Combinatorics, Probability and Computing (2016) 25, 269–299.
doi:10.1017/S0963548315000103

On the Method of Typical Bounded Differences

LUTZ WARNKE†

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,
Wilberforce Road, Cambridge CB3 0WB, UK
(e-mail: L.Warnke@dpmms.cam.ac.uk)

Received 23 December 2012; revised 22 December 2014; first published online 27 August 2015

Concentration inequalities are fundamental tools in probabilistic combinatorics and theoretical computer science for proving that functions of random variables are typically near their means. Of particular importance is the case where $f(X)$ is a function of independent random variables $X = (X_1, \dots, X_n)$. Here the well-known *bounded differences inequality* (also called McDiarmid's inequality or the Hoeffding–Azuma inequality) establishes sharp concentration if the function f does not depend too much on any of the variables. One attractive feature is that it relies on a very simple Lipschitz condition (L): it suffices to show that $|f(X) - f(X')| \leqslant c_k$ whenever X, X' differ only in X_k. While this is easy to check, the main disadvantage is that it considers *worst-case* changes c_k, which often makes the resulting bounds too weak to be useful.

In this paper we prove a variant of the bounded differences inequality which can be used to establish concentration of functions $f(X)$ where (i) the *typical* changes are small, although (ii) the worst case changes might be very large. One key aspect of this inequality is that it relies on a simple condition that (a) is easy to check and (b) coincides with heuristic considerations as to why concentration should hold. Indeed, given an event Γ that holds with very high probability, we essentially relax the Lipschitz condition (L) to situations where Γ occurs. The point is that the resulting *typical* changes c_k are often much smaller than the worst case ones.

To illustrate its application we consider the reverse H-free process, where H is 2-balanced. We prove that the final number of edges in this process is concentrated, and also determine its likely value up to constant factors. This answers a question of Bollobás and Erdős.

2010 *Mathematics subject classification*: Primary 60C05

1. Introduction

In probabilistic combinatorics and theoretical computer science it is often crucial to predict the likely value(s) of functions of random variables. More precisely, in many applications $f(X)$ is a function of independent random variables $X = (X_1, \dots, X_N)$, and we need to

† Part of the work was done while interning with the Theory Group at Microsoft Research, Redmond, WA.

prove that it is concentrated in a narrow range around its expected value, that is, $f(X)$ is typically about $\mathbb{E}f(X)$. The crux is that the functions of interest are often defined in an indirect or complicated way, so that basic bounds such as Chebyshev's inequality are either hard to evaluate or give error bounds that are too weak in applications. In this work we thus investigate easily checked conditions which ensure that the function $f(X)$ is close to its mean with very high probability, *i.e.*, that large deviations from $\mathbb{E}f(X)$ are highly unlikely.

An important paradigm in this area of research (see *e.g.* [48]) states that a random function which depends 'smoothly' on many independent random variables should be sharply concentrated, meaning that $|f(X)-\mathbb{E}f(X)| = o(\mathbb{E}f(X))$ holds with probability very close to one. In many applications (*e.g.*, the design of randomized algorithms or random graph theory) each random variable X_k takes values in a set Λ_k, and in this case a discrete Lipschitz condition for

$$f : \prod_{j\in[N]} \Lambda_j \to \mathbb{R}$$

conveniently ensures that $f(X)$ does not depend too much on any of the variables, where $[N] = \{1,\dots,N\}$. Perhaps the most famous result in this context is the *bounded differences inequality* (also called McDiarmid's inequality or the Hoeffding–Azuma inequality), which is now widely used in discrete mathematics and computer science; see *e.g.* the surveys [34, 35]. Here we only state its one-sided version since the analogous lower tail estimate $\mathbb{P}(f(X) \leqslant \mathbb{E}f(X)-t)$ follows by considering the function $-f(X)$.

Theorem 1.1 ('Bounded differences inequality'). *Let $X = (X_1,\dots,X_N)$ be a family of independent random variables with X_k taking values in a set Λ_k. Assume that the function $f : \prod_{j\in[N]} \Lambda_j \to \mathbb{R}$ satisfies the following* Lipschitz condition.

(L) *There are numbers $(c_k)_{k\in[N]}$ such that, whenever $x, \tilde{x} \in \prod_{j\in[N]} \Lambda_j$ differ only in the kth coordinate, we have*

$$|f(x)-f(\tilde{x})| \leqslant c_k. \tag{1.1}$$

Then, for all $t \geqslant 0$ we have

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X)+t) \leqslant \exp\left(-\frac{2t^2}{\sum_{k\in[N]} c_k^2}\right). \tag{1.2}$$

While the simplicity of (L) makes this inequality very intuitive and easy to apply, perhaps its main drawback is that it considers *worst case* changes. In particular, the resulting concentration bounds are rather weak (or even trivial) in situations where the worst case c_k are much larger than the *typical* changes. A standard example is $f(X)$ counting the number of triangles in the binomial random graph $G_{n,p}$: since every pair of vertices has up to $n-2$ common neighbours, the worst case is $c_k = \Theta(n)$, which is much larger than we expect from the $\Theta(np^2)$ common neighbours we usually have for $p \geqslant n^{-1/2+\varepsilon}$. In fact, here Theorem 1.1 only gives trivial estimates for $p = O(n^{-1/3})$, but

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

it seems plausible that concentration should hold in such applications where the typical changes are much smaller than the worst case ones.

This motivated a line of research [30, 27, 42, 43, 49, 50] which focused on exponential tail inequalities of the form $\mathbb{P}(|f(X)-\mathbb{E}f(X)| \geqslant t) \leqslant e^{-g(f,X,t)}$ in situations where, intuitively speaking, the *average* Lipschitz coefficients c_k are small. Pioneered by Kim and Vu [30], such results usually require $f(X)$ to have a special structure (a polynomial of independent random variables of a certain type), reducing their range of applications compared to (1.2). Furthermore, the assumptions of such techniques are often more involved (and harder to check) than the simple Lipschitz condition (L).

In contrast, much less research has been devoted to developing easy-to-use tools for proving concentration results in such situations. The Hoeffding–Azuma inequality [24, 3] implies, for example, that (1.2) essentially remains true if we relax (1.1) to worst case conditional expected changes:

$$|\mathbb{E}(f(X) \mid X_1,\dots,X_k) - \mathbb{E}(f(X) \mid X_1,\dots,X_{k-1})| \leqslant c_k. \tag{1.3}$$

While this might be useful in certain textbook examples, it typically has two main drawbacks in involved combinatorial applications: (a) conditional expectations are usually difficult to calculate and (b) it often yields no substantial improvement (for, say, $k \geqslant N/2$ the worst case in (1.3) over all choices of $X_1,\dots,X_k$ is often comparable to (1.1)). There are also some approaches which allow (1.3) to be violated occasionally [15, 16, 29, 35, 44], but these usually require knowledge about conditional probability distributions, making them particularly difficult to apply when $f(X)$ has no simple explicit form.

1.1. Typical bounded differences inequality

In this paper we develop a variant of the bounded differences inequality which can be used to establish concentration of functions $f(X)$ where (i) the *typical* changes are small although (ii) the worst case changes might be very large. One key aspect of this inequality is that it relies on a simple and attractive condition that (a) is easy to check and (b) coincides with heuristic considerations as to why concentration should hold. Indeed, given a 'good' event Γ that holds with very high probability, we essentially relax the Lipschitz condition (L) to situations where Γ occurs. More precisely, for the sake of proving concentration the following inequality usually allows us to restrict our attention to such *typical* changes, which are often much smaller than the worst case ones.

Theorem 1.2 ('Typical bounded differences inequality'). *Let $X=(X_1,\dots,X_N)$ be a family of independent random variables with X_k taking values in a set Λ_k. Let $\Gamma \subseteq \prod_{j\in[N]}\Lambda_j$ be an event and assume that the function $f:\prod_{j\in[N]}\Lambda_j \to \mathbb{R}$ satisfies the following* typical Lipschitz condition.

(TL) *There are numbers $(c_k)_{k\in[N]}$ and $(d_k)_{k\in[N]}$ with $c_k \leqslant d_k$ such that, whenever $x,\tilde{x} \in \prod_{j\in[N]}\Lambda_j$ differ only in the kth coordinate, we have*

$$|f(x)-f(\tilde{x})| \leqslant \begin{cases} c_k & \text{if } x \in \Gamma, \\ d_k & \text{otherwise.} \end{cases} \tag{1.4}$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Then, for all numbers $(\gamma_k)_{k\in[N]}$ *with* $\gamma_k \in (0,1]$, *there is an event* $\mathcal{B} = \mathcal{B}(\Gamma, (\gamma_k)_{k\in[N]})$ *satisfying*

$$\mathbb{P}(\mathcal{B}) \leqslant \mathbb{P}(X \notin \Gamma) \sum_{k\in[N]} \gamma_k^{-1} \quad \text{and} \quad \neg\mathcal{B} \subseteq \Gamma, \tag{1.5}$$

such that for all $t \geqslant 0$ *we have*

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) \leqslant \exp\left(-\frac{t^2}{2\sum_{k\in[N]}(c_k + e_k)^2}\right), \tag{1.6}$$

where $e_k = \gamma_k(d_k - c_k)$.

Remark 1. In many applications the following simple consequence of (1.5)–(1.6) suffices:

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t) \leqslant \exp\left(-\frac{t^2}{2\sum_{k\in[N]}\big(c_k + \gamma_k(d_k - c_k)\big)^2}\right) + \mathbb{P}(X \notin \Gamma) \sum_{k\in[N]} \gamma_k^{-1}. \tag{1.7}$$

Remark 2. If each X_k takes only two values (*i.e.*, when $|\Lambda_k| = 2$) the exponents in (1.6) and (1.7) may be multiplied by a factor of 4, analogous to the standard bound (1.2).

As before, this inequality is only stated for the upper tail since an application to $-f(X)$ yields the same estimate for $\mathbb{P}(f(X) \leqslant \mathbb{E}f(X) - t)$. One key property of the 'bad' event $\mathcal{B}$ is that it does *not* depend on the function $f(X)$, so that (1.6) can be used as a tail estimate in union bound arguments. We expect that the *typical* changes c_k are usually *substantially smaller* than the worst case d_k, and the 'compensation factor' γ_k is supposed to moderate the effects of the d_k in (1.6) and (1.7). Indeed, in the typical application γ_k will be very small (this choice is possible if Γ holds with very high probability), so that we can think of $e_k = \gamma_k(d_k - c_k)$ as a negligible 'error term' (see [4, 21] for recent applications in probabilistic combinatorics and theoretical computer science). With this in mind, perhaps the most important aspect of Theorem 1.2 is that it may still yield concentration in situations where Theorem 1.1 only gives trivial bounds due to very large worst case Lipschitz coefficients.

