

# A probabilistic approach to block sizes in random maps

Louigi Addario-Berry

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, Québec, H3A 0B9, Canada

E-mail address: [louigi.addario@mcgill.ca](mailto:louigi.addario@mcgill.ca)

URL: <http://problab.ca/louigi>

**Abstract.** We present a probabilistic approach to block sizes in random maps, which yields straightforward and singularity analysis-free proofs of some results of [Bender et al. \(1995\)](#); [Banderier et al. \(2001\)](#); [Gao and Wormald \(1999\)](#). The proof also yields joint convergence in distribution of the rescaled size of the  $k$ 'th largest 2-connected block in a large random map, for any fixed  $k \geq 2$ , to a vector of Fréchet-type extreme order statistics. This seems to be a new result even when  $k = 2$ .

## 1. Introduction

The paper [Banderier et al. \(2001\)](#) is reasonably called the culmination of an extended line of research into core sizes in large random planar maps. The paper is an analytic *tour de force*, proceeding via singularity analysis of generating functions and the coalescing saddlepoint method. [Banderier et al. \(2001\)](#) demonstrate how this powerful set of tools can be used to derive to local limit theorems and sharp upper and lower tail estimates. In particular, their theorems unify and strengthen the results from [Bender et al. \(1995\)](#) and [Gao and Wormald \(1999\)](#).

The purpose of this note is to explain a probabilistic approach to the study of large blocks in large random maps. We end up proving two results. One is a weakening of ([Banderier et al., 2001](#), Theorem 7), the other a strengthening of ([Banderier et al., 2001](#), Proposition 5). The main point, though, is that our approach, which is to reduce the problem to a question about outdegrees in conditioned Galton-Watson trees, feels direct and probabilistically natural (and short). A related technique for studying various observables of “decomposable” random combinatorial structures, using Boltzmann samplers, was introduced in [Panagiotou and Weißl \(2007\)](#). We discuss the relation between our approach and that of [Panagiotou and Weißl \(2007\)](#) in Section 3.

The remainder of the introduction lays out the definitions required for the remainder of the work. Section 2 recalls Tutte’s compositional approach to planar map enumeration [Tutte \(1963\)](#), and describes an associated tree decomposition of maps into higher connectivity submaps. Randomness finally arrives in Section 3, which also contains the statements and proofs of this work’s proposition, corollary, and theorem.

1.1. *Notation for maps and trees.* We refer the reader to [Lando and Zvonkin \(2004\)](#) for a careful treatment of maps on surfaces, but provide all the definitions we directly require. In this work, a (plane) map  $M$  is a planar graph  $(v(M), e(M))$  properly embedded in the sphere  $\mathbb{S}^2$ , and considered up to orientation-preserving homeomorphisms of  $\mathbb{S}^2$ . Here  $v(M)$  and  $e(M)$  are the vertices and edges of  $M$ , respectively. All maps in this work are plane, and we hereafter omit this adjective. We also write  $\bar{e}(M)$  for the set of oriented edges of map  $M$ .

---

*Received by the editors December 13, 2018.*

1991 *Mathematics Subject Classification.* Primary: 60C05. Secondary: 05C10, 05C30.

*Key words and phrases.* Random maps, extreme value theory, Tutte decomposition, condensation, stable processes.

We say a map  $M'$  is a submap of map  $M$  if  $M'$  may be obtained from  $M$  by removal of a subset of the vertices and a subset of the edges of  $M$ . Any subgraph of  $(v(M), e(M))$  induces a submap of  $M$ , and conversely any submap of  $M$  is induced by a subgraph of  $(v(M), e(M))$ . Note that the faces of a submap need not be faces of the original map.

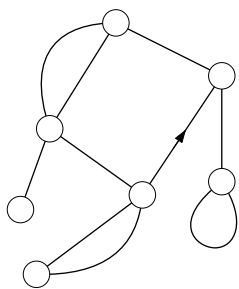
A *rooted* map is a pair  $M = (M, \rho)$ , where  $M$  is a planar map and  $\rho = \rho^- \rho^+$  is an oriented edge of  $M$  with tail  $\rho^-$  and head  $\rho^+$ . We view  $M$  as embedded in  $\mathbb{R}^2$  so that the unbounded face lies to the right of  $\rho$ ; this in particular gives meaning to the “interior” and “exterior” for cycles of  $M$  (see Figure 1.1a). When convenient we write  $v(M)$ , etcetera, instead of  $v(M)$ . The *size* of a map is its number of edges; map  $M$  is larger than map  $M'$  if  $|e(M)| \geq |e(M')|$ . The *trivial* map is the map with one vertex and no edges. We root the trivial map at its unique vertex for notational convenience.

A *plane tree* is a connected rooted map  $T = (T, \rho)$  with no cycles. We refer to  $\rho^-$  as the root of  $T$ . Children and parents are then defined in the usual way. The *outdegree* of  $v \in v(T)$  is the number of children of  $v$  in  $T$ .

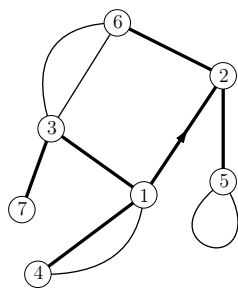
We require an ordering rule for the oriented edges of an arbitrary rooted map  $M = (M, \rho)$ . Any fixed rule would do, but for concreteness we describe a specific total order  $\prec_M$  of  $\bar{e}(M)$ . Write  $<_M$  for the total order of the *vertices*  $v(M)$  induced by a breadth first search starting from  $\rho^-$  using the counterclockwise order of edges around a vertex to determine exploration priority (see Figure 1.1b). Listing the vertices according to this order as  $v_1, v_2, \dots, v_{|v(M)|}$ , we in particular have  $v_1 = \rho^-, v_2 = \rho^+$ . We sometimes refer to  $<_M$  as lexicographic order.

