Formation of a giant component in the intersection graph of a random chord diagram

Hüseyin Acan[1]

Department of Mathematics, Rutgers University, Piscataway NJ, 08854, USA

Boris Pittel[2]

Department of Mathematics, The Ohio State University, Columbus, Ohio 43210, USA

Abstract

We study the number of chords and the number of crossings in the largest component of a random chord diagram when the chords are sparsely crossing. This is equivalent to studying the number of vertices and the number of edges in the largest component of the random intersection graph. Denoting the number of chords by n and the number of crossings by m, when $m/(n \log n)$ tends to a limit in $(0, 2/\pi^2)$, we show that the chord diagram chosen uniformly at random from all the diagrams with given parameters has a component containing almost all the crossings and a positive fraction of chords. On the other hand, when $m \leq n/14$, the size of the largest component is of order $O(\log n)$. One of the key analytical ingredients is an asymptotic expression for the number of chord diagrams with parameters n and m for $m < (2/\pi^2)n \log n$, based on the Touchard-Riordan formula and the Jacobi identity for the generating function of Euler partition function.

Keywords: chord diagram, enumeration, crossing, asymptotics, giant component
05C30, 05C80, 05C05, 34E05, 60C05

1. Introduction

A *chord diagram* of size n is a pairing of $2n$ points. It is customary to place the $2n$ points on a circle in general position, label them 1 through $2n$ clockwise, and connect the two points in the pairing with a chord. Alternatively, we can represent a chord diagram by putting the numbers $\{1, \dots, 2n\}$ on a line in increasing order and connecting the pairs of a chord diagram by an arc; we call it a *linearized chord diagram*. For an illustration, see Figure 1.

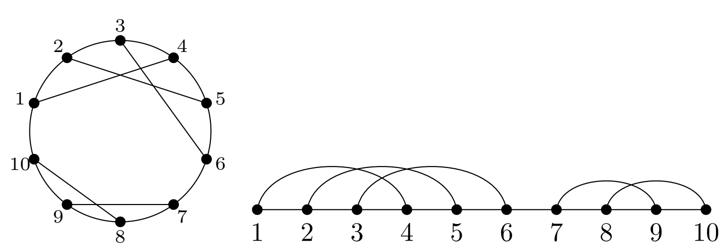

Figure 1: A circular and a linearized chord diagram. They are equivalent to each other.

Email addresses: huseyin.acan@rutgers.edu (Hüseyin Acan), bgp@math.ohio-state.edu (Boris Pittel)

[1]Supported by the National Science Foundation fellowship (Award No. 1502650). Most of this work was done while the author was at Monash University.

[2]Research of Boris Pittel is supported by NSF Grant DMS-1101237

Chord diagrams appear in various contexts in mathematics, especially in topology. For instance, Chmutov and Duzhin [19], Stoimenow [46], Bollobás and Riordan [13], and Zagier [48] used chord diagrams to bound the dimension of the space of order n Vassiliev invariants in knot theory. Rosenstiehl [43] gave a characterization of Gauss words in terms of the intersection graphs of the chord diagrams.

As another application, consider an oriented surface obtained by taking a regular $2n$-gon and gluing the edges pairwise with opposite directions. Each such gluing defines a chord diagram; simply interpret the glued edges of the $2n$-gon as pairs of endpoints of chords. In this topological context, it is natural to ask what the genus of a given chord diagram is. A remarkable formula for the generating function of the double sequence $c_g(n)$ was found by Harer and Zagier [30]; here $c_g(n)$ denotes the number of chord diagrams with n chords and genus g. Recently the second author found a relatively simple proof of the Harer-Zagier formula, [39]. Linial and Nowik [33] found the asymptotic likely value of the genus of the chord diagram chosen uniformly at random. Subsequently, Chmutov and Pittel [20] proved that, as $n \to \infty$, the genus of the random chord diagram is asymptotic to the Gaussian random variable with mean $n/2$ and variance $\frac{1}{4}\log n$. Recently Chmutov and Pittel [21] proved a similar result for the random surface obtained by gluing, uniformly at random, several polygonal discs with various numbers of sides. The case of discs with the same number of sides had been studied by Pippenger and Schleich [36], Gamburd [29], Fleming and Pippenger [28]. For detailed information about the chord diagrams and their topological and algebraic significance we refer the reader to Chmutov, Duzhin, and Mostovoy's book [18].

A chord diagram of size n can be thought of as a fixed-point-free involution of a set of $2n$ numbers. Baik and Reins [8] found the asymptotic distribution of the length of the longest decreasing subsequence of a random fixed-point-free involution. Chen et al. [17] showed that the crossing number and the nesting number of linearized chord diagrams have a symmetric joint distribution. Since the lack of a decreasing subsequence of length $2k+1$ in an involution is equivalent to the lack of $(k+1)$-nesting in the corresponding chord diagram, the result of Baik and Reins, combined with the result of Chen et al., gives the distribution for the maximum number of chords, all crossing each other, when the chord diagram is chosen uniformly at random.

In random graph theory, Bollobás and O. Riordan [12] used the random *linearized* chord diagrams to provide a precise description of the preferential attachment random graph model introduced by Barabási and Albert [9].

It is easy to see that there are $(2n-1)!!$ chord diagrams of size n. However, enumerating chord diagrams with special properties could become hard rather quickly. A classic example is counting chord diagrams with a given number of crossings. This problem was first studied by Touchard [47], who found a bivariate generating function for $T_{n,m}$, the number of chord diagrams of size n with m crossings. To this end, he considered an equivalent problem of enumerating the linearized diagrams by crossings: the $2n$ points are distributed on a straight line, and are connected in pairs by n concave arcs, all above the line; the crossings of these arcs correspond to the crossings of chords on the circle. Later, J. Riordan [41] used Touchard's formula to extract the remarkable explicit formulas for $\sum_m q^m T_{n,m}$, and $T_{n,m}$ itself. However, the latter is in the form of an alternating sum, indispensable for moderate values of m and n, but not easily yielding an asymptotic approximation for $T_{n,m}$ for $n, m \to \infty$. (We refer the reader to Aigner [4] for an eminently readable exposition of the Touchard-Riordan achievement.) A quarter century later, Flajolet and Noy [27] were able to use J. Riordan's formula for the univariate $\sum_m q^m T_{n,m}$ to show that the number of crossings in the uniformly random chord diagram is asymptotically Gaussian. Cori and Marcus [22] counted the number of isomorphism classes of chord diagrams, with two chord diagrams being isomorphic if they are rotationally equivalent.

Another way to represent a chord diagram $\mathcal{D}$ is to associate with it a graph $G_{\mathcal{D}}$, called the *intersection graph of* $\mathcal{D}$. The vertices of $G_{\mathcal{D}}$ are the chords of $\mathcal{D}$ and there is an edge between two vertices in $G_{\mathcal{D}}$ if and only if the corresponding chords cross each other in $\mathcal{D}$, see Figure 2. If, instead of labeling the endpoints of the chords, we label the chords from 1 to n in an arbitrary way, we obtain a labeled *circle graph*. Circle graphs are interesting in their own right and they have been studied widely. A characterization of circle graphs was given by Bouchet [15]. (Still, as Arratia et al. [7] pointed out in a lucid discussion, even a formula, exact or asymptotic, for the number of circle graphs remains unknown.)

A chord diagram $\mathcal{D}$ is *connected* if there is no line cutting the circle that does not intersect any of the chords and partitions the set of chords into two nonempty subsets. In other words, $\mathcal{D}$ is connected if and only if $G_{\mathcal{D}}$ is connected.

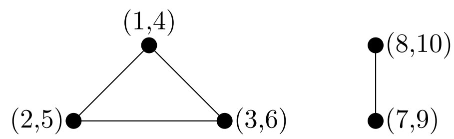

Figure 2: Intersection graph of the chord diagram given in Figure 1.

By making use of recurrence relations, Stein and Everett [45] proved that, as n tends to infinity, the probability that a random chord diagram with n chords is connected approaches $1/e$. Later, Flajolet and Noy [27] proved that almost all chord diagrams are *monolithic*, i.e. consist of a single giant component and a number of isolated chords. Having proved that in the limit the number of isolated chords was Poisson(1), they recovered Stein and Everett's result. The result of Flajolet and Noy was later extended in several directions in [2]. For example, the digraph obtained from a uniformly random chord diagram of size n by orienting the edges by flipping a fair coin was proved to be strongly connected with the limit probability $1/e^3$.

Our motivation for a probabilistic study of the random diagram comes from the realization that its intersection graph represents a structurally rich analogue of the classic random graph $G(n, m)$, i.e. the graph distributed uniformly on the set of all $\binom{\binom{n}{2}}{m}$ graphs on $[n] := \{1, \dots, n\}$ with m edges. More than half a century ago, Erdős and Rényi [24]–[25] basically created modern random graph theory by showing that $n/2$ and $n \log n/2$ are the respective thresholds of the number of edges m for appearance, with high probability, of a giant component in $G(n, m)$, and for $G(n, m)$ becoming connected whp. (An event A_n occurs with high probability (whp) if $\lim_{n\to\infty} \mathrm{P}(A_n) = 1$.) The analysis was a striking manifestation of interplay between classic graph-enumerative techniques and probabilistic, moment-based, estimates. What are then the corresponding thresholds for the intersection graphs of chord diagrams?

In this paper, we find some partial answers. In particular, we show (Theorem 5.9) that if the number of crossings $m = m(n)$ is such that $\lim m/(n \log n) \in (0, 2/\pi^2)$, then, whp there is a giant component containing almost all m crossings and a positive fraction of all vertices. We had to impose the upper bound on m since that was the range for which we were able to establish a sharp asymptotic formula for $T_{n,m}$, the number of chord diagrams with n chords and m crossings.

Earlier we proved [3] that $m'(n) = (6/\pi^2)n \log n$ is the threshold value of the number m of crossings for *connectedness* of the intersection graph of a uniformly random chord diagram, with n chords, all crossing an additional chord which is not a part of the diagram. (Here the intersection graph is actually a permutation graph for a permutation of $[n]$ induced by such a special chord diagram, with the number of edges equal to the number of inversions in that permutation.)

It is highly plausible that the result in Theorem 5.9 holds for every m dependent on n in such a way that $\lim m/(n \log n) > 0$. This would certainly follow, if one could find a way to "embed" an intersection graph with m_1 crossings into that with m_2 crossings, whenever $m_1 < m_2$. However, unlike the Erdős-Rényi random graph $G(n, m)$, such an embedding is highly problematic, if possible at all, for the random intersection graphs in question.

In this regard, our intersection graph and the one in [3] join the club of many other random graph models lacking "embedability", such as the random regular graph, or more generally, the random graph with a given degree sequence. (We did prove though in [3] that the connectedness probability for the random permutation graph increases with m.) It would be very interesting to find an algorithm that, given n, m, generates an almost uniformly random chord diagram with parameters n and m.

The only insight into the component structure of our intersection graph for m's satisfying both $m = \Omega(n \log n)$ and $m = o(n^{3/2})$ is a *gap property* for the number of crossings in the *densest* component, the one with the highest ratio of number of crossings to the number of chords. It states that whp the densest component is either of size $O((m/n) \log n)$ or it contains almost all the m crossings whence the size of the component is at least $(1 + o(1))\sqrt{2m}$.

We also show (Theorem 5.10) that if $m \leq n/14$, then whp the largest component has a size below $5 \log n/(\log \frac{225}{224})$. The bound $m \leq n/14$ may well be improved; as we mentioned, the giant-component threshold for $G(n, m)$ is $m = n/2$, see Erdős-Rényi [24]–[25], Bollobás [11]. In any event, if a deterministic threshold $m = m(n)$ for the birth of a giant component in the intersection graph exists, it is sandwiched between $n/14$ and $\varepsilon n \log n$, for an arbitrarily small $\varepsilon > 0$.

It is well known that for $G(n, m)$ the connectedness threshold is $m \sim n \log n/2$, ([24], [25], [11]). According to [27], the crossing number of the uniformly random diagram is sharply concentrated around its mean $n(n-1)/6$, with a standard deviation of order $n^{3/2}$, and the intersection graph is disconnected with probability approaching $1 - e^{-1}$ as n tends to infinity, see [45]. These results almost certainly rule out the existence of a connectedness threshold $m(n) = o(n^2)$. Still, we conjecture that $m(n) \approx n^{3/2}$ is the threshold value of m for the second largest component to be of bounded size.

Among the key ingredients of our proofs is the asymptotic formula for $T_{n,m}$ for $m < (2/\pi^2)n \log n$ (Lemma 3.4), and a bound $T_{n,m} \le C_n I_{n,m}$, where C_n is the n-th Catalan number and $I_{n,m}$ is the number of permutations of $[n]$ with m inversions (Lemma 4.4). The asymptotic formula is based on the Touchard-Riordan sum-type formula, Jacobi's identity and Freiman's asymptotic formula for the generating function of the Euler partition function. A final step in our argument is based on a rather deep formula for the number of non-crossing partitions with given block sizes, due to Kreweras [32].

We should note that the Jacobi identity had appeared prominently in Josuat-Vergés and Kim's [31] paper in the context of some new Touchard-Riordan type formulas for generating functions.

In our opinion, enumeration of chord diagrams with given genus, number of chords, and number of crossings is an exciting open problem. This is, of course, equivalent to finding the distribution of the genus of the uniformly random chord diagram with a given number of chords and a given number of crossings. The issue here is that while the genus is zero iff the number of crossings is zero, the genus is always below $\frac{n+1}{2}$, while the largest number of crossings is $\binom{n}{2}$. (Conceptually this question is connected to the work of Archdeacon and Grable [6] and Rödl and Thomas [42], who obtained strikingly sharp bounds of the genus of the Bernoulli random graph $G(n, p)$, each (i, j) being an edge with probability p, independently of all other $\binom{n}{2} - 1$ pairs.) As discussed above, without the "number of crossings" parameter, the genus problem has been addressed. We conjecture that when the number of crossings is passing through the threshold value for birth of a giant component, the genus of the surface associated with the diagram is experiencing a dramatic increase, not unlike the "double jump phenomenon" for the giant component in the random graph $G(n, m)$ (see [24], [25], [11]).

2. Paper structure

To ease the task of reading the paper, we describe here the chronological organization of the proofs. In Section 3 we use the Touchard-Riordan formula to derive the asymptotic formula for $T_{n,m}$, the number of chord diagrams with n chords and m crossings. As an illustration, this formula is used to obtain the limit distribution of the number of "cuts" in a random *linearized* chord diagram with $m = O(n)$ crossings.

In Section 4 we establish the upper and lower bounds for $T_{n,m}$ for m outside of the range covered by the asymptotic formula for $T_{n,m}$. As an application, we show that the maximum size of a cut in the random linearized chord diagram with $m = O(n)$ crossings is bounded in probability.

In Section 5 we use the asymptotics and the bounds for $T_{n,m}$ to prove our main results. Our first step is to derive a bound for the total number $\mathcal{C}_{\nu,\mu}$ of the *connected* chord diagrams with ν chords and $\mu > \nu$ crossings, going beyond the explicit formulas for $\mathcal{C}_{\nu,\nu-1}$ and $\mathcal{C}_{\nu,\nu}$. Next, we use this bound to prove a crucial, crossing-density gap result: for the crossing density $m/n > \alpha > 4e^2$, it is unlikely that the intersection graph contains a component with at least $\beta \log n$ vertices with some explicit constant $\beta = \beta(\alpha)$, whose edge density falls below $m/(\alpha n)$. Thus for $m/n \to \infty$, whp there are no components of size $\Omega(\log n)$, whose density is negligible compared with m/n. Focusing on the densest component as a potential candidate for being the largest component, we show that whp either it has $O(\log n)$ vertices, or it has almost all m crossings, whence at least $(1 + o(1))\sqrt{2m}$ vertices. In addition, we prove that if $\lim m/(n \log n) \in (0, 2/\pi^2)$, then, for b large enough, whp there is no component of size at least $b \log n$, with the number of crossings below $(1 - \varepsilon)m$, whence there can be at most one component with that many vertices. This result and an additional enumeration based on Kreweras' formula for the number of non-crossing partitions with given block sizes allow us to show that, for m in question, whp there exists a component that has almost all m crossings and a positive fraction of n chords, a genuine giant. Lastly, we show that, for $m \le n/14$, whp the size of the largest component is $O(\log n)$.

We conclude with a list of open problems.

3. Counting moderately crossing chord diagrams

Our ultimate goal is to analyze a chord diagram chosen uniformly at random from all chord diagrams with n chords and m crossings. Let $T_{n,m}$ denote the size of the sample space, i.e. the number of n-chord diagrams with m crossings, where $m \in [0, N]$, $N = \binom{n}{2}$. To this end, we will first find asymptotic formulas for $T_{n,m}$ and for the number of ℓ-tuples of chord diagrams with a total number of n chords and m crossings. A remarkable formula for the generating function of $T_{n,m}$ was given by Touchard [47]. Using this formula, Riordan [41] found an alternating sum expression for $T_{n,m}$. We use Touchard's formula to obtain a more general result than the one given by Riordan. While the number of all graphs on n vertices with m edges is the obvious $\binom{N}{2}$, the formula for $T_{n,m}$ we are about to cite is rather deep.

