MAJORITY DYNAMICS: THE POWER OF ONE

ASHWIN SAH AND MEHTAAB SAWHNEY

ABSTRACT. Consider $n = \ell + m$ individuals, where $\ell \leq m$, with ℓ individuals holding an opinion A and m holding an opinion B. Suppose that the individuals communicate via an undirected network G, and in each time step, each individual updates her opinion according to a majority rule (that is, according to the opinion of the majority of the individuals she can communicate with in the network). This simple and well studied process is known as "majority dynamics in social networks". Here we consider the case where G is a random network, sampled from the binomial model $\mathbb{G}(n, p)$, where $(\log n)^{-1/16} \leq p \leq 1 - (\log n)^{-1/16}$. We show that for $n = \ell + m$ with $\Delta = m - \ell \leq (\log n)^{1/4}$, the above process terminates whp after three steps when a consensus is reached. Furthermore, we calculate the (asymptotically) correct probability for opinion B to "win" and show it is

$$\Phi\left(\frac{p\Delta\sqrt{2}}{\sqrt{\pi p(1-p)}}\right) + O(n^{-c}),$$

where Φ is the Gaussian CDF. This answers two conjectures of Tran and Vu and also a question raised by Berkowitz and Devlin.

The proof technique involves iterated degree revelation and analysis of the resulting degreeconstrained random graph models via graph enumeration techniques of McKay and Wormald as well as Canfield, Greenhill, and McKay.

1. INTRODUCTION

Considerable effort has been devoted to understanding exchange of opinions between individuals, seeing as it plays a major role in all types of social interaction. Of course, no simple model can accurately describe the behavior of many actors in complicated situations, so analysis and understanding of natural models for this problem has generated significant interest. A natural model, which has even been of interest in biophysics [17] and psychology [7], is so-called majority dynamics. It can be briefly described as follows. Given n individuals, let the network G capture the set of interactions between participants. For each participant $i \in \{1, \ldots, n\}$ with initial opinion $A_i^{(0)} \in \{\pm 1\}$, at every time step they adopt the majority opinion of their neighbors, that is, $A_i^{(t+1)} = \text{sign}(\sum_{j\sim i} A_j^{(t)})$. The key object of study therefore is understanding the propagation of opinions and how the local structure of the network affects these dynamics. We refer the reader to [3, 21–23] for further references regarding majority dynamics.

We now precisely define majority dynamics in terms of partitions of the graph G as this will be our focus in order to analyze it. Additionally, following [23] we follow the convention that if a participant's neighborhood is equally split between opinions then they keep the same opinion.

Definition 1.1. Given a graph G with bipartition $B_0 \sqcup R_0$, the majority dynamics at time i are computed as follows. Given $B_i \sqcup R_i$, a new partition $B_{i+1} \sqcup R_{i+1}$ by swapping precisely those vertices with strictly more of their neighbors on the other side of the partition. We say a color blue or red wins by step k if $B_k = B_0 \cup R_0$ or $R_k = B_0 \cup R_0$, respectively.

Our primary object of study in the paper concerns majority dynamics on random graphs $\mathbb{G}(n, p)$. First considered by Benjamini, Chan, O'Donnell, Tamuz, and Tan [2], research has primarily focused on establishing that majority dynamics terminates in a finite number of steps (see e.g. [10] and the

Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-1745302.

very recent [8] aimed at understanding sparse graphs) or understanding the even finer question of the distribution of which color majority dynamics terminates on [3, 23]

Our primary aim is to resolve a conjecture of Tran and Vu [23] which informally states that for majority dynamics in $\mathbb{G}(n, 1/2)$ even a bias of a single extra voter is sufficient to influence the final state by a positive probability. An essentially equivalent conjecture appears in subsequent work of Berkowitz and Devlin [3, Conjecture 7].

Conjecture 1.2 ([23, Conjecture 7]). Majority dynamics on $\mathbb{G}(2n + \Delta, 1/2)$ with sets $R_0 = \{v_1, \ldots, v_{n+\Delta}\}$ and $B_0 = \{v'_1, \ldots, v'_n\}$ converges to $R_k = R_0 \cup B_0$ with probability at least $1/2 + f(\Delta)$, where $f(\Delta) > 0$, as $n \to \infty$.

Tran and Vu [23] resolved this conjecture for (even) $\Delta \geq 12$, and Berkowitz and Devlin [3] resolved it for $\Delta \geq 3$.

We resolve this conjecture in full.

Theorem 1.3. There is an absolute constant c > 0 so that the following holds. Let $n \ge 1$. Let $0 \le \Delta \le (\log n)^{1/4}$ and let $n \ge 1$, $(\log n)^{-1/16} \le p \le 1 - (\log n)^{-1/16}$. In majority dynamics on $\mathbb{G}(2n + \Delta, p)$ with $|R_0| = n + \Delta$, with probability at least $1 - O(n^{-c})$ there is a color with more vertices at step 1 and that color wins by step 3. Furthermore, $|R_3| = 2n + \Delta$ with probability

$$\Phi\left(\frac{p\Delta\sqrt{2}}{\sqrt{\pi p(1-p)}}\right) + O(n^{-c})$$

where Φ is the cdf of $\mathcal{N}(0,1)$.

Remark. In particular, the event that both colors have the same size at step 1 occurs with decaying probability. The parameters $(\log n)^{-1/16}, (\log n)^{-1/4}$ can certainly be improved substantially but we have chosen to focus on the dense regime.

We note that Theorem 1.3 additionally resolves [23, Conjecture 8] regarding monotonicity of the limiting probabilities with respect to Δ , and the proof of Theorem 1.3 essentially answers [3, Question 2] (see in particular Theorems 2.1 and 3.5 which provide fine information about the sizes of the various parts after one and two days). We also note that this is the first work which gives an exact limiting probability for a specific color winning when that probability is strictly between 0 and 1 (other than the simple symmetric case $\Delta = 0$).

We anticipate that the techniques of this paper combined with recent refined asymptotic enumeration results of [15, 16] can yield further refinements of work of [3, 10]. In particular this may allow precise understanding of the number of steps before reaching stability for wider ranges of sparse pthan currently known.

1.1. Strategy. The broad structure of this paper breaks into 2 phases. In the first we substantially refine results of [3] in order to obtain a local limit theorem of how many vertices switched from blue to red and red to blue jointly. Our techniques rely extensively on graph enumeration results and models developed for degree sequences in $\mathbb{G}(n, p)$ by McKay and Wormald [20] and random bipartite graphs by McKay and Skerman [18]. Technically, this fine-tuned local limit theorem is not necessary to complete the analysis, and one can use a (non-joint) central limit theorem for the lead [3, Theorem 1] along with a precise computation of its mean (using for example techniques similar to Appendix A or [3, Lemma 12]).

The second and third days also use the graph enumeration techniques of McKay and Wormald [19] which were extended to bipartite graphs by Canfield, Greenhill, and McKay [5]; however at these stages we will only derive coarser information about the degree sequences and the number of red and blue vertices. In particular, we prove that given a sufficiently large initial lead, on the second day the number of red and blue vertices concentrate in intervals of length $O(n^{1-\eta})$ for an

absolute constant η . Further, we show that the side leading will have developed a substantial lead (of linear order). Then a final application of degree enumeration implies that with high probability that the process terminates on the third day, because it is unlikely for any vertex to have a degree so large that it overcomes the gap between sizes. (Simpler arguments in [2,3,23] show termination by the fourth day without enumeration at this stage.)

For these two stages we rely on a modification of a concentration argument developed by the Ferber, Kwan, Narayanan, and the authors [9] where a general framework for applying the second moment method with McKay-Wormald [19] enumeration formulas were used to resolve a conjecture of Füredi on the existence of "unfriendly" partitions in G(n, 1/2). The analysis here is substantially simpler as we need to track fewer parameters to guarantee convergence to termination within a finite time horizon. In particular, the analysis of the third day only requires a large-deviation bound on the degrees of vertices from what is expected in a degree-constrained random graph model, and the analysis of the second day has substantially simpler formulas due to the setting.

Acknowledgements. We thank Asaf Ferber, Vishesh Jain, Matthew Kwan, and Bhargav Narayanan for discussions related to this project.

2. Day one

As mentioned, the analysis for the first day involves proving a local limit theorem for the sizes of parts. Although a central limit theorem was shown by Berkowitz and Devlin [3] for the size of the red partition after one step, we will require understanding of how many vertices switched from blue to red and red to blue jointly, rather than the net amount. A central limit theorem may be derivable from their method, which involves moments. We need only a joint central limit theorem but we have chosen to demonstrate a local limit theorem to demonstrate the power of these techniques, and due to its independent interest. In particular, enumeration techniques allow one to reduce this computation to a local limit theorem for certain binomial random variables and various question about the model can be derived using these techniques.

The main result of this section is the following theorem. Its proof occupies Sections 2.1 to 2.4.

Theorem 2.1. There exists constants C, c > 0 such that the following holds. Let $n \ge 2$, let $0 \le \Delta \le (\log n)^{1/4}$ and let $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Let

$$\sqrt{\frac{n}{4\pi}}x' = x - \left(\frac{1}{2} + \frac{p(\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n,$$
$$\sqrt{\frac{n}{4\pi}}y' = y - \left(\frac{1}{2} + \frac{p(-\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n.$$

In majority dynamics on $\mathbb{G}(2n + \Delta, p)$ with $|R_0| = n + \Delta$, we have that

$$\mathbb{P}[|R_0 \cap R_1| = x \land |B_0 \cap B_1| = y]$$

$$= \frac{2}{n\sqrt{\pi(2+\pi)}} \exp\left(-\frac{(1+\pi)(x')^2 - 2(x'y') + (1+\pi)(y')^2}{2\pi(2+\pi)}\right) + O(n^{-1-c})$$

Furthermore for |x'| or $|y'| \ge C\sqrt{\log n}$ we have that

$$\mathbb{P}[|R_0 \cap R_1| = x \land |B_0 \cap B_1| = y] \le n^{-5}.$$

2.1. Initial estimates. We will first need some initial estimates regarding specific distributions which will show up when computing our local limit theorem. First, we record the probability that one binomial is greater than a different binomial with similar parameters. We defer its proof, which consists mainly of binomial manipulations and applications of well-known local central limit theorems, to Appendix A.

Lemma 2.2. There is c > 0 so that the following holds. We are given $n \ge 2$, $\tau \in \mathbb{Z}$ of magnitude at most $2(\log n)^{1/4}$, and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Suppose that $q = p + \alpha/n$ and $q' = p + \beta/n$ with $|\alpha|, |\beta| \le 40\sqrt{p(1-p)\log n}$. Then

$$\mathbb{P}[\operatorname{Bin}(n+\tau,q) \ge \operatorname{Bin}(n,q')] = \frac{1}{2} + \frac{p\tau + 1/2 + \alpha - \beta}{2\sqrt{\pi p(1-p)n}} + O(n^{-3/4}).$$

Next, we need to understand the mean and standard deviation of certain conditioned binomial random variables. The level of control required can be deduced from the Berry-Esseen theorem.

Lemma 2.3. There is c > 0 so that the following holds. We are given $n \ge 2$, $\tau \in \mathbb{Z}$ of magnitude at most $2(\log n)^{1/4}$, and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Suppose that $q, q' \in p \pm 40\sqrt{p(1-p)\log n/n}$. Let $X \sim \operatorname{Bin}(n + \tau, q)$ and $Y \sim \operatorname{Bin}(n, q')$. Let X^+ be X conditional on X > Y while X^- be X conditional on $X \le Y$. Then

$$\mathbb{E}X^{+} = pn + \sqrt{\frac{p(1-p)n}{\pi}} + O(n^{1/4}), \qquad \text{Var}\,X^{+} = \left(1 - \frac{1}{\pi}\right)p(1-p)n + O(n^{3/4})$$
$$\mathbb{E}X^{-} = pn - \sqrt{\frac{p(1-p)n}{\pi}} + O(n^{1/4}), \qquad \text{Var}\,X^{-} = \left(1 - \frac{1}{\pi}\right)p(1-p)n + O(n^{3/4}).$$

Proof. By Berry-Esseen, the joint distribution $(X - pn, Y - pn)/\sqrt{p(1-p)n}$ has cumulative distribution function differing from $\mathcal{N}(0, I_2)$ by $O(1/\sqrt{np(1-p)})$ pointwise. (Note that τ is small, so the shifts are negligible.) Let $Z_1, Z_2 \sim \mathcal{N}(0, 1)$. We see that

$$\mathbb{E}X^{+} = \mathbb{E}[X|X \ge Y] = pn + \sqrt{p(1-p)n} \frac{\mathbb{E}[X \mathbb{1}_{X \ge Y}]}{\mathbb{P}[X \ge Y]}$$
$$= pn + \sqrt{p(1-p)n} \frac{\mathbb{E}[Z_1 \mathbb{1}_{Z_1 \ge Z_2}] + O(n^{-1/4})}{\frac{1}{2} + O(n^{-1/4})}$$
$$= pn + \sqrt{\frac{p(1-p)n}{\pi}} + O(n^{1/4}).$$

The error terms $O(n^{-1/4})$ come from integrating the discrepancy in cumulative distribution functions over the region where $(X - pn, Y - pn)/\sqrt{p(1-p)n}$ is bounded by $O(\sqrt{\log n})$ and using a large deviation bound for binomials outside. Similarly,

$$\mathbb{E}(X^{+} - pn)^{2} = \mathbb{E}[(X - pn)^{2} | X \ge Y] = p(1 - p)n \frac{\mathbb{E}[Z_{1}^{2} \mathbb{1}_{Z_{1} \ge Z_{2}}] + O(n^{-1/4})}{\frac{1}{2} + O(n^{-1/4})}$$
$$= p(1 - p)n + O(n^{3/4}).$$

Therefore

Var
$$X^+ = \mathbb{E}[(X^+ - pn)^2] - (\mathbb{E}[X^+ - pn])^2 = \left(1 - \frac{1}{\pi}\right)p(1 - p)n + O(n^{3/4}).$$

Above, we used $\mathbb{E}[Z_1 \mathbb{1}_{Z_1 \ge Z_2}] = 1/(2\sqrt{\pi})$ and $\mathbb{E}[Z_1^2 \mathbb{1}_{Z_1 \ge Z_2}] = 1/2$. The computation for X^- is exactly analogous so we omit it.

Next, we need a local limit theorem for sums of these conditioned binomial random variables. The proof uses log-concavity of binomial distributions, along with a technique of Bender [1] which upgrades a Berry-Essen quality central limit theorem for a log-concave variable into a local central limit theorem. Though it follows by directly citing such results, we spell out the details in order to quantify the bounds.

Proposition 2.4. There is c > 0 so that the following holds. We are given $n \ge 2$, $m, \tau_k, \tau'_k \in \mathbb{Z}$ of magnitude at most $2(\log n)^{1/4}$, and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Suppose that $q_k, q'_k \in p \pm 40\sqrt{p(1-p)\log n}/n$. Let $X_k \sim \operatorname{Bin}(n+\tau_k,q_k)$ and $Y_k \sim \operatorname{Bin}(n+\tau'_k,q'_k)$. Let X_k^+ be X_k conditional on $X_k > Y_k$ while X_k^- be X_k conditional on $X_k \le Y_k$. Fix some $i \in [n+m]$ and sequence $\epsilon_k \in \{\pm 1\}$, and let

$$S = \sum_{k=1}^{i} \epsilon_k X_k^{-} + \sum_{k=i+1}^{n+\tau'} \epsilon_k X_k^{+}.$$

Then

$$\mathbb{P}[S=s] = \frac{1}{\sqrt{2\pi\sigma_S}} \exp\left(-\frac{(s-\mu_S)^2}{2\sigma_S^2}\right) + O\left(\frac{1}{n^{1/5}\sigma_S}\right)$$

for all $i \in [n+m]$ and $s \in \mathbb{Z}$, if μ_S and σ_S are the mean and variance of S.

Proof. Note that X_k, Y_k have probabilities converging to that of a normalized Gaussian, by a local limit theorem. Combining with tail bounds, we easily see that X_k^+ has well-behaved (centered) moments: its variance is $\Theta(p(1-p)n)$ and its centered third moment is $\Theta((p(1-p)n)^{3/2})$. The same holds for X_k^- . Therefore, the Berry-Esseen theorem shows that the cumulative distribution functions of S and $\mathcal{N}(\mu_S, \sigma_S^2)$ differ by $O(1/\sqrt{n})$ everywhere.

