



Citation for published version:

Blanchet, J & Stauffer, A 2013, 'Characterizing optimal sampling of binary contingency tables via the configuration model', *Random Structures and Algorithms*, vol. 42, no. 2, pp. 159-184.
<https://doi.org/10.1002/rsa.20403>

DOI:

[10.1002/rsa.20403](https://doi.org/10.1002/rsa.20403)

Publication date:

2013

Document Version

Early version, also known as pre-print

[Link to publication](#)

This is the pre-peer reviewed version of the following article: Blanchet, J & Stauffer, A 2013, 'Characterizing optimal sampling of binary contingency tables via the configuration model' *Random Structures and Algorithms*, vol 42, no. 2, pp. 159-184, which has been published in final form at <http://dx.doi.org/10.1002/rsa.20403>

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Characterizing Optimal Sampling of Binary Contingency Tables via the Configuration Model

Jose Blanchet*

Alexandre Stauffer†

Abstract

A binary contingency table is an $m \times n$ array of binary entries with row sums $\mathbf{r} = (r_1, \dots, r_m)$ and column sums $\mathbf{c} = (c_1, \dots, c_n)$. The configuration model generates a contingency table by considering r_i tokens of type 1 for each row i and c_j tokens of type 2 for each column j , and then taking a uniformly random pairing between type-1 and type-2 tokens. We give a necessary and sufficient condition so that the probability that the configuration model outputs a binary contingency table remains bounded away from 0 as $N = \sum_{i=1}^m r_i = \sum_{j=1}^n c_j$ goes to ∞ . Our finding shows surprising differences from recent results for binary *symmetric* contingency tables.

Keywords and phrases. Contingency tables, configuration model, uniform sampling

1 Introduction

Given two natural numbers m and n , let $\mathbf{r} = (r_1, r_2, \dots, r_m)$ and $\mathbf{c} = (c_1, c_2, \dots, c_n)$ be vectors of positive integers such that $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j = N$. Let $\Omega_{\mathbf{r}, \mathbf{c}}$ be the set of matrices with binary entries such that the sum of the i -th row is given by r_i and the sum of the j -th column is given by c_j . These matrices are known as binary contingency tables. We consider the problem of sampling uniformly from $\Omega_{\mathbf{r}, \mathbf{c}}$ and of computing $|\Omega_{\mathbf{r}, \mathbf{c}}|$. This problem can also be viewed as uniformly sampling and counting bipartite graphs with $m+n$ nodes such that the node degrees in one partition are given by \mathbf{r} and the node degrees in the other partition are given by \mathbf{c} .

We study a well-known and simple algorithm for sampling contingency tables, which is usually referred to as the configuration model and was introduced by Bollobás [8]. The configuration model proceeds as follows. For each row i , consider r_i tokens of type 1, and for each column j , consider c_j tokens of type 2. Then, construct a table T by sampling uniformly a random matching between type-1 and type-2 tokens. In other words, first label both the type-1 tokens from 1 to N and the type-2 tokens from 1 to N . Next, keep the type-1 tokens fixed and draw a uniformly random permutation of the type-2 tokens. Finally, establish a matching between type-1 and type-2 tokens according to the position in the permutation. In this way, the entry $T_{i,j}$ is taken to be the number of type-1 tokens from row i that were matched to type-2 tokens from column j .

The configuration model produces a table in $\Theta(N)$ time, but may output non-binary tables. Yet, given that the table generated is binary, then the output table is a uniform sample from $\Omega_{\mathbf{r}, \mathbf{c}}$.

*Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, U.S.A. Email: blanchet@columbia.edu. Supported in part by NSF grants 0902075 and 0846816.

†Computer Science Division, University of California, Berkeley, CA 94720-1776, U.S.A. Email: stauffer@cs.berkeley.edu. Supported by a Fulbright/CAPES scholarship and NSF grants CCF-0635153 and DMS-0528488.

Therefore, if the probability that the configuration model outputs a *binary* table does not go to zero as $N \rightarrow \infty$, we obtain both an *exact* sampler for the uniform distribution on $\Omega_{\mathbf{r},\mathbf{c}}$ and, as explained in Section 3, a randomized algorithm to approximate $|\Omega_{\mathbf{r},\mathbf{c}}|$ that runs in time $\Theta(N)$. We call such a running time optimal for uniform generation on $\Omega_{\mathbf{r},\mathbf{c}}$ since it takes at least N elements to encode a given binary table.

We study the asymptotic behavior as $N \rightarrow \infty$ of the probability that the configuration model generates a binary table. For this reason, we consider input sequences $(\mathbf{r}, \mathbf{c})_{N \geq 1}$, where for each $N \geq 1$, $\mathbf{r}(N)$ and $\mathbf{c}(N)$ are vectors with cardinality $m(N)$ and $n(N)$, respectively, and whose elements are non-negative integers and satisfy $\sum_{i=1}^{m(N)} r_i(N) = \sum_{j=1}^{n(N)} c_j(N) = N$. We assume that $r_1(N) \geq r_2(N) \geq \dots \geq r_{m(N)}(N)$ and $c_1(N) \geq c_2(N) \geq \dots \geq c_{n(N)}(N)$ for all N and allow $m(N)$ and $n(N)$ to vary with N . It will be convenient in our development to consider the vectors $\mathbf{r}(N)$ and $\mathbf{c}(N)$, for fixed N , as having infinite elements. To this end, we set $r_i(N) = 0$ for $i \geq m(N) + 1$ and $c_j(N) = 0$ for $j \geq n(N) + 1$. Without loss of generality we assume that $r_1(N) \geq c_1(N)$ for all N . Finally, for each i and j , we regard $(r_i)_{N \geq 1}$ and $(c_j)_{N \geq 1}$ as sequences in their own right. For brevity, we henceforth drop any explicit dependence on N from our notation.

Our main result characterizes the class of input sequences $(\mathbf{r}, \mathbf{c})_N$ for which the configuration model takes $\Theta(N)$ time to sample uniformly from $\Omega_{\mathbf{r},\mathbf{c}}$ and to approximate $|\Omega_{\mathbf{r},\mathbf{c}}|$. Note that if $c_1 = 1$, all the tables satisfying \mathbf{r} and \mathbf{c} are binary, so not only the configuration model generates only binary contingency tables, but also $|\Omega_{\mathbf{r},\mathbf{c}}|$ can be trivially obtained. Thus, we assume that $r_1 \geq c_1 \geq 2$ for all N . In order to have a more clear statement for Theorem 1.1 below, we will use the following technical assumptions:

(A1) For each $i \geq 1$, r_i is either $\Omega(N)$ or $o(N)$.

(A2) $\limsup_{N \rightarrow \infty} c_1 = \liminf_{N \rightarrow \infty} c_1$ (they may both be ∞).

We need these assumptions only in the proof of Lemma 7.2 for the analysis of the second part of condition 2 in Theorem 1.1 below. Note that sequences that do not satisfy these assumptions are such that the asymptotic behavior of either column 1 or some row i have substantial fluctuations that make their \liminf differ strictly from their \limsup . We believe that such sequences represent pathological cases that are unlikely to appear in most practical and theoretical applications; in fact, even if the sequences do not satisfy these assumptions, our results do not apply only when the sequences of row and column sums are presented in a particularly correlated way. Incorporating these cases into Theorem 1.1 would render the statement of condition 2 below much harder to understand. We discuss this further at the end of Section 7. We now state our main result, which gives necessary and sufficient conditions for the optimality of the configuration model.

Theorem 1.1. *Let T be a table produced by the configuration model given input sequences \mathbf{r} and \mathbf{c} satisfying (A1) and (A2) above. Let κ be the first row having sum $o(N)$ (i.e., $\kappa = \min\{i \geq 1: r_i = o(N)\}$). Then, $\mathbf{P}(T \in \Omega_{\mathbf{r},\mathbf{c}}) = \Omega(1)$ if and only if the following two conditions hold:*

1. $\sum_{i=1}^m \sum_{j=1}^n r_i(r_i - 1)c_j(c_j - 1) = O(N^2)$.
2. $\sum_{i=\kappa}^{\infty} r_i = \Omega(N)$ or $\lim_{N \rightarrow \infty} c_1 < \kappa$.

Remark 1.2. We point out that κ may be ∞ . In this case, from condition 1 in Theorem 1.1 we have that $c_1 = O(1)$ and the sum in condition 2 in Theorem 1.1 is equal to 0.

When $r_1 = o(N)$, condition 2 above is always satisfied since $\kappa = 1$ and $\sum_{i=\kappa}^{\infty} r_i = N$; in this case condition 1 is both necessary and sufficient. However, note that conditions 1 and 2 are not redundant. For instance, for any sequence (\mathbf{r}, \mathbf{c}) with $r_1 = N - o(N)$, $c_1 = 2$ and $c_2 = 1$, we have $\kappa = 2$, which violates condition 2, though condition 1 holds. A substantial discussion on how our main result relates to existing work is given in Section 2

Our theoretical developments are partly driven by our desire to guide practitioners in areas of applied statistics who often deal with hypothesis testing involving graphical models and binary contingency tables (see for instance [3] and [10]). In these types of settings, data is encoded in the form of a binary table and one is interested in studying the null hypothesis that row and column sums are sufficient statistics for determining the distribution of all the entries in the table. To test this hypothesis statisticians compare the value of a given statistic of the observed table (e.g., the sum of the hamming distances of pairs of rows) with values generated by sampling tables under the distribution induced by the null hypothesis, which is precisely the uniform distribution on binary contingency tables with prescribed row and column sums. For its simplicity and small running time, the configuration model is a very appealing algorithm to be used in this setting. Our Theorem 1.1 above fully characterizes the sequences \mathbf{r} and \mathbf{c} for which the configuration model is a fast and reliable algorithm for uniform generation of binary contingency tables.

The configuration model is by now a classical, well-known algorithm that has been applied in practice, as described above, and also in more theoretical settings. For example, some asymptotic estimates for $|\Omega_{\mathbf{r},\mathbf{c}}|$ (e.g., [17] and [13]) are obtained via analyses of the configuration model. Some results on the structural properties of graphs obtained uniformly at random from $\Omega_{\mathbf{r},\mathbf{c}}$ also use the configuration model (e.g., [15, Chapter 9] and [12]). Usually, it is easier to analyze a graph obtained via the configuration model than a random sample from $\Omega_{\mathbf{r},\mathbf{c}}$, and it is important to know whether results for one model can be carried over to the other. In order to explain how our results apply to this type of questions, let A be any property that can be tested for a bipartite graph (e.g., A can be the property that the graph has a connected component with a constant fraction of the vertices, which is the property studied in [12]). If the conditions in Theorem 1.1 hold, then any property A that holds with probability $1 - o(1)$ for the configuration model also holds with probability $1 - o(1)$ for a graph obtained uniformly at random from $\Omega_{\mathbf{r},\mathbf{c}}$. This corresponds to the notion of contiguity between probability measures, which is more thoroughly explained in [15, Chapter 9]. The corollary below gives an application of our results. We remark that this can only be obtained since the configuration model is an *exact* sampler for the uniform distribution over $\Omega_{\mathbf{r},\mathbf{c}}$.