To illustrate the ease of application of Theorem 1.2, consider again the example where $f(X)$ counts the number of triangles in $G_{n,p}$ (here $X_i \in \{0,1\}$ corresponds to the presence or absence of the ith edge). Define Γ as the event that every pair of vertices has at most $\Delta = \max\{2np^2, n^\varepsilon\}$ common neighbours, which fails with probability at most $e^{-\Omega(n^\varepsilon)}$ by standard Chernoff bounds. It is straightforward to see that in this case (TL) holds with, say, $c_k = \Delta$ and $d_k = n$. Setting $\gamma_k = n^{-1}$ we thus have $e_k = o(c_k)$ and $\mathbb{P}(\mathcal{B}) \leqslant e^{-\Omega(n^\varepsilon)}$, which means that both terms are negligible for the sake of establishing concentration. It follows that for $p \geqslant n^{-2/3+\varepsilon}$ the typical bounded differences inequality (Theorem 1.2) yields tight concentration of the number of triangles (with $\mathbb{P}(f(X) \notin (1 \pm n^{-\varepsilon})\mathbb{E}f(X)) \leqslant e^{-\Omega(n^\varepsilon)}$, say), whereas Theorem 1.1 already fails for $p = n^{-1/3}$. Note that we picked all parameters in a uniform way, setting $c_k = C$, $d_k = D$ and $\gamma_k = \gamma$; this might be convenient in many applications (where $\gamma \approx C/D$ should often suffice).

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Remark 3. For simplicity we stated Theorems 1.1 and 1.2 for finite sets Λ_k, but both extend in the obvious way to subsets $\Lambda_k \subseteq \mathbb{R}$ (and, more generally, Polish spaces); see also Section 2.4 in [17]. For example, (1.4) is then replaced by the assumption that, for each $k \in [N]$, almost surely

$$\sup_{x_k \in \Lambda_k} |f(X) - f(X_1, \dots, X_{k-1}, x_k, X_{k+1}, \dots, X_N)| \leqslant \begin{cases} c_k & \text{if } \Gamma \text{ holds,} \\ d_k & \text{otherwise.} \end{cases} \tag{1.8}$$

The remainder of this paper is organized as follows. In Sections 1.1.1–1.1.4 we present several extensions of Theorem 1.2 that, for example, yield improved estimates for 0–1 variables X_i (see Theorem 1.3), introduce a convenient 'two-sided' typical Lipschitz condition (see Theorem 1.5), or weaken the independence assumption (see Theorem 1.9), for example to incorporate the uniform random graph $G_{n,m}$. In Section 1.2 we put these estimates in a wider perspective, and discuss one key application where these are crucial for answering an old question of Bollobás and Erdős. Section 2 is devoted to the proof of our new concentration inequalities, which in Section 3 are then illustrated by an application to the so-called reverse H-free process. Finally, in the Appendix we show that certain complications of Theorem 1.2 are necessary.

1.1.1. Improvement for Bernoulli random variables. If the underlying probability space is generated by independent Bernoulli random variables we can strengthen Theorem 1.2 significantly. For example, in the common situation where the success probabilities are all equal to p (as in $G_{n,p}$) the following natural extension of Theorem 1.2 essentially allows us to multiply the denominator of (1.6) by an extra factor of p (on an intuitive level one can perhaps think of this as applying Theorem 1.2 after conditioning on $\Theta(Np)$ variables being 'relevant').

Theorem 1.3 ('Typical bounded differences inequality for 0–1 variables'). *Suppose that $X = (X_1, \dots, X_N)$ is a family of independent random variables with $X_k \in \{0,1\}$ and $p_k = \mathbb{P}(X_k = 1)$. Let $\Gamma \subseteq \{0,1\}^N$ be an event and assume that the function $f : \{0,1\}^N \to \mathbb{R}$ satisfies the typical Lipschitz condition* (TL) *with $\Lambda_k = \{0,1\}$. Then, for all numbers $(\gamma_k)_{k\in[N]}$ with $\gamma_k \in (0,1]$, there is an event $\mathcal{B} = \mathcal{B}(\Gamma, (\gamma_k)_{k\in[N]})$ satisfying* (1.5) *such that for all $t \geqslant 0$ we have*

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) \leqslant \exp\left(-\frac{t^2}{2\sum_{k\in[N]}(1-p_k)p_k(c_k+e_k)^2 + 2Ct/3}\right), \tag{1.9}$$

where $e_k = \gamma_k(d_k - c_k)$ and $C = \max_{k\in[N]}(c_k + e_k)$,

Remark 4. If $f(X)$ and Γ are either both monotone increasing or decreasing, we have

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t) \leqslant \frac{\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B})}{1 - \mathbb{P}(\mathcal{B})}. \tag{1.10}$$

In typical applications of this inequality we hope to be able to ignore the 'error term' $2Ct/3$ (and select γ_k such that $c_k + e_k \approx c_k$, as before). In this case (1.9) is close to $e^{-t^2/(2\sum p_k c_k^2)}$, which for $p_k = o(1)$ is a significant improvement of the corresponding

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

$e^{-2t^2/(\sum c_k^2)}$ from Remark 2. For example, in the case of triangles in $G_{n,p}$ this allows us to extend the concentration result of the previous section to edge probabilities satisfying $p \geqslant n^{-4/5+\varepsilon}$ (see also [14] for a recent application in the context of additive combinatorics). In fact, the estimates implied by (1.9) and (1.10) are sometimes comparable to those of Janson's inequality [25, 39]: see Section 1.2.2.

Ignoring the 'good' event Γ in Theorem 1.3, we also obtain a strengthening of Theorem 1.1. Since this natural variant of the bounded differences inequality does not seem to be as widely known, we explicitly state it for ease of reference (it can be deduced from Theorem 3.8 in McDiarmid's survey [35], and a slightly weaker version of (1.11) follows directly from Theorem 3.9 therein; Alon, Kim and Spencer [2] also proved a comparable inequality that applies to small values of t only: for those the contribution of Ct to the denominator of (1.11) is negligible).

Corollary 1.4 ('Bounded differences inequality for 0–1 variables'). *Let $X = (X_1, \dots, X_N)$ be a family of independent random variables with $X_k \in \{0,1\}$ and $p_k = \mathbb{P}(X_k = 1)$. Assume that the function $f : \{0,1\}^N \to \mathbb{R}$ satisfies the Lipschitz condition* (L) *with $\Lambda_k = \{0,1\}$. Let $C = \max_{k\in[N]} c_k$. Then, for all $t \geqslant 0$ we have*

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t) \leqslant \exp\left(-\frac{t^2}{2\sum_{k\in[N]}(1-p_k)p_k c_k^2 + 2Ct/3}\right). \tag{1.11}$$

Proof. Apply Theorem 1.3 with $\Gamma = \{0,1\}^N$ and $d_k = c_k$. □

This extends Bernstein's inequality (a strengthening of the Chernoff bounds for small deviations: see *e.g.* Remark 2.9 in [26]), which applies to sums of independent random variables. One key aspect of (1.11) is that it is almost tight for linear functions $f(X) = \sum_k c_k X_k$, in which case $V = \operatorname{Var} f(X) = \sum_k p_k(1-p_k)c_k^2$. Indeed, the estimate of Corollary 1.4 is then close to $e^{-t^2/(2V)}$ for t not too large, which is exactly the tail behaviour predicted by the central limit theorem.

Remark 5. Our arguments in fact yield a slightly stronger form of (1.9) and (1.11), analogous to Bennet's sharpening of the Chernoff bounds (see *e.g.* Remark 2.9 in [26]). Indeed, for $\phi(x) = (1+x)\log(1+x) - x$ we can improve terms of the form $e^{-t^2/(2V+2Ct/3)}$ to $e^{-V/C^2\cdot\phi(Ct/V)}$, where V equals $\sum_k(1-p_k)p_k(c_k+e_k)^2$ and $\sum_k(1-p_k)p_k c_k^2$ in (1.9) and (1.11). For $t = \omega(V/C)$ these refined estimates sharpen the exponents from $\Theta(t/C)$ to $\Theta(t/C \cdot \log(Ct/V))$, that is, they yield a logarithmic improvement.

Remark 6. Theorem 1.3 and Corollary 1.4 extend with minor modifications to the case where each X_k takes values in a set Λ_k and satisfies $\max_{\eta\in\Lambda_k} \mathbb{P}(X_k = \eta) \geqslant 1 - p_k$. Indeed, these hold after replacing all occurrences of $(1-p_k)p_k$ and $c_k + e_k$ by p_k and $c_k + e_k(1-p_k)^{-1}$, respectively.

1.1.2. Two-sided Lipschitz conditions. The typical Lipschitz condition (TL) is 'one-sided': $|f(x) - f(\tilde{x})| \leqslant c_k$ is supposed to hold if $x \in \Gamma$. This keeps the formulas simple, but in

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

many applications it is easier (and perhaps more natural) to verify a 'two-sided' condition where $x, \tilde{x} \in \Gamma$ holds. The following theorem states that we may use a two-sided variant of (TL) at the cost of slightly increasing the 'error term' e_k.

Theorem 1.5 ('Two-sided typical Lipschitz condition'). *Let* $\min_{\eta \in \Lambda_k} \mathbb{P}(X_k = \eta) \geqslant q_k$. *Then Theorems 1.2, 1.3 and Remarks 2, 5, 6 remain valid with* $e_k = 2\gamma_k(d_k - c_k)q_k^{-1}$ *if the Lipschitz condition* (1.4) *of* (TL) *is replaced by the following two-sided variant:*

$$|f(x) - f(\tilde{x})| \leqslant \begin{cases} c_k & \text{if } x, \tilde{x} \in \Gamma, \\ d_k & \text{otherwise.} \end{cases} \tag{1.12}$$

Whenever q_k^{-1} is not too big, (1.12) seems the most convenient condition: it is much simpler to check than (1.4) and does not substantially deteriorate the error bounds. For example, in the random graph $G_{n,p}$ we usually have $q_k^{-1} \leqslant n^2$, which in the typical application with $\mathbb{P}(X \notin \Gamma) \leqslant n^{-\omega(1)}$ can be compensated by adapting γ_k accordingly (also note that $(1 - p_k)p_k q_k^{-1} \leqslant 1$ in the case of Theorem 1.3).

