The degree-restricted random process is far from uniform

Michael Molloy∗ and Erlang Surya† and Lutz Warnke‡

April 5, 2023; revised February 11, 2025

Abstract

The degree-restricted random process is a natural algorithmic model for generating graphs with degree sequence $d_n = (d_1, \ldots, d_n)$: starting with an empty $n$-vertex graph, it sequentially adds new random edges so that the degree of each vertex $v_i$ remains at most $d_i$. Wormald conjectured in 1999 that, for $d$-regular degree sequences $d_n$, the final graph of this process is similar to a uniform random $d$-regular graph.

In this paper we show that, for degree sequences $d_n$ that are not nearly regular, the final graph of the degree-restricted random process differs substantially from a uniform random graph with degree sequence $d_n$. The combinatorial proof technique is our main conceptual contribution: we adapt the switching method to the degree-restricted process, demonstrating that this enumeration technique can also be used to analyze stochastic processes (rather than just uniform random models, as before).

1 Introduction

Random graph processes that grow step-by-step over time are powerful in both theory and practice: they are often used to generate sophisticated combinatorial objects with surprising properties [8, 9, 19, 10, 11], and are also frequently used to model complex networks arising in applications [1, 15, 16, 6, 41]. Rather few proof techniques exist for analyzing such stochastic processes, which is why even the most basic questions about these processes are often hard to answer. In this paper we add the so-called ‘switching method’ to the list of techniques for analyzing stochastic processes.

This paper concerns the perhaps simplest random graph process that attempts to generate a graph with a given graphic degree sequence $d_n = (d_1^{(n)}, d_2^{(n)}, \ldots, d_n^{(n)})$: the degree-restricted random $d_n$-process starts with an empty $n$-vertex graph, and then step-by-step adds a new random edge subject to the constraint that the degree of each vertex $v_i$ remains at most $d_i$ (without creating loops or multiple edges). The $d_n$-process is so natural that it has been studied since the mid 1980s in chemistry [32, 2, 3], combinatorics [45, 54, 48] and statistical physics [7], often with a focus on $d$-regular degree sequences $d_n$. For graphic degree sequences $d_n$ with constant maximum degree, Ruciński and Wormald [45] showed in 1990 that the $d_n$-process is a natural algorithmic model in the sense that it typically generates a graph with degree sequence $d_n$. Conceptually the most interesting open problem remaining [1, 8, 54, 48] concerns its distribution: is the final graph of the $d_n$-process similar to a uniform random graphs with degree sequence $d_n$?

To make this question precise, let $G_{d_n}$ be the final graph of the $d_n$-process conditioned on having degree sequence $d_n$, and let $G_{d_n}$ be the uniform random graph with degree sequence $d_n$. Since the distributions

1 As usual, a degree-sequence $d_n$ is called graphic if it is the degree sequence of some simple graph. It is well known (and easy to check) that any degree sequence $d_n = (d_1^{(n)}, d_2^{(n)}, \ldots, d_n^{(n)}) \in \{0, \ldots, \Delta\}^n$ is graphic when the degree sum $\sum_{i \in [n]} d_i^{(n)}$ is even and larger than some sufficiently large constant $m_0 = m_0(\Delta)$; see Lemma 22.

2 The $d_n$-process does not always generate a graph with the desired degree sequence. For example, using the graphic degree sequence $d_4 = (2, 2, 2, 2)$ it is possible to get stuck after three steps if we create a triangle.
of $G_{d_n}^P$ and $G_{d_n}$ are not identical, it is natural to focus on similarity with respect to their typical properties. More formally, the remaining key question is whether $G_{d_n}^P$ and $G_{d_n}$ are contiguous, i.e., if every property that holds with high probability in one also holds with high probability in the other (see Section 9.6). In fact, Wormald conjectured in 1999 that $G_{d_n}^P$ and $G_{d_n}$ are contiguous for $d$-regular degree sequences $d_n$, for any fixed $d \geq 2$ (see Conjecture 6.3). Until now these contiguity questions have remained open, partly because we are lacking suitable proof techniques for such questions in random graph processes such as $G_{d_n}^P$ (which are much harder to analyze than uniform random graphs such as $G_{d_n}$).

In this paper we show that, for graphic degree sequences $d_n$ with constant maximum degree $\Delta = O(1)$, the final graph $G_{d_n}^P$ of the $d_n$-process is not contiguous to the uniform random graph $G_{d_n}$, unless the degree sequence $d_n$ is nearly regular (which means that all but $o(n)$ vertices have the same degree). In other words, the final graph of the degree-restricted random process differs substantially from a uniform random graph with degree sequence $d_n$. Here we denote the number of vertices in $d_n$ with degree $j$ by

$$n_j := n_j(d_n) := \left| \{v \in \{1, \ldots, n\} : d_v^{(n)} = j \} \right|.$$  

**Theorem 1** (The degree restricted process is far from uniform). Fix two constants: an integer $\Delta > 1$ and a real $\xi \in (0, 1)$. Assume that the graphic degree sequence $d_n = (d_1^{(n)}, \ldots, d_n^{(n)}) \in \{1, 2, \ldots, \Delta\}^n$ satisfies the ’non-regularity’ assumption $\max_{1 \leq j \leq \Delta} n_j/n \leq 1 - \xi$. Then the final graph $G_{d_n}^P$ of the degree restricted random $d_n$-process is not contiguous to the uniform random graph $G_{d_n}$ with degree sequence $d_n$.

This theorem follows from a more general result we shall present in Section 1.1, which shows that a certain edge-statistic typically differs substantially between $G_{d_n}^P$ and $G_{d_n}$. This yields a stronger discrepancy between $G_{d_n}^P$ and $G_{d_n}$ than non-contiguity, including that the total variation distance between them is nearly maximal, and also the existence of a simple algorithm that can distinguish between them.

The combinatorial proof technique used for Theorem 1 is our main conceptual contribution: we adapt the switching method to the degree-restricted random $d_n$-process, demonstrating that this enumeration technique can also be used to analyze stochastic processes (rather than just uniform random models, as before). Interestingly, this is not the first time that the $d_n$-process has stimulated the development of a new proof technique for stochastic processes: indeed, the widely-used differential equation method as well as the associated self-correction idea were first developed for the random $d_n$-process; both were crucial for later breakthroughs in Ramsey Theory, demonstrating the potential of developing new proof techniques for stochastic processes such as the random $d_n$-process.

### 1.1 Main result: discrepancy in edge-statistics

Our main result shows that the number of certain edges differs substantially between the degree restricted random process and a uniform random graph with degree sequence $d_n$. To formalize this difference, note that after rescaling $\xi$ to $\xi/2$ (as we may), the ’non-regularity’ assumption of Theorem 1 implies that

$$\sum_{1 \leq j \leq k} n_j/n \in [\xi, 1 - \xi]$$

for some integer $1 \leq k < \Delta$. For any graph $G$, let $X_k(G)$ denote the number of edges in $G$ whose endpoints are of degree at most $k$; we call such edges small edges. **Theorem 2** with $\epsilon = \beta/2$, say, shows that $X_k(G_{d_n}^P)$ and $X_k(G_{d_n})$ differ with high probability, which implies the non-contiguity result Theorem 1. Here

$$\mu = \mu(d_n, k) := \frac{(\sum_{1 \leq j \leq k} n_j)^2}{4m} \quad \text{with} \quad m = m(d_n) := \frac{\sum_{1 \leq j \leq \Delta} j n_j}{2},$$

where $\mu$ approximates the expected number of small edges in the uniform random graph $G_{d_n}$ (see Appendix A.2), and $m$ equals the number of edges of the degree sequence $d_n$.

---

3 The fact that the distributions of $G_{d_n}^P$ and $G_{d_n}$ are not identical can be seen by inspection of specific sequences on a few vertices; see Appendix A.1 for an example. For 2-regular degree sequences $d_n$ certain expectations and probabilities also differ slightly, but these minor differences do not rule out contiguity (since they do not concern high probability events).

4 As usual, we say that a graph property holds with high probability if it holds with probability tending to 1 as $n \to \infty$.

5 For example, a moment of reflection reveals that even estimating $P(G_{d_n}^P = G)$ is challenging; see Section 1.1.
Theorem 2 (Discrepancy in number of small edges). Fix two constants: an integer $\Delta > 1$ and a real $\xi \in (0, 1)$. Assume that the graphic degree sequence $d_n \in \{1, \ldots, \Delta\}^n$ satisfies for some integer $k = k(d_n) \in \{1, \ldots, \Delta - 1\}$ the ‘non-regularity’ assumption \[1\]. Then the following holds, with $\mu = \mu(d_n, k)$ as defined in \(2\):

(i) For any $\varepsilon = \varepsilon(n) \gg n^{-1/2}$, the uniform random graph $G_{d_n}$ with degree sequence $d_n$ satisfies
\[ P\left( \left| X_k(G_{d_n}) - \mu \right| < \varepsilon \mu \right) \geq 1 - e^{-\Theta(\varepsilon^2 n)}. \]  
(ii) There is a constant $\beta = \beta(\xi, \Delta) > 0$ such that the final graph $G_{d_n}'$ of the $d_n$-process satisfies
\[ P\left( \left| X_k(G_{d_n}') - \mu \right| > \beta \mu \right) \geq 1 - e^{-\Theta(n)}. \]

Invoking again Theorem 2 with $\varepsilon = \beta/2$, note that, with probability at least $1 - e^{-\Theta(n)}$, the number of small edges satisfies $|X_k(G_{d_n}') - \mu > \beta \mu$ and $|X_k(G_{d_n}) - \mu < \frac{1}{2} \beta \mu$. This striking difference between $G_{d_n}'$ and $G_{d_n}$ has a number of interesting consequences: (a) a polynomial algorithm can tell $G_{d_n}'$ and $G_{d_n}$ apart by simply counting the number of small edges, (b) the edit-distance between $G_{d_n}'$ and $G_{d_n}$ is at least $\frac{1}{2} \beta \mu = \Theta(n)$, i.e., very large, and (c) the total variation distance between $G_{d_n}'$ and $G_{d_n}$ is nearly maximal, i.e., close to one:
\[ |d_{TV}(G_{d_n}', G_{d_n}) - 1| \leq e^{-\Theta(n)}. \]

The proof of inequality (3) for the uniform random graph with degree sequence $d_n$ is standard: it is based on routine configuration model \cite{12, 54, 20} arguments, and appears in Appendix A.2. Our main contribution is the proof of inequality (4) for the $d_n$-process: it is based on an intricate edge-switching argument involving the trajectories of the random $d_n$-process, see the proof outline in Section 1.2 and the technical details in Sections 3–5. In Section 6, we also mention a potential alternative approach based on the differential equation method based, which so far has resisted rigorous analysis (which is why we resorted to the switching method).

We now give some intuition as to why the number $X_k$ of small edges differs between $G_{d_n}$ and $G_{d_n}'$. Namely, if we construct a random graph with degree sequence $d_n$ using the standard configuration model by adding edges step-by-step (one at a time), then in each step the probability of selecting a vertex $v$ as an endpoint is proportional to the current unused degree of that vertex. By contrast, in the degree restricted $d_n$-process all vertices with positive unused degree are (approximately) equally likely to be selected in each step. This ‘preferential versus uniform’ difference heuristically suggests that small edges tend to be created with higher probability in the $d_n$-process, which makes the edge-discrepancies in Theorem 2 plausible (see also Remark 4).

1.2 Proof strategy: switching method

The basic idea of the switching method is to estimate ratios of closely related set-sizes via local perturbations: namely, by defining a suitable ‘switching operation’ that maps objects in $A$ to objects in $B$, one can often estimate the set-size ratio $|A|/|B|$ fairly precisely via a double-counting argument (by estimating the number of switchings from $A$ to $B$ and the number of inverse switchings from $B$ to $A$). This has been successfully used to enumerate many combinatorial structures and analyze uniform models thereof, including graphs with a given degree sequence \cite{36, 38, 53, 26, 22, 31}, 0-1-matrices \cite{25, 37}, regular hypergraphs \cite{17}, structured graph classes \cite{53, 21, 13}, Latin rectangles \cite{23} and Latin squares \cite{33, 34}.

Our proof of Theorem 2(ii) adapts the switching method to the degree restricted random $d_n$-process, which requires several new ideas. Indeed, for uniform random structures such as $G_{d_n}$, the normalizing constants in $P(G_{d_n} \in A)/P(G_{d_n} \in B) = |A|/|B|$ cancel, and the switching method directly applies. This simplification does not happen for stochastic processes like the $d_n$-process, so we needed to develop a form of the switching method that allows for different probabilities: roughly speaking, we (i) apply switching operations directly to the trajectories of the random $d_n$-process, and then (ii) average over the probabilities of these trajectories and their corresponding graphs; both ideas should also aid analysis of other stochastic processes.

The starting point of our proof strategy is the following well-known\footnote{Given two graphs $G_1, G_2$ with the same number of vertices, the edit-distance between them is defined as the minimum number of edge-changes (addition or removal) one needs to apply to $G_1$ in order to obtain a graph that is isomorphic to $G_2$.} switching operation on graphs:
\[ \text{given a graph } G^+ \text{ with degree sequence } d_n \text{ where the two edges } ab, xy \text{ satisfy the degree constraints} \]

\[ \text{with } d_{TV}(G_{d_n}', G_{d_n}) = 1 \text{ where the sum is taken over all possible graphs } G \text{ with degree sequence } d_n. \]

This switching operation was introduced by Peterson \cite{42, 40} in year 1891 (to analyze structural properties of regular graphs).
Figure 1: Switching example: the edges $ab, xy$ in $G^+$ are replaced with the edges $ax, by$ to obtain $G^-$ (all other edges remain unchanged). Note that $G^+$ and $G^-$ have the same vertex degrees, and that $G^+$ has one more edge than $G^-$ where both endpoints have degree one (which are so-called small edges).

Since our goal is to show that the $d_n$-process prefers more small edges than the uniform model $G_{d_n}$, in view of $\mathbb{P}(G_{d_n} = G^+)/\mathbb{P}(G_{d_n} = G^-) = 1$ we thus would like to prove the probability ratio estimate

$$\frac{\mathbb{P}(G_{d_n}^p = G^+)}{\mathbb{P}(G_{d_n}^p = G^-)} \geq 1 + \epsilon$$

for some constant $\epsilon > 0$, as this would imply the desired discrepancy in the number of small edges.

In Section 1.2.1 we first sketch a switching argument for establishing (7) in the special case $k = 1$, and outline how it implies Theorem 2 (ii). Afterwards, in Section 1.2.2 we discuss how we modify this switching argument for the more interesting general case $k \geq 1$, where we need to deal with the major obstacle that (7) is not always true: see Figure 3 in Appendix A.1 for a counterexample with $k = 2$. Finally, in Section 1.2.3 we compare our approach with classical switching arguments (that do not concern stochastic processes).

### 1.2.1 Special case $k = 1$: comparison of edge-sequences of graphs

To prove the probability ratio estimate (7) in the special case $k = 1$, we look at the trajectories of the $d_n$-process, which means that we take the order of the added edges into account (this is important, since different orderings of the same edge set can occur with different probabilities). In particular, by summing over the set $\Pi_G$ of all ordered edge-sequences of a given graph $G$, the probability that the $d_n$-process produces $G$ is

$$\mathbb{P}(G_{d_n}^p = G) = \sum_{\sigma \in \Pi_G} \mathbb{P}(\sigma),$$

where $\mathbb{P}(\sigma)$ denotes the probability that the $d_n$-process produces the edge-sequence $\sigma$. Recall that, in our switching operation on graphs, the edges $ab, xy$ in $G^+$ are replaced with the edges $ax, by$ to obtain $G^-$; see Figure 1. Let $\sigma_{ab,xy}$ be any edge-sequence of $G^+$ where the edge $ab$ appears earlier than the edge $xy$. For our new switching operation on edge-sequences, it is natural to consider the following two edge-sequences of $G^-$: (i) the edge-sequence obtained by replacing $ab, xy$ with $ax, by$ in that order, which we call $\sigma_{ax,by}$, and (ii) the one obtained by replacing $ab, xy$ with $by, ax$ in that order, which we call $\sigma_{by,ax}$. To complete the symmetry, we also consider the edge-sequence of $G^+$ obtained by swapping the positions of $ab, xy$ in $\sigma_{ab,xy}$, which we call $\sigma_{xy,ab}$. By expanding $\mathbb{P}(\sigma)$ and carefully estimating the resulting formula (see (20)–(21) in Section 3.1), it turns out that one can prove the surprisingly clean ratio estimate

$$\frac{\mathbb{P}(\sigma_{ab,xy}) + \mathbb{P}(\sigma_{xy,ab})}{\mathbb{P}(\sigma_{ax,by}) + \mathbb{P}(\sigma_{by,ax})} \geq 1,$$

9See Sections 2.1–2.2 for how we deal with the possibility that this replacement might create multiple edges.
which already implies the desired probability ratio estimate \((7)\) with \(\epsilon = 0\) by noting that
\[
\mathbb{P}(G_{dn}^P = G^+) = \sum_{\sigma_{ab,xy}} \mathbb{P}(\sigma_{ab,xy} + \mathbb{P}(\sigma_{xy,ab}) \geq \sum_{\sigma_{ax,by}} \mathbb{P}(\sigma_{ax,by}) + \mathbb{P}(\sigma_{by,ax}) = \mathbb{P}(G_{dn}^P = G^-). \tag{10}
\]

With more work one can show that, for a positive proportion of the edge-sequences \(\sigma_{ab,xy}\), we have
\[
\frac{\mathbb{P}(\sigma_{ab,xy} + \mathbb{P}(\sigma_{xy,ab})}{\mathbb{P}(\sigma_{ax,by}) + \mathbb{P}(\sigma_{by,ax})} \geq 1 + \epsilon'
\]
for some constant \(\epsilon' > 0\), which by similar reasoning as for inequality \((10)\) then proves the desired probability ratio estimate \((7)\) for an appropriately defined \(\epsilon > 0\), which was our main goal.

For the interested reader, we now outline how \((7)\) implies Theorem 2(b) To this end, let
\[
\mathcal{G}_\ell := \{ G : \text{graph with degree sequence } d_n \text{ and } X_k(G) = \ell \}.
\]

Note that ‘switching’ the edges \(ab, xy\) to \(ax, by\) maps a graph from \(\mathcal{G}_{\ell+1}\) to a graph in \(\mathcal{G}_\ell\). When the number of small edges satisfies \(\ell \approx \mu\), then by counting the number of ways each graph in \(\mathcal{G}_{\ell+1}\) can be mapped to a graph in \(\mathcal{G}_\ell\), and vice versa, one can obtain the size ratio estimate \(|\mathcal{G}_\ell|/|\mathcal{G}_{\ell+1}| \approx 1\) via a double counting argument that is standard for switching arguments. Combining these ideas with \((7)\), the crux is that by a more careful double counting argument one can also obtain the probability ratio estimate
\[
\frac{\mathbb{P}(X_k(G_{dn}^P) = \ell)}{\mathbb{P}(X_k(G_{dn}^P) = \ell + 1)} = \frac{\sum_{F \in \mathcal{G}_\ell} \mathbb{P}(G_{dn}^P = F)}{\sum_{H \in \mathcal{G}_{\ell+1}} \mathbb{P}(G_{dn}^P = H) } \leq 1 - \tau \tag{13}
\]
for an appropriately defined \(\tau \in (0, 1)\); in fact, estimate \((13)\) remains true as long as \(|\ell - \mu| \leq \gamma \mu\) for some sufficiently small \(\gamma > 0\). For any \(0 \leq z \leq \gamma \mu\) we then obtain the probability ratio estimate
\[
\frac{\mathbb{P}(|X_k(G_{dn}^P) - \mu| \leq z)}{\mathbb{P}(|X_k(G_{dn}^P) - \mu| \leq z + 1)} \leq \frac{\sum_{\ell:|\ell - \mu| \leq z} \mathbb{P}(X_k(G_{dn}^P) = \ell)}{\sum_{\ell:|\ell - \mu| \leq z} \mathbb{P}(X_k(G_{dn}^P) = \ell + 1)} \leq 1 - \tau, \tag{14}
\]
and the proof of inequality \((4)\) follows readily for \(\beta := \gamma/2\): by invoking \((14)\) we have
\[
\mathbb{P}(|X_k(G_{dn}^P) - \mu| \leq \beta \mu) \leq \prod_{0 \leq i \leq \lfloor 2\beta \mu \rfloor} \frac{\mathbb{P}(|X_k(G_{dn}^P) - \mu| \leq \beta \mu + i)}{\mathbb{P}(|X_k(G_{dn}^P) - \mu| \leq \beta \mu + i + 1)} \leq (1 - \tau)^{|\beta \mu|} \leq e^{-\Theta(n)}, \tag{15}
\]
where for the last inequality we used the fact that \(\mu \geq (\xi n)/(2\Delta n) = \Theta(n)\).