Breadth-first search builds a spanning tree  $F = F(M)$  of  $M$  rooted at  $v_1 = \rho^-$ : for each  $v \neq \rho^-$ , the parent  $p(v)$  of  $v$  in  $F$  is the  $<_M$ -minimal neighbour  $w$  of  $v$ . (There may be multiple edges of  $M$  joining a node  $w$  to a child  $v$  of  $w$ , but only one of these is an edge of  $F$ ; here is how to determine which. If  $w = \rho^- = v_1$  then take the first copy of each edge leaving  $w$  in counterclockwise order around  $w$  starting from  $\rho = \rho^- \rho^+$ . If  $w \neq \rho^-$  then take the first copy of each edge leaving  $w$  in counterclockwise order starting from  $w p(w)$ ; this makes sense inductively since  $p(w) <_M w$ .)

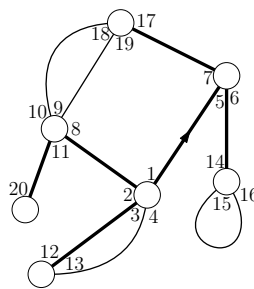
A *corner* of  $M$  is a pair  $(uv, uw)$  of oriented edges, where  $uw$  is the successor of  $uv$  in counterclockwise order around  $v$ . It is useful to identify oriented edges with corners: the corner corresponding to  $uw$  is the corner lying to the left of its tail. This is a bijective correspondence. We define the total order  $\prec_M$  on the set of corners (equivalently, the set of oriented edges) of  $M$  as follows (see Figure 1.1c): say  $uv \prec_M u'v'$  if either (a)  $u <_M u'$  or (b)  $u = u'$  and  $uv$  precedes  $u'v'$  in counterclockwise order around  $u$  starting from  $u p(u)$  (or, if  $u = v_1 = \rho^-$ , starting from  $\rho$ ).



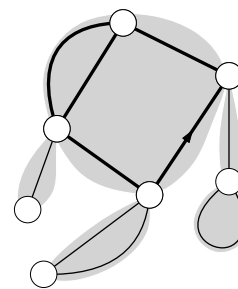
(A) A map  $M = (M, \rho)$ . The root edge  $\rho$  is drawn pointing from  $\rho^-$  to  $\rho^+$ .



(B) The breadth-first search tree of  $M$  has bold edges. Vertices are labelled in increasing order according to  $<_M$ .



(C) The oriented edges/corners are labelled in increasing order according to  $\prec_M$ .



(D) The blocks of  $M$  are shaded, and the root block has bold edges.

## 2. Planar maps as composite structures

We say a rooted map  $M$  is *separable* if there is a way to partition  $e(M)$  into nonempty sets  $E$  and  $E'$  so that there is exactly one vertex  $v$  incident to edges of both  $E$  and  $E'$ . If  $M$  is not separable

it is called *2-connected*.<sup>1</sup> Write  $\mathcal{M}$  for the set of rooted maps, and  $\mathcal{B}$  for the set of 2-connected rooted maps. Tutte (1963) showed how to count 2-connected maps by decomposing general maps into 2-connected submaps, then using Lagrange inversion. The remainder of the section presents this decomposition. We carefully define the tree structure associated to the decomposition, which is not explicitly used by Tutte, as it plays a key role in Section 3.

The maximal 2-connected submaps of  $M$  are called the *blocks* of  $M$  (hence the notation  $\mathcal{B}$ ). They are edge-disjoint, and have a natural tree structure associated to them; see Figure 1.1d. Write  $B = B(M)$  for the maximal 2-connected submap of  $M$  containing  $\rho$ ; call  $B$  the *root block*.

For each oriented edge  $uv$  of  $B$ , there is a (possibly trivial) unique maximal submap of  $M$  disjoint from  $B$  except at  $u$  and lying to the left of  $uv$ . We denote this map  $M_{uv} = (M_{uv}, \rho_{uv})$ , and call it the *pendant submap at  $uv$*  (or at the corresponding corner of  $B$ ). When  $M_{uv}$  is non-trivial,  $\rho_{uv}$  is the edge of  $M$  following  $uv$  in counterclockwise order around  $u$ . See Figure 2.2 for an illustration. We may reconstruct  $M$  from  $B$  and the  $2|e(B)|$  submaps  $\{M_{uv}, \{u, v\} \in e(B)\}$  by identifying the tail of the root edge of  $M_{uv}$  with  $u \in v(M)$  in such a way that the root edge of  $M_{uv}$  lies to the left of  $uv$ .

Compositionally, we thereby obtain that *rooted maps are 2-connected maps of rooted maps*. To formalize this, let  $\mathcal{M}_n$  (resp.  $\mathcal{C}_n$ ) be the set of rooted maps (resp. rooted 2-connected maps) with  $n$  edges, and write  $M_n = |\mathcal{M}_n|$ ,  $C_n = |\mathcal{C}_n|$ . We take  $C_0 = 1 = M_0$ . Then with  $M(z) = \sum_{n \geq 0} M_n z^n$  and  $C(z) = \sum_{n \geq 0} C_n z^n$ , we have (see Tutte (1963), equation (6.3))

$$M(z) = C(zM(z)^2). \quad (2.1)$$

Now, introduce a formal variable  $y$  with  $y^2 = z$ . Then with  $h(y) = yM(y^2) = z^{1/2}M(z)$ , by (2.1) we have  $h(y) = yC(h(y)^2)$  so, by Lagrange inversion,

$$[z^n]M(z) = [y^{2n+1}]h(y) = \frac{1}{2n+1} [y^{2n}]C(y)^{2n+1}.$$

Here is the combinatorial interpretation of this identity. Given a map  $M = (M, \rho)$ , represent the block structure of  $M$  by the following plane tree  $T_M$  defined as follows. (The construction is illustrated in Figure 2.3.) Let  $B = (B, \rho)$  be the block containing  $\rho$ , and list the *oriented* edges  $\bar{e}(B)$  according to the order  $\prec_B$  as  $a_1, \dots, a_{2|e(B)|}$ . We say that the root  $\emptyset$  of  $T_M$  represents  $B$  in  $T_M$ .