Next, note that X_k, Y_k have log-concave probability mass functions (on \mathbb{Z}) by log-concavity of binomials, hence (X_k, Y_k) has a jointly log-concave probability mass function in the sense that

$$p(a,b)p(c,d) \le p\left(\left\lfloor \frac{a+c}{2} \right\rfloor, \left\lfloor \frac{b+d}{2} \right\rfloor\right) p\left(\left\lceil \frac{a+c}{2} \right\rceil, \left\lceil \frac{b+d}{2} \right\rceil\right).$$

Conditioning on a convex set preserves log-concavity in this sense, hence (X_k, Y_k) conditional on $X_k \ge Y_k$ as well as conditional on $X_k < Y_k$ both have log-concave probability mass functions. By [11, Theorem 1.2] (which is essentially reproves to [13, Theorem 1.4] but allows functions to be 0), we see that the marginals of a distribution which is log-concave in this sense are log-concave. Therefore X_k^+ , X_k^- have log-concave probability mass functions.

Finally, convolutions of log-concave sequences are log-concave, so S has log-concave probability mass function. We established earlier that it satisfies a quantitative central limit theorem. We now quantify an argument of Bender [1] in order to deduce the desired result.

Let m_S be the mode of S. Above this value, the probability mass is nonincreasing, while below it is nondecreasing. First suppose that $s > m_S + n^{-1/4} \sigma_S$. We see that

$$\begin{split} \mathbb{P}[S=s] &\leq \frac{1}{\lceil n^{-1/4} \sigma_S \rceil} \mathbb{P}[s \leq S < s + n^{-1/4} \sigma_S] \\ &= \frac{1}{\lceil n^{-1/4} \sigma_S \rceil} \mathbb{P}[s \leq \mathcal{N}(\mu_S, \sigma_S^2) < s + n^{-1/4} \sigma_S] + O(n^{-1/4} \sigma_S^{-1}) \\ &= \frac{1}{\sqrt{2\pi} \sigma_S} \exp\left(-\frac{(s - \mu_S)^2}{2\sigma_S^2}\right) + O(n^{-1/5} \sigma_S^{-1}). \end{split}$$

The last line follows since $(s - \mu_S)^2/(2\sigma_S^2)$ is either stable up to a multiplicative factor of $(1 + O(n^{-1/5}))$ upon changing s by $\pm n^{-1/4}\sigma_S$ or is super-polynomially small (hence absorbed into the additive error term, since $\sigma_S^2 = \Theta(p(1-p)n^2)$ is polynomial). The lower bound is analogous. Furthermore, this holds for $s < m_S - n^{-1/4}\sigma_S$ by an identical argument. Therefore,

$$\mathbb{P}[S=s] = \frac{1}{\sqrt{2\pi\sigma_S}} \exp\left(-\frac{(s-\mu_S)^2}{2\sigma_S^2}\right) + O(n^{-1/5}\sigma_S^{-1})$$

as long as $s \notin m_S \pm n^{-1/4} \sigma_S$.

Finally, suppose that $m_S \leq s \leq m_S + n^{-1/4} \sigma_S$ (the symmetric case is analogous). We have

$$\mathbb{P}[S=s] \ge \mathbb{P}[S=s + \lceil n^{-1/4}\sigma_S \rceil] \\ = \frac{1}{\sqrt{2\pi}\sigma_S} \exp\left(-\frac{(s + \lceil n^{-1/4}\sigma_S \rceil - \mu_S)^2}{2\sigma_S^2}\right) + O(n^{-1/5}\sigma_S^{-1}) \\ = \frac{1}{\sqrt{2\pi}\sigma_S} \exp\left(-\frac{(s - \mu_S)^2}{2\sigma_S^2}\right) + O(n^{-1/5}\sigma_S^{-1}),$$

where the last equality uses a similar argument to above. This is in fact enough to demonstrate that $|m_S - \mu_S| = O(n^{-1/5}\sigma_S)$ (since if it were too far, then the sequence would have an increase-decrease pattern twice).

Finally, we obtain an upper bound via log-concavity:

$$\mathbb{P}[S=s] \leq \frac{\mathbb{P}[S=s+\lceil n^{-1/4}\sigma_S\rceil]^2}{\mathbb{P}[S=s+2\lceil n^{-1/4}\sigma_S\rceil]} = \frac{1}{\sqrt{2\pi\sigma_S}} \exp\left(-\frac{(s-\mu_S)^2}{2\sigma_S^2}\right) + O(n^{-1/5}\sigma_S^{-1})$$
nalogous computation and the fact that m_S, μ_S are close. The result follows.

by an analogous computation and the fact that m_S, μ_S are close. The result follows.

2.2. Degree sequence models. We now define a plethora of degree sequence models for random graphs that will be needed for the computations. At a high level, the work of McKay and Wormald [20] and McKay and Skerman [18] demonstrate that degrees of random graphs look independent conditional on, for example, total edge count. These models provide a way to encapsulate these facts quantitatively.

Definition 2.5 (Degree sequence domains). Let $I_n = \{0, \ldots, n-1\}^n$, E_n be the even sum sequences in this set, and I_n^{ℓ} be the sum ℓ sequences. We will typically denote elements of these sets by **d**. Let $I_{m,n} = \{0, \ldots, n\}^m \times \{0, \ldots, m\}^n$, $E_{m,n}$ be the sequences with equal sums on both sides, and $E_{m,n}^{\ell}$ be the sequences with equal sums ℓ . We will typically denote elements of these sets by s of length m and t of length n. We will denote random variable versions of these by capital boldface instead.

Definition 2.6 (True degree models). \mathcal{D}_p^n is the degree sequence distribution of $\mathbb{G}(n,p)$, which is a random variable supported on $E_n \subseteq I_n$. $\mathcal{D}_p^{m,n}$ is the degree sequence distribution of a bipartite graph with m vertices on one side and n on the other, each edge included independently with probability p, which is a random variable supported on $E_{m,n} \subseteq I_{m,n}$.

Definition 2.7 (Independent degree models). \mathcal{B}_p^n is the distribution of *n* independent Bin(n-1, p)random variables, supported on I_n . $\mathcal{B}_p^{m,n}$ is the distribution of m independent Bin(n,p) and n independent Bin(m, p) variables, supported on $I_{m,n}$.

Definition 2.8 (Conditioned degree models). \mathcal{E}_p^n is the distribution of \mathcal{B}_p^n conditioned on having even sum, supported on E_n . $\mathcal{E}_p^{m,n}$ is the distribution of $\mathcal{B}_p^{m,n}$ conditioned on having equal sums on both sides, supported on $E_{m,n}$.

Definition 2.9 (Integrated degree models). \mathcal{I}_p^n is the distribution sampled as follows. Sample $p' \sim \mathcal{N}(p, p(1-p)/(n^2-n))$, conditional on being in (0,1). Then sample from $\mathcal{E}_{p'}^n$. $\mathcal{I}_p^{m,n}$ is the distribution sampled as follows. Sample $p' \sim \mathcal{N}(p, p(1-p)/(2mn))$, conditional on being in (0,1). Then sample from $\mathcal{E}_{p'}^{m,n}$.

We are now ready to state the necessary results.

Theorem 2.10 (From [19, Theorem 3(ii)], [20, Theorem 3.6]). There is c > 0 and a growing function so that the following holds. Let $n \ge 2$ and suppose $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. There is an event $B_p^n \subseteq I_n$ such that $\mathbb{P}_{\mathcal{D}_p^n}[B_p^n] = n^{-\omega(1)}$ and uniformly for all $\mathbf{d} \in I_n \setminus B_p^n$ we have

$$\mathbb{P}_{\mathcal{D}_p^n}[\mathbf{D} = \mathbf{d}] = (1 + O(n^{-c}))\mathbb{P}_{\mathcal{I}_p^n}[\mathbf{D} = \mathbf{d}]$$

Theorem 2.11 (From [18, Theorem 1(a)]). There is c > 0 so that the following holds. Suppose $m, n \ge 2$ are such that $m = O(n\sqrt{\log n})$ and $n = O(m\sqrt{\log m})$. Suppose that $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Then there is an event $B_p^{m,n} \subseteq I_{m,n}$ such that $\mathbb{P}_{\mathcal{D}_p^{m,n}}[B_p^{m,n}] = O(\exp(-n^c))$ and uniformly for $(\mathbf{s}, \mathbf{t}) \in I_{m,n} \setminus B$ we have

$$\mathbb{P}_{\mathcal{D}_p^{m,n}}[\mathbf{S} = \mathbf{s} \wedge \mathbf{T} = \mathbf{t}] = (1 + O(n^{-1/3}))\mathbb{P}_{\mathcal{I}_p^{m,n}}[\mathbf{S} = \mathbf{s} \wedge \mathbf{T} = \mathbf{t}].$$

2.3. Computing a local limit result.

2.3.1. Transferring to an independent model. Now consider sampling $\mathbb{G}(2n+\Delta, p)$ and revealing the degrees among each part R_0 and B_0 as well from vertices in R_0 to B_0 and vice versa. We swap vertices purely based on this degree information. Since the sizes of the swapped parts are measurable with respect to this, which has distribution coming from three independent Erdős-Renyi graph models, we see by Theorems 2.10 and 2.11 that up to a multiplicative factor of $1 + O(n^{-c})$ and an additive error of $n^{-\omega(1)}$ it is enough to compute the relevant probabilities if the models on the parts are $\mathcal{I}_p^{n+\Delta}$, \mathcal{I}_p^n , and $\mathcal{I}_p^{n+\Delta,n}$ instead. We let **d** be the degree sequence of size $n + \Delta$, **d'** be the one of length n, and \mathbf{s}, \mathbf{t} be of length $m = n + \Delta$ and n.

At this point it is useful to define $R_0 = \{n + 1, ..., 2n + \Delta\}$ and $B_0 = [n]$ as usual and define the swapped sets R_1, B_1 purely as functions of a triple of degree sequences $(\mathbf{d}, \mathbf{d}', (\mathbf{s}, \mathbf{t}))$ from $I_{n+\Delta}, I_n$, and $I_{n+\Delta,n}$. (We define it in the obvious way so as to apply even if the total sum in I_n is not even, or the sums across both sides in $I_{n+\Delta,n}$ are not equal.)

With this in mind, the transference described above can be written quantitatively as

$$\mathbb{P}_{\mathcal{D}_{p}^{n+\Delta},\mathcal{D}_{p}^{n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y] = (1 + O(n^{-c}))\mathbb{P}_{\mathcal{I}_{p}^{n+\Delta},\mathcal{I}_{p}^{n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y] \\
\mathcal{I}_{p}^{n+\Delta,n} \\
+ O(n^{-\omega(1)}).$$
(2.1)

Furthermore,

$$\begin{aligned} \mathbb{P}_{\mathcal{I}_{p}^{n+\Delta},\mathcal{I}_{p}^{n}}[|R_{0}\cap R_{1}| &= x \land |B_{0}\cap B_{1}| = y] \\ &= \frac{1}{\int_{q_{0},q_{1},q_{2}\in[0,1]} d\mu(q_{0},q_{1},q_{2})} \int_{q_{0},q_{1},q_{2}\in[0,1]} \mathbb{P}_{\mathcal{E}_{q_{0}}^{n+\Delta},\mathcal{E}_{q_{1}}^{n}}[|R_{0}\cap R_{1}| &= x \land |B_{0}\cap B_{1}| = y]d\mu(q_{0},q_{1},q_{2}) \\ &= \int_{q_{0},q_{1},q_{2}\in p\pm 20\sqrt{p(1-p)\log n/n}} \mathbb{P}_{\mathcal{E}_{q_{0}}^{n+\Delta},\mathcal{E}_{q_{1}}^{n}}[|R_{0}\cap R_{1}| &= x \land |B_{0}\cap B_{1}| = y]d\mu(q_{0},q_{1},q_{2}) + O(n^{-10}), \end{aligned}$$

$$(2.2)$$

where μ denotes the measure of three independent Gaussians centered at p with variances $p(1-p)/((n + \Delta)^2 - (n + \Delta))$, $p(1-p)/(n^2 - n)$, and $p(1-p)/(2n(n + \Delta))$. The last line follows since such Gaussians lie in (0, 1) with exponentially good probability, and in fact are of size $p \pm 20\sqrt{p(1-p)\log n}/n$ with probability at least $1 - n^{-10}$.

At this point, we have nearly reached a model with independent Bernoulli sequences. However, we must condition on being even sum or having equal sum across two sides. To deal with this, we iteratively apply Bayes's rule to reduce to understanding genuinely independent random variables. This technique is closely related that in the proof given for [18, Theorem 8]. We have

$$\mathbb{P}_{\mathcal{E}^{n+\Delta}_{q_0},\mathcal{E}^n_{q_1}}_{\mathcal{E}^{n+\Delta,n}_{q_2}}[|R_0 \cap R_1| = x \land |B_0 \cap B_1| = y]$$

$$=\frac{\mathbb{P}_{\mathcal{B}_{q_0}^{n+\Delta},\mathcal{B}_{q_1}^n,\mathcal{B}_{q_2}^{n+\Delta,n}}[|R_0\cap R_1| = x \wedge |B_0\cap B_1| = y \wedge |\mathbf{D}|/2, |\mathbf{D}'|/2 \in \mathbb{Z} \wedge |\mathbf{S}| = |\mathbf{T}|]}{\mathbb{P}_{\mathcal{B}_{q_0}^{n+\Delta},\mathcal{B}_{q_1}^n,\mathcal{B}_{q_2}^{n+\Delta,n}}[|\mathbf{D}|/2, |\mathbf{D}'|/2 \in \mathbb{Z} \wedge |\mathbf{S}| = |\mathbf{T}|]}.$$
(2.3)

At this point, every event being considered is essentially coming from a sum of independent binomials or counting inequalities between independent binomials, so one should expect that these probabilities can be computed precisely. We can in fact do this, although we choose to iteratively simplify the expression by removing portions that "act independent".

2.3.2. Removing evenness. First, reveal $\mathcal{B}_{q_2}^{n+\Delta,n}$, that is, **S** and **T**. Further reveal $R_0 \cap R_1$ and $B_0 \cap B_1$. Clearly the remaining randomness is as follows: for $v \in R_0 \cap R_1$, we sample $d_v \sim$ $\operatorname{Bin}(n+\Delta,p_0)|_{\geq s_v}$, and similar for the other three parts. Note that with probability at least $1 - 2\exp(-\Omega(n))$ there are at least n/4 vertices $v \in R_0$ with $s_i \in pn \pm 100\sqrt{p(1-p)n}$ and at least n/4 vertices $v \in B_0$ with $t_i \in pn \pm 100\sqrt{p(1-p)n}$. For such vertices, regardless of whether it was revealed to be in $R_0 \cap R_1$ or $R_0 \setminus R_1$ (and similar for blue vertices), we see that the conditional distribution of its degree is some conditioned binomial that is easily checked to be equidistributed (mod 2) up to say an error of $O(n^{-1/4})$. If we reveal the degrees of every other vertex, then add up n/4 of these random variables, we obtain equidistribution (mod 2) where both values are attained with probability $1/2 + O(\exp(-n))$. Therefore the numerator and denominator satisfy

$$\frac{\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta},\mathcal{B}_{q_{1}}^{n},\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y \land |\mathbf{D}|/2, |\mathbf{D}'|/2 \in \mathbb{Z} \land |\mathbf{S}| = |\mathbf{T}|]}{\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta},\mathcal{B}_{q_{1}}^{n},\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|\mathbf{D}|/2, |\mathbf{D}'|/2 \in \mathbb{Z} \land |\mathbf{S}| = |\mathbf{T}|]} = \frac{(\frac{1}{4} + O(\exp(-n)))\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta},\mathcal{B}_{q_{1}}^{n},\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y \land |\mathbf{S}| = |\mathbf{T}|] + O(\exp(-\Omega(n)))}{(\frac{1}{4} + O(\exp(-n)))\mathbb{P}_{\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|\mathbf{S}| = |\mathbf{T}|] + O(\exp(-\Omega(n)))} = \frac{\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta},\mathcal{B}_{q_{1}}^{n},\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y \land |\mathbf{S}| = |\mathbf{T}|]}{\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta,n}}[|\mathbf{S}| = |\mathbf{T}|]} + O(\exp(-\Omega(n))). \tag{2.4}$$

In the last line, we used that the final denominator probability is large. This can be seen since it is the chance that two samples of $Bin(n(n + \Delta), q_2)$ equal each other. Being the same distribution supported on $[0, n(n+\Delta)]$, we see this occurs with probability at least $1/(n(n+\Delta)+1)$ by Cauchy-Schwarz.

