We consider the multiparameter random simplicial complex as a higher dimensional extension of the classical Erdős–Rényi graph. We investigate appearance of “unusual” topological structures in the complex from the point of view of large deviations. We first study upper tail large deviation probabilities for subcomplex counts, deriving the order of magnitude of such probabilities at the logarithmic scale precision. The obtained results are then applied to analyze large deviations for the number of simplices of the multiparameter simplicial complexes. Finally, these results are also used to deduce large deviation estimates for Betti numbers of the complex in the critical dimension.

**KEYWORDS**

Betti number, large deviations, multiparameter simplicial complex

1 **INTRODUCTION**

One can view a simplicial complex as a network with connections potentially involving more than 2 vertices at a time. Given a set $V$ of vertices, an undirected graph allows for the existence of only edges of the type $(v_1, v_2)$ for $v_1, v_2$ in $V$, while potentially higher-dimensional “edges” would have the form $(v_1, \ldots, v_k)$ with $k \geq 2$ for $v_1, \ldots, v_k$ in $V$. If $k > 2$, this is a hyperedge, which is not allowed in a graph but is allowed in a hypergraph. A simplicial complex is a special kind of a hypergraph, in which a subset of hyperedge is itself a hyperedge. That is, if $(v_1, \ldots, v_k)$ with $k > 2$ is a hyperedge, then so is the collection of $k - 1$ vertices obtained by removing from $(v_1, \ldots, v_k)$ any one of its $k$ vertices. When describing a simplicial complex, one typically says that $(v_1, \ldots, v_k)$ forms a $(k - 1)$-dimensional simplex (henceforth we call it a $(k - 1)$-simplex), and not a hyperedge.

In a random simplicial complex, the simplices of different dimensions are added according to a randomized rule. Some of the models of random simplicial complexes are extensions of the classical Erdős–Rényi random graph, in which potential edges between two vertices are formed with probability
$p$, independently of other potential edges. Possible rules of constructing a random simplicial complex include the flag complex (also known as the clique complex), in which a potential $k$-simplex is formed whenever a set of $k + 1$ vertices constitutes a clique in the Erdős–Rényi graph (see e.g., [14]). The Linial-Meshulam-Wallach complex of a fixed maximal dimension $k$ is a random simplicial complex, in which all of the $(k-1)$-simplices are present with probability 1, while the potential $k$-simplices are included with probability $p$, independently of other $k$-simplices ([17, 19]). The most general model in this direction is the Costa-Farber multiparameter simplicial complex, for which potential simplices are added inductively in their dimensions; for every $k = 1, 2, \ldots$, each potential $k$-simplex is included with probability $p_k$, independently of other simplices, only when all of its $(k-1)$-faces are present (see [7, 8]).

The randomness of the simplicial complexes induces randomness on the topological structure of the complex and its topological invariants, such as the Betti numbers and the Euler characteristic. The distributions of topological invariants have been a subject of recent interest for various models of random simplicial complexes. This includes the existence of a dominating dimension and central limit theorems for the Euler characteristic; see e.g., [11, 15, 16, 22]. Functional limit theorems for a dynamic version of the multiparameter model have been established in [21].

However, the previous work describes only the “usual” topological structure of random simplicial complexes, in the sense of the “average” behavior and likely deviations from the “average” behavior of the topological invariants. In contrast, the primary focus of this paper is in the situations when the topological structure of the complex is less usual, in the sense of a topological invariant being far away from the average. Such events are, by definition, rare but may have an oversized impact on the function of the network, and they are typically referred to as large deviations events. Such events are often related to situations when certain subcomplexes appear significantly more or significantly less than expected. Understanding the probabilities of such events is sometimes described as the upper (or lower) tail large deviations problem for a subcomplex count.

Within the context of the Erdős–Rényi random graphs, large deviation problems for subgraph counts have attracted much attention over the last decade in, among many others, [2, 4–6, 10, 18, 23]. In particular, [3] gives a comprehensive presentation, covering various large deviation problems for the random graph of different degrees of denseness. Moreover, [13] developed a general framework for the upper tail large deviation problems for subgraph counts in the random graph. The combinatorial part of this approach is based on a result in [1], and an extension to uniform subhypergraph counts in a random hypergraph setup is provided in [9].

The present work receives much inspiration from [13] and addresses the upper tail large deviation problems for subcomplex counts in the multiparameter random simplicial complex. Due to its high-dimensional topological structure, this problem is more involved than the analogous problems for random graphs. Although some of the results available for the Erdős–Rényi random graph have not yet been fully extended to the multiparameter random simplicial complex, the results we obtain are useful for understanding certain rare events in the latter complex. For instance, we are able to describe the upper tail large deviations for the Betti numbers at the dominating dimensions.

This paper is organized as follows. In Section 2, we present a formal definition of the multiparameter random simplicial complex. In Section 3, we address general upper tail large deviation problem for subcomplex counts. Section 4 specializes to the large deviation problem for the number of simplices of the multiparameter simplicial complexes, under the setup of [21]. Finally, Section 5 discusses the upper tail large deviations for Betti numbers at the dominating dimension.

Throughout the paper we use the notation $|A|$ for the cardinality of a set $A$. 
2 | THE MULTIPARAMETER RANDOM SIMPLICIAL COMPLEX AND THE SUBCOMPLEX COUNT PROBLEM

In this section we formally construct the multiparameter random simplicial complex introduced in [7, 8]. This complex is a model of an abstract simplicial complex on the alphabet \([n] = \{1, \ldots, n\}\), parametrized by \(p = p(n) = (p_1, \ldots, p_{n-1}) \in [0, 1]^{n-1}\). A simplex (or a face, or a word) in this complex is a nonempty collection of letters in the alphabet, and the dimension of a simplex is equal to the number of letters in the word minus 1. The \(i\)th skeleton of a complex is the subcomplex consisting of all faces of dimension \(i\) or less. The multiparameter random simplicial complex is built recursively, starting with \([n]\) as the 0th skeleton. For \(i = 1, \ldots, n-1\), once the \((i-1)\)st skeleton has been constructed, each of the potential \(i\)-simplices whose boundary is in that \((i-1)\)st skeleton, is added to the complex with probability \(p_i\), independently of other potential \(i\)-simplices. We denote the obtained random complex by \(K(n; p(n)) = K(n; p_1, \ldots, p_{n-1})\).

We are interested in the subcomplex counting problem for the multiparameter random simplicial complex. Given two simplicial complexes, \(F_1\) and \(F_2\), an ordered copy of \(F_1\) in \(F_2\) is an injective simplicial map from the vertex set of \(F_1\) to the vertex set of \(F_2\). In particular, this map has the property that the vertices of every simplex in \(F_1\) are mapped into the vertices of a simplex in \(F_2\) of the same dimension. Similarly, an unordered copy of \(F_1\) in \(F_2\) is a subcomplex \(F_3\) of \(F_2\), which is isomorphic to \(F_1\); that is, there is a bijective mapping between the vertex set of \(F_1\) and the vertex set of \(F_3\), such that a set of vertices forms a simplex in \(F_1\) if and only if the corresponding vertex set under this mapping forms a simplex in \(F_3\). If \(n_o(F_2, F_1)\) and \(n(F_2, F_1)\) are the numbers of ordered and unordered copies of \(F_1\) in \(F_2\) correspondingly, then it is clear that \(n_o(F_2, F_1)/n(F_2, F_1) = |\text{Aut}(F_1)|\), the number of automorphisms of \(F_1\) consisting of all permutations of the vertices of \(F_1\) preserving the complex.

Let \(F\) be a fixed simplicial complex of dimension \(k \geq 1\). For \(n > k\) and probabilities \(p(n)\), we denote by \(N_n(F)\) and \(N_{o,n}(F)\) the (random) numbers of unordered and ordered copies of \(F\) in the multiparameter simplicial complex \(K(n; p_1, \ldots, p_{n-1})\), correspondingly. Note that \(p_i\) with \(i > k\) do not affect \(N_n(F)\) and \(N_{o,n}(F)\), so we may assume that \(p_i = 0\) for \(i > k\) and simply write the complex as \(K(n; p_1, \ldots, p_k)\). If we denote for \(i = 0, 1, \ldots, k\),

\[
F_i = \text{the set of } i\text{-simplices in } F, \\
 s_i(F) = \text{the number of } i\text{-simplices in } F, 
\]

then

\[
\mu_{o,n}(F) := \mathbb{E}[N_{o,n}(F)] = (n)_{s_0(F)} \prod_{i=1}^k p_i^{s_i(F)}, \\
\mu_n(F) := \mathbb{E}[N_n(F)] = \mu_{o,n}(F)/|\text{Aut}(F)|, 
\]

where \((n)_{s_0(F)} := n(n-1) \cdots (n-s_0(F)+1)\).

As \(n \to \infty\), the copies of \(F\) can potentially be found in many (nearly) independent parts of the multiparameter simplicial complex \(K(n; p_1, \ldots, p_k)\), so one expects that for large \(n\) enough, \(N_n(F)\) and \(N_{o,n}(F)\) do not deviate “too much” from their corresponding means. Therefore, the upper tail large deviation probabilities

\[
P \left( N_{o,n}(F) \geq (1+\varepsilon)\mu_{o,n}(F) \right), \quad \varepsilon > 0, 
\]
and the lower tail large deviation probabilities

\[ P \left( N_{0,n}(F) \leq (1 - \varepsilon)\mu_{0,n}(F) \right), \quad 0 < \varepsilon < 1, \tag{2.4} \]

are expected to be exponentially small for large \( n \) enough. Our subject of interest is to investigate exactly how small these probabilities are. In this article, we focus only on the upper tail large deviations in (2.3). An analysis of the lower tail large deviations in (2.4) is postponed to a future publication.