For $n > 0$, introduce the generating function $T_n(x) = \sum_m T_{n,m} x^m$. Let $T_0(x) = 1$ and

$$T(x, y) = \sum_{n \geq 0} T_n(x) y^n = \sum_{n,m} T_{n,m} x^m y^n.$$

Touchard's formula states: for $|x| < 1$, $|y| \leq 1/4$,

$$T(x, (1 - x)y) = C(y) A(x, 1 - C(y)), \tag{3.1}$$

where

$$A(x, y) := \sum_{j \geq 0} x^{\binom{j+1}{2}} y^j,$$

and $C(y)$ is the generating function of Catalan numbers, that is, for $C_n = (n + 1)^{-1}\binom{2n}{n}$,

$$C(y) := \sum_{n \geq 0} C_n y^n.$$

Note that $A(x, y)$ converges for $|x| < 1$ and all y, since the ratio of two consecutive terms is $x^j y$. Also, it is well known that $C(y)$ converges for $|y| \leq 1/4$, and for those y's

$$yC^2(y) - C(y) + 1 = 0. \tag{3.2}$$

Solving (3.2) with the initial condition $C(0) = C_0 = 1$ gives

$$C(y) = \frac{1 - \sqrt{1 - 4y}}{2y} = \frac{2}{1 + \sqrt{1 - 4y}}. \tag{3.3}$$

Rewriting (3.2) as

$$C(y) - 1 = yC^2(y), \tag{3.4}$$

and using Lagrange inversion formula, we obtain

$$[y^n](C(y) - 1)^j = \frac{j}{n}[y^{n-j}](y + 1)^{2n} = \frac{j}{n}\binom{2n}{n - j}, \quad 0 < j \leq n. \tag{3.5}$$

For $j = 1$, we are back to $C_n = (n + 1)^{-1}\binom{2n}{n}$. Equation (3.5) will be used shortly in the proof of Lemma 3.2. Using (3.1), Riordan [41] obtained the following formula for $T_{n,m}$.

Theorem 3.1 (Touchard-Riordan)**.** *The number of chord diagrams with n chords and m crossings is given by*

$$T_{n,m} = \sum_j (-1)^j \binom{n + m - 1 - J(j)}{n - 1} \frac{2j + 1}{n + j + 1} \binom{2n}{n - j}, \tag{3.6}$$

where $J(j) = \binom{j+1}{2}$ and the sum is over all $j \geq 0$ such that $j \leq n$ and $J(j) \leq m$.

We will need the following more general statement.

Lemma 3.2. *Given $\ell \geq 1$,*

$$[x^m y^n]T^\ell(x,y) = \sum_{\boldsymbol{j}=(j_1,\dots,j_\ell)\geq \boldsymbol{0}} \prod_{\mu=1}^{\ell} (-1)^{j_\mu} \binom{n+m-1-U(\boldsymbol{j})}{n-1} \frac{2j+\ell}{2n+\ell}\binom{2n+\ell}{n-j}, \tag{3.7}$$

where

$$j := \sum_\nu j_\nu, \qquad U(\boldsymbol{j}) := \sum_\mu J(j_\mu), \qquad J(i) := \binom{i+1}{2},$$

and the sum in (3.7) is over $\boldsymbol{j} \geq \boldsymbol{0}$ (meaning that each component of $\boldsymbol{j}$ is nonnegative), such that $j \leq n$ and $U(\boldsymbol{j}) \leq m$.

Proof. Using (3.4), (3.5), and (3.1),

$$\begin{aligned}
[x^m y^n]T^\ell(x,y) &= [x^m y^n]\big((1-x)^{-n} C(y)^\ell A^\ell(x, 1-C(y))\big)\\
&= [x^m y^n]\left\{(1-x)^{-n} C(y)^\ell \left(\sum_{j\geq 0}(-1)^j x^{J(j)}(C(y)-1)^j\right)^\ell\right\}\\
&= [x^m y^n]\left\{(1-x)^{-n} \sum_{j_1,\dots,j_\ell\geq 0}\prod_{\mu=1}^{\ell}(-1)^{j_\mu}x^{J(j_\mu)}(C(y)-1)^{j_\mu}\left(\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}(C(y)-1)^\kappa\right)\right\}\\
&= \sum_{j_1,\dots,j_\ell\geq 0}\prod_{\mu}(-1)^{j_\mu}\left([x^m](1-x)^{-n}x^{U(\boldsymbol{j})}\right)\left(\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}[y^n](C(y)-1)^{\kappa+j}\right)\\
&= \sum_{j_1,\dots,j_\ell\geq 0}\left(\prod_{\mu=1}^{\ell}(-1)^{j_\mu}\right)\binom{n+m-1-U(\boldsymbol{j})}{n-1}\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}\frac{\kappa+j}{n}\binom{2n}{n-\kappa-j}.
\end{aligned}$$

Here

$$\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}\binom{2n}{n-\kappa-j} = \binom{2n+\ell}{n-j},$$

and

$$\begin{aligned}
\sum_{\kappa=0}^{\ell}\kappa\binom{\ell}{\kappa}\binom{2n}{n-\kappa-j} &= \ell\sum_{r=0}^{\ell-1}\binom{\ell-1}{r}\binom{2n}{n-r-1-j}\\
&= \ell\binom{2n+\ell-1}{n-1-j},
\end{aligned}$$

implying

$$\begin{aligned}
\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}\frac{\kappa+j}{n}\binom{2n}{n-\kappa-j} &= \frac{j}{n}\sum_{\kappa=0}^{\ell}\binom{\ell}{\kappa}\binom{2n}{n-\kappa-j} + \frac{1}{n}\sum_{\kappa=0}^{\ell}\kappa\binom{\ell}{\kappa}\binom{2n}{n-\kappa-j}\\
&= \frac{j}{n}\binom{2n+\ell}{n-j} + \frac{\ell}{n}\binom{2n+\ell-1}{n-1-j}\\
&= \frac{2j+\ell}{2n+\ell}\binom{2n+\ell}{n-j}.
\end{aligned}$$

This completes the proof of (3.7). $\square$

The formula in Lemma 3.2 is too unwieldy for our needs; however it enables us to derive an eminently usable, asymptotic formula for $[x^m y^n]T^\ell(x,y)$, in particular for $T_{n,m}$, when the number of crossings m is not too large compared with n. We begin with

Lemma 3.3. *Let $n \to \infty$ and $m = O(n)$. Then, setting $q = m/(m+n)$,*

$$\begin{gathered} T_{n,m} \sim \binom{n+m-1}{n-1} C_n \prod_{j\geq 1}(1-q^j)^3, \\ [x^m y^n]T^\ell(x,y) \sim \ell(2f(q))^{\ell-1} T_{n,m}; \quad f(x) := \sum_{j\geq 0}(-1)^j x^{J(j)}, \end{gathered} \tag{3.8}$$

where $J(j) = \binom{j+1}{2}$ as in the previous lemma.

Proof. The case $m = O(1)$ is easy as the sums in (3.6) and (3.7) are asymptotic to their first terms, which correspond to $j = 0$ and $\boldsymbol{j} = \mathbf{0}$, respectively. Consider now the more difficult case $m \to \infty$. We notice upfront that $1 - q$ is bounded away from 0 for $m = O(n)$. Let $S_{n,m}(\boldsymbol{j})$ denote the absolute value of the $\boldsymbol{j}$-th term of the sum in (3.7), i.e.

$$S_{n,m}(\boldsymbol{j}) = \binom{n+m-1-U(\boldsymbol{j})}{n-1} \frac{2j+\ell}{2n+\ell} \binom{2n+\ell}{n-j},$$

where

$$j = \sum_\nu j_\nu \quad \text{and} \quad U(\boldsymbol{j}) = \sum_\mu J(j_\mu).$$

We first find an upper bound and an asymptotic equation for $S_{n,m}(\boldsymbol{j})$ by analyzing the individual factors above. Observe that

$$\binom{n+m-1-U(\boldsymbol{j})}{n-1} = \binom{n+m-1}{n-1} \frac{(m)_{U(\boldsymbol{j})}}{(m+n-1)_{U(\boldsymbol{j})}}, \tag{3.9}$$

where $(a)_k$ denotes the k-th falling factorial of a. Note that $U(\boldsymbol{j}) \leq m$ for an admissible $\boldsymbol{j}$, i.e. for $\boldsymbol{j}$ meeting the conditions of Lemma 3.2. Using $(1+1/u)^v \leq (1+1/u)^u \leq e$ for any $u > 0$ and $v \leq u$, we get

$$\frac{(m)_{U(\boldsymbol{j})}}{(m+n-1)_{U(\boldsymbol{j})}} \leq \left(\frac{m}{m+n-1}\right)^{U(\boldsymbol{j})} = q^{U(\boldsymbol{j})}\left(1 + \frac{1}{m+n-1}\right)^{U(\boldsymbol{j})} \leq e q^{U(\boldsymbol{j})}, \quad (q = m/(m+n)). \tag{3.10}$$

Also, straightforward computations give

$$\frac{2j+\ell}{2n+\ell}\binom{2n+\ell}{n-j} \leq \frac{2j+\ell}{2n+\ell}\binom{2n+\ell}{n} \leq \frac{2j+\ell}{2n+\ell} 2^\ell \binom{2n}{n} \leq 2^\ell (2j+\ell) C_n.$$

Therefore, uniformly for all admissible $\boldsymbol{j}$,

$$S_{n,m}(\boldsymbol{j}) \leq_b 2^\ell (2j+\ell) \binom{n+m-1}{n-1} C_n \cdot q^{U(\boldsymbol{j})}. \tag{3.11}$$

Here and elsewhere we use $A \leq_b B$ as a shorthand for $A = O(B)$ when B is too bulky. Furthermore, $\boldsymbol{j}$ is certainly admissible if say $j < m^{1/5}$, and for those $\boldsymbol{j}$, with similar computations, it is easy to obtain

$$S_{n,m}(\boldsymbol{j}) = \big(1 + O(j^4/m)\big) 2^{\ell-1}(2j+\ell) \binom{n+m-1}{n-1} C_n \cdot q^{U(\boldsymbol{j})}. \tag{3.12}$$

Combining (3.11) and (3.12), and using the uniform convergence of the infinite series $\sum_{\boldsymbol{j}\geq\mathbf{0}} j^4 q^{U(\boldsymbol{j})}$, we get

$$[x^m y^n]T^\ell(x,y) = \binom{n+m-1}{n-1} C_n 2^{\ell-1} \times \left[\sum_{\boldsymbol{j}} (2j+\ell) \prod_{\mu=1}^{\ell} (-1)^{j_\mu} q^{J(j_\mu)} + o(1)\right], \tag{3.13}$$

the sum being taken over *all* $\boldsymbol{j} \geq \mathbf{0}$. Here

$$\sum_{\boldsymbol{j}} \prod_{\mu=1}^{\ell} (-1)^{j_\mu} q^{J(j_\mu)} = f(q)^\ell$$

and

$$\sum_{\boldsymbol{j}} 2j \prod_{\mu=1}^{\ell} (-1)^{j_\mu} q^{J(j_\mu)} = \ell f(q)^{\ell-1} \sum_{j_1 \ge 0} (-1)^{j_1} (2j_1) q^{J(j_1)},$$

where $f(x) = \sum_{j\ge 0}(-1)^j x^{J(j)}$ as defined in (3.8). Combining the last two equations, we get

$$\begin{aligned}\sum_{\boldsymbol{j}} (2j+\ell) \prod_{\mu=1}^{\ell} (-1)^{j_\mu} q^{J(j_\mu)} &= \ell f(q)^{\ell-1} \left(f(q) + 2\sum_{j_1\ge 0} (-1)^{j_1} j_1 q^{J(j_1)} \right) \\ &= \ell f(q)^{\ell-1} \sum_{j_1 \ge 0} (-1)^{j_1} (2j_1+1) q^{J(j_1)},\end{aligned}$$

and (3.13) becomes

$$[x^m y^n] T^\ell(x,y) = \binom{n+m-1}{n-1} C_n \ell \times \left[(2f(q))^{\ell-1} \sum_{j\ge 0} (-1)^j (2j+1) q^{J(j)} + o(1) \right]. \tag{3.14}$$

Since the series for $f(x)$ in (3.8) is alternating, and $q^{J(j)} \downarrow 0$, we have $f(q) > 1-q$, i.e. $f(q)$ is bounded away from zero. However, $(2j+1)q^{J(j)}$ is not monotone, and bounding the last alternating sum from below would be a rather hard task. Fortunately, there is a remarkable identity discovered by Jacobi as a corollary of the classic triple product identity, Andrews et al. [5, Page 500]:

$$\sum_{j\ge 0} (-1)^j (2j+1) x^{J(j)} = \prod_{j\ge 1} (1-x^j)^3, \quad |x|<1. \tag{3.15}$$

Since our $q = m/(m+n)$ is bounded away from 1 for $m = O(n)$, the product on the RHS of (3.15) for $x = q$ is bounded away from zero uniformly for n. With this fact in mind, Equations (3.14) and (3.15) complete the proof of Lemma 3.3. □

The reader is correct to suspect that the constraint $m = O(n)$ is unnecessarily restrictive. In our next statement we extend the asymptotic formulas to $m < (2/\pi^2) n \log n$. We hope that the considerably more technical argument can be understood more easily since we will use the proof above as a rough template.

Lemma 3.4. *Let $\ell \ge 1$ be given. If*

$$m \le \frac{2}{\pi^2} n \big(\log n - 0.5(\ell+2)\log\log n - \omega(n)\big), \tag{3.16}$$

where $\omega(n) \to \infty$ however slowly, then

$$[x^m y^n]\, T(x,y)^\ell \sim \ell (2f(q))^{\ell-1} \binom{n+m-1}{n-1} C_n \prod_{j\ge 1} (1-q^j)^3, \quad q := m/(m+n). \tag{3.17}$$

Proof. It suffices to consider the case $m/n \to \infty$, in which case $q \to 1$. We still have $f(q) > 1-q > 0$, but we need an extra effort to prove that $\lim_{n\to\infty} f(q) > 0$, a fact crucial for our argument. By the definition of $f(x)$ given in (3.8),

$$\begin{aligned} f(x) &= \sum_{\text{even } j\ge 0} \left(x^{J(j)} - x^{J(j+1)} \right) = (1-x) \sum_{\text{even } j \ge 0} x^{J(j)} (1+\cdots+x^j) = (1-x)F(x), \\ F(x) &:= \sum_{\text{even } j\ge 0} \sum_{i=J(j)}^{J(j)+j} x^i, \end{aligned}$$

where $J(j) = \binom{j+1}{2}$ as before. For a generic $\nu > 0$, the interval $[0, \nu]$ contains the $j_\nu/2$ disjoint intervals $[J(j), J(j)+j]$, j even, where j_ν is the largest even integer not exceeding $\lfloor(-3+\sqrt{4\nu+9})/2\rfloor$, and possibly a part of the $(j_\nu/2+1)$-st interval, of length $O(j_\nu)$. Since $j_\nu = O(\nu^{1/2})$, and $j_\nu^2 = 2\nu + O(\nu^{1/2})$, we have

$$\sum_{\mu\le\nu}[x^\mu]F(x) = \sum_{j=0}^{j_\nu/2}(j+1) + O(j_\nu) = \frac{\nu}{2} + O(\nu^{1/2}).$$

So, by Tauberian theorem for power series (Feller [26], Ch. XIII, Sect. 5), we have $\lim_{x\to1-}(1-x)F(x) = 1/2$, implying that

$$\lim_{x\to1-} f(x) = 1/2 > 0 \Longrightarrow \lim_{n\to\infty} f(q) = 1/2. \tag{3.18}$$

With (3.18) in mind, let us turn to the core of the proof. For the reader convenience, we restate the key identity (3.7):

$$[x^m y^n]T^\ell(x,y) = \sum_{\boldsymbol{j}=(j_1,\dots,j_\ell)\ge\boldsymbol{0}} \binom{n+m-1-U(\boldsymbol{j})}{n-1}\left(\prod_{\mu=1}^{\ell}(-1)^{j_\mu}\right)\frac{2j+\ell}{2n+\ell}\binom{2n+\ell}{n-j}, \tag{3.19}$$

where $U(\boldsymbol{j}) = \sum_\mu J(j_\mu)$ and $j = \sum_\nu j_\nu$. We focus on $S_{n,m}(\boldsymbol{j})$, the absolute value of the $\boldsymbol{j}$-th summand in (3.19). The uniform bound (3.11) continues to hold. Setting $M := \lfloor a(1-q)^{-1}\rfloor$ for some $a > 1$, we write the sum as $S_1 + S_2$, where S_1 is the contribution of $\boldsymbol{j}$'s with $\max_i j_i \le M$ and S_2 is the contribution of the remaining $\boldsymbol{j}$'s. For the terms in S_1, analogously to (3.12) we have

$$S_{n,m}(\boldsymbol{j}) = \big(1 + O(U(\boldsymbol{j})^2/m)\big)2^{\ell-1}(2j+\ell)\binom{n+m-1}{n-1}C_n\cdot q^{U(\boldsymbol{j})}.$$