2.3.3. Computing the numerator. In fact, this denominator can be computed precisely using a local limit theorem for binomial random variables. We therefore focus attention on computing the numerator. We have

$$\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta},\mathcal{B}_{q_{2}}^{n},\mathcal{B}_{q_{2}}^{n+\Delta,n}}[|R_{0}\cap R_{1}| = x \land |B_{0}\cap B_{1}| = y \land |\mathbf{S}| = |\mathbf{T}|]$$

$$= \sum_{|A|=x,|B|=y} \mathbb{P}_{\mathcal{B}}[R_{0}\cap R_{1} = A \land B_{0}\cap B_{1} = B]\mathbb{P}_{\mathcal{B}}[|\mathbf{S}| = |\mathbf{T}||R_{0}\cap R_{1} = A \land B_{0}\cap B_{1} = B]. \quad (2.5)$$

We can exactly compute the distribution of $|R_0 \cap R_1|$ and $|B_0 \cap B_1|$, which are independent. We make the following definitions for convenience going forward:

- $q_i = p + \alpha_i/n$ for $0 \le i \le 2$, where $|\alpha_i| \le 20\sqrt{p(1-p)\log n}$;
- $X_k \sim \operatorname{Bin}(n + \Delta 1, q_0)$ and $Y_k \sim \operatorname{Bin}(n, q_2)$ for $k \in [n + \Delta]$;
- Y_k^- is the distribution of Y_k conditional on $Y_k \leq X_k$ and Y_k^+ is conditional on $Y_k > X_k$;
- $Z_k \sim \operatorname{Bin}(n + \Delta, q_2)$ and $W_k \sim \operatorname{Bin}(n 1, q_1)$ for $k \in [n]$;
- Z_k^- is Z_k conditioned on $Z_k \leq W_k$ and Z_k^+ is conditioned on $Z_k > W_k$; $r = \mathbb{P}[X_1 \geq Y_1]$ and $b = \mathbb{P}[W_1 \geq Z_1]$.

We have

$$|R_0 \cap R_1| = \sum_{k=1}^{n+\Delta} \mathbb{1}[X_k \ge Y_k] \sim \operatorname{Bin}(n+\Delta, r),$$
$$|B_0 \cap B_1| = \sum_{k=1}^n \mathbb{1}[W_k \ge Z_k] \sim \operatorname{Bin}(n, b).$$

Additionally, we can compute the distributions of $|\mathbf{S}|$ and $|\mathbf{T}|$ conditional on $A = R_0 \cap R_1$ and $B = B_0 \cap B_1$, which are independent. It actually only depends on the sizes. If we condition on |A| = x and |B| = y, we have

$$|\mathbf{S}| \sim \sum_{k=1}^{x} Y_{k}^{-} + \sum_{k=x+1}^{n+\Delta} Y_{k}^{+},$$
$$|\mathbf{T}| \sim \sum_{k=1}^{y} Z_{k}^{-} + \sum_{k=y+1}^{n} Z_{k}^{+}.$$

At this point, computing the probability that $|\mathbf{S}| - |\mathbf{T}| = 0$ amounts to proving a local central limit theorem for all possible mixed sums and differences of these independent random variables. We have already done this in Proposition 2.4. Explicitly, this means that for |A| = x and |B| = y that

$$\mathbb{P}_{\mathcal{B}}[|\mathbf{S}| = |\mathbf{T}||R_0 \cap R_1 = A \land B_0 \cap B_1 = B] = \frac{1}{\sqrt{2\pi}\sigma_{x,y}} \exp\left(\frac{\mu_{x,y}^2}{2\sigma_{x,y}^2}\right) + O(n^{-1/5}\sigma_{x,y}^{-1})$$

where $\mu_{x,y}$, $\sigma_{x,y}^2$ are the mean and variance of $|\mathbf{S}| - |\mathbf{T}|$ conditional on |A| = x and |B| = y. By Lemma 2.3, we have

$$\sigma_{x,y}^2 = \left(2 - \frac{2}{\pi}\right)p(1-p)n^2 + O(n^{7/4}).$$

Therefore, let $\sigma = \sqrt{(2-2/\pi)p(1-p)}n$ and note that for |A| = x and |B| = y we have

$$\mathbb{P}_{\mathcal{B}}[|\mathbf{S}| = |\mathbf{T}||R_0 \cap R_1 = A \wedge B_0 \cap B_1 = B] = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{\mu_{x,y}^2}{2\sigma^2}\right) + O(n^{-1/5}\sigma^{-1}).$$

It remains to understand $\mu_{x,y}$. We have

$$\mu_{x,y} = x\mathbb{E}Y_k^- + (n+\Delta-x)\mathbb{E}Y_k^+ - y\mathbb{E}Z_k^- - (n-y)\mathbb{E}Z_k^+.$$

Claim 2.12. If $|x - n/2|, |y - n/2| \le \sqrt{n} \log n$ we have

$$\mu_{x,y} = \left(\frac{\alpha_1 - \alpha_0 - 2p\Delta}{\pi}\right)n + 2\sqrt{\frac{p(1-p)n}{\pi}}(x-y) + O(n^{4/5}).$$

Proof. From Lemma 2.3 we have

$$\mathbb{E}Y_k^+ = pn + \sqrt{\frac{p(1-p)n}{\pi}} + O(n^{1/4}) = \mathbb{E}Z_k^+,$$
$$\mathbb{E}Y_k^- = pn - \sqrt{\frac{p(1-p)n}{\pi}} + O(n^{1/4}) = \mathbb{E}Z_k^-.$$

The error terms are not good enough to do a direct replacement. However, from Lemma 2.2 we have

$$r = \mathbb{P}[X_1 \ge Y_1] = \frac{1}{2} + \frac{p(\Delta - 1) + 1/2 + \alpha_0 - \alpha_2}{2\sqrt{\pi p(1 - p)n}} + O(n^{-3/4}),$$

$$b = \mathbb{P}[W_1 \ge Z_1] = \frac{1}{2} + \frac{p(-\Delta - 1) + 1/2 + \alpha_1 - \alpha_2}{2\sqrt{\pi p(1 - p)n}} + O(n^{-3/4})$$

and additionally, by definition,

$$r\mathbb{E}Y_k^- + (1-r)\mathbb{E}Y_k^+ = \mathbb{E}Y_k = pn + \alpha_2,$$

$$b\mathbb{E}Z_k^- + (1-b)\mathbb{E}Z_k^+ = \mathbb{E}Z_k = pn + \alpha_2 + p\Delta + \alpha_2\Delta/n.$$

Therefore

$$\begin{split} &\mu_{x,y} - (n+\Delta)(pn+\alpha_2) + n(pn+\alpha_2 + p\Delta + \alpha_2\Delta/n) \\ &= (x - r(n+\Delta))(\mathbb{E}Y_k^- - \mathbb{E}Y_k^+) - (y - bn)(\mathbb{E}Z_k^- - \mathbb{E}Z_k^+) \\ &= (x - r(n+\Delta)) \cdot 2\sqrt{\frac{p(1-p)n}{\pi}} - (y - bn) \cdot 2\sqrt{\frac{p(1-p)n}{\pi}} + O(n^{4/5}) \\ &= 2\sqrt{\frac{p(1-p)n}{\pi}}((x - rn) - (y - bn)) + O(n^{4/5}). \end{split}$$

We deduce

$$\mu_{x,y} - 2\sqrt{\frac{p(1-p)n}{\pi}}(x-y) + O(n^{4/5})$$

= $-2\sqrt{\frac{p(1-p)n}{\pi}}n\left(\frac{p(\Delta-1) + 1/2 + \alpha_0 - \alpha_2}{2\sqrt{\pi p(1-p)n}} - \frac{p(-\Delta-1) + 1/2 + \alpha_1 - \alpha_2}{2\sqrt{\pi p(1-p)n}}\right)$
= $\left(\frac{-2p\Delta + \alpha_1 - \alpha_0}{\pi}\right)n.$

We now make the following definitions.

- Recall that $\sigma = \sqrt{(2-2/\pi)p(1-p)n}$.
- $\bullet\,$ We have

$$r^* = r^*(\alpha_0, \alpha_1, \alpha_2) = \frac{1}{2} + \frac{p(\Delta - 1) + 1/2 + \alpha_0 - \alpha_2}{2\sqrt{\pi p(1 - p)n}},$$

$$b^* = b^*(\alpha_0, \alpha_1, \alpha_2) = \frac{1}{2} + \frac{p(-\Delta - 1) + 1/2 + \alpha_1 - \alpha_2}{2\sqrt{\pi p(1 - p)n}},$$

which are within $O(n^{-3/4})$ of r, b by Lemma 2.3.

• We let

$$\mu^*(\alpha_0, \alpha_1, \alpha_2, x, y) = \left(\frac{\alpha_1 - \alpha_0 - 2p\Delta}{\pi}\right)n + 2\sqrt{\frac{p(1-p)n}{\pi}}(x-y),$$

which satisfies $\mu_{x,y} = \mu^*(\alpha_0, \alpha_1, \alpha_2, x, y) + O(n^{4/5})$ for $|x - n/2|, |y - n/2| \le \sqrt{n} \log n$ by Claim 2.12.

Now, continuing (2.5), we find for $|x - n/2|, |y - n/2| \le \sqrt{n} \log n$ that

$$\mathbb{P}_{\mathcal{B}_{q_{0}}^{n+\Delta}, \mathcal{B}_{q_{1}}^{n}, \mathcal{B}_{q_{2}}^{n+\Delta, n}}[|R_{0} \cap R_{1}| = x \land |B_{0} \cap B_{1}| = y \land |\mathbf{S}| = |\mathbf{T}|] \\
= \mathbb{P}[|R_{0} \cap R_{1}| = x]\mathbb{P}[|B_{0} \cap B_{1}| = y] \left(\frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{\mu_{x,y}^{2}}{2\sigma^{2}}\right) + O(n^{-1/5}\sigma^{-1})\right) \\
= \phi_{rn,r(1-r)n}(x)\phi_{bn,b(1-b)n}(y)\phi_{0,\sigma^{2}}(\mu_{x,y}) + O(n^{-2-1/5}) \\
= \phi_{r^{*}n,n/4}(x)\phi_{b^{*}n,n/4}(y)\phi_{0,\sigma^{2}}(\mu^{*}) + O(n^{-2-1/5}) \tag{2.6}$$

where $\phi_{a,b}$ denotes the pdf of the Gaussian with mean a and variance b. In the second line we used that the conditional probability in (2.5) given $R_0 \cap R_1$ and $B_0 \cap B_1$ depends only on their sizes. The third line used a local limit theorem for binomials and appropriately expanding out error terms. The fourth line is just manipulation of established error terms in ways that we have seen already. Note that this equality is actually true if either x or y deviates by at least $\sqrt{n} \log n$ from n/2 as then the probability $|R_0 \cap R_1| = x$ and $|B_0 \cap B_1| = y$ is super-polynomially small. Therefore, this equation is true in general.

It is also worth mentioning by similar logic that if either $|x - n/2| \ge C\sqrt{\log n}$ or $|y - n/2| \ge C\sqrt{\log n}$ then

$$\mathbb{P}_{\mathcal{B}_{q_0}^{n+\Delta}, \mathcal{B}_{q_1}^n, \mathcal{B}_{q_2}^{n+\Delta, n}}[|R_0 \cap R_1| = x \land |B_0 \cap B_1| = y \land |\mathbf{S}| = |\mathbf{T}|] \le \mathbb{P}_{\mathcal{B}}[|R_0 \cap R_1| = x] = O(n^{-10}).$$
(2.7)

2.4. Putting it together. Finally, note that the denominator of (2.4) is the probability that two samples of Bin $(n(n + \Delta), q_2)$ subtract to 0. This satisfies a local limit theorem (e.g. by [6]) and has mean 0 and variance $2q_2(1 - q_2)n(n + \Delta) = 2p(1 - p)n^2(1 + O(n^{-1/2}))$, so

$$\mathbb{P}_{\mathcal{B}_{q_2}^{n+\Delta,n}}[|\mathbf{S}| = |\mathbf{T}|] = \frac{1}{2\sqrt{\pi p(1-p)n}} + O(n^{-5/4}).$$
(2.8)

Putting together (2.1), (2.2), (2.3), (2.4), and (2.6) along with (2.8), we obtain for some absolute c > 0 that

$$\begin{split} \mathbb{P}_{\mathcal{D}_{p}^{n+\Delta},\mathcal{D}_{p}^{n}}[|R_{0}\cap R_{1}| &= x \land |B_{0}\cap B_{1}| = y] \\ &= \int_{|\alpha_{0}|,|\alpha_{1}|,|\alpha_{2}| \leq 20\sqrt{p(1-p)\log n}} \frac{\phi_{r^{*}n,n/4}(x)\phi_{b^{*}n,n/4}(y)\phi_{0,\sigma^{2}}(\mu^{*})}{1/(2\sqrt{\pi p(1-p)n})} d\nu(\alpha_{0},\alpha_{1},\alpha_{2}) + O(n^{-1-c}). \\ &= \int_{\alpha_{0},\alpha_{1},\alpha_{2} \in \mathbb{R}} \frac{\phi_{r^{*}n,n/4}(x)\phi_{b^{*}n,n/4}(y)\phi_{0,\sigma^{2}}(\mu^{*})}{1/(2\sqrt{\pi p(1-p)n})} d\nu(\alpha_{0},\alpha_{1},\alpha_{2}) + O(n^{-1-c}). \end{split}$$

where ν denotes the product measure of three independent Gaussians centered at 0 with variances p(1-p), p(1-p), p(1-p)/2, respectively. Note that the difference between sampling the α values from ν or the q values from μ is negligible.

Equivalently, we can sample $\beta_i = \alpha_i / \sqrt{p(1-p)}$ from Gaussians with variances 1, 1, 1/2 for i = 0, 1, 2, respectively. We have

$$\begin{split} &\int_{\alpha_0,\alpha_1,\alpha_2 \in \mathbb{R}} \frac{\phi_{r^*n,n/4}(x)\phi_{b^*n,n/4}(y)\phi_{0,\sigma^2}(\mu^*)}{1/(2\sqrt{\pi p(1-p)n})} d\nu(\alpha_0,\alpha_1,\alpha_2) \\ &= \frac{2\sqrt{\pi p(1-p)n}}{(\sqrt{2\pi})^5(\sqrt{\pi})(n/4)\sqrt{(2-2/\pi)p(1-p)n}} \int_{\beta} e^{-\frac{2(x-r^*n)^2+2(y-b^*n)^2}{n} - \frac{(\mu^*)^2}{(4-4/\pi)p(1-p)n^2} - \frac{\beta_0^2+\beta_1^2+2\beta_2^2}{2}} d\beta. \end{split}$$

When the values of r^*, b^*, μ^* are substituted in, this becomes a Gaussian integral in $\beta_0, \beta_1, \beta_2$. Let

$$\sqrt{\frac{n}{4\pi}}x' = x - \left(\frac{1}{2} + \frac{p(\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n,$$
$$\sqrt{\frac{n}{4\pi}}y' = y - \left(\frac{1}{2} + \frac{p(-\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n.$$

Then we deduce

$$-\frac{2(x-r^*n)^2+2(y-b^*n)^2}{n}-\frac{(\mu^*)^2}{(4-4/\pi)p(1-p)n^2}-\frac{\beta_0^2+\beta_1^2+2\beta_2^2}{2}$$

$$= -\frac{1}{2\pi p(1-p)} (x'\sqrt{p(1-p)} - \alpha_0 + \alpha_2)^2 - \frac{1}{2\pi p(1-p)} (y'\sqrt{p(1-p)} - \alpha_1 + \alpha_2)^2 - \frac{1}{4\pi(\pi-1)p(1-p)} ((x'-y')\sqrt{p(1-p)} + \alpha_1 - \alpha_0)^2 - \frac{\beta_0^2 + \beta_1^2 + 2\beta_2^2}{2} = -\frac{1}{2\pi} (x'-\beta_0 + \beta_2)^2 - \frac{1}{2\pi} (y'-\beta_1 + \beta_2)^2 - \frac{1}{4\pi(\pi-1)} (x'-y'+\beta_1 - \beta_0)^2 - \frac{\beta_0^2 + \beta_1^2 + 2\beta_2^2}{2}.$$

Changing variables via $\alpha_i = \sqrt{p(1-p)}\beta_i$ therefore yields

$$\begin{split} &\int_{\alpha_0,\alpha_1,\alpha_2 \in \mathbb{R}} \frac{\phi_{r^*n,n/4}(x)\phi_{b^*n,n/4}(y)\phi_{0,\sigma^2}(\mu^*)}{1/(2\sqrt{\pi p(1-p)n})} d\nu(\alpha_0,\alpha_1,\alpha_2) \\ &= \frac{1}{\pi^2\sqrt{\pi-1n}} \int_{\beta} e^{-\frac{1}{2\pi}(x'-\beta_0+\beta_2)^2 - \frac{1}{2\pi}(y'-\beta_1+\beta_2)^2 - \frac{1}{4\pi(\pi-1)}(x'-y'+\beta_1-\beta_0)^2 - \frac{\beta_0^2+\beta_1^2+2\beta_2^2}{2}}{d\beta} \\ &= \frac{2}{n\sqrt{\pi(2+\pi)}} \exp\left(-\frac{(1+\pi)(x')^2 - 2(x'y') + (1+\pi)(y')^2}{2\pi(2+\pi)}\right). \end{split}$$

Furthermore, if either $|x - n/2| \ge C\sqrt{\log n}$ or $|y - n/2| \ge C\sqrt{\log n}$ for appropriate C > 0 then we obtain a bound of size $O(n^{-5})$, which is easily seen using (2.7) along with Section 2.3.3. (2.1), (2.2), (2.3), (2.4), and (2.8). This completes the proof of Theorem 2.1.