Corollary 1.3. *Let A be a property that can be tested for a bipartite graph. Let $p(A)$ be the probability that a graph obtained uniformly at random from $\Omega_{\mathbf{r},\mathbf{c}}$ contains property A , and $p'(A)$ be the probability that a graph obtained via the configuration model given \mathbf{r} and \mathbf{c} contains property A . If conditions 1 and 2 in Theorem 1.1 are satisfied and $p'(A) = 1 - o(1)$, then $p(A) = 1 - o(1)$.*

Proof. Let ρ be the probability that the configuration model outputs a binary table. Note that $\rho = \Omega(1)$ if conditions 1 and 2 in Theorem 1.1 are satisfied. Since the configuration model is an exact sampler for the uniform distribution over $\Omega_{\mathbf{r},\mathbf{c}}$, we obtain $p(A) \geq 1 - \frac{1-p'(A)}{\rho} = 1 - o(1)$. \square

2 Related Work

Theorem 1.1 can be seen as an extension of recent work by Janson [14], who studied the probability that the configuration model generates a binary *symmetric* table. Letting $\Omega'_{\mathbf{r}}$ be the set of all

binary *symmetric* tables with row and column sums given by \mathbf{r} , [14, Theorem 1.1] establishes that $\mathbf{P}(T \in \Omega'_{\mathbf{r}}) = \Omega(1)$ if and only if $\sum_{i=1}^m r_i^2 = O(N)$.

To contrast Janson's result to the case of non-symmetric tables studied here, note that $(\mathbf{r}, \mathbf{c})_N$ satisfying conditions 1 and 2 give rise to a much wider class of behavior than in the symmetric case. For instance, the apparently similar conditions $\sum_{i=1}^m \sum_{j=1}^n r_i(r_i - 1)c_j(c_j - 1) = O(N^2)$ and $\sum_{i=1}^m \sum_{j=1}^n r_i^2 c_j^2 = O(N^2)$ are far from identical; if $\mathbf{c} = \{2, 1, 1, \dots, 1\}$, then the former condition is satisfied regardless of \mathbf{r} while the latter may not hold. Besides, the condition $\sum_{i=1}^m r_i^2 = O(N)$ for symmetric tables allows r_1 to grow only as $O(\sqrt{N})$, whereas our Theorem 1.1 reveals that there are sequences with r_1 as large as $N - o(N)$ for which the configuration model produces a binary table with probability $\Omega(1)$. Therefore, the growth behavior allowed for r_1 in Theorem 1.1 as $N \rightarrow \infty$ is much wider than in the symmetric case. This wider type of growth behavior makes the analysis for the non-symmetric case qualitatively different. Moreover, our proof techniques are completely different from those employed by Janson and reveal some structural properties of the tables generated with the configuration model. For example, we show that conditioning on the entries with relatively large row and column sums being binary, the probability that there is an entry with value larger than 2 is tiny (see Lemma 6.5). We believe that our techniques can be exploited in the analysis of related problems (such as efficient sampling of non-binary contingency tables).

Polynomial-time algorithms have been developed for the problem of approximating $|\Omega_{\mathbf{r}, \mathbf{c}}|$. In fact, approximating $|\Omega_{\mathbf{r}, \mathbf{c}}|$ can be reduced to the problem of computing the permanent of a binary $\ell \times \ell$ matrix with $\ell = \Theta(mn)$; a problem that enjoys a notable history and place in the theory of computation. Valiant [22] showed that computing the permanent belongs to the class of #P-complete problems, for which proving the existence of a polynomial-time algorithm would have extensive implications in complexity theory. It is still an open problem, however, to verify whether counting the number of binary contingency tables is #P-complete, though the more general problem of counting the number of (not necessarily binary) contingency tables has been shown to be #P-complete by Dyer et al. [11]. The ground-breaking work of Jerrum et al. [16] provided the first Fully Polynomial Randomized Approximation Scheme (FPRAS) [19] to compute the permanent of a binary matrix. Bézaková et al. [6] used simulated annealing techniques to develop an asymptotically faster algorithm to approximate the permanent, which runs in $O(\ell^7 \log^4 \ell)$ time for an $\ell \times \ell$ matrix. In another paper, Bézaková et al. [4] developed an algorithm that works directly with contingency tables. Their algorithm for approximately sampling binary tables runs in $O(m^2 n^2 N^3 \Delta \log^5(m+n))$ time, where Δ is the maximum over all row and column sums.

Although these algorithms are proved to run in polynomial time for all \mathbf{r} and \mathbf{c} , their efficiency is far from being useful in the types of applications described at the end of Section 1. For this reason, other approaches to uniformly sampling and counting binary contingency tables have been proposed. Chen et al. [10] developed a sequential importance sampling algorithm to count the number of contingency tables. Their algorithm applies a heuristic construction and has been observed to perform well in practice, but Bézaková et al. [5] proved that there exist \mathbf{r} and \mathbf{c} such that the heuristic of Chen et al. [10] underestimates the number of binary contingency tables by an exponential factor unless the algorithm is run for an exponential amount of time. On the other hand, Blanchet [7] provided a rigorous analysis of the heuristic of Chen et al. [10] and showed that if $r_1 = o(\sqrt{N})$, $\sum_{i=1}^m r_i^2 = O(N)$, and $c_1 = O(1)$, then this approach yields a FPRAS for counting binary contingency tables with running time $O(N^3)$. Our Theorem 1.1 significantly weakens the assumptions in [7], and drastically improves upon the running time of all the aforementioned algorithms.

In a different direction, much effort has been made to derive asymptotics for $|\Omega_{\mathbf{r}, \mathbf{c}}|$. The first result to allow the row and column sums to grow with N is the one by O'Neil [20], which is restricted to the case $n = m$ and $r_1 = O(\log^{1/4-\epsilon} n)$ for any constant $\epsilon > 0$. Later, McKay [17] considered the case

$r_1 = o(N^{1/4})$ and derived the first asymptotics for $|\Omega_{\mathbf{r},\mathbf{c}}|$ to allow r_1 to grow polynomially with N . Currently, the asymptotics for *sparse* binary tables that allows the largest range for \mathbf{r} and \mathbf{c} is the one by Greenhill et al. [13] for the case $r_1 c_1 = o(N^{2/3})$. These results by McKay [17] and Greenhill et al. [13] were obtained using the configuration model as a part of their proof technique. Similarly, the work of Blanchet discussed above [7] also uses the configuration model, as well as McKay’s estimator [17], to analyze the heuristics of Chen et al. [10]. Using different techniques, Canfield et al. [9] derived asymptotics for *dense* binary tables, and Barvinok [1] derived general lower and upper bounds for $|\Omega_{\mathbf{r},\mathbf{c}}|$ that are within a factor $(mn)^{\Theta(m+n)}$ from each other. For binary *symmetric* tables, besides the work of Janson [14] cited above, we highlight the work of Bayati et al. [2], who developed an algorithm that generates a symmetric table almost uniformly at random in time $O(r_1 N)$ as long as $r_1 = c_1 = O(N^{1/4-\epsilon})$ for any constant $\epsilon > 0$. Their analysis gives an alternative proof of a result originally derived by McKay [18].

We remark that, to the best of our knowledge, none of the existing asymptotics for $|\Omega_{\mathbf{r},\mathbf{c}}|$ applies to the whole of the spectrum of sequences \mathbf{r} and \mathbf{c} that satisfy conditions 1 and 2 in our Theorem 1.1. Furthermore, most of the known asymptotics take advantage of the configuration model in a fundamental way. Since our result fully characterizes the sequences for which the configuration model is contiguous to the uniform distribution, our conditions shed light into the whole spectrum of sequences for which analytical estimators might be obtained by directly applying the configuration model.

Under the conditions of Theorem 1.1, the configuration model gives a FPRAS for approximating $|\Omega_{\mathbf{r},\mathbf{c}}|$; thus it approximates $|\Omega_{\mathbf{r},\mathbf{c}}|$ to a precision of the form $1 + O(N^{-c})$, for an arbitrarily large constant $c > 0$ ¹, whereas asymptotics for $|\Omega_{\mathbf{r},\mathbf{c}}|$ have fixed precision. We remark that asymptotics for $|\Omega_{\mathbf{r},\mathbf{c}}|$ can also be used to produce an *almost* uniform sampling procedure for binary contingency tables. Sinclair and Jerrum [21] showed that for any self-reducible problem² an asymptotic approximation with at least constant precision can be used to produce an almost uniform sampling procedure. However, the running time of the sampling procedure depends on the mixing time of a Markov chain, which not only may be challenging to obtain precisely but also is usually too large for many practical applications. Moreover, we remark that this technique cannot be directly employed with the current asymptotics for $|\Omega_{\mathbf{r},\mathbf{c}}|$ since they impose some conditions on \mathbf{r} and \mathbf{c} . Under these conditions, the problem of sampling binary contingency tables is not guaranteed to be self-reducible: when splitting the table into smaller tables, we do not necessarily obtain that the new row and column sums satisfy the conditions of the asymptotic results.

3 Preliminaries

As mentioned in Section 1, we use the configuration model to generate a contingency table T (not necessarily binary). There are $N!$ possible matchings among the tokens, but any given *binary* contingency table generated by the configuration model corresponds to $\prod_{i=1}^m \prod_{j=1}^n r_i! c_j!$ such matchings, since permuting the tokens within each row or column does not change the final table. Therefore, we can conclude that $|\Omega_{\mathbf{r},\mathbf{c}}| \prod_{i=1}^m \prod_{j=1}^n r_i! c_j! = \mathbf{P}(T \in \Omega_{\mathbf{r},\mathbf{c}}) N!$, and the problem of computing $|\Omega_{\mathbf{r},\mathbf{c}}|$ is equivalent to evaluating $\mathbf{P}(T \in \Omega_{\mathbf{r},\mathbf{c}})$.

¹We remark that under the conditions of Theorem 1.1 the configuration model approximates $|\Omega_{\mathbf{r},\mathbf{c}}|$ to a precision of the form $1 \pm \epsilon$ for any *constant* $\epsilon > 0$ in time $\Theta(N)$, but can approximate $|\Omega_{\mathbf{r},\mathbf{c}}|$ to a precision $1 + O(N^{-c})$ for an arbitrary constant $c > 0$ in polynomial time.

²Informally, a problem is self-reducible if it can be split in parts where each part is itself a smaller instance of the same problem. In the case of sampling binary contingency tables, after generating all the entries of a given column, we can update the row and column sums properly so that generating the remaining entries translates to sampling a binary contingency table with different row and column sums.

If $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$, then we obtain a Fully Polynomial Randomized Approximation Scheme (FPRAS) for estimating $|\Omega_{\mathbf{r}, \mathbf{c}}|$ as follows (we refer the reader to [19] for more information on FPRAS). Generate a sequence of independent contingency tables using the configuration model and output the fraction of the tables that turn out to be binary. If $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$, for any constants $\delta > 0$ and $\epsilon > 0$, it suffices to generate a constant (depending polynomially on ϵ^{-1} and $\log \delta^{-1}$) number of tables such that with probability $1 - \delta$ our estimator to $|\Omega_{\mathbf{r}, \mathbf{c}}|$ has precision $1 \pm \epsilon$.

We conclude this section by introducing fundamental notation that we will use in the proof. Let I be the index set $[1, m] \times [1, n]$ and $T = (T_{i,j})_{(i,j) \in I}$ be a table generated by the configuration model. Let Z be the number of non-binary entries of T , so $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \mathbf{P}(Z = 0)$. Given two integers $k \geq 0$ and $x \geq 0$, we define $x^k = \frac{x!}{(x-k)!}$. Recall that the configuration model generates a table by taking a random matching between type-1 and type-2 tokens. We assume that each token is individually labeled and refer to a single pair of a type-1 and a type-2 token as an *edge*. We say that an edge is matched by the configuration model if the corresponding tokens are matched. A set of two edges for the same entry is referred to as a *double edge*. For $(i, j) \in I$, let $\mathcal{B}_2(i, j)$ be the set of all possible double edges that can be matched for the entry (i, j) . An element of $\mathcal{B}_2(i, j)$ has the form $\{e_1, e_2\}$, where e_1 and e_2 are disjoint edges for the entry (i, j) , that is, e_1 and e_2 correspond to 4 distinct tokens, 2 type-1 tokens from row i and 2 type-2 tokens from column j . Clearly, the cardinality of $\mathcal{B}_2(i, j)$ is given by $|\mathcal{B}_2(i, j)| = r_i^2 c_j^2 / 2!$. For any $(i, j) \in I$ and $B \in \mathcal{B}_2(i, j)$, let $\mathcal{M}(B)$ be the event that the double edge represented by B is matched by the configuration model. Note that given any specific $B \in \mathcal{B}_2(i, j)$, $\mathbf{P}(\mathcal{M}(B)) = 1/N^2$. With this notation, note that the event $\{Z \geq 1\}$ is equivalent to $\{\bigcup_{B \in \mathcal{B}_2} \mathcal{M}(B)\}$.

