立集的临界值为 $m_c = \left\lceil \frac{n(n-1)}{2} (1 - n^{-\frac{2}{k-1}}) \right\rceil$, 其中, $\lceil x \rceil$ 为取整函数. 最后, 我们进行了仿真实验分析, 实验结果表明, 当独立集节点数 $2 \leq k \ll \sqrt{n}$ 时, 随机图 $G(n, p)$ 和 $G(n, m)$ 中 k -独立集性质出现相变的临界值与仿真实验结果一致且临界值与节点总

数 n 和独立集节点数 k 有关, 但当 $k \gg \sqrt{n}$ 时, 随机图 $G(n, p)$ 和 $G(n, m)$ 中 k -独立集性质出现相变的临界值与仿真结果不一致.

参 考 文 献

[1] Coja-Oghlan A, Efthymiou C. On independent sets in random graphs [J]. Random Structures & Algorithms, 2014, 47(3): 136-144

[2] Zhao Chunxiao, Wang Guangxing. A δ -degree-based clustering algorithm in MANET [J]. Journal of Computer Research and Development, 2005, 42(5): 818-822 (in Chinese)
(赵春晓, 王光兴. 一种 δ -度约束的自组网成簇算法[J]. 计算机研究与发展, 2005, 42(5): 818-822)

[3] Garey M R, Johnson D S. Computers and Intractability: A Guide to the Theory of NP-Completeness [M]. New York: W H Freeman and Company, 1979

[4] Sloane N J A. Unsolved problem in graph theory arising from the study of codes [J]. Graph Theory Notes of New York, 1989, 18(11): 11-20

[5] Erdős P, Rényi A. On random graphs I [J]. Publicationes Mathematicae, 1959(6): 290-297

[6] Wang Xiaofan, Li Xiang, Chen Guanrong. Network Science: An Introduction [M]. Beijing: Higher Education Press, 2012: 200-201 (in Chinese)
(汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012: 200-201)

[7] Hopcroft J, Kannan R. Computer Science Theory for the Information Age [M]. Shanghai: Shanghai Jiao Tong University Press, 2013: 39-72

[8] Zdeborová L, Krzakala F. Phase transitions in the coloring of random graphs [J/OL]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(3): Article Number 031131. [2016-09-01]. <https://doi.org/10.1103/PhysRevE.76.031131>

[9] Xu Ke, Li Wei. Exact phase transitions in random constraint satisfaction problems [J]. Journal of Artificial Intelligence Research, 2000, 12(1): 93-103

[10] Xu Ke, Li Wei. Many hard examples in exact phase transitions [J]. Theoretical Computer Science, 2006, 355(3): 291-302

[11] Mertens S, Zard M, Zecchina R. Threshold values of random K-SAT from the cavity method [J]. Random Structures & Algorithms, 2006, 28(3): 340-373

[12] Nekovee M, Moreno Y, Bianconi G, et al. Theory of rumour spreading in complex social networks [J]. Physica A Statistical Mechanics & Its Applications, 2008, 374(1): 457-470

[13] Gómezgardiñes J, Lotero L, Taraskin S N, et al. Explosive contagion in networks [J/OL]. Scientific Reports, 2016, 6(1): Article Number 19767. [2016-09-01]. <https://doi.org/10.1038/srep19767>

[14] Culberson J, Gao Yong, Anton C. Phase transitions of dominating clique problem and their implications to heuristics in satisfiability search [DB]. Trier, Rheinland-Pfalz, Germany: DBLP, 2005 [2016-09-01]. <http://dblp.org/db/conf/ijcai/ijcai2005.html>

[15] Nehéz M, Olejár D, Demetrian M. A detailed study of the dominating cliques phase transition in random graphs [C] // Proc of the 9th Annual Int Conf on Theory and Applications of Models of Computation. Berlin: Springer, 2012: 594-603

[16] Bollobasi B. Random Graphs [M]. New York: Academic Press, 2001

[17] Seierstad T G. The phase transition in random graphs and random graph processes [D]. Berlin: Humboldt University, 2007

[18] Nehéz M, Olejár D, Demetrian M. On emergence of dominating cliques in random graphs [EB/OL]. 2008 [2016-09-01]. <https://arxiv.org/abs/0805.2105>

[19] Dharwadker A. The independent set algorithm [OL]. 2006 [2016-09-01]. http://www.dharwadker.org/independent_set/

[20] Nobari S, Lu X, Karras P, et al. Fast random graph generation [C] //Proc of the 14th Int Conf on Extending Database Technology. New York: ACM, 2011: 331-342



Lu Youjun, born in 1985. PhD candidate. Student member of CCF. His main research interests include computational complexity and complex networks.