3. TRACKING THE REMAINDER

Now we adapt the approach of Ferber, Kwan, Narayanan, and the authors [9] to analyze the remainder of the majority dynamics process. Note that it is key that we computed what the leads were after day one at the scale of \sqrt{n} , since the techniques in that work only constrain objects at the scale $O(n^{1-\eta})$. However, essentially the same set of coarse data that is tracked in that work, along with the information from Theorem 2.1, will allow us to perform an analysis of the remaining process via iterated revelation.

3.1. Tracking degree parameters. We first define the parameters that will be tracked, which are basically the joint degree distributions of each part of the graph to each of the other parts.

Given $k \ge 1$ and $x \in \{0,1\}^k$ of the form (x_0,\ldots,x_{k-1}) , let $V_x = \bigcap_{i=0}^{k-1} X_i(x)$ where $X_i(0) = R_i$ and $X_i(1) = B_i$. Additionally, for $v \in B_0 \cup R_0$ let

$$\deg^{(k)} v = ((\deg_{V_x} v - p|V_x|) / \sqrt{p(1-p)n})_{x \in \{0,1\}^k}.$$

Finally, for $x \in \{0, 1\}^k$ let \mathcal{L}_x be the distribution of deg^(k) v if we sample a uniform $v \in V_x$ (implicitly assuming it is nonempty).

It will be helpful to recall the following definition of Kolmogorov distance.

Definition 3.1. If \mathcal{L} and \mathcal{L}' are probability distributions on \mathbb{R}^d , the Kolmogorov distance $d_{\mathrm{K}}(\mathcal{L}, \mathcal{L}')$ is the supremum of $|\mathcal{L}(A) - \mathcal{L}'(A)|$ over all sets $A = (-\infty, a_1] \times \cdots \times (-\infty, a_d]$, where $a_1, \ldots, a_d \in \mathbb{R}$.

3.2. Additional data from day one. We now quickly derive certain coarse degree statistics arising from day one. These results are substantially less delicate than the previous section.

Lemma 3.2. There are C, c > 0 such that the following holds. Let $n \ge 2$, let $0 \le \Delta \le (\log n)^{1/4}$ and let $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. For each $x \in \{0, 1\}$ we have $d_{\mathrm{K}}(\mathcal{L}_x, \mathcal{N}(0, I_2)) \le n^{-c}$ with probability $1 - O(n^{-5})$ under majority dynamics on $\mathbb{G}(2n + \Delta, p)$ with $|R_0| = n + \Delta$. Furthermore, \mathcal{L}_x is supported on $[-C\sqrt{\log n}, C\sqrt{\log n}]$ with probability $1 - O(n^{-5})$. (Hence we can choose $\widehat{\mathcal{L}}_x$ to have the same support.) *Remark.* A version of the above result when p = 1/2 and $\Delta = 0$ with a weaker probability bound appears in [9, Section 4.1], which is also sufficient for our purposes.

Sketch. It suffices to check it for x = 0, as the remaining case is analogous. The support claim is immediate by a union bound over all vertices. The Kolmogorov distance claim follows from the degree models in Section 2. Specifically, consider the reduction from the true degree sequence model to the independent degree model, and then note that regardless of the revealed q_0, q_1, q_2 in $p\pm 20\sqrt{p(1-p)\log n/n}$, each vertex in R_0 has joint degree distribution extremely close to a correctly normalized Gaussian. Everything is now independent, so Chernoff on the number of vertices with degrees $pn + \sqrt{p(1-p)n[x, x + n^{-c}]}$, ranging over a polynomial-sized set of values x proves the desired result. We must divide by the probability that the number of edges within each part is even and that the number of edges in the bipartite part agrees across both sides, as in Sections 2.3.2 and 2.3.3, but these are polynomial probabilities which do not affect the bound significantly.

3.3. Data from day two. We are now in position to derive the necessary data for day two. We show that given a substantial lead after day one that this leads grows to a linear size on the following day with high probability. To do this we reveal certain information and condition on certain high probability outcomes.

- **E1** Reveal all $\deg_{B_0} v$ and $\deg_{R_0} v$ values, which is enough to execute day one and determine R_1, B_1 .
- **E2** Furthermore, we assume that this revelation satisfies Lemma 3.2 and we let $|R_0 \cap R_1| = x$ and $|B_0 \cap B_1| = y$, defining

$$\sqrt{\frac{n}{4\pi}}x' = x - \left(\frac{1}{2} + \frac{p(\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n,$$
$$\sqrt{\frac{n}{4\pi}}y' = y - \left(\frac{1}{2} + \frac{p(-\Delta - 1) + 1/2}{2\sqrt{\pi p(1 - p)n}}\right)n.$$

as in the statement of Theorem 2.1.

- **E3** We may assume that our revelation gave rise to values $x', y' = O(\sqrt{\log n})$ with probability at least $1 n^{-5}$ by Theorem 2.1.
- **E4** Finally, the number of edges between the two parts in the initial partition is $pn(n + \Delta) + O(n^{3/2-1/5})$ with super-polynomially high probability, so we may assume that our revelation gave rise to such a number of edges. Similarly within each part, we may assume we have $p\binom{n}{2} + O(n^{3/2-1/5})$ edges.

In order to execute day two, we reveal $\deg_T v$ for $T \in \{R_0 \cap R_1, R_0 \cap B_1, B_0 \cap R_1, B_0 \cap B_1\}$. Depending on the total degree from v to R_1 and B_1 , as well as whether $v \in B_1$ or R_1 in the case of ties, we know where v lands in the next step.

Claim 3.3. There is an absolute c > 0 such that the following holds. Given revelations and assumptions **E1** to **E4**, over the remaining randomness for each $x \in \{0,1\}^3$ the number of vertices in V_x is concentrated at a scale $O(n^{1-c})$. In particular, V_x is in an interval of length $O(n^{1-c})$ around its mean with probability at least $1 - n^{-c}$.

This implies that $|R_2|$ and $|B_2|$ are concentrated.

ŀ

To do this, we attempt to understand the degree distribution better. First, let

$$\beta_i^R = \frac{\deg_{R_0} v_i - p(n + \Delta - 1)}{\sqrt{p(1 - p)(n + \Delta - 1)}}, \qquad \beta_i^B = \frac{\deg_{B_0} v_i - pn}{\sqrt{p(1 - p)n}}$$

for $i \in R_0$ and

$$\beta_j^R = \frac{\deg_{R_0} v_j - p(n+\Delta)}{\sqrt{p(1-p)(n+\Delta)}}, \qquad \beta_j^B = \frac{\deg_{B_0} v_j - p(n-1)}{\sqrt{p(1-p)(n-1)}}$$

for $j \in B_0$.

Given $v \in R_0 \cup B_0$, look at

$$\deg^{(1)} v = (\alpha_0, \alpha_1), \qquad \deg^{(2)} v = (\rho_{00}, \rho_{01}, \rho_{10}, \rho_{11})$$

where $\alpha_i = (\deg_{V_i} v - p|V_i|)/\sqrt{p(1-p)n}$ and $\rho_{ij} = (\deg_{V_{ij}} v - p|V_{ij}|)/\sqrt{p(1-p)n}$. Note that α_0, α_1 are determined given the revealed information, and that

$$\rho_{00} + \rho_{01} = \alpha_0, \qquad \rho_{10} + \rho_{11} = \alpha_1$$

hold. Furthermore ρ_{00} and ρ_{01} are independent given the revealed information, and their probability distributions can be determined by Appendix B.1 and Appendix B.2, respectively.

By the first parts of Proposition B.4 and Proposition B.7, we see that with super-polynomially high probability the ρ_{ij} are bounded by $(\log n)^{25}/\sqrt{p(1-p)}$. Therefore we may assume that all vertices satisfy such a bound when revealing the new joint distribution of degrees. Furthermore, using the conditions on x', y', in both cases we will be able to apply Proposition B.5 or Proposition B.8, as long as we verify the necessary condition regarding $\sum \beta$ (in the notation of those propositions). Specifically, one needs for $v \in R_0$ that

$$\sum_{i\in R_0\backslash v}\beta_i^R=O(n^{5/6})=\sum_{j\in B_0}\beta_j^R$$

while for $v \in B_0$ one needs

$$\sum_{i \in R_0} \beta_i^B = O(n^{5/6}) = \sum_{j \in B_0 \setminus v} \beta_j^B.$$

This follows since we assumed the number of edges between the two parts in the initial partition is $pn(n + \Delta) + O(n^{3/2-1/5})$, and similar for within each of the two parts.

Now Proposition B.5 and Proposition B.8 show for $v \in R_0$ that

$$\mathbb{P}[\rho_{00} = \gamma] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \alpha_0 - \frac{\sum_{i \in V_{00}} \beta_i^R}{n/2}\right)^2\right),\\ \mathbb{P}[\rho_{10} = \gamma] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \alpha_1 - \frac{\sum_{j \in V_{10}} \beta_j^R}{n/2}\right)^2\right),\tag{3.1}$$

absorbing negligible errors such as the difference of Δ between the number of vertices of R_0 and B_0 .

Note that if we reveal the neighborhood of v, then any $w \in R_0$ will have essentially the same conditional distribution (the effect of revealing this neighborhood is to slightly adjust some degrees, which negligibly affects $\sum_{j \in R_{10}} \beta_j$, for instance).

Using this observation, a second-moment computation demonstrates that the number of vertices $v \in R_0 \cap R_1$ with

$$\rho_{00} + \rho_{10} - \rho_{01} - \rho_{11} \ge \frac{p}{\sqrt{p(1-p)n}} (|V_{01}| + |V_{11}| - |V_{00}| - |V_{10}|) = \frac{p}{\sqrt{p(1-p)n}} (|B_1| - |R_1|) \quad (3.2)$$

is concentrated (which corresponds to $v \in R_0 \cap R_1$ being in R_2 after day two is revealed). We forgo the computational details (for similar arguments of this form, see [9, Section 4.3.6]). The other cases are analogous. This completes the justification of Claim 3.3.

We quickly record that for $v \in B_0$, one obtains instead

$$\mathbb{P}[\rho_{00} = \gamma] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \alpha_0 - \frac{\sum_{i \in V_{00}} \beta_i^B}{n/2}\right)^2\right),\\ \mathbb{P}[\rho_{10} = \gamma] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \alpha_1 - \frac{\sum_{j \in V_{10}} \beta_j^B}{n/2}\right)^2\right).$$
(3.3)

Claim 3.4. There is an absolute c > 0 such that the following holds. Given revelations and assumptions E1 to E4, over the remaining randomness we have

$$\mathbb{E}|R_0 \cap R_2| = n\mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge \frac{2}{\sqrt{\pi}} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right] + O(n^{1-c})$$

and

$$\mathbb{E}|B_0 \cap R_2| = n\mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge -\frac{2}{\sqrt{\pi}} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right] + O(n^{1-c}),$$

$$\mathbb{E}|V_{T_1, T_0, T_2}| = \mathbb{E}|V_{T_1, T_1, T_2}| + O(n^{1-c}) \text{ for all } x_1, x_1', x_2, x_3 \in \{0, 1\}.$$

while also $\mathbb{E}|V_{x_1x_2x_3}| = \mathbb{E}|V_{x_1x_2x_3}| + O(n^{1-c})$ for all $x_1, x_1', x_2, x_3 \in \{0, 1\}$

For this we note that if v has parameters (α_0, α_1) defined above then

$$\rho_{00} + \rho_{10} - \rho_{01} - \rho_{11} = (2\rho_{00} - \alpha_0) + (2\rho_{10} - \alpha_1),$$

which is the sum of two independent discrete Gaussians of standard deviation 1 and discretization $2/\sqrt{p(1-p)n}$ by (3.1) and (3.3). A simple computation shows the sum of two such discrete Gaussians with the given error terms (and tail bounds) is a corresponding discrete Gaussian. In particular, we see for $v \in R_0$ that

$$\mathbb{P}[(2\rho_{00} - \alpha_0) + (2\rho_{10} - \alpha_1) = \tau] = \frac{1}{\sqrt{p(1-p)n}} \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{1}{4} \left(\tau - \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^R}{n/2}\right)^2\right) + O(n^{-1/2 - 1/12})$$

for τ on an appropriate integer lattice of discretization $2/\sqrt{p(1-p)n}$. A similar formula with β_i^B holds for $v \in B_0$. For any $v \in R_0$ we see that

$$\mathbb{P}[(3.2) \text{ for } v] = \mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^R}{n/2} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right] + O(n^{-1/13})$$

hence

$$\mathbb{E}|R_0 \cap R_2| = n\mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^R}{n/2} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right] + O(n^{12/13})$$

Here the error term comes from the earlier term, as well as the possibility of vertices that are exactly balanced (of which there are few by the given computations) which may go a different way depending on its day one (not day zero) affiliation.

Similarly, for any $v \in B_0$ we have

$$\mathbb{P}[(3.2) \text{ for } v] = \mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^B}{n/2} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right]$$

and thus

$$\mathbb{E}|B_0 \cap R_2| = n\mathbb{P}_{Z \sim \mathcal{N}(0,2)} \left[Z \ge \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^B}{n/2} + \sqrt{\frac{p}{(1-p)n}} (|B_1| - |R_1|) \right] + O(n^{12/13})$$

Finally, it suffices to compute the average of these β_i^R and β_i^B quantities over $V_{00} \cup V_{10} = R_1$. From Lemma 3.2 we know that the empirical normalized joint degree distributions for R_0 and B_0 are close to $\mathcal{N}(0, I_2)$. Therefore the degree distribution of R_1 is close to that of $(Z_1, Z_2) \sim \mathcal{N}(0, I_2)$ conditional on $Z_1 \geq Z_2$ (where Z_1 corresponds to the parameter $(\deg_{R_0} v - p(n + \Delta))/\sqrt{p(1 - p)(n + \Delta)})$). Therefore the expected value of this ensemble is within $O(n^{-c})$ of

$$\mathbb{E}[Z_1|Z_1 \ge Z_2] = \frac{1}{\sqrt{\pi}}.$$

This shows

$$\frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^R}{n} = \frac{1}{\sqrt{\pi}} + O(n^{-c}) = \frac{\sum_{i \in V_{00} \cup V_{10}} \beta_i^B}{n}.$$

Then the first part of Claim 3.4 follows.

The second part of Claim 3.4 follows from that fact that all of the expressions for probabilities above based on v are independent of the values (α_0, α_1) , so we can sum over $v \in R_0 \cap R_1$, for instance, by just summing over those $v \in R_0$ which have $\alpha_0 > \alpha_1$, of which there are $n/2 + O(\sqrt{n}(\log n))$ by **E2**.

Putting Claims 3.3 and 3.4 together and simplifying the sum of the expectations in Claim 3.4, we obtain the following information about the distribution of the sizes after day two.

Theorem 3.5. There is an absolute c > 0 such that the following holds. Given revelations and assumptions **E1** to **E4**, over the remaining randomness we have

$$|R_2| = n + n \int_{-\eta}^{\eta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(u - \sqrt{\frac{2}{\pi}}\right)^2\right) du + O(n^{1-c})$$

with probability at least $1 - n^{-c}$, where $\eta = \sqrt{p/(2(1-p)n)}(|R_1| - |B_1|)$.