### 1.2.2 General case \(k \geq 1\): comparison of edge-sequences of sets of graphs

With the benefit of hindsight, our switching argument for \(k = 1\) can be summarized as follows. We started with switching inequalities \((7)\) and \((11)\) for edge-sequences, which we then sequentially ‘lifted’ to the basic switching inequality \((7)\) for graphs, then to the key inequality \((13)\) for closely related sets of graphs, and finally to the desired inequality \((15)\) for edge statistics. Our general argument for \(k \geq 1\) will go through a similar (albeit more complicated) lifting of inequalities, but there is one major obstacle: the basic switching inequality \((7)\) for graphs is not always true, even with \(\epsilon = 0\); see Appendix A.1 for a counterexample with \(k = 2\).

Our strategy for overcoming this obstacle is based on the idea of averaging over suitable sets of graphs, called clusters (see Section 3.2). To this end, for any specific choice of our vertices \(a, b, x, y\), we will partition the choices for \(G^+\) (i.e., graphs containing edges \(ab, xy\)) into sets, which we call upper clusters. For each upper cluster \(C^+\), we also construct the corresponding set \(C^-\) of the choices for \(G^+\) obtained by replacing \(ab, xy\) by \(ax, by\) in each member of \(C^+\); we call these sets lower clusters. We will essentially prove that \((7)\) holds when averaging over graphs in the corresponding clusters: namely, for every cluster pair \(C^+, C^-\) we obtain
\[
\mathbb{P}(C^+) := \sum_{G \in C^+} \mathbb{P}(G_{dn}^P = G) \geq \sum_{G \in C^-} \mathbb{P}(G_{dn}^P = G) =: \mathbb{P}(C^-), \tag{16}
\]
and for a positive proportion of the cluster pairs $C^+, C^-$ we also obtain the stronger ratio estimate

\[
\frac{\mathbb{P}(C^+)}{\mathbb{P}(C^-)} \geq 1 + \epsilon.
\]

(17)

for some constant $\epsilon > 0$ (see Lemmas 6–7 in Section 3.3).

To prove the probability ratio estimates (16)–(17), we will establish switching inequalities for edge-sequences analogous to (9)–(11) from the case $k = 1$, the main difference being as follows: while in (9)–(11) we matched two edge-sequences of $G^+$ with two edge-sequences of $G^-$, for (16)–(17) we shall match two edge-sequences of each graph in $C^+$ with two edge-sequences of graphs in $C^-$ (but not necessarily from the same graph; see Section 4).

Finally, with estimates (16)–(17) in hand, we can then prove Theorem 2(ii) similarly to inequalities (13)–(15) from the case $k = 1$, but there is another difference that makes the details more complicated: whereas the graph switching inequality (14) holds for every pair of graphs $G^+, G^-$, the cluster switching inequality (17) only holds for some pairs of clusters $C^+, C^-$. (See Section 3.4 for the details.)

1.2.3 Another look at our switching argument

We close this proof strategy subsection with a high-level discussion of how our switching argument compares to previous switching arguments applied to uniform random models, bearing in mind that the switching inequality (13) is at the core of the argument leading to the telescoping product of ratios estimate (15).

In particular, if we were analyzing uniform random graphs $G_{d_n}$ with degree sequence $d_n$, then in view of $\mathbb{P}(G_{d_n} = F) = \mathbb{P}(G_{d_n} = H)$ the left-hand side of the corresponding key-inequality (13) would reduce to $|G|/|G|_{\ell+1}$. Obtaining tight asymptotic bounds on such ratios of closely related set-sizes is usually the main step in switching papers (see, e.g., [25, 26, 28, 51, 20, 22], which also use different switching operations).

By contrast, in our analysis of the random $d_n$-process we nearly automatically have $|G|/|G|_{\ell+1} \approx 1$, since we may focus on $\ell \approx \mu$. The main step in our proof of (13) thus is to compare the average value of $\mathbb{P}(G_{d_n} = F)$ over graphs $F \in G_\ell$ to the average of $\mathbb{P}(G_{d_n} = H)$ over graphs $H \in G_{\ell+1}$, which (together with the idea of looking at edge-sequences, i.e., trajectories of the $d_n$-process) is at the heart of our use of switchings.

2 Relaxed $d_n$-process: main technical result

In this section we state our main technical result Theorem 3 for a relaxed variant of the $d_n$-process, which allows for multiple edges but no loops. Of course, this technical result is formulated in a way that will imply our main result Theorem 2(ii) for the standard $d_n$-process, see Section 2.3 and Appendix A.3. Using the relaxed $d_n$-process for the proofs has two notable technical advantages: (a) by allowing multiple edges, in the switching arguments we do not have to worry about switching to an edge that already exists, and (b) by forbidding loops, we obtain a simpler and more tractable formula for the probabilities of the process.

2.1 Definition: configuration-graphs

Inspired by the configuration model [12, 51, 20], in this subsection we define a configuration-graph with degree sequence $d_n = (d_1^{(n)}, \ldots, d_n^{(n)})$ as follows. For each vertex $v_i$ in a graph we imagine a bin containing $d_i^{(n)}$ labeled points. By a configuration-graph $G$ with degree sequence $d_n$ we mean a perfect matching of all points where we do not allow points from the same vertex to be matched. Formally, $G$ is thus a perfect matching on the set of points

\[ P(G) := \{v_i^j : i \in [n] \text{ and } 1 \leq j \leq d_i^{(n)}\}, \]

subject to having no matched pairs of the form $v_i^j v_k^j$ (to avoid creating loops).

As in the configuration model, a configuration-graph $G$ naturally represents a multigraph $\pi(G)$ with degree sequence $d_n$, obtained by contracting the set of points

\[ V_i := \{v_i^j : 1 \leq j \leq d_i^{(n)}\} \]

\[ V_i := \{v_i^j : 1 \leq j \leq d_i^{(n)}\} \]

\[ V_i := \{v_i^j : 1 \leq j \leq d_i^{(n)}\} \]

Note that if we allowed for loops, then the degree-constraint would still forbid adding loops at vertices $v$ whose current degree is $\text{deg}(v) = d_i^{(n)} - 1$, which would considerably complicate the probability formulas in Section 3.1.
to vertex $v_i$ for each $i \in [n]$. Note that the multigraph $\pi(G)$ may contain multiple edges but not loops.

With the multigraph representation in mind, we say that $P(G)$ are points of $G$, while the sets $V_i$ with $i \in [n]$ are vertices of the configuration-graph $G$. Furthermore, the degree of a vertex in a configuration-graph is its size, so that vertex $V_i$ has degree $\deg(V_i) := |V_i| = d_i^{(n)}$. We also say that the degree of a point in a configuration-graph is the size of the vertex that contains it, so that point $w \in V_i$ has degree $\deg(w) := |V_i| = d_i^{(n)}$. Hence, as with graphs, $X_k(G)$ is the number of edges whose endpoints are of degree at most $k$, meaning that the endpoints are contained in vertices of size at most $k$.

### 2.2 Definition: relaxed $d_n$-process

In this subsection we formally define our relaxed $d_n$-process, which attempts to randomly generate a configuration-graph with degree sequence $d_n = (d_1^{(n)}, \ldots, d_n^{(n)})$. At any point during the process, we say that a vertex $V_i$ is unsaturated if there is at least one point $w \in V_i$ that is currently not matched.

#### Sampling random partial configuration-graph with relaxed $d_n$-process

1. $G := (P(G), E(G))$ with $E(G) := \emptyset$
2. while there are at least 2 unsaturated vertices in $G$ do
   3. Pick a uniformly random pair $V_i, V_j$ of distinct unsaturated vertices
   4. Pick uniformly random unmatched points $v^i_i \in V_i$ and $v^j_j \in V_j$
   5. Add the edge $v^i_i v^j_j$ to $E(G)$
3. end while
4. return $G$

As with the (standard) $d_n$-process, it is possible that the relaxed $d_n$-process will not complete, i.e., will not succeed in matching every point, and so the final configuration-graph will not have the desired degree sequence $d_n$. However, it turns out that this ‘bad behavior’ happens with sufficiently small probability (see Lemma 22 in Appendix A.3), so that conditioning on the relaxed $d_n$-process completing does not create any technical problems. We henceforth write $G_{d_n}^{P,\ast}$ for the final configuration-graph of the relaxed $d_n$-process conditioned on having degree sequence $d_n$, i.e., conditioned on completing (this is well-defined, since we only consider degree sequences $d_n$ that are graphic).

### 2.3 Main technical result: small edges in relaxed $d_n$-process

Our main technical result is the following theorem, which is a reformulation of our main result Theorem 2 (ii) for the relaxed $d_n$-process. It intuitively shows that the number of small edges in $G_{d_n}^{P,\ast}$ is typically far from $\mu$, i.e., the typical number of small edges in $G_{d_n}$ (see part (i) of Theorem 2).

**Theorem 3** (Main technical result). Suppose that assumptions of Theorem 2 hold. Then there is a constant $\alpha = \alpha(\xi, \Delta) > 0$ such that the final graph $G_{d_n}^{P,\ast}$ of the relaxed $d_n$-process satisfies

$$
\mathbb{P}(\left| X_k(G_{d_n}^{P,\ast}) - \mu \right| \leq \alpha \mu) \leq e^{-\Theta(n)}. \tag{18}
$$

As the reader can guess, Theorem 3 for the relaxed $d_n$-process implies Theorem 2 (ii) for the standard $d_n$-process $G_{d_n}^{P,\ast}$. While this kind of transfer from multigraphs to simple graphs is conceptually standard (much in the spirit of the standard argument for the configuration model), it turns out that the transfer argument for the $d_n$-process is more involved than usual; hence we defer the technical details to Appendix A.3. To sum up: in order to establish Theorem 2 (ii) we must prove Theorem 3 in the following Sections 3–5.

**Remark 4.** While Theorem 3 is enough to prove our main result, by combining (18) with the lower tail result Theorem 23 from Appendix A.4 (with $\epsilon = \alpha$) we obtain the stronger bound $\mathbb{P}(X_k(G_{d_n}^{P,\ast}) \leq (1 + \alpha) \mu) \leq e^{-\Theta(n)}$. Hence, with high probability, the $d_n$-process indeed contains significantly more small edges than $G_{d_n}$ (as suggested by the heuristics in Section 1.1).
3 Core switching arguments

This section is devoted to the switching-based proof of our main technical result Theorem 3.2.2 which concerns the number of small edges in the relaxed $d_n$-process. Namely, in Sections 3.1–3.2 we introduce the setup and main switching definitions, and in Section 3.3 then state several key switching and counting results, which in turn are used to prove Theorem 3.2.2 in Section 3.4.

3.1 Preliminaries: probabilities in relaxed $d_n$-process

For our upcoming switching arguments we need to be able to compare the probabilities with which the relaxed $d_n$-process produces certain configuration-graphs. Here our basic approach is to expand these probabilities, by considering all possible edge-sequences that produce the configuration-graphs (as mentioned in Section 1.2.1).

Turning to the details, let $G$ be any configuration-graph with degree sequence $d_n$, and let $\sigma$ be any ordering of its edges (to clarify: the endpoints of the edges of a configuration-graph consist of points, see Section 2.1). Consider building $G$ by adding its edges one at a time, ordered by $\sigma$. Defining

$$\Gamma_i(\sigma) := \text{number of unsaturated vertices remaining after adding the first } i \text{ edges},$$

the probability that the relaxed $d_n$-process chooses the edges of $G$ in order $\sigma$ is easily seen to be

$$\prod_{0 \leq i \leq m-1} \frac{2}{\Gamma_i(\sigma)(\Gamma_i(\sigma) - 1)} \cdot \prod_{1 \leq j \leq n} \frac{1}{d_j^{(n)}}. \quad (20)$$

If we condition on the process completing (i.e., that the final configuration-graph has degree sequence $d_n$), then the probability that the relaxed $d_n$-process chooses the edge-sequence $\sigma$ is thus proportional to

$$Z(\sigma) := \prod_{0 \leq i \leq m-1} \frac{2}{\Gamma_i(\sigma)(\Gamma_i(\sigma) - 1)}.$$ 

(21)

Therefore $P(G_{d_n}^{P_r} = G)$, the probability that the (conditional) relaxed $d_n$-process produces $G$, is proportional to

$$Z(G) := \sum_{\sigma \in \Pi_G} Z(\sigma), \quad (22)$$

where we sum over the set $\Pi_G$ of all edge-sequences of $G$. When summing over a set $S$ of configuration-graphs $S$, we henceforth also use the convenient shorthand

$$Z(S) := \sum_{G \in S} Z(G). \quad (23)$$

For later reference, we now record two simple bounds on the number $\Gamma_i(\sigma)$ of unsaturated vertices.

**Observation 5.** We have $2(m - i)/\Delta \leq \Gamma_i(\sigma) \leq 2(m - i)$ for all $0 \leq i \leq m$.

**Proof.** For any vertex $V_j$, let $\deg_i(V_j)$ denote the number of matched points in $V_j$ after adding the first $i$ edges. Since $\Gamma_i(\sigma)$ counts the number of $j \in [n]$ with $\deg_i(V_j) \leq d_j^{(n)} - 1$, using $0 \leq \deg_i(V_j) \leq d_j^{(n)} \leq \Delta$ we infer that

$$\Gamma_i(\sigma) \leq \sum_{j \in [n]} \left[ d_j^{(n)} - \deg_i(V_j) \right] \leq \Delta \cdot \Gamma_i(\sigma).$$

Noting that $\sum_{j \in [n]} d_j^{(n)} = 2m$ and $\sum_{j \in [n]} \deg_i(V_j) = 2i$ then establishes the claimed bounds. \qed

3.2 Definition: clusters of configuration-graphs

In this subsection we introduce sets of configuration-graphs called clusters, whose probabilities we shall compare in our upcoming switching arguments (as mentioned in Section 1.2.2). To motivate their definition, in Section 3.2.1 we develop some intuition about which configuration-graphs are more likely to be produced by the relaxed $d_n$-process. The reader mainly interested in the formal definition of clusters may wish to directly skip to Section 3.2.2.
3.2.1 Intuition about configuration-graphs

In order to discuss our intuition about which configuration-graphs the relaxed $\text{d}_n$-process favors, we first recall the switching setup from Section 1.2. We have a configuration-graph $G^+$ with edges $ab, xy$ that satisfy $\deg(a), \deg(b) \leq k$ and $\deg(x), \deg(y) > k$ (recall that $a, b, x, y$ are points, and that the degree of a point is the size of the vertex that contains it). We let $\deg(a, b, x, y)$ respectively, where we henceforth assume $A \neq B$ and $X \neq Y$. Furthermore, we write $G^-$ for the configuration-graph obtained by replacing the edges $ab, xy$ with $ax, by$; see Figure 1 for an example with $k = 1$.

As discussed in Section 1.2, ideally the relaxed $\text{d}_n$-process should produce $G^+$ with higher probability than $G^-$, which by Section 3.1 means that $Z(G^+) \geq Z(G^-)$. It turns out that this is true for $k = 1$, but not always true for $k \geq 2$. To gain some intuition why, we wish to understand what configuration-graphs $G$ have high values of $Z(G) = \sum_{\sigma \in \Pi G} Z(\sigma)$. By carefully inspecting (21), it is not hard to see that $Z(\sigma)$ tends to be big when the values $\Gamma_i(\sigma)$ tend to be small, i.e., when vertices tend to become saturated early. Furthermore, the impact on $Z(\sigma)$ of a vertex becoming saturated early is amplified if other vertices are also saturated early.

With these ‘early saturation’ observations in mind, we now consider the case $k = 1$, i.e., where $\deg(A) = \deg(B) = 1$ holds. If the vertices $A, B$ form the edge $ab$, then they will both become saturated at the same step, namely the step where that edge appears in the edge-sequence $\sigma$. So in edge-sequences where $A$ is saturated early, we automatically get the amplification of a second vertex becoming saturated at the same step. Of course, in edge-sequences where $A$ is saturated late then $B$ is also saturated late. But it turns out that the multiplicative nature of the formula for $Z$ results in the good edge-sequences outweighing the bad ones. After summing over all edge-sequences, it thus becomes plausible that $Z(G^+) > Z(G^-)$.

For $k \geq 2$, the reasoning is not as clean-cut since the edge $ab$ does not ensure that $A, B$ both become saturated at the same step. Nevertheless, it increases the probability that, when $A$ is saturated then $B$ will be saturated soon (and vice versa). Furthermore, the described increase in probability is intuitively higher for the edge $ab$ than for the edge $ax$, because the degree of $B$ is lower than $X$ (so fewer other points need to be selected until saturation). Hence joining $ab$ seemingly brings a similar, albeit smaller, amplification benefit as doing the same in the case $k = 1$. However, the behavior of the $\text{d}_n$-process turns out to be more complicated: other neighbours of $A, B$ can also significantly impact $Z$, and in certain situations this ‘neighborhood effect’ can even be stronger than the benefit of $A$ being adjacent to $B$. For example, it could be better for $A$ to be adjacent to a vertex $X$ with slightly higher degree than $B$, provided that the neighbors of $X$ have significantly lower degrees than the neighbors of $B$ (since this intuitively gives the amplified benefit of several neighbors of $X$ becoming saturated at around the same step that $A$ and $X$ are). Figure 3 in Appendix A.1 contains an example where this happens: we have $\deg(X) = \deg(B) + 1$, but $X$ is adjacent to more low degree vertices than $B$, and $A$ does not become saturated if $X$ becomes saturated. Note that the described problematic ‘neighborhood effect’ cannot occur for $k = 1$, since when $A, B$ are adjacent, then they have no other neighbors (which partially explains why the case $k = 1$ is more well-behaved; see also Remark 14 in Section 4).

To overcome the discussed obstacle, we have developed a more complicated form of switching. Instead of measuring the effect on $Z$ of switching edges of a single configuration-graph, we will measure the average effect of switching on a set of configuration-graphs, which we call a cluster. These clusters are specifically chosen so that, when averaging over a cluster, the neighbors of $X$ will not be any better than the neighbors of $B$, and so the described problematic ‘neighborhood effect’ will not occur. But the amplifying effect from the edge $ab$ still remains, and so $Z$ really will be larger on average in $G^+$ than in $G^-$ (see Lemma 6 in Section 3.3).

3.2.2 Formal definition of clusters

With an eye on introducing clusters, for any configuration-graph $G$ with a set of points $S$, we write

$$N_G(S) := \{v \text{ point in } G : v \notin S \text{ and there is } w \in S \text{ such that } vw \in E(G)\}$$

for the set of neighbors of $S$. Often $S$ will be specified as a vertex, or a collection of vertices; this means $S$ will be the set of points of these vertices. So $N_G(S)$ will not be the set of vertices adjacent to $S$, but the specific points within those vertices that are matched in $G$ to the points of $S$. We are now ready to formally define clusters, which are illustrated in Figure 3. Recall that $A, B, X, Y$ are distinct vertices with $\deg(A), \deg(B) \leq k$ and $\deg(X), \deg(Y) > k$, and that $a, b, x, y$ are points of $A, B, X, Y$, respectively.
Figure 2: Example of an upper cluster $C^+=C^+(G, ab, xy)$ containing ten configuration-graphs. In this case $N_G(A \cup B \cup X \cup Y) = \{v\}$, and three edges are fixed: $ab, xy$ and the only edge not incident to $A, B, X, Y$. If we start with only these three edges, one can construct any configuration-graph in the cluster $C^+$ by (i) matching $v$ to an unmatched point in $A, B, X, Y$, and then (ii) matching the remaining four unmatched points in $A, B, X, Y$. One can easily check that there are five ways to do step (i), and in each case there are exactly two ways to complete step (ii), leading to all ten depicted configuration-graphs.

**Definition (Clusters).**

(a) $G_{ab,xy}$ is the set of configuration-graphs with degree sequence $d_n$ that contain the two edges $ab, xy$. Similarly, $G_{ax,by}$ is the set of configuration-graphs with degree sequence $d_n$ that contain the two edges $ax, by$.