Therefore, $S_1 = S_{11} + R_1$, where

$$\begin{aligned} S_{11} &= \binom{n+m-1}{n-1}C_n 2^{\ell-1}\sum_{\substack{j_1,\dots,j_\ell\\ \max j_i\le M}}\left(\sum_{t=1}^{\ell}(2j_t+1)\right)\prod_{\mu=1}^{\ell}(-1)^{j_\mu}q^{J(j_\mu)} \\ &= \binom{n+m-1}{n-1}C_n\ell\sum_{j=0}^{M}(-1)^j(2j+1)q^{J(j)}\left(2\sum_{k=0}^{M}(-1)^k q^{J(k)}\right)^{\ell-1}, \end{aligned} \tag{3.20}$$

and

$$|R_1| \le_b \frac{1}{m}\binom{n+m-1}{n-1}C_n\left(\sum_{j\ge0} j^5 q^{J(j)}\right)\left(\sum_{k\ge0} q^{J(k)}\right)^{\ell-1}. \tag{3.21}$$

For the last bound we have used $U(\boldsymbol{j})^2\sum_\mu j_\mu \le_b \sum_\nu j_\nu^5$ and the fact that ℓ is fixed. Defining the functions

$$h_M(q) = \sum_{j=M+1}^{\infty}(-1)^j(2j+1)q^{J(j)};\quad f_M(q) = \sum_{k=0}^{M}(-1)^k q^{J(k)},$$

and using (3.15) on the last line of (3.20), we write

$$S_{11} = \binom{n+m-1}{n-1}C_n\ell\left(\prod_{j\ge1}(1-q^j)^3 - h_M(q)\right)(2f_M(q))^{\ell-1}. \tag{3.22}$$

Now $q^{J(j)} \le \exp(-j^2(1-q)/2)$, and $x\exp(-x^2(1-q)/2)$ attains its maximum at $(1-q)^{-1/2} \ll a(1-q)^{-1} = M$. So

$$|h_M(q)| \le \sum_{j=M}^{\infty}(2j+1)q^{J(j)} \le_b \int_M^\infty x\exp(-x^2(1-q)/2)dx = (1-q)^{-1}\exp\left(-\frac{a^2}{2(1-q)}\right). \tag{3.23}$$

Also,

$$\left|\sum_{k>M}(-1)^k q^{J(k)}\right| \le q^{J(M+1)} \le q^{a^2(1-q)^{-2}/2} \le \exp\left(-\frac{a^2}{2(1-q)}\right),$$

where the last inequality follows from $x^{1/(1-x)} \le e^{-1}$ for any $x\in(0,1)$. So using Jacobi identity,

$$\sum_{0\le j\le M}(-1)^j(2j+1)q^{J(j)} = \prod_{j\ge1}(1-q^j)^3 + O\big((1-q)^{-1}e^{-a^2/2(1-q)}\big).$$

Here, by Freiman's asymptotic formula (see Postnikov [40, Sect. 2.7], and also Pittel [37, Eq. 2.8], [38, Sect. 2]),

$$\begin{aligned}\prod_{j\ge1}(1-q^j) &= \exp\left(-\frac{\pi^2}{6z} - \tfrac{1}{2}\log\frac{z}{2\pi} + O(|z|)\right)_{z=-\log q}\\ &= \exp\left(-\frac{\pi^2}{6(1-q)} - \frac{1}{2}\log(1-q) + O(1)\right).\end{aligned}\tag{3.24}$$

Therefore

$$\sum_{0\le j\le M}(-1)^j(2j+1)q^{J(j)} = \prod_{j\ge1}(1-q^j)^3\left(1+O\big((1-q)^{1/2}e^{-(a^2-\pi^2)/2(1-q)}\big)\right).$$

Also, using $\lim f(q) = 1/2 > 0$,

$$f_M(q) = f(q) - \sum_{k>M}(-1)^k q^{J(k)} = f(q)\left(1+O\big(e^{-a^2/2(1-q)}\big)\right).$$

So, selecting $a=\pi\sqrt{3}$ say, (3.20) becomes

$$S_{11} = \left(1+O\big(e^{-(1-q)^{-1}}\big)\right)\ell\big(2f(q)\big)^{\ell-1}\binom{n+m-1}{n-1}C_n\prod_{j\ge1}(1-q^j)^3.\tag{3.25}$$

Furthermore, (3.21) together with the bounds

$$\sum_{j\ge0} j^5 q^{J(j)} = O((1-q)^{-3}),\quad 2\sum_{k\ge0} q^{J(k)} \le 2(1-q)^{-1/2}\tag{3.26}$$

yield

$$|R_1| \le_b \frac{1}{m}(1-q)^{-(\ell+5)/2}\binom{n+m-1}{n-1}C_n.\tag{3.27}$$

Let us compare the expression for S_{11} in (3.25) and the the bound (3.27) for $|R_1|$. By Freiman's formula and the condition (3.16), we have

$$\begin{aligned}\frac{m^{-1}(1-q)^{-(\ell+5)/2}}{\prod_{j\ge1}(1-q^j)^3} &\le_b \exp(\pi^2 m/2n)\,\frac{m^{(\ell+3)/2}}{n^{(\ell+5)/2}}\\ &\le_b (\log n)^{(\ell+3)/2}n^{-1}\exp(\pi^2 m/2n - 0.5\log\log n)\\ &\le_b \exp\left(\frac{\ell+2}{2}\log\log n - \log n + \frac{\pi^2 m}{2n}\right)\\ &\le e^{-\omega(n)} \to 0.\end{aligned}$$

Since $f(q)$ is bounded away from 0, it follows from (3.25) and (3.27) that S_1, the contribution to the sum in (3.19) of the terms with $\max_i j_i \le M$, is given by

$$S_1 = (1+o(1))\binom{n+m-1}{n-1}C_n\ell(2f(q))^{\ell-1}.\tag{3.28}$$

It remains to show that S_2, the contribution to the sum in (3.19) of the terms with $\max_i j_i > M$, is negligible compared to S_1. For ℓ is fixed, we have $\binom{2n+\ell}{n-j} = O\big(\binom{2n}{n}\big)$, uniformly for j. So the equation (3.7) gives

$$|S_2| \leq_b \binom{n+m-1}{n-1} C_n \sum_{\substack{j_1,\ldots,j_\ell \\ \max j_i > M}} \prod_{\mu} q^{J(j_\mu)} \sum_{t=1}^{\ell} (2j_t+1).$$

Considering $(j_1, \ldots, j_\ell)$ with $\max j_i = j_1$, and using symmetry, we see that the right hand side above is at most

$$\ell \binom{n+m-1}{n-1} C_n \left(\sum_{j \geq 0} q^{J(j)} \right)^{\ell-1} \sum_{j_1 \geq M} \ell(2j_1+1) q^{J(j_1)}.$$

Using $\sum_{j\geq 0} q^{J(j)} \leq (1-q)^{-1}$ and the upper bound given in (3.23) for the last sum above, we get

$$|S_2| \leq_b \binom{n+m-1}{n-1} C_n (1-q)^{-\ell} \exp\left(-\tfrac{a^2}{2(1-q)}\right).$$

Using this inequality, the asymptotic value of S_1 given in (3.28), and $f(q) > 1-q$, we get

$$\frac{|S_2|}{S_1} \leq_b (1-q)^{-2\ell+1} \exp\left(-\tfrac{a^2}{2(1-q)}\right) \to 0$$

as $q = m/(m+n) \to 1$. This finishes the proof. $\square$

Remark 3.5. Whether the constraint (3.16) can be relaxed to, say, $m = \Theta(n \log n)$ is, in our opinion, a hard open problem.

As the first application, we apply Lemma 3.3 to determine the limit distribution of the number of *cuts* in a random linearized chord diagram. A *cut* is a partition of $[2n]$ into two blocks $[2n_1]$ and $[2n] \setminus [2n_1]$ such that there is no chord joining two points from different blocks. Let $X_{n,m}$ be the random variable counting the cuts in the linearized chord diagram chosen uniformly at random among all diagrams with m crossings.

Theorem 3.6. *Suppose that $n \to \infty$ and $m = O(n)$. Let $q = m/(m+n)$. Then $f(q) > 1/2$ and is bounded away from $1/2$, and for each $j \geq 0$,*

$$\mathrm{P}(X_{n,m} = j) = (j+1)(1-p)^2 p^j + o(1); \quad p = 1 - (2f(q))^{-1}. \tag{3.29}$$

Remark 3.7. We have $p = 1/2$ for $m = 0$, whence

$$\lim_{n\to\infty} \mathrm{P}(X_{n,0} = j) = (j+1)2^{-(j+2)}, \quad j \geq 0.$$

Also, a byproduct of this theorem is a pure-calculus inequality $f(x) > 1/2$ for $x \in [0, 1)$, which seems hard to prove out of the context of the chord diagrams.

Proof of Theorem 3.6. Notice upfront that

$$\mathrm{P}(X_{n,m} \geq 1) \geq \frac{T_{n-1,m}}{T_{n,m}},$$

as $T_{n-1,m}$ counts the linearized chord diagrams with an arc from the point 1 to the point 2. Therefore, for $m = O(n)$ as $n \to \infty$, Lemma 3.3 implies that

$$\liminf\ \mathrm{P}(X_{n,m} \geq 1) \geq \liminf \frac{C_{n-1}}{C_n} \cdot \frac{\binom{n+m-2}{n-2}}{\binom{n+m-1}{n-1}} = \liminf \frac{n-1}{4(n+m-1)} > 0. \tag{3.30}$$

Next, observe that

$$\begin{aligned}
\mathrm{E}\left[\binom{X_{n,m}}{k}\right] &= \frac{1}{T_{n,m}} \sum_{\substack{(n_1,m_1),\dots,(n_{k+1},m_{k+1}) \\ \sum_i n_i=n,\, \sum_j m_j=m;\ n_1,\dots,n_{k+1}>0}} \prod_{i=1}^{k+1} T_{n_i,m_i} \\
&= \frac{1}{T_{n,m}} [x^m y^n](T(x,y)-1)^{k+1} \\
&= \frac{1}{T_{n,m}} \sum_{\ell=0}^{k+1} (-1)^{k+1-\ell} \binom{k+1}{\ell} [x^m y^n](T(x,y))^{\ell}.
\end{aligned}$$

So, by Lemma 3.3,

$$\begin{aligned}
\mathrm{E}\left[\binom{X_{n,m}}{k}\right] &= o(1) + \sum_{\ell=0}^{k+1} (-1)^{k+1-\ell} \ell \binom{k+1}{\ell} (2f(q))^{\ell-1} \\
&= (k+1)(2f(q)-1)^k + o(1).
\end{aligned}$$

In particular, it follows that $\liminf f(q) \geq 1/2$. If $\liminf f(q) = 1/2$ then $\mathrm{E}[X_{n,m}] \to 0$ and $\mathrm{P}(X_{n,m} > 0) \to 0$, in violation of (3.30). Thus $\liminf f(q) > 1/2$ if $m = O(n)$; this effectively proves that $f(x) > 1/2$ for all $x \in [0,1)$. By the last equation, $X_{n,m}$ is asymptotic, in distribution, to X_q such that

$$\mathrm{E}\left[\binom{X_q}{k}\right] = (k+1)(2f(q)-1)^k.$$

Notice that, for $z > 0$ small enough,

$$\begin{aligned}
\mathrm{E}\left[z^{X_q}\right] &= \mathrm{E}\left[(1+(z-1))^{X_q}\right] = \sum_{k\geq 0} (z-1)^k \mathrm{E}\left[\binom{X_q}{k}\right] \\
&= \sum_{k\geq 0} (z-1)^k (k+1)(2f(q)-1)^k = \frac{1}{\left[1-(z-1)(2f(q)-1)\right]^2} \\
&= \left(\frac{1-p}{1-zp}\right)^2; \quad p = p(q) := 1-(2f(q))^{-1};
\end{aligned}$$

here $\limsup_{n\to\infty} p(q) < 1$ since $\liminf_{n\to\infty} f(q) > 1/2$. Note that $(1-p)/(1-zp)$ is the moment generating function of a Geometric random variable Y_q with success probability $(1-p)$, that is,

$$\mathrm{P}(Y_q = j) = (1-p)p^j, \quad j \geq 0.$$

Therefore, $X_q \overset{\mathcal{D}}{\equiv} Y_q' + Y_q''$, where Y_q' and Y_q'' are independent copies of the geometric Y_q. Hence

$$\mathrm{P}(X_q = j) = \sum_{i=0}^{j} \mathrm{P}(Y_q' = i)\ \mathrm{P}(Y_q'' = j-i) = (j+1)(1-p)^2 p^j.$$

Since $X_{n,m}$ is asymptotic in distribution to X_q, the last equation implies

$$\mathrm{P}(X_{n,m} = j) = \mathrm{P}(X_q = j) + o(1) = (j+1)(1-p)^2 p^j + o(1), \quad j \geq 0. \qquad \square$$

Remark 3.8. Intuitively, whp all the cuts are relatively close to the point 1 or point $2n$, and the numbers of those "left" and "right" cuts are asymptotically independent, each close to the geometric Y. We will prove the first part of this conjecture in the next section, as an immediate application of an upper bound for $T_{\nu,\mu}$ that holds for all values of the parameters ν and μ.

4. Bounds for $T_{n,m}$

For the proofs of our main results in Section 5, in addition to the sharp asymptotic formula (3.17) for $T_{n,m}$, we will also need some bounds for $T_{n,m}$ for m not in the range covered by this formula.

These bounds will be expressed in terms of the double-index sequence of the numbers $\{I_{n,m}\}$, where $I_{n,m}$ is the number of permutations of $[n]$ with m inversions. Each of those $I_{n,m}$ permutations $p = (p_1, \dots, p_n)$ gives rise to an *inversion sequence* $\boldsymbol{x} = (x_1, \dots, x_n)$: x_j is the number of pairs (p_i, p_j) such that $i < j$ and $p_i > p_j$. Obviously $x_i \le i - 1$ and $\sum_i x_i = m$. Conversely, every such sequence $\boldsymbol{x}$ determines a unique permutation p such that $\boldsymbol{x}$ is p's inversion sequence. Existence of this bijective correspondence implies a classic identity

$$I_{n,m} = [z^m] \prod_{j=0}^{n-1} (1 + z + \cdots + z^j) = [z^m] \prod_{j=1}^{n} \frac{1 - z^j}{1 - z}. \tag{4.1}$$

Probabilistically, the number of inversions, denote it I_n, in a uniformly random permutation of $[n]$ is distributed as $R_1 + \cdots + R_n$, where $R_1, \dots, R_n$ are independent, R_j being uniform on $[j-1]$. In particular, the likely values of I_n are of order $O(n^2)$.

Clearly, $I_{n,m}$ is at most the number of nonnegative integer solutions of the equation $x_1 + \cdots + x_n = m$, i.e.

$$I_{n,m} \le \binom{n+m-1}{n-1}. \tag{4.2}$$

Now we state a technical lemma that gives a lower bound for $I_{n,m}$ for $m \ll n^2$. Equation (4.2) and this lemma will be used in Lemma 4.4, which gives upper and lower bounds for $T_{n,m}$.

Lemma 4.1. *Suppose that $m/n \to \infty$ and $m = o\big(n^{3/2}\big)$. Then,*

$$I_{n,m} \ge \binom{n+m-1}{n-1} \exp\left(-\frac{\pi^2 m}{6n} + O(\log n)\right).$$

We note that Louchard and Prodinger [34], see other references therein, used a saddle-point method to find very sharp asymptotic formulas for $I_{n,m}$ in the cases $m = O(n)$ and $m = \Theta(n^2)$. Our argument is based on reduction to a local limit theorem for the sum of independent random variables with log-concave distributions established by Bender [10] and Canfield [16] about forty years ago.

Proof. Pick $\rho \in (0,1)$, and introduce the sequence $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ of *independent* random variables such that

$$\mathrm{P}(Y_i = j) = \frac{(1-\rho)\rho^j}{1-\rho^i}, \quad 0 \le j \le i-1;$$

so

$$\mathrm{E}\big[z^{Y_i}\big] = \frac{1-\rho}{1-\rho^i} \cdot \frac{1-(\rho z)^i}{1-\rho z}, \quad 1 \le i \le n.$$

Then (4.1) becomes

$$\begin{aligned} I_{n,m} &= \rho^{-m} [z^m] \prod_{i=1}^{n} \frac{1-(\rho z)^i}{1-\rho z} = \rho^{-m} \prod_{i=1}^{n} \frac{1-\rho^i}{1-\rho} \; [z^m] \prod_{i=1}^{n} \mathrm{E}\big[z^{Y_i}\big] \\ &= \rho^{-m} \prod_{i=1}^{n} \frac{1-\rho^i}{1-\rho} \; \mathrm{P}(\|\boldsymbol{Y}\| = m), \end{aligned} \tag{4.3}$$

where $\|\boldsymbol{Y}\| := \sum_i Y_i$. In particular,

$$I_{n,m} \le \rho^{-m} \prod_{i=1}^{n} \frac{1-\rho^i}{1-\rho}, \tag{4.4}$$

and we get the sharpest upper bound by selecting ρ^* that minimizes the right side of (4.4). Since $I_{n,m}$ does not depend on ρ, this ρ^* maximizes $\mathrm{P}(\|\boldsymbol{Y}\| = m)$. Lowering the bar, we are content to prove existence of a stationary point ρ^*, for which we will be able to bound from *below* both the RHS of (4.4) and $\mathrm{P}(\|\boldsymbol{Y}\| = m)$, whence $I_{n,m}$ itself.