4 Proof of Theorem 1.1

The proof of Theorem 1.1 follows from the three propositions below, which we will prove in subsequent sections. We suggest the reader to ignore the assumptions (A1) and (A2) until the very end of the proof of Theorem 1.1 (more specifically until the proof of Lemma 7.2). We could have applied the assumptions to slightly simplify some other parts of the proof; however, we opted not to do so to emphasize exactly the places where the assumption must be used.

Proposition 4.1 below shows that condition 1 is necessary regardless of the value of r_1 ; its proof is given in Section 5.

Proposition 4.1. *If $\sum_{(i,j) \in I} r_i(r_i - 1)c_j(c_j - 1)$ is not $O(N^2)$, then $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}})$ is not $\Omega(1)$.*

The proof of Proposition 4.1 highlights the importance of the definition of *double edges*, since condition 1 in Theorem 1.1 translates to the expected number of double edges in T being uniformly bounded over N . Note that for the case of symmetric tables, condition 1 is both necessary and sufficient, while for the non-symmetric case it is just necessary. Now, we assume that $r_1 = o(N)$ and show in Proposition 4.2 that, in this case, condition 1 in Theorem 1.1 is also sufficient. The proof of Proposition 4.2 is presented in Section 6.

Proposition 4.2. *If $\sum_{(i,j) \in I} r_i(r_i - 1)c_j(c_j - 1) = O(N^2)$ and $r_1 = o(N)$ then $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$.*

If r_1 is not $o(N)$, i.e., $\limsup_{N \rightarrow \infty} r_1/N > 0$, then $(\mathbf{r}, \mathbf{c})_N$ contains a subsequence $(\mathbf{r}', \mathbf{c}')_N$ for which $r'_1 = \Omega(N)$ holds³. The next proposition deals with the case $r_1 = \Omega(N)$ and its proof is presented in Section 7.

³This is an example of a place where we could have applied assumption (A1), since it gives that if r_1 is not $o(N)$, then it is $\Omega(N)$. However, it is not crucial to use the assumption at this point and we will handle the case where r_1 is neither $o(N)$ nor $\Omega(N)$ right after the statement of Proposition 4.3.

Proposition 4.3. *If $\sum_{(i,j) \in I} r_i(r_i - 1)c_j(c_j - 1) = O(N^2)$ and $r_1 = \Omega(N)$, then $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ if and only if $\sum_{i=\kappa}^{\infty} r_i = \Omega(N)$ or $\lim_{N \rightarrow \infty} c_1 < \kappa$, where κ is defined as in Theorem 1.1.*

It is clear that Propositions 4.1, 4.2, and 4.3 establish that $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ if and only if both condition 1 and 2 in Theorem 1.1 are satisfied and r_i is either $o(N)$ or $r_1 = \Omega(N)$. We now explain the case when r_1 is neither $o(N)$ nor $\Omega(N)$, i.e., $\limsup_{N \rightarrow \infty} r_1/N > 0$ but $\liminf_{N \rightarrow \infty} r_1/N = 0$. For this, we will make use of the following technical lemma, which is also used in [15, chapter 9] and [14].

Lemma 4.4 (Subsubsequence principle). *If every subsequence $(\mathbf{r}', \mathbf{c}')_N$ of $(\mathbf{r}, \mathbf{c})_N$ contains a further subsequence $(\mathbf{r}'', \mathbf{c}'')_N$ for which $\mathbf{P}(T \in \Omega_{\mathbf{r}'', \mathbf{c}''}) = \Omega(1)$ then $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$.*

Proof. We will prove this lemma by contradiction. Assume that every subsequence $(\mathbf{r}', \mathbf{c}')_N$ of $(\mathbf{r}, \mathbf{c})_N$ contains a further subsequence $(\mathbf{r}'', \mathbf{c}'')_N$ for which $\mathbf{P}(T \in \Omega_{\mathbf{r}'', \mathbf{c}''}) = \Omega(1)$, but $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}})$ is not $\Omega(1)$. This means that $\liminf_{N \rightarrow \infty} \mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = 0$, and consequently there exists a subsequence $(\mathbf{r}''', \mathbf{c}''')_N$ of $(\mathbf{r}, \mathbf{c})_N$ such that $\lim_{N \rightarrow \infty} \mathbf{P}(T \in \Omega_{\mathbf{r}''', \mathbf{c}'''}) = 0$, which contradicts our assumption. \square

Finally, we will be able to conclude the proof of Theorem 1.1 with Lemma 4.5 below, which uses the subsubsequence principle. Since we will also apply this lemma later on, we give it in more generality than needed here.

Lemma 4.5. *Let \mathcal{Z} be some space of sequences $(\mathbf{r}, \mathbf{c})_N$ indexed by $N \geq 1$ such that if $(\mathbf{r}, \mathbf{c}) \in \mathcal{Z}$ then any subsequence of (\mathbf{r}, \mathbf{c}) is also in \mathcal{Z} . Given a sequence $(\mathbf{r}, \mathbf{c}) \in \mathcal{Z}$, let $f_N(\mathbf{r}, \mathbf{c})$ be a sequence of nonnegative real numbers indexed by $N \geq 1$. Assume that if for all $(\mathbf{r}, \mathbf{c}) \in \mathcal{Z}$ such that $f_N(\mathbf{r}, \mathbf{c}) = \Omega(1)$ or $f_N(\mathbf{r}, \mathbf{c}) = o(1)$ as $N \rightarrow \infty$, we have $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$. Then, $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ also holds for the $(\mathbf{r}, \mathbf{c}) \in \mathcal{Z}$ for which $\limsup_{N \rightarrow \infty} f_N(\mathbf{r}, \mathbf{c}) > \liminf_{N \rightarrow \infty} f_N(\mathbf{r}, \mathbf{c}) = 0$.*

Proof. We use the subsubsequence principle (Lemma 4.5). If $\limsup_{N \rightarrow \infty} f_N(\mathbf{r}, \mathbf{c}) > \liminf_{N \rightarrow \infty} f_N(\mathbf{r}, \mathbf{c}) = 0$, then for every subsequence $(\mathbf{r}', \mathbf{c}')$ of (\mathbf{r}, \mathbf{c}) it is the case that $(\mathbf{r}', \mathbf{c}') \in \mathcal{Z}$ (by the property of \mathcal{Z}) and either $\limsup_{N \rightarrow \infty} f_N(\mathbf{r}', \mathbf{c}') = 0$ or there exists a subsequence $(\mathbf{r}'', \mathbf{c}'') \in \mathcal{Z}$ of $(\mathbf{r}', \mathbf{c}')$ for which $\liminf_{N \rightarrow \infty} f_N(\mathbf{r}'', \mathbf{c}'') > 0$. In the former case, since $f_N(\mathbf{r}', \mathbf{c}') = o(1)$, we know that $\mathbf{P}(T \in \Omega_{\mathbf{r}', \mathbf{c}'}) = \Omega(1)$. In the latter case, since $f_N(\mathbf{r}'', \mathbf{c}'') = \Omega(1)$, we have $\mathbf{P}(T \in \Omega_{\mathbf{r}'', \mathbf{c}''}) = \Omega(1)$. Therefore, using the subsubsequence principle we obtain $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$. \square

We set \mathcal{Z} as the space of sequences satisfying conditions 1 and 2 from Theorem 1.1. It is easy to check that this space satisfies the condition in Lemma 4.5. Then, we set $f_N(\mathbf{r}, \mathbf{c}) = r_1/N$, and Lemma 4.5 gives that $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ for the case when r_1 is neither $\Omega(N)$ nor $o(N)$ but conditions 1 and 2 hold. This completes the proof of Theorem 1.1. \square

5 Proof of Proposition 4.1

We prove Proposition 4.1 using the second-moment method. We define the function

$$\mu(N) = \sum_{(i,j) \in I} \frac{r_i^2 c_j^2}{2N^2}, \quad (1)$$

which satisfies $\limsup_{N \rightarrow \infty} \mu(N) = \infty$ by the assumptions of Proposition 4.1, and show that under this condition $\liminf_{N \rightarrow \infty} \mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = 0$ (i.e., $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}})$ is not $\Omega(1)$). Let F be the random

variable counting the number of double edges that are matched by the configuration model, that is, $F = \sum_{B \in \mathcal{B}_2} \mathbf{1}(\mathcal{M}(B))$, where $\mathbf{1}(\cdot)$ is the indicator function. Note that

$$\mathbf{E}F = \sum_{(i,j) \in I} \sum_{B \in \mathcal{B}_2(i,j)} \frac{1}{N^2} = \sum_{(i,j) \in I} \frac{r_i^2 c_j^2}{2!N^2} = \mu(N). \quad (2)$$

Our strategy is to use Chebyshev's inequality to obtain an upper bound for $\mathbf{P}(T \in \Omega_{\mathbf{r},\mathbf{c}})$ via

$$\mathbf{P}(T \in \Omega_{\mathbf{r},\mathbf{c}}) = \mathbf{P}(F \leq 0) = \mathbf{P}(\mathbf{E}F - F \geq \mathbf{E}F) \leq \frac{\mathbf{Var}(F)}{\mathbf{E}^2 F}. \quad (3)$$

We now derive an upper bound for $\mathbf{Var}(F)$. Note that

$$\mathbf{E}F^2 = \sum_{(i,j) \in I} \sum_{B \in \mathcal{B}_2(i,j)} \sum_{(i',j') \in I} \sum_{B' \in \mathcal{B}_2(i',j')} \mathbf{P}(\mathcal{M}(B) \cap \mathcal{M}(B')),$$

from which we can write

$$\begin{aligned} \mathbf{Var}(F) &= \sum_{(i,j) \in I} \sum_{B \in \mathcal{B}_2(i,j)} \mathbf{P}(\mathcal{M}(B)) \sum_{(i',j') \in I} \sum_{B' \in \mathcal{B}_2(i',j')} (\mathbf{P}(\mathcal{M}(B') | \mathcal{M}(B)) - \mathbf{P}(\mathcal{M}(B'))) \quad (4) \\ &\leq \sum_{(i,j) \in I} \frac{r_i^2 c_j^2}{2N^2} (1 + \varphi_1(i,j) + \varphi_2(i,j) + \varphi_3(i,j) + \varphi_4(i,j)), \end{aligned}$$

where the terms 1 and $\varphi_1(i,j)$ to $\varphi_4(i,j)$ are explained next. First of all, when $(i,j) = (i',j')$ and $B = B'$ we have that $(\mathbf{P}(\mathcal{M}(B') | \mathcal{M}(B)) - \mathbf{P}(\mathcal{M}(B'))) \leq 1$. Now, the term $\varphi_1(i,j)$ corresponds to the cases where B and B' are double edges for the same entry (i,j) and also have one edge in common (i.e., $B \cup B'$ is a set of three edges). In such cases, to compute $\varphi_1(i,j)$ we shall use $\mathbf{P}(\mathcal{M}(B') | \mathcal{M}(B)) - \mathbf{P}(\mathcal{M}(B')) \leq \mathbf{P}(\mathcal{M}(B') | \mathcal{M}(B))$ and simply estimate $\mathbf{P}(\mathcal{M}(B') | \mathcal{M}(B))$. The term $\varphi_2(i,j)$ corresponds to the terms where B and B' are double edges for the same entry (i,j) but have no edge in common (i.e., $B \cup B'$ is a set of four edges). Before proceeding to describe $\varphi_3(i,j)$ and $\varphi_4(i,j)$, let us explain how to compute $\varphi_1(i,j)$ and $\varphi_2(i,j)$, which we express as

$$\varphi_1(i,j) = \frac{2(r_i - 2)(c_j - 2)}{N - 2}, \quad (5)$$

and

$$\varphi_2(i,j) = \frac{(r_i - 2)^2 (c_j - 2)^2}{2(N - 2)^2} - \frac{r_i^2 c_j^2}{2N^2} \leq 0. \quad (6)$$

Given that a double edge $\{e_1, e_2\}$ is chosen from the entry (i,j) , $\varphi_1(i,j)$ is the probability that another edge e_3 from (i,j) is chosen, which is given by $\frac{(r_i - 2)(c_j - 2)}{N - 2}$. The additional factor 2 comes from the fact that, once we fix e_3 , there are 2 possible choices of double edges for B' , namely $B' = \{e_1, e_3\}$ and $B' = \{e_2, e_3\}$. The equation for $\varphi_2(i,j)$ is obtained in a similar way, but we need to compute the probability that we choose a double edge $\{e_3, e_4\}$ from (i,j) such that $\{e_1, e_2\} \cap \{e_3, e_4\} = \emptyset$, which gives the term $\frac{(r_i - 2)^2 (c_j - 2)^2}{2(N - 2)^2}$. The last term in (6) comes from the term $\mathbf{P}(\mathcal{M}(B'))$ from (4), which is the probability that a double edge $\{e_3, e_4\}$ is chosen independently of $\{e_1, e_2\}$.