### 3 Upper Tail Large Deviations

Our approach to understanding the upper tail large deviations for subcomplex counts is inspired by [13]. If \( G \) is a fixed simplicial complex of dimension \( k \geq 1 \), we denote by \( N(m_0, m_1, \ldots, m_k; G) \) the maximum of \( n(F, G) \) taken over all simplicial complexes \( F \) with \( s_i(F) \leq m_i, \: i = 0, 1, \ldots, k \). Clearly, \( N(m_0, m_1, \ldots, m_k; G) = 0 \) unless \( s_i(G) \leq m_i, \: i = 0, 1, \ldots, k \).

The number \( N(m_0, m_1, \ldots, m_k; G) \) is often referred to as the extremal parameter and is related to a certain linear optimization problem that we now describe. Using the notation in (2.1), we consider the linear program

\[ \max \sum_{v \in G_0} x_v \tag{3.1} \]

subject to

\[ 0 \leq \sum_{v \in \sigma_i} x_v \leq \log m_i, \: \sigma_i \in G_i, \: i = 0, 1, \ldots, k. \]

Denote by \( \gamma = \gamma(m_0, m_1, \ldots, m_k; G) \) the optimal value of this problem.

**Proposition 3.1.** Assume that \( s_i(G) \leq m_i, \: i = 0, 1, \ldots, k \). Then, there are finite positive constants \( c(G), C(G) \) that depend only on \( G \), such that

\[ c(G)e^{r(m_0, m_1, \ldots, m_k; G)} \leq N(m_0, m_1, \ldots, m_k; G) \leq C(G)e^{r(m_0, m_1, \ldots, m_k; G)}. \tag{3.2} \]

**Proof.** We first prove the lower bound in (3.2). Let \((x^*_v, v \in G_0)\) be an optimal solution to the linear program (3.1). We construct a simplicial complex \( F \) as follows. Choose a constant

\[ 0 < c < \min_{0 \leq j \leq k} \left[ \left( \frac{1}{s_j(G)(1 + s_j(G))} + 1 \right)^{1/(j+1)} - 1 \right], \tag{3.3} \]

which depends only on \( G \). We start with a family of disjoint sets \((V_v)_{v \in G_0}\), where \( V_v \) consists of \( n_v := \lceil ce^v \rceil \) points for each \( v \in G_0 \). Define the vertices of \( F \) to be the points in the union \( \bigcup_{v \in G_0} V_v \), that is, we take \( F_0 = \bigcup_{v \in G_0} V_v \). Next, for every \( j \in \{1, \ldots, k\} \) and distinct vertices \( v_1, \ldots, v_j+1 \in G_0 \), every point set \((w_1, \ldots, w_{j+1}) \in \prod_{i=1}^{j+1} V_{v_i} \) forms a \( j \)-simplex in \( F \) if and only if the vertices \( v_i, i = 1, \ldots, j+1 \), form a \( j \)-simplex in \( G \).

Now, we claim that

\[ s_j(F) \leq m_j, \quad j = 0, 1, \ldots, k. \]
Consider first \( s_0(F) \). If \( s_0(G) < m_0 \), then

\[
s_0(F) = \sum_{v \in G_0} \left\lfloor ce^{v_0} \right\rfloor \leq \sum_{v \in G_0} \left( ce^{v_0} + 1 \right) \leq s_0(G)(cm_0 + 1) \leq m_0.
\]

For the last inequality above, we have applied the bound (3.3) with \( j = 0 \). Suppose next that \( s_0(G) = m_0 \). In this case, \( e^{v_0} \leq s_0(G) \) for each \( v \in G_0 \) and \( c \leq 1/s_0(G) \) lead to

\[
s_0(F) = \sum_{v \in G_0} \left\lfloor ce^{v_0} \right\rfloor \leq s_0(G) = m_0.
\]

Similarly, for any \( j = 1, \ldots, k \), if \( s_j(G) < m_j \), then

\[
s_j(F) = \sum_{(v_1, \ldots, v_{j+1}) \in G_j} \prod_{i=1}^{j+1} \left\lfloor ce^{v_i} \right\rfloor \leq \sum_{(v_1, \ldots, v_{j+1}) \in G_j} \prod_{i=1}^{j+1} \left( ce^{v_i} + 1 \right)
= \sum_{(v_1, \ldots, v_{j+1}) \in G_j} \left( 1 + \sum_{i=1}^{j+1} \sum_{A \subseteq \{1, \ldots, j+1\}, |A|=i} c^i \exp \left\{ \sum_{v \in A} x_v \right\} \right)
\leq \sum_{(v_1, \ldots, v_{j+1}) \in G_j} \left( 1 + \sum_{i=1}^{j+1} c^i (j+1) m_j \right)
= s_j(G) \left( 1 + m_j ((1 + c^i j + 1) - 1) \right).
\]

By (3.3), the last expression is bounded by \( m_j \) as required. On the other hand, if \( s_j(G) = m_j \), then for \((v_1, \ldots, v_{j+1}) \in G_j \), we have \( e^{v_i} \leq s_j(G) \), \( i = 1, \ldots, j+1 \); so the bound \( c \leq 1/s_j(G) \) again leads to

\[
s_j(F) = \sum_{(v_1, \ldots, v_{j+1}) \in G_j} \prod_{i=1}^{j+1} \left\lfloor ce^{v_i} \right\rfloor \leq s_j(G) = m_j.
\]

For the simplicial complex \( F \) constructed above, the number of ordered copies of \( G \) in \( F \) is at least

\[
\prod_{v \in G_0} n_v \geq c^{s_0(G)} \exp \left\{ \sum_{v \in G_0} x_v \right\} = c^{s_0(G)} e^{v_0}.
\]

We thus conclude that

\[
N(m_0, m_1, \ldots, m_k; G) \geq n(F, G) \geq \frac{c^{s_0(G)} e^{v_0}}{|Aut(G)|} e^{v},
\]

establishing the lower bound in (3.2).

In order to prove the upper bound in (3.2), we start with the dual problem to the optimization problem (3.1). It is the linear program

\[
\min \left[ \sum_{v \in G_0} y_v \log m_0 + \sum_{i=1}^{k} \sum_{\sigma_i \in G_i} z_{i, \sigma_i} \log m_i \right] \tag{3.4}
\]

subject to
\[
y_v + \sum_{i=1}^{k} \sum_{\sigma_i \in G_i} z_{\sigma_i}^{(i)} \geq 1 \text{ for any } v \in G_0,
\]
\[
y_v \geq 0, \ z_{\sigma_i}^{(i)} \geq 0 \text{ for all } v \in G_0 \text{ and } \sigma_i \in G_i, \ i = 1, \ldots, k.
\]

The optimal value of the dual problem (3.4) equals \( \gamma = \gamma(m_0, m_1, \ldots, m_k; G) \); that is, it has the same optimal value as the original linear program in (3.1). For later use, let \((y_v^*), (z_{\sigma_i}^{(i)*})\) be an optimal solution to the dual problem in (3.4), so that

\[
\gamma = \sum_{v \in G_0} y_v^* \log m_0 + \sum_{i=1}^{k} \sum_{\sigma_i \in G_i} z_{\sigma_i}^{(i)*} \log m_i.
\] (3.5)

Now, let us fix a simplicial complex \( F \) of dimension \( k \), satisfying \( s_i(F) \leq m_i, \ i = 0, 1, \ldots, k \). Then, the upper bound in (3.2) is obtained as an immediate consequence of the bound

\[
n_o(F, G) \leq C(G)e^\gamma \quad (3.6)
\]

for some constant \( C(G) \) that does not depend on \( F \). For the proof of (3.6), we consider a partition

\[
F_0 = \bigcup_{v \in G_0} V_v \quad (3.7)
\]

of the vertex set of \( F \) into subsets indexed by the vertices of \( G \). Denote by \( \mathcal{H} = \mathcal{H}(F, G) \) the collection of all ordered copies of \( G \) in \( F \), so that \(|\mathcal{H}(F, G)| = n_o(F, G)\). Further, let \( \mathcal{W} = \mathcal{W}(F, G) \) be a subset of \( \mathcal{H} \), such that each \( v \in G_0 \) is mapped into one of the vertices in \( V_v \).

We create a random partition (3.7) as follows. To each vertex \( w \in F_0 \), assign a vertex \( U(w) \in G_0 \) uniformly randomly, and independently from each other. Now, let

\[
V_v = \{w \in F_0 : U(w) = v\}, \ v \in G_0
\]

(some sets \( V_v \) may be empty). In this setting, \( \mathcal{W} \) is a random subset of \( \mathcal{H} \), so that

\[
\mathbb{E}[|\mathcal{W}|] = \sum_{\varphi \in \mathcal{H}} \mathbb{P}(\varphi \in \mathcal{W}) = \sum_{\varphi \in \mathcal{H}} \mathbb{P}(U(\varphi(v)) = v \text{ for each } v \in G_0) = s_0(G)^{-s_0(G)}n_o(F, G).
\]

This indicates that there exists a nonrandom partition (3.7) of the vertex set of \( F \), for which

\[
|\mathcal{W}| \geq s_0(G)^{-s_0(G)}n_0(F, G). \quad (3.8)
\]