Remark 7. As pointed out by Oliver Riordan, it is possible to bootstrap (1.12) from (1.4) by modifying the good event. Indeed, defining $\Gamma' \subseteq \Gamma$ such that for $x \in \Gamma'$ any single coordinate change results in a sample point satisfying $\tilde{x} \in \Gamma$, it follows that the one-sided condition $x \in \Gamma'$ implies the two-sided condition $x, \tilde{x} \in \Gamma$. Using the bound $\mathbb{P}(X \notin \Gamma') \leqslant \mathbb{P}(X \notin \Gamma) \sum_{k \in [N]} q_k^{-1}$ this approach often leads to estimates that are comparable with Theorem 1.5 (in fact, monotonicity of Γ also transfers to Γ').

In some applications (a) the q_k are very small, and (b) exploiting $x \in \Gamma$ when bounding $|f(x) - f(\tilde{x})|$ is difficult, in which case neither the two-sided condition (1.12) nor the one-sided condition (1.4) seem to be suitable. In an attempt to deal with such situations we introduce an intermediate variant, which is 'locally' two-sided: it only requires the (one-sided) typical Lipschitz condition (1.4) to hold when each coordinate of *both* sample points $x, \tilde{x}$ satisfies some local 'good' event $x_j, \tilde{x}_j \in \Gamma_j$.

Theorem 1.6 ('Typical bounded differences inequality with truncation'). *Suppose* $X = (X_1, \dots, X_N)$ *is a family of independent random variables with* X_k *taking values in a set* Λ_k. *Suppose* $(\Gamma_k)_{k \in [N]}$ *and* $\Gamma \subseteq \prod_{j \in [N]} \Gamma_j$ *are events with* $\Gamma_k \subseteq \Lambda_k$. *Assume that the function* $f : \prod_{j \in [N]} \Lambda_j \to \mathbb{R}$ *satisfies the Lipschitz condition* (1.4) *of* (TL) *only for all* $x, \tilde{x} \in \prod_{j \in [N]} \Gamma_j$ *that differ only in the kth coordinate, and that* $|f(x) - f(\tilde{x})| \leqslant s$ *for all* $x, \tilde{x} \in \prod_{j \in [N]} \Lambda_j$ *that differ in at least one coordinate. Then for all numbers* $(\gamma_k)_{k \in [N]}$ *with* $\gamma_k \in (0, 1]$ *there is an event* $\mathcal{B} = \mathcal{B}(\Gamma, (\gamma_k)_{k \in [N]})$ *satisfying* (1.5) *such that for all* $t \geqslant 0$ *we have*

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t + \Delta \text{ and } \neg\mathcal{B}) \leqslant \exp\left(-\frac{t^2}{2\sum_{k \in [N]}(c_k + e_k)^2}\right), \tag{1.13}$$

where $\Delta = s\mathbb{P}(X \notin \Gamma)$ *and* $e_k = \gamma_k(d_k - c_k)$.

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Remark 8. If $\Gamma = \prod_{j\in[N]} \Gamma_j$ holds we may set $\mathcal{B} = \neg\Gamma$, $e_k = 0$ and multiply the exponent of (1.13) by a factor of 4. If certain monotonicity properties hold we can remove the Δ term: for example, we can set $\Delta = 0$ if $f(x) \geqslant f(\tilde{x})$ whenever $x, \tilde{x} \in \prod_{j\in[N]} \Lambda_j$ differ only in $x_k \in \Lambda_k \setminus \Gamma_k$ and $\tilde{x}_k \in \Gamma_k$.

Theorem 1.6 seems particularly useful when the underlying random variables are grouped into larger blocks B_k, so that each X'_k now takes values in its own product space $\Lambda'_k = \prod_{j\in B_k} \Lambda_j$ (by construction the X'_k are again independent). For example, the so-called 'vertex exposure' of $G_{n,p}$ uses $n-1$ blocks, where X'_k corresponds to the group of edges $E_k = (v_k v_{k+1}, \dots, v_k v_n)$. In this case $q_k \leqslant p^{n-k}$ and the 'good' event Γ of, say, having at most $\Sigma = \max\{2np, n^{\varepsilon}\}$ neighbours can dramatically fail after changing the kth coordinate (the degree of v_k can change up to $n-k$). Here we can overcome these issues using the 'local' event Γ_k that at most Σ edges of E_k are present, so that after a one-coordinate change of $x \in \prod_j \Gamma_j$ from x_k to $\tilde{x}_k \in \Gamma_k$ the degree of every vertex changes by at most Σ (if $x \in \Gamma$ then every vertex has at most 2Σ neighbours). In other words, the local Γ_k and global Γ can complement each other in order to moderate the large worst case effects (in particular when many variables are associated with each coordinate).

Theorem 1.6 also allows us to routinely apply certain truncation arguments (without *ad hoc* calculations). A typical example is $f(X) = \sum_k X_k$ with X_k having exponential tails, where one often first proves concentration of, say, $\sum_k \min\{X_k, C\log N\}$, and then transfers this result to the original sum; see *e.g.* [1, 13]. Here (1.13) almost immediately yields concentration of $f(X)$ via the local events Γ_k for which $X_k \leqslant C\log N$ occurs (setting $\Gamma = \prod_j \Gamma_j$, $d_k = c_k = C\log N$ and $\gamma_k = 1$).

1.1.3. Dynamic exposure of the variables. The previous inequalities can be refined by exposing the values of the random variables X_i one by one in an *adaptive* order. Intuitively this allows us to exploit that after having learned the values of certain variables, some other X_j may no longer influence the value of $f(X)$. This approach was introduced by Alon, Kim and Spencer [2], and is particularly useful whenever we can determine $f(X)$ without knowing the value of all random variables. More formally, a *strategy* sequentially exposes $X_{q_1}, X_{q_2}, \dots$, where each index $q_i = q_i(X_{q_1}, \dots, X_{q_{i-1}})$ may depend on the previous outcomes and indices (we use the convention that $q_{k+1} = q_k$ if $f(X)$ is determined by $(X_{q_1}, \dots, X_{q_k})$ with $k < N$); every strategy has a natural representation in the form of a decision tree (see also Section 3 in [2]). With a fixed strategy in mind, for every possible outcome $X = (X_1, \dots, X_N)$ we obtain a set of queried indices $Q \subseteq [N]$, and by $\mathcal{Q}$ we denote the set of all possible such query sets Q. The resulting key improvement is that in most inequalities we essentially may replace $k \in [N]$ with $k \in Q$ for some 'worst case' set of indices $Q \in \mathcal{Q}$ (note that $\gamma_k = \gamma$ is a typical choice in applications).

Theorem 1.7 ('Dynamic exposure of the variables'). *Let $\gamma_k = \gamma$ for all $k \in [N]$. Then, for any strategy, Theorems 1.1, 1.2, 1.3, 1.5, 1.6, Corollary 1.4 and Remarks 2, 5, 6, 8 remain valid with $\sum_{k\in[N]}$ replaced by $\max_{Q\in\mathcal{Q}} \sum_{k\in Q}$ and $\max_{k\in[N]}$ replaced by $\max_{Q\in\mathcal{Q}} \max_{k\in Q}$, with the addition that $\mathcal{B}$ depends on the query strategy.*

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Theorem 1.8 ('Monotone dynamic exposure of the variables'). *Consider any strategy satisfying $q_{i+1} \geqslant q_i$ in each step. Then Theorems 1.1, 1.2, 1.3, 1.5, 1.6, Corollary 1.4 and Remarks 2, 4, 5, 6, 8 remain valid with $\sum_{k\in[N]}$ replaced by $\max_{Q\in\mathcal{Q}} \sum_{k\in Q}$ and $\max_{k\in[N]}$ replaced by $\max_{Q\in\mathcal{Q}} \max_{k\in Q}$, with the exception that* (1.5) *remains unchanged* (*here $\mathcal{B}$ does not depend on the query strategy*).

Applied to Corollary 1.4, Theorem 1.3 and Remark 5, these results tighten and extend an inequality of Alon, Kim and Spencer [2], which is based on the Lipschitz condition (L). In certain applications dynamic exposure yields significant improvements, and for an illustration we refer to Claim 2 in [2], where it is crucial to reduce (the order of magnitude of) the number of queried variables. Further refinements are possible by using adaptive Lipschitz bounds c_k, which is perhaps most easily exploited by tailoring the arguments of Section 2 to the specific application.

One key feature of Theorem 1.8 is that the 'bad' event $\mathcal{B}$ does not depend on the strategy used, making it particularly useful in union bound arguments. As an illustration consider the example where $f(X) = f_U(X)$ counts, in $G_{n,p}$, the number of triangles in a subset $U \subseteq V$ of the vertices. Define Γ and Δ as in Section 1.1. Since $f(X)$ depends only on edges in U, using Theorems 1.3 and 1.8 (sequentially exposing all edges in U) we infer for $|U| = u \geqslant u_0 = u_0(n,p)$ and $\tilde{\Delta} = \min\{\Delta, u\}$ that

$$\mathbb{P}(f(X) \notin (1 \pm n^{-\varepsilon})\mathbb{E}f(X) \text{ and } \neg\mathcal{B}) \leqslant \exp\left(-\Theta\left(\frac{n^{-2\varepsilon}u^6p^6}{u^2p\tilde{\Delta}^2 + n^{-\varepsilon}u^3p^3\tilde{\Delta}}\right)\right) \leqslant n^{-\omega(u)}.$$

Taking a union bound, the probability that some $U \subseteq V$ with $|U| \geqslant u_0$ has the 'wrong' number of triangles is at most $\mathbb{P}(\mathcal{B}) + n^{-\omega(u_0)}$. Here we crucially exploited that $\mathcal{B}$ is a 'global' event not depending on $f(X)$ or U, so that $\mathbb{P}(\mathcal{B}) \leqslant e^{-\Omega(n^\varepsilon)}$ does not need to 'compete' with the n^u choices for the subsets (this issue often makes traditional bad events ineffective in union bound arguments).

1.1.4. Weakening the independence assumption. The concentration results discussed so far extend to certain dependent random variables $X = (X_1, \dots, X_N)$ that are generated by a sequence of 'nearly' independent (or uniform) random choices. As we shall see, they apply to random permutations $\pi \in S_n$ and uniform random graphs $G_{n,m}$, for example. To motivate the new (GL) condition below, we consider independent random variables, in which case the mapping $\rho_k : \Sigma_a \to \Sigma_b$ that changes the value of the kth coordinate from a to b is a bijection. Here (L) yields $|f(x) - f(\rho_k(x))| \leqslant c_k$ and independence implies

$$\mathbb{P}(X = x \mid X \in \Sigma_a) = \mathbb{P}(X = \rho_k(x) \mid X \in \Sigma_b).$$

With this in mind (1.14) can be viewed as a natural analogue of (1.4) in which the outcomes x and $\tilde{x} = \rho_k(x)$ may differ in more than just one coordinate, and (1.15) accounts for the fact that the variables are not necessarily independent.

