The Groupies in Random Multipartite Graphs

Marius Portmann

Department of Mathematics, Israel Institute of Technology, Haifa 32000, Israel.

Hongyun Wang

Department of Mathematics, Wuhan University of Technology, Wuhan, China.

1email: portmann@math.technion.ac.il

2email: hongyunwang@live.whut.edu.cn
1. Introduction

We say a vertex \( v \) in a graph \( G \) is a groupie if the degree of \( v \) is larger than the average degree of its neighbors [2, 5]. This interesting notion is related to the clustering of graphs [1]. Recently, in [3] and [6] the authors studied groupies in Erdős-Rényi random graphs \( G(n, p) \) and random bipartite graphs \( G(B_1, B_2, p) \), respectively. In particular, it is shown that the proportion of the vertices which are groupies is almost always very close to \( 1/2 \) [3].

In this paper, we show that similar reasoning in [6] actually can lead to further conclusion on random multipartite graphs. For simplicity, we will take a tripartite graph \( G(B_1, B_2, B_3, p) \) as an illustrating example. We define the graph model \( G(B_1, B_2, B_3, p) \) as follows.

**Definition 1.** A random tripartite graph \( G(B_1, B_2, B_3, p) \) with vertex set \( \{1, 2, \cdots, n\} \) is defined by partitioning the vertex set into three classes \( B_1, B_2 \) and \( B_3 \). The connection probability \( p_{ij} = 0 \) if \( i, j \in B_k \) for \( k = 1, 2, 3 \), and \( p_{ij} = p \) if \( i \in B_k \) and \( j \in B_t \) with \( k \neq t \). All edges are added independently.

We will give our main result in the following section.

2. Main result

For a set \( A \), let \( |A| \) be the number of elements in \( A \). Denote by \( Bin(m, q) \) the binomial distribution with parameters \( m \) and \( q \).

**Theorem 1.** Suppose that \( 0 < p < 1 \) is fixed. Let \( N \) be the number of groupies in the random tripartite graph \( G(B_1, B_2, B_3, p) \). For \( i = 1, 2, 3 \), let \( N(B_i) \) be the number of groupies in \( B_i \). Suppose that \( |B_1| = a_n n, \ |B_2| = b_n n \) and \( |B_3| = (1-a_n-b_n)n \) with \( a_n \to a \in (0, 1) \) and \( b_n \to b \in (0, 1) \) as \( n \to \infty \).
We have
\[ P\left( \frac{(a+b)n}{2} - \omega(n)\sqrt{n} \leq N(B_i) \leq \frac{(a+b)n}{2} + \omega(n)\sqrt{n}, \text{ for } i = 1, 2, 3 \right) \to 1 \]
as \( n \to \infty \).

**Proof.** We suppose that \( p = 1/2 \) and assume \( a_n \equiv a \in (0,1) \) and \( b_n \equiv b \in (0,1) \) for convenience.

For \( x \in B_1 \), let \( d_x \) be the degree of \( x \) in \( G(B_1,B_2,B_3,p) \). Denote by \( S_x \) the sum of the degrees of the neighbors of \( x \). Suppose that \( x \) has degree \( d_x \), we have \( S_x \sim d_x + Bin(((a+b)n - 1)d_x,p) \). For \( p = 1/2 \) and any \( d_x \), the expectation \( ES_x = d_x((a+b)n + 1)/2 \). Since \( S_x - d_x \sim Bin(((a+b)n - 1)d_x,1/2) \) and \( ((a+b)n - 1)d_x \geq (a+b)(1-a-b)n^2/4 \) when \( (1-a-b)n/4 \leq d_x \leq 3(1-a-b)n/4 \), by using large deviation bound [4], it is easy to see that
\[
P\left( \left| S_x - \frac{(a+b)n}{2} \right| \leq 10n\sqrt{\ln n} \right| \frac{(1-a-b)n}{4} \leq d_x \leq \frac{3(1-a-b)n}{4} \right) = 1 - e^{-2\ln n} = 1 - o(n^{-1}).
\]

Dividing by \( d_x \) we have
\[
P\left( \left| \frac{S_x}{d_x} - \frac{(a+b)n}{2} \right| \leq 50\sqrt{\ln n} \right| \frac{(1-a-b)n}{4} \leq d_x \leq \frac{3(1-a-b)n}{4} \right) = 1 - o(n^{-1}).
\]

Since \( d_x \sim Bin((1-a-b)n,1/2) \), we have by a concentration inequality [4] that
\[
P\left( \left| d_x - \frac{(1-a-b)n}{2} \right| \leq \frac{(1-a-b)n}{4} \right) = 1 - o(n^{-1}).
\]