The term $\varphi_3(i,j)$ corresponds to the terms where B and B' are double edges for the same row but different columns or for the same column but different rows. Using similar reasoning we obtain

$$\varphi_3(i,j) = \sum_{i' \neq i} \left(\frac{r_{i'}^2 (c_j - 2)^2}{2(N - 2)^2} - \frac{r_{i'}^2 c_j^2}{2N^2} \right) + \sum_{j' \neq j} \left(\frac{(r_i - 2)^2 c_{j'}^2}{2(N - 2)^2} - \frac{r_i^2 c_{j'}^2}{2N^2} \right) \leq 0, \quad (7)$$

Ultimately, $\varphi_4(i, j)$ corresponds to terms where B and B' represent disjoint rows and columns and is given by

$$\varphi_4(i, j) = \sum_{i' \neq i} \sum_{j' \neq j} \left(\frac{r_{i'}^2 c_{j'}^2}{2(N-2)^2} - \frac{r_{i'}^2 c_{j'}^2}{2N^2} \right). \quad (8)$$

Now we simplify the equations above. For $\varphi_2(i, j)$ and $\varphi_3(i, j)$ we simply use the fact that they are at most 0. For $\varphi_1(i, j)$, we have that (1) implies

$$\varphi_1(i, j) \leq 2 \frac{\sqrt{r_i(r_i-1)c_j(c_j-1)}}{N-2} \leq 2\sqrt{2\mu(N)}(1 + O(1/N)), \quad (9)$$

for $r_i, c_j \geq 2$. Clearly, the entries (i, j) with $r_i \leq 1$ or $c_j \leq 1$ do not contribute to F . Since $1/(N-2)^2 - 1/N^2 = O(1/N^3)$ we can write

$$\varphi_4(i, j) \leq \sum_{i' \neq i} \sum_{j' \neq j} \frac{r_{i'}^2 c_{j'}^2}{2} O(1/N^3) = O(\mu(N)/N).$$

Therefore, the variance of F translates to

$$\begin{aligned} \mathbf{Var}(F) &\leq \sum_{(i,j) \in I} \frac{r_i^2 c_j^2}{2N^2} \left(1 + 2\sqrt{2\mu(N)} + O(\mu(N)/N) \right) \\ &= \mu(N) O(\sqrt{\mu(N)} + \mu(N)/N). \end{aligned} \quad (10)$$

Since $r_i, c_j \leq N$ for all $(i, j) \in I$, we have $\mu(N) = O(N^2)$, and consequently, $\sqrt{\mu(N)} = O(N)$. Now, plugging (2) and (10) into (3), we obtain that there are constants C and N_0 such that for all $N \geq N_0$ we obtain

$$\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \mathbf{P}(F \leq 0) \leq \frac{C}{\sqrt{\mu(N)}} + \frac{C}{N}.$$

Taking the lim inf as $N \rightarrow \infty$ concludes the proof of Proposition 4.1 since $\limsup_{N \rightarrow \infty} \mu(N) = \infty$. \square

6 Proof of Proposition 4.2

In the proof of Proposition 4.2 we assume that $2 \leq c_1 \leq r_1 = o(N)$ and that there exists a constant $C > 0$ such that for all sufficiently large N it holds that

$$\sum_{(i,j) \in I} \frac{r_i^2 c_j^2}{2N^2} \leq C. \quad (11)$$

We split the table T into two regions I_L and I_S . Let $\epsilon > 0$ be a small number that we will set later, and define the index sets $I_L = \{(i, j) \in I : (r_i - 1)(c_j - 1) \geq \epsilon N\}$ and $I_S = I \setminus I_L$. We remark that the set I_L may be empty. Intuitively, I_L represents the entries of T with large row and column sums. Since the r_i 's and the c_j 's are assumed to be non-increasing, a useful conceptual diagram for the definition of I_L is given by the shaded area in Figure 1(a).

Let Z_L be the number of non-binary entries in I_L , and Z_S be the number of non-binary entries in I_S . Clearly, $Z = Z_L + Z_S$. Let W_L be the sum of the entries in I_L (i.e., $W_L = \sum_{(i,j) \in I_L} T_{i,j}$). Note that $\{W_L = 0\} \subseteq \{Z_L = 0\}$, which gives $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) \geq \mathbf{P}(W_L = 0) \mathbf{P}(Z_S = 0 \mid W_L = 0)$. We will

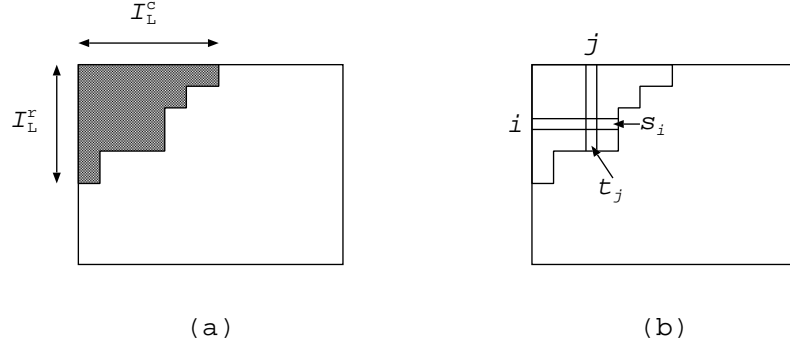


Figure 1: Diagram illustrating the entries in I_L (shaded area in part (a)) and the definition of I_L^r , I_L^c , s_i , and t_j .

deal with the terms $\mathbf{P}(W_L = 0)$ and $\mathbf{P}(Z_S = 0 \mid W_L = 0)$ separately. We will need the following definitions. Let I_L^r be the set of rows with at least one entry in I_L (i.e., $I_L^r = \{i: (i, j) \in I_L \text{ for some } j\}$), and I_L^c be the set of columns with at least one entry in I_L (see Figure 1(a)). Lemma 6.1 below deals with the term $\mathbf{P}(W_L = 0)$.

Lemma 6.1. *If (11) is satisfied and $r_1 = o(N)$, then we obtain $|I_L| = O(1)$ and $\mathbf{P}(W_L = 0) = \Omega(1)$.*

Proof. We assume that I_L is not empty (otherwise the lemma vacuously holds) and that (11) is satisfied. Note the importance of the assumption $c_1 \leq r_1 = o(N)$; if for instance $r_i = \Omega(N)$ for all i , then all the entries from column 1 could be contained in I_L and, therefore, we would have $\mathbf{P}(W_L = 0) = 0$, violating the statement of the lemma.

Let $\gamma = \sum_{(i,j) \in I_L} r_i c_j / N$. We show that $\gamma, |I_L| = O(1)$. For all sufficiently large N we have

$$C \geq \sum_{(i,j) \in I_L} \frac{r_i^2 c_j^2}{2N^2} \geq \sum_{(i,j) \in I_L} \frac{r_i c_j \epsilon}{2N} = \frac{\gamma \epsilon}{2}, \quad (12)$$

where the second inequality is obtained from $(r_i - 1)(c_j - 1) \geq \epsilon N$, for all $(i, j) \in I_L$ by the definition of I_L . On the other hand, from the definition of I_L we obtain

$$\gamma = \sum_{(i,j) \in I_L} \frac{r_i c_j}{N} \geq \sum_{(i,j) \in I_L} \frac{(r_i - 1)(c_j - 1)}{N} \geq |I_L| \epsilon.$$

Consequently, combining the previous estimate with (12) we conclude that $\gamma, |I_L| = O(1)$.

It is useful to see Figure 1(a,b) throughout this discussion. Let $s_i = \sum_{j': (i,j') \in I_L} c_{j'}$ and $t_j = \sum_{i': (i',j) \in I_L} r_{i'}$. In words, for any row $i \in I_L^r$, s_i is the sum of the column sums over all entries in I_L corresponding to row i . Similarly, t_j is defined for any column $j \in I_L^c$ as the sum of the row sums over all entries in I_L corresponding to column j . Note that $\gamma N = \sum_{i \in I_L^r} r_i s_i = \sum_{j \in I_L^c} c_j t_j$.

We now derive a lower bound for $\mathbf{P}(W_L = 0)$. Consider the row 1, which belongs to I_L^r . The number of ways we can match the r_1 type-1 tokens of the first row with the $N - s_1$ type-2 tokens available

outside I_L is $(N - s_1)^{r_1}$. Employing this reasoning for each row in I_L^c , we obtain (let $\ell = |I_L^c|$)

$$\begin{aligned} \mathbf{P}(W_L = 0) &= \frac{(N - s_1)^{r_1} (N - r_1 - s_2)^{r_2} \cdots (N - r_1 - \cdots - r_{\ell-1} - s_\ell)^{r_\ell}}{N^{r_1} (N - r_1)^{r_2} \cdots (N - r_1 - \cdots - r_{\ell-1})^{r_\ell}} \\ &\geq \left(1 - \frac{s_1}{N - \sum_{i' \in I_L^c} r_{i'}}\right)^{r_1} \cdots \left(1 - \frac{s_\ell}{N - \sum_{i' \in I_L^c} r_{i'}}\right)^{r_\ell} \\ &= \left(1 - \frac{s_1}{N - o(N)}\right)^{r_1} \cdots \left(1 - \frac{s_\ell}{N - o(N)}\right)^{r_\ell} \end{aligned} \quad (13)$$

$$\begin{aligned} &\geq \exp\left(-\sum_{i \in I_L^c} \frac{r_i s_i}{[N - o(N)][1 - O(s_i/N)]}\right) \\ &\geq \exp(-\gamma) - o(1), \end{aligned} \quad (14)$$

where in (13) we used the fact that $\sum_{i' \in I_L^c} r_{i'} \leq r_1 |I_L^c| = o(N)$, and (14) comes from the fact that $(1 - x) \geq \exp(-x/(1 - x))$ for all $x \in [0, 1]$, which we apply with $x = s_i/(N - o(N))$ for each $i \in I_L^c$. Moreover, since $r_1 \geq c_1$ and (11) holds, then $c_j = O(\sqrt{N})$ for all j , and consequently, $s_i \leq c_1 |I_L^c| = O(\sqrt{N})$. This completes the proof of the lemma, since $\gamma = O(1)$. \square

For the term $\mathbf{P}(Z_S = 0 \mid W_L = 0)$, we use the fact that I_L contains only $O(1)$ entries to conclude that conditioning on I_L has a small effect. Then, we can carry out the analysis as if no conditioning is being made, and we use the fact that $(r_i - 1)(c_j - 1) < \epsilon N$ for $(i, j) \in I_S$ to simplify the calculations. The following lemma, which we prove in a moment, deals with this case.

