

Research Article

On the Isolated Vertices and Connectivity in Random Intersection Graphs

Yilun Shang

Institute for Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249, USA

Correspondence should be addressed to Yilun Shang, shylmath@hotmail.com

Received 10 January 2011; Accepted 5 April 2011

Academic Editor: Liying Kang

Copyright © 2011 Yilun Shang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study isolated vertices and connectivity in the random intersection graph $G(n, m, p)$. A Poisson convergence for the number of isolated vertices is determined at the threshold for absence of isolated vertices, which is equivalent to the threshold for connectivity. When $m = \lfloor n^\alpha \rfloor$ and $\alpha > 6$, we give the asymptotic probability of connectivity at the threshold for connectivity. Analogous results are well known in Erdős-Rényi random graphs.

1. Introduction

The classical random graph $G(n, p)$, introduced by Erdős and Rényi in the late 1950s, consists of a fixed set of n vertices and edges that exist with a certain probability p , independently from each other. Since then many other random graph models with dependent edges have been developed. Among them, random intersection graph [1, 2] is defined as follows. Consider a set V with n vertices and another universal set W with m elements. Define a bipartite graph $B(n, m, p)$ with independent vertex sets V and W . Edges between $v \in V$ and $w \in W$ exist independently with probability p . The random intersection graph $G(n, m, p)$ derived from $B(n, m, p)$ is defined on the vertex set V with vertices $v_1, v_2 \in V$ adjacent if and only if there exists some $w \in W$ such that both v_1 and v_2 are adjacent to w in $B(n, m, p)$.

Appropriately scaling the parameter m as $m = \lfloor n^\alpha \rfloor$ with some $\alpha > 0$, Singer-Cohen [1] establishes connectivity thresholds for $G(n, m, p)$: the threshold lies at $p = (\ln n)/m$ and $\sqrt{(\ln n)/nm}$ for $\alpha \leq 1$ and $\alpha > 1$, respectively. The result also reveals an asymptotic equivalence of graph connectivity and absence of isolated vertices in $G(n, m, p)$, that is, the zero-one law for the absence of isolated vertices is equal to that for connectivity. This is familiar in Erdős-Rényi model; see [3, 4] for more details. The study in the present paper is in continuation of Chapter 3 in [1]. Taking our cue from existing results for Erdős-Rényi

graphs (e.g., [4, Corollary 3.31] and [3, Theorem 7.3]), we aim to explore similar results for the properties of isolated vertices and connectivity in $G(n, m, p)$.

The connectivity thresholds of another class of random intersection graphs $G(n, m, k)$, called random key graphs or uniform random intersection graphs, have been investigated recently [5, 6]. Both $G(n, m, p)$ and $G(n, m, k)$ can be viewed as subclasses of a general model [7]. In [8], the authors determine a zero-one law for the absence of isolated vertices in $G(n, m, k)$, which again turns out to be equivalent to that for graph connectivity [6]. Moreover, they show a Poisson convergence for the number of isolated vertices, which refines the corresponding zero-one law and leads to a “double exponential” result.

In this paper, we deal with the asymptotic distribution of the number of isolated vertices and address the connectivity probability in $G(n, m, p)$ with $m = \lfloor n^\alpha \rfloor$, $\alpha > 6$. A Poisson approximation result (Theorem 2.1) for the number of isolated vertices is obtained by utilizing the Stein-Chen method, which yields convergence to a Poisson random variable. The isolated vertices threshold [1, Proposition 3.2] now readily follows from our Theorem 2.1 by an easy monotonicity argument. In addition, based on a strong equivalence theorem [9] relating the $G(n, m, p)$ and $G(n, p)$ models we derive an approximation of the probability of connectivity at the threshold when $\alpha > 6$ (see Theorem 2.3), which is analogous to the well-known “double exponential” result of Erdős and Rényi [10].

Other related works regarding $G(n, m, p)$ model have been reported. For example, [11, 12] examines the limiting distribution of the degree of a typical vertex, [13] treats the evolution of the order of the largest component, and random weights are assigned to the vertices in [14] to get general degree distributions.