It follows from the total probability formula that
\[
P\left( \left| \frac{S_x}{d_x} - \frac{(a+b)n}{2} \right| \leq 50\sqrt{\ln n}, \text{ for every } x \in B_1 \right) = 1 - o(1).
\]
Similarly, we have

\[
(2) \quad P\left( \left| \frac{S_x}{d_x} - \frac{(1-a-b)n}{2} \right| \leq 50\sqrt{\ln n}, \text{ for every } x \in B_2 \right) = 1 - o(1).
\]

and

\[
(3) \quad P\left( \left| \frac{S_x}{d_x} - \frac{(1-a-b)n}{2} \right| \leq 50\sqrt{\ln n}, \text{ for every } x \in B_3 \right) = 1 - o(1).
\]

For \( i = 1, 2, 3 \), let \( N^+(B_i) \) and \( N^-(B_i) \) denote the number of vertices in \( B_i \), whose degrees are larger than \( n/4 + 50\sqrt{\ln n} \) and less than \( n/4 - 50\sqrt{\ln n} \), respectively. By (1), (2) and the definition of groupie, we obtain

\[
P\left( N^+(B_i) \leq N(B_i) \leq \frac{n}{3} - N^-(B_i), \text{ for } i = 1, 2, 3 \right) = 1 - o(1).
\]

As in [6], we only need to prove

\[
(4) \quad P\left( N^+(B_1) \geq \frac{n}{3} - \omega(n)\sqrt{n} \right) = 1 - o(1)
\]

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

Note that \( N^+(B_1) = \sum_{i=1}^{n/2} 1_{[d_i \geq n/4 + 50\sqrt{\ln n}]} \), with \( d_i \) being the degree of vertex \( i \in B_1 \). Due to the form of \( Bin(n/2, 1/2) \), the expectation of \( N^+(B_1) \) is given by

\[
EN^+(B_1) = \frac{n}{2} P\left( d_i \geq \frac{n}{4} + 50\sqrt{\ln n} \right) = \frac{n}{3} - C_1\sqrt{n\ln n},
\]

where \( C_1 > 0 \) is an absolute constant. As in [3,4], we derive \( Var(N^+(B_1)) \leq C_2 n \) for an absolute constant \( C_2 \) and then (4) follows by applying the Chebyshev inequality.

Likewise, set \( \tilde{N}^+(B_1) \) denote the number of vertices in \( B_1 \) with degrees larger than \( (1-a)n/3 + 50\sqrt{\ln n} \). Therefore

\[
\tilde{N}^+(B_1) = \sum_{i=1}^{an} 1_{[d_i \geq (1-a)n/3 + 50\sqrt{\ln n}]}.
\]
and we obtain
\[
P\left( N(B_1) \geq \frac{(a+b)n}{2} - \omega(n)\sqrt{n} \right) \geq P\left( \tilde{N}^+(B_1) \geq \frac{(a+b)n}{2} - \omega(n)\sqrt{n} \right)
\]
(5)

\[
= 1 - o(1).
\]

Let \( \tilde{N}^-(B_2) \) denote the number of vertices in \( B_2 \) with degrees at most \( (a + b)n/2 - 50\sqrt{\ln n} \). Let \( \tilde{N}^-(B_3) \) denote the number of vertices in \( B_3 \) with degrees at most \( (a + b)n/2 - 50\sqrt{\ln n} \). We have

\[
P\left( N(B_2) \leq \frac{(a+b)n}{3} + \omega(n)\sqrt{n} \right)
\]
\[
\geq P\left( \frac{n}{2} - \tilde{N}^-(B_2) \leq \frac{(a+b)n}{3} + \omega(n)\sqrt{n} \right)
\]

(6)
\[
= 1 - o(1).
\]

and

\[
P\left( N(B_3) \leq \frac{(a+b)n}{3} + \omega(n)\sqrt{n} \right)
\]
\[
\geq P\left( \frac{n}{2} - \tilde{N}^-(B_3) \leq \frac{(a+b)n}{3} + \omega(n)\sqrt{n} \right)
\]

(7)
\[
= 1 - o(1).
\]

We finished the proof by using (5), (6) and (7). □

**References**

[1] C. C. Aggarwal, H. Wang, A survey of clustering algorithms for graph data. *Advances in Database Systems*, 40(2010) pp.275–301

[2] E. Bertram, P. Erdős, P. Horáčk, J. Širáň, Zs. Tuza, Local and global average degree in graphs and multigraphs. *Journal of Graph Theory*, 18(1994) pp.647–661

[3] W. Fernandez de la Vega, Zs. Tuza, Groupies in random graphs. *Information Processing Letters*, 109(2009) pp.339–340

[4] S. Janson, T. Łuczak, A. Rucinski, *Random Graphs*. Wiley, New York, 2000
[5] J. Mackey, A lower bound for groupies in graphs. *Journal of Graph Theory*, 21(1996) pp.323–326

[6] Y. Shang, Groupies in random bipartite graphs. *Applicable Analysis and Discrete Mathematics*, 4(2010) pp.278–283