Fixing a collection \( \mathcal{W} \) that satisfies (3.8), we define \( \mathcal{W}_0 := \{H_0 : H \in \mathcal{W}\} \) to be a collection of vertex sets of complexes in \( \mathcal{W} \). Note that \( \mathcal{W}_0 \) can be viewed as a hypergraph on \( F_0 \), that is, a collection of distinct subsets (i.e., hyperedges) of a vertex set \( F_0 \). Clearly, \(|\mathcal{W}| \leq |\mathcal{W}_0| |\text{Aut}(G)| \). For a subset \( U \subseteq F_0 \), define the trace of \( \mathcal{W}_0 \) on \( U \) by

\[
\text{Tr}(\mathcal{W}_0, U) = \{H_0 \cap U : H_0 \in \mathcal{W}_0\}.
\]
We choose a large positive integer \( t \) and define

\[
l_0(v) = \left\lfloor tv^* \right\rfloor, \quad v \in G_0, \quad l_i(\sigma_i) = \left\lfloor tz_{\sigma_i}^{(i,v)} \right\rfloor, \quad \sigma_i \in G_i, \quad i = 1, \ldots, k,
\]

where \( y^*_v \) and \( z_{\sigma_i}^{(i,v)} \) are given in (3.5). Referring to the partition (3.7) of \( F_0 \) constructed above that satisfies (3.8), we now construct a family \( U_1, \ldots, U_s \) of subsets of \( F_0 \) as follows. Take each set \( V_v \) in (3.7) exactly \( l_0(v) \) times for every \( v \in G_0 \). Next, for each \( \sigma_i = \{v_1, \ldots, v_{j+1}\} \in G_i, \quad i = 1, \ldots, k \), take the union \( V_{v_1} \cup \ldots \cup V_{v_{j+1}} \) exactly \( l_i(\sigma_i) \) times. Finally, we enumerate these subsets as \( U_1, \ldots, U_s \), where

\[
s = \sum_{v \in G_0} l_0(v) + \sum_{i=1}^k \sum_{\sigma_i \in G_i} l_i(\sigma_i).
\]

Then, for every \( v \in G_0 \), each of the vertices in \( V_v \) appears exactly \( l_0(v) + \sum_{i=1}^k \sum_{\sigma_i \in G_i, \quad v \in \sigma_i} l_i(\sigma_i) \) times in the sets \( U_1, \ldots, U_s \). By the constraint of the dual problem (3.4),

\[
l_0(v) + \sum_{i=1}^k \sum_{\sigma_i \in G_i, \quad v \in \sigma_i} l_i(\sigma_i) \geq t \left( y^*_v + \sum_{i=1}^k \sum_{\sigma_i \in G_i, \quad v \in \sigma_i} z_{\sigma_i}^{(i,v)} \right) \geq t.
\]

This implies that every vertex in \( F \) appears at least \( t \) times in the sets \( U_1, \ldots, U_s \). By Lemma A.1 in the Appendix,

\[
|\mathcal{W}'|/|\text{Aut}(G)'| \leq |\mathcal{W}_0'| \leq \prod_{m=1}^s |\text{Tr}(\mathcal{W}_0, U_m)|
\]

\[
= \prod_{v \in G_0} \left| \text{Tr}(\mathcal{W}_0, V_v) \right| l_0(v) \prod_{j=1}^k \prod_{\sigma_i = \{v_1, \ldots, v_{j+1}\} \in G_i} \left| \text{Tr}(\mathcal{W}_0, V_{v_1} \cup \ldots \cup V_{v_{j+1}}) \right| l_i(\sigma_i). \tag{3.10}
\]

By the definition of \( \mathcal{W} \), each \( H_0 \in \mathcal{W}_0 \) has at most one element in \( V_v \) for each \( v \in G_0 \), so

\[
\left| \text{Tr}(\mathcal{W}_0, V_v) \right| \leq |V_v| \leq s_0(F) \leq m_0. \tag{3.11}
\]

Similarly, for each \( \sigma_j = \{v_1, \ldots, v_{j+1}\} \in G_j \), the intersection of \( H_0 \in \mathcal{W}_0 \) and \( \bigcup_{i=1}^{j+1} V_{v_i} \) either forms a \( j \)-simplex in \( F \) or becomes an empty set. Therefore,

\[
\left| \text{Tr}(\mathcal{W}_0, V_{v_1} \cup \ldots \cup V_{v_{j+1}}) \right| \leq s_j(F) \leq m_j, \quad j = 1, \ldots, k. \tag{3.12}
\]

Substituting the bounds in (3.11) and (3.12) back into (3.10), and recalling that each \( l_i \) depends on \( t \) via (3.9), we have, as \( t \to \infty \),

\[
|\mathcal{W}| \leq |\text{Aut}(G)| \prod_{v \in G_0} m_0 \prod_{j=1}^k \prod_{\sigma_i = \{v_1, \ldots, v_{j+1}\} \in G_i} m_j \tag{3.13}
\]

\[
= |\text{Aut}(G)| \prod_{v \in G_0} y^*_v \prod_{j=1}^k \prod_{\sigma_i = \{v_1, \ldots, v_{j+1}\} \in G_i} m_j = |\text{Aut}(G)| e^*,
\]
where the last equality follows from (3.5). Combining (3.8) and (3.13), we have

\[ n_0(F, G) \leq s_0(G^{\gamma(G)}|\text{Aut}(G)|e^{\gamma'}, \]

which establishes the bound (3.6), as desired.

The following lemma is a useful consequence of Proposition 3.1. It is a higher-dimensional version of Lemma 2.1 in [13].

**Lemma 3.2.** Let \( H \) be a \( k \)-dimensional subcomplex of \( G \). Then, there exists a constant \( C_H \in (1, \infty) \) such that if \( 0 \leq m_1 < m_2 \leq \frac{m_0^{k+1}}{s_k(G)} \),

\[ N(m_0, m_1s_1(G), \ldots, m_1s_k(G); H) \leq C_H \left( \frac{m_1}{m_2} \right)^{1/(k+1)} N(m_0, m_2s_1(G), \ldots, m_2s_k(G); H). \]

**Proof.** It is enough to consider the case \( m_1 > 0 \). By Proposition 3.1,

\[ N(m_0, m_1s_1(G), \ldots, m_1s_k(G); H) \leq C_1(H)e^{\gamma_1} \]

for some \( C_1(H) \in (1, \infty) \), where

\[ \gamma_1 = \max \sum_{v \in H_0} x_v \]

subject to

\[ 0 \leq x_v \leq \log m_0, \ v \in H_0, \]

\[ \sum_{v \in \sigma_i} x_v \leq \log s_i(G) + \log m_1, \ \sigma_i \in H_i, \ i = 1, \ldots, k. \]

Let \((x^{*}_v, v \in H_0)\) be an optimal solution for this problem. Consider all \( v \in H_0 \) that belong to a \( k \)-simplex in \( H \), and choose among them a vertex \( \tilde{v} \in H_0 \) with the smallest value of \( x^{*}_v \); that is,

\[ x^{*}_{\tilde{v}} = \min_{\sigma_i \in H_0, v \in \sigma_i} x^{*}_v. \]

By the feasibility of \((x^{*}_v)\), we have

\[ x^{*}_{\tilde{v}} \leq \frac{1}{k+1} \log s_k(G) + \frac{1}{k+1} \log m_1. \]

Define for \( v \in H_0 \),

\[ x^{**}_v = x^{*}_v \text{ for } v \neq \tilde{v}, \quad x^{**}_{\tilde{v}} = x^{*}_{\tilde{v}} + \frac{1}{k+1} \log \frac{m_2}{m_1}. \]

Then for \( \sigma_i \in H_i, \ i = 1, \ldots, k, \) we have

\[ \sum_{v \in \sigma_i} x^{**}_v \leq \sum_{v \in \sigma_i} x^{*}_v + \frac{1}{k+1} \log \frac{m_2}{m_1} \]

\[ \leq \log s_i(G) + \log m_1 + \frac{1}{k+1} \log \frac{m_2}{m_1} \leq \log s_i(G) + \log m_2. \]
while, by (3.15),
\[
x_v^{**} \leq \frac{1}{k+1} \log s_k(G) + \frac{1}{k+1} \log m_1 + \frac{1}{k+1} \log \frac{m_2}{m_1}
\]
\[
= \frac{1}{k+1} \log s_k(G) + \frac{1}{k+1} \log m_2 \leq \log m_0.
\]
We thus conclude that \((x_v^{**})\) is a feasible solution to the linear program
\[
\gamma_2 = \max \sum_{v \in H_0} x_v
\]
subject to
\[
0 \leq x_v \leq \log m_0, \ v \in H_0,
\]
\[
\sum_{v \in \sigma_i} x_v \leq \log s_i(G) + \log m_2, \ \sigma_i \in H_i, \ i = 1, \ldots, k.
\]
Therefore,
\[
\gamma_2 \geq \sum_{v \in H_0} x_v^{**} = \gamma_1 + \frac{1}{k+1} \log \frac{m_2}{m_1}.
\]
Appealing once again to Proposition 3.1, as well as (3.14), we have for some constant \(C_2(H) \in (0, 1)\),
\[
N (m_0, m_2s_1(G), \ldots, m_2s_k(G); H) \geq C_2(H) e^{\gamma_2} \geq C_2(H) \left( \frac{m_2}{m_1} \right)^{1/(k+1)} e^{\gamma_2}
\]
\[
\geq \frac{C_2(H)}{C_1(H)} \left( \frac{m_2}{m_1} \right)^{1/(k+1)} N (m_0, m_1s_1(G), \ldots, m_1s_k(G); H),
\]
as required.

For numbers \(0 \leq p_i \leq 1, \ i = 1, \ldots, k,\) and a simplicial complex \(G\) of dimension \(k\), denote
\[
\Psi_{G,n} := n^{s_0(G)} \prod_{i=1}^k p_i^{s_i(G)} \sim \mu_{G,n}(G), \ n \to \infty,
\]
and define
\[
M_{G,n}^*(p_1, \ldots, p_k) := \max \left\{ 1 \leq m \leq \left( \frac{n}{s_k(G)} \right)^{k+1} : N (n, m s_1(G), \ldots, m s_k(G); H) \leq \Psi_{H,n} \right\}
\]
for every nonempty subcomplex \(H\) of \(G\).

The following theorem is the main result of this section. It is an extension of Theorem 1.2 of [13] to the multiparameter random simplicial complexes. Its statement uses the notation \(K_{k,n}\) for the complete complex of dimension \(k\) of \(n\) vertices (i.e., a complex on \(n\) vertices containing all possible simplices of dimensions \(k\) and smaller).