The node  $\emptyset$  has  $2|e(B)|$  children in  $T_M$ . List them from left to right as  $1, \dots, 2|e(B)|$ . Fix  $i \in \{1, \dots, 2|e(B)|\}$ . If the counterclockwise successor  $e_i = e_i^- e_i^+$  of  $a_i$  around  $a_i^-$  in  $M$  is also in  $\bar{e}(B)$  then the corner formed by  $a_i$  and  $e_i$  contains no pendant submap. In this case  $i$  is a leaf in  $T_M$ . Otherwise,  $e_i \in \bar{e}(M) \setminus \bar{e}(B)$ . In this case write  $M_i$  for the connected component of  $(v(M), e(M) \setminus e(B))$  containing  $\{e_i^-, e_i^+\}$ , and let  $M_i = (M_i, e_i)$ . The subtree of  $T_M$  rooted at  $i$  is recursively defined to be the tree  $T_{M_i}$ . Figure 2.3a and 2.3c show a map  $M$  and a schematic representation of its blocktree. Figure 2.3b shows the corresponding tree  $T_M$ .

If  $M$  is 2-connected then  $T_M$  is simply a root of outdegree  $2|e(M)|$  whose children are all leaves. More generally, for each block  $B$  of  $M$ , there is a corresponding node of  $T_M$  with exactly  $2|e(B)|$  children. In other words, given the tree  $T_M$ , the block sizes in  $M$  are known.

Given the map  $B_\rho$ , the map  $M$  may be reconstructed by identifying  $e_i^-$  (the tail of the root edge of  $M_i = (M_i, e_i)$ ) and  $a_i^-$  so that  $e_i$  follows  $a_i$  in counterclockwise order around  $a_i^-$ . (This was explained in the paragraph preceding (2.1).) It follows recursively that  $M$  is uniquely specified by  $T_M$  together with the set of maps  $(B_v, v \in v(T_M))$ , where  $B_v$  is the block of  $M$  represented by  $v$  in  $T_M$ . If  $v$  is a leaf, take  $B_v$  to be the trivial map. Note that every node  $v$  has precisely  $2|e(B_v)|$  children in  $T_M$ , so  $|e(T_M)| = 2|e(M)|$ . For the map  $M$  from Figure 2.3a, the nontrivial blocks represented by nodes of  $T_M$  are shown with identifying labels in Figure 2.3d.

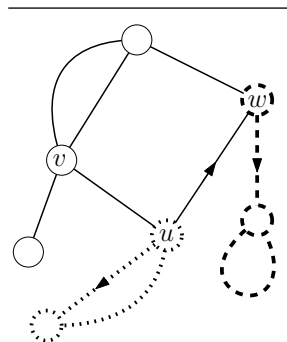


FIGURE 2.2.  $M_{uv}$  and  $M_{wu}$  are respectively dotted and dashed.

<sup>1</sup>The terminology of graphs and of maps are slightly at odds here. Many graph theorists would consider the “lollipop” graph, with one loop and one non-loop edge, to be 2-connected. As a map, it is not.

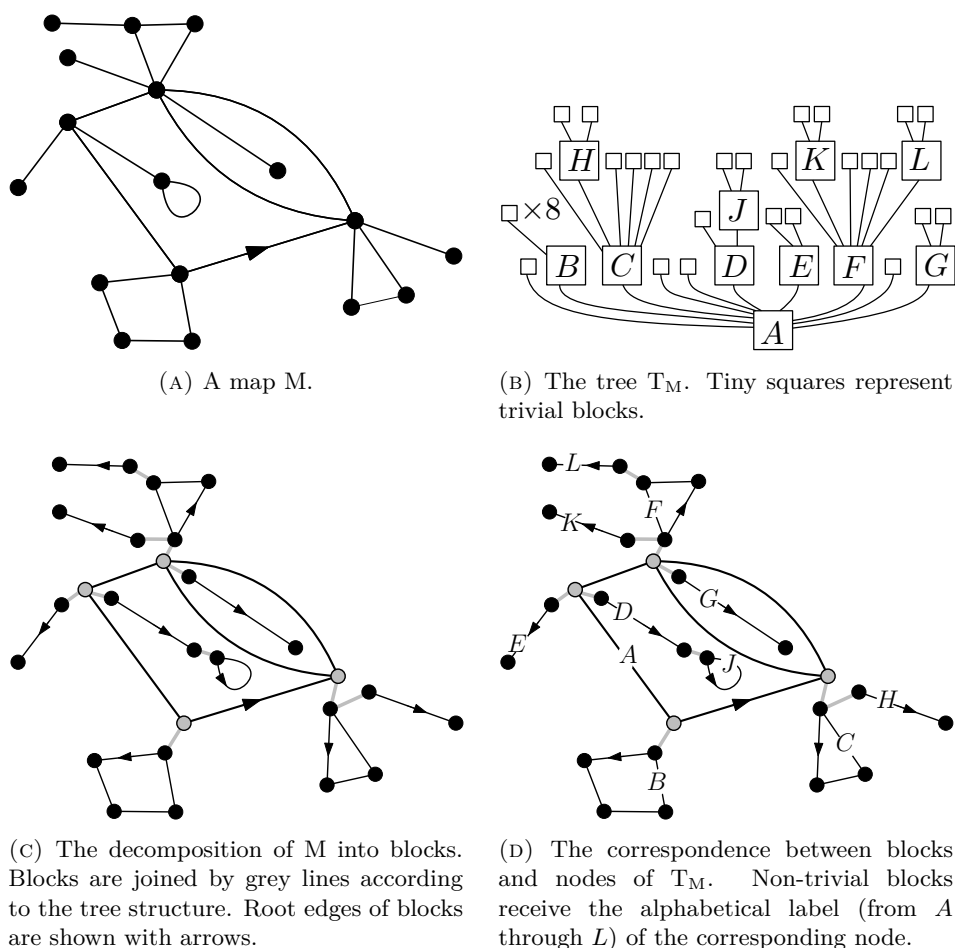


FIGURE 2.3. The relation between a map  $M$  and the plane tree  $T_M$ .