(b) We define an equivalence relation $\sim$ on $G_{ab,xy}$ and $G_{ax,by}$, where $G_1 \sim G_2$ if $N_{G_1}(A \cup B \cup X \cup Y) = N_{G_2}(A \cup B \cup X \cup Y)$ holds and the set of edges not incident to $A, B, X, Y$ are identical.

(c) The equivalence classes of $\sim$ on $G_{ab,xy}$ are called upper clusters. The equivalence classes of $\sim$ on $G_{ax,by}$ are called lower clusters.

For a configuration-graph $G \in G_{ab,xy}$, we denote by $C^+(G, ab, xy)$ the upper cluster containing $G$ that corresponds to an equivalence class in $G_{ab,xy}$. Similarly for a configuration-graph $G \in G_{ax,by}$, we denote by $C^-(G, ax, by)$ the lower cluster containing $G$ that corresponds to an equivalence class in $G_{ax,by}$. Thus, if $H$ is another configuration-graph in $C^+(G, ab, xy)$ then $C^+(G, ab, xy) = C^+(H, ab, xy)$.

We can think of an upper cluster as follows. Start with any configuration-graph $G$ with edges $ab, xy$. Remove all edges of $G$ with at least one endpoint in $A, B, X, Y$, except for $ab, xy$. Then replace them with any other edges, each having at least one endpoint in $A, B, X, Y$, subject to the final degree-sequence being $d_n$. So if a vertex $Z$ (distinct from $A, B, X, Y$) had $\ell$ neighboring points in $A, B, X, Y$, then $Z$ will still have $\ell$ neighboring points in those vertices. The set of configuration-graphs that can be obtained is the upper cluster; cf. Figure 2. We can think of lower clusters in the same manner, replacing $ab, xy$ with $ax, by$.

**Definition (Switching-partners).** Given $G^+ \in G_{ab,xy}$, let $G^- \in G_{ax,by}$ be the graph obtained by replacing the edges $ab, xy$ in $G^+$ with $ax, by$. There is a natural bijection between configuration-graphs in $C^+ = C^+(G^+, ab, xy)$ and $C^- = C^-(G^-, ax, by)$. Namely, for every configuration-graph in $C^+$, replace the edges $ab, xy$ with $ax, by$. The resulting set of configuration-graphs is the lower cluster $C^-$. We call this pair of clusters $C^+, C^-$ switching-partners.
We remark that comparing the values of $Z$ averaged over the configuration-graphs of two switching-partners (see Lemma 6 below) is the more complicated form of switching that we alluded to in Section 3.2.1. The crux is that in this average, the vertices in $A, B, X, Y$ each have the same potential neighbors outside of $\{A, B, X, Y\}$, which eliminates the problematic ‘neighborhood effect’ described in Section 3.2.1.

3.3 Key lemmas: switching and counting results for clusters

In this subsection we state our main switching results for clusters, along with relevant counting results.

We start with the following switching-type result between upper and lower clusters, whose proof we defer to Section 4.1. Recall that $Z = \sum_{G \in C} Z(G) = \sum_{G \in C} \sum_{\sigma \in \Pi_G} Z(\sigma)$ is proportional to the probability that the relaxed $d_n$-process produces a graph from the cluster $C$, see (22)–(23). So Lemma 6 intuitively states that, up to some local averaging\textsuperscript{11}, the relaxed $d_n$-process prefers the edges $ab, xy$ over the edges $ax, by$, i.e., prefers configuration-graphs with more small edges (which is consistent with Remark 4 and the heuristics from Section 1.1).

**Lemma 6 (Switching of clusters).** For all switching-partners $C^+, C^-$, we have $Z(C^+) \geq Z(C^-)$.

We would like to strengthen the conclusion of Lemma 6 to $Z(C^+) \geq (1 + \epsilon)Z(C^-)$ for some $\epsilon > 0$, but for technical reasons we only prove it for clusters with special properties, defined below as good clusters (which suffices, since a significant proportion of clusters turns out to be good). The main difficulty in proving it for all clusters is that when the vertex $A$ or $B$ become saturated during the final few steps of the relaxed $d_n$-process, the number of unsaturated vertices $\Gamma_i$ is so small that a small (additive) change to it could have a significant effect on $Z(\sigma)$; cf. equation (21) and Observation 5. The definition of good clusters will intuitively ensure that such ‘late saturation effects’ have limited impact on our arguments. Recall that $G_{ab, xy}$ is the set of configuration-graphs with degree sequence $d_n$ that contain the edges $ab, xy$, see Section 3.2.2.

**Definition (Good edge-sequences and clusters).** An edge-sequence $\sigma$ of $G \in G_{ab, xy}$ is called good with respect to $a, b, x, y$ if the following two conditions hold:

(i) no points of $X, Y$ are adjacent to any points of $A, B$ in $G$, and
(ii) the vertices $A$ and $B$ are both saturated in the first $(1 - \zeta)m$ steps, where

$$\zeta := \frac{\xi^2}{16\Delta^3}. \quad (25)$$

An upper cluster $C^+ = C^+(G, ab, xy)$ is called good if at least a sixteenth of its total $Z$-value comes from good edge-sequences $\sigma$, i.e., if

$$\sum_{G \in C^+} \sum_{\sigma \in \Pi_G: (\sigma, ab, xy) \in G} Z(\sigma) \geq \frac{1}{16} Z(C^+), \quad (26)$$

where

$$\mathcal{G} := \{ (\sigma, ab, xy) : \sigma \text{ is good with respect to } a, b, x, y \}. \quad (27)$$

The next lemma, whose proof we defer to Section 4.1, shows that for good clusters we can indeed strengthen the switching-type estimate of Lemma 6 by an extra factor of $1 + \epsilon$.

**Lemma 7 (Switching of good clusters).** There are $\epsilon, n_0 > 0$ such that the following holds for all $n \geq n_0$: if $C^+, C^-$ are switching-partners and $C^+ = C^+(G, ab, xy)$ is good, then we have

$$Z(C^+) \geq (1 + \epsilon)Z(C^-).$$

Recall that $X_k(G)$ is the number of small edges in a configuration-graph $G$, and that $\mu$ is approximately the expected number of small edges in a uniform configuration-graph with degree sequence $d_n$. Our main

\textsuperscript{11}Since $C^+$ and $C^-$ have the same cardinality, Lemma 6 shows that $\sum_{G \in C^+} Z(G)/|C^+| \geq \sum_{G \in C^-} Z(G)/|C^-|$, which in concrete words means that $Z(G)$ averaged over $G \in C^+$ is at least as large as $Z(G)$ averaged over $G \in C^-$. 11
technical result Theorem 3 bounds the probability that, in the graph chosen by the relaxed $d_n$-process, the number of small edges has a small (but significant) distance from $\mu$. We thus define

$$N_z := \{G : \text{ configuration-graph with degree sequence } d_n \text{ and } |X_k(G) - \mu| \leq z\}. \quad (28)$$

and henceforth focus on the configuration-graphs $G \in N_{\mu}$ where $\gamma > 0$ is a constant. For such $G$, different choices for $a, b, x, y$ can yield different upper clusters $C^+(G, ab, xy)$, and similarly for lower clusters. Note that while switching-partners form a bijection upon specifying $a, b, x, y$, each configuration-graph can have different switching-partners for different choices of those four vertices. One particularly noteworthy example arises when we exchange $x, y$. Note that $C^+(G, ab, xy) = C^+(G, ab, yx)$, but $C^-(G, ax, by)$ and $C^-(G, ay, bx)$ are different. Similarly when we exchange $a, b$. This could potentially create a problem with our definition of switching-partners, since the switching-partner of $C^+(G^+, ab, xy)$ is defined to be $C^-(G^-, ax, by)$, while the switching-partner of $C^+(G^+, ab, yx)$ is defined to be $C^-(G^-, ay, bx)$. To account for this problem we specify the following convention, which will be key for the counting result Lemma 8 below (as well as Observation 10 in Section 5).

Notational Convention. $C^+(G, ab, xy), C^+(G, ab, yx), C^+(G, ba, xy), C^+(G, ba, yx)$ are considered to be four different clusters, despite the fact that they are identical. Similarly $C^-(G, ax, by), C^-(G, by, ax)$ are considered to be two different clusters and so are $C^-(G, ay, bx), C^-(G, bx, ay)$. On the other hand, $C^-(G, ax, by)$ and $C^-(G, xa, by)$ are considered the same; note that this does not lead to any problems in our definition of switching-partners, since we list the edge as “$ax$” rather than “$xa$” in our definitions of lower clusters and switching-partners.

When we discuss the number of clusters containing $G$, we mean the number of choices of $a, b, x, y$ such that $G$ lies in a cluster $C^+(G, ab, xy)$ or $C^-(G, ax, by)$; i.e., the number of choices of $a, b, x, y$ such that $E(G)$ contains $ab, xy$ or $ax, by$. From the notational convention above, we see that clusters obtained by permuting the order of $a, b$ or $x, y$ are considered to be distinct, even if they contain the same set of configuration-graphs (e.g., $C^+(G, ab, xy)$ and $C^+(G, ab, yx)$ are counted as two different clusters). With this convention in mind, for any configuration-graph $G$, we define $U_G$ to be the number of upper clusters that contain $G$. We similarly define $L_G$ to be the number of lower clusters that contain $G$.

The next lemma, whose proof we defer to Section 5.2, shows that for any configuration-graph with approximately $\mu$ small edges, the number of upper and lower clusters containing it are roughly equal for which it is intuitively important that switching-partners form a bijection, as discussed above).

Lemma 8 (Counting clusters). There are $\gamma_0, n_0, D > 0$ such that the following holds for all $0 < \gamma < \gamma_0$ and $n \geq n_0$: we have $|L_G/U_H - 1| \leq D\gamma$ and $L_G, U_H \geq 1$ for all configuration-graphs $G, H \in N_{\gamma_0}$.

For technical reasons we shall also need the following lemma, whose proof we defer to Section 5.2. It intuitively says that, for any edge-sequence $\sigma$ of a configuration-graph $G$ with approximately $\mu$ small edges, many upper clusters $C^+ = C^+(G, ab, xy)$ containing $G$ yield a good edge-sequence $(\sigma, ab, xy) \in \mathcal{G}$, where $\mathcal{G}$ is defined as in (27).

Lemma 9 (Counting upper clusters yielding good edge-sequences). There are $\gamma, n_0 > 0$ such that the following holds for all $n \geq n_0$: for all configuration-graphs $G \in N_{\gamma_0}$ and edge-sequences $\sigma \in \Pi_G$ we have

$$|\{C^+ = C^+(G, ab, xy) : (\sigma, ab, xy) \in \mathcal{G}\}| \geq \frac{1}{8} U_G. \quad (29)$$

We close this subsection with an observation about the number of small edges in switching-partners.

Observation 10 (Small edges in switching-partners). For any pair of switching-partners $C^+, C^-$ the following holds: if $C^-$ contains any configuration-graph $G \in N_z$, then $C^+ \subseteq N_{z+4\Delta}$.

Proof. By construction, any two configuration-graphs in $C^+ \cup C^-$ can differ in at most $4\Delta$ edges. Therefore the number of small edges of any configuration-graph in $C^+$ can differ by at most $4\Delta$ from the number of small edges of $G \in N_z$, which in turn establishes $C^+ \subseteq N_{z+4\Delta}$. \qed
3.4 Double-counting argument: proof of Theorem 3

This subsection is devoted to the proof of our main technical result Theorem 3, and we start by outlining our proof strategy. Our main step will be to prove that, when \( z \) is sufficiently small, decreasing \( z \) by an additive constant will decrease \( Z(N_z) \) be a multiplicative constant. Specifically, we will prove that the following key inequality holds (for all sufficiently large \( n \)): there is a constant \( \tau \in (0, 1) \) such that

\[
Z(N_z) \leq (1 - \tau)Z(N_{z+5\Delta}) \quad \text{for all } 0 \leq z \leq \gamma\mu, \tag{30}
\]

where \( \gamma \) comes from Lemmas 8 and 9 (and the usage of the additive constant \( 5\Delta \) is closely related to Observation 10, as we shall see). Iterating (30) implies that decreasing \( z \) by \( \Theta(n) \) will decrease \( Z(N_z) \) by a factor that is exponentially small in \( \Theta(n) \). In particular, setting \( \alpha := \gamma/2 \) and noting that \( \mu \geq (\xi n)^2/(2\Delta n) = \Theta(n) \), we thus readily obtain that

\[
Z(N_{\alpha µ}) \leq e^{-\Theta(n)} \cdot Z(N_{2\alpha µ}).
\]

Since \( \mathbb{P}(G_{d_n}^{\ell_2} \in N_z) \) is proportional to \( Z(N_z) \), see Section 3.1, this immediately implies Theorem 3.

Our strategy for proving the key inequality (30) focuses on clusters. Recall that each configuration-graph \( G \) lies in \( L_G \) lower clusters and \( U_G \) upper clusters. Lemma 8 tells us that the values of \( L_G, U_G \) are very close over all configuration-graphs \( G \in N_{\gamma\mu} \); close enough that (30) will follow from proving

\[
\sum_{G \in N_z} L_G Z(G) \leq (1 - 2\tau) \sum_{G \in N_{z+5\Delta}} U_G Z(G), \tag{31}
\]

and taking \( \gamma \) to be sufficiently small in terms of \( \tau \). For technical reasons, we then define a suitable set \( \mathcal{L} \) of lower clusters \( C^- \) with the property that \( N_z \subseteq \bigcup_{C^- \in \mathcal{L}} C^- \). Furthermore, we let \( \mathcal{U} \) contain all upper clusters \( C^+ \) that are switching-partners of the lower clusters \( C^- \) in \( \mathcal{L} \), which will turn out to satisfy \( \bigcup_{C^+ \in \mathcal{U}} C^+ \subseteq N_{z+5\Delta} \) by Observation 10. Hence inequality (31) would be implied by

\[
\sum_{C^- \in \mathcal{L}} Z(C^-) \leq (1 - 2\tau) \sum_{C^+ \in \mathcal{U}} Z(C^+). \tag{32}
\]

Since \( \mathcal{U} \) consists of the switching-partners of the lower clusters in \( \mathcal{L} \), Lemma 8 implies the following weaker bound: \( \sum_{C^- \in \mathcal{L}} Z(C^-) \leq \sum_{C^+ \in \mathcal{U}} Z(C^+) \). If a positive proportion of the clusters in \( \mathcal{U} \) are good, then Lemma 7 allows us to improve this bound to obtain (32) and thus inequality (31). In the remaining case where rather few clusters of \( \mathcal{U} \) are good, inequality (32) might not hold. But it turns out that, using Lemma 9, we are able to compare \( \sum_{C^+ \in \mathcal{U}} Z(C^+) \) with \( \sum_{G \in N_{\gamma\mu}} L_G Z(G) \) via the contributions of good edge-sequences (and here the careful definition of \( \mathcal{L} \) in (35) below with respect to \( N_{z+1} \)) rather than \( N_z \) will matter). This will allow us to complete the proof of inequality (31) by a technical case analysis; see (38)–(42) below for the details.

**Proof of Theorem 3.** As mentioned in the above proof outline, our main goal is to prove inequality (30). Indeed, by iterating this key inequality, in view of Section 3.1 and \( \mu = \Theta(n) \) it then follows that

\[
\mathbb{P}(|X_k(G_{d_n}^{\ell_2}) - \mu| \leq \gamma\mu/2) \leq \frac{\mathbb{P}(|X_k(G_{d_n}^{\ell_2}) - \mu| \leq \gamma\mu/2 + 5\Delta \cdot \gamma\mu/(10\Delta))}{Z(N_{\gamma\mu}(2\Delta \cdot \gamma\mu/(10\Delta)))} \cdot (1 - \tau)^{(\gamma\mu/(10\Delta))} \leq e^{-\Theta(n)}, \tag{33}
\]

which completes the proof with \( \alpha := \gamma/2 \).

It thus remains to prove inequality (30). To this end we will henceforth tacitly assume that \( 0 \leq z \leq \gamma\mu \), and also that \( \gamma > 0 \) is sufficiently small (whenever necessary). We shall first switch the focus of inequality (30) from configuration-graphs to clusters, leveraging that any \( G \in N_{\gamma\mu} \) is contained in roughly the same number of clusters. Indeed, recalling the definition (23) of \( Z(N_z) \), Lemma 8 implies that

\[
\frac{Z(N_z)}{Z(N_{z+5\Delta})} = \frac{\sum_{G \in N_z} Z(G)}{\sum_{G \in N_{z+5\Delta}} Z(G)} \leq (1 + O(\gamma)) \cdot \frac{\sum_{G \in N_{z+5\Delta}} L_G Z(G) - \sum_{G \in N_{z+5\Delta}} U_G Z(G)}{U_G Z(G)}, \tag{34}
\]
where the implicit constant in $O(\gamma)$ is uniform i.e., does not depend on $\gamma$. The next step is to further rewrite the right-hand side of inequality (34) in terms of clusters, for which we define

$$
\mathcal{L} := \{ C^- : \text{lower cluster with } C^- \cap \mathcal{N}_{z+1} \neq \emptyset \},
\mathcal{U} := \{ C^+ : \text{upper cluster that is switching partner of } C^- \in \mathcal{L} \}.
$$

(35)

Since every $G \in \mathcal{N}_{z+1}$ lies in at least one lower cluster by Lemma 8, we have $\mathcal{N}_z \subseteq \mathcal{N}_{z+1} \subseteq \bigcup_{C^- \in \mathcal{L}} C^-$. Observation 10 (applied with $z$ replaced by $z+1$) implies $\bigcup_{C^+ \in \mathcal{U}} C^+ \subseteq \mathcal{N}_{z+1} \subseteq \mathcal{N}_{z+5\Delta}$. It follows that

$$
\sum_{G \in \mathcal{N}_z} L_G Z(G) \leq \sum_{C^- \in \mathcal{L}} Z(C^-) =: Z(\mathcal{L}),
\sum_{G \in \mathcal{N}_{z+5\Delta}} U_G Z(G) \geq \sum_{C^+ \in \mathcal{U}} Z(C^+) =: Z(\mathcal{U}).
$$

(36)

With the cluster based inequalities from (36) in hand, we are now in a position to estimate the right-hand side of (34) from above. Let $\mathcal{G}$ denote the set of good clusters, where good clusters are defined around (26). Decomposing $Z(\mathcal{L})$ into two parts, by invoking Lemmas 6 and 7 separately for $C^+ \in \mathcal{G}$ and $C^+ \in \mathcal{G}^c$ it follows that

$$
Z(\mathcal{L}) = \sum_{C^- : C^+ \in \mathcal{U} \cap \mathcal{G}} Z(C^-) + \sum_{C^- : C^+ \in \mathcal{U} \cap \mathcal{G}^c} Z(C^-)
\leq \sum_{C^+ \in \mathcal{U} \cap \mathcal{G}} Z(C^+) + \frac{1}{1+\epsilon} \sum_{C^+ \in \mathcal{U} \cap \mathcal{G}^c} Z(C^+)
= Z(\mathcal{U}) - \frac{\epsilon}{1+\epsilon} \sum_{C^+ \in \mathcal{U} \cap \mathcal{G}^c} Z(C^+),
$$

(37)

where in the summations $C^-$ denotes the switching-partner of $C^+$ (defined by the bijection discussed earlier).

We now distinguish two cases. If $\sum_{C^+ \in \mathcal{U} \cap \mathcal{G}} Z(C^+) \geq Z(\mathcal{U})/32$, then inequality (37) implies

$$
Z(\mathcal{L}) \leq (1-\nu) Z(\mathcal{U})
$$

(38)

for $\nu := \epsilon/[32(1+\epsilon)] \in (0,1)$, say.