Theorem 1.9 ('General bounded differences inequality'). *Let $X = (X_1, \dots, X_N)$ be a family of random variables with X_k taking values in a set Λ_k. Let $\Gamma \subseteq \prod_{j\in[N]} \Lambda_j$ be an event. Then the conclusions of Theorem 1.2 remain valid if instead of* (TL) *the function $f : \prod_{j\in[N]} \Lambda_j \to \mathbb{R}$ satisfies the following* general Lipschitz condition.

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

(GL) *There are numbers* $(c_k)_{k\in[N]}$ *and* $(d_k)_{k\in[N]}$ *with* $c_k \leqslant d_k$ *such that the following holds for any possible sequences of outcomes* $a_1,\dots,,a_{k-1},a$ *and* $a_1,\dots,a_{k-1},b$ *of* $X_1,\dots,X_k$. *Defining*

$$\begin{aligned}\Sigma_z &= \Sigma_z(a_1,\dots,a_{k-1})\\ &= \Big\{x=(a_1,\dots,a_{k-1},z,x_{k+1},\dots,x_N)\in\prod_{j\in[N]}\Lambda_j \;:\; \mathbb{P}(X=x)>0\Big\},\end{aligned}$$

there is an injection $\rho_k=\rho_k(\Sigma_a,\Sigma_b):\Sigma_a\to\Sigma_b$ *such that for all* $x\in\Sigma_a$ *we have*

$$|f(x)-f(\rho_k(x))|\leqslant\begin{cases}c_k & \text{if } x\in\Gamma,\\ d_k & \text{otherwise,}\end{cases}\qquad \text{and} \tag{1.14}$$

$$\mathbb{P}(X=x\mid X\in\Sigma_a)\leqslant\mathbb{P}(X=\rho_k(x)\mid X\in\Sigma_b). \tag{1.15}$$

Remark 9. The proof shows that ρ_k must be a bijection with equality in (1.15). Furthermore, if X_k takes at most two values conditioned on $X_1,\dots,X_{k-1}$, then the exponent in (1.6) may be multiplied by factor of 4. In fact, (1.2) holds if $\Gamma=\prod_{j\in[N]}\Lambda_j$ (or $r_k=0$ below). In addition, for (1.6) to hold with $e_k=\gamma_k r_k\geqslant 0$ it suffices if we relax (GL) to the average Lipschitz condition

$$|\mathbb{E}(f(X)|X\in\Sigma_a)-\mathbb{E}(f(X)\mid X\in\Sigma_b)|\leqslant c_k+r_k\mathbb{P}(X\notin\Gamma\mid X\in\Sigma_a). \tag{1.16}$$

To illustrate the application of the (GL) condition we consider uniform permutations $\pi\in S_n$, which are generated by sequentially choosing each $\pi(k)$ randomly from

$$[n]\setminus\{\pi(1),\dots,\pi(k-1)\}.$$

Here Σ_z contains all π with $\pi(k)=z$ and $\pi(j)=a_j$ for $1\leqslant j<k$. In this case a bijection $\rho_k:\Sigma_a\to\Sigma_b$ is defined by the transposition of a and b (which is trivial when $a=b$), so that $\pi'=\rho_k(\pi)$ satisfies $\pi'(k)=b$, $\pi'(\pi^{-1}(b))=a$ and $\pi'(i)=\pi(i)$ for $\pi(i)\notin\{a,b\}$. Using $|\Sigma_a|=|\Sigma_b|=(n-k)!$ and the uniform measure it is not hard to check that (1.15) holds with equality. We see that to establish (1.14) it suffices to bound $|f(\pi)-f(\pi')|$ whenever π and π' are related via a transposition, which is an intuitive and easily checked condition (this may correspond to changing two coordinates).

One key aspect of (GL) is that it often maintains the simplicity of (L) and (TL). Here uniform probability measures are particularly convenient, for which it suffices to first define bijections $\rho_k:\Sigma_a\to\Sigma_b$ and then check (1.14) only (using $\mathbb{P}(X=x\mid X\in\Sigma_a)=\mathbb{P}(X=x)/\mathbb{P}(X\in\Sigma_a)$, these must satisfy (1.15) with equality). Indeed, extending the permutations example, for random sequences $T=(t_1,\dots,t_m)$ of m distinct elements from W it is enough to estimate $|f(T)-f(T')|$ whenever both sequences are related by changing one coordinate (*i.e.*, $t_k\neq t'_k$) or interchanging the order of two coordinates (*i.e.*, $t_k=t'_j$ and $t_j=t'_k$). Note that this example includes the random graph process and various hypergraph processes as special cases. Since every set with m elements gives rise to $m!$ ordered sequences, the above result also readily carries over to uniform random subsets $S\subseteq W$ of size $|S|=m$: it suffices to bound $|f(S)-f(S')|$ whenever the sets are

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

minimally different, *i.e.*, satisfy $|S \cap S'| = m - 1$ (note that for $m > |W|/2$ better results are obtained by choosing the complement uniformly at random). Here the uniform random graph $G_{n,m}$ and uniform hypergraphs are special cases. Note that the above construction also extends to multiple (independent) random objects; for example, if M random subsets $X = (S_1, \dots, S_M)$ with $S_i \in W_i$ and $|S_i| = m_i$ are chosen independently, it suffices to consider $|f(X) - f(X')|$ only for the cases where X and X' are minimally different in one set; *cf.* [36]. Finally, with similar easily checked conditions (GL) also applies, for example, to finite metric spaces, perfect matchings and the configuration model G_{d}^*; see *e.g.* [37, 53, 10].

Several extensions of Theorem 1.2 carry over to Theorem 1.9 with some minor modifications, and results analogous to those of Sections 1.1.1 and 1.1.2, including a two-sided Lipschitz condition, are stated below (Remark 5 also applies to (1.17) after adjusting V accordingly).

Theorem 1.10 ('General bounded differences inequality for asymmetric variables'). *Let $X = (X_1, \dots, X_N)$ be a family of random variables with X_k taking values in a set Λ_k, where $\max_{\eta \in \Lambda_k} \mathbb{P}(X_k = \eta \mid X_1, \dots, X_{k-1}) \geqslant 1 - p_k$ holds. Let $\Gamma \subseteq \prod_{j \in [N]} \Lambda_j$ be an event and assume that the function $f : \prod_{j \in [N]} \Lambda_j \to \mathbb{R}$ satisfies the general Lipschitz condition* (GL)*. Then for all numbers $(\gamma_k)_{k \in [N]}$ with $\gamma_k \in (0, 1]$ there is an event $\mathcal{B} = \mathcal{B}(\Gamma, (\gamma_k)_{k \in [N]})$ satisfying* (1.5) *such that for all $t \geqslant 0$ we have*

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) \leqslant \exp\left(-\frac{t^2}{2\sum_{k \in [N]} p_k(c_k + e_k)^2 + 2Ct/3}\right), \tag{1.17}$$

where $e_k = \gamma_k(d_k - c_k)(1 - p_k)^{-1}$ and $C = \max_{k \in [N]}(c_k + e_k)$.

Theorem 1.11 ('Two-sided general Lipschitz condition'). *Theorems 1.9, 1.10 and Remark 9 remain valid with $e_k = 2\gamma_k(d_k - c_k)q_k^{-1}$ and $\min_{\eta \in \Lambda_k} \mathbb{P}(X_k = \eta \mid X_1, \dots, X_{k-1}) \geqslant q_k$ if the Lipschitz condition* (1.14) *of* (GL) *is replaced by the following two-sided variant:*

$$|f(x) - f(\rho_k(\tilde{x}))| \leqslant \begin{cases} c_k & \text{if } x, \rho_k(\tilde{x}) \in \Gamma, \\ d_k & \text{otherwise.} \end{cases} \tag{1.18}$$

In addition, $q_k \leqslant |\Lambda_k|^{-1}$ suffices when all possible outcomes occur with the same probability.

The sufficient condition $q_k \leqslant |\Lambda_k|^{-1}$ often makes the two-sided Lipschitz condition of Theorem 1.11 easy to apply. For example, for random permutations $\pi \in S_n$ and random graphs $G_{n,m}$ (or the random graph process) we may take $q_k = n^{-1}$ and $q_k = n^{-2}$, respectively; see also Theorem 3.3 for an application in the context of constrained graph processes.

1.2. Discussion and applications

1.2.1. A wider perspective. As discussed, in probabilistic combinatorics and the analysis of randomized algorithms we frequently need to prove that a random function is not too far from its mean, *e.g.*, that $f(X) \approx \mathbb{E}f(X)$ or $f(X) \leqslant 2\mathbb{E}f(X)$ holds. A common feature of many recent applications is that the functions of interest are only 'smooth enough' on

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

a *high probability* event, whereas their *deterministic* worst case changes are too large for the standard bounded differences inequality (Theorem 1.1) to be effective.

In these cases there is no general method, but in the past certain *ad hoc* arguments have been successfully used in such situations (see *e.g.* [8, 11, 31, 32, 40]). Usually the key idea is to construct a random function $g(X)$ that is a *smooth approximation* of $f(X)$, which, in particular, ensures by definition that the Lipschitz coefficients are always small (approximation usually means that $f(X) \approx g(X)$ holds with high probability, but often $\mathbb{E}f(X) \approx \mathbb{E}g(X)$ is also needed). Here smoothness makes it possible to apply concentration inequalities (often the bounded differences inequality) to $g(X)$, whereas the approximation property ensures that concentration transfers from $g(X)$ to $f(X)$. The main disadvantage of this approach is that it relies on *ad hoc* arguments (which can be involved); in particular, finding suitable approximation functions may require ingenuity.

One aim of this paper is to provide easily applied tools which can *routinely* deal with such situations, establishing concentration in a rather simple way. For example, in the frequent case where the good event Γ holds with probability at least $1 - N^{-\omega(1)}$, we can typically choose $\gamma_k^{-1} = \max|f(X)|$ and then completely ignore the worst case effects; see *e.g.* [4, 14, 21] or the proof of Theorem 3.3 (this approach also applies, for example, to Lemma 14 in [40] and parts of the martingale-based proof of Theorem 2.2 in [32]). In other words, the crucial advantage of our new inequalities is that they can often remove the need for sometimes difficult *ad hoc* arguments using only minimal calculation (typically even coinciding with heuristic considerations).