Crucially, the distribution of $\sum_i Y_i$ is log-concave, i.e.

$$\mathrm{P}(\|\boldsymbol{Y}\| = j)^2 \geq \mathrm{P}(\|\boldsymbol{Y}\| = j-1)\,\mathrm{P}(\|\boldsymbol{Y}\| = j+1), \quad j \geq 0. \tag{4.5}$$

The reason is that each Y_i has a log-concave distribution and the convolution of log-concave distributions is log-concave as well, Menon [35]. Even stronger, in terminology of Canfield [16], the distribution of $\|\boldsymbol{Y}\|$ is *properly* log-concave, meaning that (a) the range of $\|\boldsymbol{Y}\|$ has no gaps, and (b) the equality in (4.5) holds only if $\mathrm{P}(\|\boldsymbol{Y}\| = j) = 0$.

Indeed, $Y_1 = 0$ is properly log-concave distributed, and (induction step) proper log-concavity of $Z_{s+1} := \sum_{r=1}^{s+1} Y_r$ for $s \geq 1$ follows from proper log-concavity of $Z_s := \sum_{r=1}^{s} Y_r$ and the identity [16]

$$P_{s+1,\nu}^2 - P_{s+1,\nu-1}P_{s+1,\nu+1} = \sum_{\alpha<\beta} \big(P_{s,\alpha}P_{s,\beta-1} - P_{s,\alpha-1}P_{s,\beta}\big)\big(p_{s+1,\nu-\alpha}\,p_{s+1,\nu-\beta+1} - p_{s+1,\nu-\alpha+1}\,p_{s+1,\nu-\beta}\big), \tag{4.6}$$

$P_{t,\mu} := \mathrm{P}(Z_t = \mu)$, $p_{t,\mu} := \mathrm{P}(Y_t = \mu)$. (See the remark following the proof.) Here is how. Each summand on the RHS of (4.6) is non-negative as both Z_s and Y_{s+1} are log-concave, and their respective ranges, $[0, 1, \dots, \binom{s}{2}]$ and $[0, 1, \dots, s]$, have no gaps. If $\nu \leq \binom{s}{2}$, we see that the summand for $\alpha = \nu$, $\beta = \nu + 1$ is

$$(P_{s,\nu}^2 - P_{s,\nu-1}P_{s,\nu+1})p_{s+1,0}^2 > 0,$$

because, by inductive hypothesis, $\{P_{s,t}\}$ is properly log-concave and $P_{s,\nu} > 0$. If $\binom{s}{2} < \nu \leq \binom{s+1}{2}$, we consider $\alpha = \nu - s$ and $\beta = \alpha + 1$. Then

$$\alpha = \nu - s \leq \binom{s+1}{2} - s = \binom{s}{2},$$
$$\alpha = \nu - s \geq \binom{s}{2} + 1 - s = \binom{s-1}{2} \geq 0,$$

whence $P_{s,\alpha} > 0$. Then the corresponding summand on the RHS of (4.6) is

$$(P_{s,\alpha}^2 - P_{s,\alpha-1}P_{s,\alpha+1})(p_{s+1,s}^2 - p_{s+1,s+1}p_{s+1,s-1}) > 0,$$

because, by inductive hypothesis, $P_{s,\alpha}^2 - P_{s,\alpha-1}P_{s,\alpha+1} > 0$, and $p_{s+1,s} > 0$, $p_{s+1,s+1} = 0$.

For $x \in (0, 1)$, introduce

$$\begin{aligned} L(x) :&= \log\left(x^{-m}\prod_{i=1}^{n}\frac{1-x^i}{1-x}\right) \\ &= -m\log x + \sum_{i=1}^{n}(\log(1-x^i) - \log(1-x)). \end{aligned}$$

The stationary points of $L(x)$ are the roots, *if any exist*, of

$$L'(x) = \frac{1}{x}\left(n\frac{x}{1-x} - m - \sum_i \frac{ix^i}{1-x^i}\right) = 0. \tag{4.7}$$

As an approximation for a possible root of $L'(x) = 0$, pick a constant $A > 0$ and introduce $\rho(A) = q(1 + A/n)$, $q := m/(m+n)$. Since $m/n^2 \to 0$, we have

$$1 - \rho = (1-q)\big(1 + O(m/n^2)\big),$$

whence

$$n\rho/(1-\rho) - m = A\,(m/n)^2 + O(m/n). \tag{4.8}$$

Further, approximating the sum $\sum_i i\rho^i/(1-\rho^i)$ by the corresponding integral we obtain

$$\sum_i \frac{i\rho^i}{1-\rho^i} = \frac{1}{(\log\rho)^2}\int_0^\infty \frac{x}{e^x-1}\,dx + O((1-\rho)^{-1})$$
$$= \left(\pi^2/6\right)(m/n)^2 + O(m/n). \tag{4.9}$$

Comparing (4.8) and (4.9), and using (4.7), we see that, for n large enough, $L'(\rho(A)) < 0$ for $A < \pi^2/6$ and $L'(\rho(A)) > 0$ if $A > \pi^2/6$. Thus the equation $L'(x) = 0$ does have a root ρ such that $\rho = q\big(1 + O(n^{-1})\big)$. Furthermore, uniformly for x between ρ and q,

$$L''(x) = \frac{m}{x^2} + \frac{n}{(1-x)^2} + \sum_i \left(\frac{ix^{i-2}}{1-x^i} - \frac{i^2x^i}{(1-x^i)^2}\right)$$
$$= O(m + m^2/n + m^3/n^3) = O(m^2/n).$$

Therefore

$$L(\rho) = L(q) + O\big(m^2n^{-1}(\rho-q)^2\big) = L(q) + O(m^2/n^3) = L(q) + o(1),$$

since $m = o(n^{3/2})$. This is equivalent to

$$\rho^{-m}\prod_{i=1}^n \frac{1-\rho^i}{1-\rho} = \big(1 + O(m^2/n^3)\big)q^{-m}\prod_{i=1}^n \frac{1-q^i}{1-q}.$$

Here, by Lemma 3.11 in [3], we have

$$\prod_{j=1}^n (1-q^j) = \big(1 + O(m^2/n^3)\big)\prod_{j=1}^\infty (1-q^j),$$

and, by the proof of Lemma 3.14 in [3], we have

$$\prod_{j=1}^\infty (1-q^j) \sim K\cdot\exp\big(-(\pi^2/6)(m/n) + (1/2)\log(m/n)\big)\,,$$

where $K = \sqrt{2\pi}\cdot e^{-\pi^2/12}$. Finally,

$$q^{-m}(1-q)^{-n} = \frac{(m+n)^{m+n}}{m^m n^n} \ge \binom{n+m}{n} \ge \binom{n+m-1}{n-1}.$$

Combining the last four equations, we conclude that

$$\rho^{-m}\prod_{i=1}^n \frac{1-\rho^i}{1-\rho} \ge \binom{n+m-1}{n-1}\exp\left(-\frac{\pi^2 m}{n} + O(\log n)\right). \tag{4.10}$$

It remains to evaluate $\mathrm{P}(\|\mathbf{Y}\| = m)$. By independence of $Y_1,\dots,Y_n$, using Berry-Esseen inequality (Feller [26], Ch. XVI, Section 5),

$$\max_{x\in\mathbb{R}}\left|\,\mathrm{P}\left(\|\boldsymbol{Y}\| \le \mathrm{E}[\|\boldsymbol{Y}\|] + x\sigma(\|\boldsymbol{Y}\|)\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-y^2/2}\,dy\right| \le 6\frac{r_3}{\sigma^3},$$

where

$$\sigma^2 = \mathrm{Var}(\|\boldsymbol{Y}\|) = \sum_{i=1}^n \mathrm{Var}(Y_i) = \sum_{i=1}^n \mathrm{E}\big[(Y_i - \mathrm{E}[Y_i])^2\big],$$
$$r_3 = \sum_{i=1}^n \mathrm{E}\big[|Y_i - \mathrm{E}[Y_i]|^3\big].$$

To compute σ^2 we use

$$\mathrm{E}\big[z^{\|\boldsymbol{Y}\|}\big]=\prod_{i=1}^{n}\mathrm{E}\big[z^{Y_i}\big]=\prod_{i=1}^{n}\frac{1-\rho}{1-\rho^i}\cdot\frac{1-(\rho z)^i}{1-\rho z},$$

and

$$\frac{d^2}{dz^2}\mathrm{E}\big[z^{\|\boldsymbol{Y}\|}\big]\bigg|_{z=1}=\mathrm{E}\big[(\|\boldsymbol{Y}\|)_2\big].$$

Computing the derivative and bounding the resulting sum by the integral we obtain

$$\begin{aligned}\sigma^2&=\mathrm{E}\big[(\|\boldsymbol{Y}\|)_2\big]+\mathrm{E}[\|\boldsymbol{Y}\|]-\big(\mathrm{E}[\|\boldsymbol{Y}\|]\big)^2\\&=n\,\frac{\rho}{(1-\rho)^2}-\sum_{i=1}^{n}\frac{i^2\rho^i}{(1-\rho^i)^2}\\&=\big(1+o(1)\big)m^2/n-\Theta((m/n)^3)=\big(1+o(1)\big)m^2/n.\end{aligned}$$

Similar, but more protracted, computations lead to

$$\begin{aligned}r_4:=\sum_{i=1}^{n}\mathrm{E}\big[(Y_i-\mathrm{E}[Y_i])^4\big]&=(1+o(1))\frac{n}{(1-\rho)^4}\\&=\big(1+o(1)\big)m^4/n^3.\end{aligned}$$

Therefore,

$$r_3\le n^{1/4}(r_4)^{3/4}=(1+o(1))m^3/n^2.$$

Consequently, for n large enough,

$$\max_{x\in\mathbb{R}}\left|\,\mathrm{P}\left(\|\boldsymbol{Y}\|\le\mathrm{E}[\|\boldsymbol{Y}\|]+x\sigma(\|\boldsymbol{Y}\|)\right)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\,dy\right|\le 7n^{-1/2}.\tag{4.11}$$

If we write $7n^{-1/2}=K/\sigma$, then

$$\frac{K}{\sqrt{\sigma}}=7\sqrt{\frac{\sigma}{n}}\le 8\sqrt{\frac{m}{n^{3/2}}}\to 0$$

since $m\ll n^{3/2}$. To finish the proof we will use the following local limit theorem by Canfield.

Theorem 4.2. *(Canfield [16]) Suppose that X has a properly log-concave distribution and that*

$$\sup_{x\in\mathbb{R}}\left|\,\mathrm{P}\big(X\le\mathrm{E}[X]+x\sigma(X)\big)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\,dy\right|\le\frac{K}{\sigma(X)}.$$

If $K>7$, $K/\sigma(X)<10^{-7}$, and $K/\sigma(X)^{1/2}<10^{-2}$, then

$$\sup_{m}\left|\,\mathrm{P}(X=m)-\frac{1}{\sqrt{2\pi}\sigma(X)}\exp\left(-\frac{(m-\mathrm{E}[X])^2}{2\sigma^2(X)}\right)\right|\le\frac{c}{\sigma(X)^{3/2}},$$

with $c:=14.5K+4.87$.

In light of (4.11), it follows from Canfield's theorem that

$$\mathrm{P}(\|\boldsymbol{Y}\|=m)=\frac{1+o(1)}{\sqrt{2\pi\mathrm{Var}(\|\boldsymbol{Y}\|)}}=\Theta(n^{1/2}m^{-1}).\tag{4.12}$$

Combining (4.3), (4.12) and (4.10), we complete the proof. $\square$

Remark 4.3. In Canfield [16], the striking identity (4.6) was used to show that the convolution operation preserves the proper log-concavity. We had to use this identity differently, i.e. inductively, because none of $Y_3,\dots,Y_n$ is *properly* log-concave. Notice also that for $\rho=1$ our claim reduces to proper logconcavity of $I_{n,m}$ for every $n\ge 1$. The usual logconcavity of this sequence is long known, of course. More recently Bóna [14] found a purely combinatorial proof of this property, a proof that does not rely on Menon's theorem.

The next lemma provides an upper bound for $T_{n,m}$ applicable to all m, and a lower bound for $T_{n,m}$ in the case when m meets the condition of Lemma 4.1, i.e. far beyond the constraint of Lemma 3.4.

Lemma 4.4. *(i) For all $m, n \geq 0$,*

$$T_{n,m} \leq C_n I_{n,m} \leq C_n \binom{n+m-1}{n-1}. \tag{4.13}$$

(ii) If $n \to \infty$ and $m = o(n^{3/2})$, then

$$T_{n,m} \geq_b \exp\left[-2(m/n)(\log n)\right] C_n \binom{n+m-1}{n-1}. \tag{4.14}$$

Proof. (i) If a chord connects i and j, where $i < j$, we call i and j the *initial point* and the *terminal point* of the chord, respectively. We label the chords of a chord diagram according to the ordering of initial points; the initial point of the i-th chord is smaller than the initial point of the $(i+1)$-st one for $1 \leq i \leq n-1$. Let $\mathcal{D}$ be a chord diagram with n chords and m crossings. Let y_i be the number of terminal points between the initial points of the i-th and $(i+1)$-st chords of $\mathcal{D}$. Clearly, we have

$$y_1 + \cdots + y_k \leq k, \ \forall k < n \quad \text{and} \quad y_1 + \cdots + y_n = n. \tag{4.15}$$

Note that the number of sequences satisfying (4.15) is given by the n-th Catalan number C_n and we call such sequences as *Catalan sequences*. Also, let x_i be the number of chords among the first $i-1$ chords of $\mathcal{D}$ that cross the $i-$th chord. Clearly, $0 \leq x_i \leq i-1$ and $x_1 + \cdots + x_n = m$ and so $(x_1, \ldots, x_n)$ is an inversion sequence. Hence $\mathcal{D}$ gives a pair of n-long sequences $(\boldsymbol{y}, \boldsymbol{x}) = (\boldsymbol{y}(\mathcal{D}), \boldsymbol{x}(\mathcal{D}))$ consisting of a Catalan sequence and an inversion sequence with a total of m inversions.

In order to conclude the proof of the first part, we need to show that $(\boldsymbol{y}(\mathcal{D}_1), \boldsymbol{x}(\mathcal{D}_1)) \neq (\boldsymbol{y}(\mathcal{D}_1), \boldsymbol{x}(\mathcal{D}_1))$ for $\mathcal{D}_1 \neq \mathcal{D}_2$. First, if $\boldsymbol{y}(\mathcal{D}_1) = \boldsymbol{y}(\mathcal{D}_2)$, then the initial points of $\mathcal{D}_1$ agree with the initial points of $\mathcal{D}_2$. Now note that, for $i < j$, the i-th and j-th chords cross each other if the terminal point of the i-th chord lies between the initial and terminal points of the j-th chord. Consequently, if $\boldsymbol{x}(\mathcal{D}_1) = \boldsymbol{x}(\mathcal{D}_2)$ in addition to $\boldsymbol{y}(\mathcal{D}_1) = \boldsymbol{y}(\mathcal{D}_2)$, then the terminal points of all the chords in the two diagrams also agree and hence $\mathcal{D}_1 = \mathcal{D}_2$, which finishes the first part.

(ii) In the first part we showed that each chord diagram gives a unique pair $(\boldsymbol{y}, \boldsymbol{x})$. To find a lower bound on $T_{n,m}$, we need to work the other way around. Let $\mathcal{D}$ be a chord diagram and let $(\boldsymbol{y}, \boldsymbol{x})$ be the corresponding pair of sequences. Let F_j and L_j correspond to the initial (first) and the terminal (last) points of the j-th chord of $\mathcal{D}$. The chord labeled with $(n-k)$ intersects x_{n-k} chords of smaller label and thus at least x_{n-k} of the terminal points from $L_1, \ldots, L_{n-k-1}$ must appear after F_{n-k}. On the other hand, a total number of $y_{n-k} + \cdots + y_n$ terminal points appear after F_{n-k}, of which $k+1$ of them are $L_{n-k}, \ldots, L_n$. Thus, we get the following set of inequalities:

$$x_{n-k} \leq y_{n-k} + y_{n-k+1} + \cdots + y_n - (k+1), \quad 0 \leq k \leq n-1. \tag{4.16}$$

We now claim that, conversely, a pair $(\boldsymbol{y}, \boldsymbol{x})$ satisfying the inequalities in (4.16) corresponds to a chord diagram. To prove this, first note that $\boldsymbol{y}$ alone determines the initial points $F_1, \ldots, F_n$ of the chords, so we only need to recover the terminal points $L_1, \ldots, L_n$ using (4.16). Now we find $L_n, L_{n-1}, \ldots, L_1$ in this order as follows: L_n is the (x_n+1)-st number bigger than F_n and once $L_n, \ldots, L_{n-k+1}$ are determined, to ensure that the $(n-k)$-th chord intersects x_{n-k} chords of smaller label, we must choose L_{n-k} as (in their natural order) the $(x_{n-k}+1)$-st available number bigger than F_{n-k}, that is, $(x_{n-k}+1)$-st number in $\{F_{n-k}+1, \ldots, 2n\} \setminus \{F_{n-k+1}, \ldots, F_n, L_{n-k+1}, \ldots, L_n\}$. This choice is possible by (4.16) since there are $y_{n-k} + \cdots + y_n - k$ available numbers after choosing $L_n, \ldots, L_{n-k-1}$.