The rest of the paper is organized as follows. Our main results are presented in Section 2. Sections 3 and 4 contain technical proofs of Theorems 2.1 and 2.3, respectively. Throughout the paper we set $m = \lfloor n^\alpha \rfloor$ for some $\alpha > 0$.

2. Main Results

In this section we provide our main results. Let X denote the number of isolated vertices in $G(n, m, p)$ and let $\text{Poi}(\lambda)$ be a Poisson random variable with parameter λ . Denote by $E(X)$ and $\text{Var}(X)$ the mean and variance of random variable X , respectively. Recall that the Poisson random variable has the unusual property that the mean and variance are both equal to the parameter λ .

Theorem 2.1. *In the model $G(n, m, p)$, let*

$$p = \begin{cases} \frac{\ln n + \beta_n}{m}, & \alpha \leq 1, \\ \sqrt{\frac{\ln n + \beta_n}{nm}}, & \alpha > 1, \end{cases} \quad (2.1)$$

where $\beta_n \in \mathbb{R}$. If $\lim_{n \rightarrow \infty} \beta_n = \beta \in \mathbb{R}$, then one has

$$X \xrightarrow{D} \text{Poi}(e^{-\beta}) \quad (2.2)$$

as $n \rightarrow \infty$, where \xrightarrow{D} represents convergence in distribution.

The upcoming corollary is immediate from Theorem 2.1.

Corollary 2.2. *In the model $G(n, m, p)$ with p determined through (2.1), suppose $\lim_{n \rightarrow \infty} \beta_n = \beta \in \mathbb{R}$. Then one gets*

$$\lim_{n \rightarrow \infty} P(G(n, m, p) \text{ contains no isolated vertices}) = e^{-e^{-\beta}}. \quad (2.3)$$

For a parallel “double exponential” result for connectivity when α is large, we have the following.

Theorem 2.3. *In the model $G(n, m, p)$ with $\alpha > 6$ and p determined through (2.1) (i.e., $p = \sqrt{(\ln n + \beta_n)/nm}$), assume that $\lim_{n \rightarrow \infty} \beta_n = \beta \in \mathbb{R}$. Then one has*

$$\lim_{n \rightarrow \infty} P(G(n, m, p) \text{ is connected}) = e^{-e^{-\beta}}. \quad (2.4)$$

These results complement those presented in [1] and get further insight into the evolutionary similarities and differences between $G(n, m, p)$ and $G(n, p)$ models. A natural question would be to ask what happens for connectivity probability when α is small. This is currently under investigation.

3. Proof of Theorem 2.1

For $i = 1, \dots, n$, let $X_i = 1_{[\text{vertex } i \text{ is isolated in } G(n, m, p)]}$ and $X = \sum_{i=1}^n X_i$. Therefore, X counts the number of isolated vertices in $G(n, m, p)$ as defined in Section 2. We will demonstrate the asymptotic Poisson distribution of X by employing the Stein-Chen method [15].

Before proceeding, we first introduce some definitions and notations. Let $q = 1 - p$ and $|S|$ denote the cardinality of a set S . For two integer-valued random variables X and Y , the total variation distance between them (more correctly, between their distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$) is given by

$$d_{\text{TV}}(X, Y) = d_{\text{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{A \subseteq \mathbb{Z}} |P(X \in A) - P(Y \in A)|. \quad (3.1)$$

Let Γ be a finite set of indices and let $(I_a)_{a \in \Gamma}$ be a family of random indicator variables. We say $(I_a)_{a \in \Gamma}$ are positively related (c.f. [15]) if, for each $a \in \Gamma$, there exist random indicator variables $(J_{ba})_{b \in \Gamma \setminus \{a\}}$ with the distributions

$$\mathcal{L}\left((J_{ba})_{b \in \Gamma \setminus \{a\}}\right) = \mathcal{L}\left((I_b)_{b \in \Gamma \setminus \{a\}} \mid I_a = 1\right), \quad (3.2)$$

such that $J_{ba} \geq I_b$ for every $b \neq a$. It is notable that “positively related” is much stronger than “positively correlated”. Suppose $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ are sequences of positive real numbers. We write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

A useful result obtained by the Stein-Chen method is the following.