**Theorem 3.3.** For every \(\epsilon > 0\), there exists \(C(\epsilon, G) > 0\) so that for all \(n \geq 1\),
\[
\mathbb{P} \left( N_{G,n} (G) \geq (1 + \epsilon) \mu_{G,n}(G) \right) \leq \exp \left\{ -C(\epsilon, G) M_{G,n}^*(p_1, \ldots, p_k) \right\}.
\]
Moreover, if \((1 + \varepsilon) \mu_{o,n}(G) \leq n_o(K_{k,n}, G)\), there exists \(B(\varepsilon, G) > 0\), such that for all \(n \geq 2k + 1\),

\[
\mathbb{P} \left( N_{o,n}(G) \geq \left(1 + \varepsilon\right) \mu_{o,n}(G) \right) \geq \frac{1}{4 \left( \prod_{j=1}^{k} p_j \right)^{B(\varepsilon, G)M_{G_o}(p_1, \ldots, p_k)}}. \tag{3.17}
\]

**Remark 3.4.** Theorem 3.3 identifies the order of magnitude of the upper tail large deviation probability at the logarithmic scale precision. The logarithmic order of magnitude differs between the upper bound (3.16) and the lower bound (3.17) by a factor of \(\log(\prod_{j=1}^{k} p_j)\). We note that under a common setup \(p_i = n^{-\alpha_i}\) for some \(\alpha_i \in [0, \infty)\), \(i \geq 1\), as in Section 4 below, the factor \(\log(\prod_{j=1}^{k} p_j)\) is logarithmic in \(n\). In contrast, the main term \(M_{G_o}(p_1, \ldots, p_k)\) typically grows polynomially, determining largely the order of magnitude of the large deviation probability.

Note also that the condition \((1 + \varepsilon) \mu_{o,n}(G) \leq n_o(K_{k,n}, G)\) is rarely restrictive. In fact, it is equivalent to \((1 + \varepsilon) \prod_{j=1}^{k} s_j(G) \leq 1\); this, however, trivially holds whenever \(p_j \to 0\) as \(n \to \infty\) for some \(j\).

**Proof.** We start with proving the upper bound in (3.16). Let \(G_1, \ldots, G_M\) be the ordered copies of \(G\) in \(K_{k,n}\) where \(M = (n)_{\text{an}(G)} = n(n-1) \cdots (n-s_0(G)+1)\). Clearly,

\[
N_{o,n}(G) = \sum_{j=1}^{M} I_j,
\]

where

\[
I_j = \mathbb{1} \{ G_j \text{ is a subcomplex of } K(n; p_1, \ldots, p_k) \}, \quad j = 1, \ldots, M.
\]

Therefore, for each \(m = 1, 2, \ldots\), we have

\[
\mathbb{E}[N_{o,n}(G)^m] = \sum_{1 \leq i_1, \ldots, i_m \leq M} \mathbb{E}[I_{i_1} \cdots I_{i_m}] = \sum_{1 \leq i_1, \ldots, i_m \leq M} \prod_{j=1}^{k} p_j^{s_j(G_{i_j} \cup \ldots \cup G_{i_m})}, \tag{3.18}
\]

with \(s_j(\cdot)\) as in (2.1). For a fixed \(i^{(m-1)} = (i_1, \ldots, i_{m-1}) \in \{1, \ldots, M\}^{m-1}\), denote \(F_{i^{(m-1)}} = G_{i_1} \cup \cdots \cup G_{i_{m-1}}\). Then (3.18) becomes

\[
\mathbb{E}[N_{o,n}(G)^m] = \sum_{i_j=1}^{k} \prod_{j=1}^{i_{m-1}} p_j^{s_j(F_{i^{(m-1)}} \cup \ldots \cup G_{i_{m-1}} \cap G_{i_m})}.
\]

For every fixed \(i^{(m-1)}\), if \(F_{i^{(m-1)}} \cap G_{i_m}\) contains at least one simplex of positive dimension, then this intersection is isomorphic to some subcomplex \(H\) of \(G\) of positive dimension. We thus conclude that

\[
\mathbb{E}[N_{o,n}(G)^m] \leq \sum_{i_j=1}^{k} \prod_{j=1}^{i_{m-1}} p_j^{s_j(F_{i^{(m-1)}})} \left[ M \prod_{j=1}^{k} p_j^{s_j(G_{i_j} \cup \cdots \cup G_{i_{m-1}} \cap G_{i_m})} + \sum_{i_m: F_{i^{(m-1)}} \cap G_{i_m} \neq \emptyset} \prod_{j=1}^{k} p_j^{s_j(G_{i_j} \cup \cdots \cup G_{i_{m-1}} \cap G_{i_m})} \right] = \sum_{i_j=1}^{k} \prod_{j=1}^{i_{m-1}} p_j^{s_j(F_{i^{(m-1)}})} \left[ \mu_{o,n}(G) + \sum_{H \subseteq G} \prod_{j=1}^{k} p_j^{s_j(G_{i_j} \cup \cdots \cup G_{i_{m-1}} \cap G_{i_m})} \left[ i: F_{i^{(m-1)}} \cap G_{i} \cong H \right] \right],
\]
where $\mu_{o,n}(G)$ is given in (2.2), and the sum $\sum_{H \subseteq G}$ is taken over all subcomplexes of $G$ with positive dimension, and $\cong$ means isomorphism between simplicial complexes. For every subcomplex $H$ of $G$, there are at most

$$N(s_0(F_{p^{(n-1)}}, s_1(F_{p^{(n-1)}}, \ldots, s_k(F_{p^{(n-1)}}); H) \leq N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H)$$

ways to choose an unordered copy of $H$ in $F_{p^{(n-1)}}$. To bound the number of ordered copies of $G$ in $K_{k,n}$ whose intersection with $F_{p^{(n-1)}}$ is isomorphic to $H$ (i.e., $\{|i: F_{p^{(n-1)}} \cap G \cong H\}$), notice that each choice of an unordered copy of $H$ in $F_{p^{(n-1)}}$ determines $s_0(H)$ vertices in the copy of $G$; thus, the number of ways to select the remaining vertices of the copy of $G$ is at most $(n - s_0(H))s_1(G) - s_0(H) = (n - s_0(H))s_1(G)/(n - s_0(H))$. Finally, the vertices of the copy of $G$ can be numbered in at most $s_0(G)!$ ways. From these observations, we conclude that

$$\left| \{i: F_{p^{(n-1)}} \cap G \cong H\} \right| \leq N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H)\frac{(n - s_0(G))}{(n - s_0(H))} s_0(G)!.$$ 

Now (2.2) gives us

$$\mathbb{E}[N_{o,n}(G)^m] \leq \sum_{i = 1}^{k} p_i^s(F_{p^{(n-1)}}) \times \mu_{o,n}(G) \left[ 1 + s_0(G)! \sum_{H \subseteq G} N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H) \right].$$

Using (3.18) with $m$ replaced by $m - 1$ results in

$$\mathbb{E}[N_{o,n}(G)^m] \leq \mathbb{E}[N_{o,n}(G)^{m-1}] \mu_{o,n}(G) \left[ 1 + s_0(G)! \sum_{H \subseteq G} N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H) \right].$$

By the monotonicity of the function $N$ in all of its arguments, an inductive argument gives us the bound

$$\mathbb{E}[N_{o,n}(G)^m] \leq \mu_{o,n}(G)^m \left[ 1 + s_0(G)! \sum_{H \subseteq G} N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H) \right]^{m-1}, \quad (3.19)$$

for every $m \geq 1$.

For $\theta \in (0, 1)$ to be determined in the sequel, we take $m = [\theta M_{G,n}^*(p_1, \ldots, p_k)] = [\theta M_{G,n}^*]$. Note that by Lemma 3.2,

$$N(n, (m-1)s_1(G), \ldots, (m-1)s_k(G); H) \leq N(n, \theta M_{G,n}^*s_1(G), \ldots, \theta M_{G,n}^*s_k(G); H) \leq C_H\theta^{1/(k+1)}N(n, M_{G,n}^*s_1(G), \ldots, M_{G,n}^*s_k(G); H) \leq \theta^{1/(k+1)}C_H \Psi_{H,n}.$$ 

Therefore, using (3.19) with $m = [\theta M_{G,n}^*]$ and Markov’s inequality, we obtain for $\varepsilon > 0$,

$$\mathbb{P}(N_{o,n}(G) \geq (1 + \varepsilon)\mu_{o,n}(G)) \leq (1 + \varepsilon)^{-\theta M_{G,n}^*} \left[ 1 + \theta^{1/(k+1)}s_0(G)! \sum_{H \subseteq G} C_H \Psi_{H,n} \mu_{o,n}(H) \right]^{\theta M_{G,n}^*}.$$
Denote these ordered copies of $G$ as being $\text{rooted ordered copies of } G$. Given a subcomplex $H$ of $G$, such that

\begin{align*}
n(H, i, \ldots, k, m) \geq 1, \ldots, k, m \geq 1, \text{ and a subcomplex } H \text{ of } G, \text{ such that} \end{align*}

\begin{align*}
N(n, a_1, \ldots, a_k m; H) \geq 2(1 + \epsilon) \Psi_{H,n}. \tag{3.20}
\end{align*}

This implies that there is a complex $F$ on at most $n$ nodes with $s_i(F) \leq a_i m, i = 1, \ldots, k$, such that

\begin{align*}
n(F, H) \geq 2(1 + \epsilon) \Psi_{H,n} \geq 2(1 + \epsilon) \mu_{o,n}(H). \tag{3.21}
\end{align*}

Our first goal is to show that under the assumption (3.20),

\begin{align*}
\mathbb{P}\left( N_{o,n}(G) \geq (1 + \epsilon) \mu_{o,n}(G) \right) \geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G) + a_j m}. \tag{3.22}
\end{align*}