It remains to show that the right hand side of (4.14) is a lower bound on the number pairs $(\boldsymbol{y}, \boldsymbol{x})$ satisfying (4.16). For a given $\boldsymbol{x}$, let $N(\boldsymbol{x})$ denote the number of $\boldsymbol{y}$'s meeting the constraint (4.16).

Claim. Let $M = M(\boldsymbol{x})$ denote the maximum of the n terms in the inversion sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$. Then, $N(\boldsymbol{x}) \geq C_{n-M}$.

Proof of the claim. Let $\mathcal{A} = \mathcal{A}(M)$ be the set of Catalan sequences $\boldsymbol{y} = (y_1, \dots, y_n)$ such that $y_i = 0$ for $i \le M$ and $\boldsymbol{y}' := (y_{M+1}, \dots, y_n - M)$ is also a Catalan sequence. Since $\boldsymbol{y}'$ is a Catalan sequence, by (4.15), for $0 \le k < n - M$,

$$y_{n-k} + \cdots + y_{n-1} + (y_n - M) \ge k + 1.$$

This inequality, together with the fact that $\max_i x_i = M$, gives

$$x_{n-k} \le M \le y_{n-k} + \cdots + y_{n-1} + y_n - (k+1).$$

For $n - M \le k < n$, we have

$$x_{n-k} \le n - k - 1 = y_{M+1} + \cdots + y_n - (k+1) = y_{n-k} + \cdots + y_n - (k+1),$$

where the inequality follows from the fact that $\boldsymbol{x}$ is an inversion sequence and the equalities follow from the definition of $\mathcal{A}$. By the last two equations, any $\boldsymbol{y} \in \mathcal{A}$ satisfies the condition (4.16). This finishes the proof since $|\mathcal{A}| = C_{n-M}$. □

By the claim above, we have

$$T_{n,m} = \sum_{\boldsymbol{x}} N(\boldsymbol{x}) \ge \sum_{\boldsymbol{x}} C_{n-M(\boldsymbol{x})}. \tag{4.17}$$

In the proof of Lemma 3.4 in [3] it was shown that, whp, the maximum $M(\boldsymbol{x})$ does not exceed $(1+\varepsilon)(m/n)\log n$ when $\boldsymbol{x} = (x_1, \dots, x_n)$ is chosen uniformly at random from inversion sequences with m inversions. Using this fact and (4.17), we get: for $M_0 = \lceil ((1+\varepsilon)m/n)\log n \rceil$,

$$\limsup \frac{I_{n,m} C_{n-M_0}}{T_{n,m}} \le 1, \tag{4.18}$$

Also, by the Stirling's formula for the factorials,

$$C_{n-M_0} \sim 4^{-M_0} \cdot C_n = \exp(-M_0 \log 4) C_n. \tag{4.19}$$

Combining (4.18) and (4.19), with small enough ε, and Lemma 4.1 we complete the proof. □

Remark 4.5. A closer look shows that, in fact, $M(\boldsymbol{x})$ is asymptotic to $(m/n)\log n$ in probability, and that $\mathrm{P}(M(\boldsymbol{x}) \le (1-\varepsilon)(m/n)\log n) \le \exp(-cn^{\varepsilon})$, which is much smaller than $\exp(-\Theta(m/n))$. Thus the choice of M_0 in (4.18) is asymptotically the best possible if we want the fraction $I_{n,m,M_0}/I_{n,m}$ to be at least $e^{-bm/n}$ for some constant $b > 0$; here the I_{n,m,M_0} denotes the number of permutations with m inversions and $\max x_i \le M_0$.

Remark 4.6. Our, admittedly limited, numerical experiments seem to indicate that, for $m = \Theta(n \log n)$, $T_{n,m}$ is at least of order $e^{-b(m/n)} C_n \binom{n+m-1}{n-1}$ for some constant $b > 0$, a bound that matches qualitatively the asymptotic formula for $T_{n,m}$ for $m < (2/\pi^2) n \log n$ in Lemma 3.4. However, the exponential factor in the lower bound (4.14) is much smaller, namely $e^{-\Theta((m/n)^2)}$. So far we have not been able to replace this factor by anything substantially larger. At the moment, it seems that $n^{3/2}$ is actually the threshold value of m for validity of the lower bound. Here is a quick-and-dirty argument to lend some support for this conjecture. A pair $(\boldsymbol{x}, \boldsymbol{y})$ determines a chord diagram if and only if the condition (4.16) is satisfied. For a typical Catalan sequence $\boldsymbol{y}$, we have

$$\max_{0 \le k < n} \{y_{n-k} + y_{n-k+1} + \cdots + y_n - (k+1)\} = O(\sqrt{n}).$$

On the other hand, an average x_i is of order m/n, which is much larger than $\sqrt{n}$ for $m \gg n^{3/2}$. Consequently, the probability that (4.16) is satisfied for a random $\boldsymbol{x}$ and a random $\boldsymbol{y}$ is extremely small. However, we do not know how to handle non-typical Catalan sequences; so we cannot exclude the possibility that the conjecture is false.

Remark 4.7. By (4.13), for $x, y > 0$,

$$\begin{aligned} T(x,y) = \sum_{m,n} T_{n,m} x^m y^n &\le \sum_{m,n} \binom{n+m-1}{n-1} C_n x^m y^n \\ &= \sum_n y^n C_n \sum_m x^m \binom{n+m-1}{n-1}. \end{aligned}$$

For $x < 1$, the innermost series converges to $(1-x)^{-n}$, and then the double series converges to $C(y/(1-x))$ if $y/(1-x) < 1/4$. Therefore, we have an elementary proof that the bivariate generating function series $T(x, y)$ converges if $x, y > 0$ and $y/(1-x) < 1/4$.

Here is an illustration of the power of the upper bound (4.13) combined with Lemma 3.3. Consider again the uniformly random linearized chord diagram on $[2n]$ with m crossings. For a cut $\mathcal{K} = [2n_1] \cup ([2n] \setminus [2n_1])$, we set $n_2 = n - n_1$, define $|\mathcal{K}| = \min\{n_1, n_2\}$, and finally define $Y_{n,m} = \max_{\mathcal{K}} |\mathcal{K}|$.

Lemma 4.8. *If $m = O(n)$ then $Y_{n,m}$ is bounded in probability.*

Proof. Given $n_1 + n_2 = n$ and $m_1 + m_2 = m$, where $n_1, n_2 > 0$, the expected number of cuts with parts $[2n_1]$ and $[2n_1 + 1, 2n]$, and the number of crossings in the left subdiagram and the right subdiagram equal m_1 and m_2, respectively, is

$$Z_{\boldsymbol{n},\boldsymbol{m}} := \frac{T_{n_1,m_1} T_{n_2,m_2}}{T_{n,m}},$$

where $\boldsymbol{n} = (n_1, n_2)$ and $\boldsymbol{m} = (m_1, m_2)$. By Lemma 3.3,

$$T_{n,m} \sim \binom{n+m-1}{n-1} C_n \prod_{j\ge 1} (1-q^j)^3,$$

and, by (4.13),

$$T_{n_i,m_i} \le \binom{n_i+m_i-1}{n_i-1} C_{n_i}, \quad i = 1, 2.$$

Hence,

$$Z_{\boldsymbol{n},\boldsymbol{m}} \le_b \frac{\prod_i \binom{n_i+m_i-1}{n_i-1} C_{n_i}}{\binom{n+m-1}{n-1} C_n} \le \frac{\prod_i \binom{n_i+m_i}{n_i} C_{n_i}}{\binom{n+m}{n} C_n}.$$

Therefore, since $C_\nu = \Theta\big(\nu^{-3/2} 4^\nu\big)$,

$$Z_{\boldsymbol{n},\boldsymbol{m}} \le_b \frac{n^{3/2}}{n_1^{3/2} n_2^{3/2}} \cdot \frac{\prod_i \binom{n_i+m_i-1}{n_i-1}}{\binom{n+m-1}{n-1}}. \tag{4.20}$$

Observe that

$$\sum_{\boldsymbol{m}: m_1+m_2=m} \prod_i \binom{n_i+m_i-1}{n_i-1} = \binom{n+m-1}{n-1}.$$

Indeed, the RHS is the total number of non-negative integer solutions of

$$\sum_{j=1}^{n_1} x_j + \sum_{j=n_1+1}^{n} x_j = m,$$

and each such solution is a pair $(x_1, \dots, x_{n_1})$, $(x_{n_1+1}, \dots, x_{n_1+n_2})$ of solutions, each of the corresponding equation

$$\sum_{j=}^{n_1} x_j = m_1, \qquad \sum_{j=n_1+1}^{n} x_j = m_2,$$

for the unique choice of m_1, m_2 satisfying $m_1 + m_2 = m$. So summing (4.20) over $\boldsymbol{m}$, we get

$$\sum_{\boldsymbol{m}: m_1+m_2=m} Z_{\boldsymbol{n},\boldsymbol{m}} \le_b \frac{n^{3/2}}{n_1^{3/2} n_2^{3/2}}.$$

Consequently, as $A \to \infty$,

$$\begin{aligned} \mathrm{P}(Y_{n,m} \ge A) &\le \sum_{\boldsymbol{n}: \min\{n_1,n_2\}\ge A} \; \sum_{\boldsymbol{m}: m_1+m_2=m} Z_{\boldsymbol{n},\boldsymbol{m}} \\ &\le_b n^{3/2} \sum_{\min\{n_1,n_2\}\ge A} n_1^{-3/2} n_2^{-3/2} \\ &\le_b \sum_{A \le n_1 \le n/2} n_1^{-3/2} = O(A^{-1/2}) \to 0. \end{aligned}$$

□

5. The Largest Component

We now turn our attention to the component sizes of chord diagrams with given number m of crossings. Throughout this section, unless otherwise stipulated, we will assume that m satisfies the condition (3.16) with $\ell = 1$, so that $T_{n,m}$ is given by the asymptotic formula (3.17) with $\ell = 1$.

Like the classic case of the Erdős-Rényi random graph, (Bollobás [11]), we need a usable bound for $\mathcal{C}_{\nu,\mu}$, the total number of connected chord diagrams on $[2\nu]$ with μ crossings. $\{\mathcal{C}_{\nu,\mu}\}$ and its bivariate generating function $\mathcal{C}(x, y)$ below will, hopefully, not be confused with the Catalan number C_n and its generating function $C(y)$. It was first found by Dulucq and Peanud [23] (see also Stanley [44, Exercise 5.46]) that

$$\mathcal{C}_{\nu,\nu-1} = \frac{1}{2\nu-1}\binom{3\nu-3}{\nu-1},$$

and it was proved in [1] that

$$\mathcal{C}_{\nu,\nu} = 2 + \sum_{j=1}^{\min(6,\nu-3)} \frac{\nu}{3}\binom{6}{j}\frac{j}{\nu-3}\binom{3\nu-9}{\nu-3-j} + 2\sum_{k=4}^{\nu-1}\frac{\nu}{k}\sum_{j=1}^{\min(\nu-k,2k)}\frac{j}{\nu-k}\binom{2k}{j}\binom{3\nu-3k}{\nu-k-j}.$$

Using

$$j\binom{d}{j} = d\binom{d-1}{j-1}, \quad \sum_j \binom{a}{j}\binom{b}{c-j} = \binom{a+b}{c},$$

the above formula is simplified to

$$\mathcal{C}_{\nu,\nu} = 2 + \binom{3\nu-4}{\nu-3} + 2\sum_{k=4}^{\nu-1}\binom{3\nu-k-1}{2\nu-1}.$$

Thus, as $\nu \to \infty$,

$$\mathcal{C}_{\nu,\nu-1} \sim \frac{2}{3^3}\nu^{-1}\binom{3\nu}{\nu}, \qquad \mathcal{C}_{\nu,\nu} \sim \frac{2^2}{3^4}\binom{3\nu}{\nu}.$$

We conjecture that, more generally, for $\mu = O(\nu)$,

$$\mathcal{C}_{\nu,\mu} = \nu^{\mu-\nu+o(\nu)}\binom{3\nu}{\nu}.$$

For a chord diagram $\mathcal{D}$, we call the component of $\mathcal{D}$ that contains point 1, the *root component*. The endpoints of the root component's chords determine pairwise disjoint arcs, each of which contains a chord diagram, possibly an empty one. Note that we cannot have any chord joining two points from two different arcs, as otherwise, that chord would belong to the root component. If $\mathcal{D}$ has parameters (n, m), the root component has parameters (ν, μ), and the 2ν diagrams determined by the root component have parameters $(n_1, m_1), \ldots, (n_{2\nu}, m_{2\nu})$, we have

$$n_1 + \cdots + n_{2\nu} = n - \nu, \quad m_1 + \cdots + m_{2\nu} = m - \mu.$$

Thus,

$$T_{n,m} = \sum_{\substack{\nu\ge 1\\ \mu\ge\nu-1}} \mathcal{C}_{\nu,\mu} \sum_{\substack{n_1+\cdots+n_{2\nu}=n-\nu\\ m_1+\cdots+m_{2\nu}=m-\mu}} \prod_{j=1}^{2\nu} T_{n_j,m_j}.$$

Setting $\mathcal{C}_{0,0} = 1$, we get

$$\begin{aligned}\sum_{n,m} T_{n,m}x^m y^n &= 1 + \sum_{\nu,\mu}\mathcal{C}_{\nu,\mu}x^\mu y^\nu \sum_{\substack{n_1+\cdots+n_{2\nu}\ge 0\\ m_1,\ldots,m_{2\nu}\ge 0}} \prod_{j=1}^{2\nu} T_{n_j,m_j}x^{m_j}y^{n_j}\\ &= \sum_{\nu,\mu}\mathcal{C}_{\nu,\mu}x^\mu y^\nu \left(\sum_{n_1\ge 0, m_1\ge 0} T_{n_1,m_1}x^{m_1}y^{n_1}\right)^{2\nu}.\end{aligned}$$

Equivalently,

$$T(x,y) = \mathcal{C}(x, yT^2(x,y)), \tag{5.1}$$

where $\mathcal{C}(x,y) := \sum_{\mu,\nu} \mathcal{C}_{\nu,\mu} x^\mu y^\nu$ denotes the bivariate generating function for the sequence $\{\mathcal{C}_{\nu,\mu}\}$. For $x = 1$ we get the identity proved earlier by Flajolet and Noy [27].

Equation (5.1) implies a Chernoff-type bound for $\mathcal{C}_{\nu,\mu}$:

$$\mathcal{C}_{\nu,\mu} \le \frac{T(x,y)}{x^\mu y^\nu [T(x,y)]^{2\nu}}, \quad \forall\, x < 1,\ y < \frac{1-x}{4}. \tag{5.2}$$

Since $T(x,y)$ increases with y, the best estimate, for a given $x < 1$, is obtained by letting $y \uparrow (1-x)/4$. From (3.1), and $C(1/4) = 2$, it follows that

$$\lim_{y \uparrow (1-x)/4} T(x,y) = 2f(x), \quad f(x) = \sum_{j \ge 0} (-1)^j x^{\binom{j+1}{2}}.$$

Using these and $1 - x < f(x) < 1$ in (5.2), we obtain

$$\mathcal{C}_{\nu,\mu} \le \frac{2f(x)}{x^\mu (1-x)^\nu f(x)^{2\nu}} \le \frac{2}{x^\mu (1-x)^{3\nu}}, \quad \forall\, x < 1.$$

The RHS is minimized at $x = \mu/(3\nu + \mu)$, and we get

$$\mathcal{C}_{\nu,\mu} \le 2\, \frac{(3\nu+\mu)^{3\nu+\mu}}{(3\nu)^{3\nu} \mu^\mu}. \tag{5.3}$$

In particular,

$$\mathcal{C}_{\nu,\nu-1} \le_b \frac{(4\nu-1)^{4\nu-1}}{(3\nu)^{3\nu}(\nu-1)^{\nu-1}} \le_b \nu^{1/2} \binom{4\nu}{\nu},$$

similar to, but noticeably worse than the exact formula for the number of trees, which is $\frac{1}{2\nu-1}\binom{3\nu-3}{\nu-1}$. For μ/ν large, we get a bound better than (5.3) by using the obvious inequality $\mathcal{C}_{\nu,\mu} < T_{\nu,\mu}$ and (4.13):

$$\mathcal{C}_{\nu,\mu} \le C_\nu \binom{\mu+\nu-1}{\nu-1} \le_b \frac{4^\nu}{\nu^2} \cdot \frac{(\mu+\nu)^{\mu+\nu}}{\mu^\mu \nu^\nu} \le 4^\nu \frac{(\mu+\nu)^{\mu+\nu}}{\mu^\mu \nu^\nu}. \tag{5.4}$$

Combining (5.3) and (5.4), we obtain

$$\mathcal{C}_{\nu,\mu} \le_b \min\left\{ \frac{4^\nu}{\nu^2} \cdot \frac{(\mu+\nu)^{\mu+\nu}}{\mu^\mu \nu^\nu}, \frac{(3\nu+\mu)^{3\nu+\mu}}{(3\nu)^{3\nu}\mu^\mu} \right\}. \tag{5.5}$$

In combination with the bounds on $T_{n,m}$, these enumerative results will enable us to gain an insight into the component structure of the random intersection graph, and to prove eventually the main result on formation of its giant component.