### 3. Random maps

Let  $M_n \in_u \mathcal{M}_n$ ; this notation means that  $M_n$  is a random variable uniformly distributed over the (finite) set  $\mathcal{M}_n$ . We now describe the law of the tree  $T_{M_n}$ . Recall that  $M_n = |\mathcal{M}_n|$  and  $C_n = |\mathcal{C}_n|$ , and that

$$M_n = \frac{2 \cdot 3^n (2n)!}{(n+2)!n!}.$$

Using this, the compositional equation (2.1), and a little thought (see [Tutte \(1963\)](#), Section 6 or [Goulden and Jackson \(2004\)](#), pages 152-153), Lagrange inversion yields

$$C_0 = 1, \quad C_k = \frac{2(3k-3)!}{k!(2k-1)!} \quad \text{for } k \geq 1. \quad (3.1)$$

(The formulas for  $M_n$  and  $C_n$  are due to [Tutte \(1963\)](#); see also [Brown \(1963\)](#).) Using Stirling's approximation, the formula (3.1) for  $C_k = |\mathcal{C}_k|$  implies that  $C(z)$  has radius of convergence  $4/27$ . Furthermore, it is straightforward to calculate that  $C(4/27) = 4/3$ , and that  $\widehat{C}(4/27) := \sum_{k \geq 0} k(4/27)^k \cdot C_k = 4/9$ . The fact  $C(4/27)$  is finite is used straightaway; the second identity is noted for later use.

Fix  $z \in (0, 4/27]$  and define a probability measure  $\mu^z$  on the non-negative integers by

$$\mu^z(\{2k\}) = \frac{C_k z^k}{C(z)}.$$

Let  $T^z$  be a Galton-Watson tree with offspring distribution  $\mu^z$ , and let  $T_n^z$  be a random tree whose law is that of  $T^z$  conditional on  $|e(T^z)| = 2n$ .

**Proposition 3.1.** *For all  $z \in (0, 4/27]$ , the trees  $T_n^z$  and  $T_{M_n}$  have the same law.*

*Proof of Proposition 3.1:* Fix a rooted plane tree  $t$  with  $2n$  edges, and list the outdegrees in  $t$  in lexicographic order as  $d_1, \dots, d_{2n+1}$ ; we assume all these are even. We saw in Section 2 that a map  $M$  is uniquely specified by the tree  $T_M$  together with 2-connected maps  $(B_i, 1 \leq i \leq 2n)$ , where  $B_i$  has  $d_i/2$  edges. It follows that the number of maps  $M$  with  $T_M = t$  is precisely

$$m(t) = \prod_{i=1}^{2n} C_{\frac{d_i}{2}}.$$

Therefore,  $\mathbf{P}\{T_{M_n} = t\}$  is proportional to  $m(t)$ . It is easily seen that this is also true for  $\mathbf{P}\{T_n^z = t\}$  whatever the value of  $z \in (0, 4/27]$ .  $\square$

For the remainder of the section, let  $(X_i, i \geq 1)$  be iid with law  $\mu$ , and write  $S_k = \sum_{i=1}^k X_i$ . Now write  $\mu = \mu^{4/27}$  and  $T_n = T_n^{4/27}$ .

**Corollary 3.2.** *List the outdegrees in  $T_n$  as in lexicographic order as  $D_1, \dots, D_{2n+1}$ , and let  $\sigma$  be a uniformly random cyclic shift of  $\{1, \dots, 2n+1\}$ . Then the conditional law of  $(X_1, \dots, X_{2n+1})$  given that  $S_{2n+1} = 2n$  is precisely that of  $(D_{\sigma(1)}, \dots, D_{\sigma(2n+1)})$ .*

*Proof:* This follows immediately from Proposition 3.1 and the cycle lemma (Pitman, 2006, Lemma 6.1).  $\square$

The corollary allows statistics about block sizes in  $M_n$  to be deduced by studying a sequence of iid random variables conditioned on its sum. Pitman (2006) explains a quite general link between probabilistic analysis of composite structures and randomly stopped sums; he calls this *Kolchin's representation of Gibbs partitions*. In a sense, the point of this note is to place the study of block sizes in maps within the latter framework.

We now state our main and only theorem. Let  $A$  be a Stable(3/2) random variable, characterized by its Laplace transform:

$$\mathbf{E}[e^{-tA}] = e^{\Gamma(-3/2)t^{3/2}} = e^{(4\pi^{1/2}/3)t^{3/2}}.$$

This distribution is also called a *map-Airy* distribution. The above scaling is used for the map-Airy distribution in (cite banderier et al); a similar scaling is used for general stable laws in (cite feller vol 2 pages 581-583). Also, let  $(G_k, k \geq 1)$  be the ordered atoms of a rate one Poisson point process on  $[0, \infty)$ , so  $G_k$  is  $\Gamma(k)$ -distributed.

**Theorem 3.3.** *Let  $M_n \in_u \mathcal{M}_n$ , and for  $k \geq 1$  let  $L_{n,k}$  be the number of edges in the  $k$ 'th largest block of  $M_n$ . Then as  $n \rightarrow \infty$ ,*

$$\frac{n/3 - L_{n,1}}{2^{7/6}/(27\pi)^{1/2}n^{2/3}} \xrightarrow{d} A,$$

and, jointly with the previous convergence, for any fixed  $k$ , as  $n \rightarrow \infty$ ,

$$\left( \frac{L_{n,j}}{(2/3)^{5/3}\pi^{-1/3}n^{2/3}}, 2 \leq j \leq k \right) \xrightarrow{d} \left( G_{j-1}^{-3/2}, 2 \leq j \leq k \right).$$

Before proving the theorem, we introduce a small amount of notation. Given sequences  $(Y_n)$  and  $(Z_n)$  of random vectors, write  $Y_n \stackrel{d}{\approx} Z_n$  if  $d_{TV}(Y_n, Z_n) \rightarrow 0$  as  $n \rightarrow \infty$ , where  $d_{TV}$  is total variation distance. Also, for a random vector  $Y$  and an event  $E$ , we write  $(Y|E)$  for a random vector whose law is the conditional law of  $Y$  given that  $E$  occurs.

