**SpringerPlus**

# Groupies in multitype random graphs

Yilun Shang[*]

*Correspondence:
shyl@tongji.edu.cn
Department of Mathematics,
Tongji University,
Shanghai 200092, China

## Abstract

A groupie in a graph is a vertex whose degree is not less than the average degree of its neighbors. Under some mild conditions, we show that the proportion of groupies is very close to 1/2 in multitype random graphs (such as stochastic block models), which include Erdős-Rényi random graphs, random bipartite, and multipartite graphs as special examples. Numerical examples are provided to illustrate the theoretical results.

**Keywords:** Random graph, Degree, Groupie, Multitype

**Mathematics Subject Classification:** 05C07, 05C80

## Background

A vertex in a graph $G$ is said to be a *groupie* if its degree is not less than the average degree of its neighbors. Various properties of groupies have been investigated in deterministic graph theory (Ajtai et al. 1980; Bertram et al. 1994; Ho 2007; Mackey 1996; Poljak et al. 1995). For example, it was proved in Mackey (1996) that there are at least two groupies in any simple graphs with at least two vertices. Groupies were even found to be related to Ramsey numbers (Ajtai et al. 1980). More recently, Fernandez de la Vega and Tuza (2009) showed that, in Erdős-Rényi random graphs $G(n, p)$, the proportion of vertices that are groupies is almost always very near to 1/2 as $n \to \infty$. Later the author Shang (2010) obtained a result of similar flavor in random bipartite graphs $G(n_1, n_2, p)$. It was shown that the proportion of groupies in each partite set is almost always very close to 1/2 if $G(n_1, n_2, p)$ is balanced, namely, $n_1 = n_2$.

In this paper, we consider groupies in a more general random graph model, which we call *multitype random graphs*. Let $q$ be a positive integer. Denote $[q] := \{1, \dots, q\}$. Define the 'gene' for a multitype random graph as a weighted complete graph $K_q$ (having a loop at each vertex) on the vertex set $[q]$, with a weight $\alpha_i > 0$ associated to each vertex, and a weight $0 \le \beta_{ij} \le 1$ associate to each edge $ij$. Note that $\beta_{ij} = \beta_{ji}$ since we deal with undirected graphs. We assume $\sum_{i=1}^{q} \alpha_i = 1$. The multitype random graph $G(n, K_q)$ with gene $K_q$ is generated as follows. Let $n$ be much larger than $q$, and let $[n]$ be its vertex set. We partition $[n]$ into $q$ sets $V_1, \dots, V_q$ by putting vertex $v$ in $V_i$ with probability $\alpha_i$ independently. Each pair of vertices $v \in V_i$ and $u \in V_j$ are connected with probability $\beta_{ij}$ independently (all the decisions on vertices and edges are made independently).

For $i = 1, \dots, q$, let $N_i$ represent the number of the groupies in $V_i$. Thus, $N := \sum_{i=1}^{q} N_i$ is the number of groupies in the multitype random graph $G(n, K_q)$. Denote by

$\alpha = (\alpha_i) \in \mathbb{R}^q$ and $\beta = (\beta_{ij}) \in \mathbb{R}^{q \times q}$. For generality, we will usually think of $\beta$ and $\alpha$ as functions of $n$ in the same spirit of random graph theory (Bollobás 2001; Janson et al. 2000). Let $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^q$ be the all-one vector. All the asymptotic notations used in the paper such as $O$, $o$, and $\Omega$ are standard, see e.g. Janson et al. (2000). Our first result is as follows.

**Theorem 1** *Let $q \geq 2$. Assume that $\beta\alpha = (\theta + o(\sqrt{\ln n}/n))\mathbf{1}$, where $\theta > 0$ is a constant. If $\min_{i \neq j}\{\alpha_i, \beta_{ij}\} > c$ for some constant $c > 0$, and $\max_i\{\beta_{ii}\} = o(\sqrt{\ln n}/n)$, then*

$$\mathrm{P}\left(\frac{\alpha_i n}{2} - \omega(n)\sqrt{n} \leq N_i \leq \frac{\alpha_i n}{2} + \omega(n)\sqrt{n}, \text{ for } i = 1, \ldots, q\right) \to 1$$

*as $n \to \infty$, where $\omega(n) = \Omega(\ln n)$ is any function tending to infinity. Hence,*

$$\mathrm{P}\left(\frac{n}{2} - \omega(n)\sqrt{n} \leq N \leq \frac{n}{2} + \omega(n)\sqrt{n}\right) \to 1$$

*as $n \to \infty$, where $\omega(n) = \Omega(\ln n)$ is any function tending to infinity.*

When $\beta$ and $\alpha$ are independent of $n$, the following corollary is immediate.