Lemma 5.1 (Crossing-density gap)**.** *Define the crossing density of a chord diagram as the ratio of its number of crossings to the number of chords. Let α be a constant greater than $4e^2$ and let $\beta = \frac{5}{\log\alpha - \log(4e^2)}$. For crossing density m/n exceeding α, whp, the intersection graph has no component of size above $\beta \log n$ whose crossing density is below $m/(\alpha n)$.*

Proof. Let $\mathcal{X}_{\nu,\mu}$ denote the number of components with parameters ν and μ in a random circular diagram with n chords and m crossings. We first bound the expected number of $\mathcal{X}_{\nu,\mu}$. The probability $P_{\nu,\mu}$ that the root component has parameters (ν,μ) is given by

$$P_{\nu,\mu} = \frac{\mathcal{C}_{\nu,\mu}}{T_{n,m}} \sum_{\substack{n_1+\cdots+n_{2\nu}=n-\nu \\ m_1+\cdots+m_{2\nu}=m-\mu}} \prod_{j=1}^{2\nu} T_{n_j,m_j} = \frac{\mathcal{C}_{\nu,\mu}}{T_{n,m}} \,[x^{m-\mu} y^{n-\nu}]\, T(x,y)^{2\nu}.$$

By cyclic symmetry,

$$\mathrm{E}[\mathcal{X}_{\nu,\mu}] = P_{\nu,\mu}\,\frac{n}{\nu} = \frac{n\,\mathcal{C}_{\nu,\mu}}{\nu\,T_{n,m}}\,[x^{m-\mu}y^{n-\nu}]\,T(x,y)^{2\nu}. \tag{5.6}$$

Let us see what we can get from (5.6). By (3.1),

$$[x^{m-\mu}y^{n-\nu}]\,T(x,y)^{2\nu} \le \frac{\big[C(y/(1-x))A(x,1-C(y/(1-x)))\big]^{2\nu}}{x^{m-\mu}y^{n-\nu}},$$

for all $x<1$, $y<(1-x)/4$. Letting $y\uparrow(1-x)/4$, using $C(1/4)=2$ and $A(x,-1)=f(x)\le 1$, and setting $x=(m-\mu)/(m-\mu+n-\nu)$, we obtain

$$\begin{aligned}
[x^{m-\mu}y^{n-\nu}]\,T(x,y)^{2\nu} &\le 4^n\,\frac{1}{x^{m-\mu}(1-x)^{n-\nu}}\\
&= 4^n\,\frac{(m-\mu+n-\nu)^{m-\mu+n-\nu}}{(m-\mu)^{m-\mu}(n-\nu)^{n-\nu}}.
\end{aligned}$$

Consequently, the identity (5.6) yields

$$\mathrm{E}[\mathcal{X}_{\nu,\mu}] \le n\,\mathcal{C}_{\nu,\mu}\,\frac{4^n}{T_{n,m}}\,\frac{(m-\mu+n-\nu)^{m-\mu+n-\nu}}{(m-\mu)^{m-\mu}(n-\nu)^{n-\nu}}. \tag{5.7}$$

Now, by Lemma 3.4,

$$T_{n,m} \ge_b \binom{n+m-1}{n-1} C_n \exp\left(-\frac{\pi^2}{2(1-q)}\right), \quad q := \frac{m}{m+n},$$

provided that

$$m \le \frac{2n}{\pi^2}\big(\log n - (3/2)\log\log n - \omega(n)\big), \tag{5.8}$$

where $\omega(n)\to\infty$ however slowly. For such an m, we have

$$\exp\left(\tfrac{\pi^2}{2(1-q)}\right) \le \frac{n}{\log n} \quad \text{and} \quad \binom{n+m-1}{n-1} \ge \frac{\binom{n+m}{n}}{\log n}.$$

Using the two inequalities above and Stirling's formula for the Catalan number C_n, (5.7) becomes

$$\mathrm{E}[\mathcal{X}_{\nu,\mu}] \le_b \frac{n^{7/2}\cdot\mathcal{C}_{\nu,\mu}}{\binom{n+m}{n}}\cdot\frac{(m-\mu+n-\nu)^{m-\mu+n-\nu}}{(m-\mu)^{m-\mu}(n-\nu)^{n-\nu}}.$$

Now, using

$$c\,b^{-1/2}\frac{b^b}{a^a(b-a)^{b-a}} \le \binom{b}{a} \le \frac{b^b}{a^a(b-a)^{b-a}},$$

$c>0$ being an absolute constant, and log-concavity of $f(a,b) := \frac{b^b}{a^a(b-a)^{b-a}}$, we replace the last bound with a cruder version. Namely, if m satisfies the inequality (5.8), then

$$\mathrm{E}[\mathcal{X}_{\nu,\mu}] \le_b n^4\mathcal{C}_{\nu,\mu}\,\frac{n^\nu m^\mu}{(n+m)^{\nu+\mu}} \tag{5.9}$$

uniformly for all $\nu\le n$ and $\nu-1\le\mu\le m$, or using (5.4),

$$\begin{aligned}
\mathrm{E}[\mathcal{X}_{\nu,\mu}] &\le_b n^4\,\frac{4^\nu}{\nu^2}\,\frac{(\mu+\nu)^{\mu+\nu}}{\mu^\mu\nu^\nu}\,\frac{n^\nu m^\mu}{(n+m)^{\nu+\mu}}\\
&= n^4\,(4^\nu/\nu^2)F(\mu/\nu,m/n)^\nu, \qquad (5.10)\\
F(x,y) &:= \frac{(1+x)^{1+x}}{x^x}\cdot\frac{y^x}{(1+y)^{1+x}}.
\end{aligned}$$

Let $y = m/n \ge \alpha$ with $\alpha > 4e^2$, and $x = \mu/\nu \le y/\alpha$. Taylor-expanding $z \log z$ about $z = x$ and using $x \ge (\nu-1)/\nu$,

$$\begin{aligned}\log F(x,y) &= (1+x)\log(1+x) - x\log x + x\log y - (1+x)\log(1+y)\\ &\le (1+\log x) + \frac{1}{2x} - \log y \le 2 + \log x - \log y = \log\frac{xe^2}{y}\\ &\le \log\frac{ye^2}{\alpha y} = \log\frac{e^2}{\alpha} < \log\frac{1}{4}.\end{aligned}$$

So (5.10) becomes:

$$\mathrm{E}[\mathcal{X}_{\nu,\mu}] \le_b n^4\rho^\nu/\nu^2, \quad \rho := \frac{4e^2}{\alpha},$$

uniformly for all $\mu \ge \nu - 1$ with $\mu/\nu \le \alpha^{-1}m/n$. For $\beta = -5/\log\rho$,

$$\begin{aligned}\sum_{\nu\ge\beta\log n,\, \mu/\nu\le\alpha^{-1}m/n} \mathrm{E}[\mathcal{X}_{\nu,\mu}] &\le_b n^4(m/n)\sum_{\nu\ge\beta\log n}\rho^\nu/\nu\\ &\le_b n^4\log n\cdot\rho^{\beta\log n}/\log n = 1/n \to 0.\end{aligned}$$ □

Lemma 5.1 shows that, for the random diagram with density m/n sufficiently large, whp there are no components of size $\Omega(\log n)$ with density smaller by a constant factor than m/n. We anticipate that, for $m/n \to \infty$, whp there exists a large component and that a likely candidate is a component with the maximum density. Let us focus on such components. Given parameters ν and μ, let $A_{\nu,\mu}$ denote the event "there is a maximum density component with ν chords and μ crossings". Needless to say, on the event $A_{\nu,\mu}$, the maximum density is μ/ν.

Lemma 5.2. *Suppose $m/n \to \infty$ and m satisfies (3.16) in Lemma 3.4. Let $c \in (1,2)$ be fixed. Define $\alpha = 7\max\{\log(1/ce^{-c}), \log(1/0.99)\}$. Then,*

$$\lim_{n,m\to\infty}\sum_{\nu,\mu}\mathrm{P}(A_{\nu,\mu}) = 0, \tag{5.11}$$

where the sum is over all pairs (ν,μ) such that

$$\nu \ge \alpha\log n, \qquad \mu \le (2-c)m.$$

In words, it is very unlikely that the densest component has size exceeding $\alpha\log n$ and that its number of crossings scaled by m is strictly below 1.

Proof. Notice upfront that $\mathrm{P}(A_{\nu,\mu}) = 0$ if $\mu/\nu < m/n$. Thus, in (5.11), the terms of interest are those with $\mu/\nu \ge m/n$. As in the proof of Lemma 5.1, a component with parameters ν and μ induces the partition of the remaining set of $2(n-\nu)$ points into 2ν isolated subdiagrams with parameters n_j, m_j, $1 \le j \le 2\nu$. If a chosen component is of maximum density μ/ν, then, in addition, we must have $m_j/n_j \le \mu/\nu$. So, instead of (5.6), we obtain

$$\mathrm{P}(A_{\nu,\mu}) \le \frac{n\,\mathcal{C}_{\nu,\mu}}{\nu\,T_{n,m}}\,[x^{m-\mu}y^{n-\nu}]\,T_{\mu/\nu}(x,y)^{2\nu},$$

where

$$T_{\mu/\nu}(x,y) := 1 + \sum_{0<i/j\le\mu/\nu} T_{i,j}x^iy^j.$$

Here

$$[x^{m-\mu}y^{n-\nu}]\,T_{\mu/\nu}(x,y)^{2\nu} \le \frac{T_{\mu/\nu}(x,y)^{2\nu}}{x^{m-\mu}y^{n-\nu}}, \quad \forall\, x>0,\ y>0.$$

Let

$$x := \frac{m-\mu}{m-\mu+n-\nu}, \quad y := \frac{1}{4}(1-x)$$

and observe that $x \to 1$ from below since $m - \mu \ge m(c-1) \gg n$. Similar to (5.9), we obtain

$$\begin{aligned}
\mathrm{P}(A_{\nu,\mu}) \le\ & \frac{n\,\mathcal{C}_{\nu,\mu}4^{n-\nu}}{\nu\,T_{n,m}}\,\frac{(m-\mu+n-\nu)^{m-\mu+n-\nu}}{(m-\mu)^{m-\mu}(n-\nu)^{n-\nu}}\,T_{\mu/\nu}(x,y)^{2\nu}\\
\le_b\ & n^4 4^{-\nu}\mathcal{C}_{\nu,\mu}\,\frac{m^\mu n^\nu}{(m+n)^{\mu+\nu}}\,T_{\mu/\nu}(x,y)^{2\nu}\\
\le\ & \frac{n^4}{\nu^2}\cdot\frac{(\mu+\nu)^{\mu+\nu}}{\mu^\mu\nu^\nu}\cdot\frac{m^\mu n^\nu}{(m+n)^{\mu+\nu}}\,T_{\mu/\nu}(x,y)^{2\nu},
\end{aligned}\tag{5.12}$$

where we use (5.4) in the last step. Let us bound the last factor in (5.12). Using the upper bound (4.13) in Lemma 4.4, we have

$$\begin{aligned}
T_{\mu/\nu}(x,y) &\le 1 + \sum_{0<i/j\le\mu/\nu}\binom{i+j-1}{j-1}C_j x^i y^j\\
&= \sum_{i,j\ge 0}\binom{i+j-1}{j-1}C_j x^i y^j - \sum_{\substack{j>0\\ i/j>\mu/\nu}}\binom{i+j-1}{j-1}C_j x^i y^j\\
&=: \Sigma_1 - \Sigma_2,
\end{aligned}\tag{5.13}$$

Here

$$\begin{aligned}
\Sigma_1 &= \sum_{j\ge 0}C_j y^j\sum_{i\ge 0}\binom{i+j-1}{j-1}x^i\\
&= \sum_{j\ge 0}C_j y^j(1-x)^{-j} = \sum_{j\ge 0}C_j(1/4)^j = C(1/4) = 2.
\end{aligned}\tag{5.14}$$

Turn to Σ_2. For a given $j > 0$, introduce $i_0 = i_0(j) := \min\{i : i > j\mu/\nu\}$, and write

$$\sum_{i>j\mu/\nu}\binom{i+j-1}{j-1}x^i = x^{i_0}\sum_{i\ge i_0}\binom{i+j-1}{j-1}x^{i-i_0} := x^{i_0}\,\Sigma_2^*.$$

We are going to use Abelian summation by parts to bound Σ_2^* from below. Using

$$\sum_{b=a}^{a+N-1}\binom{b}{a} = \binom{a+N}{a+1},$$

we have: for $N > 0$,

$$\begin{aligned}
S_{N,j} := \sum_{i=i_0}^{i_0+N-1}\binom{i+j-1}{j-1} &= \sum_{i=0}^{i_0+N-1}\binom{i+j-1}{j-1} - \sum_{i=0}^{i_0-1}\binom{i+j-1}{j-1}\\
&= \binom{i_0+j+N-1}{j} - \binom{i_0+j-1}{j},
\end{aligned}$$

and $S_{0,j} = 0$. Using

$$\binom{i+j-1}{j-1} = S_{i-i_0+1,j} - S_{i-i_0,j},$$

we get

$$\begin{aligned}
\Sigma_2^* = \sum_{i\ge i_0}\big[S_{i-i_0+1,j} - S_{i-i_0,j}\big]x^{i-i_0} &= (1-x)\sum_{i\ge i_0}S_{i-i_0+1,j}x^{i-i_0}\\
&= (1-x)\sum_{i\ge i_0}\left[\binom{i+j}{j} - \binom{i_0+j-1}{j}\right]x^{i-i_0}\\
&\ge (1-x)\sum_{i\ge i_0}\binom{i-i_0+j}{j}x^{i-i_0}\\
&= (1-x)\cdot(1-x)^{-j-1} = (1-x)^{-j}.
\end{aligned}\tag{5.15}$$

Explanation for the inequality above: For $j \geq 1$, we have

$$\binom{i_0+j-1}{j} \leq \binom{i_0+j}{j} - 1 \leq \binom{i+j}{j} - \binom{i-i_0+j}{j},$$

where the last inequality follows from

$$\binom{|A|}{x} + \binom{|B|}{x} - \binom{|A \cap B|}{x} \leq \binom{|A \cup B|}{x}$$

for any two sets A and B and any nonegative integer x. (Take $x = j$ and let A and B be two sets such that $|A| = i_0 + j,\ |B| = i - i_0 + j,\ |A \cap B| = j$.) As $x \to 1$, we have $x^{i_0} = x^{j\mu/\nu}(1 + O(1-x))$. Using (3.3) and (5.15), we have

$$\begin{aligned}\Sigma_2 &= \sum_{j \geq 0} C_j y^j x^{i_0(j)} \Sigma_2^* \geq (1 + O(1-x))) \sum_{j \geq 0} C_j y^j \left(\frac{x^{\mu/\nu}}{1-x}\right)^j \\ &= (1 + O(1-x)) \sum_{j \geq 0} C_j (x^{\mu/\nu}/4)^j = (1 + O(1-x)) C(x^{\mu/\nu}/4) \\ &= (1 + O(1-x)) \frac{2}{1 + \sqrt{1 - x^{\mu/\nu}}}.\end{aligned} \tag{5.16}$$

Combining (5.14) and (5.16) we transform (5.13) into

$$\begin{aligned}T_{\mu/\nu}(x,y) &\leq 2 - (1 + O(1-x)) \frac{2}{1 + \sqrt{1 - x^{\mu/\nu}}} \\ &= \frac{2\sqrt{1 - x^{\mu/\nu}}}{1 + \sqrt{1 - x^{\mu/\nu}}} \cdot (1 + O(\sqrt{1-x}\,)).\end{aligned} \tag{5.17}$$

Using (5.17) we replace (5.12) with

$$\mathrm{P}(A_{\nu,\mu}) \leq_b n^4 \big[R_{\nu,\mu} + o(1)\big]^\nu / \nu^2, \tag{5.18}$$

$$R_{\nu,\mu} := \frac{4(1+\mu/\nu)^{1+\mu/\nu}\,(m/n)^{\mu/\nu}}{(\mu/\nu)^{\mu/\nu}\,(1+m/n)^{1+\mu/\nu}} \cdot \left(\frac{\sqrt{1 - x^{\mu/\nu}}}{1 + \sqrt{1 - x^{\mu/\nu}}}\right)^2. \tag{5.19}$$

Define $X = \frac{\mu/\nu}{m/n}$. Here, since $\mu/\nu \geq m/n \to \infty$,

$$\begin{aligned}\frac{4(1+\mu/\nu)^{1+\mu/\nu}}{(\mu/\nu)^{\mu/\nu}} \cdot \frac{(m/n)^{\mu/\nu}}{(1+m/n)^{1+\mu/\nu}} &= 4X\big(1 + 1/(\mu/\nu)\big)^{1+\mu/\nu} \cdot \big(1 - 1/(1+m/n)\big)^{1+\mu/\nu} \\ &= 4X e^{1+o(1) - X(1+o(1))},\end{aligned} \tag{5.20}$$

uniformly over X. We have two cases.