*Proof of Theorem 3.3:* We begin with some straightforward facts about the the random variables  $(X_i, i \geq 1)$ . The values of  $C(4/27)$  and  $\widehat{C}(4/27)$  imply that  $\mathbf{E}X_1 = \sum_{j \geq 0} 2j\mu(\{2j\}) = 2/3$ . Furthermore, as  $j \rightarrow \infty$ , by Stirling's formula we have

$$\mu(\{2j\}) \sim \left( \frac{8}{27\pi} \right)^{1/2} j^{-5/2}.$$

Writing  $c = (\frac{8}{27\pi})^{1/2}$ , it thus follows from (Feller, 1971, Theorem XVII.5.2) that as  $m \rightarrow \infty$ ,

$$\frac{S_m - 2m/3}{cm^{2/3}} \xrightarrow{d} A. \tag{3.2}$$

Next, for  $m \geq 1$  let  $X^{m,1}, \dots, X^{m,m}$  be the decreasing rearrangement of  $X_1, \dots, X_m$ . Then by classic results in extreme value theory (see, e.g., (Leadbetter et al., 1983, Section 2.2)), or by a straightforward computation, it follows that for any fixed  $k$ ,

$$\left(\frac{3/2}{cm}\right)^{2/3} (X^{m,i}, i \leq k) \xrightarrow{d} (G_i^{-2/3}, i \geq k). \quad (3.3)$$

Now list the blocks of  $M_n$  in decreasing order of size (number of edges) as  $C_1, \dots, C_K$ , breaking ties arbitrarily, so that  $L_{n,k} = |e(C_k)|$ . By Proposition 3.1, the sequence  $(2L_{n,k}, 1 \leq k \leq K)$  has the same law as the decreasing rearrangement of non-zero outdegrees in  $T_n$ . By Corollary 3.2, it follows that for all  $i$  and  $k$  we have

$$\mathbf{P}\{L_{n,k} = i\} = \mathbf{P}\{X^{2n+1,k} = 2i \mid S_{2n+1} = 2n\}, \quad (3.4)$$

The large values in such collections of conditioned random variables have been studied in detail by Janson (2012). Many of the results in Janson (2012) are phrased in terms of statistics of random balls-into-boxes configurations; the connection between this and outdegrees in conditioned Galton-Watson trees is made explicit in (Janson, 2012, Section 8). One of the themes running through that work is that of *condensation*: for heavy-tailed random variables, conditioning a sum  $S_m$  to be large is often equivalent to conditioning on having a single exceptionally large summand. (See Armendáriz and Loulakis (2011); Ferrari et al. (2007); Kortchemski (2015) for other instances of this phenomenon in related settings.)

In (Janson, 2012, Theorem 19.34), Janson provides several results regarding conditional distributions such as that in (3.4). Recalling the notation introduced just before the proof, the specific result from that theorem which we use is that

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid S_{2n+1} = 2n) \approx (2n - S_{2n}, X^{2n,1}, \dots, X^{2n,2n}).$$

For expository purposes, we include a proof of this result (in Proposition 3.4, below), which closely follows that in Janson (2012).

Together with (3.2) and (3.4), the asymptotic distributional equivalence of the first coordinate above implies that

$$\frac{n/3 - L_{n,1}}{2^{-1/3}cn^{2/3}} \approx \frac{S_{2n} - 2(2n/3)}{c(2n)^{2/3}} \xrightarrow{d} A.$$

The first convergence follows since  $2^{-1/3}c = 2^{7/6}/(27\pi)^{1/2}$ . Similarly, using the above asymptotic distributional equivalence together with (3.3) and (3.4) yields that for any fixed  $k \geq 2$ ,

$$\left(\frac{3/2}{2cn}\right)^{2/3} (L_{n,i}, 2 \leq i \leq k) \approx \left(\frac{3/2}{2cn}\right)^{2/3} (X_{2n}(i), 1 \leq i \leq k-1) \xrightarrow{d} (G_i^{-2/3}, 1 \leq i \leq k-1),$$

which completes the proof since  $(3/(4c))^{2/3} = (3/2)^{5/3}\pi^{1/3}$ .  $\square$

**Proposition 3.4.** *As  $n \rightarrow \infty$ ,*

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid S_{2n+1} = 2n) \approx (2n - S_{2n}, X^{2n+1,1}, \dots, X^{2n+1,2n}).$$

We first state and prove an auxiliary lemma, before proving Proposition 3.4.

**Lemma 3.5.** *Fix a decreasing sequence  $(\delta_n)$  with  $\delta_n \rightarrow 0$  slowly. Let*

$$E_n = \{S_{2n+1} = 2n, |X^{2n+1,1} - 2n/3| < \delta_n n, X^{2n+1,2} < n/10\}.$$

*If  $\delta_n \rightarrow 0$  sufficiently slowly then  $\mathbf{P}\{E_n \mid S_{2n+1} = 2n\} \rightarrow 1$ .*

*Proof:* Recall that the  $X_i$  are iid with  $\mathbf{P}\{X_i = 2m\} \sim cm^{-5/2}$  and  $\mathbf{E}X_i = 2/3$ .

Write  $N = \#\{i \leq 2n+1 : X_i \geq n/10\}$ . By symmetry, if  $N = 1$  then each entry of  $(X_i, 1 \leq i \leq 2n+1)$  is equally likely to be the unique maximum. Also, for  $n$  large, if  $|X_i - 2n/3| < \delta_n n$  then  $X_i \geq n/10$ . Provided  $\delta_n \rightarrow 0$  sufficiently slowly, by the law of large numbers,

$\mathbf{P}\{|S_{2n} - 4n/3| < \delta_n n\} \rightarrow 1$ , so

$$\begin{aligned} \mathbf{P}\{E_n\} &= \mathbf{P}\{S_{2n+1} = 2n, |X^{2n+1,1} - 2n/3| < \delta_n n, N = 1\} \\ &= (2n+1)\mathbf{P}\{S_{2n+1} = 2n, |X_{2n+1} - 2n/3| < \delta_n n, N = 1\} \\ &\geq (2n+1)\mathbf{P}\{S_{2n+1} = 2n, |X_{2n+1} - 2n/3| \leq \delta_n n\} \\ &\geq (2n+1)\mathbf{P}\{|S_{2n} - 4n/3| \leq \delta_n n\} \cdot \inf_{m: |m-2n/3| \leq \delta_n n} \mathbf{P}\{X_{2n+1} = m\} \\ &\geq Cn^{-3/2}, \end{aligned}$$

for an absolute constant  $C > 0$ .