Lemma 6.2. *If (11) is satisfied, then $\mathbf{P}(Z_S = 0 \mid W_L = 0) = \Omega(1)$.*

Proposition 4.2 follows immediately from Lemmas 6.1 and 6.2. \square

Proof of Lemma 6.2

We devote the remainder of this section to prove Lemma 6.2. The proof is rather delicate and will also require additional lemmas. We will use the following quantity

$$\lambda = \lambda(N) = \sum_{(i,j) \in I_S} \frac{r_i^2 c_j^2}{2N^2}.$$

Clearly, $\limsup_{N \rightarrow \infty} \lambda \leq C$ given (11), and $\lambda = \mathbf{E}[Z_S]$.

Before proceeding to the proof, we need to introduce some notation. For an integer $k \geq 1$ and any $(i, j) \in I$, let $\mathcal{B}_k(i, j)$ be the set of all sets of k disjoint edges between row i and column j , which generalizes the definition of $\mathcal{B}_2(i, j)$ from Section 3. A typical element of $\mathcal{B}_k(i, j)$ has the form $\{e_1, e_2, \dots, e_k\}$, where e_1, e_2, \dots, e_k are disjoint edges between row i and column j . Recall that the edges e_1, e_2, \dots, e_k are said to be disjoint when each edge e_ℓ , $1 \leq \ell \leq k$, corresponds to a pair of tokens $\{\beta_\ell, \beta'_\ell\}$ such that $\beta_1 \neq \beta_2 \neq \cdots \neq \beta_k$ and $\beta'_1 \neq \beta'_2 \neq \cdots \neq \beta'_k$, that is, the edges e_1, e_2, \dots, e_k do not share tokens. Intuitively, $\mathcal{B}_k(i, j)$ is the set of all possible ways we can match k edges between row i and column j . So, clearly $|\mathcal{B}_k(i, j)| = \frac{r_i^k c_j^k}{k!}$, for all $(i, j) \in I$ and all k .

For any set of edges B , we define $\mathcal{M}(B)$ as the event that all the edges in B are matched by the configuration model. Note that the occurrence of $\mathcal{M}(B)$ for $B \in \mathcal{B}_k(i, j)$ means that the entry $T_{i,j} \geq$

k . For all k , define $\mathcal{B}_k = \bigcup_{(i,j) \in I_S} \mathcal{B}_k(i,j)$. Finally, for all $k \geq 1$, define the event $\mathcal{P}_k = \bigcup_{B \in \mathcal{B}_k} \mathcal{M}(B)$. In words, \mathcal{P}_k is the event that there is an entry $T_{i,j} \geq k$ in I_S . Using our notation we can write

$$\mathbf{P}(Z_S \geq 1 \mid W_L = 0) = \mathbf{P}(\mathcal{P}_2 \mid W_L = 0). \quad (15)$$

When (11) is satisfied, we can show that the probability that \mathcal{P}_3 occurs is small. We exploit this fact to simplify the calculations via the following trivial inequalities

$$\mathbf{P}(\mathcal{P}_2 \mid W_L = 0) \geq \mathbf{P}(\mathcal{P}_2 \cap \mathcal{P}_3^c \mid W_L = 0) \quad (16)$$

and

$$\mathbf{P}(\mathcal{P}_2 \mid W_L = 0) \leq \mathbf{P}(\mathcal{P}_2 \cap \mathcal{P}_3^c \mid W_L = 0) + \mathbf{P}(\mathcal{P}_3 \mid W_L = 0). \quad (17)$$

We start the proof by stating a lemma that indicates that for any two disjoint sets of disjoint edges B and B' , the probability that the edges in B are matched by the configuration model conditioning on B' being matched and $W_L = 0$ is essentially the same as without conditioning.

Lemma 6.3. *Let B and B' be disjoint sets of disjoint edges from I_S such that the event $\mathcal{M}(B \cup B')$ has non-zero probability when conditioned on $W_L = 0$. Let k and k' be the number of edges in B and B' , respectively. If $k, k' = O(1)$, we obtain*

$$\mathbf{P}(\mathcal{M}(B) \mid \mathcal{M}(B'), W_L = 0) = \frac{1 + o(1)}{(N - k')^{\underline{k}}}. \quad (18)$$

Proof. Before proceeding to the proof we show how to sample a table T under the condition $W_L = 0$ and $\mathcal{M}(B')$. If we were to condition only on $\mathcal{M}(B')$, it would suffice to disregard the tokens corresponding to the edges in B' and take a random pairing for the remaining tokens in a standard fashion. However, conditioning on $W_L = 0$ is more delicate. For example, conditioning on $W_L = 0$ implies that a pair of tokens from row i and column j for which $(i,j) \in I_L$ cannot be matched since all the entries in I_L are 0.

Define $s_i = \sum_{j': (i,j') \in I_L} c_{j'}$ and $t_j = \sum_{i': (i',j) \in I_L} r_{i'}$ (refer to Figure 1(b) for a pictorial illustration of s_i and t_j). Let ρ_i be the number of edges in B' corresponding to row i , and let ρ'_j be the number of edges in B' corresponding to column j . We sample T in a column-by-column manner, but the order according to which we sample the columns will matter. For each column j , there is a set X_j of type-1 tokens that can be matched to type-2 tokens from column j given $W_L = 0$ and B' . Also, since the columns are in non-increasing order, we obtain that $X_j \subseteq X_{j'}$ for all $j' \in [j, n]$. Our strategy is to sample the columns in a non-increasing order, starting from column 1 until column $|I_L^c|$. (Recall the definition $I_L^c = \{j: (i,j) \in I_L \text{ for some } i\}$.) Then, at that moment, all entries corresponding to a column in I_L have already been assigned. Therefore, sampling the remaining entries of the table does not depend on $W_L = 0$ and we carry out the sampling trivially using the standard procedure for the configuration model.

We now describe how to sample the entries from a column $j \in I_L^c$ given that all columns from 1 to $j - 1$ have already been sampled. Note that there are $c_j - \rho'_j$ type-2 tokens still unmatched for column j . The main property we use is that $X_j \subseteq X_{j'}$ for all $j' \in [j, n]$, that is, the type-1 tokens that can be matched to the type-2 tokens from column j can also be matched to any other column j' that have not yet been sampled. Therefore, it follows that each possible way to match the tokens from column j is equally likely; we can take a uniformly random permutation of the type-1 tokens in X_j that have not been matched to any column $j' < j$ and select the first $c_j - \rho'_j$ to be matched to the type-2 tokens from column j .

Recall that k and k' are the cardinalities of B and B' , respectively. Now we proceed to the proof of (18). First, assume $k = 1$, i.e., $B = \{e_1\} = \{(\beta_1, \beta'_1)\}$, where β_1 is a type-1 token from some row i and β'_1 is a type-2 token from some column j . We denote by \mathcal{I}_j the set of entries in I_L corresponding to column $j' \leq j - 1$. Formally, $\mathcal{I}_j = \{(i', j') \in I_L : j' \in [1, j - 1]\}$. We need to consider two cases.

Case 1: β'_1 corresponds to a column $j \in I_L^c$.

We write the probability that β_1 is matched to β'_1 as $q_1 q_2$, where q_1 is the probability that β_1 has not been matched to any column $j' \leq j - 1$ and q_2 is the probability that β_1 is matched to β'_1 given that it was not matched to any column $j' \leq j - 1$. We start with q_2 . Note that there are $\zeta_j = N - t_j - \sum_{j'=1}^{j-1} c_{j'} - \left(k' - \sum_{j'=1}^{j-1} \rho'_{j'}\right)$ type-1 tokens available to be matched to β'_1 . Note that $t_j \leq r_1 |I_L| = o(N)$ and $\sum_{j'=1}^{j-1} c_{j'} - \sum_{j'=1}^{j-1} \rho'_{j'} = O(j\sqrt{N})$. Therefore, since $j = O(1)$, we have

$$q_2 = \frac{1}{\zeta_j} = \frac{1}{N - k' - o(N)}.$$

For q_1 , note first that the probability that β_1 is not matched to any type-2 token from a column j' is

$$\frac{(\zeta_{j'} - 1)^{c_{j'} - \rho'_{j'}}}{\zeta_{j'}^{c_{j'} - \rho'_{j'}}} = \frac{\zeta_{j'} - c_{j'} + \rho'_{j'}}{\zeta_{j'}} = 1 - O(1/\sqrt{N}),$$

since $k' = O(1)$. Recall that i is the row associated with the token β_1 . Clearly, when assigning a column j' for which $(i, j') \in I_L$, we have that β_1 will not be matched to any token from column j' since we are conditioning on $W_L = 0$. Therefore, we obtain for q_1

$$q_1 = \prod_{j' \leq j-1: (i, j') \notin I_L} \left(1 - O(1/\sqrt{N})\right) = 1 - O(1/\sqrt{N}),$$

since $j = O(1)$. We then obtain that β_1 is matched to β'_1 with probability $\frac{1}{N-k'}(1 + o(1))$.

Case 2: β'_1 corresponds to a column $j \notin I_L^c$.

Again, with probability $1 - O(1/\sqrt{N})$, β_1 was not matched to any type-2 token from columns in I_L^c . When this happens, there are still $N - \sum_{j' \in I_L^c} c_{j'} - \left(k' - \sum_{j' \in I_L^c} \rho'_{j'}\right) = N - k' - O(\sqrt{N})$ type-1 tokens to be matched to β'_1 and the probability that β_1 and β'_1 are matched is $\frac{1}{N-k'}(1 + O(1/\sqrt{N}))$.

Therefore, for $k = 1$ and $k' = O(1)$, we obtain $\mathbf{P}(\mathcal{M}(B) \mid \mathcal{M}(B'), W_L = 0) = \frac{1}{N-k'}(1 + o(1))$. When $k \geq 2$, let $B = \{e_1, e_2, \dots, e_k\}$ and B_ℓ be the first ℓ edges in B , i.e., $B_\ell = \{e_1, e_2, \dots, e_\ell\}$. For convenience, define $B_0 = \emptyset$. Therefore

$$\begin{aligned} \mathbf{P}(\mathcal{M}(B) \mid \mathcal{M}(B'), W_L = 0) &= \prod_{\ell=1}^k \mathbf{P}(\mathcal{M}(\{e_\ell\}) \mid \mathcal{M}(B_{\ell-1}), \mathcal{M}(B'), W_L = 0) \\ &= \prod_{\ell=1}^k \left(\frac{1}{N - k' - \ell + 1} (1 + o(1)) \right), \end{aligned}$$

which concludes the proof of Lemma 6.3 since $k, k' = O(1)$. \square

To simplify the calculations to follow, we first solve the simpler case when $\lambda = o(1)$. For this we prove Lemma 6.4 below, which gives a stronger result.