1.2.2. Comparison with Janson's inequality. In this section we demonstrate that in certain applications our inequalities give exponential estimates that (i) are tight and (ii) successfully compete with the well-known Janson's inequality. To this end we focus on subgraph counts in the binomial random graph $G_{n,p}$ since a concrete example seems more illustrative to us. Henceforth we assume that H is a fixed 2-*balanced* graph, *i.e.*, where H has $e_H \geqslant 2$ edges and all its proper subgraphs $G \subsetneq H$ with $v_G \geqslant 3$ vertices satisfy

$$\frac{e_G - 1}{v_G - 2} \leqslant \frac{e_H - 1}{v_H - 2} = d_2(H). \tag{1.19}$$

This class of graphs includes, for example, trees, complete graphs and cycles of arbitrary size. Let Y_H count the number of H copies in $G_{n,p}$. For 2-balanced graphs it is well known (see *e.g.* [26]) that, assuming $p \geqslant n^{-1/d_2(H)}$, Janson's inequality gives

$$\mathbb{P}(Y_H \leqslant \mathbb{E}Y_H - t) \leqslant \exp\left(-\Theta\left(\frac{t^2}{\mu^2/(n^2p)}\right)\right), \tag{1.20}$$

where $\mu = \mathbb{E}Y_H$. Spencer [45] proved that (the following statement is implied by the proof of Theorem 3 in [45], in turn based on Theorem 5 therein), assuming $p \geqslant n^{-1/d_2(H)}(\log n)^b$ with $0 < b = b(H) < 2$, for every $c > 0$ the following holds with probability at least $1 - n^{-c}$ for $n \geqslant n_0(c, H)$: every pair xy of vertices is contained in at most $\Delta = O(n^{v_H-2}p^{e_H-1})$ extensions to copies of H (for which adding the edge xy completes a copy of H containing xy). The latter event will be our decreasing Γ, which allows us to use

$$c_k = \Theta(n^{v_H-2}p^{e_H-1}) = \Theta(\mu/(n^2p))$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

as well as $d_k = n^{v_H}$ and $\gamma_k = n^{-v_H}$ in our typical bounded differences inequality, so that $e_k = \gamma_k(d_k - c_k) = o(c_k)$. Applying Spencer's result with $c = v_H + 3$, we have $\mathbb{P}(\mathcal{B}) \leqslant n^{-1}$ by (1.5). Note that, since for the lower tail we have $t = O(\mu)$, it follows that

$$\sum_k pc_k^2 + t \max_k c_k = \Theta(\mu^2/(n^2 p)).$$

For the decreasing function $f = -Y_H$, a combination of (1.9) and (1.10) now yields

$$\mathbb{P}(Y_H \leqslant \mathbb{E}f(X) - t) \leqslant \exp\left(-\Theta\left(\frac{t^2}{\mu^2/(n^2 p)}\right)\right),$$

which asymptotically matches (1.20), that is, the estimate of Janson's inequality. In fact, for $t = \Theta(\mu)$ this bound is best possible (up to constants in the exponent) since $G_{n,p}$ contains no edges (and thus no copies of H) with probability $e^{-\Theta(n^2 p)}$; see also [28].

1.2.3. Application: the reverse H-free process. The following variations of the classical random graph processes were proposed by Bollobás and Erdős at the 1990 *Quo Vadis* conference on graph theory in an attempt to improve Ramsey numbers [9, 12].

(i) The *H-free process*, where, starting with an empty graph on n vertices, in each step a new edge is added, chosen uniformly at random from all pairs whose addition does not complete a copy of H.

(ii) The *reverse H-free process*, where, starting with a complete graph on n vertices, in each step an edge is removed, chosen uniformly at random from all edges that are contained in a copy of H.

(iii) The *H-removal process*, where, starting with a complete graph on n vertices, in each step all e_H edges of a copy of H are removed, which is selected uniformly at random from all copies of H.

All of these processes end with an H-free graph, and Bollobás and Erdős asked (among other structural properties) what their typical final number of edges is [9, 12].

These variations have received considerable attention in recent years, in particular the H-free process. Its typical final number of edges is now known up to logarithmic factors [7, 38] for the class of *strictly* 2-*balanced* graphs H, where in (1.19) the inequality is strict. Matching bounds up to constant factors have only been established for some special forbidden graphs and the class of C_ℓ-free processes; see *e.g.* [51, 52]. The final graph of the K_s-free process also yields the best known lower bounds on the Ramsey numbers $R(s,t)$ with $s \geqslant 4$; see [5, 7]. Recently Makai [33] determined the (asymptotic) final number of edges of the reverse H-free process for the class of strictly 2-balanced graphs, but its final graph yields no new estimates for $R(s,t)$. Although the related H-removal process has been studied in several papers, the final number of edges is known up to multiplicative $n^{o(1)}$ factors only in the special case $H = K_3$; see *e.g.* [6, 41, 46].

Using our typical bounded differences inequality, in Section 3 we show that the final number of edges in the reverse H-free process is sharply concentrated when H is 2-balanced (we do not assume *strictly* 2-balanced), and also determine the likely number of edges up to constants. This is in contrast to all known results for the widely studied H-free and H-removal processes. Indeed, in these (a) no sharp concentration results are known, (b) the order of magnitude of the final number of edges is open for most strictly

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

2-balanced graphs, and (c) no general results apply to the class of 2-balanced graphs. As we shall see, when H is a matching the expected final number of edges in the reverse H-free process is $O(1)$. When it comes to concentration we thus restrict our main attention to all other 2-balanced graphs H, which in fact satisfy $d_2(H) \geqslant 1$ (with equality for trees) and $e_H \geqslant 2$. Here our next result shows that the reverse H-free process typically ends with $\Theta(n^{2-1/d_2(H)})$ edges, answering (up to constant factors) the aforementioned question of Bollobás and Erdős from 1990.

Theorem 1.12. *Let H be a 2-balanced graph. There x are constants $a, A > 0$ such that the final number of edges M_n in the reverse H-free process has expectation satisfying $an^{2-1/d_2(H)} \leqslant \mathbb{E}M_n \leqslant An^{2-1/d_2(H)}$. Further, for every $c > 0$ and $n \geqslant n_0(c, H)$, we have $|M_n - \mathbb{E}M_n| \leqslant \sqrt{\mathbb{E}M_n}(\log n)^{4e_H}$ with probability at least $1 - n^{-c}$.*

Our arguments partially generalize to arbitrary graphs. Set $d_2(K_2) = 1/2$ and

$$m_2(H) = \max_{G \subseteq H, e_G \geqslant 1} d_2(G),$$

so that $m_2(H) = d_2(H)$ for 2-balanced graphs H. We show that for any graph H with $e_H \geqslant 2$ the expected final number of edges in the reverse H-free process is $\Theta(n^{2-1/m_2(H)})$, and prove concentration under certain conditions (satisfied, for example, by a clique K_r with an extra edge hanging off); see Section 3. The proof of Theorem 1.12 also extends to a finite family of forbidden graphs $\mathcal{H}$, which for the H-free process was considered in [38]. Indeed, defining the reverse $\mathcal{H}$-free process in the obvious way (always removing a random edge that is contained in a copy of some $H \in \mathcal{H}$) we obtain, for example, the following generalization.

Theorem 1.13. *Let $\mathcal{H} = \{H_1, \dots, H_r\}$ be a family of 2-balanced graphs. Define $H, J \in \mathcal{H}$ such that $d_2(H) = \min_{F \in \mathcal{H}} d_2(F)$ and $e_J = \max_{F \in \mathcal{H}} e_F$. There are constants $a, A > 0$ such that the final number of edges M_n in the reverse $\mathcal{H}$-free process has expectation satisfying $an^{2-1/d_2(H)} \leqslant \mathbb{E}M_n \leqslant An^{2-1/d_2(H)}$. Furthermore, for every $c > 0$ and $n \geqslant n_0(c, \mathcal{H})$ we have $|M_n - \mathbb{E}M_n| \leqslant \sqrt{\mathbb{E}M_n}(\log n)^{4e_J}$ with probability at least $1 - n^{-c}$.*

2. Proofs of the concentration inequalities

We start by proving two general martingale inequalities. These are applied in Section 2.2, where we establish our variants of the bounded differences inequality.

2.1. Martingale inequalities

Our concentration results are based on the following variants of Hoeffding–Azuma/Bernstein-type martingale inequalities. Since they are not stated exactly in this form in the literature, we give short proofs for the reader's convenience (following the slick approach of Freedman [20]). In both we assume that $(\mathcal{F}_k)_{0 \leqslant k \leqslant N}$ is an increasing sequence of σ-algebras, and $(M_k)_{0 \leqslant k \leqslant N}$ is an $(\mathcal{F}_k)_{0 \leqslant k \leqslant N}$-adapted martingale.

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Lemma 2.1 ('Bounded differences martingale inequality'). *Suppose that L_k and U_k are $\mathcal{F}_{k-1}$-measurable variables satisfying $L_k \leqslant M_k - M_{k-1} \leqslant U_k$. Set $S_k = \sum_{i\in[k]}(U_i - L_i)^2$. For every $t \geqslant 0$ and $S > 0$ we have*

$$\mathbb{P}(M_k \geqslant M_0 + t \text{ and } S_k \leqslant S \text{ for some } k \in [N]) \leqslant e^{-2t^2/S}. \tag{2.1}$$

Lemma 2.2 ('Bounded variances martingale inequality'). *Let U_k be an $\mathcal{F}_{k-1}$-measurable variable satisfying $M_k - M_{k-1} \leqslant U_k$. Set $C_k = \max_{i\in[k]} U_i$ and*

$$V_k = \sum_{i\in[k]} \mathrm{Var}(M_i - M_{i-1} \mid \mathcal{F}_{i-1}).$$

Let $\phi(x) = (1+x)\log(1+x) - x$. For every $t \geqslant 0$ and $V, C > 0$ we have

$$\begin{aligned}\mathbb{P}(M_k \geqslant M_0 + t,\ V_k \leqslant V \text{ and } C_k \leqslant C \text{ for some } k \in [N]) &\leqslant e^{-V/C^2\cdot\phi(Ct/V)} \\ &\leqslant e^{-t^2/(2V+2Ct/3)}.\end{aligned} \tag{2.2}$$

Remark 10. Note that V_k generalizes S_k since $\mathrm{Var}(M_i - M_{i-1} \mid \mathcal{F}_{i-1}) = \mathbb{E}((M_i - M_{i-1})^2 \mid \mathcal{F}_{i-1})$ holds (it is not hard to check that $V_k \leqslant S_k/4$). In fact, Lemmas 2.1 and 2.2 extend with minor modifications to supermartingales: defining $V_k = \sum_{i\in[k]} \mathbb{E}((M_i - M_{i-1})^2 \mid \mathcal{F}_{i-1})$ suffices.

Observe that we allow for (accumulative) random bounds on the one-step changes (and other quantities), which for Lemma 2.1 is the main difference to the usual formulation of the classical Hoeffding–Azuma inequality [24, 3]. Lemma 2.2 extends the related Theorem 2.2.2 of Kim and Vu [30] (see also Lemma 3.1 in Vu's survey [50]), and tightens Theorem 3.15 of McDiarmid [35].