Xu Daoyun, born in 1959. Professor and PhD supervisor at the College of Computer Science and Technology, Guizhou University. Senior member of CCF. His main research interests include computability theory and computational complexity.

2018 年《计算机研究与发展》专题(正刊)征文通知

——新型存储系统前沿技术

近年来,随着国家和社会信息化发展的不断加速,对信息存储的提出了越来越高的要求.一方面,大数据时代,数据存储的规模和处理需求越来越高,亟需新型存储系统和技术以提供更高的性能和更好的可扩展性.另一方面,由于各种人工智能系统及相关技术的出现,现有的存储技术和系统难以满足上层系统和技术的需求.因此,存储系统结构技术研究面临诸多新的机遇和挑战.

《计算机研究与发展》开辟了“计算机体系结构前沿技术”系列,并计划于 2018 年出版“新型存储系统结构前沿技术”专题.本专题将集中讨论面向非易失存储器件的存储体系结构技术、海量信息存储系统研究与技术、新型工艺支撑的计算机存储体系结构技术、存储前沿技术等,相关研究可以涉及不同应用领域、环境、场景的存储系统,也可以涉及存储系统的不同层次和部件,还可以包括基于新型工艺和器件的存储结构设计等.

本次专题受到中国计算机学会信息存储专委会的支持.专题录用的文章作者将被邀请参加 2018 年 9 月全国信息存储学术会议,邀请作者到会演讲并参与讨论.欢迎相关领域的专家学者和科研人员踊跃投稿.现将专题论文征集的有关事项通知如下.

征文范围(但不限于)

- 1) 面向非易失存储器件的存储体系结构技术,如:面向非易失存储器件的持久性内存(memory)和存储(storage)结构与技术;面向持久性内存的编程模型、空间管理、文件系统和数据库等;固态 SSD 存储与技术;新型存储系统及其应用技术.
- 2) 海量信息存储系统技术,如:高性能计算机存储系统、分布式文件系统;网络存储、数据存储网络及相关技术;海量信息存储技术;网络存储系统可用性、安全性、可靠性及容灾研究;重复数据删冗、归档、分级等存储技.
- 3) 新型工艺支撑的计算机存储体系结构技术,如:非易失存储器结构技术;PIM (Processing in Memory) 技术;3-维堆叠结构技术.
- 4) 存储前沿技术,如:软件定义存储与存储超融合;类脑存储;量子存储.

稿件内容也可以为上述所列内容的交叉.本刊也鼓励作者讨论与上述领域相关但此处未予提及的重要问题、创造性的研究思路和探索视角.

投稿要求

- 1) 论文应属于作者的科研成果,未在国内外公开发行的刊物或会议上发表,不存在一稿多投问题.作者在投稿时,需向编辑部提交版权转让协议.
- 2) 论文一律用 Word 格式排版,格式体例参考近期出版的《计算机研究与发展》的要求(<http://crad.ict.ac.cn/下载区域>).
- 3) 论文通过期刊网站(<http://crad.ict.ac.cn>)投稿,投稿时需提供作者的联系方式.作者投稿时请务必在留言中注明“存储系统前沿 2018 专题”(否则按自由来稿处理).

重要日期

征文截止日期:2018 年 4 月 5 日

录用通知日期:2018 年 5 月 31 日

修改稿提交日期:2018 年 6 月 15 日

出版日期:2018 年 9 月

特邀编委

刘志勇 研究员 中国科学院计算技术研究所 zyliu@ict.ac.cn

舒继武 教授 清华大学 shujw@tsinghua.edu.cn

联系方式

编辑部:crad@ict.ac.cn, 010-62620696, 010-62600350

通信地址:北京 2704 信箱《计算机研究与发展》编辑部 邮政编码:100190