Given a subcomplex $H$ of $G$ satisfying (3.20), we see that each of the ordered copies of $G$ in $K_{k,n}$ has a unique corresponding ordered copy of $H$ in $K_{k,n}$; if the latter is also in $F$, we refer to that ordered copy of $G$ as being $F$-rooted. Since there are $n_o(F, H) = |\text{Aut}(H)|n(F, H)$ ordered copies of $H$ in $F$, the number of $F$-rooted ordered copies of $G$ in $K_{k,n}$ is

\begin{align*}
J := n_o(F, H) (n - s_0(H)) s_0(G) - s_0(H).
\end{align*}

Denote these ordered copies of $G$ in $K_{k,n}$ by $G_1, \ldots, G_J$. Since $|\text{Aut}(H)| \geq 1$, it follows from (3.21) that

\begin{align*}
\prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)} \geq 2(1 + \epsilon) \mu_{o,n}(H) (n - s_0(H)) s_0(G) - s_0(H) \prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)} = 2(1 + \epsilon) \mu_{o,n}(G). \tag{3.23}
\end{align*}

Let $K_F(n; p_1, \ldots, p_k)$ be the multiparameter simplicial complex $K(n; p_1, \ldots, p_k)$ conditioned on $F \subseteq K(n; p_1, \ldots, p_k)$. For $i = 1, \ldots, J$, let $Z_i$ be the indicator function of the event that $G_i$ is a subcomplex of $K_F(n; p_1, \ldots, p_k)$. Then,

\begin{align*}
\mathbb{P}(Z_i = 1) = \prod_{j=1}^{k} p_j^{s_j(G \setminus F)} \geq \prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)}. \tag{3.24}
\end{align*}
Since the rightmost term in (3.24) is independent of $i$, the lower bound in Lemma A.2 of the Appendix gives us

$$
P\left(N_{o,n}(G) \geq \frac{1}{2} \prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)} \mid F \subseteq K(n; p_1, \ldots, p_k)\right)
\geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)} \geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G)}.
$$

Therefore, by (3.23),

$$
P\left(N_{o,n}(G) \geq (1 + \epsilon)\mu_{o,n}(G)\right) \geq P\left(N_{o,n}(G) \geq \frac{1}{2} \prod_{j=1}^{k} p_j^{s_j(G) - s_j(H)} \right)
\geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G)} \P\left(F \subseteq K(n; p_1, \ldots, p_k)\right)
= \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G)} \prod_{j=1}^{k} p_j^{s_j(F)} \geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G) + a_j m},
$$

establishing (3.22), as desired.

Now, we are ready to prove the lower bound in (3.17). Suppose first that

$$2(2(1 + \epsilon)C_H)^{k+1} M_{G,n}^* \leq s_k(G)^{-1} \left(\frac{n}{k+1}\right),
$$

where $C_H > 1$ is the constant in Lemma 3.2, that is increased, without loss of generality, to be the nearest positive integer. Then, $M_{G,n}^* < \lfloor s_k(G)^{-1} \left(\frac{n}{k+1}\right) \rfloor$, so there is a subcomplex $H$ of $G$ such that

$$N\left(n, (M_{G,n}^* + 1) s_1(G), \ldots, (M_{G,n}^* + 1) s_k(G); H\right) > \Psi_{H,n}.$$

Therefore, by Lemma 3.2,

$$2(1 + \epsilon)\Psi_{H,n} < 2(1 + \epsilon)N(n, (M_{G,n}^* + 1) s_1(G), \ldots, (M_{G,n}^* + 1) s_k(G); H)
\leq 2(1 + \epsilon)N(n, 2M_{G,n}^* s_1(G), \ldots, 2M_{G,n}^* s_k(G); H)
\leq N(n, 2(1 + \epsilon)C_H)^{k+1} M_{G,n}^* s_1(G), \ldots, 2(1 + \epsilon)C_H)^{k+1} M_{G,n}^* s_k(G); H).$$

Since the condition (3.20) is now satisfied with $a_i = s_i(G)$, $i = 1, \ldots, k$, and $m = 2(2(1 + \epsilon)C_H)^{k+1} M_{G,n}^*$, we conclude by (3.22) that

$$P\left(N_{o,n}(G) \geq (1 + \epsilon)\mu_{o,n}(G)\right) \geq \frac{1}{4} \prod_{j=1}^{k} p_j^{s_j(G)} + 2(2(1 + \epsilon)C_H)^{k+1} M_{G,n}^* s_j(G)
\geq \frac{1}{4} \left(\prod_{j=1}^{k} p_j\right)^{B_1(\epsilon, G) M_{G,n}^*},
$$

where

$$B_1(\epsilon, G) = \max_{j=1, \ldots, k} s_j(G) \left[1 + 2 \left(2(1 + \epsilon)\max_{H \subseteq G} C_H\right)^{k+1}\right].$$
Next, we need to consider the case when (3.25) does not hold. In this case, it follows from the assumption 
\[(1 + \epsilon)\mu_{o,n}(G) \leq N(K_{k,n}, G)\] 
that, for \(n \geq 2k + 1\),
\[
\mathbb{P}(N_{o,n}(G) \geq (1 + \epsilon)\mu_{o,n}(G)) \geq \mathbb{P}(N_{o,n}(G) \geq N(K_{k,n}, G)) \geq \mathbb{P}(K(n; p_1, \ldots, p_k) = K_{k,n}) = \prod_{j=1}^{k} p_j^{n \choose j+1} = \prod_{j=1}^{k} p_j^{n \choose k+1} \geq \left(\prod_{j=1}^{k} p_j\right)^{B_2(\epsilon, G)M^*_G_{o,n}}.
\]
where
\[
B_2(\epsilon, G) = 2\left(2(1 + \epsilon)\max_{H \subseteq G} C_H\right)^{k+1} s_k(G).
\]

Now, (3.17) follows from (3.26) and (3.27).

4 | SIMPLICES OF THE MULTIPARAMETER SIMPLICIAL COMPLEXES

Distributional limit theorems for the multiparameter simplicial complex \(K(n; p)\) were obtained in [21]. These results are obtained under the assumption
\[
p_i = n^{-\alpha_i}, \ i \geq 1,
\]
for \(\alpha_i \in [0, \infty], \ i \geq 1.\) In this section we retain this assumption and, instead of distributional results, we investigate large deviation probabilities for the number of certain simplices in \(K(n; p)\). We will use the general results obtained in the previous section.

In this section we are interested in counting the simplices of dimension \(k\), satisfying
\[
q := \min\{i \geq 1 : \alpha_i > 0\} \leq k, \ \text{such that} \ \left(\frac{k}{q}\right) \alpha_q < 1;
\]
this introduces a minor unimportant abuse of notation by conflating the dimension of the simplex with the largest dimension of an entire complex.

Let \(\sigma_k\) be a simplex of dimension \(k\) satisfying (4.2). The following proposition, a part of which requires an extra assumption on the parameters, describes the size of a crucial ingredient in the logarithmic order of magnitude of the upper tail large deviations probability: \(\mathbb{P}(N_{o,n}(\sigma_k) > (1 + \epsilon)\mu_{o,n}(\sigma_k))\) for \(\epsilon > 0.\) Notice that the extra assumption (4.3) below, as well as the second inequality in (4.2), will be used only for proving a lower bound in (4.4); see (4.16) and (4.20).

**Proposition 4.1.** Suppose that (4.2) holds. In the case of \(q < k\), suppose also that for any \(k_0 \in \{q + 1, \ldots, k\},\)
\[
\frac{k - q}{k + 1} \left(\frac{k + 1}{q + 1}\right) \alpha_q + \sum_{j=q+1}^{k_0} \left(\frac{k + 1}{j + 1}\right) \alpha_j < k_0 - q.
\]

Then, for large enough \(n,\)
\[
C_k^{-1} n^{q+1-(q+1)\alpha_q} \leq M^*_{\alpha,n}(n^{-\alpha_1}, \ldots, n^{-\alpha_q}) \leq C_k n^{q+1-(q+1)\alpha_q}
\]
for some $C_k \geq 1$. In particular, for $\varepsilon > 0$ and all large enough $n$,

$$\exp \left\{ -C_k'(\varepsilon)n^{q+1 - \left( \frac{1}{q} \right) \alpha_q} \log n \right\} \leq \mathbb{P} \left( N_{\alpha,n}(\sigma_k) \geq (1 + \varepsilon)\mu_{\alpha,n}(\sigma_k) \right) \leq \exp \left\{ -C_k''(\varepsilon)n^{q+1 - \left( \frac{1}{q} \right) \alpha_q} \right\}$$

(4.5)

for some positive constants $C_k'(\varepsilon)$, $C_k''(\varepsilon)$.

Remark 4.2. It is common to say that a dimension $k$ is critical if

$$\sum_{i=1}^{k} \binom{k}{i} \alpha_i < 1 < \sum_{i=1}^{k+1} \binom{k+1}{i} \alpha_i, \quad \text{and} \quad q \leq k;$$

(4.6)

see [11, 21]. Counting the simplices at the critical dimension is particularly important since their numbers largely determine the behaviour of the Euler characteristic of $K(n; p)$ and the Betti numbers of the same dimension; see [15, 21, 22] for more information on how to relate simplex counts to these topological invariants. The assumption (4.2) is weaker than (4.6), and in certain scenarios, Proposition 4.1 can apply to simplex counts in dimensions above the critical dimension.

Remark 4.3. Proposition 4.1 indicates that, at least under an extra condition, it is the skeleton of dimension $q$, the lowest nontrivial dimension of the complex, that plays a crucial role in determining the rate of decay of the upper large deviation probabilities for the number of simplices of dimension $k \geq q$. The reason appears to be the fact that “flipping” of a $q$-simplex from “on” to “off” or vice versa affects the topology of the complex more than does any flipping in other dimensions. The same phenomenon has been observed in the central limit theorem for the simplex counts; see Proposition 3.6 in [21].