Otherwise $\sum_{C^+ \in \mathcal{U} \cap \mathcal{G}} Z(C^+) \leq Z(\mathcal{U})/32$ holds, in which case we shall focus on the contribution of good edge-sequences to $Z(\mathcal{U})$. Using the definition of good upper clusters (see page 11), it follows that the contribution of good edge-sequences is at most

$$
\sum_{C^+ \in \mathcal{U}} \sum_{ \sigma \in \Pi_{G^c} \cap \mathcal{G}^c \cap \{ (\sigma,ab,xy) \} } Z(\sigma) \leq \sum_{C^+ \in \mathcal{U}} \left(1_{(C^+ \in \mathcal{G})} Z(C^+) + 1_{(C^+ \in \mathcal{G}^c)} \frac{1}{16} Z(C^+) \right)
$$

$$
= \sum_{C^+ \in \mathcal{U} \cap \mathcal{G}} Z(C^+) + \frac{1}{16} \sum_{C^+ \in \mathcal{U} \cap \mathcal{G}^c} Z(C^+) \leq \frac{3}{32} Z(\mathcal{U}),
$$

(39)

where the edges $ab,xy$ in the third summation are determined by the corresponding upper cluster $C^+ = C^+(G,ab,xy)$. To bound the contribution of good edge-sequences from below, we next analyze the the left-hand side of (39) more carefully. Lemma 8 implies that every $G^+ \in \mathcal{N}_z$ is contained in at least one upper cluster. We now claim that any upper cluster $C^+$ containing some $G^+ \in \mathcal{N}_z$ satisfies $C^+ \in \mathcal{U}$. To see this, note that by construction its switching-partner $G^- \in C^-$ (see page 10) contains $X_k(G^-) = X_k(G^+) - 1$ small edges, which in view of (28) implies $G^- \in \mathcal{N}_{z+1}$. Recalling the definition (35) of $\mathcal{L}$ it follows that $C^- \in \mathcal{L}$, which by construction gives $C^+ \in \mathcal{U}$, as claimed. This in particular implies $\mathcal{N}_z \subseteq \bigcup_{C^+ \in \mathcal{U}} C^+$, and using Lemma 8 it follows that

$$
\sum_{C^+ \in \mathcal{U}} \sum_{ \sigma \in \Pi_{G^c} \cap \mathcal{G}^c \cap \{ (\sigma,ab,xy) \} } Z(\sigma) \geq \sum_{G \in \mathcal{N}_z} \sum_{\sigma \in \Pi_G} Z(\sigma) \cdot \left| \{ C^+ = C^+(G,ab,xy) : (\sigma,ab,xy) \in \mathcal{G} \} \right|
\geq \sum_{G \in \mathcal{N}_z} Z(G) \cdot \frac{1}{8} U_G.
$$

(40)
Combining (40) and (39), using Lemma 8 we infer that
\[
Z(\mathcal{U}) \geq \frac{4}{3} \sum_{G \in \mathcal{N}_z} U_G Z(G) \geq (1 - O(\gamma)) \cdot \frac{4}{3} \sum_{G \in \mathcal{N}_z} L_G Z(G).
\]
(41)

Finally, by combining the two above-discussed cases with (36), using inequalities (38) and (41) it follows that we always have
\[
\frac{\sum_{G \in \mathcal{N}_z} L_G Z(G)}{\sum_{G \in \mathcal{N}_z} U_G Z(G)} \leq \min \left\{ \frac{Z(\mathcal{L})}{Z(\mathcal{U})}, \frac{\sum_{G \in \mathcal{N}_z} L_G Z(G)}{\sum_{G \in \mathcal{N}_z} U_G Z(G)} \right\} \leq \max \left\{ 1 - \nu, \frac{1}{4/3 - O(\gamma)} \right\},
\]
(42)
which together with estimate (34) establishes inequality (30) for sufficiently small \( \gamma > 0 \).

4 Switching results: deferred proofs

This section is devoted to the deferred proofs of the ‘cluster switching’ results Lemma 6 and 7 from Section 3.3. To avoid clutter, we henceforth always tacitly assume that \( n \geq n_0(\Delta, \xi) \) is sufficiently large (whenever necessary). Furthermore, recalling that there is a natural bijection between the upper clusters and the lower clusters, for the rest of the section we fix an upper cluster \( C^+ = C^+(G^+, ab, xy) \) and its switching-partner lower cluster \( C^- = C^-(G^-, ax, by) \), as defined in Observation 3.2.2. We also define
\[
\mathcal{P}^- := \bigcup_{G \in C^-} \Pi_G \quad \text{and} \quad \mathcal{P}^+ := \bigcup_{G \in C^+} \Pi_G
\]
(43)
to be the set of edge-sequences of configuration-graphs that lie in \( C^- \) and \( C^+ \), respectively. Thus, if \( \sigma \in \mathcal{P}^+ \) then \( \sigma \) contains the edges \( ab, xy \), while if \( \sigma \in \mathcal{P}^- \) then \( \sigma \) contains the edges \( ax, by \).

In our upcoming switching based proofs it will be important to understand when certain points are chosen. More specifically, at any point in our proof, when we are comparing two edge-sequences the step at which any vertex beside \( A, B, X, Y \) saturates (to clarify: we say that vertex \( A \) saturates at step \( i \), if \( A \) is saturated for the first time after the first \( i \) edges are added) will be the same in both edge-sequences (where \( A, B, X, Y \) denote the vertices that contain the points \( a, b, x, y \), as usual). This allows us to focus on the step where points of \( A, B, X, Y \) are chosen. For any edge-sequence \( \sigma = (e_1, \ldots, e_m) \) of the edges of a configuration-graph \( G \), we let
\[
\sigma(e) = \{1, \ldots, m\}
\]
(44)
be the unique edge-number such that \( e = e_{\sigma(e)} \). Similarly, for any point \( v \) of \( G \), we let \( \sigma(v) = \{1, \ldots, m\} \) be the unique edge-number such that \( v \in e_{\sigma(v)} \). For our analysis, we will need to focus on the last step at which a point of the vertex \( A \) besides \( a \) is chosen in \( \sigma \), which we formally denote as
\[
t_a = t_a(\sigma) := \begin{cases} \max \{ \sigma(a') : a' \in A \setminus \{a\} \} & \text{if } |A| > 1, \\ 0 & \text{otherwise}. \end{cases}
\]
(45)
The crux will be that \( \sigma(a) \) and \( t_a \) together determine the step \( \max \{ \sigma(a), t_a \} \) at which vertex \( A \) saturates in \( \sigma \), which will be important for understanding the contributions of the vertex \( A \) to \( Z(\sigma) \). We define \( t_b, t_x, t_y \) analogously to \( t_a \).

With the notation above, Lemma 6 can be stated as
\[
\sum_{\sigma \in \mathcal{P}^+} Z(\sigma) \geq \sum_{\sigma \in \mathcal{P}^-} Z(\sigma).
\]
(46)
Our proof strategy for inequality (46) is as follows: in Section 4.3 we will construct pairs of edge-sequences \( (\sigma_1, \sigma_2) \in \mathcal{P}^+ \times \mathcal{P}^- \), called twins, which will satisfy \( Z(\sigma_1) = Z(\sigma_2) \). Hence (46) holds with equality when summing only over those paired edge-sequences. The unpaired \( \sigma \in \mathcal{P}^+ \cup \mathcal{P}^- \) will have the very useful property that \( t_a \leq t_y \) and \( t_b \leq t_x \). In such case the edge \( ab \), especially if located early in the sequence, will help to saturate \( A \) and \( B \) early on. This suggests that \( Z(\sigma) \) should typically have higher values when \( ab \) is in \( \sigma \),
i.e., $\sigma \in \mathcal{P}^\dagger$. Using this property, we will see that \([46]\) also holds when summing only over those unpaired members. Combining these estimates will eventually establish \([46]\) and thus Lemma \(6\) see Section \(4.1\).

Recall that two configuration-graphs $G^+ \in C^+$ and $G^- \in C^-$ are switching-partners if $G^-$ is obtained from $G^+$ by replacing edges $ab, xy$ with $ax, by$, see Section \(3.2.2\). This notion extends naturally to edge-sequences of configuration-graphs that lie in $C^+$ and $C^-$, respectively.

**Definition (Counterparts).** We say that the edge-sequences $\sigma \in \mathcal{P}^+$ and $\sigma' \in \mathcal{P}^-$ are counterparts if $\sigma'$ is obtained from $\sigma$ by replacing $ab$ with $ax$ and $xy$ with $by$.

Note that two counterparts have the same values of $t_a, t_b, t_x$ and $t_y$; this is one of the motivations for defining those values. Note also that the vertices $A, Y$ and $B, X$ become saturated at the same step of two counterparts, but that $B$ and $X$ may become saturated at different steps. This means that, in general, $Z(\sigma)$ differs from $Z(\sigma')$. When defining pairs of twins in Section \(4.3\), we will account for this by rearranging some of the edges involving $b, x$ so that $t_b, t_x$ exchange values, while $t_a, t_y$ remain the same. As a result, every vertex except $B$ and $X$ becomes saturated at the same step in $\sigma$ and in its twin. Furthermore, the step at which $B$ and $X$ saturates will swap between twins. Therefore, the twins will both have the same $Z$-value, as desired; see Section \(4.3\) for the details.

Besides replacing edges as in counterparts, we also need an operation that changes the relative order of appearance of the switching-relevant edges. First, for $\sigma \in \mathcal{P}^+$, we define $\overline{\sigma}$ to be the edge-sequence obtained from $\sigma$ by swapping the positions of $ab$ and $xy$. Similarly, for $\sigma \in \mathcal{P}^-$ we define $\overline{\sigma}$ to be the edge-sequence obtained from $\sigma$ by swapping the positions of $ax$ and $by$. For later reference, we now record some basic properties of these natural operations on edge-sequences (whose routine verification we omit).

**Observation 11.** For all counterparts $\sigma \in \mathcal{P}^+$ and $\sigma' \in \mathcal{P}^-$ the following properties hold:

(i) The edge-sequences $\sigma$ and $\sigma'$ are counterparts.

(ii) The values $t_a, t_b, t_x, t_y$ are the same for $\sigma, \sigma', \overline{\sigma}, \overline{\sigma'}$.

(iii) The map $\sigma \mapsto \overline{\sigma}$ is a bijection between $\mathcal{P}^+$ and $\mathcal{P}^+$, and a bijection between $\mathcal{P}^-$ and $\mathcal{P}^-$.

In the following lemma, $\mathcal{T}$ denotes the set of edge-sequences which have a twin; see Section \(4.3.4\). As indicated above, an edge-sequence $\sigma \in \mathcal{T}$ and its twin will have the same $Z$-value, which allows us to obtain part \((i)\) in the following lemma (the technical property from part \((iii)\) was also discussed above).

**Lemma 12 (Properties of twins).** There exists $\mathcal{T} \subseteq \mathcal{P}^+ \cup \mathcal{P}^-$ satisfying the following properties:

(i) $\sum_{\sigma \in \mathcal{T} \cap \mathcal{P}^+} Z(\sigma) = \sum_{\sigma \in \mathcal{T} \cap \mathcal{P}^-} Z(\sigma)$.

(ii) For all counterparts $\overline{\sigma}, \overline{\sigma'}$ we have: $\sigma \in \mathcal{T}$ iff $\overline{\sigma} \in \mathcal{T}$ iff $\sigma' \in \mathcal{T}$ iff $\overline{\sigma'} \in \mathcal{T}$.

(iii) If $\sigma \not\in \mathcal{T}$ then $t_a \leq t_y$ and $t_b \leq t_x$ in $\sigma$.

(iv) If in $\sigma \in \mathcal{P}^+ \cup \mathcal{P}^-$ all edges incident to $X, Y$ appear after all edges incident to $A, B$, then $\sigma \not\in \mathcal{T}$.

Recall that in order to prove Lemma \(6\) our strategy is to first sum over edge-sequences with a twin, for which we shall invoke part \((i)\) of Lemma \(12\). When summing over the remaining edge-sequences with no twin, which we call non-twins, in Section \(4.1\) we will invoke part \((i)\) of the following lemma (which applies due to part \((iii)\) of Lemma \(12\)). For Lemma \(7\) we will use a variant of this strategy, exploiting the stronger conclusion of part \((ii)\) of Lemma \(13\) in order to obtain the desired extra $1 + \epsilon$ factor.

**Lemma 13 (Switching of edge-sequences).** There is $\epsilon' = \epsilon'(\zeta) > 0$ such that, for all counterparts $\sigma \in \mathcal{P}^+$ and $\sigma' \in \mathcal{P}^-$ satisfying $t_a \leq t_y$ and $t_b \leq t_x$, the following holds:

(i) $Z(\sigma) + Z(\overline{\sigma}) \geq Z(\sigma') + Z(\overline{\sigma'})$.

(ii) If $\min\{\sigma(xy), t_x, t_y\} - \max\{\sigma(ab), t_a, t_b\} \geq \zeta m/3$, then $Z(\sigma) + Z(\overline{\sigma}) \geq (1 + \epsilon')[Z(\sigma') + Z(\overline{\sigma'})]$.

Before giving the proofs of the above two auxiliary lemmas, we shall first use both to prove the ‘cluster switching’ results Lemmas \(6\) and \(7\) in Section \(4.1\). Afterwards, in Section \(4.2\) we use edge-switching arguments to prove Lemma \(13\) and in Section \(4.3\) we then define twins and prove Lemma \(12\).

**Remark 14 (Special case $k = 1$).** In the case $k = 1$ we have $t_a = t_b = 0$, and so every pair of counterparts $\sigma, \sigma'$ satisfies $t_a \leq t_y$ and $t_b \leq t_x$, and thus Lemma \(13\) applies. Hence there is no need for Lemma \(12\) or even the concept of twins. This is one place where the case $k = 1$ is much simpler than the general case.
4.1 Switching of clusters: proofs of Lemmas 6 and 7

To prove the basic cluster switching result Lemma 6 we shall invoke Lemma 12 for twins \( \sigma \) and Lemma 13 for non-twins \( \sigma \), as discussed.

**Proof of Lemma 6.** Recalling Observation 11 (iii) as well as twins properties (ii) and (iii) from Lemma 12, using part (i) of Lemma 13 it follows that non-twins satisfy

\[
\sum_{\sigma \in P^+ \setminus T} Z(\sigma) = \frac{1}{2} \sum_{\sigma \in P^+ \setminus T} [Z(\sigma) + Z(\sigma')] \geq \frac{1}{2} \sum_{\sigma \in P^- \setminus T} [Z(\sigma) + Z(\sigma)] = \sum_{\sigma \in P^- \setminus T} Z(\sigma).
\]

Combining this inequality with part (i) of Lemma 12 for twins, we infer that

\[
Z(C^+) = \sum_{\sigma \in P^+ \cap T} Z(\sigma) + \sum_{\sigma \in P^+ \cap S^c} Z(\sigma) \geq \sum_{\sigma \in P^- \cap T} Z(\sigma) + \sum_{\sigma \in P^- \cap S^c} Z(\sigma) = Z(C^-),
\]

as claimed. \( \square \)

Turning to the proof of the stronger switching result Lemma 7 for good clusters \( C^+ = C^+(G, ab, xy) \), to work with good edge-sequences in \( G \in C^+ \) it is convenient to define the shorthand

\[
G_{ab, xy} := \{ \sigma : (\sigma, ab, xy) \in G \},
\]

where \( G \) is defined as in (27). Our strategy will be to construct a ‘super nice’ subset \( S \subseteq P^+ \cap G_{ab, xy} \cap T^c \) of non-twins that satisfy the ‘gap’ hypothesis of Lemma 13 (ii). In comparison with the above proof of Lemma 6 this eventually allows us to gain the desired extra \( 1 + \epsilon \) factor by invoking the stronger conclusion of Lemma 13 (ii) for non-twins \( \sigma \in S \).

**Proof of Lemma 7.** Let \( S \subseteq P^+ \) denote the set of edge-sequences \( \sigma \in P^+ \) with the following two properties: (a) that \( \min \{\sigma(xy), t_x, t_y\} - \max \{\sigma(ab), t_a, t_b\} \geq \zeta m/3 \) holds, and (b) that all edges incident to \( A, B \) appear after all edges incident to \( A, B \). By property (iv) of twins (see Lemma 12) it follows that \( S \subseteq T^c \). Recalling that the map \( \sigma \mapsto \sigma' \) is a bijection between \( P^+ \) and \( P^- \) (see Observation 11), we shall now decompose \( Z(C^-) = \sum_{\sigma \in C^-} Z(\sigma') \) into several parts, and apply different inequalities to each part. Recalling Observation 11 and twins property (ii) by invoking parts (i) and (ii) of Lemma 13 separately for \( \sigma \in S^c \) and \( \sigma \in S \), together with Lemma 12 it follows that there is \( \epsilon' = \epsilon'(\zeta) > 0 \) such that

\[
Z(C^-) = \frac{1}{2} \sum_{\sigma \in S} [Z(\sigma') + Z(\sigma)] + \frac{1}{2} \sum_{\sigma \in P^+ \cap T \cap S^c} [Z(\sigma') + Z(\sigma)] + \sum_{\sigma \in P^+ \cap T} Z(\sigma') \\
\leq \frac{1}{1+\epsilon'} \frac{1}{2} \sum_{\sigma \in S} [Z(\sigma) + Z(\sigma)] + \frac{1}{2} \sum_{\sigma \in P^+ \cap T \cap S^c} [Z(\sigma) + Z(\sigma)] + \sum_{\sigma \in P^+ \cap T} Z(\sigma) \quad (48)
\]

It thus suffices to prove that \( \sum_{\sigma \in S} Z(\sigma) \geq \rho Z(C^+) \) for some \( \rho = \rho(\zeta, \Delta) > 0 \). Indeed, for \( \epsilon := \epsilon'/(1+\epsilon') \cdot \rho/2 \) we then infer the desired result \( Z(C^+) \geq (1+\epsilon)Z(C^-) \) by noting that

\[
Z(C^-) \leq (1-\epsilon)Z(C^+) \leq Z(C^+)/ (1+\epsilon). \quad (49)
\]

In the remaining proof of \( \sum_{\sigma \in S} Z(\sigma) \geq \rho Z(C^+) \), we shall work with a fixed lexicographic ordering of all possible edges. The following mapping \( f : P^+ \cap G_{ab, xy} \rightarrow S \) is at the heart of our argument. Given an edge-sequence \( \sigma \in P^+ \cap G_{ab, xy} \), we write \( e_1, \ldots, e_\ell \) for an enumeration of all edges incident to \( X, Y \) (including \( xy \)) according to our lexicographic ordering. For \( i = 1, \ldots, \ell \) we now sequentially proceed as follows: if the edge \( e_i \) is not in the last \( \frac{2}{3}\zeta m - 2\Delta \) steps of our current edge-sequence, then we repeatedly shift it back by exactly \( \lfloor \frac{1}{4}\zeta m \rfloor \) steps until it is. Writing \( f(\sigma) \) for the edge-sequence resulting from this shifting process, we define

\[
S_f := \{ f(\sigma) : \sigma \in P^+ \cap G_{ab, xy} \}. \quad (50)
\]
We claim that $S_f \subseteq S$. Fix $\sigma^* \in S_f$, where by construction $\sigma^* = f(\sigma)$ for some $\sigma \in \mathcal{P}^+ \cap \mathcal{G}_{ab,xy}$. Note that $\sigma \in \mathcal{G}_{ab,xy}$ implies $\sigma^* \in \mathcal{G}_{ab,xy}$, and that the edges $e_1, \ldots, e_\ell$ all appear in the last $\frac{3}{2}\zeta m$ steps of the new edge-sequence $\sigma^*$ (since $\ell \leq 2\Delta$). Furthermore, since $A, B$ are not adjacent to $X, Y$ in $\sigma \in \mathcal{P}^+ \cap \mathcal{G}_{ab,xy}$ by definition of good clusters, it follows that $A, B$ are saturated in the first $(1-\zeta)\Delta$ steps in $\sigma$ and thus $\sigma^*$ (since we do not shift edges incident to $A, B$). Putting things together, we infer that $\sigma^*$ satisfies both properties (a) and (b). Therefore $\sigma^* \in S$, establishing the claim that $S_f \subseteq S$.