Lemma 3.1 (see [4, 15]). Suppose that $Y = \sum_{a \in \Gamma} I_a$, where the $(I_a)_{a \in \Gamma}$ are positively related random indicator variables. Then one has

$$d_{\text{TV}}(Y, \text{Poi}(EY)) \leq \frac{1 - e^{-EY}}{EY} \left[\text{Var } Y - EY + 2 \sum_{a \in \Gamma} (EI_a)^2 \right]. \quad (3.3)$$

The next lemma collects some well-known approximations that are used in this paper.

Lemma 3.2. If $mp \rightarrow 0$, then $(1 - p)^m \sim 1 - mp$; and if $mp^2 \rightarrow 0$, then $(1 - p)^m \sim e^{-mp}$.

In the sequel, we estimate the expectation of random variable X .

Lemma 3.3. Suppose $\alpha > 0$. Under the assumptions of Theorem 2.1, one gets

$$\lim_{n \rightarrow \infty} EX = e^{-\beta}. \quad (3.4)$$

Proof. The probability that a vertex i is isolated can be computed as

$$EX_i = \sum_{s=0}^m \binom{m}{s} p^s (1-p)^{m-s} (1-p)^{(n-1)s} = \left[1 - p + p(1-p)^{n-1} \right]^m, \quad (3.5)$$

where the index s represents the number of vertices in W which are adjacent to i in $B(n, m, p)$. Hence

$$EX = n \left[1 - p + p(1-p)^{n-1} \right]^m. \quad (3.6)$$

For $\alpha \leq 1$, we have

$$\begin{aligned} mp^2 (1 - q^{n-1})^2 &\leq mp^2 = \frac{(\ln n + \beta_n)^2}{m} \longrightarrow 0, \\ pmq^{n-1} &= (\ln n + \beta_n) \left(1 - \frac{\ln n + \beta_n}{m} \right)^{n-1} \leq (\ln n + \beta_n) e^{-((n-1)/m)(\ln n + \beta_n)} \longrightarrow 0 \end{aligned} \quad (3.7)$$

as $n \rightarrow \infty$. Thus by Lemma 3.2, we obtain

$$\begin{aligned} EX &\sim ne^{-pm(1-q^{n-1})} = ne^{-pm} e^{pmq^{n-1}} \\ &\sim ne^{-pm} = ne^{-(\ln n + \beta_n)} \longrightarrow e^{-\beta}, \end{aligned} \quad (3.8)$$

as $n \rightarrow \infty$.

For $\alpha > 1$, note that

$$\begin{aligned} np &= n\sqrt{\frac{\ln n + \beta_n}{nm}} = \sqrt{\frac{\ln n + \beta_n}{n^{\alpha-1}}} \longrightarrow 0, \\ m(np^2)^2 &= m\left(\frac{\ln n + \beta_n}{m}\right)^2 = \frac{(\ln n + \beta_n)^2}{m} \longrightarrow 0 \end{aligned} \quad (3.9)$$

as $n \rightarrow \infty$. By using Lemma 3.2, we have

$$\begin{aligned} EX &\sim n[1 - p + p(1 - np)]^m = n(1 - np^2)^m \\ &\sim ne^{-np^2m} = ne^{-(\ln n + \beta_n)} \longrightarrow e^{-\beta}, \end{aligned} \quad (3.10)$$

as $n \rightarrow \infty$. The proof is then complete. \square

Proof of Theorem 2.1. The triangular inequality for the total variation distance implies

$$d_{\text{TV}}(X, \text{Poi}(e^{-\beta})) \leq d_{\text{TV}}(X, \text{Poi}(EX)) + d_{\text{TV}}(\text{Poi}(EX), \text{Poi}(e^{-\beta})). \quad (3.11)$$