Note that L_k, U_k are $\mathcal{F}_{k-1}$-measurable, whereas $M_k - M_{k-1}$ is $\mathcal{F}_k$-measurable. This difference sometimes causes subtle off-by-one errors. For example, as pointed out by Oliver Riordan, $\sum_k \mathbb{P}(|M_k - M_{k-1}| > c_k) \leqslant \eta$ does *not* imply

$$\mathbb{P}(M_N \geqslant M_0 + t) \leqslant e^{-t^2/(2\sum_k c_k^2)} + \eta, \tag{2.3}$$

although this is claimed by Theorem 8.4 in [16]. The problem is that, conditional on $\mathcal{F}_{k-1}$, in the next step it sometimes is *always* possible for $|M_k - M_{k-1}| \leqslant c_k$ to fail (although this might be unlikely). With this in mind, we see that (2.3) holds, for example under the assumption

$$\sum_k \mathbb{P}(\text{it is possible, given } M_1, \dots, M_{k-1}, \text{ that } |M_k - M_{k-1}| > c_k) \leqslant \eta.$$

In fact, assuming that $|M_k - M_{k-1}| \leqslant C_k$ always holds, the approach of [15, 44] implies (2.3) when

$$\sum_k (1 + 2C_k/t) \cdot \mathbb{P}(|M_k - M_{k-1}| > c_k/4) \leqslant \eta.$$

2.1.1. Proof of Lemmas 2.1 and 2.2. Our proofs use the following (standard) inequalities due to Hoeffding [24] and Steiger [47]; they follow, *e.g.*, from the proofs of Lemmas 2.4, 2.6 and 2.8 in McDiarmid's survey [35].

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Lemma 2.3. *Let X be a random variable with $\mathbb{E}(X \mid \mathcal{F}) = 0$. Let L, U be $\mathcal{F}$-measurable random variables. Set $g(x) = (e^x - 1 - x)/x^2$ for $x \neq 0$ and $g(0) = 1/2$. For all $\lambda \geqslant 0$ the following holds:*

$$L \leqslant X \leqslant U \implies \mathbb{E}(e^{\lambda X} \mid \mathcal{F}) \leqslant e^{\lambda^2 (U-L)^2/8} \quad \text{and} \tag{2.4}$$

$$X \leqslant U \implies \mathbb{E}(e^{\lambda X} \mid \mathcal{F}) \leqslant e^{\lambda^2 g(\lambda U) \operatorname{Var}(X \mid \mathcal{F})}. \tag{2.5}$$

Furthermore, $g(x)$ is a non-negative increasing function. □

Lemma 2.4. *Set $\phi(x) = (1+x)\log(1+x) - x$. We have $\phi(x) \geqslant x^2/(2 + 2x/3)$ for all $x \geqslant 0$.* □

Proof of Lemmas 2.1 and 2.2. Set

$$W_k = \sum_{i \in [k]} g(\lambda U_i) \operatorname{Var}(M_i - M_{i-1} \mid \mathcal{F}_{i-1}).$$

The key point is that M_{k-1} and L_k, U_k, S_k, W_k are $\mathcal{F}_{k-1}$-measurable for $k \geqslant 1$. So, by applying (2.4) and (2.5) to $M_k - M_{k-1}$, using $S_0 = W_0 = 0$ we see that

$$Y_k = e^{\lambda(M_k - M_0) - \lambda^2 S_k/8} \quad \text{and} \quad Z_k = e^{\lambda(M_k - M_0) - \lambda^2 W_k}$$

satisfy $\mathbb{E}(Y_k \mid \mathcal{F}_{k-1}) \leqslant Y_{k-1}$ and $\mathbb{E}(Z_k \mid \mathcal{F}_{k-1}) \leqslant Z_{k-1}$, that is, they are supermartingales. We define the stopping time T as the minimum of N and the smallest $k \in [N]$ with $M_k - M_0 \geqslant t$; as usual, we write $i \wedge T$ as shorthand for $\min\{i, T\}$. By construction $(Y_{k \wedge T})_{0 \leqslant k \leqslant N}$ and $(Z_{k \wedge T})_{0 \leqslant k \leqslant N}$ are both supermartingales. In particular, using $S_0 = W_0 = 0$ we have

$$\mathbb{E} Y_{N \wedge T} \leqslant \mathbb{E} Y_0 = 1 \quad \text{and} \quad \mathbb{E} Z_{N \wedge T} \leqslant \mathbb{E} Z_0 = 1.$$

Let $\mathcal{E}$ denote the event that $M_k \geqslant M_0 + t$ and $S_k \leqslant S$ for some $k \in [N]$. Note that $\mathcal{E}$ implies $Y_{N \wedge T} = Y_T \geqslant e^{\lambda t - \lambda^2 S/8}$. So, for $\lambda = 4t/S$, Markov's inequality gives

$$\mathbb{P}(\mathcal{E}) \leqslant \mathbb{P}(Y_{N \wedge T} \geqslant e^{\lambda t - \lambda^2 S/8}) \leqslant e^{\lambda^2 S/8 - \lambda t} = e^{-2t^2/S},$$

which establishes (2.1) and thus Lemma 2.1.

We proceed similarly for $(Z_{k \wedge T})_{0 \leqslant k \leqslant N}$ and let $\mathcal{E}'$ denote the event that $M_k \geqslant M_0 + t$, $V_k \leqslant V$ and $C_k \leqslant C$ for some $k \in [N]$. Using $V_k \geqslant 0$ and monotonicity of $g(x) \geqslant 0$, we see that $\mathcal{E}'$ implies

$$Z_{N \wedge T} \geqslant e^{\lambda t - \lambda^2 g(\lambda C) V_k} \geqslant e^{\lambda t - \lambda^2 g(\lambda C) V}.$$

Using Markov's inequality, $\phi(x) = (1+x)\log(1+x) - x$ and Lemma 2.4, for $\lambda = \log(1 + Ct/V)/C$ it follows that

$$\mathbb{P}(\mathcal{E}') \leqslant e^{\lambda^2 g(\lambda C) V - \lambda t} = e^{-V/C^2 \cdot \phi(Ct/V)} \leqslant e^{-t^2/(2V + 2Ct/3)},$$

which establishes (2.2) and thus Lemma 2.2. □

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

2.2. Bounded differences inequalities

The textbook proof of Theorem 1.1 is based on the Hoeffding–Azuma inequality [24, 3], and essentially uses the 'worst case' Lipschitz condition (1.1) to apply Lemma 2.1 with $|U_k - L_k| \leqslant c_k$. We need some modifications to deal with the obstacle that the 'good' event Γ and thus the 'typical case' in (1.4) does *not* always hold, and these are partially inspired by the seminal work of Shamir and Spencer [44] from 1987.

When $\mathbb{P}(X \notin \Gamma) \leqslant \eta$ holds one might be tempted to add η to the error bound and then always assume that Γ holds. The problem is that in the martingale-based proof one needs to estimate *conditional* expected changes as in (1.3). So, informally speaking, despite $\omega \in \Gamma$ the 'good' event can still fail 'inside' the corresponding expectations. One can try to overcome this by conditioning on Γ, but this usually introduces a new technical problem: then the variables are, in general, not conditionally independent (in which case Lipschitz conditions comparable to (1.4) no longer suffice to bound the expected changes). These technicalities seem to cause some confusion in [18, 22], for example.

We sidestep these issues by noting that for good bounds on conditional *expected* one-step changes it suffices that the conditional probabilities of large changes are small. One key aspect of our approach is that we can always guarantee this via the 'global' event Γ only, that is, *without* having any knowledge about the corresponding conditional distributions.

2.2.1. The general approach. We now introduce the set-up used in all subsequent proofs. We consider the increasing sequence of sub-σ-fields $\mathcal{F}_k$ generated by $X_1, \dots, X_k$. Using Doob's construction, the sequence $Y_k = \mathbb{E}(f(X) \mid \mathcal{F}_k)$ is a martingale with $Y_0 = \mathbb{E}f(X)$ and $Y_N = f(X)$. Now, for each $k \in [N]$, we define $\mathcal{B}_{k-1}$ as the $\mathcal{F}_{k-1}$-measurable event that

$$\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}) > \gamma_k \tag{2.6}$$

holds. Let $\mathcal{B} = \neg\Gamma \cup \bigcup_{k\in[N]} \mathcal{B}_{k-1}$. Note that $\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_0) = \mathbb{P}(X \notin \Gamma)$ yields $\mathbb{P}(\mathcal{B}_0) = 0$ if $\gamma_1 \geqslant \mathbb{P}(X \notin \Gamma)$ and $\mathbb{P}(\mathcal{B}_0) = 1$ otherwise. Using $\gamma_1 \in (0, 1]$ we infer

$$\mathbb{P}(\neg\Gamma \cup \mathcal{B}_0) \leqslant \gamma_1^{-1}\mathbb{P}(X \notin \Gamma).$$

Observing that

$$\mathbb{P}(X \notin \Gamma) = \mathbb{E}(\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1})) \geqslant \gamma_k \mathbb{P}(\mathcal{B}_{k-1}),$$

the union bound now gives

$$\mathbb{P}(\mathcal{B}) \leqslant \mathbb{P}(\neg\Gamma \cup \mathcal{B}_0) + \sum_{2\leqslant k\leqslant N} \mathbb{P}(\mathcal{B}_{k-1}) \leqslant \sum_{k\in[N]} \gamma_k^{-1} \cdot \mathbb{P}(X \notin \Gamma). \tag{2.7}$$

Let the stopping time T be the minimum of N and the smallest $0 \leqslant k < N$ for which $\mathcal{B}_k$ holds (note that $T \leqslant k-1$ is $\mathcal{F}_{k-1}$-measurable). Setting $M_k = Y_{k\wedge T}$, it follows that the sequence $(M_k)_{0\leqslant k\leqslant N}$ is a martingale with $Y_0 = M_0$. Since $\neg\mathcal{B}$ implies $T = N$, recalling $Y_N = f(X)$ we see that

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) = \mathbb{P}(Y_N \geqslant Y_0 + t \text{ and } \neg\mathcal{B}) \leqslant \mathbb{P}(M_N \geqslant M_0 + t). \tag{2.8}$$

It remains to establish suitable tail estimates for $\mathbb{P}(M_N \geqslant M_0 + t)$, and via Lemmas 2.1 and 2.2 this reduces to proving (deterministic) upper bounds on the random variables S_N,

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

V_N and C_N. To this end we consider the martingale difference sequences $\Delta M_k = M_k - M_{k-1}$ and $\Delta Y_k = Y_k - Y_{k-1}$, which satisfy $\mathbb{E}(\Delta M_k \mid \mathcal{F}_{k-1}) = 0$ and $\mathbb{E}(\Delta Y_k \mid \mathcal{F}_{k-1}) = 0$. Set $e_k = \gamma_k(d_k - c_k)$ and $\Delta_k = c_k + e_k$.