We now bound the preimage size $|f^{-1}(\sigma^*)|$ for $\sigma^* \in S_f$. Note that $\sigma$ and $\sigma^* = f(\sigma)$ contain the same sets of edges, so from $\sigma^*$ we can uniquely recover the lexicographic enumeration of the edges $e_1, e_2, \ldots, e_\ell$ incident to $X, Y$. The crux is that if we are given $\sigma^*$ and the number of times each edge $e_i$ is shifted backward (which is at most $\lceil m/(1/2\zeta m) \rceil$ by construction), then we can uniquely recover $\sigma$. Since $\ell \leq 2\Delta$, it follows that

$$|f^{-1}(\sigma^*)| \leq \left( \left\lceil \frac{m}{1/2\zeta m} \right\rceil \right)^\Delta \leq \left( \frac{6}{7} \right)^{2\Delta} .$$

(51)

We next bound the ratio $Z(\sigma^*)/Z(\sigma)$ for $\sigma^* = f(\sigma)$. Since the shifted edges are shifted back by $\lceil \frac{1}{3}\zeta m \rceil$ steps until they are in the last $\frac{3}{2}\zeta m - 2\Delta$ steps, it follows that none of them are shifted back into the last $\frac{3}{2}\zeta m - 2\Delta$ steps. As a consequence we have $\Gamma_i(\sigma) = \Gamma_i(\sigma^*)$ for $i > \lceil (1 - \zeta/3)m \rceil + 2\Delta$, since $\sigma, \sigma^*$ are identical in the last $2\Delta$ steps of edges, so from $\sigma, \sigma^*$ we do not shift edges incident to $X, Y$. The crux is that if we are given $\sigma^*$ and the number of times each edge $e_i$ is shifted backward (which is at most $\lceil m/(1/2\zeta m) \rceil$ by construction), then we can uniquely recover $\sigma$. Since $\ell \leq 2\Delta$, it follows that

$$\frac{Z(\sigma^*)}{Z(\sigma)} = \prod_{0 \leq i \leq \lceil (1 - \zeta/3)m \rceil + 2\Delta} \frac{\Gamma_i(\sigma)(\Gamma_i(\sigma) - 1)}{\Gamma_i(\sigma^*)(\Gamma_i(\sigma^*) - 1)} \geq \prod_{0 \leq i \leq \lceil (1 - \zeta/3)m \rceil + 2\Delta} \left( 1 - \frac{2\Delta + 2}{\Gamma_i(\sigma^*)} \right) m \left( 1 - \frac{2\Delta + 2}{\Gamma_i(\sigma^*)} \right)^m \geq e^{-48\Delta^2/\zeta} .$$

(52)

Finally, combining $S_f \subseteq S$ with the definition of good clusters and (51)–(52), we infer that

$$\frac{\sum_{\sigma^* \in S_f} Z(\sigma^*)}{Z(C^+)} \geq \frac{\sum_{\sigma^* \in S_f} Z(\sigma^*)}{Z(C^+)} \sum_{\sigma \in \mathcal{P}^+ \cap \mathcal{G}_{ab,xy}} Z(\sigma) \frac{Z(\sigma)}{Z(C^+)} \geq \frac{\sum_{\sigma^* \in S_f} Z(\sigma^*)}{Z(C^+)} \sum_{\sigma \in S_f} Z(\sigma) \cdot \frac{1}{16} \geq e^{-48\Delta^2/\zeta} \frac{1}{(6/\zeta)^{2\Delta}} \cdot \frac{1}{16} =: \varrho,$$

which completes the proof of Lemma 13 as discussed.

4.2 Switching of edge-sequences: proof of Lemma 13

For an edge-sequence $\sigma \in \Pi_G$ of $G$, recall that the definition (21) of the probability parameter $Z(\sigma)$ is

$$Z(\sigma) = \prod_{0 \leq i \leq m-1} \frac{2}{\Gamma_i(\sigma)(\Gamma_i(\sigma) - 1)} ,$$

where $\Gamma_i = \Gamma_i(\sigma)$ denotes the number of unsaturated vertices after adding the first $i$ edges of the edge-sequence $\sigma$. To prove the switching result Lemma 13 (which is at the heart of our approach), we will analyze how the denominators $\Gamma_i(\Gamma_i - 1)$ differ between the edge-sequences $\sigma, \overline{\sigma}, \sigma', \overline{\sigma'}$. A close inspection of these transformations reveals that $A, B, X, Y$ are the only vertices whose saturation steps might differ between $\sigma, \overline{\sigma}, \sigma', \overline{\sigma'}$, so our analysis reduces to tracking when $A, B, X, Y$ saturate in these edge-sequences (recall that $A, B, X, Y$ denote the vertices that contain the points $a, b, x, y$).
Proof of Lemma 13. Recall from Observation 11 [ii] that the steps \( t_a, t_b, t_x, t_y \) defined around (45) are the same for all the four edge-sequences \( \sigma, \mathcal{F}, \sigma', \mathcal{F}' \). Note that, without loss of generality, we may henceforth assume that \( \sigma(ab) < \sigma(xy) \) holds: indeed, part [ii] is otherwise vacuous, and for part [i] we can otherwise exchange the role of \( \sigma \) and \( \mathcal{F} \) (which swaps the order of \( \sigma(ab) \) and \( \sigma(xy) \)), and also exchanges the roles of \( \sigma' \) and \( \mathcal{F}' \) by Observation 11). Our upcoming estimates formally use a case distinction depending on whether \( t_a \leq t_b \) or \( t_b \leq t_a \) holds. Both cases can be handled by the same argument (with routine notational changes), so we shall henceforth only consider the case where \( t_a \leq t_b \) holds.

Recall that \( \Gamma_i(\sigma) \) is defined to be the number of unsaturated vertices after the addition of edge \( i \). Thus, if a vertex \( v \) becomes saturated with the addition of edge \( uv \) then \( v \) is not counted in \( \Gamma_i(\sigma) \) for \( i = \sigma(uv) \). For example, if \( \sigma(ab) < t_a \) then \( a \) is not counted in \( \Gamma_i(\sigma) \).

As mentioned above, a careful inspection reveals that \( A, B, X, Y \) are the only vertices whose saturation steps might differ between \( \sigma, \mathcal{F}, \sigma', \mathcal{F}' \). Based on the way the edge-replacements and order-swaps of the edge-sequences \( \sigma, \mathcal{F}, \sigma', \mathcal{F}' \) (see Definition 4 and above Observation 11) alter only the order of the edges \( \{ab, xy\} \) or \( \{ax, by\} \) added in steps \( \sigma(ab) \) and \( \sigma(xy) \), it thus follows that for \( i \notin \{\sigma(ab), \sigma(xy) - 1\} \) the number \( \Gamma_i(\pi) \) of unsaturated vertices is the same for all edge-sequences \( \pi \in \{\sigma, \mathcal{F}, \sigma', \mathcal{F}'\} \) (in fact, the same set of vertices are saturated in all four sequences). With foresight, we now introduce

\[
i_a := \begin{cases} 
\sigma(ab), & \text{if } \sigma(ab) \geq t_a, \\
t_a, & \text{if } \sigma(ab) < t_a < \sigma(xy), \\
\sigma(xy), & \text{if } t_a \geq \sigma(xy),
\end{cases}
\]  

(53)

and define \( i_b, i_x, i_y \) similarly. Intuitively, \( i_a \) locates \( t_a \) with respect to the interval \( [\sigma(ab), \sigma(xy)] \). Note that the assumptions \( t_a \leq t_x \) and \( t_b \leq t_y \) of Lemma 13 ensure together with our extra assumption \( t_a \leq t_b \) that \( t_a = \min\{t_a, t_b, t_x, t_y\} \). Thus, \( i_a = \min\{i_a, i_b, i_x, i_y\} \) by inspecting (53). The following property of the number of unsaturated vertices in the four edge-sequences \( \sigma, \mathcal{F}, \sigma', \mathcal{F}' \) is crucial for our argument.

Claim 15. For each \( i \notin \{i_a, \max\{i_x, i_y\} - 1\} \), the number \( \Gamma_i(\pi) \) of unsaturated vertices is the same for all edge-sequences \( \pi \in \{\sigma, \mathcal{F}, \sigma', \mathcal{F}'\} \).

Proof. The claim about the number \( \Gamma_i(\pi) \) of unsaturated vertices follows for \( i < \sigma(ab) \) and \( i \geq \sigma(xy) \) by the discussion above (53), and follows for the remaining \( i \) by the case-by-case analysis below (keeping in mind that \( A, B, X, Y \) are the only vertices whose saturation steps might differ).

(i) If \( \sigma(ab) \leq i < i_a \), then by inspecting (53) it follows that \( i < i_a \leq t_a = \min\{t_a, t_b, t_x, t_y\} \), and so none of the vertices \( A, B, X, Y \) are saturated in any of the four edge-sequences.

(ii) If \( \max\{i_x, i_y\} \leq i < \sigma(xy) \), then exactly two of the vertices \( A, B, X, Y \) are unsaturated in each of the four edge-sequences. To see this, note that in each sequence the two vertices chosen in step \( \sigma(xy) \) are not saturated by the end of step \( i \). Furthermore, using the hypothesis \( t_a \leq t_x \) and \( t_b \leq t_y \) of Lemma 13 we infer that \( \max\{t_x, t_y\} \geq \max\{t_a, t_b\} \), which in turn implies \( \max\{i_x, i_y\} \geq \max\{i_a, i_b, i_x, i_y\} \) by inspecting (53). Hence in each of the two sequences the two vertices chosen in step \( i(ab) \) will be saturated by the end of step \( i \).

From Claim 15 in view of the definition (21) of \( Z(\sigma) \) we readily infer that, defining

\[
Z'(\pi) := \prod_{i_a \leq i < \max\{i_x, i_y\}} \frac{2}{\Gamma_i(\pi)(\Gamma_i(\pi) - 1)},
\]  

(54)

we have

\[
\frac{Z'(\sigma)}{Z(\sigma)} = \frac{Z'(\mathcal{F})}{Z(\mathcal{F})} = \frac{Z'(\sigma')}{Z(\sigma')} = \frac{Z'(\mathcal{F}')}{Z(\mathcal{F}')},
\]  

(55)

It follows that inequality \( Z(\sigma) + Z(\mathcal{F}) \geq Z(\sigma') + Z(\mathcal{F}') \) from Lemma 13 part [i] is equivalent to

\[
Z'(\sigma) + Z'(\mathcal{F}) \geq Z'(\sigma') + Z'(\mathcal{F}'),
\]  

(56)

and that the inequality from Lemma 13 part [ii] is equivalent to \( Z(\sigma) + Z(\mathcal{F}) \geq (1 + e')[Z(\sigma') + Z(\mathcal{F}')] \). Using several case distinctions, we shall now prove these equivalent inequalities.

To avoid clutter in the upcoming estimates, for brevity we define \( \Gamma_i^* \) to be \( \Gamma_i(\sigma) \) plus the number of saturated vertices among \( A, B, X, Y \) at the end of step \( i \) of \( \sigma \). Thus, \( \Gamma_i^* \) is the size of the set of all unsaturated
vertices at the end of step \(i\), unioned with \(\{A, B, X, Y\}\), and this is true for step \(i\) of any of the four edge-sequences \(\sigma, \tau, \sigma', \tau'\). This allows us to express \(\Gamma_i\), for any of those sequences, as \(\Gamma_i^*\) minus the number of \(\{A, B, X, Y\}\) that are saturated at the end of step \(i\).

Recall that we have \(t_a \leq t_b \leq t_x\) and \(t_a \leq t_y\) (as discussed above). In the proof of the above-mentioned inequalities, we shall consider three possible cases for the location of \(t_y\).

**Case** \(t_a \leq t_b \leq t_x \leq t_y\): Note that this implies \(\tau(ab) \leq i_a \leq i_b \leq i_x \leq i_y \leq \tau(xy)\) by (53). Recalling (56), we thus only need to consider \(i_a \leq i \leq \max\{i_x, i_y\} = i_y\) in (54). By tracking when the vertices \(A, B, X, Y\) saturate, it follows that

\[
Z'(\sigma) = \prod_{i_a \leq i < i_b} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)},
\]

\[
Z'(\tau) = \prod_{i_a \leq i < i_b} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_x} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_y} \frac{2}{\Gamma_i^* - 2},
\]

as well as

\[
Z'(\sigma') = \prod_{i_a \leq i < i_b} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)}.
\]

\[
Z'(\tau') = \prod_{i_a \leq i < i_b} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)}.
\]

Note that \(B_1 \geq B_2, C_3 \geq C_2, D_1 \geq D_2\) and \(C_1C_2 \geq C_3^2\). Using elementary estimates\(^1\) we then obtain that

\[
\frac{Z'(\sigma) + Z'(\tau)}{Z'(\sigma') + Z'(\tau')} = \frac{B_1C_3D_2 + 1}{B_1C_3D_2} \geq \frac{C_3^2}{C_2^3} \geq \frac{C_3^2}{C_2^2} \geq 1,
\]

establishing inequality (56).

For part (ii) of Lemma 13, note that for \(i_b \leq i < i_x\) we used \(\Gamma_i^* - 2 = \Gamma_i(\sigma)\) above (cf. \(C_1\) above). Furthermore, we have \(\Gamma_i(\sigma) \leq 2(m - i)\) by Observation 4. Combined with the fact that \(i_x - i_b \geq \min\{\tau(xy), t_x\} - \max\{\tau(ab), t_b\}\) satisfies \(i_x - i_b \geq \zeta m/3\) by the ‘gap’ hypothesis of part (iii), it follows that

\[
\frac{C_3}{C_2} = \frac{\prod_{i_a \leq i < i_x} \left(1 + \frac{2}{\Gamma_i^* - 2}\right)}{\prod_{i_a \leq i < i_x} \left(1 + \frac{1}{m - i}\right)} = 1 + \frac{i_x - i_b}{m - i} \geq 1 + \frac{\zeta m/3}{m + 1} \geq 1 + \frac{\zeta}{6},
\]

which enables us to improve the final inequality of (57) to \(\geq 1 + \epsilon'\) for suitable \(\epsilon' = \epsilon'(\zeta) > 0\).

**Case** \(t_a \leq t_b \leq t_y \leq t_x\): Note that this implies \(\tau(ab) \leq i_a \leq i_b \leq i_y \leq i_x \leq \tau(xy)\) by (53), so this time we only need to consider \(i_a \leq i < \max\{i_x, i_y\} = i_x\) in (54). Proceeding analogously to the previous case, by tracking when the vertices \(A, B, X, Y\) saturate it follows that

\[
Z'(\sigma) = \prod_{i_a \leq i < i_b} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)},
\]

\[
Z'(\tau) = \prod_{i_a \leq i < i_b} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_x} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_y} \frac{2}{\Gamma_i^* - 2},
\]

\[
Z'(\sigma') = \prod_{i_a \leq i < i_b} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)},
\]

\[
Z'(\tau') = \prod_{i_a \leq i < i_b} \frac{2}{\Gamma_i^* - 1} \cdot \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \cdot \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)}.
\]

\(^1\)In equation (57) we exploit that for all \(\alpha, \beta \geq 1\) have \(\alpha^2(\beta - 1) \geq \beta - 1\), which implies \(2(\alpha^2\beta + 1) \geq (\alpha^2\beta + 1) + (\alpha^2 + \beta) = (\alpha^2 + 1)(\beta + 1)\) and thus \((\alpha^2\beta + 1)/(\alpha(\beta + 1)) \geq (\alpha^2 + 1)/(2\alpha)\).
as well as

\[
Z'(\sigma') = \prod_{i_a \leq i < i_b} \left( \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right),
\]

\[
Z'(\sigma) = \prod_{i_a \leq i < i_b} \left( \prod_{i_a \leq i < i_y} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 1)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \right).
\]

Note that \(B_1 \geq B_2, C_1C_2 \geq C_3^2, C_3 \geq C_2,\) and \(D_1 \geq D_2.\) Similarly to (57), using elementary estimates\(^{13}\) we then obtain that

\[
\frac{Z'(\sigma) + Z'(\sigma)}{Z'(\sigma') + Z'(\sigma')} \geq \frac{C_1^2}{C_2^2} \frac{B_1D_2 + 1}{B_2D_1} + \frac{1}{2} \frac{(C_2^2)^2 + 1}{2C_2^2} \geq 1,
\]

establishing (59). For part (ii) using \(i_y - i_b \geq \zeta m/3\) we obtain analogously to (58) that

\[
\frac{C_3}{C_2} = \prod_{i_b \leq i < i_x} \left( 1 + \frac{2}{\Gamma_i^* - 2} \right) \geq \prod_{i_b \leq i < i_x} \left( 1 + \frac{1}{m - i} \right) = 1 + \frac{i_x - i_b}{m - i} \geq 1 + \frac{\zeta}{6},
\]

which again enables us to improve the final inequality of (59) to \(\geq 1 + \epsilon'.\)

**Case** \(i_a \leq i_y \leq i_b \leq i_x: \) Note that this implies \(\sigma(ab) \leq i_a \leq i_y \leq i_b \leq i_x \leq \sigma(xy)\) by (58), so we again only need to consider \(i_a \leq i < \text{max}\{i_x, i_y\} = i_x\) in (51). Proceeding analogously to the previous cases, it follows that

\[
Z'(\sigma) = \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 3)(\Gamma_i^* - 2)} \right),
\]

\[
Z'(\sigma) = \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 1)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 2)} \right),
\]

as well as

\[
Z'(\sigma') = \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 2)} \right),
\]

\[
Z'(\sigma') = \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 1)} \right) \cdot \prod_{i_a \leq i < i_y} \left( \prod_{i_a \leq i < i_x} \frac{2}{(\Gamma_i^* - 1)(\Gamma_i^* - 2)} \right) \cdot \prod_{i_b \leq i < i_y} \left( \prod_{i_b \leq i < i_x} \frac{2}{(\Gamma_i^* - 2)(\Gamma_i^* - 3)} \right).
\]

Noting that \(B_1 \geq B_2\) and \(D_1 \geq D_2\) imply \((B_1 - B_2)(D_1 - D_2) \geq 0\), it follows that

\[
\frac{Z'(\sigma) + Z'(\sigma)}{Z'(\sigma') + Z'(\sigma')} = \frac{B_1D_1 + B_2D_2}{B_1D_2 + B_2D_1} \geq 1.
\]

For part (ii) there is nothing to show in this case, since \(i_y \leq i_b\) implies \(\min\{\sigma(xy), t_x, t_y\} - \max\{\sigma(ab), t_a, t_b\} \leq 0 < \zeta m/3\), contradicting the ‘gap’ hypothesis of part (ii). \(\square\)

---

\(^{13}\) Similarly to Footnote 12 in equation (59), we exploit that for all \(\alpha, \beta, \gamma \geq 1\) we have \(\alpha^2 \beta (\gamma - 1) \geq \gamma - 1\) and \(\alpha^2 \beta (\gamma - 1) \geq \beta - 1\), which implies \(2(\alpha^2 \beta + 1) \geq (\alpha^2 + \gamma) + (\alpha^2 + \beta) = (\alpha^2 + 1)(\beta + \gamma)\) and thus \((\alpha^2 + \beta + 1)/(\alpha^2 + \beta + 1) \geq (\alpha^2 + 1)/(\alpha^2 + 1)\).
4.3 Pairing of edge-sequences: twins and proof of Lemma 12

Expanding on the beginning of Section 4, the majority of this subsection is devoted to the rigorous construction of twins, i.e., pairs of edge-sequences \((\sigma_1, \sigma_2) \in (\mathcal{P}^+ \times \mathcal{P}^-) \cup (\mathcal{P}^- \times \mathcal{P}^+)\) that satisfy \(Z(\sigma_1) = Z(\sigma_2)\) and a number of additional technical properties that are useful for the proof of Lemma 12.

Before giving the formal details of our construction, we start with a simple yet illustrative example for the case \(\sigma_1 \in \mathcal{P}^+\) with \(t_x(\sigma_1) < t_y(\sigma_1)\). Suppose that \(B\) and \(X\) are not adjacent, with \(\deg(B) = 4\) and \(\deg(X) = 6\), and that the ten edges incident to \(B\) and \(X\) appear in the edge-sequence \(\sigma_1\) in the following manner:

\[
..., qx_1, ..., rb_1, ..., ab, ..., sx_2, ..., tx_3, ..., ux_4, ..., vx_5, ..., xy, ..., wb_2, ..., zb_3, ...
\]

We then replace the edges \(ab, xy\) with \(ax, by\) (which can be thought of as swapping the points \(b\) and \(x\)), and further swap some of the points in \(B\) with some of the points in \(X\) (say, swapping \(b_2\) and \(x_3\), and swapping \(b_3\) and \(x_5\)) to obtain the edge-sequence \(\sigma_2 \in \mathcal{P}^-\) (clusters are defined to accommodate these operations; in particular, swapping of \(b\) and \(x\) ensures \(\sigma_2 \in \mathcal{P}^-\)):

\[
..., qx_1, ..., rb_1, ..., ax, ..., sx_2, ..., tb_2, ..., ux_3, ..., vb_3, ..., by, ..., wx_4, ..., zv_5, ...
\]

There are two important properties to observe in this example:

(i) Only points of \(B\) and \(X\) move to different positions, meaning that any vertex \(V \notin \{B, X\}\) becomes saturated during the same step as \(\sigma_2\) as does in \(\sigma_1\).