Case $X \geq c$. Since

$$\frac{\sqrt{1 - x^{\mu/\nu}}}{1 + \sqrt{1 - x^{\mu/\nu}}} \leq \frac{1}{2},$$

we have

$$R_{\nu,\mu} \leq X e^{1+o(1) - X(1+o(1))} \leq \rho + o(1), \quad \rho := ce^{1-c} < 1,$$

as $c > 1$. Thus,

$$\mathrm{P}(A_{\nu,\mu}) \leq_b n^4 \big[R_{\nu,\mu} + o(1)\big]^\nu \leq n^4 (\rho + o(1))^\nu,$$

so that

$$\begin{aligned}\sum_{\nu,\mu : X \geq c} \mathrm{P}(A_{\nu,\mu}) &\leq_b \sum_{\nu \geq \alpha \log n} \sum_{\mu \geq \nu - 1} n^4 (\rho + o(1))^\nu \\ &\leq n^6 \sum_\nu (\rho + o(1))^\nu \leq_b n^6 (\rho + o(1))^{\alpha \log n} \to 0,\end{aligned} \tag{5.21}$$

since $\alpha \geq 7/\log(1/\rho)$.

Case $X \leq c$. The function $\phi(z) = \frac{\sqrt{1-z}}{1+\sqrt{1-z}}$ is decreasing on $(0,1)$, so to find an upper bound for $\phi(x^{\mu/\nu})$, we want to bound $x^{\mu/\nu}$ from below. We have

$$\begin{aligned} x^{\mu/\nu} &= \exp\bigl[-(\mu/\nu)(1-x) + O((\mu/\nu)(1-x)^2)\bigr] \\ &= \exp\bigl[-(\mu/\nu)(1-x) + O(n/m)\bigr]. \end{aligned}$$

Further, using

$$\mu/\nu - m/n \leq (c-1)(m/n), \quad m - \mu + n - \nu \geq m - \mu \geq (1-c)m,$$

we compute

$$\begin{aligned} -\frac{\mu}{\nu}(1-x) &= -\frac{\mu}{\nu} \cdot \frac{n-\nu}{m-\mu+n-\nu} \\ &= -X - \frac{\mu}{\nu}\left(\frac{n-\nu}{m-\mu+n-\nu} - \frac{n}{m}\right) \\ &= -X - \frac{\mu}{\nu} \cdot \frac{n\nu(\mu/\nu - m/n) - n(n-\nu)}{(m-\mu+n-\nu)m} \\ &\geq -X - \frac{\mu}{\nu} \cdot \frac{n\nu(\mu/\nu - m/n)}{(m-\mu+n-\nu)m} \\ &\geq -X - \frac{\mu}{\nu} \cdot \frac{n\nu(c-1)(m/n)}{(c-1)m^2} \\ &= -X - \mu/m \geq -c - (2-c) = -2. \end{aligned}$$

Consequently, $x^{\mu/\nu} \geq e^{-3}$, and $\phi(x^{\mu/\nu}) \leq \phi(e^{-3}) \leq 0.494$. Since Xe^{1-X} is decreasing on $(1,\infty)$ and takes the value 1 for $X = 1$, we have

$$Xe^{1+o(1)-X(1+o(1))} \leq 1 + o(1)$$

for $1 \leq X \leq c$. Therefore,

$$R_{\nu,\mu} \leq 4 \times (0.495)^2 \leq 0.981$$

and

$$\mathrm{P}(A_{\nu,\mu}) \leq_b n^4\bigl[R_{\nu,\mu} + o(1)\bigr]^\nu \leq n^4(0.99)^\nu$$

As in the previous case,

$$\begin{aligned} \sum_{\nu,\mu: X \leq c} \mathrm{P}(A_{\nu,\mu}) &\leq_b \sum_\nu \sum_\mu n^4(0.99)^\nu \\ &\leq n^6 \sum_\nu (0.99)^\nu \leq_b n^6(0.99)^{\alpha \log n} \to 0, \end{aligned} \tag{5.22}$$

since $\alpha > 7\log(1/0.99)$. The equations (5.21) and (5.22) imply that

$$\lim_{n\to\infty} \sum_{\nu,\mu: X \leq c} \mathrm{P}(A_{\nu,\mu}) = 0.$$

□

Letting $c \downarrow 1$, we arrive at

Corollary 5.3. *Suppose $m/n \to \infty$ and m satisfies (3.16) in Lemma 3.4. Then whp*
- *either the densest component is of size $O(\log n)$,*
- *or its number of crossings is almost m, whence its size is at least $(1+o(1))\sqrt{2m}$.*

Remark 5.4. This is a good place to notice that the sole reason for $\log n$ to appear in the first alternative was that we confined ourselves to $m = O(n \log n)$ meeting the constraint (5.8), in which case $T_{n,m}$ is bounded from below by $C_n\binom{n+m-1}{n-1}\exp(-\gamma \log n)$. For the constraint $n \log n \ll m \ll n^{3/2}$ we still have the lower bound (4.14) from Lemma 4.4,

$$T_{n,m} \ge_b \exp\bigl(-\Theta((m/n)\log n)\bigr) C_n \binom{n+m-1}{n-1}.$$

To off-set this exponential factor, we could have confined ourselves to ν of order $(m/n)\log n$, at least, arriving at the counterpart of Corollary 5.3 with the first alternative becoming "either the densest component is of size $O((m/n)\log n)$", but with the second alternative remaining unchanged. In other words, the gap property for the crossing density of the densest component continues to hold for $n \log n \ll m \ll n^{3/2}$.

Now if $m = \Theta(n \log n)$, and the densest component has size ν then for the number of crossings we have

$$\frac{\nu(\nu-1)}{2} \ge \mu \ge \nu \frac{m}{n},$$

implying $\nu \ge 2m/n = \Theta(\log n)$. So, if $\nu = O(\log n)$, that is, if the first alternative in Corollary 5.3 holds, then $\nu = \Theta(\log n)$ and $\mu = \Theta((\log n)^2)$, and the maximum density μ/ν is of order m/n exactly. This is the reason why in the rest of the paper we continue to stick with $m = \Theta(n \log n)$. Our goal is to eliminate, eventually, the case $\nu = O(\log n)$.

Lemma 5.5. *Given fixed $c \ge 1$, $b > 1$, let $B_{n,m} = B_{n,m}(c, b)$ denote the event: the maximum density is below $c\, m/n$ and there is a (ν, μ)-component meeting the constraints*

$$\nu \ge b \log n, \quad \mu \le (1 - b^{-1/3})m. \tag{5.23}$$

For every $c \ge 1$, there exists $b = b(c) > 1$ such that $\mathrm{P}(B_{n,m}) \to 0$. Thus whp every component either has size below $b \log n$, or has at least $(1 - b^{-1/3})m$ edges.

Proof. First of all, in view of Lemma 5.1, by choosing b sufficiently large we can consider only (ν, μ)-components with $\mu/\nu \ge dm/n$, with some fixed $d > 0$. Also, for μ satisfying (5.23),

$$\frac{m-\mu}{n-\nu} \ge \frac{mb^{-1/3}}{n} = \Theta(b^{-1/3}\log n) \to \infty.$$

Arguing as in the proof of Lemma 5.2, we obtain

$$\mathrm{P}(B_{n,m}) \le_b n^4 \sum_{\nu,\mu} \bigl[R_{\nu,\mu}(1 + O(b^{-1}))\bigr]^{\nu}, \tag{5.24}$$

where the sum is over all (ν, μ) satisfying (5.23), but instead of (5.19) we get

$$\begin{aligned} R_{\nu,\mu} &:= 4\frac{(1+\mu/\nu)^{1+\mu/\nu}}{(\mu/\nu)^{\mu/\nu}} \cdot \frac{(m/n)^{\mu/\nu}}{(1+m/n)^{1+\mu/\nu}} \cdot \left(\frac{\sqrt{1-x^{cm/n}}}{1+\sqrt{1-x^{cm/n}}}\right)^2 \\ &\le \frac{(1+\mu/\nu)^{1+\mu/\nu}}{(\mu/\nu)^{\mu/\nu}} \cdot \frac{(m/n)^{\mu/\nu}}{(1+m/n)^{1+\mu/\nu}}. \end{aligned}$$

Here as before

$$x = (m-\mu)/(m-\mu+n-\nu) = 1 - O(b^{-1}). \tag{5.25}$$

(The remainder $O(b^{-1})$ in (5.25) is the reason for the same remainder in (5.24).) Again, set $X = \frac{\mu/\nu}{m/n}$. Since $m/n \to \infty$ and $\mu/\nu \to \infty$,

$$R_{\nu,\mu}(1 + O(b^{-1})) \le X e^{1-X+O(b^{-1})},$$

The log-concave function $H(X) := Xe^{1-X}$ attains its absolute maximum 1 at $X = 1$. Let $A > 0$ be a constant and first consider the contribution of X's with $|X - 1| \ge Ab^{-1/2}$. We have

$$\max\{H(X) : |X-1| \ge Ab^{-1/2}\} \le \max\big\{H(1-Ab^{-1/2}), H(1+Ab^{-1/2})\big\} \le \exp\big[-A^2/(3b)\big].$$

Thus, for this range of X,

$$R_{\nu,\mu}(1+O(b^{-1})) \le \exp\big[-A^2/(4b)\big]$$

if we choose A sufficiently large. So

$$n^4 \cdot \sum_{\substack{\nu \ge b\log n\\ |X-1|\ge Ab^{-1/2}}} \big[R_{\nu,\mu}(1+O(b^{-1}))\big]^{\nu} \le n^4 \cdot \sum_{\nu \ge b\log n} \nu^2 \exp\big[-\nu A^2/(4b)\big] \to 0,$$

if $b(A^2/(4b)) > 5$, that is, if $A^2 > 20$. The factor ν^2 in the second sum is due to the fact that there are at most $\binom{\nu}{2}$ values of μ.

Now consider the contribution of (ν, μ) where $|X - 1| \le Ab^{-1/2}$. We have

$$x^{cm/n} = \exp\big[-(1-x)cm/n + O((1-x)^2 m/n)\big]$$

and

$$\begin{aligned}(1-x)\frac{m}{n} &= \frac{m}{m+n}\left[1 + \frac{\nu(\mu/\nu - m/n)}{m-\mu+n-\nu}\right]\\ &= \frac{m}{m+n}\big[1 + O(\mu b^{-1/2}/(m-\mu))\big]\\ &= \frac{m}{m+n}\big[1 + O(b^{-1/6})\big] \,=\, 1+o(1).\end{aligned}$$

Therefore, introducing $\rho = 2\frac{\sqrt{1-e^{-c}}}{1+\sqrt{1-e^{-c}}} < 1$, we obtain

$$R_{\nu,\mu}(1+O(b^{-1})) \le \rho\big(1+O(b^{-1/6})\big)Xe^{1-X} \le \rho^{1/2},$$

if b is large enough. We conclude that

$$n^4 \sum_{\substack{\nu \ge b\log n\\ |X-1|\le Ab^{-1/2}}} \big[R_{\nu,\mu}(1+O(b^{-1}))\big]^{\nu} \le n^4 \sum_{\nu \ge b\log n} \nu^2 (\rho^{1/2})^{\nu} \to 0,$$

if b is sufficiently large. □

Lemma 5.6. *Suppose that* $\lim_{n\to\infty} m/(n\log n) \in (0, 2/\pi^2)$. *Whp,*

- *either there exists a (necessarily unique) component that contains almost all m crossings, whence has at least $(1+o(1))\sqrt{2m}$ vertices,*
- *or there is no component of size ν with $\nu/\log n$ exceeding a large constant.*

Proof. It follows directly from Lemma 5.2 and Lemma 5.5. □

If we rule out the second alternative, we will be able to claim that whp there is a component containing almost all m crossings. To do so, we need some enumerative groundwork.

Given $k, \ell > 1$ and $s \le k$, let $T_{n,m}(k, \ell, s)$ denote the total number of diagrams with k components, each of size not exceeding ℓ, and with exactly s components of size 1, i.e. isolated chords. Obviously,

$$T_{n,m}(k,\ell,s) = 0 \quad \text{if} \quad \ell(k-s) < n-s. \tag{5.26}$$

Lemma 5.7. *Introducing*

$$I_j(x) := \sum_{\mu\ge 0} I_{j,\mu}x^\mu = (1+x)\cdots(1+x+\cdots+x^{j-1}),$$

we have:

$$\max_{\ell}\ T_{n,m}(k,\ell,s) \le \frac{(2n)_{k-1}}{(k-s)!s!}\,[x^m y^{n-s}]\left(\sum_{j=2}^{\infty} C_j I_j(x) y^j\right)^{k-s}. \tag{5.27}$$

Proof. For a generic diagram with parameters n and m, with k components, of size not exceeding ℓ, and s components of size 1, let s_j denote the total number of components of size j; so $\boldsymbol{s} = (s_1, s_2, \dots, s_n)$ meets the conditions:

$$s_1 = s; \quad (\forall j > \ell)\ s_j = 0; \quad \sum_{j=2}^{n} s_j = k - s; \quad \sum_{j=2}^{n} j s_j = n - s. \tag{5.28}$$

For such a diagram to exist, it is necessary that the point sets of the components form a non-crossing partition of $[2n]$. By Kreweras' formula [32], the total number of such partitions is $(2n)_{k-1}/[s_1!s_2!\cdots]$. In addition, for each $2 \le j \le n$, and $1 \le t \le s_j$, let $m_{j,t}$ denote the number of crossings of the t-th component from the arbitrarly ordered list of all components of size j. Clearly, $\boldsymbol{m} = \{m_{j,t}\}$ meets the condition

$$\sum_{j=2}^{n}\sum_{t=1}^{s_j} m_{j,t} = m. \tag{5.29}$$

Then,

$$\begin{aligned}
T_{n,m}(k,\ell,s) &\le \sum_{\boldsymbol{s}\text{ meets }(5.28)} \frac{(2n)_{k-1}}{s_1!\cdots s_n!} \sum_{\boldsymbol{m}\text{ meets }(5.29)} \prod_{\substack{2\le j\le n\\ 1\le t\le s_j}} T_{j,m_{j,t}}\\
&\le \frac{1}{s!} \sum_{\boldsymbol{s}\text{ meets }(5.28)} \frac{(2n)_{k-1}}{s_2!\cdots s_n!} \sum_{\boldsymbol{m}\text{ meets }(5.29)} \prod_{\substack{2\le j\le n\\ 1\le t\le s_j}} C_j I_{j,m_{j,t}}\\
&= \frac{(2n)_{k-1}}{s!} \sum_{\boldsymbol{s}\text{ meets }(5.28)} \frac{1}{s_2!\cdots s_n!}\,[x^m]\prod_{j=2}^{n}\left(C_j\sum_{\mu\ge 0} I_{j,m}x^\mu\right)^{s_j}\\
&= \frac{(2n)_{k-1}}{s!}\,[x^m] \sum_{\boldsymbol{s}\text{ meets }(5.28)} \prod_{j=2}^{n} \frac{(C_j I_j(x))^{s_j}}{s_j!}.
\end{aligned} \tag{5.30}$$

Here the last sum is at most

$$\begin{aligned}
\sum_{\substack{\boldsymbol{s}\ge \boldsymbol{0}\\ \sum_{j\ge 2} j s_j<\infty}} \prod_{j=2}^{\infty} \frac{(y^j z C_j I_j(x))^{s_j}}{s_j!} &= [y^{n-s}z^{k-s}]\exp\left(z\sum_{j\ge 2} y^j C_j I_j(x)\right)\\
&= [y^{n-s}]\,\frac{1}{(k-s)!}\left(\sum_{j=2}^{\infty} y^j C_j I_j(x)\right)^{k-s}.
\end{aligned} \tag{5.31}$$

Equations (5.30) and (5.31) imply (5.27), which finishes the proof. $\square$

Lemma 5.27 enables us to obtain an explicit bound for $T_{n,m}(k,\ell,s)$. First of all, using

$$(1+x)\times\cdots\times(1+x+\cdots+x^{j-1}) = (1-x)^{-j}(1-x)\cdots(1-x^j)$$

we get: for $k-s \ge (n-s)/\ell$,

$$T_{n,m}(k,\ell,s) \le \frac{(2n)_{k-1}}{(k-s)!s!}\,[x^m y^{n-s}]\left(\sum_{j=2}^{\infty} C_j\left(\frac{y}{1-x}\right)^j \prod_{t=1}^{j}(1-x^t)\right)^{k-s}.$$

The bivariate series on the RHS has positive coefficients, and converges for $x \in (0,1)$, $y \in (0,(1-x)/4]$. So, by Chernoff-type bound with $x \in (0,1)$ and $y=(1-x)/4$, we obtain

$$\begin{aligned}
T_{n,m}(k,\ell,s) &\le \frac{(2n)_{k-1}}{(k-s)!s!}\,x^{-m}y^{-(n-s)}\left((1-x)(1-x^2)\sum_{j=2}^{\infty}\frac{C_j}{4^j}\prod_{t=3}^{j}(1-x^t)\right)^{k-s}\\
&\le \frac{4^{n-s}(2n)_{k-1}}{(k-s)!s!}\,x^{-m}(1-x)^{-(n+s-2k)}\big[1/4+O(1-x)\big]^{k-s}\\
&= \frac{4^{n-k}(2n)_{k-1}}{(k-s)!s!}\,x^{-m}(1-x)^{-(n+s-2k)}\big[1+O(1-x)\big]^{k-s}.
\end{aligned}$$

Choosing $x=m/(m+n)$ we get

$$T_{n,m}(k,\ell,s) \le \binom{k}{s}\frac{4^{n-k}(2n)_{k-1}}{k!}\cdot\frac{(m+n)^{m+n+s-2k}}{m^m n^{n+s-2k}}\big[1+O(n/(m+n))\big]^{k-s}. \tag{5.32}$$