Proof of Theorem 1.2. It suffices to show $\Delta M_k \in [-\Delta_k, \Delta_k]$ for each $k \in [N]$: then the claim follows by applying Lemma 2.1 with $S = \sum_{k\in[N]}(2\Delta_k)^2$. The following argument is written with an eye on the upcoming proofs (where various modifications are needed). Note that $\Delta M_k = 0$ if $T \leqslant k-1$ and $\Delta M_k = \Delta Y_k$ if $T \geqslant k$. So it is enough to prove that $|\Delta Y_k| \leqslant \Delta_k$ whenever $T \geqslant k$. For brevity, for $z \in \Lambda_k$ and $y = (y_{k+1}, \dots, y_N) \in \prod_{k<j\leqslant N} \Lambda_j$ we write $f_y(z)$ for $f(X_1, \dots, X_{k-1}, z, y_{k+1}, \dots, y_N)$. Note that

$$\begin{aligned}\mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = a) \\ = \sum_{y_{k+1},\dots,y_N} f_y(a)\, \mathbb{P}(X_{k+1} = y_{k+1}, \dots, X_N = y_N \mid \mathcal{F}_{k-1}, X_k = a).\end{aligned}$$

Defining $|\Delta Y_k(a,b)|$ via the next equation, since $X_1, \dots, X_N$ are independent it follows that

$$\begin{aligned}|\Delta Y_k(a,b)| &= |\mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = a) - \mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = b)| \\ &\leqslant \sum_{y_{k+1},\dots,y_N} |f_y(a) - f_y(b)|\, \mathbb{P}(X_{k+1} = y_{k+1}, \dots, X_N = y_N \mid \mathcal{F}_{k-1}, X_k = a).\end{aligned} \tag{2.9}$$

By distinguishing between $X \in \Gamma$ and $X \notin \Gamma$, each time applying (1.4) as appropriate, we infer

$$\begin{aligned}|\Delta Y_k(a,b)| &\leqslant c_k \mathbb{P}(X \in \Gamma \mid \mathcal{F}_{k-1}, X_k = a) + d_k \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = a) \\ &= c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = a).\end{aligned} \tag{2.10}$$

Recall that $\mathcal{B}_{k-1}$ fails if $T \geqslant k$. Using $\sum_{y_k} \mathbb{P}(X_k = y_k \mid \mathcal{F}_{k-1}) = 1$ together with (2.10) and (2.6), for $T \geqslant k$ we deduce

$$\begin{aligned}|\Delta Y_k| &= |\mathbb{E}(f(X) \mid \mathcal{F}_{k-1}) - \mathbb{E}(f(X) \mid \mathcal{F}_k)| \leqslant \sum_{y_k} |\Delta Y_k(y_k, X_k)|\, \mathbb{P}(X_k = y_k \mid \mathcal{F}_{k-1}) \\ &\leqslant c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}) \leqslant c_k + \gamma_k(d_k - c_k) = \Delta_k.\end{aligned} \tag{2.11}$$

As explained, this completes the proof. □

Here we could have used the classical Hoeffding–Azuma inequality [24, 3] since the proof yields (deterministic) bounds for each individual ΔM_k. We decided to apply Lemma 2.1 since the forthcoming modifications needed for the 'dynamic exposure' of Section 1.1.3 do use its full strength, *i.e.*, that accumulative estimates of the ΔM_k suffice.

Proof of Remark 2. Following the approach of McDiarmid [34], we now modify the proof of Theorem 1.2 whenever X_k takes only two values, say, $\Lambda_k = \{0, 1\}$. We focus on the relevant case $T \geqslant k$, where $\Delta M_k = \Delta Y_k$. Define L_k and U_k as the minimum and maximum of

$$\mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = z) - \mathbb{E}(f(X) \mid \mathcal{F}_{k-1}), \quad \text{for } z \in \{0, 1\}.$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Clearly L_k and U_k are $\mathcal{F}_{k-1}$-measurable and satisfy $L_k \leqslant \Delta Y_k \leqslant U_k$. The key observation is that, using $T \geqslant k$, there exists an $\mathcal{F}_{k-1}$-measurable $\alpha \in \{0,1\}$ satisfying

$$\gamma_k \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}) \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = \alpha). \tag{2.12}$$

So, since $X_k \in \{0,1\}$ takes only two values, using (2.10) and (2.12) we infer for $T \geqslant k$ that

$$|U_k - L_k| = |\Delta Y_k(\alpha, 1-\alpha)| \leqslant c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = \alpha) \leqslant \Delta_k. \tag{2.13}$$

This completes the proof (by applying Lemma 2.1 with $S = \sum_{k\in[N]} \Delta_k^2$). □

In fact, Theorem 1.1 follows by a similar modification ((2.10) implies $\max_{a,b} |\Delta Y_k(a,b)| \leqslant c_k$).

Proof of Theorem 1.3. In the proof of Theorem 1.2 we established $|\Delta M_k| \leqslant \Delta_k$ for every $k \in [N]$. In view of this it suffices to show

$$\mathrm{Var}(\Delta M_k \mid \mathcal{F}_{k-1}) \leqslant (1-p_k)p_k\Delta_k^2$$

for each $k \in [N]$: then the claim follows by applying Lemma 2.2 with

$$V = \sum_{k\in[N]} (1-p_k)p_k\Delta_k^2 \quad \text{and} \quad C = \max_{k\in[N]} \Delta_k.$$

Observe that $\mathbb{E}(\Delta M_k \mid \mathcal{F}_{k-1}) = 0$ implies

$$\mathrm{Var}(\Delta M_k \mid \mathcal{F}_{k-1}) = \mathbb{E}(\Delta M_k^2 \mid \mathcal{F}_{k-1}).$$

Recall that $\Delta M_k = 0$ if $T \leqslant k-1$ and $\Delta M_k = \Delta Y_k$ if $T \geqslant k$. Combining these facts, it is enough to prove that $\mathbb{E}(\Delta Y_k^2 \mid \mathcal{F}_{k-1}) \leqslant (1-p_k)p_k\Delta_k^2$ whenever $T \geqslant k$. Set

$$D_k = \mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = 1) - \mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = 0),$$

and observe that

$$\Delta Y_k = D_k\mathbb{P}(X_k = 0 \mid \mathcal{F}_{k-1})\mathbb{1}_{\{X_k=1\}} - D_k\mathbb{P}(X_k = 1 \mid \mathcal{F}_{k-1})\mathbb{1}_{\{X_k=0\}}.$$

Hence

$$|\Delta Y_k| \leqslant |D_k| \sum_{\beta\in\{0,1\}} \mathbb{P}(X_k = \beta \mid \mathcal{F}_{k-1})\mathbb{1}_{\{X_k=1-\beta\}}. \tag{2.14}$$

Arguing as in (2.12) and (2.13), we readily obtain $|D_k| = |\Delta Y_k(1,0)| \leqslant \Delta_k$ when $T \geqslant k$, and thus infer

$$\Delta Y_k^2 \leqslant \Delta_k^2 \sum_{\beta\in\{0,1\}} \mathbb{P}(X_k = \beta \mid \mathcal{F}_{k-1})^2\mathbb{1}_{\{X_k=1-\beta\}}.$$

Using the independence of $X_1, \ldots, X_N$, it follows that for $T \geqslant k$ we have

$$\mathbb{E}(\Delta Y_k^2 \mid \mathcal{F}_{k-1}) \leqslant \Delta_k^2 \sum_{\beta\in\{0,1\}} \mathbb{P}(X_k = \beta)^2\mathbb{P}(X_k = 1-\beta) = (1-p_k)p_k\Delta_k^2, \tag{2.15}$$

where we used $\mathbb{P}(X_k = 1) = p_k$ and $(1-x)^2x + x^2(1-x) = (1-x)x$ for the last inequality. □

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Note that (2.14) implies $\Delta M_k \leqslant \max\{1-p_k, p_k\} \cdot \Delta_k$, but the resulting minor improvement of C usually has negligible effect.

Proof of Remark 6. In the more general situation where each X_k takes values in a set Λ_k and satisfies $\max_{\eta\in\Lambda_k} \mathbb{P}(X_k = \eta) \geqslant 1-p_k$, we first show that Theorem 1.3 holds after replacing $(1-p_k)p_k$ and $\Delta_k = c_k + e_k$ by p_k and $\tilde{\Delta}_k = c_k + e_k \cdot (1-p_k)^{-1}$, respectively. With the proof of Theorem 1.3 in mind it suffices to show $\mathrm{Var}(\Delta Y_k \mid \mathcal{F}_{k-1}) \leqslant p_k \tilde{\Delta}_k^2$ whenever $T \geqslant k$. For $\beta \in \Lambda_k$ satisfying $\mathbb{P}(X_k = \beta) \geqslant 1-p_k$ set $\tau_k = \mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = \beta)$ and $D_k = \mathbb{E}(f(X) \mid \mathcal{F}_k) - \tau_k$. Note that, using the independence of $X_1, \dots, X_N$, we have

$$\mathbb{P}(D_k \neq 0 \mid \mathcal{F}_{k-1}) \leqslant \mathbb{P}(X_k \neq \beta) \leqslant p_k.$$

We claim that it suffices to show $|D_k| \leqslant \tilde{\Delta}_k$. Indeed, since Y_{k-1} and τ_k are $\mathcal{F}_{k-1}$-measurable, we have

$$\mathrm{Var}(\Delta Y_k \mid \mathcal{F}_{k-1}) = \mathrm{Var}(D_k \mid \mathcal{F}_{k-1}) \leqslant \mathbb{E}(D_k^2 \mid \mathcal{F}_{k-1}) \leqslant \tilde{\Delta}_k^2 \mathbb{P}(D_k \neq 0 \mid \mathcal{F}_{k-1}) \leqslant p_k \tilde{\Delta}_k^2.$$

To bound $|D_k|$, first note that $T \geqslant k$ and independence of $X_1, \dots, X_N$ yields

$$\gamma_k \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}) \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = \beta)(1-p_k). \tag{2.16}$$

So, using (2.10) and (2.16), for $T \geqslant k$ we infer

$$\begin{aligned} |D_k| &= |\mathbb{E}(f(X) \mid \mathcal{F}_{k-1}, X_k = \beta) - \mathbb{E}(f(X) \mid \mathcal{F}_k)| = |\Delta Y_k(\beta, X_k)| \\ &\leqslant c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = \beta) \leqslant c_k + \gamma_k (1-p_k)^{-1} \cdot (d_k - c_k) = \tilde{\Delta}_k, \end{aligned} \tag{2.17}$$

establishing the claim.