Remark 4.4. Suppose $q = 1$ in (4.2) and assume the subcriticality condition

$$\sum_{i=1}^{k} \binom{k}{i} \alpha_i < 1,$$

(4.7)

which is one of the inequalities in (4.6). Then, if $k = 1$, the statement of Proposition 4.1 follows from Corollary 1.7 in [13]. For $k = 2$, it is easy to see that condition (4.3) follows from (4.7), but that is no longer the case for $k \geq 3$. However, if $k = 3$, one can still directly compute the value of $M^*_{\alpha,n}(n^{-\alpha_1}, n^{-\alpha_2}, n^{-\alpha_3})$ and verify the inequalities in (4.4), without using condition (4.3). To summarize, the claim of Proposition 4.1 holds at least for $k \in \{1, 2, 3\}$, under the assumption (4.7) only. We do not know if one can deduce the same conclusion for $k \geq 4$.

Proof. Since (4.5) follows from (4.4) and Theorem 3.3, we only need to prove the bounds in (4.4). We start with the upper bound. Recall that

$$M^*_{\alpha,n}(n^{-\alpha_1}, \ldots, n^{-\alpha_k}) = \min_{H: \text{subcomplex of } \sigma_k} K_H,$$

where for a subcomplex $H$ of $\sigma_k$,

$$K_H = \max \left\{ m \leq \binom{n}{k+1}: N \left( n, m \binom{k+1}{2}, m \binom{k+1}{3}, \ldots, m \binom{k+1}{k}, m; H \right) \leq \Psi_{H,n} \right\}.$$

(4.8)
Therefore, to prove the upper bound in (4.4) we only need to detect a specific subcomplex $H$ of $\sigma_k$, such that

$$K_H \leq C_k n^{q+1-\left(\frac{k}{q}\right)a_q}. \quad (4.9)$$

Let us take $H$ to be the $q$-skeleton of $\sigma_k$. For this $H$, in the obvious notation,

$$K_H = \max \left\{ m \leq \left( \begin{array}{c} n \\ k+1 \end{array} \right) : N \left( n, m \left( \begin{array}{c} k+1 \\ 2 \end{array} \right), \ldots, m \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) ; H \right) \leq n^{k+1-\left(\frac{k+1}{q+1}\right)a_q} \right\}. \quad (4.10)$$

By Proposition 3.1,

$$a_k e^\gamma \leq N \left( n, m \left( \begin{array}{c} k+1 \\ 2 \end{array} \right), \ldots, m \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) ; H \right) \leq b_k e^\gamma \quad (4.11)$$

for some $a_k, b_k > 0$, where

$$\gamma = \max \sum_{v=1}^{k+1} x_v \quad (4.12)$$

subject to

$$0 \leq x_v \leq \log n, \ v = 1, \ldots, k+1,$$

$$\sum_{v \in \sigma_i} x_v \leq \log \left\{ m \left( \begin{array}{c} k+1 \\ i+1 \end{array} \right) \right\}, \ \sigma_i \in H_i, \ i = 1, \ldots, q.$$ 

First, suppose $\left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) m > n^{q+1}$, in which case, we have $\left( \begin{array}{c} k+1 \\ j+1 \end{array} \right) m > n^{q+1}, j = 1, \ldots, q$. Then, $x_v = \log n, \ v = 1, \ldots, k+1$, is easily seen to be an optimal solution to (4.12). It then follows from (4.11) that

$$N \left( n, m \left( \begin{array}{c} k+1 \\ 2 \end{array} \right), \ldots, m \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) ; H \right) \geq a_k n^{k+1}. \quad (4.13)$$

However, as $\alpha_q > 0$, there is no $m \in \mathbb{N}$ that satisfies (4.13) and the inequality in (4.10).

Therefore, we only need to consider the case $\left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) m \leq n^{q+1}$. Then, one can see that

$$x_v = \frac{1}{q+1} \log \left\{ m \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) \right\}, \ v = 1, \ldots, k+1,$$

is an optimal solution to the linear program (4.12), so that

$$\gamma = \frac{k+1}{q+1} \log \left\{ m \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right) \right\}. \quad (4.12)$$

Therefore, by (4.11),

$$b_k^{-\frac{q+1}{k+1}} \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right)^{-1} n^{q+1-\left(\frac{k}{q}\right)a_q} \leq K_H \leq a_k^{-\frac{q+1}{k+1}} \left( \begin{array}{c} k+1 \\ q+1 \end{array} \right)^{-1} n^{q+1-\left(\frac{k}{q}\right)a_q},$$

and (4.9) follows.
We now prove the lower bound in (4.4). For this purpose we need to prove that for every subcomplex $H$ of $\sigma_k$,

$$K_H \geq C_k^{-1} n^{q+1-\binom{q}{q}^d}.$$

(4.14)

Consider first a subcomplex $H$ of dimension $1, \ldots, q-1$. In this case, it is clear that $\Psi_{H,n} = n^{k+1}$, and thus, $K_H = \binom{n}{k+1}$ and (4.14) trivially holds. Consider next a subcomplex $H$ of dimension $q$. Let $\overline{H}$ be a $(q+1)$-uniform hypergraph on $k+1$ vertices with its hyperedges identified as a $q$-simplex in $H$. Given another hypergraph $\overline{F}$, define $\overline{n}(\overline{F}, \overline{H})$ to be the number of unordered copies (as a hypergraph) of $\overline{H}$ in $\overline{F}$. Define also

$$\overline{N} \left(n, m \left(\frac{k+1}{q+1}\right); \overline{H}\right) := \max \left\{ \overline{n}(\overline{F}, \overline{H}) : v_F \leq n, \ e_F \leq m \left(\frac{k+1}{q+1}\right) \right\},$$

where $v_F$ is the number of vertices in $\overline{F}$ and $e_F$ the number of hyperedges in $\overline{F}$. Then, by construction, we have

$$\overline{N} \left(n, m \left(\frac{k+1}{q+1}\right); \overline{H}\right) \geq N \left(n, m \left(\frac{k+1}{2}\right), \ldots, m \left(\frac{k+1}{q+1}\right); H\right).$$

By virtue of this inequality together with $v_{\overline{H}} = k+1$ and $e_{\overline{H}} = s_q(H)$,

$$K_H = \max \left\{ m \leq \binom{n}{k+1} : N \left(n, m \left(\frac{k+1}{2}\right), \ldots, m \left(\frac{k+1}{q+1}\right); H\right) \leq n^{k+1-s_q(H)\alpha_q} \right\} \geq \max \left\{ m \leq \binom{k+1}{q+1}^{-1} \binom{n}{q+1} : \overline{N} \left(n, m \left(\frac{k+1}{q+1}\right); \overline{H}\right) \leq n^{\psi^\pi p_q^{-1}} \right\}.$$  

(4.15)

Observe that $\overline{H}$ is seen to be a subhypergraph of a $\binom{k}{q}$-regular, $(q+1)$-uniform hypergraph. Moreover, by (4.2),

$$p_q = n^{-q} \geq n^{-\binom{q}{q}^{-1}} > n^{-\binom{q+1}{q}^{-1}}.$$  

(4.16)

Hence, Lemma A.3 in the Appendix with $\ell = q+1$, $d = \binom{k}{q}$, and $p = p_q$, implies that the last quantity in (4.15) is at least

$$C_n^{q+1} p_q^{\binom{q}{q}} = C n^{q+1-\binom{q}{q}^d} \alpha_q,$$

for some constant $C$, as desired for (4.14).

Now, it remains to establish (4.14) for subcomplexes $H$ of dimension $k_0 \in \{q+1, \ldots, k\}$. By (4.8), we need to show that there exists $C_k > 0$ such that for any subcomplex $H$ of $\sigma_k$ on $k+1$ vertices, and all $n$ large enough,

$$N \left(n, \left[ C_k^{-1} n^{q+1-\binom{q}{q}^d} \alpha_q \right] \left(\frac{k+1}{2}\right), \ldots, \left[ C_k^{-1} n^{q+1-\binom{q}{q}^d} \alpha_q \right] \left(\frac{k+1}{k_0+1}\right); H\right) \leq n^{k+1-\sum_{j=k_0+1}^{k} \alpha_j \alpha_q}.$$
It follows from Proposition 3.1, together with the dual formulation (3.4), that for a given subcomplex $H$ and $n$ large enough, it is sufficient to exhibit non-negative numbers $y_v, v = 1, \ldots, k + 1$ and $z^{(i)}_{\sigma_i}, \sigma_i \in H_i, i = 1, \ldots, k_0$, such that

$$y_v + \sum_{i=1}^{k_0} \sum_{\sigma_i \in H_i} z^{(i)}_{\sigma_i} \geq 1 \text{ for any } v = 1, \ldots, k + 1,$$

(4.17)

and

$$\sum_{v=1}^{k+1} y_v \log n + \sum_{i=1}^{k_0} \sum_{\sigma_i \in H_i} z^{(i)}_{\sigma_i} \log \left[ C_k^{-1} n^{q+1-\left(\frac{k}{q}\right)\alpha_q} \right] \left( k + 1 \right) \leq \left( k + 1 - \sum_{j=q}^{k_0} s_j(H)\alpha_j \right) \log n + B,$$

(4.18)

where $B$ is a $k$-dependent constant. It is clear that if we can choose these numbers in such a way that

$$\sum_{v=1}^{k+1} y_v = k + 1 - \sum_{j=q}^{k_0} s_j(H)\alpha_j,$$

(4.19)

then (4.18) will be satisfied for large $n$, regardless of the constant $C_k$ above. Specifically, we choose the numbers $(y_v)$ and $(z^{(i)}_{\sigma_i})$ as follows.

$$y_v = 1 - \frac{s_{k_0,v}}{s_{k_0}(H)}, \quad v = 1, \ldots, k + 1,$$

$$z^{(k_0)}_{\sigma_{k_0}} = \frac{1}{s_{k_0}(H)}, \quad i = k_0, \quad z^{(i)}_{\sigma_i} = 0, \quad i \neq k_0,$$

where for a vertex $v$, $s_{k_0,v}$ is the number of $k_0$-simplices in $H$ to which $v$ belongs. It is elementary to check that these variables satisfy the constraints in (4.17) as equalities. Moreover, it is evident that

$$\sum_{i=1}^{k_0} \sum_{\sigma_i \in H_i} z^{(i)}_{\sigma_i} = 1,$$

while

$$\sum_{v=1}^{k+1} y_v = k + 1 - \frac{1}{s_{k_0}(H)} \sum_{v=1}^{k+1} s_{k_0,v} = k + 1 - (k_0 + 1) = k - k_0,$$

since every $k_0$-simplex contributes to exactly $k_0 + 1$ vertices. Therefore, (4.19) reduces to

$$k - k_0 + q + 1 - \left( \begin{array}{c} k \\ q \end{array} \right) \alpha_q < k + 1 - \sum_{j=q}^{k_0} s_j(H)\alpha_j.$$