Lemma 5.8. *Suppose that* $\lim m/(n\log n) \in (0,2/\pi^2)$. *Then whp there exists a component that has almost all* m *crossings.*

Proof. By Lemma 5.6, it suffices to prove that, for every $A>0$, whp there is a component of size exceeding $\ell := A\log n$. Let $\mathcal{X}$ denote the total number of isolated chords in the random diagram; so $\mathcal{X}=\mathcal{X}_{1,0}$, where $\mathcal{X}_{\nu,\mu}$ is as defined in the proof of Lemma 5.1). Clearly,

$$\mathrm{E}[\mathcal{X}] \le \frac{2n}{T_{n,m}}\sum_{\substack{n_1+n_2=n-1\\ m_1+m_2=m}} T_{n_1,m_1}T_{n_2,m_2} = \frac{2n}{T_{n,m}}\,[x^m y^{n-1}]\,T(x,y)^2. \tag{5.33}$$

Using (3.17) with $\ell=2$,

$$\mathrm{E}[\mathcal{X}] \le_b n\frac{\binom{n+m-2}{n-2}}{\binom{n+m-1}{n-1}} = \frac{n(n-1)}{n+m-1} = O(n(\log n)^{-1}).$$

Hence whp $\mathcal{X} \le n/(\log n)^{1-\varepsilon}$ for any fixed $\varepsilon \in (0,1)$. Thus it suffices to show

$$\sum_{k,s}\frac{T_{n,m}(k,\ell,s)}{T_{n,m}} \to 0, \tag{5.34}$$

where the sum is over all k,s such that

$$s \le s(n) := \frac{n}{(\log n)^{1-\varepsilon}}, \qquad k-s \ge \frac{n-s}{\ell}. \tag{5.35}$$

(Indeed, $T_{n,m}(k,\ell,s)/T_{n,m}$ is the probability that the diagram has k components, with exactly s components of size 1, and all other components of size *not exceeding* ℓ.) Combining the asymptotic formula (3.17) for $T_{n,m}$ in Lemma 3.4, the bound (5.32) for $T_{n,m}(k,\ell,s)$, and the constraints (5.35) we obtain:

$$\begin{aligned}
\frac{T_{n,m}(k,\ell,s)}{T_{n,m}} &\ll \binom{2n}{k}\left(\frac{n}{m+n}\right)^{k+n/(2\ell)} \le \frac{(2n)^k}{k!}\left(\frac{n}{m}\right)^k \times \left(\frac{n}{m}\right)^{n/(2\ell)}\\
&\le \exp(2n^2/m)\cdot\exp\left(-\frac{n}{2A\log n}\log(m/n)\right)\\
&\le \exp\left(O(n/\log n) - \gamma\,\frac{n\log\log n}{\log n}\right),
\end{aligned}$$

$\gamma > 0$ being fixed. For the third line in the above inequality, we used $y^k/k! \le e^y$. The last quantity approaches 0 super-polynomially fast, so does the expression in (5.34). $\square$

Finally,

Theorem 5.9. *Suppose that* $\lim m/(n\log n) \in (0, 2/\pi^2)$. *Then whp there exists a component that has almost all* m *crossings and a positive fraction of* n *chords.*

Proof. Given ε, $\delta \in (0,1)$, let $N_{\delta,\varepsilon}$ denote the total number of the (ν,μ)-components with $\nu \le \delta n$ and $\mu \ge (1-\varepsilon)m$. In light of Lemma 5.8, it suffices to show that, for $\varepsilon < 1/2$, and δ sufficiently small, $\mathrm{E}[N_{\delta,\varepsilon}] \to 0$. Let $\delta < 1$ be such that

$$\delta < \frac{(1/2 - e^{-1})(1-\varepsilon)}{\log(4e)}. \tag{5.36}$$

By (5.10)

$$\mathrm{E}[N_{\delta,\varepsilon}] \le_b \sum_{\substack{\sqrt{2\mu} \le \nu \le \delta n \\ \mu \ge (1-\varepsilon)m}} (n^4/\nu^2) \exp\bigl[\nu H(x_{\nu,\mu})\bigr], \quad x_{\nu,\mu} := \frac{\mu}{\nu};$$

$$H(x) := \log 4 + (1+x)\log\frac{1+x}{1+m/n} + x\log\frac{m/n}{x}.$$

Observe that, for ν, μ in question,

$$x_{\nu,\mu} \ge \frac{m}{n}\,\frac{1-\varepsilon}{\delta} > y := \frac{m}{n}$$

since $1-\varepsilon > \delta$. Further,

$$\begin{aligned}
H(x) &= \log 4 + (1+x)\log\frac{x}{y} + (1+x)\log\frac{1+1/x}{1+1/y} + x\log\frac{y}{x}\\
&\le \log 4 + \log\frac{x}{y} + (1+x)\left(\frac{1+1/x}{1+1/y} - 1\right)\\
&\le \log 4 + 1 + \log\frac{x}{y} - \frac{x/y}{1+1/y}\\
&\le \log(4e) + \log\frac{x}{y} - \frac{1}{2}\,\frac{x}{y}\\
&\le \log(4e) - (1/2 - e^{-1})\,\frac{x}{y},
\end{aligned}$$

the last inequality following from $\log z \le e^{-1}z$. Therefore

$$\begin{aligned}
&H(x_{\nu,\mu}) \le -\gamma(\varepsilon,\delta),\\
&\gamma(\varepsilon,\delta) := (1/2 - e^{-1})\frac{1-\varepsilon}{\delta} - \log(4e) > 0,
\end{aligned}$$

see (5.36). Therefore, as $n \to \infty$,

$$\mathrm{E}[N_{\delta,\varepsilon}] \le_b n^4 \sum_{\nu \ge \sqrt{m}} \exp\bigl[-\nu\gamma(\varepsilon,\delta)\bigr] \to 0.$$ $\square$

To complete the picture, turn now to $m = \Theta(n)$.

Theorem 5.10. *If* $m \le n/14$, *then there exists a constant* $A > 0$ *such that whp the size of the largest component is at most* $A\log n$.

Proof. Let $A > 0$ to be specified shortly. For E_n, the expected number of components of size exceeding $A\log n$, (by (5.10) again), we have

$$E_n \le_b n^4 \sum_{\nu \ge A\log n} \sum_{\mu \ge \nu - 1} \nu^{-2} \cdot \exp\bigl[\nu H(x_{\nu,\mu})\bigr];$$

here $x_{\nu,\mu} = \mu/\nu \geq 1 - 1/(A \log n)$. Since $H(x)$ is concave,

$$\begin{aligned} H(x_{\nu,\mu}) &\leq H(1) + H'(1)(x_{\nu,\mu} - 1) \\ &\leq \log \frac{16m/n}{(1+m/n)^2} + O((\log n)^{-1}). \end{aligned}$$

Now $16z/(1+z)^2 < 1$ for $0 < z < z^* := 7 - \sqrt{48} > 1/14$. So if $m/n \leq 1/14$, then

$$E_n \leq_b n^4 m \sum_{\nu \geq A \log n} \exp\left[\nu \left(\log \frac{224}{225} + O((\log n)^{-1})\right)\right] \to 0,$$

if $A > 5/(\log 225/224)$. □

6. Concluding Remarks

Although chord diagrams have been studied widely, there are still many open problems about them, particularly of enumerative-probabilistic nature. The results presented in this paper provide partial solutions to some, in our opinion interesting, problems. We conclude this paper with some questions for possible extensions of our results.

An asymptotic expression for the number $T_{n,m}$ of chord diagrams with a given number of crossings has been found in Theorem 3.4 for the case $m < (2/\pi^2) n \log n$, but its extension for larger m is still to be found. It would be quite useful just to strengthen, and extend the bounds for $T_{n,m}$ given in Lemma 4.4.

Our main goal in this paper was to observe a kind of phase transition for the largest component of a random chord diagram. Theorem 5.9 tells us that when $m/n \log n$ has a limit in $(0, 2/\pi^2)$, there is a giant component containing almost all the crossings (edges in the intersection graph) and a positive fraction of chords. In Erdős-Rényi graphs, coupling $G(n,m)$ with $G(n,m+1)$ with a graph process yields immediately that having a giant component is a monotone property. While finding a similar coupling for chord diagrams is highly problematic, the existence of a giant component (whp) for $m = \Omega(n \log n)$ would follow from a much simpler claim, namely that the probability of the giant component is monotone increasing with m. Judging by our experience with the random permutation graph [3], it might be very helpful to prove that, for each n, the sequence $\{T_{n,m}\}$ is log-concave, just like $\{I_{n,m}\}$. Furthermore, it is still unclear whether whp there is a giant component for $n \ll m = o(n \log n)$.

Lastly, there are two other classes of graphs nontrivially related to chord diagrams: circle graphs and interlace graphs. A (labeled) circle graph is obtained by labeling a set of chords of a circle, where the edges are determined by the crossing relation. An *interlace graph* with vertex set $[n]$ is obtained from a permutation of the multiset $\{1, 1, 2, 2, \dots, n, n\}$, where two vertices i and j are adjacent if the corresponding symbols are interlaced in the permutation, i.e. if the permutation looks like $\dots i \dots j \dots i \dots j \dots$ or $\dots j \dots i \dots j \dots i \dots$ As Arratia et al. [7] pointed out, each circle graph is an interlace graph, and the number of interlace graphs is bounded above by the number of permutations of the multiset, which is $(2n)!/2^n$. A chord diagram corresponds to a *standard permutation*, in which the first occurence of i is always before the first occurrence of j for all pairs $i < j$. However, the number of interlace graphs of standard permutations is not the same as the number of intersection graphs due to the fact that the same interlace graph might come from many different standard permutations, whereas the intersection graphs that we consider uniquely determine the chord diagrams. For example, there are $\binom{2n}{n}/(n+1)$ standard permutations producing the empty graph on $[n]$. We are curious if the results in this paper hold for these two important classes of graphs, or at least shed some light on the respective thresholds for the appearance of a giant component.

Acknowledgement We thank Sergei Chmutov for encouraging us to study random chord diagrams and for helpful discussions. The three hard-working referees provided us with an invaluable, critical feedback, which allowed us to improve, substantially, the presentation of our study. We also thank the editors for organizing an expert refereeing of the manuscript.

References

[1] H. Acan, *An enumerative-probabilistic study of chord diagrams*, Ph.D. Thesis. The Ohio State University (2013).

[2] H. Acan, *On a uniformly random chord diagram and its intersection graph*, Discrete Math. (2016), http://dx.doi.org/10.1016/j.disc.2016.11.004.

[3] H. Acan and B. Pittel, *On the connected components of a random permutation graph with a given number of edges*, J. Combin. Theory Ser. A **120** (2013), 1947–1975.

[4] M. Aigner, *A Course in Enumeration*, Graduate Texts in Mathematics 238, Springer, 2010.

[5] G. E. Andrews, R. Askey and R. Roy, *Special Functions*, Encyclopedia of Mathematics and Its Applications 71, Cambridge University Press, 1999.

[6] D. Archdeacon and D. A. Grable, *The genus of a random graph*, Discrete Math. **142** (1995), 21–37.

[7] R. Arratia, B. Bollobás, D. Coppersmith and G. B. Sorkin, *Euler circuits and DNA sequencing by hybridization*, Discr. Appl. Math. **104** (2000), 63–96.

[8] J. Baik, E. Rains, *The asymptotics of monotone subsequences of involutions*, Duke Math. J. **109** (2001), no. 2, 205-282.

[9] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), 173–187.

[10] E. A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, J. Combin. Theory Ser. A **15** (1973), 91–111.

[11] B. Bollobás, *Random graphs*, 2nd ed., Cambridge University Press, 2001.

[12] B. Bollobás and O. Riordan, *Linearized chord diagrams and an upper bound for Vassiliev invariants*, J. Knot Theory Ramifications **9** (2000), 847–853.

[13] B. Bollobás and O. Riordan, *The diameter of a scale free random graph*, Combinatorica **24** (2004), no.1, 5–34.

[14] M. Bóna, *A combinatorial proof of the log-concavity of a famous sequence counting permutations*, Electronic J. Comb., **11** (2005), no.2, N2.

[15] A. Bouchet, *Circle graph obstructions*, J. Combin. Theory Ser. B, **60** (1994) 107–144.

[16] E. R. Canfield, *Applications of the Berry–Esseen inequality to combinatorial estimates*, J. Comb. Theory (A) **28** (1980), 17–25.

[17] W. Y. C. Chen, E. Y. P. Deng, R. R. X. Du, R. P. Stanley, C. H. Yan, *Crossings and nestings of matchings and partitions*, Trans. Amer. Math. Soc. **359** (2007), no. 4, 1555–1575.

[18] S. Chmutov, S. Duzhin, and J. Mostovoy, *Introduction to Vassiliev knot invariants*, Cambridge University Press, 2012.

[19] S. Chmutov and S. Duzhin, *An upper bound for the number of Vassiliev knot invariants*, J. Knot Theory Ramifications **3** (1994), 141–151.

[20] S. Chmutov and B. Pittel, *The genus of a random chord diagram is asymptotically normal*, J. Combin Theory Ser. A **120** (2013), no. 1, 102–110.

[21] S. Chmutov and B. Pittel, *On a surface formed by randomly gluing together polygonal discs*, Adv. in Appl. Math. **73** (2016), 23–42.

[22] R. Cori and M. Marcus, *Counting non-isomorphic chord diagrams*, Theoret. Comput. Sci. **204** (1998), no.1, 55–73.

[23] S. Dulucq, J.-P. Penaud, *Cordes, arbres et permutations*, Discrete Math. **117** (1993) 89–105.

[24] P. Erdős and A. Rényi, *On random graphs I*, Publ. Math. Debrecen **6** (1959), 290–297.

[25] P. Erdős and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5** (1960), 17–61.

[26] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, New York (1971).

[27] P. Flajolet and M. Noy, *Analytic combinatorics of chord diagrams*, Formal Power Series and Algebraic Combinatorics, 12th International Conference, FPSAC'00, Moscow (2000) 191–201, Springer, Berlin.

[28] K. Fleming and and N. Pippenger, *Large deviations and moments for the Euler characteristic of a random surface*, Random Structures and Algorithms **37** (2010), 465–476.

[29] A. Gamburd, *Poisson-Dirichlet distribution for random Belyi surfaces*, Ann. Probab. **34** (2006), 1827–1848.

[30] J. Harer and D. Zagier, *The Euler characteristic of the moduli spaces of curves*, Invent. Math. **85** (1986), 457–486.

[31] M. Josuat-Vergés and J.S. Kim, *Touchard-Riordan formulas, T-fractions, and Jacobi's triple product identity*, Ramanujan J. **30** (2013), 341–378.

[32] G. Kreweras, *Sur les partitions non croisées d'un cycle*, Discrete Math. **1** (1972), no. 4, 333–350.

[33] N. Linial and T. Nowik, *The expected genus of a random chord diagram*, Discrete Comput. Geom. **45** (2011), no. 1, 161–180.

[34] G. Louchard and H. Prodinger, *The number of inversions in permutations: a saddle point approach*, J. Integer Sequences **6**, Article 03.2.8 (2003).

[35] K. V. Menon, *On the convolution of logarithmically concave sequences*, Proc. Amer. Math. Society, **23** (1969), no. 2, 439–441.

[36] N. Pippenger and K. Schleich, *Topological characteristics of random triangulated surfaces*, Random Structures and Algorithms **28** (2006), 247–288.

[37] B. Pittel, *On a likely shape of the random Ferrers diagram*, Adv. in Appl. Math. **18** (1997), no.4, 432-488.

[38] B. Pittel, *On dimensions of a random solid diagram*, Comb. Probab. Comp. **14** (2005), 873-895.

[39] B. Pittel, *Another proof of Harer-Zagier formula*, Electronic J. Comb., **23(1)**, (2016), #P1.21.

[40] A. G. Postnikov, *Introduction to Analytic Number Theory*, Vol. 68 of *Translations of Mathematical Monographs*, AMS, Providence, RI, 1988.

[41] J. Riordan, *The distribution of crossings of chords joining pairs of* $2n$ *points on a circle*, Math. Comp. **29** (1975), 215–222.

[42] V. Rödl and R. Thomas, *On the genus of a random graph*, Random Struct. Algorithms, **1** (1995), 1–12.

[43] P. Rosenstiehl, *Solution algébrique du problème de gauss sur la permutation des points d'intersection d'une ou plusieurs courbes fermées du plan*, C.R. Acad. Sci. **283** (1976), (A): 551–553.

[44] R. P. Stanley, *Enumerative Combinatorics, II*, Cambridge University Press, 1999.

[45] P. R. Stein, J. A. Everett, *On a class of linked diagrams. II. Asymptotics* Discrete Math., **21** (1978), 309–318.

[46] A. Stoimenow, *Enumeration of chord diagrams and an upper bound for Vassiliev invariants*, J. Knot Theory Ramifications **7** (1998), 93–114 .

[47] J. Touchard, *Sur un problème de configurations et sur les fractions continues*, Canad. J. Math. **4** (1952), 2–25.

[48] D. Zagier, *Vassiliev invariants and a strange identity related to the Dedekind eta-function*, Topology **40** (2001), 945–960.