Lemma 6.4. *If $\limsup_{N \rightarrow \infty} \lambda < 1$, we have $\mathbf{P}(Z_S = 0 \mid W_L = 0) = \Omega(1)$.*

Proof. We will show that $\liminf_{N \rightarrow \infty} \mathbf{P}(Z_S = 0 \mid W_L = 0) > 0$. Using Markov's inequality, we can write

$$\liminf_{N \rightarrow \infty} \mathbf{P}(Z_S = 0 \mid W_L = 0) = 1 - \limsup_{N \rightarrow \infty} \mathbf{P}(Z_S \geq 1 \mid W_L = 0) \geq 1 - \limsup_{N \rightarrow \infty} \mathbf{E}[Z_S \mid W_L = 0].$$

Using linearity of expectation, we obtain $\mathbf{E}[Z_S \mid W_L = 0] = \sum_{(i,j) \in I_S} \sum_{B \in \mathcal{B}_2(i,j)} \mathbf{P}(\mathcal{M}(B) \mid W_L = 0)$. From Lemma 6.3 we have $\mathbf{P}(\mathcal{M}(B) \mid W_L = 0) = (1 + o(1))/N^2$, which gives

$$\liminf_{N \rightarrow \infty} \mathbf{P}(Z_S = 0 \mid W_L = 0) \geq 1 - \limsup_{N \rightarrow \infty} \sum_{(i,j) \in I_S} \frac{r_i^2 c_j^2}{2N^2} (1 + o(1)) \geq 1 - \limsup_{N \rightarrow \infty} \lambda.$$

□

From now on, we will assume that $\lambda = \Theta(1)$. The case where λ is neither $o(1)$ nor $\Omega(1)$ can be handled by Lemma 4.5 by setting $f_N(\mathbf{r}, \mathbf{c}) = \lambda$ and \mathcal{Z} as the space of sequences satisfying the conditions in Proposition 4.2.

Now we use Lemma 6.3 to show that the bounds in (16) and (17) are tight up to smaller-order terms. This simplification is the main reason for treating I_L and I_S separately.

Lemma 6.5. *Conditional on $W_L = 0$, the probability that the configuration model creates three edges for any entry in I_S can be upper bounded by $\mathbf{P}(\mathcal{P}_3 \mid W_L = 0) \leq \lambda\epsilon/3 + o(1)$.*

Proof. For any $(i, j) \in I_S$, the number of ways to match three edges from (i, j) is $r_i^3 c_j^3 / 3!$, and

$$\mathbf{P}(\mathcal{P}_3 \mid W_L = 0) \leq \sum_{(i,j) \in I_S} \sum_{B \in \mathcal{B}_3(i,j)} \mathbf{P}(\mathcal{M}(B) \mid W_L = 0).$$

We use $(r_i - 1)(c_j - 1) < \epsilon N$ and Lemma 6.3 to conclude that for all $(i, j) \in I_S$

$$\mathbf{P}(\mathcal{P}_3 \mid W_L = 0) \leq \sum_{(i,j) \in I_S} \frac{r_i^3 c_j^3}{3!} \frac{(1 + o(1))}{N^3} < \sum_{(i,j) \in I_S} \frac{r_i^2 c_j^2}{6N^2} \epsilon (1 + o(1)),$$

which together with (11) yields the validity of Lemma 6.5. □

It remains to derive a bound for the term $\mathbf{P}(\mathcal{P}_2 \cap \mathcal{P}_3^c \mid W_L = 0)$. We apply the inclusion-exclusion principle. Define the set \mathcal{D}_ℓ to contain all possible sets of the form $\{d_1, d_2, \dots, d_\ell\}$ with d_1, d_2, \dots, d_ℓ being distinct double edges (i.e., distinct elements from \mathcal{B}_2). Ideally, we would like $(d_1, d_2, \dots, d_\ell) \in \mathcal{D}_\ell$ to represent a possible choice of ℓ double edges for ℓ distinct entries of the table. However, it may be the case that, say, d_1 and d_2 are double edges for the same entry. In this case, if d_1 and d_2 have one edge in common, then they correspond to having 3 edges matched for that entry. Otherwise, if d_1 and d_2 do not have an edge in common, then they correspond to having 4 edges matched for that entry. In any of these cases, we count these terms in the event \mathcal{P}_3 , which we treat separately using Lemma 6.5. That is the reason why we derive the probability for $\mathcal{P}_2 \cap \mathcal{P}_3^c$ instead of working directly with \mathcal{P}_2 . This is also the main benefit we obtain from considering the entries in I_L and I_S separately. The elements of \mathcal{D}_ℓ that count for the event $\mathcal{P}_2 \cap \mathcal{P}_3^c$ are only those corresponding to ℓ double edges from ℓ distinct entries. There is also one additional case. There exist terms of the form

$D = \{d_1, d_2, \dots, d_\ell\} \in \mathcal{D}_\ell$ such that, say, d_1 and d_2 share a token. This happens if for example d_1 is a double edge for the entry (i, j) , d_2 is a double edge for the entry (i, j') , and one of the type-1 tokens from row i contained in d_1 is also contained in d_2 . However, should that be the case, then d_1 and d_2 cannot occur simultaneously and the event that all double edges in D are matched by the configuration model has probability 0.

Abusing notation slightly, for an element $D \in \mathcal{D}_\ell$, we denote by $\mathcal{M}(D)$ the event that all ℓ double edges in D are matched by the configuration model. Therefore, using the inclusion-exclusion principle, we obtain

$$\mathbf{P}(\mathcal{P}_2 \cap \mathcal{P}_3^c \mid W_L = 0) = \sum_{\ell \geq 1} (-1)^{\ell+1} p_\ell, \quad (19)$$

where $p_\ell = \sum_{D \in \mathcal{D}_\ell} \mathbf{P}(\mathcal{M}(D) \cap \mathcal{P}_3^c \mid W_L = 0)$. We will take a value $L > 0$ sufficiently large that we will set later and consider the value of p_ℓ for $\ell \leq L$. The following lemma gives lower and upper bounds for p_ℓ .

Lemma 6.6. *Assume $\lambda = \Omega(1)$ and fix L . Let $\xi > 0$ be an arbitrarily small constant. We can set $\epsilon = \epsilon(\lambda, L, \xi)$ sufficiently small in the definition of I_L so that for all $\ell \leq L$ we have*

$$\frac{\lambda^\ell}{\ell!} (1 - \xi - o(1)) \leq p_\ell \leq \frac{\lambda^\ell}{\ell!} + o(1) \quad (20)$$

Proof. Let \mathcal{J}_ℓ be the set of all sets of ℓ distinct elements from I_S , that is, $J \in \mathcal{J}_\ell$ has the form $J = \{(i_1, j_2), (i_2, j_2), \dots, (i_\ell, j_\ell)\}$, where each (i_k, j_k) corresponds to a distinct entry from I_S . Now let $\mathcal{B}_2(J)$, for $J \in \mathcal{J}_\ell$, be the set of all possible ways we can choose one double edge from each element of J . That is, if $J = \{(i_1, j_2), (i_2, j_2), \dots, (i_\ell, j_\ell)\}$, then a typical element from $B \in \mathcal{B}_2(J)$ has the form $B = \{d_1, d_2, \dots, d_\ell\}$, where d_k is a double edge corresponding to the entry (i_k, j_k) , for $1 \leq k \leq \ell$. Recall that in the summation in (20), we obtain $\mathbf{P}(\mathcal{M}(D) \cap \mathcal{P}_3^c \mid W_L = 0) = 0$ for all $D \in \mathcal{D}_\ell$ that contain two or more distinct double edges that share a token. Therefore, we obtain the following equality

$$p_\ell = \sum_{D \in \mathcal{D}_\ell} \mathbf{P}(\mathcal{M}(D) \cap \mathcal{P}_3^c \mid W_L = 0) = \sum_{J \in \mathcal{J}_\ell} \sum_{B \in \mathcal{B}_2(J)} \mathbf{P}(\mathcal{M}(B) \cap \mathcal{P}_3^c \mid W_L = 0), \quad (21)$$

where the last term translates to

$$\mathbf{P}(\mathcal{M}(B) \cap \mathcal{P}_3^c \mid W_L = 0) = \mathbf{P}(\mathcal{P}_3^c \mid \mathcal{M}(B), W_L = 0) \mathbf{P}(\mathcal{M}(B) \mid W_L = 0). \quad (22)$$

We start with the term $\mathbf{P}(\mathcal{P}_3^c \mid \mathcal{M}(B), W_L = 0) = 1 - \mathbf{P}(\mathcal{P}_3 \mid \mathcal{M}(B), W_L = 0)$. For each $(i, j) \in J$, B contains at least one double edge for the entry (i, j) . Therefore, the probability that \mathcal{P}_3 happens can be upper bounded by the probability that another edge corresponding to some entry in J is matched plus the probability that 3 edges for some entry not in J are matched, that is, $\mathbf{P}(\mathcal{P}_3 \mid \mathcal{M}(B), W_L = 0)$ is at most

$$\sum_{(i,j) \in J} \sum_{B' \in \mathcal{B}_1(i,j) \cup B} \mathbf{P}(\mathcal{M}(B') \mid \mathcal{M}(B), W_L = 0) + \sum_{(i,j) \notin J} \sum_{B' \in \mathcal{B}_3(i,j)} \mathbf{P}(\mathcal{M}(B') \mid \mathcal{M}(B), W_L = 0),$$

where the notation $\cup B$ represents the union of all elements in $B = \{d_1, d_2, \dots, d_\ell\}$ (recall that the d_i 's are double edges). Note that $\cup B$ contains 2ℓ edges. Therefore, for $\ell = O(1)$ we can use Lemma 6.3 for the first term and a derivation similar to the proof of Lemma 6.5 for the second term to obtain

$$\begin{aligned} & \mathbf{P}(\mathcal{P}_3 \mid \mathcal{M}(B), W_L = 0) \\ & \leq \sum_{(i,j) \in J} \frac{(r_i - 2)(c_j - 2)}{N - 2\ell} (1 + o(1)) + \sum_{(i,j) \notin J} \frac{r_i^{\frac{3}{2}} c_j^{\frac{3}{2}}}{3!(N - 2\ell)^{\frac{3}{2}}} (1 + o(1)) \end{aligned} \quad (23)$$

$$\leq \epsilon(L + \lambda/3) + o(1), \quad (24)$$

uniformly in J and B .

Now we turn to the term $\sum_{J \in \mathcal{J}_\ell} \sum_{B \in \mathcal{B}_2(J)} \mathbf{P}(\mathcal{M}(B) \mid W_L = 0)$, which corresponds to an upper bound for the right hand side of (21). Our goal is to write this term recursively in ℓ . First, notice that the case $\ell = 1$ reduces to

$$\sum_{J \in \mathcal{J}_1} \sum_{B \in \mathcal{B}_2(J)} \mathbf{P}(\mathcal{M}(B) \mid W_L = 0) = \sum_{(i,j) \in I_S} \frac{r_i^2 c_j^2}{2N^2} (1 + o(1)) = \lambda + o(1),$$

where $\lambda = \sum_{(i,j) \in I_S} r_i^2 c_j^2 / (2N^2)$ as defined in Lemma 6.2.