**Corollary 1** *Let $q \geq 2$. Assume that $\beta\alpha = \theta\mathbf{1}$ for $\theta > 0$, and $\beta_{ii} = 0$ for all $i$. Then*

$$\mathrm{P}\left(\frac{\alpha_i n}{2} - \omega(n)\sqrt{n} \leq N_i \leq \frac{\alpha_i n}{2} + \omega(n)\sqrt{n}, \text{ for } i = 1, \ldots, q\right) \to 1$$

*as $n \to \infty$, where $\omega(n) = \Omega(\ln n)$ is any function tending to infinity.*

Clearly, by taking $q = 2$, $\alpha_1 = \alpha_2 = 1/2$, and $\beta_{11} = \beta_{22} = 0$, we recover the result in Shang (2010, Thm. 1) for balanced random bipartite graphs.

Theorem 1 requires that the edges between sets $V_i$, $i = 1, \ldots, q$ are dense, namely, the multitype random graph $G(n, K_q)$ in question resembles a dense 'multipartite' graph. For sparse random graphs on the other hand, we have the following result.

**Theorem 2** *Let $q \geq 1$. Assume that $\beta\alpha = (\theta + o(\sqrt{\ln n}/n))\mathbf{1}$, where $\theta = \theta(n)$ is a function of $n$. If $\min_i\{\alpha_i\} > c$ for some constant $c > 0$, $\min_{i \neq j}\{\beta_{ij}\} \gg (\ln n)^2/n$, and $\max_i\{\beta_{ii}\} = o(\sqrt{\ln n}/n)$, then*

$$\mathrm{P}\left(\frac{\alpha_i n(1 - \varepsilon(n))}{2} \leq N_i \leq \frac{\alpha_i n(1 + \varepsilon(n))}{2}, \text{ for } i = 1, \ldots, q\right) \to 1$$

*as $n \to \infty$, where $\varepsilon(n) = \Omega(\ln n/\sqrt{n})$ is any function tending to zero. Hence,*

$$\mathrm{P}\left(\frac{n(1 - \varepsilon(n))}{2} \leq N \leq \frac{n(1 + \varepsilon(n))}{2}\right) \to 1$$

*as $n \to \infty$, where $\varepsilon(n) = \Omega(\ln n/\sqrt{n})$ is any function tending to zero.*

It follows from Theorem 2 that we may reproduce the result for sparse Erdős-Rényi random graphs Fernandez de la Vega and Tuza (2009, Thm. 2) by taking $q = \alpha_1 = 1$,

$\beta_{11} = o(\sqrt{\ln n}/n)$; and the result for sparse balanced random bipartite graphs Shang (2010, Thm. 2) by taking $q = 2$, $\alpha_1 = \alpha_2 = 1/2$, $\beta_{11} = \beta_{22} = 0$ and $\beta_{12} \gg (\ln n)^2/n$.

The multitype random graph $G(n, K_q)$ is generated through a double random process. In the following, we will also consider a closely related 'random-free' model $G'(n, K_q)$. Given a gene $K_q$ defined as above, the *random-free multitype random graph* $G'(n, K_q)$ (a.k.a. stochastic block model Holland et al. 1983) is constructed by partitioning $[n]$ into $q$ sets $V_1, \ldots, V_q$ with $|V_i| = \alpha_i n$. Recall that $\sum_{i=1}^q \alpha_i = 1$. We draw an edge $vu$ with probability $\beta_{ij}$ independently for $v \in V_i$ and $u \in V_j$; thus the first random step in the original construction disappears, which explains the name 'random-free'.

In "Proof of the main results" section, we will show Theorems 1 and 2 by first proving analogous results for the random-free version $G'(n, K_q)$. To illustrate our theoretical results, a numerical example is presented in "Numerical simulations" section.