In view of this lower bound on  $\mathbf{P}\{E_n\}$ , in order to prove the lemma it suffices to establish that  $\mathbf{P}\{S_{2n+1} = 2n, E_n^c\} = o(n^{-3/2})$ . We first bound the probability that  $S_{2n+1} = 2n$  and  $N = 1$  but  $E_n$  does not occur:

$$\begin{aligned} &\mathbf{P}\{S_{2n+1} = 2n, N = 1, |X^{2n+1,1} - 2n/3| \geq \delta_n n\} \\ &= (2n+1)\mathbf{P}\{S_{2n+1} = 2n, X^{2n+1,2} < n/10, X_{2n+1} \geq n/10, |X_{2n+1} - 2n/3| \geq \delta_n n\} \\ &= (2n+1) \sum_{m \geq n/10: |m-2n/3| \geq \delta_n n} \mathbf{P}\{X_{2n+1} = m\} \mathbf{P}\{S_{2n} = 2n - m\} \\ &\leq (2n+1)\mathbf{P}\{|S_{2n} - 4n/3| \geq \delta_n n\} \sup_{m \geq n/10: |m-2n/3| \geq \delta_n n} \mathbf{P}\{X_{2n+1} = m\} \\ &= o(n^{-3/2}), \end{aligned}$$

the last bound holding since  $\mathbf{P}\{|S_{2n} - 4n/3| \geq \delta_n n\} \rightarrow 0$ .

It remains to prove that  $\mathbf{P}\{S_{2n+1} = 2n, N \neq 1\} = o(n^{-3/2})$ . The case  $N \geq 2$  is simpler: since  $\mathbf{P}\{X_1 \geq m\} = O(m^{-3/2})$ ,

$$\mathbf{P}\{S_{2n+1} = 2n, N \geq 2\} \leq \mathbf{P}\{N \geq 2\} \leq \binom{2n+1}{2} \mathbf{P}\{X_1 > n/10, X_2 \geq n/10\} = O(n^{-3}).$$

In order to bound  $\mathbf{P}\{S_{2n+1} = 2n, N = 0\}$ , write  $X'_i = X_i \mathbf{1}_{[X_i < n/10]}$ , and  $S' = \sum_{i \leq 2n+1} X'_i$ . Then for any  $t > 0$ , by Markov's inequality and the independence of the  $X'_i$ ,

$$\mathbf{P}\{S_{2n+1} = 2n, N = 0\} = \mathbf{P}\{S' = 2n\} \leq e^{-2nt} \mathbf{E}e^{tS'} = e^{-2nt} \cdot \left(\mathbf{E}e^{tX'_1}\right)^{2n+1}.$$

We apply this with  $t = 3 \log n/n$ . To bound  $\mathbf{E}e^{tX'_1}$ , we use that for  $x \in [0, 5]$ ,  $e^x - 1 - x = O(x^2)$ . We thus have

$$\begin{aligned} \mathbf{E}e^{tX'_1} &= 1 + t\mathbf{E}X'_1 + \sum_{k < n/10} \mathbf{P}\{X_1 = k\} (e^{t(k-1)} - 1 - tk) \\ &\leq 1 + 2t/3 + C \sum_{k < n/10: tk \leq 5} k^{-5/2} (tk)^2 + C \sum_{k < n/10: tk > 5} k^{-5/2} e^{tk}. \end{aligned}$$

The first sum on the final line is  $O(t^{3/2}) = o(1/n)$ . For the second note that when  $tk > 5$ ,

$$\frac{k^{-5/2} e^{tk}}{(k+1)^{-5/2} e^{t(k+1)}} = \left(1 + \frac{1}{k}\right)^{5/2} e^{-t} < e^{5/(2k)-t} < e^{-t/2},$$

so the second sum is bounded by

$$e^{tn/10} (n/10)^{-5/2} \sum_{i \geq 0} e^{-it/2} = O(n^{-5/2} e^{tn/10}/t) = o(1/n).$$

Thus, for  $t = 3 \log n/n$  we obtain that  $\mathbf{E}e^{tX'_1} = 1 + 2 \log n/n + o(1/n)$ , so

$$\mathbf{P}\{S_{2n+1} = 2n, N = 0\} \leq e^{-2nt} \cdot \left(\mathbf{E}e^{tX'_1}\right)^{2n+1} = n^{-6} \left(1 + \frac{2 \log n + o(1)}{n}\right)^{2n+1} = o(n^{3/2}).$$

This completes the proof.  $\square$

*Proof of Proposition 3.4:* For  $1 \leq i \leq 2n+1$  write  $E_{n,i} = E_n \cap \{X_i = X^{2n+1,1}\}$ . Then let

$$A = \left\{ (x_1, \dots, x_{2n+1}) \in \mathbb{Z}^{2n+1} : \right. \\ \left. \forall i \leq 2n, 0 \leq x_i < n/10, \left| \sum_{i=1}^{2n} x_i - 4n/3 \right| \leq \delta_n n, x_{2n+1} = 2n - \sum_{i=1}^{2n} x_i \right\}.$$

For  $n$  large, if  $(x_1, \dots, x_{2n+1}) \in A$  then  $2n - \sum_{i \leq 2n} x_i \geq (2/3 - \delta_n)n \geq n/10 \geq \max_{i \leq 2n} x_i$ , so  $E_{n,2n+1} = \{(X_1, \dots, X_{2n+1}) \in A\}$ .