By a coupling argument ([16, page 58]) and Lemma 3.3, we have

$$d_{\text{TV}}(\text{Poi}(EX), \text{Poi}(e^{-\beta})) \leq |EX - e^{-\beta}| \longrightarrow 0 \quad (3.12)$$

as $n \rightarrow \infty$. Combining this with (3.11), we now only need to prove

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(X, \text{Poi}(EX)) = 0. \quad (3.13)$$

First, we claim that $(X_i)_{i=1}^n$ are positively related. To see this, define

$$X_{ji} = 1_{[\text{vertex } j \text{ is isolated in } G(n, m - |S_i|, p)]} \quad (3.14)$$

for every $j \neq i$, where $S_i \subseteq W$ represents the elements in W which are adjacent to i in $B(n, m, p)$ (S_i is possibly empty). The random graphs $G(n, m - |S_i|, p)$ and $G(n, m, p)$ are coupled in a natural way. Conditional on the isolation of vertex i in $G(n, m, p)$, any vertex j ($j \neq i$) is not adjacent to vertices of S_i in $B(n, m, p)$. Hence, we have

$$\mathcal{L}\left((X_{ji})_{j=1, j \neq i}^n\right) = \mathcal{L}\left((X_j)_{j=1, j \neq i}^n \mid X_i = 1\right). \quad (3.15)$$

For every $j \neq i$, if $X_j = 1$ then $X_{ji} = 1$. Consequently, we get $X_{ji} \geq X_j$.

By Lemma 3.1, the binary nature and exchangeability of the random variables involved, we find that

$$\begin{aligned}
 d_{\text{TV}}(X, \text{Poi}(EX)) &\leq \frac{1 - e^{-EX}}{EX} \left[\text{Var } X - EX + 2 \sum_{i=1}^n (EX_i)^2 \right] \\
 &\leq \frac{1}{EX} \left[\text{Var } X - EX + 2 \sum_{i=1}^n (EX_i)^2 \right] \\
 &= \frac{1}{EX} \left[E(X^2) - (EX)^2 - EX + 2n(EX_1)^2 \right] \\
 &= \frac{1}{EX} \left[n(n-1)E(X_1X_2) - n(n-2)(EX_1)^2 \right].
 \end{aligned} \tag{3.16}$$

The cross term $E(X_1X_2)$ in (3.16) is shown to be given by

$$\begin{aligned}
 E(X_1X_2) &= \sum_{s=0}^m 2^{m-s} \binom{m}{s} (1-p)^{2s} (1-p)^{(m-s)(n-2)} [p(1-p)]^{m-s} \\
 &= \left[(1-p)^2 + 2p(1-p)^{n-1} \right]^m,
 \end{aligned} \tag{3.17}$$

where s counts the number of vertices in W adjacent to neither 1 or 2 in $B(n, m, p)$, leaving $m-s$ vertices in W adjacent to exactly one of 1, 2.

Combining (3.5), (3.16), and (3.17) readily gives

$$\begin{aligned}
 d_{\text{TV}}(X, \text{Poi}(EX)) \\
 \leq \frac{(n-1) \left[(1-p)^2 + 2p(1-p)^{n-1} \right]^m - (n-2) \left[1-p + p(1-p)^{n-1} \right]^{2m}}{\left[1-p + p(1-p)^{n-1} \right]^m}.
 \end{aligned} \tag{3.18}$$

For $\alpha \leq 1$, we have similarly as in the proof of Lemma 3.3,

$$mp^2(2-p-2q^{n-1})^2 \leq 4mp^2 \rightarrow 0, \quad pmq^{n-1} \rightarrow 0 \tag{3.19}$$

as $n \rightarrow \infty$. Thereby, it follows from Lemma 3.2 that

$$\begin{aligned}
 (n-1) \left[(1-p)^2 + 2p(1-p)^{n-1} \right]^m &= (n-1) \left[1-p(2-p-2q^{n-1}) \right]^m \\
 &\sim ne^{-mp(2-p-2q^{n-1})} \\
 &= ne^{-2mp} e^{mp(p+2q^{n-1})} \\
 &\sim ne^{-2mp}.
 \end{aligned} \tag{3.20}$$