## Proof of the main results

**Proposition 1**  *Theorem 1 holds verbatim for the random-free model $G'(n, K_q)$.*

*Proof*  Without loss of generality, we consider $i = 1$, other values of $i$ being completely similar. Take vertex $v \in V_1$ and denote by $d_v$ the degree of $v$ in $G'(n, K_q)$. Therefore, $d_v = \sum_{i=1}^q d_i$, where $d_i$ means the number of neighbors of $v$ in $V_i$. Let $S_v$ represent the sum of degrees of the neighbors of $v$. Write $\mathrm{Bin}(n, p)$ for a Binomial variable with parameters $n$ and $p$. Assuming that $v$ has degree $d_v$, we obtain

$$
\begin{aligned}
S_v \sim d_v &+ 2 \sum_{i=1}^q \mathrm{Bin}\left( \frac{d_i(d_i - 1)}{2}, \beta_{ii} \right) + 2 \sum_{1 \le i < j \le q} \mathrm{Bin}(d_i d_j, \beta_{ij}) \\
&+ \sum_{j=1}^q \mathrm{Bin}\big( d_j(\alpha_1 n - d_1 - 1), \beta_{1j} \big) + \sum_{i=2}^q \sum_{j=1}^q \mathrm{Bin}\big( d_j(\alpha_i n - d_i), \beta_{ij} \big),
\end{aligned}
\tag{1}
$$

where the second and third terms on the right-hand side evaluate the contribution of degrees within the neighborhood, while the last two terms correspond to the sum of out-going degrees. Here, $\sim$ means identity of distribution by convention.

For any $d_v$, the expectation of $S_v$ can be computed as

$$
\mathrm{E}S_v = d_v + \sum_{i=1}^q d_i^2 \beta_{ii} - \sum_{i=1}^q d_i \beta_{ii} - \sum_{j=1}^q d_j \beta_{1j} + \sum_{i=1}^q \sum_{j=1}^q d_j \alpha_i n \beta_{ij}.
\tag{2}
$$

It follows from the assumption $\max_i\{\beta_{ii}\} = o(\sqrt{\ln n}/n)$ and the reverse Cauchy–Schwarz inequality Pólya and Szegö (1972, p. 71) that $\sum_{i=1}^q d_i^2 \beta_{ii} = o(d_v^2 \sqrt{\ln n}/n)$. Using $\beta\alpha = (\theta + o(\sqrt{\ln n}/n))\mathbf{1}$ and the symmetry of $\beta$, we obtain $\sum_{i=1}^q \sum_{j=1}^q d_j \alpha_i n \beta_{ij} = \sum_{j=1}^q d_j \sum_{i=1}^q \alpha_i n \beta_{ji} = d_v(\theta n + o(\sqrt{\ln n}))$. Consequently, (2) becomes $\mathrm{E}S_v = d_v \theta n + \Theta(d_v) + o(d_v^2 \sqrt{\ln n}/n) + o(d_v \sqrt{\ln n})$. Define the event $\mathcal{A}_v = \{\alpha_1 \beta_{11} n - (\ln n)\sqrt{n\beta_{11}} \le d_1 \le \alpha_1 \beta_{11} n + (\ln n)\sqrt{n\beta_{11}}, \; \alpha_i \beta_{1i} n - (\ln n)\sqrt{n} \le d_i \le \alpha_i \beta_{1i} n + (\ln n)\sqrt{n}, \text{ for } i = 2, \cdots, q\}$. Set $\Phi = \sum_{i=1}^q d_i(d_i - 1)/2 + \sum_{i<j} d_i d_j + \sum_{j=1}^q d_j(\alpha_1 n - d_1 - 1) + \sum_{i=2}^q \sum_{j=1}^q d_j(\alpha_i n - d_i)$. In view of

(1), the distribution of $S_v - d_v$ is identical to that of the sum of $\Phi$ independent random variables, each of which is bounded above by 2. This number is $\Theta(n^2)$ when the event $\mathcal{A}_v$ occurs. Thus, the large deviation bound Janson et al. (2000, p. 29) gives

$$P(|S_v - d_v\theta n| \leq 10qn\sqrt{\ln n}\,|\mathcal{A}_v) \geq P(|S_v - d_v - E(S_v - d_v)| \leq 8qn\sqrt{\ln n}\,|\mathcal{A}_v)$$
$$\geq 1 - e^{-3\ln n} = 1 - o(n^{-1}).$$