Remark. The integral is a signed integral. In particular its sign is the same as η .

3.4. Finishing on day three. To finish we now use rather coarse consequences of degree enumeration to prove that every vertex is of the appropriate color. Note that Theorem 2.1 tells us the distribution of the lead $|R_1| - |B_1|$, and Theorem 3.5 tells us, in terms of the lead after day one, what the lead after day two is concentrated at. Furthermore, if the lead at day one is sufficiently positive then so will be the lead at day two with high probability. For example, for $p \leq 1/2$ a lead of \sqrt{n} yields a lead of $\Omega(n\sqrt{p})$.

Using arguments in [2,3,23] one can immediately prove that the side leading after day two has colored all the vertices in two further days. This along with Theorems 2.1 and 3.5 will immediately justify Theorem 1.3 except that we can only guarantee it ends by day four. The arguments in [2,3,23] appear not sufficiently refined to deliver the day three result.

Proof of Theorem 1.3. Make revelations as in **E1** to **E4**. Then reveal the information deg⁽²⁾ v for all $v \in R_0 \cup B_0$, which allows us to determine the parts up to the end of day two. Reveal such that Claims 3.3 and 3.4 and Theorem 3.5 are satisfied. Let

$$\eta = \sqrt{\frac{p}{2(1-p)n}}(|R_1| - |B_1|)$$

and note

$$|R_2| - |B_2| = 2n \int_{-\eta}^{\eta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(u - \sqrt{\frac{2}{\pi}}\right)^2\right) du + O(n^{1-c})$$

from Theorem 3.5, for some small absolute constant $c \in (0, 1/4)$.

We wish to show that over all of the randomness (including the revealed randomness), if $\eta > 0$ then red will win in 3 days while if $\eta < 0$ then red will win in 3 days.

First if $\eta \ge 10 + \sqrt{2 \log p}$ then $|B_2| \le pn/5$. We see that red wins after day three with extremely high probability since the initial graph has minimum degree at least pn/2 with probability at least $1 - O(\exp(-\Omega(pn)))$, and this forces every vertex to have more neighbors on the red side than blue side after day two is finished.

Similarly, if $-\eta \ge 10 + \sqrt{2\log p}$ then blue wins after day three with extremely high probability.

The case $|\eta| \leq n^{-c/2}$ occurs with probability $O(n^{-c/4})$ by the local limit theorem of Theorem 2.1, so we ignore it.

Finally, without loss of generality we consider the case $n^{-c/2} < \eta < 10 + \sqrt{2 \log p}$ (the opposite case being analogous except with red and blue switched).

Now to determine what happens on day three, we reveal $\deg^{(3)} v$ for all $v \in R_0 \cup B_0$. This comes from another ensemble of degree-constrained distributions, so we apply the results of Appendix B again, between all pairs of the four parts V_x for $x \in \{0, 1\}^2$. First, by Claims 3.3 and 3.4 we have

$$|V_x| \ge n \mathbb{P}_{Z \sim \mathcal{N}(0,2)}[Z \ge \frac{2}{\sqrt{\pi}} + \eta \sqrt{2}] + O(n^{1-c}) \ge p^2 n \ge n/(\log n)^{1/8}$$

for all $x \in \{0,1\}^3$. Thus we are in position to apply Propositions B.4 and B.7 (as this guarantees the condition on the parameter h).

We need to check that the values β are indeed of size $O((\log n)^2)$. This follows from the results of Section 3.3; recall that we showed the ρ_{ij} corresponding to each v was bounded by $(\log n)^{25}/\sqrt{p(1-p)}$ with super-polynomially high probability, and otherwise there was an exact formula which guarantees the appropriate boundedness with super-polynomially high probability.

Now Propositions B.4 and B.7 show that with super-polynomially high probability, for all $v \in R_0 \cup B_0$ and $x \in \{0, 1\}^3$ we have

$$\deg_{V_x} v = p|V_x| + O(n^{1/2}(\log n)^{25}).$$

Therefore

$$\deg_{R_2} v - \deg_{B_2} v = p(|R_2| - |B_2|) + O(n^{1/2} (\log n)^{25}) \ge n^{1-c/3}$$

for all $v \in R_0 \cup B_0$ with high probability. Thus every vertex will be on the red side after day three, as desired.

We have shown that with high probability (namely, as long as **E1** to **E4** hold and $|\eta| > n^{-c/2}$, and over the randomness of certain degree revelations over three days), some color has the lead after the first day and it wins in three days. This probability is in fact polynomially good.

Finally, Theorem 2.1 tells us the probability that $\eta > 0$ to a high degree of accuracy, and simple computation with normal distributions shows it is

$$\mathbb{P}_{Z \sim \mathcal{N}(0,1)} \left[Z \le \frac{p\Delta\sqrt{2}}{\sqrt{\pi p(1-p)}} \right] + O(n^{-c})$$

if c > 0 is a small enough absolute constant. We are done.

References

- Edward A. Bender, Central and local limit theorems applied to asymptotic enumeration, J. Combinatorial Theory Ser. A 15 (1973), 91–111. 4, 5
- [2] Itai Benjamini, Siu-On Chan, Ryan O'Donnell, Omer Tamuz, and Li-Yang Tan, Convergence, unanimity and disagreement in majority dynamics on unimodular graphs and random graphs, Stochastic Processes and their Applications 126 (2016), 2719–2733. 1, 3, 16
- [3] Ross Berkowitz and Pat Devlin, Central limit theorem for majority dynamics: Bribing three voters suffices, arXiv:2010.08172. 1, 2, 3, 16
- [4] E. Bolthausen, An estimate of the remainder in a combinatorial central limit theorem, Z. Wahrsch. Verw. Gebiete 66 (1984), 379–386. 21
- [5] E. R. Canfield, C. Greenhill, and B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with specified line sums, J. Combin. Theory Ser. A 115 (2008), 32–66. 2, 20, 25, 26
- [6] E. Rodney Canfield, Application of the Berry-Esséen inequality to combinatorial estimates, J. Combin. Theory Ser. A 28 (1980), 17–25. 11
- [7] Dorwin Cartwright and Frank Harary, Structural balance: a generalization of Heider's theory., Psychological Review 63 (1956), 277. 1
- [8] Debsoumya Chakraborti, Jeong Han Kim, Joonkyung Lee, and Tuan Tran, Majority dynamics on sparse random graphs, arXiv:2105.12709.
- [9] Asaf Ferber, Matthew Kwan, Bhargav Narayanan, Ashwin Sah, and Mehtaab Sawhney, Friendly bisections of random graphs, to appear. 3, 12, 13, 14, 20, 26
- [10] Nikolaos Fountoulakis, Mihyun Kang, and Tamás Makai, Resolution of a conjecture on majority dynamics: Rapid stabilization in dense random graphs, Random Structures & Algorithms 57 (2020), 1134–1156. 1, 2

- [11] Diana Halikias, Bo'az Klartag, and Boaz A Slomka, Discrete variants of brunn-minkowski type inequalities, arXiv:1911.04392.
- [12] S. Janson, T. Łuczak, and A. Rucinski, Random graphs, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. 22
- Bo'az Klartag and Joseph Lehec, Poisson processes and a log-concave Bernstein theorem, Studia Math. 247 (2019), 85–107. 5
- [14] M. Kwan, B. Sudakov, and T. Tran, Anticoncentration for subgraph statistics, J. Lond. Math. Soc. 99 (2019), 757–777. 20
- [15] A. Liebenau and N. Wormald, Asymptotic enumeration of digraphs and bipartite graphs by degree sequence, arXiv:2006.15797. 2
- [16] A. Liebenau and N. Wormald, Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph, arXiv:1702.08373. 2
- [17] Warren S McCulloch and Walter Pitts, A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics 5 (1943), 115–133. 1
- [18] Brendan D. McKay and Fiona Skerman, Degree sequences of random digraphs and bipartite graphs, J. Comb. 7 (2016), 21–49. 2, 6, 7
- [19] Brendan D. McKay and Nicholas C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, European J. Combin. 11 (1990), 565–580. 2, 3, 6, 20, 26
- [20] Brendan D. McKay and Nicholas C. Wormald, The degree sequence of a random graph. I. The models, Random Structures Algorithms 11 (1997), 97–117. 2, 6
- [21] Elchanan Mossel, Joe Neeman, and Omer Tamuz, Majority dynamics and aggregation of information in social networks, Autonomous Agents and Multi-Agent Systems 28 (2014), 408–429. 1
- [22] Elchanan Mossel and Omer Tamuz, Opinion exchange dynamics, Probab. Surv. 14 (2017), 155–204. 1
- [23] Linh Tran and Van Vu, Reaching a consensus on random networks: The power of few, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020), Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. 1, 2, 3, 16
- [24] R. Vershynin, High-dimensional probability, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018, An introduction with applications in data science, With a foreword by Sara van de Geer. 20

APPENDIX A. MISCELLANEOUS CALCULATIONS

We now include a proof of Lemma 2.2.

) > D'

 $I \setminus 1$

Proof of Lemma 2.2. We have

m[**D** · (

$$\begin{aligned} \mathbb{P}[\mathrm{Bm}(n+\tau,q) &\geq \mathrm{Bm}(n,q)] \\ &= \sum_{i=0}^{n} \sum_{j=i}^{n+\tau} \binom{n+\tau}{j} \left(p + \frac{\alpha}{n} \right)^{j} \left(1 - p - \frac{\alpha}{n} \right)^{n+\tau-j} \binom{n}{i} \left(p + \frac{\beta}{n} \right)^{i} \left(1 - p - \frac{\beta}{n} \right)^{n-i} \\ &= \sum_{i=0}^{n} \sum_{j=i}^{n+\tau} \binom{n+\tau}{j} p^{j} (1-p)^{n+\tau-j} \binom{n}{i} p^{i} (1-p)^{n-i} \\ &\times \left(1 + \frac{\alpha}{pn} \right)^{j} \left(1 - \frac{\alpha}{(1-p)n} \right)^{n+\tau-j} \left(1 + \frac{\beta}{pn} \right)^{i} \left(1 - \frac{\beta}{(1-p)n} \right)^{n-i} \\ &= \sum_{i=0}^{n} \sum_{j=i}^{n+\tau} \binom{n+\tau}{j} p^{j} (1-p)^{n+\tau-j} \binom{n}{i} p^{i} (1-p)^{n-i} \\ &\times \left(1 + O\left(\frac{(\log n)^{2}}{n} \right) \right) \exp\left(\frac{j\alpha}{pn} - \frac{(n+\tau-j)\alpha}{(1-p)n} + \frac{i\beta}{pn} - \frac{(n-i)\beta}{(1-p)n} \right). \end{aligned}$$

The last line comes from the bounds on α, β, p . Note that terms where $|i-pn| \ge C\sqrt{p(1-p)n\log n}$ or $|j-pn| \ge C\sqrt{p(1-p)n\log n}$ can contribute at most O(1/n) to the total mass, if C is a large enough constant. Let $i = pn + x\sqrt{p(1-p)n}$ and $j = pn + y\sqrt{p(1-p)n}$ where x, y range over

appropriate values of magnitude at most $C\sqrt{\log n}$. The exponential term in the above can be expanded if i, j have the above bounded criterion, and otherwise the total contribution is negligible anyway. We thus find that the above equals

$$\begin{split} &\sum_{i=0}^{n} \sum_{j=i}^{n+\tau} \binom{n+\tau}{j} p^{j} (1-p)^{n+\tau-j} \binom{n}{i} p^{i} (1-p)^{n-i} + O(n^{-1} (\log n)^{2}) \\ &+ \sum_{i=0}^{n} \sum_{j=i}^{n+\tau} \left(\frac{j\alpha}{pn} - \frac{(n+\tau-j)\alpha}{(1-p)n} + \frac{i\beta}{pn} - \frac{(n-i)\beta}{(1-p)n} \right) \binom{n+\tau}{j} p^{j} (1-p)^{n+\tau-j} \binom{n}{i} p^{i} (1-p)^{n-i} \\ &= \mathbb{P}[\operatorname{Bin}(n+\tau,p) \geq \operatorname{Bin}(n,p)] + O(n^{-1} (\log n)^{2}) \\ &+ \sum_{x} \sum_{y \geq x} \left(\frac{y\alpha + x\beta}{\sqrt{p(1-p)n}} \right) \binom{n+\tau}{j} p^{j} (1-p)^{n+\tau-j} \binom{n}{i} p^{i} (1-p)^{n-i} \\ &= \mathbb{P}[\operatorname{Bin}(n+\tau,p) \geq \operatorname{Bin}(n,p)] + \frac{1}{\sqrt{p(1-p)n}} \int_{u=-\infty}^{\infty} \int_{v=u}^{\infty} (v\alpha + u\beta) \frac{1}{2\pi} \exp(-(u^{2}+v^{2})/2) dudv \\ &+ O((\log n)^{2}n^{-1}) \\ &= \mathbb{P}[\operatorname{Bin}(n+\tau,p) \geq \operatorname{Bin}(n,p)] + \frac{\mathbb{E}_{Z_{1},Z_{2} \sim \mathcal{N}(0,1)} [\alpha Z_{1} + \beta Z_{2} | Z_{1} \geq Z_{2}]}{2\sqrt{p(1-p)n}} + O((\log n)^{2}n^{-1}) \\ &= \mathbb{P}[\operatorname{Bin}(n+\tau,p) \geq \operatorname{Bin}(n,p)] + \frac{\alpha - \beta}{2\sqrt{\pi p(1-p)n}} + O((\log n)^{2}n^{-1}) \end{split}$$

In the second line we dropped the τ term as $(\tau \alpha)/((1-p)n)$ is negligible, and in the third line we used the multidimensional Berry-Esseen theorem. For the final line, we used

$$\mathbb{E}_{Z_1, Z_2 \sim \mathcal{N}(0,1)}[Z_1 - Z_2 | Z_1 \ge Z_2] = \mathbb{E}_{G \sim \mathcal{N}(0,2)}|G| = \frac{2}{\sqrt{\pi}}$$

and $\mathbb{E}[Z_1 + Z_2 | Z_1 \ge Z_2] = \mathbb{E}[Z_1 + Z_2] = 0.$

Now let $m = \min(n + \tau, n)$. We compute by a local central limit theorem for $|k| \leq |\tau|$ that for two independent binomial samples,

$$\mathbb{P}_{X,X'\sim\operatorname{Bin}(m,p)}[X' = X + k] = \mathbb{P}_{X,X'}[X' - X = k]$$

= $(1 + O(n^{-1/4}))\frac{1}{\sqrt{4\pi p(1-p)m}}\exp\left(-\frac{k^2}{4p(1-p)m}\right)$
= $\frac{1}{\sqrt{4\pi p(1-p)n}} + O(n^{-3/4}).$

Thus if $0 < k \le |\tau|$ we have by symmetry that

$$2\mathbb{P}_{X,X'\sim\operatorname{Bin}(m,p)}[X' \ge X+k] + (2k-1)\left(\frac{1}{\sqrt{4\pi p(1-p)n}} + O(n^{-3/4})\right) = 1$$

hence

$$\mathbb{P}_{X,X'\sim\operatorname{Bin}(m,p)}[X' \ge X+k] = \frac{1}{2} - \frac{2k-1}{4\sqrt{\pi p(1-p)n}} + O(n^{-3/4})$$

for all $0 < k \le |\tau|$. In fact this holds for $|k| \le |\tau|$ by a similar argument. Therefore if $\tau \ge 0$ we have

$$\mathbb{P}[\operatorname{Bin}(n+\tau,p) \ge \operatorname{Bin}(n,p)] = \mathbb{E}_{k\sim\operatorname{Bin}(\tau,p)}[\mathbb{P}_{X,X'\sim\operatorname{Bin}(m,p)}[X' \ge X-k]]$$
$$= \mathbb{E}_{k\sim\operatorname{Bin}(\tau,p)}\left[\frac{1}{2} + \frac{2k+1}{4\sqrt{\pi p(1-p)n}} + O(n^{-3/4})\right]$$

$$= \frac{1}{2} + \frac{2p\tau + 1}{4\sqrt{\pi p(1-p)n}} + O(n^{-3/4}).$$

A similar argument shows the same formula for $\tau < 0$.