Applying this to (3.18), we obtain

$$\begin{aligned} d_{\text{TV}}(X, \text{Poi}(EX)) &\leq \frac{(1 + o(1))ne^{-2mp} - (1 + o(1))ne^{-2mp}}{(1 + o(1))e^{-mp}} \\ &= o(1)ne^{-mp} = o(1)e^{-\beta_n} \longrightarrow 0, \end{aligned} \quad (3.21)$$

as $n \rightarrow \infty$.

For $\alpha > 1$, we get as in the proof of Lemma 3.3,

$$np \longrightarrow 0, \quad m(np^2)^2 \longrightarrow 0 \quad (3.22)$$

as $n \rightarrow \infty$. Hence, from Lemma 3.2 we have

$$\begin{aligned} (n-1)\left[(1-p)^2 + 2p(1-p)^{n-1}\right]^m &= (n-1)\left[1 - 2p + p^2 + 2p(1-p)^{n-1}\right]^m \\ &\sim n\left[1 - 2p + p^2 + 2p(1-np)\right]^m \\ &\sim n\left(1 - 2np^2\right)^m \\ &\sim ne^{-2nmp^2}. \end{aligned} \quad (3.23)$$

Applying this to (3.18), we have

$$\begin{aligned} d_{\text{TV}}(X, \text{Poi}(EX)) &\leq \frac{(1 + o(1))ne^{-2nmp^2} - (1 + o(1))ne^{-2nmp^2}}{(1 + o(1))e^{-nmp^2}} \\ &= o(1)ne^{-nmp^2} = o(1)e^{-\beta_n} \longrightarrow 0, \end{aligned} \quad (3.24)$$

as $n \rightarrow \infty$, which concludes the proof. \square

4. Proof of Theorem 2.3

Let

$$\hat{p} = 1 - \left(1 - \frac{p^2}{q^2 + npq + \binom{n}{2}p^2}\right)^m. \quad (4.1)$$

The following lemma drawn from [9] states an equivalence of $G(n, m, p)$ and $G(n, \hat{p})$ models.

Lemma 4.1 (see [9]). *Let $\alpha > 6$ and p be such that*

$$\frac{\omega}{n\sqrt{m}} \leq p \leq \sqrt{\frac{2 \ln n - \omega}{m}} \quad (4.2)$$

for some $\omega \rightarrow \infty$. For any $a \in [0, 1]$ and any graph property \mathcal{A} , as $n \rightarrow \infty$ it follows that

$$P(G(n, m, p) \in \mathcal{A}) \longrightarrow a \quad \text{if and only if} \quad P(G(n, \hat{p}) \in \mathcal{A}) \longrightarrow a. \quad (4.3)$$

We recall the following classical result for connectivity threshold of $G(n, p)$.

Lemma 4.2 (see [10]). *Let $c \in \mathbb{R}$ be fixed and $p = (\ln n + c + o(1))/n$. Then*

$$P(G(n, p) \text{ is connected}) \longrightarrow e^{-e^{-c}}, \quad (4.4)$$

as $n \rightarrow \infty$.

Proof of Theorem 2.3. In view of Lemmas 4.1 and 4.2, it suffices to prove that

$$n\hat{p} - \ln n \longrightarrow \beta, \quad (4.5)$$

as $n \rightarrow \infty$.