(ii) Inspecting the definition (44)–(45), we have \(t_x(\sigma_2) = t_b(\sigma_1)\) and \(t_b(\sigma_2) = t_x(\sigma_1)\), as well as \(\sigma_2(ax) = \sigma_1(ab)\) and \(\sigma_2(by) = \sigma_1(xy)\). This implies that the steps at which \(B\) and \(X\) become saturated are swapped between \(\sigma_1\) and \(\sigma_2\).

From these properties, it follows that the number of unsaturated vertices satisfies \(\Gamma(t)(\sigma_1) = \Gamma(t)(\sigma_2)\) for all steps \(1 \leq t \leq m\), which in turn leads to \(Z(\sigma_1) = Z(\sigma_2)\). In this case, we will call \(\sigma_2\) the \(BX\) twin of \(\sigma_1\), and vice versa; see Section 4.3.1. In general, we will carefully swap some points in \(B\) and \(X\) but not in a \(BX\) edge such that the two properties stated above holds (to clarify: we will not swap the endpoints of any edges between \(B\) and \(X\), since that could create loops).

With that example in mind, we now turn to the formal details of our construction, where the following auxiliary observation will be used to swap points.

Observation 16. Consider any integers satisfying \(0 < d_b < d_x \leq t < d_x + d_b\). Let \(\Theta_1\) denote the set of sequences consisting of \(d_x\) many \(X\)’s and \(d_b\) many \(B\)’s such that: (i) the last character is \(B\), (ii) the \(t^{th}\) character is \(X\), and (iii) every character after the \(t^{th}\) is \(B\). Let \(\Theta_2\) denote the set of sequences consisting of \(d_x\) many \(X\)’s and \(d_b\) many \(B\)’s such that: (i) the last character is \(X\), (ii) the \(t^{th}\) character is \(B\), and (iii) every character after the \(t^{th}\) is \(X\). Then \(|\Theta_1| \leq |\Theta_2|\).

Proof. The main idea is that restriction (iii) is stricter for \(\Theta_1\) than for \(\Theta_2\) because of \(d_x > d_b\). More formally, by considering the first \(t - 1\) character in the sequence, it follows that \(|\Theta_1| = (t - 1) \leq (t - 1) = |\Theta_2|\), where the inequality holds since \(t \leq d_x + d_b - 1\).

Observation 16 implies that there is an injection \(I_{t,d_b,d_x}\) mapping \(\Theta_1\) into \(\Theta_2\). There are many possible injections, but we henceforth fix one arbitrarily (and use it in what follows).

4.3.1 BX-twins

Consider any edge-sequence \(\sigma_1 \in \mathcal{P}^+\) with \(t_x(\sigma_1) < t_y(\sigma_1)\). We will construct its \(BX\) twin \(\sigma_2 \in \mathcal{P}^-\) by generalizing the example above, where we swapped \(b\) and \(x\), but did not swap points of \(B\) and \(X\) in a \(BX\) edge. To this end we define \(\lambda_{BX}\) as the number of edges between \(B\) and \(X\) (which we will call \(BX\) edge), and introduce the parameters

\[
d_b := \deg(B) - 1 - \lambda_{BX} \quad \text{and} \quad d_x := \deg(X) - 1 - \lambda_{BX},
\]

which corresponds to the number of \(B\) and \(X\) points that we might swap. Using \(\deg(B) < \deg(X)\), we infer that \(d_b < d_x\). With a view towards applying Observation 16, we consider the \(d_b + d_x\) edges in \(\sigma_1\) incident to \(B\) or \(X\) but not \(ab\) or \(xy\) and not a \(BX\)-edge. By considering these edges in the order in
which they appear in $\sigma_1$, we naturally obtain a sequence containing $d_b$ many $B$'s and $d_x$ many $X$'s, where the $i^{th}$ entry is $B$ (resp. $X$) if the $i^{th}$ edge is has $B$ (resp. $X$) as an endpoint. The resulting sequence $S_1$ is well-defined, since none of these edges has both $B$ and $X$ as endpoints. Since $t_e(\sigma_1) < t_b(\sigma_1)$, it follows that the last entry of $S_1$ is a $B$. Let $t$ be the position in $S_1$ of the last $X$. Thus $t \geq d_x$. So, using the notation from Observation 16, we have $S_1 \in \Theta_1$ and we then set $S_2 := I_{t,d_b,d_x}(S_1)$ using the injection we fixed above. (As an illustration, in the example above we have $d_b = 3$ and $d_x = 5$ as well as $S_1 = X, B, X, X, X, B, B$ and $S_2 = X, B, X, B, X, B, X, X$.) With $S_1$ and $S_2$ in hand, we are now ready to construct $\sigma_2$ by modifying $\sigma_1$ via the following steps:

(i) In $\sigma_1$, replace $ab$ with $ax$ and $xy$ with $by$.
(ii) We temporarily ignore the label of points in $X$. For every edge $e$ counted by $d_x$, i.e., edges incident to $X$ that is not $xy$ and not a $BX$ edge, let $i_e$ denote the position in $S_1$ corresponding to $e$; thus $S_1$ has an $X$ in position $i_e$. If $S_2$ has a $B$ in position $i_e$, then we modify $e$ by changing endpoint $X$ to $B$; keep this edge in the same position in $\sigma_2$ as $e$ is in $\sigma_1$. (If $S_2$ has an $X$ in position $i_e$, then we do nothing.)
(iii) We temporarily ignore the label of points in $B$. For every edge $e$ counted by $d_b$, i.e., edges incident to $B$ that is not $ab$ and not a $BX$ edge, let $i_e$ denote the position in $S_1$ corresponding to $e$; thus $S_1$ has a $B$ in position $i_e$. If $S_2$ has a $X$ in position $i_e$, then we modify $e$ by changing endpoint $B$ to $X$; keep this edge in the same position in $\sigma_2$ as $e$ is in $\sigma_1$. (If $S_2$ has a $B$ in position $i_e$, then we do nothing.)
(iv) Label the points of $B$ not equal to $b$ in $\sigma_2$ with the same relative ordering as (labeled) points of $B$ not equal to $b$ in $\sigma_1$. Similarly, label the points of $X$ not equal to $x$ in $\sigma_2$ with the same relative ordering as (labeled) points of $X$ not equal to $x$ in $\sigma_1$.

The resulting edge-sequence $\sigma_2$ is called a $BX$-twin of $\sigma_1$, and from the definition of clusters it follows that $\sigma_2 \in \mathcal{P}^-$ (because we did not change the set of edges not incident to $B$ and $X$, and we did not change the union of neighboring points of $B$ and $X$). The following lemma records some additional properties.

Lemma 17. The $BX$-twin $\sigma_2 \in \mathcal{P}^-$ constructed above satisfies the following properties:

(a) $Z(\sigma_1) = Z(\sigma_2)$.
(b) $\sigma_2 \in \mathcal{P}^-$ and $t_b(\sigma_2) < t_x(\sigma_2)$.
(c) $t_a(\sigma_2) = t_a(\sigma_1)$, $t_y(\sigma_2) = t_y(\sigma_1)$, $t_b(\sigma_2) = t_x(\sigma_1)$, $t_x(\sigma_2) = t_b(\sigma_1)$.
(d) Every $\sigma \in \mathcal{P}^-$ with $t_b(\sigma) < t_x(\sigma)$ has at most one $BX$-twin.

Proof. We first argue that $t_b(\sigma_1) = t_x(\sigma_2)$. Since $t_x(\sigma_1) < t_b(\sigma_1)$, the last edge in $\sigma_1$ which has $B$ as an endpoint cannot be a $BX$ edge. So $t_b(\sigma_1)$ is the index corresponding to the last element of $\sigma_1$. That edge is modified to have endpoint $X$ in $\sigma_2$, and no edges following it in $\sigma_1$, except possibly $ab$ and $xy$, are modified. Also, no edges following it, except possibly $xy$, have $X$ as an endpoint, since $t_x(\sigma_1) < t_b(\sigma_1)$. So that index becomes $t_x(\sigma_2)$, establishing $t_b(\sigma_1) = t_x(\sigma_2)$.

Similar reasoning shows that $t_b(\sigma_2) = t_x(\sigma_1)$. The two key additional insights are: (i) every character after that point in $S_2$ is an $X$. Thus (ii) any edges after that point in $\sigma_2$ that have $B$ as an endpoint must be $BX$ edges or $ab$. The index of the latest such edge that is not $ab$, if any exist, would be $t_x(\sigma_1)$ and $t_b(\sigma_2)$.

Now, since the positions of all copies of $a, y$ are the same in $\sigma_2$ as in $\sigma_1$, it follows that the findings of the first two paragraphs imply part (a).

Part (b) immediately implies part (c).

For part (a) note that, since $ab, xy$ in $\sigma_1$ become $ax, by$ in $\sigma_2$, the step at which $B$ is saturated in $\sigma_1$ is equal to the step at which $X$ is saturated in $\sigma_2$, and vice versa. Furthermore, every other vertex becomes saturated at the same step of $\sigma_1$ and $\sigma_2$ (since they all appear on the same edges and in the same position in both). Hence $\Gamma_i(\sigma_1) = \Gamma_i(\sigma_2)$ for all $1 \leq i \leq m$, which implies $Z(\sigma_1) = Z(\sigma_2)$, establishing part (a).

Part (d) follows because $I_{t,d_b,d_x}$ is an injection. Indeed, given $\sigma_2 \in \mathcal{P}^-$ with $t_b(\sigma_2) < t_x(\sigma_2)$, we look at the $d_b$ edges incident to $B$ and $d_x$ edges incident to $X$, other than $ax, by$ and any $BX$ edges. This specifies the sequence $S_2 \in \Theta_2$ and the latest $B$ in that sequence establishes the value $t$. Now we obtain $S_1 = I_{t,d_b,d_x}^{-1}(S_2)$, which then allows us to uniquely reconstruct the $BX$-twin of $\sigma_2$, concluding the proof.

Next consider any $\sigma_1 \in \mathcal{P}^+$ with $t_x(\sigma_1) < t_b(\sigma_1)$. We can construct a $BX$-twin $\sigma_2 \in \mathcal{P}^+$ in the same manner, with only the following change: In $\sigma_1$, replace $ax$ with $ab$ and $by$ with $xy$. The obvious analogue of Lemma 17 holds, by the same proof. This will be encompassed by the more general Lemma 18 below.
4.3.2 $AY$-twins and properties of twins

Consider any $\sigma_1 \in \mathcal{P}^+ \cup \mathcal{P}^-$ with $t_y(\sigma_1) < t_a(\sigma_1)$. We construct its $AY$-twin $\sigma_2 \in \mathcal{P}^-$ using the same construction as for $BX$-twins, where we this time swap the points of $A$ and $Y$ not in an $AY$ edge (instead of the points of $B$ and $X$ not in an $BX$ edge). Besides notational changes, this really just corresponds to interchanging the roles of the vertices $B$ and $A$ as well as the vertices $X$ and $Y$ in the construction from Section 4.3.1. In particular, we here use a sequence of $d_a := \text{deg}(A) - 1 - \lambda_{AY}$ many $A$'s and $d_y := \text{deg}(Y) - 1 - \lambda_{AY}$ many $Y$'s. We replace step (i) with the following: (a) if $\sigma_1 \in \mathcal{P}^+$ then, in $\sigma_1$, replace $ab$ with $by$ and $xy$ with $ax$. and (b) if $\sigma_1 \in \mathcal{P}^-$ then, in $\sigma_1$, replace $ax$ with $xy$ and $by$ with $ab$. In steps (ii) and (iii) we use the injection $I_{t,d_a,d_y}$, where $t$ is defined analogously as in the $BX$-twin case (the assumptions of Section 16 hold, since $\text{deg}(A) < \text{deg}(Y)$ implies $d_a < d_y$), and modify step (iv) in the obvious way.

We henceforth use the term twins to refer to both $BX$-twins and $AY$-twins. Then the obvious analogue of Lemma 17 holds (by the same proof), and we now record these key properties for later reference.

**Lemma 18** (Key properties of twins). Defining $\mathcal{P} := \mathcal{P}^- \cup \mathcal{P}^+$, the following holds for twins:

(a) If $(\sigma_1, \sigma_2)$ are twins, then $Z(\sigma_1) = Z(\sigma_2)$.

(b) Every $\sigma \in \mathcal{P}$ with $t_x(\sigma) < t_b(\sigma)$ has exactly one $BX$-twin, and every $\sigma \in \mathcal{P}$ with $t_y(\sigma) < t_a(\sigma)$ has exactly one $AY$-twin.

(c) Every $\sigma \in \mathcal{P}$ with $t_y(\sigma) < t_x(\sigma)$ has at most one $BX$-twin, and every $\sigma \in \mathcal{P}$ with $t_a(\sigma) < t_y(\sigma)$ has at most one $AY$-twin.

(d) If $\sigma$ and $\sigma'$ are counterparts then either (i) they both do not have $BX$-twins, or (ii) their $BX$-twins are counterparts of each other. The same is true for $AY$-twins.

**Proof.** Parts (a) and (c) are simply restatements of parts (a) and (d) of Lemma 17 and its analogues. Part (d) holds by construction of twins. Part (b) follows from the fact that, as counterparts, both $\sigma$ and $\sigma'$ will have the same values of $d_b, d_x, t_b, t_x$ and the same ordering of edges incident to $B$ and $X$ except $ab, xy, ax, by$. So if one of them is a $BX$-twin then either $t_x < t_b$ or we can reconstruct their $BX$-twin using $I_{t,d_x,d_b}$ as in the proof of Lemma 17. In both cases both $\sigma$ and $\sigma'$ are $BX$-twins.

4.3.3 Quadruplets

Consider $\sigma_1 \in \mathcal{P}^+ \cup \mathcal{P}^-$ with $t_x(\sigma_1) < t_b(\sigma_1)$ and $t_y(\sigma_1) < t_a(\sigma_1)$. Then $\sigma_1$ has both a $BX$-twin $\sigma_2$ and an $AY$-twin $\sigma_3$. By Lemma 17 and its analogues, $t_x(\sigma_2) < t_y(\sigma_2)$ and $t_x(\sigma_3) < t_b(\sigma_3)$. Therefore, $\sigma_2$ has an $AY$-twin and $\sigma_3$ has a $BX$-twin. In the proof of Lemma 12 the following result will eventually allow us to uniquely pair up edge-sequences with their twins (this is not obvious, because there could exist edge-sequences $\sigma_1, \sigma_2$ and $\sigma_3$ that have two twins).

**Lemma 19.** The $AY$-twin of $\sigma_2$ is equal to the $BX$-twin of $\sigma_3$.

**Proof.** When we construct a $BX$-twin, we do not alter any endpoints that are $A$ or $Y$, and the edges containing those endpoints remain in the same position in the edge-sequences. It follows that when we construct the $AY$-twin of $\sigma_2$ and the $AY$-twin of $\sigma_1$, we alter the exact same sequence $S_1$ of $A$'s and $Y$'s and, since we always use the same injection $I_{t,d_x,d_y}$, we alter it in the same way. The only difference in the two constructions is that, for one twin we change $(ab, xy)$ to $(by, ax)$ and for the other we change $(ax, by)$ to $(xy, ab)$. Similarly for when we construct the $BX$-twins of $\sigma_1$ and $\sigma_3$.

Thus, the sequence of two modifications: $\sigma_1 \rightarrow \sigma_2 \rightarrow AY$-twin of $\sigma_2$ makes the same alterations as the sequence $\sigma_1 \rightarrow \sigma_3 \rightarrow BX$-twin of $\sigma_3$, with the following exception: In the first sequence, if $\sigma_1 \in \mathcal{P}^+$, we have $(ab, xy) \rightarrow (ax, by) \rightarrow (xy, ab)$, and in the other sequence, $(ab, xy) \rightarrow (by, ax) \rightarrow (xy, ab)$. If $\sigma_1 \in \mathcal{P}^-$ then we have $(ax, by) \rightarrow (ab, xy) \rightarrow (by, ax)$ and $(ax, by) \rightarrow (xy, ab) \rightarrow (by, ax)$. So in both sequences, we end up with the same pair, concluding the proof.

We henceforth refer to $\sigma_1, \sigma_2, \sigma_3$ and the $AY$-twin of $\sigma_2$ (i.e., the $BX$-twin of $\sigma_3$) as quadruplets.

4.3.4 Summing over twins and quadruplets: proof of of Lemma 12

In the following proof of Lemma 12 we construct the desired pairs of edge-sequences $(\sigma_2, \sigma_1)$, defining $T \subseteq \mathcal{P}^+ \cup \mathcal{P}^-$ as the set of all edge-sequences $\sigma$ which have a twin.
Proof of Lemma 12. The goal of part (i) is to show that
\[
\sum_{\sigma \in T \cap P^+} Z(\sigma) = \sum_{\sigma \in T \cap P^-} Z(\sigma).
\] (61)
To prove this equation, we apply Lemmas 18 and 19 to pair terms from the left-hand side of (61) with terms from the right-hand side of (61). More precisely, for each set of quadruplets, we arbitrarily choose one of the two ways to split the four into two pairs, where each pair \((\sigma_2, \sigma_1) \in (P^+ \times P^-) \cup (P^- \times P^+)\) is a set of twins. For each remaining edge-sequence \(\sigma_2 \in P^+\) that has a twin (and is not part of a quadruplet), we then form the pair \((\sigma_2, \sigma_1)\), where \(\sigma_1 \in P^-\) is the unique twin of \(\sigma_2\). By Lemma 18(i), all paired edge-sequences each have the same \(Z\)-value. So in (61) all terms cancel out, establishing that the left-hand side of (61) is equal to the right-hand side of (61), concluding the proof of part (i).

Part (iii) follows from Lemma 18(d) and the fact that the position of the edges \(ab, xy, ax, by\) in \(\sigma\) plays no role in the construction of twins.

Part (iv) follows from Lemma 18(iii).

For part (iv), for the sake of contradiction suppose \(\sigma_1 \in P\) is a BX-twin of \(\sigma\). Then \(t_x(\sigma_1) < t_b(\sigma_1)\) and no points of \(X, Y\) are adjacent to any points of \(A, B\) in \(\sigma_1\). Then by Observation 16 (ii), \(\sigma\) satisfies the following: there is some \(t \geq d_x > d_b\) such that the \(t\)th edge in the sequence of edges incident to \(B\) and \(X\) is incident to \(B\). In particular we see that not all edges incident to \(X\) appear after all edges incident to \(B\), a contradiction. Therefore \(\sigma\) is not a BX-twin, and similarly \(\sigma\) is not an AY-twin, which completes the proof.

\[\square\]

5 Counting results: deferred proofs

This section is devoted to the deferred proofs of the counting results Lemma 8 and 9 from Section 3.3, which give bounds on the number of clusters containing a given configuration-graph and the number of good edge-sequences. Both proofs are based on counting arguments and the following two observations. For brevity, we henceforth always tacitly assume that \(\gamma > 0\) is sufficiently small and that \(n > n_0(\gamma, \Delta, \xi)\) is sufficiently large (whenever necessary).

Observation 20. There are constants \(c = c(\xi, \Delta) > 1/2\) and \(\nu = \nu(\xi, \Delta) > 0\) such that \(m \geq \max\{c \sum_{j \in [k]} jn_j, \mu\}\)
and \(\mu + m - \sum_{j \in [k]} jn_j \geq \nu \max\{m, \mu\}\).

Proof. Since by assumption \(\sum_{k < j \leq \Delta} n_i = n - \sum_{j \in [k]} jn_j \geq \xi n\), it follows that \(m \geq c \sum_{j \in [k]} jn_j\) for some constant \(c = c(\xi, \Delta) > 1/2\).
Hence
\[
\mu + m - \sum_{j \in [k]} jn_j = \left(1 - \frac{\sum_{j \in [k]} jn_j}{2m}\right)^2 \cdot m \geq \left(1 - \frac{1}{2c}\right)^2 \cdot m,
\]
which completes the proof by noting that \(m^2 \geq (\sum_{j \in [k]} jn_j)^2/4\) implies \(m \geq \mu\).

