GROUPIES IN RANDOM BIPARTITE GRAPHS

Yilun Shang

A vertex $v$ of a graph $G$ is called a groupie if its degree is not less than the average of the degrees of its neighbors. In this paper we study the influence of bipartition $(B_1, B_2)$ on groupies in random bipartite graphs $G(B_1, B_2, p)$ with both fixed $p$ and $p$ tending to zero.

1. INTRODUCTION

A vertex of a graph $G$ is called a groupie if its degree is not less than the arithmetic mean of the degrees of its neighbors. Some results concerning groupies have been obtained in deterministic graph theory; see e.g. [1, 5, 6]. Recently, Fernández de la Vega and Tuza [3] investigate groupies in Erdős-Rényi random graphs $G(n, p)$ and show that the proportion of the vertices which are groupies is almost always very near to $1/2$.

In this letter, we follow the idea of [3] and deal with groupies in random bipartite graph $G(B_1, B_2, p)$. Our results indicate the proportion of groupies depends on the bipartition $(B_1, B_2)$. First, we give a formal definition for $G(B_1, B_2, p)$ as follows.

Definition 1. A random bipartite graph $G(B_1, B_2, p)$ with vertex set $[n] = \{1, 2, \ldots, n\}$ is defined by partitioning the vertex set into two classes $B_1$ and $B_2$ and taking $p_{ij} = 0$ if $i, j \in B_1$ or $i, j \in B_2$, while $p_{ij} = p$ if $i \in B_1$ and $j \in B_2$ or vice versa. Here, independently for each pair $i, j \in [n]$, we add the edge $ij$ to the random graph with probability $p_{ij}$.

By convention, for a set $A$, let $|A|$ denote the number of elements in $A$. We denote by $\text{Bin}(m, q)$ the binomial distribution with parameters $m$ and $q$. 

2000 Mathematics Subject Classification. 05C80.

Keywords and Phrases. Groupie, random graph, bipartite graph, average degree.
2. MAIN RESULTS

Theorem 1. Suppose that $0 < p < 1$ is fixed. Let $N$ denote the number of groupies in the random bipartite graph $G(B_1, B_2, p)$. For $i = 1, 2$, let $N(B_i)$ denote the number of groupies in $B_i$. Then the following is true

(i) Assume $|B_1| = an$ and $|B_2| = (1 - a)n$ for some $a \in (0, 1)$. If $a = 1/2$, then

$$P\left( \frac{n}{4} - \omega(n)\sqrt{n} \leq N(B_i) \leq \frac{n}{4} + \omega(n)\sqrt{n}, \text{ for } i = 1, 2 \right) \to 1$$

as $n \to \infty$, where $\omega(n) = \Omega(\ln n)$. If $a < 1/2$, then

$$P\left( \frac{an}{2} - \omega(n)\sqrt{n} \leq N(B_1) \text{ and } N(B_2) \leq \frac{an}{2} + \omega(n)\sqrt{n} \right) \to 1$$

as $n \to \infty$, where $\omega(n)$ is defined as above.

(ii) Assume $|B_1| = bn$ and $|B_2| = (1 - b)n$ with $\ln n/n \ll 1 - b, n \to 0$, as $n \to \infty$. Then

$$P(N = N(B_2) = |B_2|) \to 1$$

as $n \to \infty$.

Proof. For (i) we take vertex $x \in B_1$ and let $d_x$ denote the degree of $x$ in $G(B_1, B_2, p)$. Denote by $S_x$ the sum of the degrees of the neighbors of $x$. Assuming that $x$ has degree $d_x$, we have $S_x \sim d_x + \text{Bin}((an - 1)d_x, p)$, where $\sim$ represents identity of distribution. For any $d_x$, the expectation of $S_x$ is $E S_x = d_x[1 + (an - 1)p]$. Since $S_x - d_x \sim \text{Bin}((an - 1)d_x, p)$ and $(an - 1)d_x \geq a(1-a)np/2$ when $(1-a)np/2 \leq d_x \leq 3(1-a)np/2$, by using a large deviation bound (see [4] pp.29, Remark 2.9), we get

$$P\left( |S_x - d_x anp| \leq 10n \sqrt{\ln n} \bigg| \frac{(1-a)np}{2} \leq d_x \leq \frac{3(1-a)np}{2} \right) \geq 1 - e^{-2\ln n} = 1 - o(n^{-1}).$$