By the assumptions, we find that

$$\begin{aligned} & n\hat{p} - \ln n \\ & \sim n \left[1 - \left(1 - \frac{\mathfrak{D}}{(1 - \sqrt{\mathfrak{D}})^2 + n\sqrt{\mathfrak{D}}(1 - \sqrt{\mathfrak{D}}) + n^2/2(\mathfrak{D})} \right)^m \right] - \ln n \\ & \sim n \left[1 - \left(1 - \frac{\ln n + \beta_n}{nm + n\sqrt{nm(\ln n + \beta_n)} + (\ln n + \beta_n)(1 - n + n^2/2)} \right)^m \right] - \ln n, \end{aligned} \quad (4.6)$$

where \mathfrak{D} denotes $(\ln n + \beta_n)/nm$. Since

$$m \left(\frac{\ln n + \beta_n}{nm + n\sqrt{nm(\ln n + \beta_n)} + (\ln n + \beta_n)(1 - n + n^2/2)} \right)^2 \longrightarrow 0 \quad (4.7)$$

as $n \rightarrow \infty$, by Lemma 3.2, the right-hand side of (4.6)

$$\begin{aligned} & \sim \frac{nm(\ln n + \beta_n)}{nm + n\sqrt{nm(\ln n + \beta_n)} + (\ln n + \beta_n)(1 - n + n^2/2)} - \ln n \\ & = \frac{\beta_n - \ln n \sqrt{n(\ln n + \beta_n)/m} - \ln n(\ln n + \beta_n)(1/nm - 1/m + n/2m)}{1 + \sqrt{n(\ln n + \beta_n)/m} + (\ln n + \beta_n)(1/nm - 1/m + n/2m)} \\ & = \frac{\beta_n + o(1)}{1 + o(1)} \longrightarrow \beta, \end{aligned} \quad (4.8)$$

which concludes the proof. \square

Acknowledgment

The author would like to thank Karen Singer-Cohen for her warm encouragement and supplying a copy of [1].

References

- [1] K. B. Singer, *Random intersection graphs*, Ph.D. thesis, The Johns Hopkins University, Baltimore, Md, USA, 1995.
- [2] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen, "On random intersection graphs: the subgraph problem," *Combinatorics, Probability and Computing*, vol. 8, no. 1-2, pp. 131–159, 1999.
- [3] B. Bollobás, *Random Graphs*, vol. 73 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, UK, 2nd edition, 2001.
- [4] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, NY, USA, 2000.
- [5] S. R. Blackburn and S. Gerke, "Connectivity of the uniform random intersection graph," *Discrete Mathematics*, vol. 309, no. 16, pp. 5130–5140, 2009.
- [6] O. Yağan and A. M. Makowski, "Zero-one laws for connectivity in random key graphs," ISR Technical Report 2009-1, 2009, <http://hdl.handle.net/1903/8716>.
- [7] E. Godehardt and J. Jaworski, "Two models of random intersection graphs for classification," in *Exploratory Data Analysis in Empirical Research*, M. Schwaiger and O. Opitz, Eds., pp. 67–81, Springer, Berlin, Germany, 2003.
- [8] O. Yağan and A. M. Makowski, "On the random graph induced by a random key predistribution scheme under full visibility," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 544–548, Toronto, Canada, 2008.
- [9] J. A. Fill, E. R. Scheinerman, and K. B. Singer-Cohen, "Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models," *Random Structures & Algorithms*, vol. 16, no. 2, pp. 156–176, 2000.
- [10] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.
- [11] Y. Shang, "Typical vertex degrees in dense generalized random intersection graphs," *Mathematica Applicata*, vol. 23, no. 4, pp. 767–773, 2010.
- [12] D. Stark, "The vertex degree distribution of random intersection graphs," *Random Structures & Algorithms*, vol. 24, no. 3, pp. 249–258, 2004.
- [13] M. Behrisch, "Component evolution in random intersection graphs," *Electronic Journal of Combinatorics*, vol. 14, no. 1, p. R17, 2007.
- [14] Y. Shang, "Degree distributions in general random intersection graphs," *Electronic Journal of Combinatorics*, vol. 17, no. 1, p. R23, 2010.
- [15] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, vol. 2 of *Oxford Studies in Probability*, Oxford University Press, New York, NY, USA, 1992.
- [16] T. Lindvall, *Lectures on the Coupling Method*, Dover, Mineola, NY, USA, 2002.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

