The probability that a random multigraph is simple

Svante Janson

If $n \geq 1$ and $(d_i)_1^n$ is a sequence of non-negative integers, we let $G(n, (d_i)_1^n)$ be the random (simple) graph with the n vertices $1, \dots, n$, and with vertex degrees $d_1, \dots, d_n$, uniformly chosen among all such graphs (provided that there are any such graphs at all; in particular, $\sum_i d_i$ has to be even). A standard method to study $G(n, (d_i)_1^n)$ is to consider the related random multigraph $G^*(n, (d_i)_1^n)$ defined by taking a set of d_i *half-edges* at each vertex i and then joining the half-edges into edges by taking a random partition of the set of all half-edges into pairs (we tacitly assume that $\sum_i d_i$ is even). This is known as the configuration model, and such a partition of the half-edges is known as a *configuration*; this was introduced by Bollobás [2], see also Section II.4 of [3]. (See Bender and Canfield [1] and Wormald [11, 12] for related arguments.)

We obtain $G(n, (d_i)_1^n)$ by conditioning $G^*(n, (d_i)_1^n)$ on being a simple graph. It is then of crucial importance to be able to estimate the probability that $G^*(n, (d_i)_1^n)$ is simple, and in particular to decide whether

$$\liminf_{n\to\infty} \mathbb{P}\bigl(G^*(n, (d_i)_1^n) \text{ is simple}\bigr) > 0 \tag{1}$$

for given sequences $(d_i)_1^n = (d_i^{(n)})_1^n$ (depending on $n \geq 1$). (Note that (1) implies that any statement holding for $G^*(n, (d_i)_1^n)$ with probability tending to 1 does so for $G(n, (d_i)_1^n)$ too.)

A natural condition that has been used by several authors using the configuration method (including myself [6]) as a sufficient condition for (1) is

$$\sum_{i=1}^n d_i = \Theta(n) \qquad \text{and} \qquad \sum_{i=1}^n d_i^2 = O(n) \tag{2}$$

together with some bound on $\max_i d_i$. (Recall that $A = \Theta(B)$ means that both $A = O(B)$ and $B = O(A)$ hold.) Results showing, or implying, that (2) and a condition on $\max_i d_i$ imply (1) have also been given by several authors, for example Bender and Canfield [1] with $\max_i d_i = O(1)$; Bollobás [2], see also Section II.4 in [3], with $\max_i d_i \leq \sqrt{2 \log n} - 1$; McKay [9] with $\max_i d_i = o(n^{1/4})$; McKay and Wormald [10] with $\max_i d_i = o(n^{1/3})$. (Some of these papers give sharp estimates of the probability that $G^*(n, (d_i)_1^n)$ is simple also when (2) does not hold.) Similar results have also been proved for bipartite graphs [8], digraphs [5], and hypergraphs [4].

Indeed, it is not difficult to see that the method used by Bollobás [2, 3] works, assuming (2), provided only $\max_i d_i = o(n^{1/2})$. This has undoubtedly been noted by several experts, but we have not been able to find a reference to it in print when we have needed one.

One of our main result is that, in fact, (2) is sufficient for (1) without any assumption on $\max_i d_i$. Moreover, (2) is essentially necessary.

Let N be the number of edges in $G^*(n,(d_i)_1^n)$. Thus

$$2N = \sum_{i=1}^{n} d_i. \tag{3}$$

It turns out that it is more natural to state our results in terms of N than n (the number of vertices).

We consider a sequence of random multigraphs $G^*_\nu = G^*\big(n_\nu,(d_i^{(\nu)})_1^n\big)$ and consider asymptotics as $\nu \to \infty$, but for notational simplicity we will omit the index ν. (In typical application, the graphs are indexed by n, the number of vertices; moreover, typically $N = \Theta(n)$, and then N can be replaced by n in the conditions below.) We can state our first result as follows.

Theorem 1. *Let $N := \frac{1}{2}\sum_i d_i$ be the number of edges in $G^*(n,(d_i)_1^n)$, and assume that $N \to \infty$. Then*

(i) $\liminf \mathbb{P}(G^*(n,(d_i)_1^n)$ *is simple*$) > 0$ *if and only if* $\sum_i d_i^2 = O(N)$;
(ii) $\lim \mathbb{P}(G^*(n,(d_i)_1^n)$ *is simple*$) = 0$ *if and only if* $\sum_i d_i^2/N \to \infty$.

Our second main result is an asymptotic formula for the probability that $G^*(n,(d_i)_1^n)$ is simple.