Now, consider a fixed $J \in \mathcal{J}_\ell$. Note that we can write $J = J' \cup (i, j)$ where $J' \in \mathcal{J}_{\ell-1}$ and $(i, j) \in I_S \setminus J'$. Note also that there are ℓ possible such values for $J' \subset J$. For a fixed J' , let $\eta_i = \eta_i(J')$ be the number of elements in J' corresponding to an entry in row i , that is, $\eta_i = |\{(k, k') \in J' : k = i\}|$. Likewise, let $\eta'_j = \eta'_j(J')$ be the number of elements in J' corresponding to an entry in column j . Therefore, given $B' \in \mathcal{B}_2(J')$, the sum of row i becomes $r_i - 2\eta_i$ and the sum of column j becomes $c_j - 2\eta'_j$. For $\ell = O(1)$, we can apply Lemma 6.3 to derive the following equality

$$\begin{aligned} & \sum_{J \in \mathcal{J}_\ell} \sum_{B \in \mathcal{B}_2(J)} \Pr[\mathcal{M}(B) \mid W_L = 0] \\ &= \frac{1}{\ell} \sum_{J' \in \mathcal{J}_{\ell-1}} \sum_{B' \in \mathcal{B}_2(J')} \mathbf{P}(\mathcal{M}(B') \mid W_L = 0) \sum_{(i,j) \in I_S \setminus J'} \sum_{B'' \in \mathcal{B}_2(i,j)} \mathbf{P}(\mathcal{M}(B'') \mid \mathcal{M}(B'), W_L = 0) \\ &= \frac{1}{\ell} \sum_{J' \in \mathcal{J}_{\ell-1}} \sum_{B' \in \mathcal{B}_2(J')} \mathbf{P}(\mathcal{M}(B') \mid W_L = 0) (1 + o(1)) \sum_{(i,j) \in I_S \setminus J'} \frac{(r_i - 2\eta_i)^2 (c_j - 2\eta'_j)^2}{2(N - 2\ell + 2)^2}. \end{aligned} \quad (25)$$

Note that only pairs (i, j) with $r_i, c_j \geq 2$ count for the last sum in (25). So in what follow we assume that $r_i, c_j \geq 2$. Note that $\sum_i \eta_i = \sum_j \eta'_j = \ell - 1$, and letting

$$X = \sum_{(i,j) \in I_S \setminus J'} \frac{r_i^2 c_j^2}{2(N - 2\ell + 2)^2} - \sum_{(i,j) \in I_S \setminus J'} \frac{(r_i - 2\eta_i)^2 (c_j - 2\eta'_j)^2}{2(N - 2\ell + 2)^2}, \quad (26)$$

we have

$$\sum_{(i,j) \in I_S \setminus J'} \frac{(r_i - 2\eta_i)^2 (c_j - 2\eta'_j)^2}{2(N - 2\ell + 2)^2} = \sum_{(i,j) \in I_S} \frac{r_i^2 c_j^2}{2(N - 2\ell + 2)^2} - \sum_{(i,j) \in J'} \frac{r_i^2 c_j^2}{2(N - 2\ell + 2)^2} - X.$$

If we apply the condition $(r_i - 1)(c_j - 1) < \epsilon N$ and use the inequality $y \leq 2(y - 1)$ valid for all $y \geq 2$, we can write the second term in the right hand side above as

$$\sum_{(i,j) \in J'} \frac{r_i^2 c_j^2}{2(N - 2\ell + 2)^2} \leq \sum_{(i,j) \in J'} \frac{2(r_i - 1)^2 (c_j - 1)^2}{N^2} (1 + O(1/N)) \leq 2\epsilon^2 (\ell - 1) + O(1/N).$$

If we expand X in (26) we obtain $X = X_1 - X_2$, where

$$X_1 = \sum_{(i,j) \in I_S \setminus J'} \frac{2\eta'_j (2c_j - 1) r_i^2 + 2\eta_i (2r_i - 1) c_j^2 + 8\eta_i \eta_j'^2 (2r_i - 1) + 8\eta_i^2 \eta'_j (2c_j - 1)}{2(N - 2\ell + 2)^2}$$

and

$$X_2 = \sum_{(i,j) \in I_S \setminus J'} \frac{4\eta_j'^2 r_i^2 + 4\eta_i^2 c_j^2 + 4\eta_i \eta'_j (2r_i - 1)(2c_j - 1) + 16\eta_i^2 \eta_j'^2}{2(N - 2\ell + 2)^2}.$$

Since $r_i, c_j \geq 2$, we have that $(r_i - 1)(2c_j - 1)$ and $(2r_i - 1)(c_j - 1)$ can both be upper bounded by $3(r_i - 1)(c_j - 1)$. Then, applying $(r_i - 1)(c_j - 1) < \epsilon N$ to the first two terms of X_1 and using the fact that $\sum_{j=1}^n \eta_i \eta_j'^2 \leq \ell^2 \sum_{j=1}^n \eta_j' \leq \ell^3$ and similarly $\sum_{i=1}^m \eta_i^2 \eta_j' \leq \ell^2 \sum_{i=1}^m \eta_i \leq \ell^3$ for the last two terms of X_1 , we have $X_1 \leq 6\epsilon\ell + O(1/N)$, and using the simple fact $X_2 \geq 0$ we get

$$\sum_{(i,j) \in I_S \setminus J'} \frac{(r_i - 2\eta_i)^2 (c_j - 2\eta_j')^2}{2(N - 2\ell + 2)^2} \geq \lambda - 2\epsilon^2\ell - 6\epsilon\ell - O(1/N). \quad (27)$$

Putting (25) and (27) together, and iterating this procedure ℓ times for $\ell = O(1)$ we obtain

$$\frac{(\lambda - 2\epsilon^2L - 6\epsilon L)^\ell}{\ell!} - o(1) \leq \sum_{J \in \mathcal{J}_\ell} \sum_{B \in \mathcal{B}_2(J)} \mathbf{P}(\mathcal{M}(B) \mid W_L = 0) \leq \frac{\lambda^\ell}{\ell!} + o(1). \quad (28)$$

Since $\lambda = \Omega(1)$ and ϵ is sufficiently small, we can find a constant c such that

$$\frac{(\lambda - 2\epsilon^2L - 6\epsilon L)^\ell}{\ell!} \geq \frac{\lambda^\ell}{\ell!} (1 - c\epsilon L\ell) \geq \frac{\lambda^\ell}{\ell!} (1 - c\epsilon L^2).$$

Putting (22), (24), and (28) together, and plugging the result into (21), we obtain

$$\frac{\lambda^\ell}{\ell!} (1 - c\epsilon L^2 - \epsilon(L + \lambda/3) - o(1)) \leq p_\ell \leq \frac{\lambda^\ell}{\ell!} + o(1).$$

This concludes the proof of the lemma since we can set ϵ sufficiently small so that $\xi \leq \epsilon(cL^2 + L + \lambda/3)$. \square

For some fixed constant L , we can use Bonferroni inequality to obtain a lower bound for $1 - \mathbf{P}(\mathcal{P}_2 \cap \mathcal{P}_3^c \mid W_L = 0) = 1 - \sum_{\ell \geq 1} (-1)^{\ell+1} p_\ell$ via

$$1 - \sum_{\ell=1}^L (-1)^{\ell+1} p_\ell \geq 1 - \sum_{\ell=1}^L (-1)^{\ell+1} \frac{\lambda^\ell}{\ell!} - \sum_{\ell=1}^L \frac{\lambda^\ell \xi}{\ell!} - o(1) \geq e^{-\lambda} - \lambda^L/L! - e^\lambda \xi - o(1),$$

where ξ is obtained from Lemma 6.6.

Recall that (11) implies $\lambda = O(1)$. For an arbitrarily small constant $\delta > 0$ independent of N (as long as N is sufficiently large) we can set L large enough so that $\lambda^L/L! \leq \delta/4$. We can also set ξ in Lemma 6.6 so that $e^\lambda \xi \leq \delta/4$. Now, having fixed L and ξ , we can set ϵ small enough so that Lemma 6.6 can be applied and in addition we have $\lambda\epsilon/3 \leq \delta/4$. Then we put together (15) and (17), and use Lemma 6.5 and (19) to obtain for large enough N

$$\begin{aligned} \mathbf{P}(Z_S = 0 \mid W_L = 0) &\geq 1 - \sum_{\ell=1}^L (-1)^{\ell+1} p_\ell - \lambda\epsilon/3 - o(1) \\ &\geq e^{-\lambda} - \frac{3\delta}{4} - o(1) \\ &\geq e^{-\lambda} - \delta, \end{aligned}$$

which concludes the proof of Lemma 6.2. \square

7 Proof of Proposition 4.3

Recall the definition $\kappa = \min\{i \geq 1: r_i = o(N)\}$ from Theorem 1.1 and let $\kappa' = \min\{j \geq 1: \limsup_{N \rightarrow \infty} c_j \leq 1\}$. Throughout this section, we assume that $\sum_{(i,j) \in I} r_i^2 c_j^2 = O(N^2)$, $c_1 \geq 2$, and $r_1 = \Omega(N)$. These conditions immediately imply that $\kappa, \kappa' \geq 2$ and $c_1, \kappa' = O(1)$. We define κ and κ' since it suffices to look at the entries in $[1, \kappa - 1] \times [1, \kappa' - 1]$.

We prove Proposition 4.3 in two parts. In the first part, which we treat in the following lemma, we show that condition $\sum_{i=\kappa}^{\infty} r_i = \Omega(N)$ is sufficient since it implies that all entries in $[1, \kappa - 1] \times [1, \kappa' - 1]$ are zero with constant probability.

Lemma 7.1. *Assume $\sum_{(i,j) \in I} r_i^2 c_j^2 = O(N^2)$ and $r_1 = \Omega(N)$. Then, if $\sum_{i=\kappa}^{\infty} r_i = \Omega(N)$, we obtain $\Pr[T \in \Omega_{\mathbf{r}, \mathbf{c}}] = \Omega(1)$.*

Proof. We first show that with constant probability, $T_{i,j} = 0$ for each $i \in [1, \kappa - 1]$ and $j \in [1, \kappa' - 1]$. Let $X = \sum_{i=1}^{\kappa-1} r_i$. Note that by assumption $N - X = \Omega(N)$. Therefore, there exists a constant $\alpha \in (0, 1)$ such that

$$\mathbf{P} \left(\bigcap_{i=1}^{\kappa-1} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} = 0\} \right) = \prod_{j=1}^{\kappa'-1} \frac{(N - X - \sum_{j'=1}^{j-1} c_{j'})^{c_j}}{(N - \sum_{j'=1}^{j-1} c_{j'})^{c_j}} \geq \alpha, \quad (29)$$

since $c_j = O(1)$ and $\sum_{j'=1}^{j-1} c_{j'} \leq c_1 \kappa' = O(1)$ for $j \leq \kappa' - 1$. Now, for all $j \geq \kappa'$ and sufficiently large N , we have $c_j \leq 1$, and all the entries (i, j) for which $i \leq \kappa - 1$ and $j \geq \kappa'$ are binary with probability 1. We can then conclude that the probability that all the entries for rows $i \leq \kappa - 1$ are binary is at least α . Once we have sampled all the rows for which $i \leq \kappa - 1$, we can then remove these rows and obtain new sequences \mathbf{r}' and \mathbf{c}' for which $\max_i r'_i = o(N)$ (by the definition of κ) and $\sum_i r'_i = \sum_{i=\kappa}^{\infty} r_i = \Omega(N)$. Note that \mathbf{r}' and \mathbf{c}' fall into the setting of Proposition 4.2. Therefore, letting T' be a table generated from \mathbf{r}' and \mathbf{c}' , we obtain

$$\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) \geq \alpha \mathbf{P}(T' \in \Omega_{\mathbf{r}', \mathbf{c}'}) = \Omega(1), \quad (30)$$

from (29) and Proposition 4.2. \square

Now we assume that $\sum_{i=\kappa}^{\infty} r_i$ is not $\Omega(N)$ (i.e., $\liminf_{N \rightarrow \infty} \sum_{i=\kappa}^{\infty} r_i / N = 0$) and see that $\lim_{N \rightarrow \infty} c_1 < \kappa$ is both necessary and sufficient. The following lemma, which concludes the proof of Proposition 4.3, is the only place where we use the technical assumptions (A1) and (A2) from Section 1. A discussion about the case where the assumptions do not hold is given at the end of this section.