Dividing by $d_v$ and noting that $q \geq 2$, we obtain for any constant $C_1 \geq \frac{11}{c^2}$

$$P\left(\left|\frac{S_v}{d_v} - \theta n\right| \leq C_1\sqrt{\ln n}\,\Big|\mathcal{A}_v\right) = 1 - o(n^{-1}).$$

Furthermore, it is straightforward to check that the event $\mathcal{A}_v$ holds with probability $1 - o(n^{-1})$ using the Chernoff bound Janson et al. (2000, p. 27) and the fact $d_1 \sim \text{Bin}(\alpha_1 n - 1, \beta_{11})$ and $d_i \sim \text{Bin}(\alpha_i n, \beta_{1i})$, $i = 2, \cdots, q$. Therefore, an application of the total probability formula yields

$$P\left(\left|\frac{S_v}{d_v} - \theta n\right| \leq C_1\sqrt{\ln n}, \text{ for every } v \in V_1\right) = 1 - o(1). \tag{3}$$

Now denote by $N_1^+$ the number of vertices in $V_1$, whose degrees are at least $\theta n + C_1\sqrt{\ln n}$. Similarly, denote by $N_1^-$ the number of vertices in $V_1$, whose degrees are at most $\theta n - C_1\sqrt{\ln n}$. The estimation (3) implies that

$$P(N_1^+ \leq N_1 \leq \alpha_1 n - N_1^-) = 1 - o(1),$$

where we recall the definition of $N_1$ as the number of groupies in $V_1$. To complete the proof, it suffices to show

$$P\left(N_1^+ \geq \frac{\alpha_1 n}{2} - \omega(n)\sqrt{n}\right) = 1 - o(1) \tag{4}$$

and the analogous statement for $N_1^-$, where $\omega(n) = \Omega(\ln n)$ is any function tending to infinity.

We write $N_1^+$ as the sum of indicators, namely, $N_1^+ = \sum_{v \in V_1} 1_{\{d_v \geq \theta n + C_1\sqrt{\ln n}\}}$. Notice that $d_v \sim \text{Bin}(\alpha_1 n - 1, \beta_{11}) + \sum_{i=2}^{q} \text{Bin}(\alpha_i n, \beta_{1i})$ is a sum of independent binomial variables. Since $d_v$ is flat around its maximum (Butler and Stephens 1993; Drezner and Farnum 2007), we obtain

$$EN_1^+ = \alpha_1 n \cdot P(d_v \geq \theta n + C_1\sqrt{\ln n}) = \frac{\alpha_1 n}{2} - \Theta(\sqrt{n \ln n}).$$

Based on the bounded difference inequality (see e.g. Bollobás (2001, p. 24) with the difference $c_k \equiv 1$), we obtain for any $\omega(n) = \Omega(\ln n)$,

$$P\left(N_1^+ \leq \frac{\alpha_1 n}{2} - \omega(n)\sqrt{n}\right) = P(N_1^+ \leq EN_1^+ - \omega'(n)\sqrt{n}) \leq e^{-\frac{2\omega'^2(n)n}{n}} = o(1),$$

where $\omega'(n)$ is a function tending to infinity as $n \to \infty$. This proves (4). Following the same reasoning we can show $P(N_1^- \geq \alpha_1 n/2 - \omega(n)\sqrt{n}) = o(1)$, which concludes the proof. $\qquad\square$

**Proposition 2** *Theorem 2 holds verbatim for the random-free model $G'(n, K_q)$, except that we herein allow $\varepsilon(n) = \Omega(\sqrt{\ln n/n})$ as any function tending to zero.*