Dividing by $d_x$, we have for some absolute constant $C_1 > 20/[(1-a)p]$

$$P\left( \frac{|S_x - d_x anp|}{d_x anp} \leq C_1 \sqrt{\ln n} \bigg| \frac{(1-a)np}{2} \leq d_x \leq \frac{3(1-a)np}{2} \right) = 1 - o(n^{-1}).$$

Note that $d_x \sim \text{Bin}((1-a)n, p)$ and a concentration inequality (see [4] pp.27, Corollary 2.3) yields

$$P\left( |d_x - (1-a)np| \leq \frac{(1-a)np}{2} \right) = 1 - o(n^{-1}).$$

Hence, recalling the total probability formula we obtain

$$P\left( \frac{|S_x - anp|}{d_x anp} \leq C_1 \sqrt{\ln n}, \text{ for every } x \in B_1 \right) = 1 - o(1).$$

(2)
Likewise,

\[ P \left( \frac{S_x}{d_x} - (1 - a)np \leq C_1 \sqrt{\ln n}, \text{ for every } x \in B_2 \right) = 1 - o(1). \]

We treat the following two scenarios separately.

**Case 1.** \( a = 1/2 \). For \( i = 1, 2 \), let \( N^+(B_i) \) (resp. \( N^-(B_i) \)) denote the number of vertices in \( B_i \), whose degrees are at least \( np/2 + C_1 \sqrt{\ln n} \) (resp. at most \( np/2 - C_1 \sqrt{\ln n} \)). From (2), (3) and the definition of a groupie, it follows that

\[ P \left( N^+(B_i) \leq N(B_i) \leq \frac{n}{2} - N^-(B_i), \text{ for } i = 1, 2 \right) = 1 - o(1). \]

Therefore, it suffices to prove

\[ P \left( N^+(B_1) \geq \frac{n}{4} - \omega(n) \sqrt{n} \right) = 1 - o(1) \]

and the analogous statements for \( N^+(B_2), N^-(B_1) \) and \( N^-(B_2) \).

Note that \( N^+(B_1) = \sum_{x \in B_1} 1_{\{d_x \geq np/2 + C_1 \sqrt{\ln n}\}} \). Since Bin \((n/2, p)\) is flat about its maximum, the expectation of \( N^+(B_1) \) is seen to be given by

\[ EN^+(B_1) = \frac{n}{2} P \left( d_x \geq \frac{np}{2} + C_1 \sqrt{\ln n} \right) = \frac{n}{4} - \Theta(\sqrt{n \ln n}). \]

Arguing as in [3], we derive \( Var(N^+(B_1)) \leq C_2 n \) for some absolute constant \( C_2 \) and then (4) follows by applying the Chebyshev inequality. Alternatively, we may deduce (4) by the bounded difference inequality (see [2] pp.24, Theorem 1.20) without estimating the variance.

**Case 2.** \( a < 1/2 \). Let \( \tilde{N}^+(B_1) \) denote the number of vertices in \( B_1 \) with degrees at least \((1 - a)np + C_1 \sqrt{\ln n}\). Hence \( \tilde{N}^+(B_1) = \sum_{x \in B_1} 1_{\{d_x \geq (1 - a)np + C_1 \sqrt{\ln n}\}} \), and reasoning similarly as in Case 1, we get

\[ P \left( N(B_1) \geq \frac{an}{2} - \omega(n) \sqrt{n} \right) \geq P \left( \tilde{N}^+(B_1) \geq \frac{an}{2} - \omega(n) \sqrt{n} \right) = 1 - o(1) \]