Theorem 2. *Consider $G^*(n,(d_i)_1^n)$ and assume that $N := \frac{1}{2}\sum_i d_i \to \infty$. Let $\lambda_{ij} := \sqrt{d_i(d_i-1)d_j(d_j-1)}/(2N)$; in particular $\lambda_{ii} = d_i(d_i-1)/(2N)$. Then*

$$\mathbb{P}\big(G^*(n,(d_i)_1^n) \text{ is simple}\big) = \exp\Big(-\tfrac{1}{2}\sum_i \lambda_{ii} - \sum_{i<j}\big(\lambda_{ij} - \log(1+\lambda_{ij})\big)\Big) + o(1);$$

equivalently,

$$\begin{aligned}&\mathbb{P}\big(G^*(n,(d_i)_1^n) \text{ is simple}\big) \\ &= \exp\Big(-\frac{1}{4}\Big(\frac{\sum_i d_i^2}{2N}\Big)^2 + \frac{1}{4} + \frac{\sum_i d_i^2(d_i-1)^2}{16N^2} + \sum_{i<j}\big(\log(1+\lambda_{ij}) - \lambda_{ij} + \tfrac{1}{2}\lambda_{ij}^2\big)\Big) \\ &\qquad + o(1).\end{aligned}$$

In the case $\max_i d_i = o(N^{1/2})$, Theorem 2 simplifies as follows.

Corollary 3. *Assume that $N \to \infty$ and $\max_i d_i = o(N^{1/2})$. Let*

$$\Lambda := \frac{1}{2N}\sum_{i=1}^{n}\binom{d_i}{2} = \frac{\sum_i d_i^2}{4N} - \frac{1}{2}.$$

Then

$$\begin{aligned}\mathbb{P}\big(G^*(n,(d_i)_1^n) \text{ is simple}\big) &= \exp\big(-\Lambda - \Lambda^2\big) + o(1) \\ &= \exp\Big(-\frac{1}{4}\Big(\frac{\sum_i d_i^2}{2N}\Big)^2 + \frac{1}{4}\Big) + o(1).\end{aligned}$$

This formula is well known, at least under stronger conditions on $\max_i d_i$, see, for example, Bender and Canfield [1], Bollobás [3, Theorem II.16], McKay [9] and McKay and Wormald [10, Lemma 5.1]. In this case, one can use the method by Bollobás [2, 3] and show by the method of moments that the number of loops plus the number of pairs of parallel edges is asymptotically Poisson distributed with mean $\Lambda + \Lambda^2$, which yields a direct proof of Corollary 3.

To prove the theorems under the weaker condition $\max_i d_i = O(N^{1/2})$, this method fails, because this number no longer has to be approximatively Poisson distributed; for example, two vertices with degree $N^{1/2}$ will have roughly a Poisson distributed number of edges between them, and thus the number of pairs of such edges will be more like the square of a Poisson variable. Instead we count vertices with at least one loop and pairs of vertices with at least two edges between them, disregarding the number of parallel loops or edges. We show that the indicators that a vertex or a pair of vertices is bad in this sense are asymptotically independent. This uses a more complicated conditioning argument; we would like to condition on the event that a certain pair of vertices is bad, but instead we condition on the event that it has k edges given by specific pairs of half-edges, which is easier to study; by an inclusion-exclusion type argument, we then get the necessary estimates for the conditioning we want.

Details are given in [7].

References

[1] E. A. Bender & E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, **24** (1978), no. 3, 296–307.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Comb.* **1** (1980), 311–316.

[3] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge Univ. Press, Cambridge, 2001.

[4] C. Cooper, The cores of random hypergraphs with a given degree sequence. *Random Structures Algorithms* **25** (2004), no. 4, 353–375.

[5] C. Cooper & A. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.* **13** (2004), no. 3, 319–337.

[6] S. Janson & M. Luczak, A simple solution to the k-core problem. *Random Structures Algorithms*, to appear. `http://arxiv.org/math.CO/0508453`

[7] S. Janson, The probability that a random multigraph is simple. `http://arxiv.org/math.CO/0609802`

[8] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and design (Waterloo, Ont., 1982)*, pp. 225–238, Academic Press, Toronto, ON, 1984.

[9] B. D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combin.* **19A** (1985), 15–25.

[10] B. D. McKay & N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica* **11** (1991), no. 4, 369–382.

[11] N. C. Wormald, Some problems in the enumeration of labelled graphs. Ph. D. thesis, University of Newcastle, 1978.

[12] N. C. Wormald, The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** (1981), no. 2, 168–182.