(4.20)

Since

$$s_j(H) \leq \left( \begin{array}{c} k + 1 \\ j + 1 \end{array} \right), \quad j = q, \ldots, k_0,$$

(4.20) follows from (4.3).
5 | THE BETTI NUMBER AT THE CRITICAL DIMENSION

In this section we will use the results of Section 4 to derive large deviation results for the Betti number at the critical dimension. We still assume that the probabilities \((p_i, i \geq 1)\) are given by (4.1). Recall that a dimension \(k^* \geq 1\) is called critical if

\[
\sum_{i=1}^{k^*} \binom{k^*}{i} \alpha_i < 1 < \sum_{i=1}^{k^*+1} \binom{k^*+1}{i} \alpha_i, \quad \text{and} \quad q \leq k^*.
\]

(5.1)

We consider the Betti number \(\beta_{k^*,n}\) at the critical dimension. By the Morse inequalities (see e.g., Exercise 1, p. 61 in [20]), for any dimension \(j\),

\[
N_n(\sigma_j) - N_n(\sigma_{j-1}) - N_n(\sigma_{j+1}) \leq \beta_{j,n} \leq N_n(\sigma_j),
\]

(5.2)

where \(N_n(\sigma_j)\) is the number of \(j\)-simplices in the multiparameter simplicial complex. It follows from Proposition 3.1 in [21] that

\[
\mathbb{E}[N_n(\sigma_j)] \sim \frac{n^{\tau_j}}{(j+1)!}, \quad n \to \infty,
\]

(5.3)

with

\[
\tau_j = j + 1 - \sum_{i=1}^{j} \binom{j+1}{i+1} \alpha_i, \quad j = 1, 2, \ldots,
\]

and the sequence \(\tau_1, \tau_2, \ldots\) reaches its unique maximum at the critical dimension \(k^*\). It thus follows from (5.2) and (5.3) that

\[
\mathbb{E}[\beta_{k^*,n}] \sim \frac{n^{\tau_{k^*}}}{(k^*+1)!}, \quad n \to \infty.
\]

The following theorem considers the upper tail large deviations for \(\beta_{k^*,n}\). For the precise statement, one has to add an extra technical condition, (5.5) below. The extra condition is only needed for a lower bound in (5.8). We do not yet have a full picture of what happens if (5.5) is removed.

**Theorem 5.1.** As in Proposition 4.1, we assume that for any \(k_0 \in \{q + 1, \ldots, k^*\}\),

\[
\frac{k^* - q}{k^* + 1} \binom{k^* + 1}{q + 1} \alpha_q + \sum_{j=q+1}^{k_0} \binom{k^* + 1}{j+1} \alpha_j < k_0 - q.
\]

(5.4)

Assume also that

\[
\sum_{i=q}^{k^*+1} \binom{k^* + 1}{i} \alpha_i - 1 \geq \frac{q(k^* + 2)!}{(q+1)!(k^* + 1 - q)!(k^* + 1)} \alpha_q.
\]

(5.5)

Then for any \(\epsilon > 0\), there exist positive constants \(C_1(\epsilon), C_2(\epsilon) > 0\), such that

\[
\exp \left\{ -C_1(\epsilon)n^{q+1-\binom{\epsilon}{q}a_q} \log n \right\} \leq \mathbb{P} \left\{ \beta_{k^*,n} \geq (1 + \epsilon) \frac{n^{\tau_{k^*}}}{(k^*+1)!} \right\} \leq \exp \left\{ -C_2(\epsilon)n^{q+1-\binom{\epsilon}{q}a_q} \right\}
\]

(5.6)
for all \( n \) large enough.

**Proof.** Using (5.2) and Proposition 4.1, we have

\[
\Pr \left( \beta_{k^*,n} \geq (1 + \varepsilon) \frac{n^{v_q}}{(k^* + 1)!} \right) \leq \Pr \left( N_n(\sigma_{k^*}) \geq (1 + \varepsilon) \frac{n^{v_q}}{(k^* + 1)!} \right) \leq \exp \left\{ -C(\varepsilon)n^{q+1} \left( \frac{1}{q!} \right)^{a_q} \right\},
\]

for some constant \( C(\varepsilon) > 0 \). For the application of Proposition 4.1, we need not only (5.4) but also (4.2) with \( k^* \) replaced with \( k^* \); the latter condition however directly follows from the criticality condition in (5.1). Now, the upper bound in (5.6) has been obtained.

To prove the lower bound, let \( D \) be a positive constant to be determined in a moment and take

\[
m = \left( \frac{D(1 + \varepsilon))}{(k^* + 1)!} \right)^{1/(k^* + 1)} n^{1 - \frac{1}{q!} \left( \frac{1}{q!} \right)^{a_q}}.
\]

Fix arbitrary \( m \) vertices out of \( n \) (say, the vertices numbered \( 1, \ldots, m \)). Let \( A \) be the event for which all \( \binom{m}{q+1} \) simplices of dimension \( q \) based on these \( m \) vertices are present in \( K(n; p) \). Given a simplicial complex, its simplex is said to be maximal if it is not in the boundary of any simplex of a larger dimension in that complex. Define \( F_{k^*,n} \) to be the number of maximal \( k^* \)-simplices, supported on \( \{1, \ldots, m\} \in K(n; p) \). Then, conditionally on \( A \), a set of \( k^* + 1 \) vertices in \( \{1, \ldots, m\} \) forms a maximal \( k^* \)-simplex in \( K(n; p) \) with probability

\[
\prod_{i=1}^{k^*} p_i^{\binom{m}{q+1}} \left( 1 - \prod_{i=q+1}^{k^*} p_i^{\binom{m}{q+1}} \right)^{m-(k^*+1)} \left( 1 - \prod_{i=q}^{k^*} p_i^{\binom{m}{q+1}} \right)^{n-m}.
\]

Indeed, conditionally on \( A \), the first term in (5.7) is the probability that a set \( \sigma \) of \( k^* + 1 \) vertices in \( \{1, \ldots, m\} \) forms a \( k^* \)-simplex. The second term in (5.7) is the probability that \( \sigma \) does not form a \( (k^* + 1) \)-simplex with any vertex \( v \in \{1, \ldots, m\} \setminus \sigma \). Finally, the last term in (5.7) is the probability that \( \sigma \) does not form a \( (k^* + 1) \)-simplex with any vertex \( v \in \{m+1, \ldots, n\} \), with everything still being conditional on \( A \).

Note that the probability (5.7) is independent of the choice of \( k^* + 1 \) vertices in \( \{1, \ldots, m\} \). Since there are \( \binom{m}{k^*+1} \) such potential maximal \( k^* \)-simplices in \( \{1, \ldots, m\} \), it follows from Lemma A.2 in the Appendix that

\[
\Pr \left( F_{k^*,n} \geq \frac{1}{2} \binom{m}{k^* + 1} \prod_{i=1}^{k^*} p_i^{\binom{m}{q+1}} \left( 1 - \prod_{i=q+1}^{k^*} p_i^{\binom{m}{q+1}} \right)^{m-(k^*+1)} \left( 1 - \prod_{i=q}^{k^*} p_i^{\binom{m}{q+1}} \right)^{n-m} \mid A \right) \geq \frac{1}{4} \prod_{i=q+1}^{k^*} p_i^{\binom{m}{q+1}} \left( 1 - \prod_{i=q+1}^{k^*} p_i^{\binom{m}{q+1}} \right)^{m-(k^*+1)} \left( 1 - \prod_{i=q}^{k^*} p_i^{\binom{m}{q+1}} \right)^{n-m}.
\]

Note that

\[
\left( 1 - \prod_{i=q+1}^{k^*} p_i^{\binom{m}{q+1}} \right)^{m-(k^*+1)} \geq \left( 1 - n \sum_{i=q+1}^{k^*+1} \binom{m}{q+1} n^{1/\binom{m}{q+1}} \right)^{n-m} \left( 1 - n \sum_{i=q+1}^{k^*+1} \binom{m}{q+1} n^{1/\binom{m}{q+1}} \right)^{n-m}.
\]
It then follows from (5.5) that

\[
\sum_{i=q+1}^{k^*+1} \binom{k^* + 1}{i} \alpha_i > 1 - \frac{1}{q+1} \binom{k^*}{q} \alpha_q,
\]

hence, there exists a constant \( \zeta > 0 \), such that

\[
\left( 1 - \prod_{i=q+1}^{k^*+1} p_i^{(i+1)} \right)^{m-(k^*+1)} \geq \zeta.
\] (5.8)

Further, by the criticality condition (5.1),

\[
\left( 1 - \prod_{i=q+1}^{k^*+1} p_i^{(i+1)} \right)^{n-m} \geq \left( 1 - n^{-\sum_{i=q+1}^{k^*+1} \alpha_i} \right)^n \to 1, \text{ as } n \to \infty.
\]

Now, combining these results,

\[
\left( 1 - \prod_{i=q+1}^{k^*+1} p_i^{(i+1)} \right)^{m-(k^*+1)} \left( 1 - \prod_{i=q}^{k^*+1} p_i^{(i+1)} \right)^{n-m} \geq \rho,
\]

for some constant \( \rho > 0 \). Choosing \( D > 6/\rho \), one can obtain the following bound:

\[
\frac{1}{2} \left( m \right)_{k^*+1} \prod_{i=q+1}^{k^*} p_i^{(i+1)} \left( 1 - \prod_{i=q+1}^{k^*+1} p_i^{(i+1)} \right)^{m-(k^*+1)} \left( 1 - \prod_{i=q}^{k^*+1} p_i^{(i+1)} \right)^{n-m} \geq \frac{\rho m^{k^*+1}}{3(k^*+1)!} \prod_{i=q+1}^{k^*} p_i^{(i+1)} \geq \frac{D \rho (1+\epsilon)}{3(k^*+1)!} n^{r_{k^*}} > 2(1+\epsilon) \frac{n^{r_{k^*}}}{(k^*+1)!}.
\]

Now, putting all of these results gives us

\[
\mathbb{P}\left( F_{k^*,n} \geq 2(1+\epsilon) \frac{n^{r_{k^*}}}{(k^*+1)!} \right) \geq \frac{\rho}{4} \prod_{i=q+1}^{k^*} p_i^{(i+1)} \mathbb{P}(A) = \frac{\rho}{4} \prod_{i=q+1}^{k^*} p_i^{(i+1)} \frac{m^{(m+1)}_{q+1}}{(q+1)!} \geq \frac{\rho}{4} \exp \left\{ - \left( \sum_{i=q+1}^{k^*} (i+1) \alpha_i + \frac{m}{q+1} \alpha_q \right) \log n \right\} \geq \frac{\rho}{4} \exp \left\{ -C(\epsilon) n^{q+1-\left(\frac{i}{q}\right)_{aq}} \log n \right\},
\]

for some constant \( C(\epsilon) > 0 \).