Putting it all together, we see

$$\mathbb{P}[\operatorname{Bin}(n+\tau,q) \ge \operatorname{Bin}(n,q')] = \frac{1}{2} + \frac{p\tau + 1/2 + \alpha - \beta}{2\sqrt{\pi p(1-p)n}} + O(n^{-3/4}).$$

APPENDIX B. PROBABILITIES IN DEGREE-CONSTRAINED MODELS

We will first require slight modifications of [9, Propositions A.1, A.6].

B.1. Graph model. First we compute the probability of having certain neighborhood sizes in a degree-constrained model of graphs. This result follows from a delicate but straightforward argument that utilizes graph enumeration results from [19] (the bipartite model, discussed later, utilizes bipartite graph enumeration results from [5]). We state those results precisely in Appendix B.3.

Proposition B.1. There are $\epsilon_{B.1}, C_{B.1} > 0$ so the following holds. Let $n \ge C_{B.1}$ and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Let $\mathbf{d} \in E_n$ such that each $d_w = pn + O(\sqrt{p(1-p)}n^{1/2+\epsilon}B.1)$. Let G be a uniformly random graph on vertex set $W = \{v_1, \ldots, v_n\}$ with this degree sequence. Consider a size h subset $V \subseteq W$ satisfying $\min(h, n - h) \ge n/(\log n)^{1/8}$, and an integer $t \in [0, d_n]$. For $w \in W$, define β_w by $d_w = p(n-1) + \beta_w \sqrt{p(1-p)(n-1)}$. Then

$$\mathbb{P}[\deg_V(v_n) = t]$$

$$= (1 + O(n^{-1/6})) \exp\left(\frac{(\sum_{i=1}^{n} \beta_i)(\sum_{i=1}^{n} \beta_i - 2n\beta_n)}{2n^2}\right) \frac{\binom{(h-\mathbb{1}_V(v_n)}{t}\binom{(n-h-\mathbb{1}_{V^c}(v_n))}{d_n-t}}{\binom{(n-1)}{d_n-t}} \times \\ \mathbb{E}_{\substack{S_1 \sim \binom{V \setminus v_n}{t} \\ S_2 \sim \binom{V^c \setminus v_n}{d_n-t}}_{S=S_1 \cup S_2}} \exp\left(-\sqrt{\frac{p}{1-p}} \sum_{i \in W \setminus v_n} \left(-\frac{1-p}{p}\right)^{\mathbb{1}_S(i)} \frac{\beta_i}{\sqrt{n-1}} - \frac{1}{2} \sum_{i \in W \setminus v_n} \left(\frac{1-p}{p}\right)^{2\mathbb{1}_S(i)-1} \frac{\beta_i^2}{n-1}\right).$$

Here S_1, S_2 are uniform over their respective domains.

We will defer the proof of this to Appendices B.3 and B.4. We now turn to various consequences of this formula. To proceed, we will need to understand expressions as appearing in the right side of Proposition B.1. To this end, we state the following general results about sums of random variables constrained to live on a slice.

Lemma B.2. Let $a_1, \ldots, a_n \in \mathbb{R}$ and let $X = \sum_{i=1}^n a_i \xi_i$, where $\xi = (\xi_1, \ldots, \xi_n)$ is uniform on the subset of $\{0,1\}^n$ with sum s. Furthermore assume that $\eta^2 = \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2/n$. We have

$$\mathbb{P}[|X - \mathbb{E}X| \ge t] \le 2\exp(-t^2/(4\eta^2))$$

and

 $\mathbb{E}e^{X-\mathbb{E}X} \le 2e^{O(\eta^2)}.$

Proof. The first part is by Azuma-Hoeffding (see [14, Lemma 2.2] for a detailed proof). The second part follows from integrating the first (see [24, Proposition 2.5.2]). \Box

Lemma B.3. Let $a_1, \ldots, a_n \in \mathbb{R}$ and let $X = \sum_{i=1}^n a_i \xi_i$, where $\xi = (\xi_1, \ldots, \xi_n)$ is uniform on the subset of $\{0,1\}^n$ with sum s such that $\min(s, n-s) \ge n(\log n)^{-2}$. Furthermore assume that $|a_i| \le n^{-1/2} (\log n)^2$ and $\eta^2 = \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2 / n \le \sqrt{\log n}$. We have

$$\mathbb{E}e^X = \exp\left(\mathbb{E}X + \frac{1}{2}\operatorname{Var}X + O(n^{-1/9})\right).$$

Proof. Let $\mu = \mathbb{E}X$ and $\sigma^2 = \text{Var }X$. Clearly

$$\sigma^{2} = \sum_{i \neq j} a_{i}a_{j} (\mathbb{E}[\xi_{i}\xi_{j}] - \mathbb{E}\xi_{i}\mathbb{E}\xi_{j}) + \sum_{i} a_{i}^{2} (\mathbb{E}\xi_{i}^{2} - (\mathbb{E}\xi_{i})^{2})$$
$$= \sum_{i \neq j} a_{i}a_{j} \left(\frac{s(s-1)}{n(n-1)} - \frac{s^{2}}{n^{2}}\right) + \sum_{i} a_{i}^{2} \left(\frac{s}{n} - \frac{s^{2}}{n^{2}}\right) = \frac{s(n-s)}{n(n-1)}\eta^{2}.$$

First by Lemma B.2 we have

$$\mathbb{P}[|X - \mathbb{E}X| \ge t] \le 2\exp(-t^2/(4\eta^2))$$

for all $t \ge 0$. Now

$$\mathbb{E}e^{X-\mathbb{E}X} = \int_{-\infty}^{\infty} e^t \mathbb{P}[X - \mathbb{E}X \ge t] dt = \int_{-\infty}^{8\eta\sqrt{\log n}} e^t \mathbb{P}[X - \mathbb{E}X \ge t] dt + O\left(\int_{4\eta\sqrt{\log n}}^{\infty} e^{t-t^2/(4\eta^2)} dt\right)$$
$$= \int_{-\infty}^{8\eta\sqrt{\log n}} e^t \mathbb{P}[X - \mathbb{E}X \ge t] dt + O\left(\int_{8\eta\sqrt{\log n}}^{\infty} e^{-t^2/(8\eta^2)} dt\right)$$
$$= \int_{-\infty}^{8\eta\sqrt{\log n}} e^t \mathbb{P}[X - \mathbb{E}X \ge t] dt + O(n^{-4}).$$

If $\sigma \leq n^{-1/8}$, then η is similarly bounded and we obtain an upper bound of the form $1 + O(n^{-1/9})$. Combining with $\mathbb{E}e^X \geq e^{\mathbb{E}X}$, the result follows. The result follows. If $\sigma > n^{-1/8}$ then a combinatorial central limit theorem of Bolthausen [4] shows

$$d_{K}(X - \mathbb{E}X, \mathcal{N}(0, \sigma^{2})) = O\left(\sum_{i=1}^{n} |a_{i}|^{3} / \sigma^{3}\right) = O(n^{-2/17}).$$

This allows us the replace the integrand above with the CDF of a Gaussian, and we easily derive

$$\mathbb{E}e^{X-\mathbb{E}X} = e^{\frac{\sigma^2}{2}} + O(n^{-2/17}) \cdot O(e^{\eta\sqrt{\log n}}) = \exp(\sigma^2/2 + O(n^{-1/9})).$$

We now use this information to explicitly compute the formula in Proposition B.1 under some slight additional hypotheses.

Proposition B.4. Assume the hypotheses of Proposition B.1. Assume additionally that $\beta_w = O((\log n)^2)$ for all $w \in W$. Then if $|t - ph| \ge n^{1/2} (\log n)^{25}$ we have

$$\mathbb{P}[\deg_V(v_n) = t] \le \exp(-\Omega((t - ph)^2/n)).$$

If $|t-ph| \le n^{3/5}$ and furthermore $(\sum_{i=1}^n \beta_i^2)/n \le (\log n)^{1/9}$ then we have

$$\begin{split} \mathbb{P}[\deg_{V}(v_{n}) &= t] \\ &= (1 + O(n^{-1/10})) \frac{\binom{h}{t} \binom{n-h-1}{d_{n}-t}}{\binom{n-1}{d_{n}}} \exp\left[\frac{(\sum_{i=1}^{n} \beta_{i})(\sum_{i=1}^{n} \beta_{i} - 2n\beta_{n})}{2n^{2}} \\ &- \sqrt{\frac{p}{1-p}} \bigg(\sum_{i \in V \setminus v_{n}} \left(1 - \frac{t}{ph}\right) \frac{\beta_{i}}{\sqrt{n-1}} + \sum_{i \in V^{c} \setminus v_{n}} \left(1 - \frac{d_{n}-t}{p(n-h)}\right) \frac{\beta_{i}}{\sqrt{n-1}}\bigg) - \frac{1}{2} \sum_{i \in W \setminus v_{n}} \frac{\beta_{i}^{2}}{n-1} \\ &+ \frac{1}{2nh} \sum_{i < j \in V \setminus v_{n}} (\beta_{i} - \beta_{j})^{2} + \frac{1}{2n(n-h)} \sum_{i < j \in V^{c} \setminus v_{n}} (\beta_{i} - \beta_{j})^{2} \bigg]. \end{split}$$

Proof. We apply Proposition B.1. Let

$$X = -\sqrt{\frac{p}{1-p}} \sum_{i \in W \setminus v_n} \left(-\frac{1-p}{p} \right)^{\mathbb{1}_S(i)} \frac{\beta_i}{\sqrt{n-1}} - \frac{1}{2} \sum_{i \in W \setminus v_n} \left(\frac{1-p}{p} \right)^{2\mathbb{1}_S(i)-1} \frac{\beta_i^2}{n-1},$$

where $S_1 \sim {\binom{V \setminus v_n}{t}}$ and $S_2 \sim {\binom{V^c \setminus v_n}{d_n - t}}$. The point will be that in typical cases X is a random variable with sub-Gaussian tails, and that it is converging to a Gaussian, which are together enough to compute its exponential moment. When t is far from ph, we will instead obtain a tail bound. We have

 $\mathbb{E}X$

$$\begin{split} &= -\sqrt{\frac{p}{1-p}} \sum_{i \in V \setminus v_n} \left(\frac{t(-(1-p))}{hp} + \frac{h-t}{h} \right) \frac{\beta_i}{\sqrt{n-1}} + O(n^{-1/3}) \\ &- \sqrt{\frac{p}{1-p}} \sum_{i \in V \setminus v_n} \left(\frac{(d_n - t)(-(1-p))}{(n-h)p} + \frac{(n-h) - (d_n - t)}{n-h} \right) \frac{\beta_i}{\sqrt{n-1}} + O(n^{-1/3}) \\ &- \frac{1}{2} \sum_{i \in V \setminus v_n} \left(\frac{t(1-p)}{hp} + \frac{(h-t)p}{h(1-p)} \right) \frac{\beta_i^2}{n-1} + O(n^{-1/3}) \\ &- \frac{1}{2} \sum_{i \in V \setminus v_n} \left(\frac{(d_n - t)(1-p)}{(n-h)p} + \frac{((n-h) - (d_v - t))p}{(n-h)(1-p)} \right) \frac{\beta_i^2}{n-1} + O(n^{-1/3}) \\ &= -\sqrt{\frac{p}{1-p}} \left(\sum_{i \in V \setminus v_n} \left(1 - \frac{t}{ph} \right) \frac{\beta_i}{\sqrt{n-1}} + \sum_{i \in V \cap v_n} \left(1 - \frac{d_n - t}{p(n-h)} \right) \frac{\beta_i}{\sqrt{n-1}} \right) \\ &- \frac{1}{2} \sum_{i \in W \setminus v_n} \frac{\beta_i^2}{n-1} + O(n^{-1/3} + |t-ph|(\log n)^2/n). \end{split}$$

The initial additive error terms $O(n^{-1/3})$ come from the fact that $v_n \in V$ or $v_n \in V^c$ slightly change the fractions listed above, but not by much.

At this point, if $|t - ph| \ge n^{3/5}$, we have

$$\frac{\binom{(h-\mathbb{1}_{V}(v_{n})}{t}\binom{(n-h-\mathbb{1}_{V^{c}}(v_{n})}{d_{n}-t}}{\binom{(n-1)}{d_{n}}} \le \exp(-\Omega((t-ph)^{2}/n))$$

by tail bounds for the hypergeometric distribution (see e.g. [12, Theorem 2.10]). The initial exponential term is bounded by $\exp(O((\log n)^4))$, and we are left with $\mathbb{E}\exp(X)$. Now Lemma B.2 demonstrates $\mathbb{E}\exp(X) \leq \exp(\mathbb{E}X + O((\log n)^4))$ since the coefficient variance in X is $O((\log n)^4/n)$ by the given conditions. But the above demonstrates

$$|\mathbb{E}X| = O\left(\frac{|t-ph|}{\sqrt{n}}(\log n)^2\right).$$

This immediately gives a bound of the claimed quality.

From now on we assume $|t - ph| \le n^{3/5}$. Note that the error term computed on $\mathbb{E}X$ is now of quality $O(n^{-1/3})$ uniformly. We next compute the variance of X. It is straightforward to see that X and

$$X' = -\sqrt{\frac{p}{1-p}} \sum_{i \in W \setminus v_n} \left(-\frac{1-p}{p} \right)^{\mathbb{I}_S(i)} \frac{\beta_i}{\sqrt{n-1}}$$
22

have $|\operatorname{Var} X - \operatorname{Var} X'| = O(n^{-1/4})$. From the proof of Lemma B.3, we see

$$\operatorname{Var} X = \frac{1}{p(1-p)(n-1)} \left(\frac{t(h-t)}{h(h-1)} \frac{\sum_{i < j \in V \setminus v_n} (\beta_i - \beta_j)^2}{h} + \frac{(d_n - t)((n-h) - (d_n - t))}{(n-h)(n-h-1)} \frac{\sum_{i < j \in V^c \setminus v_n} (\beta_i - \beta_j)^2}{n-h} \right) + O(n^{-1/4}),$$

where we again use that the fraction $t/|V \setminus v_n|$ is close to t/h regardless of if $v_n \in V$. Using $t = ph + O(n^{3/5})$ and $d_n = pn + O(\sqrt{p(1-p)n}(\log n)^2)$, we find

$$\operatorname{Var} X = \frac{1}{nh} \sum_{i < j \in V \setminus v_n} (\beta_i - \beta_j)^2 + \frac{1}{n(n-h)} \sum_{i < j \in V^c \setminus v_n} (\beta_i - \beta_j)^2 + O(n^{-1/4}).$$

Note that $\operatorname{Var} X \leq \sum_{i=1}^{n} \beta_i^2 / \min(h, n-h) = O(n(\log n)^{1/9} / \min(h, n-h))$. Now apply Lemma B.3 to the two slices defining X. Note that the condition $\eta^2 \leq \sqrt{\log n}$ follows from the inequalities $(n/h)(p(1-p))^{-1}(\log n)^{1/9} < \sqrt{\log n}$ and the relation between σ^2, η^2 in the proof of Lemma B.3. Therefore

$$\mathbb{E}e^{X} = \exp\left(\mathbb{E}X + \frac{1}{2}\operatorname{Var}X + O(n^{-1/10})\right).$$

Finally, using Proposition B.1, we obtain

$$\begin{split} \mathbb{P}[\deg_{V}(v_{n}) &= t] \\ &= (1 + O(n^{-1/10})) \frac{\binom{h}{t} \binom{n-h-1}{d_{n}-t}}{\binom{n-1}{d_{n}}} \exp\left[\frac{(\sum_{i=1}^{n} \beta_{i}) (\sum_{i=1}^{n} \beta_{i} - 2n\beta_{n})}{2n^{2}} \\ &- \sqrt{\frac{p}{1-p}} \bigg(\sum_{i \in V \setminus v_{n}} \left(1 - \frac{t}{ph}\right) \frac{\beta_{i}}{\sqrt{n-1}} + \sum_{i \in V^{c} \setminus v_{n}} \left(1 - \frac{d_{n}-t}{p(n-h)}\right) \frac{\beta_{i}}{\sqrt{n-1}} \bigg) - \frac{1}{2} \sum_{i \in W \setminus v_{n}} \frac{\beta_{i}^{2}}{n-1} \\ &+ \frac{1}{2nh} \sum_{i < j \in V \setminus v_{n}} (\beta_{i} - \beta_{j})^{2} + \frac{1}{2n(n-h)} \sum_{i < j \in V^{c} \setminus v_{n}} (\beta_{i} - \beta_{j})^{2} \bigg]. \end{split}$$

We used that the product of binomials changes by a small factor upon swapping between $v \in V$ and $v \in V^c$.