Recall that an edge is small if both endpoints have degree at most \(k\) (see Section 1.1). We say an edge is large if both endpoints have degree greater than \(k\), and mixed if one endpoint’s degree is at most \(k\) and the other is greater than \(k\). With these definitions in hand, we now record the following consequence of our notational convention from Section 3.3 (see page 12).

Observation 21. For all configuration-graphs \(G\) the following holds:
(i) The number \(U_G\) of upper clusters containing \(G\) equals four times the number of pairs consisting of small and large edges in \(G\).
(ii) The number \(L_G\) of lower clusters containing \(G\) equals twice the number of pairs consisting of mixed edges in \(G\) whose endpoints are contained in four distinct vertices.

Proof. For part (i) note that for every small edge \(e\) and large edge \(e'\) in a graph \(G\), there are four upper clusters of the form \(C^+(G, e, e')\). For part (ii) note that for every pair of edges \(e, e'\) in a graph \(G\), each with one endpoint of degree at most \(k\) and another of degree greater than \(k\), there are two lower clusters of the form \(C^-(G, e, e')\) if the endpoints of \(e\) and \(e'\) are contained in four distinct vertices (note that the distinct vertices condition is baked into the definition of clusters, as we insist on \(A, B, X, Y\) being distinct vertices).

\[\square\]
5.1 Counting clusters: proof of Lemma 8

Recall that $U_G$ and $L_G$ count the number of upper and lower clusters that contain the configuration-graph $G$ (see Section 3.3). To prove Lemma 8 we shall combine Observation 21 with counting arguments to show that $L_G$ and $U_H$ are approximately equal when $G$ and $H$ contain similar numbers of small edges.

Proof of Lemma 8. We first estimate $L_G$ and $U_G$ for any configuration-graph $G$ with $\ell = X_k(G)$ small edges. Recall that an edge is called mixed if one endpoint’s degree is at most $k$ and the other is greater than $k$. By Observation 21, the number $L_G$ of lower clusters containing $G$ equals to twice the number of pairs $(e, e')$ of mixed edges in $G$ where the endpoints of $e, e'$ lie on four different vertices. Since there are $2 \cdot X_k(G) = 2\ell$ small points whose neighbor is also small, we infer that there are $\sum_{j \in [k]} j n_j - 2\ell$ mixed edges. It follows that

$$\left(\sum_{j \in [k]} j n_j - 2\ell\right) - \left(\sum_{j \in [k]} j n_j - 2\ell\right) \cdot 2\Delta \leq \frac{L_G}{2} \leq \left(\sum_{j \in [k]} j n_j - 2\ell\right).$$

By Observation 21, the number $U_G$ of upper clusters containing $G$ equals to four times the number of pairs of edges $(e, e')$ where $e$ is small and $e'$ is large in $G$. Since there are $\ell$ small edges and $m - \ell - \sum_{j \in [k]} j n_j$ large edges, it follows that

$$U_G = 4\ell \left(\ell + m - \sum_{j \in [k]} j n_j\right).$$

We next simplify inequality (62) for $L_G$ when $G$ has $\ell \leq (1 + \gamma)\mu$ small edges. Using Observation 20 we infer $2\ell \leq (1 + \gamma)\sum_{j \in [k]} j n_j/(2c)$ for some $c > 1/2$, so that $\sum_{j \in [k]} j n_j - 2\ell = \Omega(\sum_{j \in [k]} j n_j) = \omega(1)$ for all sufficiently small $\gamma > 0$. Hence inequality (62) implies

$$L_G = (1 + o(1)) \cdot \left(\sum_{j \in [k]} j n_j - 2\ell\right)^2 = \omega(1).$$

Finally, combining (63)–(64) with $4m\mu = (\sum_{j \in [k]} j n_j)^2$ and Observation 20 for all sufficiently small $\gamma > 0$ it follows that for all configuration-graphs $G, H \in \mathcal{N}_{\gamma}$, we have

$$L_G \leq \frac{1}{(1 + O(\gamma)) \cdot 4\mu(1 + O(\gamma))} \left(\sum_{j \in [k]} j n_j - 2[1 + O(\gamma)]\mu\right)^2$$

$$= (1 + O(\gamma)) \cdot \left(1 + \left(\sum_{j \in [k]} j n_j - 4m\mu + O(\gamma)(\sum_{j \in [k]} j n_j + \mu)\right)\right)$$

$$= (1 + O(\gamma)) \cdot \left(1 + \frac{O(\gamma m)}{\nu \max\{m, \mu\} - O(\gamma \mu)}\right) = 1 + O(\gamma),$$

which together with (64) also implies $U_H = \omega(1)$, completing the proof. 

5.2 Counting good edge-sequences: proof of Lemma 9

Recall that in any upper cluster $C^+ = C^+(G, ab, xy)$ the edge $ab$ is small and the edge $xy$ is large (see Section 3.2.2). By Observation 21, we know that $U_G$ counts four times the number of pairs consisting of small and large edges, and so it follows (with room to spare) that the desired estimate (29) of Lemma 9 would be implied by

$$\Lambda_{G, \sigma} := \frac{|\{ab, xy \in E(G) : ab \text{ small, } xy \text{ large, } (\sigma, ab, xy) \in G\}|}{|\{ab, xy \in E(G) : ab \text{ small, } xy \text{ large}\}|} \geq \frac{1}{2}.$$ 

Intuitively (66) says that at least half of the choices of small edges $ab \in E(G)$ and large edges $xy \in E(G)$ yield a good edge-sequence $(\sigma, ab, xy)$. To prove this estimate, we shall first use counting arguments to show that many small edges have endpoints lying in vertices that do not become saturated during the last $\zeta m$ steps of the edge-sequence $\sigma$. Then, for each such small edge $ab$, we shall use counting arguments to show that almost all large edges $xy$ yield a good edge-sequence $(\sigma, ab, xy) \in G$, i.e., $X$ and $Y$ are not adjacent to $A$ and $B$ (recall that $A, B, X, Y$ denote the vertices that contain the points $a, b, x, y$). As we shall see, this together implies that many edge-sequences are good, which eventually establishes (66) and thus (29).
Proof of Lemma 9. We henceforth fix a configuration-graph $G \in \mathcal{N}_{\gamma \mu}$ and an edge-sequence $\sigma \in \Pi_G$ of the edges of $E(G)$. A small edge $ab \in E(G)$ is called early if both vertices $A, B$ saturate during the first $(1 - \zeta)m$ steps. Inspecting $A_{G,\sigma}$ from (66), by definition of good edge-sequences $G$ (see Section 3.3) we have

$$A_{G,\sigma} \geq \frac{|\{ab \in E(G) : ab \text{ small, } ab \text{ early}\}|}{|\{ab \in E(G) : ab \text{ small}\}|} \cdot \min_{ab \in E(G): \text{ ab small, } ab \text{ early}} \frac{|\{xy \in E(G) : xy \text{ large, } (\sigma, ab, xy) \in G\}|}{|\{xy \in E(G) : xy \text{ large}\}|}. \quad (67)$$

It remains to bound the two fractions in the right-hand side of (67) from below, and we start with the first fraction. As before, we write $\ell = X_k(G)$ for the number of small edges of $G$, where $\ell \geq (1 - \gamma)\mu$ by assumption. During the final $\zeta m$ steps of the process, at most $2 \cdot \zeta m \leq \zeta n \Delta$ small vertices become saturated. Noting that $2m \leq n\Delta$ and $\gamma \leq 1/2$ imply $\ell \geq (1 - \gamma)\mu \geq (\sum_{j \in [k]} jn_j)^2/8m \geq \xi^2 n/(4\Delta)$, using the definition (25) of $\zeta = \xi^2/(16\Delta^3)$ it follows that

$$\frac{|\{ab \in E(G) : ab \text{ small, } ab \text{ early}\}|}{|\{ab \in E(G) : ab \text{ small}\}|} \geq \frac{\ell - \zeta n \Delta \cdot \Delta}{\ell} = 1 - \frac{\xi^2 n}{16\Delta^2} \geq \frac{3}{4}. \quad (68)$$

Turning to the second fraction in the right-hand side of (67), let $ab \in E(G)$ be a small and early edge. There are $\ell + m - \sum_{j \in [k]} jn_j$ large edges $xy \in E(G)$, as established above (69). Furthermore, there are at most $2k \cdot \Delta \leq 2\Delta^2$ edges $xy \in E(G)$ with the property that at least one point of $X, Y$ is adjacent to a point in $A, B$ (recall that $A, B, X, Y$ denote the vertices containing the points $a, b, x, y$). For sufficiently small $\gamma > 0$ Observation 20 implies that $\mu + m - \sum_{j \in [k]} jn_j \geq 2\gamma \mu$, so using $\ell \geq (1 - \gamma)\mu$ and $\mu = \Theta(n)$ it follows that

$$\frac{|\{xy \in E(G) : xy \text{ large, } (\sigma, ab, xy) \in G\}|}{|\{xy \in E(G) : xy \text{ large}\}|} \geq \frac{\ell + m - \sum_{j \in [k]} jn_j - 2\Delta^2}{\ell + m - \sum_{j \in [k]} jn_j} \geq 1 - \frac{2\Delta^2}{\gamma \mu} \geq \frac{2}{3},$$

which together with inequalities (67)–(68) establishes (66) and thus the desired estimate (29). \qed

6 Epilogue: alternative approach

In this final section we briefly discuss an alternative approach to Theorems 1 and 2, which instead of switching is based on the differential equation method [52, 53, 50]. Our primary interest is the intriguing possibility that this approach can be made rigorous, which would lead to a conceptually simpler proof of our main result.

The basic idea is straightforward: standard combinatorial methods (see Appendix A.2) show that in the uniform model the number of small edges $X_k(G_{d_n})$ is concentrated around $\mu = (\sum_{j \in [k]} jn_j)^2/(4mn)$, and a routine application of the differential equation method shows that in the standard $d_n$-process the number of small edges $X_k(G'_{d_n})$ is concentrated around $\rho_k n$, where the quantity $\rho_k$ has an analytic description in terms of differential equations. To prove a discrepancy it thus suffices to show that the two rescaled means $\mu/n$ and $\rho_k$ differ in the limit as $n \to \infty$. While this discrepancy is easy to verify for any concrete degree sequence $d_n$ covered by Theorem 2 (by numerically solving the relevant differential equations), we are lacking an analytic proof technique that can prove this for all such $d_n$ based only on the relevant differential equations.

To stimulate further research into such analytic proof techniques (which would most likely enhance the power of the differential equation method), as a proof of concept we now outline an analytic argument that works for degree sequences $d_n$ that are sufficiently irregular, i.e., under much stronger assumptions than Theorem 2. To this end we shall assume that in $d_n$ the number of vertices of degree $j$ satisfies $n_j/n \to r_j$ as $n \to \infty$ for all $j \in [\Delta]$, which in view of the (standard) combinatorial part (1) of Theorem 2 implies that in the uniform model we typically have

$$\frac{X_k(G_{d_n})}{n} \sim \frac{\mu}{n} \sim \frac{(\sum_{j \in [\Delta]} jr_j)^2}{2 \sum_{j \in [\Delta]} jr_j} =: \hat{\rho}_k \quad \text{and} \quad \frac{m}{n} \sim \frac{1}{2} \sum_{j \in [\Delta]} jr_j =: T,$$

where $a_n \sim b_n$ is the usual shorthand for $a_n = (1 + o(1))b_n$ as $n \to \infty$. To analyze the number of small edges $X_k(G'_{d_n})$ in the standard $d_n$-process, we shall track the following random variables after each step $i$: $U(i)$ counts the number of unsaturated vertices, $U_A(i)$ with $a \in [\Delta]$ counts the number of unsaturated
vertices $v$ with degree $d_v^{(n)} = a$ in $d_n$, and $X_{a,b}(i)$ with $a, b \in [\Delta]$ counts the number of edges whose two endvertices have degrees $a$ and $b$ in $d_n$. A standard application of the differential equation method shows that typically the following holds: for all steps $0 \leq i \leq m$ we have $U(i)/n = u(t) + o(1)$, $U_j(i)/n = u_j(t) + o(1)$, $X_{a,b}(i)/n = x_{a,b}(t) + o(1)$ with $t = i/n$, where the functions $u(t)$, $u_j(t)$ and $x_{a,b}(t)$ are the solutions to suitable differential equations that depend only on $r_1, \ldots, r_\Delta$. In the standard $d_n$-process we thus typically have

$$\frac{X_k(G_{d_n}^P)}{n} = \sum_{1 \leq a \leq b \leq k} \frac{X_{a,b}(m)}{n} = \rho_k + o(1) \quad \text{for} \quad \rho_k := \sum_{1 \leq a \leq b \leq k} x_{a,b}(T).$$

To prove a discrepancy in the number of small edges in both models, by our above discussion it thus remains to analytically show that $\rho_k \neq \tilde{\rho}_k$ (which we know to be true by our switching based Theorem 2). The key to proving this discrepancy is to prove suitable bounds on the values of $x_{a,b}(T)$. We know how to do this when the degree sequence $d_n$ is sufficiently far from regular, in a sense specified below. But new ideas are needed to do so for all the degree sequences $d_n$ covered by Theorem 2. We close this paper with further elaboration of this analytic approach, and a description of how it works on some degree sequences $d_n$. The relevant differential equations satisfy $u_j(0) = r_j$ and $u_j'(t) \geq -2u_j(t)/u(t)$, from which it follows that

$$u_j(t) \geq r_j e^{-\lambda(t)} \quad \text{with} \quad \lambda(t) := \int_0^t \frac{2}{u(x)} \, dx.$$ 

The simple degree-counting argument underlying Observation 5 shows that $U(i) \geq 2(m - i)/\Delta$, and one can similarly show that $u(t) \geq 2(T - t)/\Delta$, which by integration implies $\lambda(t) \leq \Delta \ln(T/(T - t))$. Inserting these estimates and $u(t) \leq 1$ into the relevant differential equations $x_{a,b}(0) = 0$ and $x_{a,b}'(t) = (1 + \mathbb{1}_{a \neq b})u_a(t)u_b(t)/u^2(t)$, we obtain that

$$x_{a,b}(T) = \int_0^T x_{a,b}'(t) \, dt \geq (1 + \mathbb{1}_{a \neq b})r_ar_b \int_0^T \left( \frac{T - t}{T} \right)^{2\Delta} \, dt = (1 + \mathbb{1}_{a \neq b})r_ar_b \frac{T}{2\Delta + 1},$$

which in turn implies that, say,

$$\rho_k = \sum_{1 \leq a \leq b \leq k} x_{a,b}(T) \geq \frac{T(\sum_{j \in [k]} r_j)^2}{2\Delta + 1} \geq \frac{T(\sum_{j \in [\Delta]} j^2)^2}{(2\Delta + 1)k^2}.$$ 

Comparing $\rho_k$ with $\tilde{\rho}_k$, we see that the desired discrepancy in the number of small edges is implied by $T/(2\Delta + 1) > k^2/(4T)$, or equivalently $\sum_{j \in [\Delta]} j^2 > k\sqrt{2\Delta + 1}$. While this sufficient condition is weaker than the one from Theorem 2 it does apply to sufficiently irregular degree sequences $d_n$. For example, this condition establishes non-contiguity for $\Delta \geq 7$ when half of the vertices of $d_n$ have degree one and half have degree $\Delta$ (i.e., where $r_1 = r_\Delta = 1/2$ and $k = 1$). We leave it as an open problem to find an improved analytic argument that recovers Theorems 1 and 2 based on analysis of the relevant differential equations.

Acknowledgements. We are grateful to the referees for a careful reading and helpful suggestions. Lutz Warnke would also like to thank Nick Wormald for many stimulating discussions about the behavior of the degree-restricted random $d_n$-process.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (2002), 47.
[2] K. Balińska and L. Quintas. Random graph models for physical systems. In *Graph Theory and Topology in Chemistry* (Athens, Ga., 1987), volume 51 of Studies in Physical and Theoretical Chemistry, pages 349–361. Elsevier, Amsterdam, 1987.
[3] K. Balińska and L. Quintas. The sequential generation of random $f$-graphs. Line maximal 2-, 3- and 4-graphs. *Computers & Chemistry* 14 (1990), 323–328.
[4] K. Balińska and L. Quintas. Big cycles in random edge maximal 2-graphs. In *Proceedings of the Twenty-Fourth Southeastern International Conference on Combinatorics, Graph Theory, and Computing* (Boca Raton, FL, 1993), volume 93, pages 131–141, 1993.

\[\text{With more work one can show that } u_j(t) = r_j e^{-\lambda(t)} \sum_{0 \leq \ell < j} \lambda^\ell(t)/\ell! \text{ and } u(t) = \sum_{j \in \Delta} u_j(t) \text{ hold.}\]
Figure 3: Counterexample to the simple switching heuristic from Sections 1.2 and 3.2.1 (we depict configuration-graphs as defined in Section 2.1 where each vertex is represented by an ellipse containing $\deg(v)$ many points): the graph $G^-$ is obtained from $G^+$ by the natural switching operation which replaces the edges $ab, xy$ with $ax, by$, but despite the degree difference $\max\{\deg(a), \deg(b)\} < \min\{\deg(x), \deg(y)\}$ the resulting graph $G^-$ is more likely than $G^+$; see Appendix A.1 for more details.

[44] R. Robinson and N. Wormald. Almost all regular graphs are Hamiltonian. *Random Structures and Algorithms* 5 (1994), 363–374.
[45] A. Ruciński and N. Wormald. Random graph processes with degree restrictions. *Combinatorics, Probability and Computing* 1 (1992), 169–180.
[46] A. Ruciński and N. Wormald. Random graph Processes with Maximum Degree 2. *The Annals of Applied Probability* 7 (1997), 183–199.
[47] A. Ruciński and N. C. Wormald. Connectedness of graphs generated by a random $d$-process. *Journal of the Australian Mathematical Society*, 72(1):67–85, 2002.
[48] A. Telcs, N. Wormald and S. Zhou. Hamiltonicity of random graphs produced by 2-processes. *Random Structures and Algorithms* 31 (2007), 450–481.
[49] L. Warnke. On the method of typical bounded differences. *Combinatorics, Probability and Computing* 25 (2016), 269–299.
[50] L. Warnke. On Wormald’s differential equation method. *Combinatorics, Probability and Computing*, to appear. arXiv:1905.08928
[51] L. Warnke and N. Wormald. Phase transition in evolving random graphs with bounded degrees. Manuscript (2022).
[52] N. Wormald. Differential equations for random processes and random graphs. *Annals of Applied Probability* 5 (1995), 1217–1235.
[53] N. Wormald. The perturbation method and triangle-free random graphs. *Random Structures and Algorithms* 9 (1996), 253–270.
[54] N. Wormald Models of random regular graphs. In *Surveys in Combinatorics*, pp. 239–298, Cambridge University Press, Cambridge (1999).
[55] N. Wormald. The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithm*, PWN, Warsaw (1999), pp. 73–155.

A Appendix

A.1 Counterexample to simple switching heuristic

In Sections 1.2 and 3.2.1 we discussed a simple switching heuristic which suggested that, when the degrees of a (configuration) graph $G^+$ satisfy $\max\{\deg(a), \deg(b)\} < \min\{\deg(x), \deg(y)\}$, then the switching operation which replaces the edges $ab, xy$ with $ax, by$ should decrease the probability of the resulting (configuration) graph $G^-$, i.e., $G^+$ should be more likely than $G^-$. Figure 3 depicts an example where this heuristic fails: with a computer program one easily checks that in the relaxed $d_n$-process (see Section 2.2) we have

$$\frac{\mathbb{P}(G^+_{d_n} \neq G^+)}{\mathbb{P}(G^-_{d_n} \neq G^-)} = \frac{Z(G^+)}{Z(G^-)} \approx 0.95652 < 1.$$ 

For the reader worried that this counterexample might be an artifact of using the relaxed $d_n$-process, we remark that if we interpret $G^+$ and $G^-$ as normal graphs (by contracting all points in an ellipse to one vertex), then with a computer program one easily checks that in the standard $d_n$-process we also have

$$\frac{\mathbb{P}(G^+_{d_n} = G^+)}{\mathbb{P}(G^-_{d_n} = G^-)} \approx 0.82164 < 1.$$ 

30
For switching arguments, the example from Figure 3 thus suggests that very local consideration on the level of pairs of graphs do not suffice, i.e., that looking at larger sets of graphs at once seems to be necessary (in this paper we implement this by considering clusters, see Section 3.2).