*Proof* We sketch the proof as it is similar. As in the proof of Proposition 1, we consider $i = 1$ and obtain the expectation of $S_v$ for $v \in V_1$ as $ES = d_v\theta n + \Theta(d_v) + o(d_v^2\sqrt{\ln n}/n) + o(d_v\sqrt{\ln n})$. Define the event $\mathcal{B}_v = \{\alpha_i\beta_{1i}n - (\ln n)\sqrt{n\beta_{1i}} \le d_i \le \alpha_i\beta_{1i}n + (\ln n)\sqrt{n\beta_{1i}}$, for $i = 1, \ldots, q\}$. The Chernoff bound Janson et al. (2000, p. 27) implies that $\mathcal{B}_v$ holds with probability $1 - o(n^{-1})$. Using the large deviation bound Janson et al. (2000, p. 29) we obtain

$$P\left(|S_v - d_v\theta n| \le 10q\ln^{\frac{5}{2}} n \,\middle|\, \mathcal{B}_v\right) \ge P\left(|S_v - d_v - E(S_v - d_v)| \le 8q\ln^{\frac{5}{2}} n \,\middle|\, \mathcal{B}_v\right)$$
$$\ge 1 - e^{-2\ln n} = 1 - o(n^{-1}).$$

Dividing by $d_v$, we obtain similarly for some constant $C_2 > 0$

$$P\left(\left|\frac{S_v}{d_v} - \theta n\right| \le C_2\sqrt{\ln n} \,\middle|\, \mathcal{B}_v\right) = 1 - o(n^{-1}),$$

and

$$P\left(\left|\frac{S_v}{d_v} - \theta n\right| \le C_2\sqrt{\ln n}, \text{ for every } v \in V_1\right) = 1 - o(1). \tag{5}$$

Denote by $N_1^+$ the number of vertices in $V_1$, whose degrees are at least $\theta n + C_2\sqrt{\ln n}$. Denote by $N_1^-$ the number of vertices in $V_1$, whose degrees are at most $\theta n - C_2\sqrt{\ln n}$. The result (5) again implies that

$$P(N_1^+ \le N_1 \le \alpha_1 n - N_1^-) = 1 - o(1).$$

It remains to show

$$P\left(N_1^+ \ge \frac{\alpha_1 n(1 - \varepsilon(n))}{2}\right) = 1 - o(1) \tag{6}$$

and the analogous statement for $N_1^-$, where $\varepsilon(n) = \Omega(\sqrt{\ln n/n})$ is any function tending to zero.

Set $N_1^+ = \sum_{v \in V_1} 1_{\{d_v \ge \theta n + C_2\sqrt{\ln n}\}}$. As in the proof of Proposition 1, we arrive at

$$EN_1^+ = \alpha_1 n \cdot P(d_v \ge \theta n + C_2\sqrt{\ln n}) = \frac{\alpha_1 n}{2} - \Theta(\sqrt{n\ln n}).$$

Invoking the bounded difference inequality Bollobás (2001, p. 24) and taking $\omega(n) := \varepsilon(n)\sqrt{n} \to \infty$, we obtain for any $\varepsilon(n) = \Omega(\sqrt{\ln n/n})$,

$$P\left(N_1^+ \le \frac{\alpha_1 n(1 - \varepsilon(n))}{2}\right) = P(N_1^+ \le EN_1^+ - \omega(n)\sqrt{n}) \le e^{-2\omega^2(n)} = o(1),$$

as $n \to \infty$. This completes the proof of (6). Likewise, we have $P(N_1^- \geq \alpha_1 n (1 - \varepsilon(n))/2) = o(1)$ as desired. □

*Proof of Theorem 1 and Theorem 2*  These results can be proven in the similar way as Propositions 1 and 2 by noting that, in the $G(n, K_q)$ model, $P(||V_i| - \alpha_i n| \geq \ln n \sqrt{n}) = o(n^{-1})$ for all $i = 1, \ldots, q$. □

## Numerical simulations

To illustrate our theoretical results, in this section we present a numerical example for the $G(n, K_q)$ model with $q = 3$.

Set $\alpha = (0.45, 0.35, 0.2)^T$, $\beta = \begin{pmatrix} 0 & 8/21 & 1/3 \\ 8/21 & 0 & 1/7 \\ 1/3 & 1/7 & 0 \end{pmatrix}$, and $\theta = 0.2$. In Fig. 1 we plot the

numbers of groupies $N_i$ for $i = 1, 2, 3$ as functions of $n$, (i) with the above constant $\beta$; and (ii) with perturbed $\beta + \Delta\beta$, where $\Delta\beta = (\ln^{1/4} n)/n \mathbf{1}\mathbf{1}^T$. Clearly, the conditions in Theorem 1 hold for both situations (i) and (ii). Fig. 1 shows that the agreement between the simulations and the theoretical prediction of Theorem 1 is excellent.