Lemma 7.2. *Assume that the following three conditions hold: $\sum_{(i,j) \in I} r_i^2 c_j^2 = O(N^2)$, $r_1 = \Omega(N)$, and $\liminf_{N \rightarrow \infty} \sum_{i=\kappa}^{\infty} r_i / N = 0$. Then, $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ if and only if $\lim_{N \rightarrow \infty} c_1 < \kappa$.*

Proof. We first show that $\lim_{N \rightarrow \infty} c_1 < \kappa$ is necessary when $\liminf_{N \rightarrow \infty} \sum_{i=\kappa}^{\infty} r_i / N = 0$. We assume that κ is finite and explain the case when $\kappa = \infty$ in Remark 7.3. We have the following result for the probability that all entries (i, j) with $i \geq \kappa$ and $j \leq \kappa' - 1$ are 0

$$\limsup_{N \rightarrow \infty} \mathbf{P} \left(\bigcap_{i=\kappa}^{\infty} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} = 0\} \right) = \limsup_{N \rightarrow \infty} \prod_{j=1}^{\kappa'-1} \frac{(\sum_{i=1}^{\kappa-1} r_i - \sum_{j'=1}^{j-1} c_{j'})^{c_j}}{(N - \sum_{j'=1}^{j-1} c_{j'})^{c_j}} = 1, \quad (31)$$

since $c_j = O(1)$ and $\sum_{j'=1}^{j-1} c_{j'} = O(1)$ for $j \leq \kappa' - 1$. When this happens, the entry $T_{i,1}$ is not zero only if r_i is not $o(N)$ (i.e., $i \leq \kappa - 1$). Since there are $\kappa - 1$ such rows, if $\lim_{N \rightarrow \infty} c_1 \geq \kappa$, then there will be a non-binary entry $(i, 1)$ with $i \leq \kappa - 1$ for a sufficiently large N . This gives that

$$\liminf_{N \rightarrow \infty} \mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) \leq 1 - \limsup_{N \rightarrow \infty} \mathbf{P}\left(\bigcap_{i=\kappa}^{\infty} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} = 0\}\right) = 0.$$

It remains to show that $\lim_{N \rightarrow \infty} c_1 < \kappa$ is sufficient. We first consider the case $\sum_{i=\kappa}^{\infty} r_i = o(N)$ and show that $\lim_{N \rightarrow \infty} c_1 < \kappa$ implies $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$. When $\sum_{i=\kappa}^{\infty} r_i = o(N)$, a derivation similar to (31) gives that $\mathbf{P}\left(\bigcap_{i=\kappa}^{\infty} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} = 0\}\right) = 1 - o(1)$; therefore when assigning a column $j \leq \kappa' - 1$, only the entries for rows $i \leq \kappa - 1$ will be non-zero with probability $1 - o(1)$. Recall that the technical assumption (A1) gives that $r_i = \Omega(N)$ for all $i \leq \kappa - 1$. Since $\sum_{j=1}^{\kappa'-1} c_j = O(1)$, for each type-2 token for a column $j \leq \kappa' - 1$, there is a constant α such that the probability that this token is matched to a type-1 token from a given row $i \leq \kappa - 1$ is at least α uniformly over all possible matchings for the other type-2 tokens from columns in $[1, \kappa - 1]$. Therefore,

$$\mathbf{P}\left(\bigcap_{i=1}^{\kappa-1} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} \leq 1\} \mid \bigcap_{i=\kappa}^{\infty} \bigcap_{j=1}^{\kappa'-1} \{T_{i,j} = 0\}\right) \geq \prod_{j=1}^{\kappa'-1} \binom{\kappa-1}{c_j} \alpha^{c_j} = \Omega(1). \quad (32)$$

We then obtain that with constant probability all the entries for columns in $[1, \kappa' - 1]$ are binary. For sufficiently large N , the remaining entries are all binary with probability 1 since for $j \geq \kappa'$ we have $c_j = 1$. The case $\sum_{i=\kappa}^{\infty} r_i$ being neither $o(N)$ nor $\Omega(N)$ can be handled by Lemma 4.5 with $f_N(\mathbf{r}, \mathbf{c}) = \sum_{i=\kappa}^{\infty} r_i/N$ and \mathcal{Z} being the space of sequences satisfying $\sum_{(i,j) \in I} r_i^2 c_j^2 = O(N^2)$, $r_1 = \Omega(N)$ and $\lim_{N \rightarrow \infty} c_1 < \kappa'$. This completes the proof. \square

Remark 7.3. We now briefly explain how to handle the case when $\kappa = \infty$. For this, we take $\kappa(N) = \min\{\kappa, m(N)\}$. (Recall that $m(N)$ is the number of rows.) Clearly, $\kappa = \lim_{N \rightarrow \infty} \kappa(N)$. Also, when $\kappa = \infty$ and $\sum_{(i,j) \in I} r_i^2 c_j^2 = O(N^2)$, we have $c_1 = O(1)$. Therefore, there exists a value of N sufficiently large such that $c_1 < \kappa(N)$. The proof above then follows using this choice of N if we replace κ by $\kappa(N)$.

The proof of Proposition 4.3 follows from Lemmas 7.1 and 7.2 and Remark 7.3. \square

The case where assumptions (A1) and (A2) do not hold

In the remainder of this section we consider the case when assumptions (A1) and (A2) do not hold. We restrict our discussion here to the case when κ is finite; see Remark 7.3 for the case when $\kappa = \infty$.

First, consider the case where (A2) does not hold (i.e., $\liminf_{N \rightarrow \infty} c_1 < \limsup_{N \rightarrow \infty} c_1$). The assumptions in Lemma 7.2 imply that $c_1 = O(1)$. However, if $\liminf_{N \rightarrow \infty} c_1 < \kappa \leq \limsup_{N \rightarrow \infty} c_1$, then the sequence $(\mathbf{r}, \mathbf{c})_N$ could be such that for every subsequence $(\mathbf{r}', \mathbf{c}')_N$ for which $\sum_{i=\kappa}^{\infty} r'_i = \Omega(N)$ we have $\limsup_{N \rightarrow \infty} c_1 \geq \kappa$, and for every subsequence $(\mathbf{r}', \mathbf{c}')_N$ for which $\liminf_{N \rightarrow \infty} \sum_{i=\kappa}^{\infty} r'_i/N = 0$ we have $\limsup_{N \rightarrow \infty} c_1 < \kappa$. This correlation between the asymptotic behavior of $\sum_{i=\kappa}^{\infty} r_i/N$ and c_1 yields $\mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = \Omega(1)$ even with $\limsup_{N \rightarrow \infty} c_1 \geq \kappa$.

Now we consider the case where (A1) does not hold (i.e., there exists a row i for which r_i is neither $o(N)$ nor $\Omega(N)$). We can assume that (A2) holds, and we consider the case in the proof of Lemma 7.2 where $\sum_{i=\kappa}^{\infty} r_i = o(N)$. In this case, any type-2 token for a column $j \leq \kappa' - 1$ will be matched to a

type-1 token for a row $i \leq \kappa - 1$ with high probability. Thus, the table will be binary with constant probability if the number of rows having sum $\Omega(N)$ is at least as large as c_1 . Recall that the number of rows whose sum is not $o(N)$ is $\kappa - 1$. However, since (A1) does not hold, there are subsequences of (\mathbf{r}, \mathbf{c}) for which the number of rows whose sum is not $o(N)$ is smaller than $\kappa - 1$, and may indeed be smaller than c_1 . For these subsequences, the output table will be non-binary with probability $1 - o(1)$, and we would have $\liminf_{N \rightarrow \infty} \mathbf{P}(T \in \Omega_{\mathbf{r}, \mathbf{c}}) = 0$ even with $\lim_{N \rightarrow \infty} c_1 < \kappa$.

8 Conclusions

We have characterized the input sequences for which the configuration model is suitable for uniformly sampling and counting binary contingency tables in optimal time (i.e., linear as a function of the number N of unit entries in the table). Surprisingly, given known results for the case of symmetric tables, having a bounded number of expected double edges in the table is just a necessary condition for optimality but not sufficient. It turns out that the full characterization for optimality in the non-symmetric case relates to the behavior of very big rows (i.e. rows of size $\Omega(N)$). Allowing such type of growth introduces significant qualitative differences between symmetric and non-symmetric tables. In turn, such differences give rise to technical challenges that are not present in the symmetric case. Our results also have important practical implications in applied settings that demand the development of easy-to-implement and fast algorithms for uniform generation of binary contingency tables.

We conclude by mentioning two open problems. Since, as we mentioned in Section 1, there is no need to employ a complicated sequential importance sampling procedure when the conditions above hold, it is interesting to know whether one can construct more specialized importance sampling procedures to obtain low complexity polynomial algorithm when the conditions in Theorem 1.1 are not satisfied. Another open problem consists in counting the number of (not necessarily binary) contingency tables. In particular, it would be interesting to know whether a necessary and sufficient condition like the one obtained in this paper can be derived for that case. We remark that for general contingency tables, the configuration model does not generate a table uniformly at random, which makes the problem more challenging.

Acknowledgments

We are thankful to Alistair Sinclair for bringing reference [14] to our attention and for helpful discussions on contingency tables.

References

- [1] A. Barvinok. On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries. *Advances in Mathematics*, 224(1):316–339, 2010.
- [2] M. Bayati, J.H. Kim, and A. Saberi. A sequential algorithm for generating random graphs. In *Proceedings of 11th Workshop on Randomization and Computation (RANDOM)*, pages 326–340, 2007.
- [3] J. Besag and P. Clifford. Generalized Monte Carlo significance tests. *Biometrika*, 76(4):633–642, 1989.

- [4] I. Bezáková, N. Bhatnagar, and E. Vigoda. Sampling binary contingency tables with a greedy start. *Random Structures and Algorithms*, 30(1-2):168–205, 2007.
- [5] I. Bezáková, A. Sinclair, and D. Štefankovič. Negative examples for sequential importance sampling of binary contingency tables. In *Proceedings of the European Symposium on Algorithms*, pages 136–147, 2006.
- [6] I. Bezáková, D. Štefankovič, V. Vazirani, and E. Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM Journal of Computing*, 37(5):1429–1454, 2008.
- [7] J. Blanchet. Efficient importance sampling for binary contingency tables. *Annals of Applied Probability*, 19(3):949–982, 2009.
- [8] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316, 1980.
- [9] E.R. Canfield, C. Greenhill, and B.D. McKay. Asymptotic enumeration of dense 0-1 matrices with specified line sums. *Journal of Combinatorial Theory, Series A*, 115:32–66, 2008.
- [10] Y. Chen, P. Diaconis, S.P. Holmes, and J.S. Liu. Sequential Monte Carlo methods for statistical analysis of tables. *Journal of the American Statistical Association*, 100(469):109–120, 2005.
- [11] M. Dyer, R. Kannan, and J. Mount. Sampling contingency tables. *Random Structures and Algorithms*, 10(4):487–506, 1997.
- [12] C. Greenhill, F.B. Holt, and N. Wormald. Expansion properties of a random regular graph after random vertex deletions. *European Journal of Combinatorics*, 29:1139–1150, 2008.
- [13] C. Greenhill, B.D. McKay, and X. Wang. Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums. *Journal of Combinatorial Theory (Series A)*, 113:291–324, 2006.
- [14] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, 18(1-2):205–225, 2009.
- [15] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. John Wiley & Sons, 2000.
- [16] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the Association for Computing Machinery*, 51(4):671–697, 2004.
- [17] B.D. McKay. Asymptotics for 0-1 matrices with prescribed line sums. In *Enumeration and Design*, pages 225–238. Academic Press, Canada, 1984.
- [18] B.D. McKay. Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combinatoria*, 19(A):15–25, 1985.
- [19] M. Mitzenmacher and E. Upfal. *Probability and Computing: randomized algorithms and probabilistic analysis*. Cambridge University Press, Cambridge, UK, 2005.
- [20] P.E. O’Neil. Asymptotics and random matrices with row-sum and column sum-restrictions. *Bulletin of the American Mathematical Society*, 75(6):1276–1282, 1969.
- [21] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

- [22] L. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.