### A.2 Proof of Theorem 2(i) number of small edges \( X_k \) in uniform model

In this appendix we, for completeness, prove Theorem 2(i) using routine configuration model arguments.

**Proof of Theorem 2(i)** Let \( G^C_{d_n} \) be the random graph obtained from the configuration model \([12, 54, 20]\) on the degree sequence \( d_n \). As usual, \( X_k(G) \) denotes the number of small edges in \( G \). Since the maximum degree is \( \Delta = O(1) \), it follows from well-known transfer results (see, e.g., [20, Theorem 11.3]) that

\[
P( |X_k(G_{d_n}) - \mu| \geq \epsilon \mu) \leq O(1) \cdot P( |X_k(G^C_{d_n}) - \mu| \geq \epsilon \mu),
\]

so that we may henceforth focus on the number \( X_k := X_k(G^C_{d_n}) \) of small edges (which include loops) in the configuration model. To this end, let \( E_k \) denote the set of all possible small edges. Note that \( |E_k| = \left( \sum_{j \in [k]} jn_j \right) \). Writing \( I_e \) is the indicator variable for the event that the edge \( e \) is in \( G^C_{d_n} \), using (1) and \( \epsilon \gg n^{-1/2} \) it follows that \( \min \{ \sum_{j \in [k]} jn_j, m \} = \Omega(n) \) as well as

\[
E X_k = \sum_{e \in E_k} E I_e = \frac{|E_k|}{2m - 1} = \left(1 + O(n^{-1})\right) \frac{\left( \sum_{j \in [k]} jn_j \right)^2}{4m} = (1 + o(\epsilon))\mu.
\]

Gearing up towards an application of the bounded differences inequality for the configuration model (which follows from a simple modification of [54, Theorem 2.19] for regular \( d_n \); see also [20, Section 1.1.4]), note that a switching operation can change the number \( X_k \) of small edges by at most two. In view of \( \epsilon \gg n^{-1/2}, \Delta = \Theta(1) \) and \( \mu = \Omega(n) \), see (1), a routine application of the bounded differences inequality to \( X_k = X_k(G^C_{d_n}) \) yields

\[
P( |X_k - \mu| \geq \epsilon \mu) \leq 2 \exp \left( -\Theta(\epsilon^2 \mu^2) \right) \leq \exp \left( -\Theta(\epsilon^2 n) \right),
\]

which together with \( X_k = X_k(G^C_{d_n}) \) and inequality (69) completes the proof of the desired estimate (3).

\( \square \)

### A.3 Transfer of Theorem 3 to Theorem 2(ii)

In this appendix we show that our main technical result Theorem 3 for the relaxed \( d_n \)-process transfers to the standard \( d_n \)-process, i.e., implies our main result Theorem 2(ii). This kind of transfer argument is conceptually standard, but the technical details are more involved than usual. Recall that, in both the final graph \( G^P_{d_n} \) of the standard \( d_n \)-process and the final graph \( G^P_{d^{n*}} \) of the relaxed \( d_n \)-process, we each time condition on having degree sequence \( d_n \), i.e., we condition on the underlying process completing. The following lemma shows that the probability of completing is always at least a positive constant (stronger bounds can be deduced from [45] and the arguments therein, but we do not need this).

**Lemma 22** (Completion probability). For any integer \( \Delta \geq 1 \) there are \( m_0, c > 0 \) such that, for any degree sequence \( d_n = (d_1^{(n)}, \ldots, d_n^{(n)}) \in \{ 0, \ldots, \Delta \}^n \) with even degree sum \( \sum_{i \in [n]} d_i^{(n)} \) and at least \( m = m(d_n) = \frac{1}{2} \sum_{i \in [n]} d_i^{(n)} \geq m_0 \) edges, the (unconditional) standard and relaxed \( d_n \)-process each complete with probability at least \( c \).

As we shall see, this lemma (whose elementary proof we defer to the end of this appendix section) and Theorem 3 effectively reduce the proof of Theorem 2(ii) to a comparison argument between the unconditional models. To this end we below write \( H^P_{d_n,i} \) for the graph of the (unconditional) standard \( d_n \)-process after step \( i \), and also write \( H^P_{d^{n*},i} \) for the (unconditional) relaxed \( d_n \)-process after step \( i \).

**Proof of Theorem 2(ii)** Fix \( \alpha > 0 \) as in Theorem 3. The main idea is to compare the (unconditional) standard and relaxed \( d_n \)-process after \( m_0 := m - \lfloor \min \{ \alpha \mu / 4, m / 2 \} \rfloor \) steps, exploiting that during the final
\( m - m_0 \leq \alpha \mu / 4 \) steps the number of small edges can only change by at most most \( \alpha \mu / 4 \). Since the graph \( G_{d_n}^P \) has the same distribution as \( H_{d_{n,m}}^P \) conditioned on completing, by invoking Lemma \[22\] it follows that

\[
\mathbb{P} (|X_k(G_{d_n}^P) - \mu| \leq \alpha \mu / 4) \leq \frac{\mathbb{P} (|X_k(H_{d_{n,m}}^P) - \mu| \leq \alpha \mu / 4)}{\mathbb{P}(H_{d_{n,m}}^P \text{ completes})} \leq O(1) \cdot \mathbb{P} (|X_k(G_{d_n}^P) - \mu| \leq \alpha \mu / 2). \tag{70}
\]

Suppose that we condition on the graph \( H_{d_{n,m}}^P \) of the (unconditional) relaxed \( d_n \)-process after \( m_0 \) steps. Then the remaining \( m - m_0 = \Theta(\min(\mu, n)) \gg 1 \) steps of the relaxed \( d_n \)-process again correspond to a relaxed \( d'_{n,m} \)-process with a modified degree sequence \( d' \) (where the vertex degrees after \( m_0 \) steps are subtracted off) with even degree sum, so by Lemma \[22\] the conditional probability of completing is at least \( c = \Omega(1) \). Since the graph \( H_{d_{n,m}}^P \) conditioned on completing has the same distribution as \( G_{d_n}^P \), by invoking Theorem \[3\] it follows that

\[
\mathbb{P} (|X_k(H_{d_{n,m}}^P) - \mu| \leq \alpha \mu / 2) \leq O(1) \cdot \mathbb{P} (|X_k(G_{d_n}^P) - \mu| \leq \alpha \mu) \leq e^{-\Theta(n)}. \tag{71}
\]

In view of (70)–(71) it remains to compare the probabilities in both unconditional processes after \( m_0 \) steps, for which we shall use a standard step-by-step comparison argument (similar to \[24, 43, 30, 51\]). To this end, for mathematical convenience we henceforth write

\[
H_{d_{n,i}}^{P,+} := \pi(H_{d_{n,i}}^P)
\]

for the multigraph representation of the relaxed \( d_n \)-process after step \( i \), where the natural mapping \( \pi \) is defined as in Section \[2.1\] (it simply corresponds to the points of the relaxed \( d_n \)-process to vertices). We now fix a graph sequence \((G_0)_{0 \leq i \leq m_0}\) satisfying \(|X_k(G_{m_0}) - \mu| \leq \alpha \mu / 2\) that can be attained by \((H_{d_{n,m}}^P)_{0 \leq i \leq m_0}\). Comparing the probability with which the next edge is added in each process, for \( 0 \leq i < m_0 \) it follows that

\[
\mathbb{P}\left( H_{d_{n,i+1}}^P = G_{i+1} \bigg| \bigcap_{0 \leq j \leq i} \{ H_{d_{n,j}}^P = G_j \} \right) = \left( \frac{\Gamma_i}{\mid Q_i \mid} \cdot \mathbb{P}\left( H_{d_{n,i+1}}^{P,+} = G_{i+1} \bigg| \bigcap_{0 \leq j \leq i} \{ H_{d_{n,j}}^{P,+} = G_j \} \right) \right), \tag{73}
\]

where \( Q_i = \left( \frac{\Gamma_i}{\mid Q_i \mid} \right) / E(G_i) \) denotes the set of edges that can be added to \( H_{d_{n,i}}^{P,+} \), and \( \Gamma_i \) denotes the number of unsaturated vertices in \( G_i \) (as before). Note that \( |E(G_i) \cap \left( \frac{\Gamma_i}{2} \right)| \leq \Gamma_i \cdot \Delta \). The argument of Observation \[5\] gives \( \Gamma_i \geq 2(\mu - i) / \Delta \geq 2(\mu - m_0) / \Delta \gg 1 \). Using \( 1 - x \geq e^{-2x} \) for \( x \in [0,1/2] \), it follows that

\[
|Q_i| \geq \left( \frac{\Gamma_i}{2} \right) - \Gamma_i \Delta = \left( \frac{\Gamma_i}{2} \right) \cdot \left( 1 - \frac{2\Delta}{\Gamma_i - 1} \right) \geq \left( \frac{\Gamma_i}{2} \right) \cdot e^{-4\Delta^2 / (m - m_0)}. \tag{74}
\]

Since initially \( \mathbb{P}(H_{d_{n,0}}^P = G_0) = 1 = \mathbb{P}(H_{d_{n,0}}^{P,+} = G_0) \), by multiplying (73) it follows that

\[
\mathbb{P}\left( \bigcap_{0 \leq i \leq m_0} \{ H_{d_{n,i}}^P = G_i \} \right) \leq \exp\left( \frac{4\Delta^2 m_0}{m - m_0} \right) \cdot \mathbb{P}\left( \bigcap_{0 \leq i \leq m_0} \{ H_{d_{n,i}}^{P,+} = G_i \} \right). \tag{74}
\]

Recalling \( \Delta = O(1) \), by definition of \( m_0 \) we have \( \Delta^2 m_0 / (m - m_0) = O(1) \). Summing (74) over all graph sequences \((G_0)_{0 \leq i \leq m_0}\) satisfying \(|X_k(G_{m_0}) - \mu| \leq \alpha \mu / 2\) that can be attained by \((H_{d_{n,m}}^P)_{0 \leq i \leq m_0}\), it follows that

\[
\mathbb{P}(|X_k(H_{d_{n,m_0}}^P) - \mu| \leq \alpha \mu / 2) \leq O(1) \cdot \mathbb{P}(|X_k(H_{d_{n,m_0}}^{P,+}) - \mu| \leq \alpha \mu / 2) \leq O(1) \cdot \mathbb{P}(|X_k(H_{d_{n,m_0}}^{P,+}) - \mu| \leq \alpha \mu / 2), \tag{75}
\]

where for the last step we used \( X_k(H_{d_{n,m_0}}^{P,+}) = X_k(H_{d_{n,m_0}}^{P,+}) \), i.e., that \( H_{d_{n,m_0}}^{P,+} \) and its multigraph representation \( H_{d_{n,m_0}}^{P,+} = \pi(H_{d_{n,m_0}}^{P,+}) \) contain the same number of small edges.

Finally, combining (75) with (70)–(71) completes the proof of inequality (4) with \( \beta := \alpha / 4 \).
We conclude by giving the deferred proof of Lemma 22 which is based on counting arguments.

**Proof of Lemma 22.** We first consider the standard \( \mathbf{d}_n \)-process. When this process fails to complete, we are left with an edge-sequence \( \hat{\sigma} \) with less than \( m \) edges, to which no more edges can be added without violating the degree constraint (not allowing multiedges). We denote the set of such ‘stuck’ edge-sequences by \( S = S(\mathbf{d}_n) \), and write \( \mathbb{P}(\sigma) \) for the probability that the standard \( \mathbf{d}_n \)-process produces the edge-sequence \( \sigma \).

We claim that there is a function \( g \) such that the following holds for any \( \hat{\sigma} \in S \): the edge-sequence \( g(\hat{\sigma}) \) completes, is identical with \( \hat{\sigma} \) in the first \( m_\Delta := m - 3\Delta^3 \) steps, and we also have
\[
\mathbb{P}(\hat{\sigma}) \leq D \cdot \mathbb{P}(g(\hat{\sigma})) \quad \text{and} \quad |g^{-1}(g(\hat{\sigma}))| \leq D \quad \text{for} \quad D := 6\Delta^3 \cdot 6\Delta^3.
\]
Writing \( \text{Im}(g) \) for the image of \( g \), using this claim it follows that
\[
\mathbb{P}(\mathbf{d}_n \text{-process completes}) \geq \frac{\sum_{\sigma \in \text{Im}(g)} \mathbb{P}(\sigma)}{\sum_{\sigma \in \text{Im}(g)} \mathbb{P}(\sigma) + \sum_{\sigma \in \mathbb{P}(\hat{\sigma})} \mathbb{P}(\sigma)} \geq \frac{1}{1 + D^2},
\]
which establishes the desired lower bound on the probability of completing, with \( c := 1/(1 + D^2) \).

In the remainder we show the existence of such \( g \) with the claimed properties (to ensure that all described steps are feasible, we shall always tacitly use that the number of edges \( m \geq m_0 \) is sufficiently large, whenever necessary). Consider \( \hat{\sigma} \in S \) and the graph \( H \) containing all edges in \( \hat{\sigma} \). Note that the remaining set \( U \) of unsaturated vertices of \( H \) are pairwise adjacent (otherwise we can add more edges). Since the maximum degree in \( \mathbf{d}_n \) is \( \Delta \), it follows that there are at most \( |U| \leq \Delta \) unsaturated vertices at the end. Hence the number \( \ell \) of missing edges at most \( \ell \leq |U| \cdot \Delta \leq \Delta^2 \). We shall next pick \( \ell \) pairwise disjoint edges whose endpoints are not adjacent to any vertices of \( U \). To this end we initialize \( F = \mathcal{N}(U) \) as the set of forbidden vertices. At each step \( i = 1, \ldots, \ell \), we then pick from \( \hat{\sigma} \) the latest edge \( u_iv_i \) whose endpoints are disjoint from \( F \), and add \( u_iv_i \) to \( F \). Let \( (a_1, \ldots, a_\ell) \) be an arbitrary but deterministic enumeration of the multiset of unsaturated vertices of \( \hat{\sigma} \) where unsaturated vertex \( v \) appears \( d_v(n) - \text{deg}_H(v) \) times. We now construct the edge-sequence \( g(\hat{\sigma}) \) by replacing the edge \( u_iv_i \) in \( \hat{\sigma} \) with the sequence of edges \( u_{i+1}, v_i, a_{i+1}, \ldots, a_{i+\ell} \). By construction, the resulting new edge-sequence \( g(\hat{\sigma}) \) completes. Furthermore, since each edge \( u_iv_i \) appears after step \( m - (\ell \Delta + \ell \cdot 2\Delta) \geq m - 3\Delta^3 = m_\Delta \) in \( \hat{\sigma} \), the edge-sequence \( g(\hat{\sigma}) \) is identical with \( \hat{\sigma} \) in the first \( m_\Delta \) steps. Since after \( m_\Delta \) steps there at most \( 2(m - m_\Delta) = 6\Delta^3 \) unsaturated vertices left, it follows that
\[
|g^{-1}(g(\hat{\sigma}))| \leq \binom{6\Delta^3}{2} \frac{m - m_\Delta}{2} \leq (6\Delta^3 \cdot 6\Delta^3) = D.
\]
We write \( \mathcal{Q}_i(\sigma) \) for the set of edges that can be added in step \( i \) of the standard \( \mathbf{d}_n \)-process after adding the first \( i \) edges in \( \sigma \). By construction, it follows that
\[
\frac{\mathbb{P}(\hat{\sigma})}{\mathbb{P}(g(\hat{\sigma}))} = \frac{\prod_{m_\Delta \leq \ell < m} |\mathcal{Q}_i(\hat{\sigma})|^{-1}}{\prod_{m_\Delta \leq \ell < m} |\mathcal{Q}_i(g(\hat{\sigma}))|^{-1}} \leq \prod_{m_\Delta \leq \ell < m} |\mathcal{Q}_i(g(\hat{\sigma}))| \leq \binom{6\Delta^3}{2} 
\]
which completes the proof of the lower bound [76], as discussed.

Finally, we sketch the similar argument for the relaxed \( \mathbf{d}_n \)-process, where it again suffices to show that any stuck edge-sequence can be changed into a completing edge-sequence without changing the choices made in the first \( m_\Delta = m - O(1) \) steps. Note that in the relaxed \( \mathbf{d}_n \)-process only one vertex remains unsaturated, so the number of missing edges is \( \ell \leq \Delta \). Then we can again find the last \( \ell \) edges in \( \hat{\sigma} \) not incident to the unsaturated vertex, and use a similar edge-replacement procedure to construct suitable \( g(\hat{\sigma}) \), which then establishes the desired constant lower bound on the probability of completing. \( \square \)

### A.4 Proof of Remark 4: lower tail of \( X_k \) in degree restricted process

In this appendix we establish Remark 4 by complementing Theorem 3 with a lower tail result for the number \( X_k(G^\mu_n) \) of small edges, for which a simpler variant of the proof of Theorem 3 suffices. In particular, we can avoid the notion of good clusters and only rely on the basic cluster switching result Lemma 6 by exploiting that configuration-graphs \( G \) which at most \( X_k(G) \leq (1 - \epsilon)\mu \) small edges are contained in considerably more lower than upper clusters (in contrast to the proof of Theorem 3, where the relevant configuration-graphs \( G \) with \( X_k(G) \approx \mu \) are contained in approximately the same number of lower and upper clusters).
Theorem 23 (Number of small edges: lower tail). Suppose that assumptions of Theorem 2 hold. Then, for any $\epsilon \in (0, 1)$, the relaxed random graph process $G^{P*,*}_{ dn}$ satisfies $\mathbb{P}(X_k(G^{P*,*}_{ dn}) \leq (1 - \epsilon)\mu) \leq e^{-\Omega(n)}$.

Proof. We proceed similarly to Theorem 3, though some details are simpler here. We henceforth write $X_k(G)$ for the number of small edges in $G$, and assume that $\ell \leq (1 - \epsilon/2)\mu$. Inspired by Section 3.4, let $L$ denote the set of all lower clusters $C^-$ that contain some configuration-graph $G$ with $X_k(G) \leq \ell$ small edges, and let $U$ contain all upper clusters $C^+$ that are switching-partners of the lower clusters $C^-$ in $L$. For any $G$ with $X_k(G) \leq \ell$, note that inequality (64) remains valid, which in turn implies $L_G \geq 1$ and thus $G \in S_{C^+} \subseteq U$. The proof of Observation 10 shows that any $G \in S_{C^+} \subseteq U$ has $X_k(G) \leq \ell + 4\Delta$ small edges. Proceeding similarly to (37) and (36), by invoking Lemma 6 to all switching-partners it follows that

$$X_k(G) \leq \ell + 4\Delta$$

Since $\epsilon \mu = \Theta(n)$, we have $\ell + 4\Delta \leq (1 - 2\epsilon/3)\mu$, say. For any configuration-graph $G$ with $X_k(G) \leq \ell + 4\Delta$ small edges, by carefully inspecting equations (63)–(65) for $U_G$ and $L_G$ it is routine to see (by combining that (64) implies $L_G/U_G = 1 + o(1)$ for $\ell = \mu$, with the observation that the estimates (63) for $U_G$ and (64) for $L_G$ are increasing and decreasing in $\ell$, respectively) that there is $\tau = \tau(\epsilon) \in (0, 1)$ such that

$$L_G \geq (1 - \tau)^2(\sum_{j \in [k]} jn_j - 2\mu)^2 \quad \text{and} \quad U_G \leq (1 - \tau)^2(\sum_{j \in [k]} jn_j - 2\mu)^2.$$ 

Putting things together, for any $\ell \leq (1 - \epsilon/2)\mu$ it follows that

$$\sum_{G: X_k(G) \leq \ell} Z(G) \leq (1 - \tau)^2 \sum_{G: X_k(G) \leq \ell + 4\Delta} Z(G).$$ 

By iterating this inequality, it then follows similarly to equation (33) that

$$\mathbb{P}(X_k(G^{P*,*}_{ dn}) \leq (1 - \epsilon)\mu) \leq \frac{\sum_{G: X_k(G) \leq (1 - \epsilon)\mu} Z(G)}{\sum_{G: X_k(G) \leq (1 - \epsilon)\mu + 4\Delta} Z(G)} \leq (1 - \tau)^{|\epsilon \mu/(8\Delta)|} \leq e^{-\Omega(n)},$$

completing the proof.

Finally, the lower tail result Theorem 23 for $G^{P*,*}_{ dn}$ of course also transfers to the standard $d_n$-process $G^{P}_{ dn}$ (the comparison arguments from Section A.3 go through mutatis mutandis).