Next, let  $\hat{E}_n = \{(X_1, \dots, X_{2n}, 2n - S_{2n}) \in A\}$ . Note that  $(X_1, \dots, X_{2n+1}) \in A$  if and only if  $(X_1, \dots, X_{2n}, 2n - S_{2n}) \in A$  and  $X_{2n+1} = 2n - S_{2n}$ . Also, for all vectors  $(x_1, \dots, x_{2n+1}) \in A$  we have  $|x_{2n+1} - 2n/3| \leq \delta_n n$ , and for such values  $x_{2n+1}$ ,

$$\mathbf{P}\{X_{2n+1} = x_{2n+1}\} \sim c(n/10)^{-5/2},$$

where as before we write  $c = \left(\frac{8}{27\pi}\right)^{1/2}$ . Thus, uniformly over  $B \subseteq A$ ,

$$\begin{aligned} & \mathbf{P}\{(X_1, \dots, X_{2n+1}) \in B\} \\ &= \sum_{(x_1, \dots, x_{2n+1}) \in B} \mathbf{P}\{(X_1, \dots, X_{2n}) = (x_1, \dots, x_{2n})\} \mathbf{P}\{X_{2n+1} = x_{2n+1}\} \\ &= (1 + o(1))c(n/10)^{-5/2} \sum_{(x_1, \dots, x_{2n+1}) \in B} \mathbf{P}\{(X_1, \dots, X_{2n}) = (x_1, \dots, x_{2n})\} \\ &= (1 + o(1))c(n/10)^{-5/2} \mathbf{P}\{(X_1, \dots, X_{2n}, 2n - S_{2n}) \in B\}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbf{P}\{(X_1, \dots, X_{2n+1}) \in B \mid E_{n,2n+1}\} &= \mathbf{P}\{(X_1, \dots, X_{2n+1}) \in B \mid (X_1, \dots, X_{2n+1}) \in A\} \\ &= \frac{\mathbf{P}\{(X_1, \dots, X_{2n+1}) \in B\}}{\mathbf{P}\{(X_1, \dots, X_{2n+1}) \in A\}} \\ &= (1 + o(1)) \frac{\mathbf{P}\{(X_1, \dots, X_{2n}, 2n - S_{2n}) \in B\}}{\mathbf{P}\{(X_1, \dots, 2n - S_{2n}) \in A\}} \\ &= (1 + o(1)) \mathbf{P}\{(X_1, \dots, X_{2n}, 2n - S_{2n}) \in B \mid \hat{E}_n\}, \end{aligned}$$

so

$$((X_1, \dots, X_{2n+1}) \mid E_{n,2n+1}) \stackrel{d}{\approx} ((X_1, \dots, X_{2n}, 2n - S_{2n}) \mid \hat{E}_n).$$

For  $n$  large, on  $\hat{E}_n$  we have  $2n - S_{2n} > \max_{i \leq 2n} X_i = X^{2n,1}$ , so

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid E_{n,2n+1}) \stackrel{d}{\approx} ((2n - S_{2n}, X^{2n,1}, \dots, X^{2n,2n}) \mid \hat{E}_n).$$

By symmetry, the distribution of the decreasing rearrangement of  $X_1, \dots, X_{2n+1}$  does not depend on the index at which the maximum occurs, so for all  $1 \leq i \leq 2n+1$ ,

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid E_{n,i}) \stackrel{d}{\approx} ((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid E_n).$$

Moreover, Lemma 3.5 implies that

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid E_n) \stackrel{d}{\approx} ((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid S_{2n+1} = 2n).$$

Finally,

$$\begin{aligned} \mathbf{P}\{\hat{E}_n\} &= \mathbf{P}\left\{|S_{2n} - 4n/3| \leq \delta_n n, \max_{i \leq 2n} X_i < n/10\right\} \\ &\geq 1 - \mathbf{P}\{|S_{2n} - 4n/3| > \delta_n n\} - 2n\mathbf{P}\{X_1 \geq n/10\} \\ &= 1 - o(1), \end{aligned}$$

provided  $\delta_n \rightarrow 0$  sufficiently slowly, using the law of large numbers to bound the first probability on the final line, and the bound  $\mathbf{P}\{X_1 \geq n/10\} = O(n^{-3/2})$  for the second. Together with the three preceding asymptotic distributional identities, this yields that

$$((X^{2n+1,1}, \dots, X^{2n+1,2n+1}) \mid S_{2n+1} = 2n) \stackrel{d}{\approx} (2n - S_{2n}, X^{2n,1}, \dots, X^{2n,2n}). \quad \square$$



## Remarks

- (1) The second statement – the convergence of the random variables  $L_{n,k}$  after rescaling when  $k \geq 2$  – seems to be new. The fact that  $(n^{-2/3}L_{n,2}, n \geq 1)$  is a tight family of random variables, or in other words that the second largest block has size  $O(n^{2/3})$  in probability, is proved in [Gao and Wormald \(1999\)](#) in some cases, and in [Banderier et al. \(2001\)](#) in greater generality.
- (2) [Panagiotou and Weiß \(2007\)](#) showed how to use compositional schemas together with Boltzmann sampling techniques to derive information about maximal node degrees and block sizes in several families of random graphs. A similar method was later used in [Panagiotou and Steger \(2009\)](#) to derive bounds on maximal and near-maximal block sizes in random planar graphs. The method from [Panagiotou and Weiß \(2007\)](#); [Panagiotou and Steger \(2009\)](#) shares aspects with our own but yields slightly different information. In particular, it does not yield results on limiting distributions (which ours does), but does yield bounds on tail probabilities (which ours does not).
- (3) The convergence of  $L_{n,1}$  is related to results from [Bender et al. \(1995\)](#) and [Gao and Wormald \(1999\)](#). A stronger, local limit theorem for  $L_{n,1}$ , with explicit estimates on the rate of convergence, is given in [Banderier et al. \(2001, Theorem 3\)](#). As mentioned earlier, the initial motivation for the current work was to show how results in this direction may be straightforwardly obtained by probabilistic arguments. With a little care, the definition of the block tree may be altered to accommodate any of the compositional schemas considered in [Banderier et al. \(2001\)](#).
- (4) In view of the preceding comment, the same line of argument should yield a version of the theorem (with constants altered appropriately) corresponding to any reasonable decomposition of a map into submaps of higher connectivity. Indeed, it seems that composite structures should in general fit within the current analytic framework. (Of course, the sorts of limit theorems one may expect will depend on the combinatorics of the specific problem. As far as I am aware, the fact that the combinatorics of maps always lead to  $O(n^{2/3})$  fluctuations and Airy-type limits is thus far an empirical fact rather than a provable necessity.)