Note that the indicator function of each maximal \( k^* \)-simplex is an independent \( k^* \)-dimensional cocycle. If we denote by \( \gamma_{k^*,n} \) the rank of the group of \( k^* \)-dimensional cocycles in \( K(n; p) \), then we have proved that for some constant \( C(\epsilon) > 0 \),

\[
\mathbb{P}\left( \gamma_{k^*,n} \geq 2(1+\epsilon) \frac{n^{r_{k^*}}}{(k^*+1)!} \right) \geq \frac{\rho}{4} \exp \left\{ -C(\epsilon) n^{q+1-\left(\frac{i}{q}\right)_{aq}} \log n \right\}.
\] (5.9)
In order to deduce the lower bound in (5.6) from (5.9), we only need to show that the rank of the group of \(k^*\)-dimensional coboundaries in \(K(n; p)\) exceeds \((1 + \epsilon)n^{k^*} / (k^* + 1)!\) only on an event whose probability is of smaller order than the probability in (5.9). The rank of the latter group does not exceed the rank of the group of \((k^* - 1)\)-dimensional cochains in \(K(n; p)\), which is, of course, equal to the number \(N_n(\sigma_{k^*-1})\) of \((k^* - 1)\)-simplices in \(K(n; p)\).

We know from Proposition 3.1 in [21] that

\[
\mathbb{E}[N_n(\sigma_{k^*-1})] \sim \frac{n^{k^*-1}}{k^*!}, \quad n \to \infty,
\]

by the criticality of the dimension \(k^*\). It remains to notice that for any \(\epsilon > 0\),

\[
\mathbb{P}(N_n(\sigma_{k^*-1}) \geq (1 + \epsilon)\mathbb{E}[N_n(\sigma_{k^*-1})]) \leq \exp \left\{ -Cn^{q+1-(\frac{k^*-1}{q})q} \right\}
\]

by the upper bound in (4.5). The right-hand side above exhibits a smaller order than the probability in (5.9), as required.

\[\blacksquare\]

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous referees for useful comments.

FUNDING INFORMATION

Samorodnitsky’s research is partially supported by the NSF grant DMS-2015242 and AFOSR grant FA9550-22-1-0091 at Cornell University. Owda’s research is partially supported by the NSF grant DMS-1811428 and AFOSR grant FA9550-1-0238 at Purdue University.

REFERENCES

1. N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, Israel J. Math. 38 (1981), 116–130.
2. B. B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao, Upper tails and independence polynomials in random graphs, Adv. Math. 319 (2017), 313–347.
3. S. Chatterjee, Large deviations for random graphs, Springer, Cham, Switzerland, 2017.
4. S. Chatterjee and A. Dembo, Nonlinear large deviations, Adv. Math. 299 (2016), 396–450.
5. S. Chatterjee and S. R. S. Varadhan, The large deviation principle for the Erdös–Rényi random graph, Eur. J. Comb. 32 (2011), 1000–1017.
6. N. Cook and A. Dembo, Large deviations of subgraph counts for sparse Erdös–Rényi graphs, Adv. Math. 373 (2020), 107289.
7. A. Costa and M. Farber, Large random simplicial complexes, I, J. Topol. Anal. 8 (2016), 399–429.
8. A. Costa and M. Farber, Large random simplicial complexes, III the critical dimension, J. Knot Theory Ramificat. 26 (2017), 1–26.
9. A. Dudek, J. Polcyn, and A. Ruciński, Subhypergraph counts in extremal and random hypergraphs and the fractional q-independence, J. Comb. Optim. 19 (2010), 184–199.
10. R. Eldan, Gaussian-width gradient complexity, reverse log-sobolev inequalities and nonlinear large deviations, Geom. Funct. Anal. 28 (2018), 1548–1596.
11. C. F. Fowler, Homology of multi-parameter random simplicial complexes, Discrete Comput. Geom. 62 (2019), 87–127.
12. E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another, Israel J. Math. 105 (1998), 251–256.
13. S. Janson, K. Oleszkiewicz, and A. Ruciński, Upper tails for subgraph counts in random graphs, Israel J. Math. 142 (2004), 61–92.
14. M. Kahle, *Topology of random clique complexes*, Discrete Math. 309 (2009), 1658–1671.
15. M. Kahle and E. Meckes, *Limit theorems for Betti numbers of random simplicial complexes*, Homology Homotopy Appl. 15 (2013), 343–374.
16. S. Kanazawa, *Law of large numbers for betti numbers of homogeneous and spatially independent random simplicial complexes*, Random Struct. Algorithm. 60 (2022), 68–105.
17. N. Linial and R. Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica 26 (2006), 475–487.
18. E. Lubetzky and Y. Zhao, *On the variational problem for upper tails in sparse random graphs*, Random Struct. Algorithm. 50 (2017), 420–436.
19. R. Meshulam and N. Wallach, *Homological connectivity of random k-dimensional complexes*, Random Struct. Algorithm. 34 (2009), 408–417.
20. J. R. Munkres, Elements of algebraic topology, Addison-Wesley Publishing Company, Menlo Park, CA, 1984.
21. T. Owada, G. Samorodnitsky, and G. Thoppe, *Limit theorems for topological invariants of the dynamic multi-parameter simplicial complex*, Stoch. Process. Appl. 138 (2021), 56–95.
22. G. C. Thoppe, D. Yogeshwaran, and R. J. Adler, *On the evolution of topology in dynamic clique complexes*, Adv. Appl. Probab. 48 (2016), 989–1014.
23. J. Yan, *Nonlinear large deviations: Beyond the hypercube*, Ann. Appl. Probab. 30 (2020), 812–846.

How to cite this article: G. Samorodnitsky and T. Owada, *Large deviations for subcomplex counts and Betti numbers in multiparameter simplicial complexes*, Random Struct. Alg. 63 (2023), 533–556. https://doi.org/10.1002/rsa.21146

APPENDIX

We collect in this section, for convenience of the reader, several known auxiliary results we use in the proofs of Proposition 3.1, Theorem 3.3, Proposition 4.1, and Theorem 5.1.

First, let $\mathcal{G}$ be a hypergraph on a vertex set $V$, viewed as a collection of distinct subsets (i.e., hyperedges) of a vertex set $V$. For a subset $U \subseteq V$, the trace of $\mathcal{G}$ on $U$ is defined as $\text{Tr}(\mathcal{G}, U) := \{ G \cap U : G \in \mathcal{G} \}$.

**Lemma A.1** (Lemma 1.2 in [12]). Let $(U_m)_{m \in \{1, \ldots, s\}}$ be (not necessarily distinct) subsets of $V$ such that each $x \in V$ belongs to at least $t$ of the $U_m$’s. Then, the number of (distinct) hyperedges in $\mathcal{G}$, denoted by $|\mathcal{G}|$, satisfies

$$|\mathcal{G}| \leq \prod_{m=1}^{s} |\text{Tr}(\mathcal{G}, U_m)|^{1/t}.$$

The next result is useful for a lower bound of the probability that the sum of possibly dependent indicator functions exceeds a fraction of its mean.

**Lemma A.2** (Lemma 3.3 in [13]). Let $X = \sum_{i=1}^{N} X_i$, where $(X_i)$ are indicator random variables, such that for each $i$, $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) \geq p_0$ for some $p_0 \geq 0$. Then,

$$\mathbb{P}\left( X > \frac{1}{2} N p_0 \right) \geq \frac{p_0}{4}.$$

Finally we need a technical result that allows us to obtain the lower bound in (4.14).
Let $H$ be a hypergraph. Given another hypergraph $F$, define $\bar{n}(F, H)$ to be the number of unordered copies (as a hypergraph) of $H$ in $F$. Define

$$\bar{N}(n, m; H) := \max \{ \bar{n}(F, H) : v_F \leq n, \ e_F \leq m \},$$

where $v_F$ is the number of vertices in $F$ and $e_F$ the number of hyperedges in $F$. Let $G$ be a $d$-regular and $\ell'$-uniform hypergraph. For any $n \geq 1$ and $0 < p < 1$, let

$$\bar{M}^*_{G,n}(p) := \max \left\{ m \leq \binom{n}{\ell'} : \bar{N}(n, m; H) \leq n^v p^e \text{ for every nonempty subhypergraph } H \text{ of } G \right\}.$$

**Lemma A.3** (Proposition 4.3 in [9]). *Suppose that $p \geq n^{-\ell'/d}$. Then,*

$$\bar{M}^*_{G,n}(p) \geq C n^\ell p^d$$

*for some $C > 0$.***