Next, let \( \tilde{N}^-(B_2) \) denote the number of vertices in \( B_2 \) with degrees at most \( anp - C_1 \sqrt{\ln n} \). Similarly, we have

\[ P \left( N(B_2) \leq \frac{an}{2} + \omega(n) \sqrt{n} \right) \geq P \left( \tilde{N}^-(B_2) \leq \frac{an}{2} + \omega(n) \sqrt{n} \right) = 1 - o(1) \]

We then conclude the proof in this case by combining (5) and (6). It is worth noting that the upper bound on \( N(B_1) \) and the lower bound on \( N(B_2) \) can not be obtained by using the above techniques.
For (ii) we need to prove the following two statements:
(a) Almost surely none of the vertices in $B_1$ is a groupie; and
(b) Almost surely every vertex in $B_2$ is a groupie.

In what follows we prove (a) only, as (b) may be proved similarly.

Fix a vertex $x \in B_1$ and assume that $x$ has degree $d_x$, we then have $S_x \sim d_x + \text{Bin}(b_n nd_x, p)$. For any $d_x$, the $\mathbb{E} S_x = d_x (b_n np + 1)$. Since $S_x - d_x \sim \text{Bin}(b_n nd_x, p)$ and $b_n nd_x \geq b_n (1 - b_n) n^2 p / 2$ when $(1 - b_n) np / 2 \leq d_x \leq 3(1 - b_n) np / 2$, as in situation (i) we obtain

$$P \left( \left| S_x - b_n nd_x p \right| \leq 10 n \sqrt{\ln n} \left| \frac{(1 - b_n) np}{2} \leq d_x \leq \frac{3(1 - b_n) np}{2} \right. \right) = 1 - o(n^{-1}).$$

Dividing by $d_x$ we have

$$P \left( \left| \frac{S_x}{d_x} - b_n np \right| \leq \frac{20 \sqrt{\ln n}}{(1 - b_n) p} \left| \frac{(1 - b_n) np}{2} \leq d_x \leq \frac{3(1 - b_n) np}{2} \right. \right) = 1 - o(n^{-1}).$$

Since $d_x \sim \text{Bin}((1 - b_n) n, p)$ and $\ln n / n \ll 1 - b_n$, we get

$$P \left( \left| d_x - (1 - b_n) np \right| \leq \frac{(1 - b_n) np}{2} \right) = 1 - o(n^{-1})$$

by exploiting a concentration inequality (see [4] pp.27, Corollary 2.3). From (7) and (8), it follows

$$P \left( \left| \frac{S_x}{d_x} - b_n np \right| \leq \frac{20 \sqrt{\ln n}}{(1 - b_n) p} \right) = 1 - o(n^{-1}).$$

We have

$$P \left( d_x \geq b_n np - \frac{20 \sqrt{\ln n}}{(1 - b_n) p} \right) \leq P \left( d_x - (1 - b_n) np \geq \frac{3}{2} \sqrt{(1 - b_n) n \ln n} \right)$$

$$\leq e^{-(3/2) \ln n} = o(n^{-1})$$

where the second inequality follows by an application of Theorem 2.1 of [4] (pp.26). Consequently, (9) and (10) yield

$$P(x \text{ is a groupie}) = o(n^{-1}),$$

which clearly concludes the proof of statement (a).

We remark that the assumption $\ln n / n \ll 1 - b_n$ given in Theorem 1 Case (ii) is not very stringent, since we must have $1 - b_n = \Omega(n^{-1})$ in our situation. The following theorem can be proved similarly.