As pointed out by a referee, the block tree construction may be viewed as an instantiation of the “enriched” trees of Labelle [Labelle \(1981\)](#), and the framework of enriched trees might be a natural one to use if one wished to generalize the arguments of the current paper; perhaps this might also shed some light on the questions implicit in the preceding paragraph.

- (5) It seems likely that instead of using the results of [Janson \(2012\)](#), one could appeal to Theorem 1 of [Armendáriz and Loulakis \(2011\)](#), using (2.7) from [Armendáriz and Loulakis \(2011\)](#) to control  $L_{n,1}$ . However, the language in [Janson \(2012\)](#) is closer to that of the current paper.

Here are two final thoughts. First, as mentioned above, the paper [Banderier et al. \(2001\)](#) proves a local limit theorem for  $L_{n,1}$ , with explicit error bounds in the rate of convergence. It would be interesting to recover such bounds by probabilistic methods. Second, that paper also proves essentially sharp bounds for the upper and lower tail probabilities of  $L_{n,1}$ ; see Theorems 1 and 5. Similar tail bounds should apply in the more general settings of [Armendáriz and Loulakis \(2011\)](#); [Janson \(2012\)](#). This seems like a fundamental question in large deviations of functions of iid random variables. The main result of [Denisov et al. \(2008\)](#) seems quite pertinent, but pertains specifically to sums rather than to more general functions.

## 4. Acknowledgements

I send my thanks to two anonymous referees, and to Mark Noy, for many useful comments and corrections.

This work was largely written while I was visiting the Isaac Newton Institute for Mathematical Sciences during the Random Geometry programme, supported by EPSRC Grant Number EP/K032208/1. I would like to thank the Newton Institute and the Simons Foundation for their hospitality and support during this time. I also thank the University of Oxford and the Leverhulme

Trust for their support during parts of this work. Finally, in all stages of this work my research was supported by NSERC and by FQRNT, I thank both institutions.

## References

- Inés Armendáriz and Michail Loulakis. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl.*, 121(5):1138–1147, 2011. URL <http://arxiv.org/abs/0912.1516>. 5, 8
- Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures Algorithms*, 19(3-4): 194–246, 2001. URL <http://algo.inria.fr/flajolet/Publications/BaFlScSo01.pdf>. 0, 8
- Edward A. Bender, L. Bruce Richmond, and Nicholas C. Wormald. Largest 4-connected components of 3-connected planar triangulations. *Random Structures Algorithms*, 7(4):273–285, 1995. 0, 8
- William G. Brown. Enumeration of non-separable planar maps. *Canad. J. Math.*, 15:526–545, 1963. URL <http://cms.math.ca/cjm/a145389>. 3
- D. Denisov, A. B. Dieker, and V. Shneer. Large deviations for random walks under subexponentiality: the big-jump domain. *Ann. Probab.*, 36(5):1946–1991, 2008. URL <http://dx.doi.org/10.1214/07-AOP382>. 8
- William Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971. 4
- Pablo A. Ferrari, Claudio Landim, and Valentin V. Sisko. Condensation for a fixed number of independent random variables. *J. Stat. Phys.*, 128(5):1153–1158, 2007. URL <http://arxiv.org/abs/math/0612856>. 5
- Zhicheng Gao and Nicholas C. Wormald. The size of the largest components in random planar maps. *SIAM J. Discrete Math.*, 12(2):217–228, 1999. 0, 8
- Ian P. Goulden and David M. Jackson. *Combinatorial enumeration*. Dover Publications, Inc., Mineola, NY, 2004. 3
- Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. URL <http://dx.doi.org/10.1214/11-PS188>. 5, 8
- I. Kortchemski. Limit theorems for conditioned non-generic Galton-Watson trees. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(2):489–511., 2015. <http://arxiv.org/abs/1205.3145>. 5
- Gilbert Labelle. Une nouvelle démonstration combinatoire des formules d’inversion de Lagrange. *Adv. in Math.*, 42(3):217–247, 1981. ISSN 0001-8708. URL [https://doi.org/10.1016/0001-8708\(81\)90041-4](https://doi.org/10.1016/0001-8708(81)90041-4). 8
- Sergei K. Lando and Alexander K. Zvonkin. *Graphs on surfaces and their applications*, volume 141 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. ISBN 3-540-00203-0. doi: 10.1007/978-3-540-38361-1. URL <http://dx.doi.org/10.1007/978-3-540-38361-1>. With an appendix by Don B. Zagier, Low-Dimensional Topology, II. 0
- M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983. ISBN 0-387-90731-9. URL <http://dx.doi.org/10.1007/978-1-4612-5449-2>. 5
- Konstantinos Panagiotou and Angelika Steger. Maximal biconnected subgraphs of random planar graphs. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 432–440. SIAM, Philadelphia, PA, 2009. 8
- Konstantinos Panagiotou and Andreas Weiß. Properties of random graphs via Boltzmann samplers. In *2007 Conference on Analysis of Algorithms, AofA 07*, Discrete Math. Theor. Comput. Sci. Proc., AH, pages 159–168. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2007. 0, 8
- J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. URL <http://www.stat.berkeley.edu/~pitman/621.pdf>. 4
- W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963. URL <https://cms.math.ca/10.4153/CJM-1963-029-x>. 0, 2, 3