Finally we note a massive simplification of this formula in the case when h is near n/2 and the total number of edges is close to $p\binom{n}{2}$.

Proposition B.5. Assume the hypotheses of the second part of Proposition B.4. Assume additionally that $|h - n/2| = O(\sqrt{n \log n})$ and $\sum_{i \in W} \beta_i = O(n^{5/6})$. Then for $|\gamma| \leq n^{1/10}$ with $ph + \gamma \sqrt{p(1-p)n} \in \mathbb{Z}$, we have

$$\mathbb{P}[\deg_V(v_n) = ph + \gamma \sqrt{p(1-p)n}] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \beta_n - \frac{\sum_{i \in V} \beta_i}{n/2}\right)^2\right).$$

Proof. Apply Proposition B.4 to $t = ph + \gamma \sqrt{p(1-p)n}$. First, since $\sum_{i=1}^{n} \beta_i = O(n^{3/4})$ and $|\beta_i| = O((\log n)^2)$, we see that the initial exponential term is small. Next, we have from $d_n = p(n-1) + \beta_n \sqrt{p(1-p)(n-1)}$ and $|h-n/2| = O(\sqrt{n \log n})$ that

$$-\sqrt{\frac{p}{1-p}} \left(\sum_{i \in V \setminus v_n} \left(1 - \frac{t}{ph}\right) \frac{\beta_i}{\sqrt{n-1}} + \sum_{i \in V^c \setminus v_n} \left(1 - \frac{d_n - t}{p(n-h)}\right) \frac{\beta_i}{\sqrt{n-1}}\right)$$
23

$$= \sum_{i \in V \setminus v_n} \frac{\gamma \beta_i}{n/2} + \sum_{i \in V^c \setminus v_n} \frac{(\beta_n - \gamma)\beta_i}{n/2} + O(n^{-1/3})$$
$$= \frac{2\gamma - \beta_n}{n/2} \sum_{i \in V \setminus v_n} \beta_i + O(n^{-1/8})$$

Similarly, in the last three terms of the formula in Proposition B.4, we can replace h by n/2 and n-1 by n in return for a negligible additive error. Therefore the terms in the exponential add up to

$$\begin{aligned} &\frac{2\gamma - \beta_n}{n/2} \sum_{i \in V \setminus v_n} \beta_i - \frac{1}{2} \sum_{i \in W \setminus v_n} \frac{\beta_i^2}{n} + \frac{1}{n^2} \sum_{i < j \in V \setminus v_n} (\beta_i - \beta_j)^2 + \frac{1}{n^2} \sum_{i < j \in V^c \setminus v_n} (\beta_i - \beta_j)^2 + O(n^{-1/8}) \\ &= \frac{2\gamma - \beta_n}{n/2} \sum_{i \in V \setminus v_n} \beta_i + \frac{1}{n^2} \sum_{i < j \in V \setminus v_n} (-2\beta_i\beta_j) + \frac{1}{n^2} \sum_{i < j \in V^c \setminus v_n} (-2\beta_i\beta_j) + O(n^{-1/8}) \\ &= \frac{2\gamma - \beta_n}{n/2} \sum_{i \in V \setminus v_n} \beta_i - \frac{1}{n^2} \left(\sum_{i \in V \setminus v_n} \beta_i \right)^2 - \frac{1}{n^2} \left(\sum_{i \in V^c \setminus v_n} \beta_i \right)^2 + O(n^{-1/8}) \\ &= \frac{2\gamma - \beta_n}{n/2} \sum_{i \in V \setminus v_n} \beta_i - \frac{2}{n^2} \left(\sum_{i \in V \setminus v_n} \beta_i \right)^2 + O(n^{-1/8}). \end{aligned}$$

Furthermore, the ratio of binomial coefficients can be computed as follows. If $|a - pb| = O(b^{3/5})$ then by Stirling's formula,

$$\begin{split} p^{a}(1-p)^{b-a} \begin{pmatrix} b \\ a \end{pmatrix} \\ &= (1+O(b^{-1/2})) \frac{1}{\sqrt{2\pi p(1-p)b}} \left(\frac{a}{pb}\right)^{-a} \left(\frac{b-a}{(1-p)b}\right)^{-(b-a)} \\ &= (1+O(b^{-1/2})) \frac{1}{\sqrt{2\pi p(1-p)b}} \left(1+\frac{a-pb}{pb}\right)^{-a} \left(1-\frac{a-pb}{(1-p)b}\right)^{-(b-a)} \\ &= \frac{1+O(b^{-1/6})}{\sqrt{2\pi p(1-p)b}} \exp\left(-a\frac{a-pb}{pb}+\frac{1}{2}a\left(\frac{a-pb}{pb}\right)^{2}+(b-a)\frac{a-pb}{(1-p)b}+\frac{1}{2}(b-a)\left(\frac{a-pb}{(1-p)b}\right)^{2}\right) \\ &= \frac{1+O(b^{-1/6})}{\sqrt{2\pi p(1-p)b}} \exp\left(-a\frac{a-pb}{pb}+\frac{1}{2}pb\left(\frac{a-pb}{pb}\right)^{2}+(b-a)\frac{a-pb}{(1-p)b}+\frac{1}{2}(b-pb)\left(\frac{a-pb}{(1-p)b}\right)^{2}\right) \\ &= \frac{1+O(b^{-1/6})}{\sqrt{2\pi p(1-p)b}} \exp\left(-\frac{(a-pb)^{2}}{2p(1-p)b}\right), \end{split}$$

so that a local central limit theorem holds. This allows us to compute

$$\frac{\binom{h}{d_n-t}\binom{n-h-1}{d_n-t}}{\binom{n-1}{d_n}} = \frac{p^t(1-p)^{h-t}\binom{h}{t}p^{d_n-t}(1-p)^{(n-h-1)-(d_n-t)}\binom{n-h-1}{d_n-t}}{p^{d_n}(1-p)^{(n-1)-d_n}\binom{n-1}{d_n}}$$
$$= \frac{\sqrt{2}+O(n^{-1/8})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{(t-ph)^2}{2p(1-p)h} - \frac{((d_n-t)-p(n-h-1))^2}{2p(1-p)(n-h-1)} + \frac{(d_n-p(n-1))^2}{2p(1-p)(n-1)}\right)$$
$$= \frac{\sqrt{2}+O(n^{-1/8})}{\sqrt{\pi p(1-p)n}} \exp\left(-\gamma^2 - (\beta_n - \gamma)^2 + \frac{\beta_n^2}{2}\right)$$

$$=\frac{\sqrt{2}+O(n^{-1/8})}{\sqrt{\pi p(1-p)n}}\exp\bigg(-\frac{(2\gamma-\beta_n)^2}{2}\bigg).$$

Putting it all together in Proposition B.4 we obtain the result, noting that β_n is small so a difference of $\beta_n/(n/2)$ is negligible in the final formula.

B.2. **Bigraph model.** Now we compute the probability of having certain neighborhood sizes in a degree-constrained model of bipartite graphs. This time we use [5] to derive the necessary initial probability computation.

Proposition B.6. There are $\epsilon_{B.6}$, $C_{B.6} > 0$ so the following holds. Let $n \ge C_{B.6}$ and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. Suppose $(\log n)^{-1/4} \le m/n \le (\log n)^{1/4}$. Let $(\mathbf{s}, \mathbf{t}) \in E_{m,n}$ (so \mathbf{s} has length m) such that each $s_w = pn + O(\sqrt{p(1-p)n}^{1/2+\epsilon}B.6)$ and each $t_w = pm + O(\sqrt{p(1-p)n}^{1/2+\epsilon}B.6)$. Let G be a uniformly random bigraph on vertex sets $W' = \{v'_1, \ldots, v'_m\}$ and $W = \{v_1, \ldots, v_n\}$ with these degree sequences between the parts. Consider a size h subset $V \subseteq W$ satisfying $\min(h, n-h) \ge n/(\log n)^{1/8}$, and an integer $t \in [0, s_m]$. For $w \in W$, define β_w by $t_w = pm + \beta_w \sqrt{p(1-p)m}$. Let $s_m = pn + \alpha \sqrt{p(1-p)n}$. Then

$$\begin{split} \mathbb{P}[\deg_{V}(v'_{m}) &= t] \\ &= (1 + O(n^{-1/8})) \exp\left(\frac{(\sum_{i=1}^{n} \beta_{i})(\sum_{i=1}^{n} \beta_{i} - 2\sqrt{mn}\alpha)}{2mn}\right) \frac{\binom{h}{t}\binom{n-h}{s_{m}-t}}{\binom{n}{s_{m}}} \times \\ \mathbb{E}_{\substack{S_{1} \sim \binom{V}{t} \\ S_{2} \sim \binom{W \setminus V}{s_{m}-t} \\ S = S_{1} \cup S_{2}}} \exp\left(-\sqrt{\frac{p}{1-p}} \sum_{i \in W} \left(-\frac{1-p}{p}\right)^{\mathbb{1}_{S}(i)} \frac{\beta_{i}}{\sqrt{m}} - \frac{1}{2} \sum_{i \in W} \left(\frac{1-p}{p}\right)^{2\mathbb{1}_{S}(i)-1} \frac{\beta_{i}^{2}}{m}\right). \end{split}$$

Here S_1, S_2 are uniform over their respective domains.

As in Appendix B.1, there are various corollaries of this fact by computing out what the expectation term yields. The proofs are exactly analogous to the ones given before and consist of routine computation given those ideas. Therefore, we leave out the proofs and merely record the necessary results.

Proposition B.7. Assume the hypotheses of Proposition B.1. Assume additionally that $\beta_w = O((\log n)^2)$ for all $w \in W$ and $\alpha = O((\log n)^2)$. Then if $|t - ph| \ge n^{1/2} (\log n)^5$ we have

$$\mathbb{P}[\deg_V(v'_m) = t] \le \exp(-\Omega((t - ph)^2/n))$$

If $|t - ph| \le n^{3/5}$ and furthermore $(n/m) \cdot (\sum_{i=1}^n \beta_i^2/n) \le (\log n)^{1/9}$ then we have

$$\begin{split} \mathbb{P}[\deg_{V}(v'_{m}) &= t] \\ &= (1 + O(n^{-1/10})) \frac{\binom{h}{t}\binom{n-h}{s_{m}-t}}{\binom{n}{s_{m}}} \exp\left[\frac{(\sum_{i=1}^{n}\beta_{i})(\sum_{i=1}^{n}\beta_{i} - 2\sqrt{mn\alpha})}{2mn} \right. \\ &- \sqrt{\frac{p}{1-p}} \left(\sum_{i \in V} \left(1 - \frac{t}{ph}\right) \frac{\beta_{i}}{\sqrt{m}} + \sum_{i \in W \setminus V} \left(1 - \frac{s_{m}-t}{p(n-h)}\right) \frac{\beta_{i}}{\sqrt{m}}\right) - \frac{1}{2} \sum_{i \in W} \frac{\beta_{i}^{2}}{m} \\ &+ \frac{1}{2mh} \sum_{i < j \in V} (\beta_{i} - \beta_{j})^{2} + \frac{1}{2m(n-h)} \sum_{i < j \in V^{c}} (\beta_{i} - \beta_{j})^{2} \right]. \end{split}$$

Remark. Note that in a random bipartite graph with part sizes m and n the condition on $\sum_{i=1}^{n} \beta_i^2/n$ can only hold when m and n are within a small power of log n factor.

Proposition B.8. Assume the hypotheses of the second part of Proposition B.7. Assume additionally that $|h - n/2| = O(\sqrt{n \log n})$ and $\sum_{i=1}^{n} \beta_i = O(n^{5/6})$. Then for $|\gamma| \leq n^{1/10}$ with $ph + \gamma \sqrt{p(1-p)n} \in \mathbb{Z}$, we have

$$\mathbb{P}[\deg_V(v'_m) = ph + \gamma \sqrt{p(1-p)n}] = \frac{\sqrt{2} + O(n^{-1/10})}{\sqrt{\pi p(1-p)n}} \exp\left(-\frac{1}{2}\left(2\gamma - \alpha - \frac{\sum_{i \in V} \beta_i}{\sqrt{mn/2}}\right)^2\right).$$

B.3. Computational preliminaries. Now we turn to justifying Propositions B.1 and B.6. We first record the graph and bigraph enumeration results that will be used.

Theorem B.9 ([19]). There exists a fixed constant $\varepsilon > 0$ such that the following holds. Consider a degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ of length n such that each $|d_i - \overline{d}| \le n^{1/2+\varepsilon}$, where $\overline{d} = (1/n) \sum_{i=1}^n d_i$ satisfies $\overline{d} \ge n/\log n$. Letting $r = \overline{dn}/2 \in \mathbb{Z}$, $\mu = \overline{d}/(n-1)$, and $\gamma_2^2 = (1/(n-1)^2) \sum_{i=1}^n (d_i - \overline{d})^2$, the number of labelled graphs with degree sequence \mathbf{d} is

$$(1+O(n^{-1/4}))\exp\left(\frac{1}{4}-\frac{\gamma_2^2}{4\mu^2(1-\mu)^2}\right)\binom{n(n-1)/2}{r}\binom{n(n-1)}{2r}^{-1}\prod_{i=1}^n\binom{n-1}{d_i}.$$

Theorem B.10 ([5]). There exists a fixed constant $\varepsilon > 0$ such that the following holds. For a pair of integers $n, m \in \mathbb{N}$ with $n/(\log n)^{1/2} \leq m \leq n(\log n)^{1/2}$, fix a pair of degree sequences $\mathbf{s} = (s_1, \ldots, s_n), \mathbf{t} = (t_1, \ldots, t_m)$ such that each $|s_i - \overline{s}| \leq n^{1/2+\varepsilon}$ and $|t_i - \overline{t}| \leq m^{1/2+\varepsilon}$, where $\overline{s} = (1/n) \sum_{i=1}^n s_i$ and $\overline{t} = (1/m) \sum_{i=1}^m t_i$ satisfy $\overline{s} \geq n/(\log n)^{1/2}$ and $\overline{t} \geq m/(\log m)^{1/2}$. Let $\gamma_2(s)^2 = (1/n^2) \sum_{i=1}^n (s_i - \overline{s})^2$, $\gamma_2(t)^2 = (1/m^2) \sum_{i=1}^m (t_i - \overline{t})^2$ and $\mu = \sum_{i=1}^n s_i/(mn) = \sum_{i=1}^m t_i/(mn)$. Let $r = \mu mn$. Then the number of labelled bipartite graphs whose partition classes have degree sequences \mathbf{s} and \mathbf{t} is

$$(1+O(n^{-1/8}))\exp\left(-\frac{1}{2}\left(1-\frac{\gamma_2(s)^2}{\mu(1-\mu)}\right)\left(1-\frac{\gamma_2(t)^2}{\mu(1-\mu)}\right)\right)\binom{mn}{r}^{-1}\prod_{i=1}^n\binom{m}{s_i}\prod_{i=1}^m\binom{n}{t_i}.$$

We next compute a certain ratio of binomials that will show up when computing probabilities via graph enumeration. An analogous result for p = 1/2 was shown in [9, Lemma B.3].