**Theorem 2.** Suppose that $np^2 \gg \ln n$, as $n \to \infty$. Let $N$ denote the number of groupies in the random bipartite graph $G(B_1, B_2, p)$. For $i = 1, 2$, let $N(B_i)$ denote the number of groupies in $B_i$. Then the following is true:
(i) Assume $|B_1| = an$ and $|B_2| = (1 - a)n$ for some $a \in (0, 1)$. If $a = 1/2$, then
\[
P\left(\frac{n(1 - \varepsilon(n))}{4} \leq N(B_i) \leq \frac{n(1 + \varepsilon(n))}{4}, \text{ for } i = 1, 2\right) \to 1
\]
as $n \to \infty$, where $\varepsilon(n)$ is any function tending to zero sufficient slowly. If $a < 1/2$, then
\[
P\left(\frac{an(1 - \varepsilon(n))}{2} \leq N(B_1) \text{ and } N(B_2) \leq \frac{an(1 + \varepsilon(n))}{2}\right) \to 1
\]
as $n \to \infty$, where $\varepsilon(n)$ is defined as above.

(ii) Assume $|B_1| = b_n n$ and $|B_2| = (1 - b_n)n$ with $1 - b_n = \Omega(1/\sqrt{\ln n})$ and $b_n \to 1$, as $n \to \infty$. Then
\[
P(N = N(B_2) = |B_2|) \to 1
\]
as $n \to \infty$.

**Proof.** We sketch the proof as follows. For (i) the inequality (1) holds following the same reasoning as in the proof of Theorem 1. Therefore, we get
\[
P\left(\frac{S_x}{d_x} - anp \leq \frac{2\sqrt{\ln n}}{(1-a)p}, \text{ for every } x \in B_1\right) = 1 - o(1),
\]
The following two large deviation statements hold similarly:
\[
P\left(\frac{S_x}{d_x} - anp \leq \frac{2\sqrt{\ln n}}{(1-a)p}, \text{ for every } x \in B_1\right) = 1 - o(1),
\]
and
\[
P\left(\frac{S_x}{d_x} - (1-a)np \leq \frac{2\sqrt{\ln n}}{(1-a)p}, \text{ for every } x \in B_2\right) = 1 - o(1).
\]

**Case 1.** $a = 1/2$. For $i = 1, 2$, let $N^+(B_i)$ (resp. $N^-(B_i)$) denote the number of vertices in $B_i$, whose degrees are at least $np/2 + 20\sqrt{\ln n}/[(1-a)p]$ (resp. at most $np/2 - 20\sqrt{\ln n}/[(1-a)p]$). As in the proof of Theorem 1, in the sequel we shall prove that
\[
P\left(N^+(B_1) \geq \frac{n(1 - \varepsilon(n))}{4}\right) = 1 - o(1).
\]

Note that $N^+(B_1) = \sum_{x \in B_1} 1_{\{d_x \geq np/2 + 20\sqrt{\ln n}/[(1-a)p]\}}$. Since Bin $(n/2, p)$ is flat about its maximum, the expectation of $N^+(B_1)$ is given by
\[
EN^+(B_1) = \frac{n}{2} P\left(d_x \geq np/2 + 20\sqrt{\ln n}/(1-a)p\right) = \frac{n}{4} - \Theta\left(\frac{\sqrt{\ln n}}{p}\right).
\]
By using the assumption $np^2 \gg \ln n$, we may also obtain $Var(N^+(B_1)) \leq C_3 n$ for some absolute constant $C_3$. Since $\varepsilon(n)$ is a function tending to zero sufficient slowly,
we have $\sqrt{n \ln n} / p \ll \varepsilon n$ and $1/\varepsilon^2 n \to 0$, as $n \to \infty$. Combining these estimations, we get (11) by employing the Chebyshev inequality as in [3].

**Case 2.** $a < 1/2$. Let $\tilde{N}^+(B_1)$ denote the number of vertices in $B_1$ with degrees at least $(1 - a)np + 20\sqrt{\ln n} / ((1 - a)p)$ and the proof follows similarly as before.

For (ii), note that our assumptions imply $\ln n/n \ll 1 - b_n \to 0$ as $n \to \infty$, and the corresponding proof in Theorem 1 holds verbatim. □

**REFERENCES**


Department of Mathematics, Shanghai Jiao Tong University, 800 DongChuan Road, Shanghai, China
E-mail: shyl@sjtu.edu.cn