## Conclusion

In this paper, we have studied the groupies in multitype random graphs. It is discovered that the proportion of groupies is very close to 1/2 in multitype random graphs, which include Erdős-Rényi random graphs, random bipartite, and multipartite graphs as special examples. We mention that there are several possibilities to continue this line of research, both by considering other more realistic random network models as well as by analyzing the limit distribution of groupies in random graphs. For example, a natural question could be to ask if there are similar results for $q = q(n)$ or edge-independent random graphs (e.g. Shang 2016)?
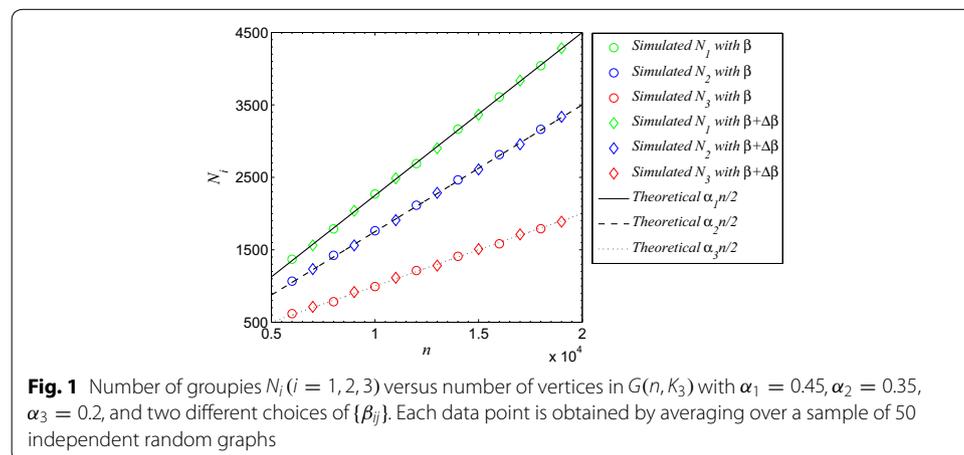


**Fig. 1** Number of groupies $N_i$ ($i = 1, 2, 3$) versus number of vertices in $G(n, K_3)$ with $\alpha_1 = 0.45$, $\alpha_2 = 0.35$, $\alpha_3 = 0.2$, and two different choices of $\{\beta_{ij}\}$. Each data point is obtained by averaging over a sample of 50 independent random graphs

**References**
Ajtai M, Komlós J, Szemerédi E (1980) A note on Ramsey numbers. J Comb Theory Ser A 29:354–360
Bertram E, Erdős P, Horák P, Širáň J, Tuza Zs (1994) Local and global average degree in graphs and multigraphs. J Graph Theory 18:647–661
Bollobás B (2001) Random graphs. Cambridge University Press, Cambridge
Butler K, Stephens M (1993) The distribution of a sum of binomial random variables. Technical Report No. 467. Department of Statistics, Stanford University
Drezner Z, Farnum N (2007) A generalized binomial distribution. Commun Stat Theory Methods 22:3051–3063
Fernandez de la Vega W, Tuza Zs (2009) Groupies in random graphs. Inform Process Lett 109:339–340
Ho PT (2007) On groupies in graphs. Aust J Combin 38:173–177
Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. Soc Netw 5:109–137
Janson S, Luczak T, Ruciński A (2000) Random graphs. Wiley, New York
Mackey J (1996) A lower bound for groupies in graphs. J. Graph Theory 21:323–326
Pólya G, Szegö G (1972) Problems and theorems in analysis, vol I: series, integral calculus, theory of functions. Translated from the German by D. Aeppli Die Grundlehren der mathematischen Wissenschaften, Band 193. Springer, New York
Poljak S, Szabó T, Tuza Zs (1995) Extremum and convergence of local average degrees in graphs. Congr Numer 112:191–198
Shang Y (2010) Groupies in random bipartite graphs. Appl Anal Discrete Math 4:278–283
Shang Y (2016) Bounding extremal degrees of edge-independent random graphs using relative entropy. Entropy 18:art. 53