Lemma B.11. Suppose that $n/(\log n)^{1/2} \le m \le n(\log n)^{1/2}$ and $(\log n)^{-1/4} \le p \le 1 - (\log n)^{-1/4}$. • If $r = pmn + \Delta_1$ and $d = pm + \Delta_2$ where $\Delta_1 = O(n^{8/5})$ and $\Delta_2 = O(n^{3/5})$, then

$$\frac{\binom{m(n-1)}{r-d}^{-1}}{\binom{mn}{r}^{-1}} p^d (1-p)^{m-d} = \exp\left(\frac{\Delta_1(\Delta_1 - 2n\Delta_2)}{2p(1-p)mn^2} + O(n^{-1/6})\right).$$
• If $r = p\binom{n}{2} + \Delta_1$ and $d = p(n-1) + \Delta_2$ where $\Delta_1 = O(n^{8/5})$ and $\Delta_2 = O(n^{3/5})$, then
$$\frac{\binom{(n-1)(n-2)/2}{r-d}\binom{(n-1)(n-2)}{2r-2d}^{-1}}{\binom{n(n-1)/2}{r}\binom{n(n-1)}{2r}^{-1}} p^d (1-p)^{n-1-d} = \exp\left(\frac{2\Delta_1(\Delta_1 - n\Delta_2)}{p(1-p)n^3} + O(n^{-1/6})\right).$$

Proof. For the first expression, we have

$$\frac{\binom{m(n-1)}{r-d}^{-1}}{\binom{mn}{r}^{-1}} p^d (1-p)^{m-d}$$

$$= (1+O(n^{-1/2})) \frac{\left(\frac{r-d}{m(n-1)}\right)^{r-d} \left(1-\frac{r-d}{m(n-1)}\right)^{m(n-1)-(r-d)}}{\left(\frac{r}{mn}\right)^r \left(1-\frac{r}{mn}\right)^{mn-r}} p^d (1-p)^{m-d}$$

$$= (1+O(n^{-1/2})) \left(\frac{n}{n-1}\right)^{m(n-1)} \left(1-\frac{d}{r}\right)^{r-d} \left(1-\frac{m-d}{mn-r}\right)^{m(n-1)-(r-d)} \left(\frac{pmn}{r}\right)^d \left(\frac{(1-p)mn}{mn-r}\right)^{m-d}$$

$$\begin{split} &= \exp\left(m - \frac{m}{2(n-1)} - d + \frac{d^2}{r} - \frac{d^2(r-d)}{2r^2} - (m-d) + \frac{(m-d)^2}{mn-r} - \frac{(m-d)^2(mn-r-m+d)}{2(mn-r)^2} \\ &+ \frac{d(pmn-r)}{r} - \frac{d(r-pmn)^2}{2r^2} + \frac{(m-d)(r-pmn)}{mn-r} - \frac{(m-d)(r-pmn)^2}{2(mn-r)^2} + O(n^{-1/6})\right) \\ &= \exp\left(-\frac{m}{2n} + \frac{(pm)^2}{pmn} - \frac{(pm)^2(pmn)}{2(pmn)^2} + \frac{(m-pm)^2}{mn-pmn} - \frac{(m-pm)^2(mn-pmn)}{2(mn-pmn)^2} \\ &+ \frac{d(pmn-r)}{r} - \frac{d(r-pmn)^2}{2(pmn)^2} + \frac{(m-d)(r-pmn)}{mn-r} - \frac{(m-d)(r-pmn)^2}{2(mn-pmn)^2} + O(n^{-1/6})\right) \\ &= \exp\left(\frac{m(r-dn)(r-pmn)}{r(mn-r)} - \frac{1}{2}\left(\frac{d}{p^2} + \frac{m-d}{(1-p)^2}\right)\left(\frac{r-pmn}{mn}\right)^2 + O(n^{-1/6})\right) \\ &= \exp\left(\frac{m(r-dn)(r-pmn)}{pmn(mn-pmn)} - \frac{m}{2p(1-p)}\left(\frac{r-pmn}{mn}\right)^2 + O(n^{-1/6})\right) \end{split}$$

We have used Stirling's formula and that $1 + x = \exp(x - x^2/2 + O(x^3))$ repeatedly. Now using the definition of Δ_1, Δ_2 finishes. For the second expression, write

$$\frac{\binom{(n-1)(n-2)/2}{r-d}\binom{(n-1)(n-2)}{2r-2d}^{-1}}{\binom{n(n-1)/2}{r}\binom{n(n-1)}{2r}^{-1}}p^d(1-p)^{n-1-d} = \frac{\frac{\binom{m_1(n_1-1)}{r_1-d_1}^{-1}}{\binom{m_1(n_1)}{r_1}^{-1}}p^{d_1}(1-p)^{m_1-d_1}}{\frac{\binom{m_2(n_2-1)}{r_1}^{-1}}{\binom{m_2(n_2-1)}{r_2}^{-1}}}p^{d_2}(1-p)^{m_2-d_2}$$

where $m_1 = 2m_2 = 2(n-1)$ and $n_1 = n_2 = n/2$, and $r_1 = 2r_2 = 2r$ and $d_1 = 2d_2 = 2d$. Now apply the first part twice.

B.4. Proof of Propositions B.1 and B.6. We first compute the graph version.

Proof of Proposition B.1. We have that our vertex is the last vertex v_n , corresponding to degree d_n . Given $S \subseteq W \setminus v_n$ of size d_n (which we abusively identify with a set of integers), let $\mathbf{d}_S = (d_1 - \mathbb{1}_{1 \in S}, \ldots, d_{n-1} - \mathbb{1}_{n-1 \in S})$. As in Theorem B.9, let

$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} d_i; \qquad \overline{d}_S = \frac{1}{n-1} \sum_{i=1}^{n-1} d_{S,i} = \frac{n}{n-1} \overline{d} - \frac{2d_n}{n-1}$$

$$r = \frac{\overline{d}n}{2}; \qquad r_S = \frac{\overline{d}_S(n-1)}{2} = r - d_n,$$

$$\mu = \frac{\overline{d}}{n-1}; \qquad \mu_S = \frac{\overline{d}_S}{n-2} = \frac{n}{n-2} \mu - \frac{2d_n}{(n-1)(n-2)},$$

$$\gamma_2^2 = \frac{1}{(n-1)^2} \sum_{i=1}^{n} (d_i - \overline{d})^2; \qquad \gamma_2^2(S) = \frac{1}{(n-2)^2} \sum_{i=1}^{n-1} (d_{S,i} - \overline{d}_S)^2.$$

Note that **d** and each \mathbf{d}_S clearly satisfies the conditions of Theorem B.9 due to our given hypotheses. Note that

$$\gamma_2^2(S) = \gamma_2^2 + O(n^{-1/4}), \mu_S = \mu + O(1/n)$$
27

due to the given hypotheses. Now define

$$A = \frac{\binom{(n-1)(n-2)/2}{r-d_n} \binom{(n-1)(n-2)}{2r-2d_n}^{-1}}{\binom{n(n-1)/2}{r} \binom{n(n-1)}{2r}^{-1}} p^{d_n} (1-p)^{n-1-d_n}$$

and recall $d_i = p(n-1) + \beta_i \sqrt{p(1-p)(n-1)}$. We have

$$r - p\binom{n}{2} = \frac{1}{2} \sum_{i=1}^{n} (d_i - p(n-1)) = \frac{\sqrt{p(1-p)(n-1)}}{2} \sum_{i=1}^{n} \beta_i.$$

By the given hypotheses and Lemma B.11 we therefore derive

$$A = \exp\left(\frac{(\sum_{i=1}^{n} \beta_i)(\sum_{i=1}^{n} \beta_i - 2n\beta_n)}{2n^2} + O(n^{-1/6})\right)$$

We therefore see from Theorem B.9 that

$$\begin{split} \mathbb{P}[N(v_n) &= S] \\ &= (1+O(n^{-1/4})) \frac{\exp\left(\frac{1}{4} - \frac{\gamma_2^2(S)}{4\mu_S^2(1-\mu_S)^2}\right) \left(\binom{(n-1)(n-2)/2}{r_S} \left(\binom{(n-1)(n-2)}{2r_S}\right)^{-1} \prod_{i=1}^{n-1} \left(\binom{n-2}{d_i-1_S(i)}\right)}{\exp\left(\frac{1}{4} - \frac{\gamma_2^2(S)}{4\mu_S^2(1-\mu_S)^2}\right) \left(\binom{(n-1)/2}{r_S} \left(\binom{(n-1)/2}{2r_S}\right)^{-1} \prod_{i=1}^{n-1} \left(\binom{n-2}{d_i-1_S(i)}\right)}{2r_S}} \\ &= (1+O(n^{-1/4})) \frac{\left(\binom{(n-1)(n-2)/2}{r_S} \left(\binom{(n-1)(n-2)}{2r_S}\right)^{-1} \prod_{i=1}^{n-1} \left(\binom{n-2}{d_i-1_S(i)}\right)}{\binom{(n(n-1)/2}{2r_S} \left(\binom{n-1}{2r_S}\right)^{-1} \prod_{i=1}^{n-1} \left(\binom{n-2}{d_i-1_S(i)}\right)}}{2r_S} \\ &= (1+O(n^{-1/4})) \frac{A}{\binom{n-1}{d_n}} p^{-d_n} (1-p)^{-(n-1-d_n)} \prod_{i\in S} \frac{d_i}{n-1} \prod_{i\notin S} \frac{n-1-d_i}{n-1} \\ &= (1+O(n^{-1/4})) \frac{A}{\binom{n-1}{d_n}} \prod_{i\in S} \left(1+\beta_i \sqrt{\frac{1-p}{p(n-1)}}\right) \prod_{i\notin S} \left(1-\beta_i \sqrt{\frac{p}{(1-p)(n-1)}}\right) \\ &= \frac{A}{\binom{n-1}{d_n}} \exp\left(-\sqrt{\frac{p}{1-p}} \sum_{i=1}^{n-1} \left(-\frac{1-p}{p}\right)^{1_S(i)} \frac{\beta_i}{\sqrt{n-1}} - \frac{1}{2} \sum_{i=1}^{n-1} \left(\frac{1-p}{p}\right)^{21_S(i)-1} \frac{\beta_i^2}{n-1} + O(n^{-1/4})\right) \end{split}$$

for each $S \in {\binom{[n-1]}{d_n}}$. Therefore

$$(1+O(n^{-1/4}))\frac{\binom{n-1}{d_n}}{A\binom{h-1}{t}\binom{v(v_n)}{d_n-t}\binom{n-1-1}{d_n-t}\mathbb{P}[\deg_V(v_n)=t]} \mathbb{P}[\deg_V(v_n)=t]$$

$$=\mathbb{E}_{\substack{S_1\sim\binom{V\setminus v_n}{d_n-t}\\S_2\sim\binom{V^{-1}v_n}{d_n-t}\\S=S_1\cup S_2}} \exp\left(-\sqrt{\frac{p}{1-p}}\sum_{i=1}^{n-1}\left(-\frac{1-p}{p}\right)^{\mathbb{I}_S(i)}\frac{\beta_i}{\sqrt{n-1}}-\frac{1}{2}\sum_{i=1}^{n-1}\left(\frac{1-p}{p}\right)^{2\mathbb{I}_S(i)-1}\frac{\beta_i^2}{n-1}\right),$$

where we are taking the uniform distribution for the sets S_1, S_2 over their domains. Rearranging gives the desired result.

Now we compute the bipartite version.

Proof of Proposition B.6. We have that our vertex is the last vertex v'_m , corresponding to degree s_m . Given $S \subseteq W$ of size s_m (which we abusively identify with a set of integers), let $\mathbf{t}_S =$

 $(t_1 - \mathbb{1}_{1 \in S}, \dots, t_n - \mathbb{1}_{n \in S})$. As in Theorem B.10, let

$$\overline{t} = \frac{1}{n} \sum_{i=1}^{n} t_i; \qquad \overline{t}_S = \frac{1}{n} \sum_{i=1}^{n} t_{S,i} = \overline{t} - \frac{s_m}{n}$$

$$r = \overline{t}n; \qquad r_S = \overline{t}_S n = r - s_m,$$

$$\mu = \frac{\overline{t}}{m}; \qquad \mu_S = \frac{\overline{t}_S}{m-1} = \frac{m}{m-1} \mu - \frac{s_m}{(m-1)n},$$

$$\gamma_2^2(t) = \frac{1}{n^2} \sum_{i=1}^{n} (t_i - \overline{t})^2; \qquad \gamma_2^2(t_S) = \frac{1}{n^2} \sum_{i=1}^{n} (t_{S,i} - \overline{t}_S)^2.$$

Let \mathbf{s}' be \mathbf{s} restricted to the first m-1 values, and let

$$\overline{s}' = \frac{1}{m-1} \sum_{i=1}^{m-1} s_i, \qquad \gamma_2^2(s') = \frac{1}{(m-1)^2} \sum_{i=1}^{m-1} (s_i - \overline{s}')^2.$$

Note that (\mathbf{s}, \mathbf{t}) and each $(\mathbf{s}', \mathbf{t}_S)$ clearly satisfy the conditions of Theorem B.10 due to our given hypotheses. Note that

$$\gamma_2^2(t) = \gamma_2^2(t_S) + O(n^{-1/4}), \qquad \mu_S = \mu + O(n^{-3/4}), \qquad \gamma_2^2(s') = \gamma_2^2(s) + O(n^{-1/4})$$

due to the given hypotheses.

Now define

$$A = \frac{\binom{(m-1)n}{r-s_m}^{-1}}{\binom{mn}{r}^{-1}} p^{s_m} (1-p)^{n-s_m}$$

and recall $t_i = pm + \beta_i \sqrt{p(1-p)m}$ and $s_m = pn + \alpha \sqrt{p(1-p)n}$. We have

$$r - pmn = \sum_{i=1}^{n} (t_i - pm) = \sqrt{p(1-p)m} \sum_{i=1}^{n} \beta_i$$

By the given hypotheses and Lemma B.11 (with m, n switched) we therefore derive

$$A = \exp\left(\frac{\left(\sum_{i=1}^{n} \beta_i\right)\left(\sum_{i=1}^{n} \beta_i - 2\sqrt{mn\alpha}\right)}{2mn} + O(n^{-1/6})\right)$$

We therefore see from Theorem B.10 (with m and n switched) that

$$\begin{split} \mathbb{P}[N(v_m') = S] \\ &= (1 + O(n^{-1/8})) \frac{\exp\left(-\frac{1}{2}\left(1 - \frac{\gamma_2(s')^2}{\mu(1-\mu)}\right)\left(1 - \frac{\gamma_2(t_S)^2}{\mu(1-\mu)}\right)\right) \binom{(m-1)n}{r} \prod_{i=1}^{m-1} \binom{n}{s_i} \prod_{i=1}^n \binom{m-1}{t_i - 1_{S(i)}}}{\exp\left(-\frac{1}{2}\left(1 - \frac{\gamma_2(s')^2}{\mu(1-\mu)}\right)\left(1 - \frac{\gamma_2(t)^2}{\mu(1-\mu)}\right)\right) \binom{mn}{r}^{-1} \prod_{i=1}^m \binom{n}{s_i} \prod_{i=1}^n \binom{m}{t_i}}{\binom{mn}{r}^{-1} \prod_{i=1}^m \binom{n}{s_i} \prod_{i=1}^n \binom{m-1}{t_i - 1_{S(i)}}} \\ &= (1 + O(n^{-1/8})) \frac{\binom{(m-1)n}{r}^{-1} \prod_{i=1}^m \binom{n}{s_i} \prod_{i=1}^n \binom{m}{t_i}}{\binom{mn}{r}^{-1} \prod_{i=1}^m \binom{n}{s_i} \prod_{i=1}^n \binom{m}{t_i}} \\ &= (1 + O(n^{-1/8})) \frac{A}{\binom{n}{s_m}} p^{-s_m} (1 - p)^{-(n-s_m)} \prod_{i \in S} \frac{t_i}{m} \prod_{i \notin S} \frac{m - t_i}{m} \\ &= (1 + O(n^{-1/8})) \frac{A}{\binom{n}{s_m}} \prod_{i \in S} \left(1 + \beta_i \sqrt{\frac{1-p}{pm}}\right) \prod_{i \notin S} \left(1 - \beta_i \sqrt{\frac{p}{(1-p)m}}\right) \\ &= \frac{A}{\binom{n}{s_m}} \exp\left(-\sqrt{\frac{p}{1-p}} \sum_{i=1}^n \left(-\frac{1-p}{p}\right)^{1_S(i)} \frac{\beta_i}{\sqrt{m}} - \frac{1}{2} \sum_{i=1}^n \left(\frac{1-p}{p}\right)^{21_S(i)-1} \frac{\beta_i^2}{m} + O(n^{-1/8})\right) \\ &= 29 \end{split}$$

for each $S \in {[n] \choose s_m}$. Therefore

$$(1+O(n^{-1/8}))\frac{\binom{n}{s_m}}{A\binom{h}{t}\binom{n-h}{s_m-t}}\mathbb{P}[\deg_V(v'_m)=t] = \mathbb{E}_{\substack{S_1 \sim \binom{V}{t}\\S_2 \sim \binom{W \setminus V}{s_m-t}\\S=S_1 \cup S_2}} \exp\left(-\sqrt{\frac{p}{1-p}}\sum_{i=1}^n \left(-\frac{1-p}{p}\right)^{\mathbbm{1}_S(i)}\frac{\beta_i}{\sqrt{m}} - \frac{1}{2}\sum_{i=1}^n \left(\frac{1-p}{p}\right)^{2\mathbbm{1}_S(i)-1}\frac{\beta_i^2}{m}\right),$$

where we are taking the uniform distribution for the sets S_1, S_2 over their domains. Rearranging gives the desired result.

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA *Email address*: {asah,msawhney}@mit.edu