A similar argument shows that Corollary 1.4 holds after replacing $(1-p_k)p_k$ by p_k. The point is that in Corollary 1.4 there is no 'good' event Γ. Consequently, when invoking (2.10) in (2.17) the standard line of reasoning (using (1.1) instead of (1.4)) yields $|D_k| \leqslant c_k$, and the claim follows. □

Proof of Theorem 1.5. The crux is that (2.10) and (2.11) are at the heart of all previous proofs. In the following we essentially exploit that both can be adapted when (1.12) instead of (1.4) holds: it suffices if $e_k = 2\gamma_k(d_k - c_k)q_k^{-1}$ is used, where $\min_{\eta\in\Lambda_k} \mathbb{P}(X_k = \eta) \geqslant q_k$.

We start by modifying the proof of Theorem 1.2. Analogous to (2.16), if $T \geqslant k$ then for all $\beta \in \Lambda_k$ we have

$$\gamma_k \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}) \geqslant \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = \beta) q_k.$$

The key point of (1.12) is that $|f(x) - f(\tilde{x})| \leqslant c_k$ only holds if $x, \tilde{x} \in \Gamma$. So, using (1.12) as appropriate, a short calculations reveals that, for $T \geqslant k$, the corresponding variant of (2.10) is

$$\begin{aligned} |\Delta Y_k(a,b)| &\leqslant c_k + (d_k - c_k)\big[\mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = a) + \mathbb{P}(X \notin \Gamma \mid \mathcal{F}_{k-1}, X_k = b)\big] \\ &\leqslant c_k + (d_k - c_k) \cdot 2\gamma_k q_k^{-1} = \tilde{\Delta}_k. \end{aligned} \tag{2.18}$$

Now, for $T \geqslant k$ we readily obtain $|\Delta Y_k| \leqslant \tilde{\Delta}_k$, which is the natural analogue of (2.11). As discussed, this establishes the claimed variant of Theorem 1.2.

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

In the proofs of Remark 2 and Theorem 1.3 we only need to adapt the bounds for $|U_k - L_k| = |\Delta Y_k(\alpha, 1-\alpha)|$ and $|D_k| = |\Delta Y_k(1,0)|$, which are both at most $\tilde{\Delta}_k$ by (2.18). Similarly, in the proof of Remark 6 it suffices to modify (2.17), where $|D_k| \leqslant \tilde{\Delta}_k$ follows from (2.18), completing the proof (in fact, our argument shows that $e_k(1-p)^{-1}$ can be replaced by e_k). □

2.2.2. Some extensions.

Proof of Remark 4. Note that $f(X) \geqslant \mathbb{E}f(X) + t$ is increasing (decreasing) if $f(X)$ is increasing (decreasing). Furthermore, in view of (2.6) it is easy to check that $\mathcal{B}_{k-1}$ is increasing (decreasing) if Γ is decreasing (increasing). Using the definition $\mathcal{B}$ and the assumptions of Remark 4, it follows that $f(X) \geqslant \mathbb{E}f(X) + t$ and $\neg\mathcal{B}$ are either both increasing or decreasing. So Harris's inequality [23] yields

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) \geqslant \mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t) \cdot \mathbb{P}(\neg\mathcal{B}),$$

which readily establishes (1.10). □

Proof of Theorem 1.6. The basic idea is to use a truncation that maps every $x_k \notin \Gamma_k$ to some fixed $z_k \in \Gamma_k$. As before, we work with the sub-σ-fields $\mathcal{F}_k$ generated by $X_1, \dots, X_k$. Recall that $\mathcal{B}$ is defined via (2.6) and satisfies $\neg\mathcal{B} \subseteq \Gamma$. Since

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t \text{ and } \neg\mathcal{B}) \leqslant \mathbb{P}(X \in \Gamma),$$

we may assume that $\mathbb{P}(X \in \Gamma) > 0$ and fix some

$$z = (z_1, \dots, z_N) \in \Gamma \subseteq \prod_{j \in [N]} \Gamma_j.$$

For $x = (x_1, \dots, x_N)$, we now define $x^* = (x_1^*, \dots, x_N^*)$ via

$$x_k^* = \begin{cases} x_k & \text{if } x_k \in \Gamma_k, \\ z_k & \text{if } x_k \notin \Gamma_k. \end{cases} \tag{2.19}$$

The key properties of this construction are (a) that $x \in \Gamma$ implies $x^* = x$, and (b) that $X^* = (X_1^*, \dots, X_N^*)$ is a family of independent random variables. Set $\mu = \mathbb{E}f(X)$ and $\mu^* = \mathbb{E}f(X^*)$. We have

$$|\mu - \mu^*| \leqslant \mathbb{E}|f(X) - f(X^*)| \leqslant s\mathbb{P}(X \notin \Gamma) = \Delta$$

(this is not best possible but keeps the formulas simple), and so $\neg\mathcal{B} \subseteq \Gamma$ yields

$$\mathbb{P}(f(X) \geqslant \mathbb{E}f(X) + t + \Delta \text{ and } \neg\mathcal{B}) \leqslant \mathbb{P}(f(X^*) \geqslant \mu^* + t \text{ and } \neg\mathcal{B}). \tag{2.20}$$

Now we estimate the right-hand side of (2.20) for $Y = f(X^*)$ as in the proof of Theorem 1.2 via (2.8), and there are only two minor differences. The first is that due to the projection (2.19) we have $X^* \in \prod_{j \in [N]} \Gamma_j$, so that the 'refined' Lipschitz coefficients c_k and d_k of Theorem 1.6 always apply. The second concerns the case distinction $X^* \in \Gamma$ and $X^* \notin \Gamma$. Here we use that $X^* \notin \Gamma$ implies $X \notin \Gamma$ pointwise, which yields $\mathbb{P}(X^* \notin \Gamma \mid \mathcal{F}_{k-1}) \leqslant \mathbb{P}(X \notin$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

$\Gamma \mid \mathcal{F}_{k-1}$). With this estimate the conclusion of (2.11) readily carries over, completing the proof. □

Here we could have estimated $f(X^*) \geqslant \mu^* + t$ via Theorem 1.2 (with Λ_k replaced by Γ_k), using that $\mathbb{P}(X^* \notin \Gamma) \leqslant \mathbb{P}(X \notin \Gamma)$. The advantage of our more pedestrian approach is that it uses the same 'bad' event $\mathcal{B}$ as all other proofs (by applying Theorem 1.2 it would depend on $z = (z_1, \dots, z_k)$).

Proof of Remark 8. The monotonicity property implies $\mathbb{E}f(X) \geqslant \mu^*$ (writing $f(X) - f(X^*)$ as a difference sequence of coordinate changes), so $\Delta = 0$ suffices using $\mathbb{E}f(X) + t \geqslant \mu^* + t$ in (2.20). Turning to the special case $\Gamma = \prod_{j\in[N]} \Gamma_j$, note that $\mathcal{B} = \neg\Gamma$ suffices to establish (2.20). Now, since $X^* = (X_1^*, \dots, X_N^*)$ is a family of independent random variables with $X_k^* \in \Gamma_k$ satisfying (L), the claimed variant readily follows from Theorem 1.1. □

2.2.3. Variants using dynamic exposure. In the following we briefly sketch how to modify the proofs of Sections 2.2.1 and 2.2.2 when the variables are exposed in a dynamic order, which will eventually establish Theorems 1.7 and 1.8 (in contrast to [2], our approach is based on general martingale inequalities). Recall that the strategies introduced in Section 1.1.3 sequentially expose $X_{q_1}, X_{q_2}, \dots$ with $q_i = q_i(X_{q_1}, \dots, X_{q_{i-1}})$. Furthermore, we use the convention that $q_{k+1} = q_k$ if $f(X)$ is determined by $(X_1, \dots, X_{q_k})$ with $k < N$. For technical reasons we slightly modify these strategies so that (always) all variables are queried. More precisely, for the proof of Theorem 1.7 we set $\tilde{q}_k = q_k$ until $f(X)$ is determined by $(X_1, \dots, X_{q_k})$; afterwards $\tilde{q}_{k+1}, \dots, \tilde{q}_N$ equals the remaining 'useless' indices $[N] \setminus \{q_1, \dots, q_k\}$ in ascending order, say (for the proof of Theorem 1.8 we simply use the fixed order $\tilde{q}_k = k$ for all $k \in [N]$). We consider an increasing sequence of sub-σ-fields, where $\mathcal{F}_k$ is generated by $X_{\tilde{q}_1}, \dots, X_{\tilde{q}_k}$. Note that each index $\tilde{q}_k$ is $\mathcal{F}_{k-1}$-measurable. Furthermore, our modification ensures that the following two key properties hold: X is $\mathcal{F}_N$-measurable (this is needed to apply Γ since the value of $f(X)$ may not uniquely determine X), and conditional on $\mathcal{F}_{k-1}$ all X_j with $j \notin \{\tilde{q}_1, \dots, \tilde{q}_{k-1}\}$ are independent random variables. Define $R = \max_{Q\in\mathcal{Q}} |Q|$ and $\mathcal{B} = \neg\Gamma \cup \bigcup_{k\in[R]} \mathcal{B}_{k-1}$ (for Theorem 1.8 we set $R = N$, so $\mathcal{B}$ remains unchanged). Since the definition of $\mathcal{B}_{k-1}$ via (2.6) involves $\mathcal{F}_{k-1}$, it follows that $\mathcal{B}$ depends on the query strategy (unless the fixed order $\tilde{q}_k = k$ is used, as in the proof of Theorem 1.8).

With these changes in mind, all arguments of Sections 2.2.1 and 2.2.2 essentially carry over word by word, the only exception being the proof of Remark 4 (the monotonicity argument needs a fixed order such as $\tilde{q}_k = k$). The crucial observation is that for every variable X_ℓ not queried by the original strategy we know that its value will *not* change the outcome of $f(X)$. To be more formal, the key point is that whenever such a 'useless' variable is queried in step i we have $\Delta M_i = 0$ (note that for $i > R$ this is always the case), that is, the indices of these variables do not contribute to S_N, V_N or C_N. Observe that due to the dynamic exposure we only have a connection between γ_k and the index $\tilde{q}_k$ (not index k, as before). We overcome this minor complication using the assumption that $\gamma_k = \gamma$ for all $k \in [N]$ (we may allow for different γ_k if $\tilde{q}_k = k$ is used), which also ensures

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

that

$$\max_{Q\in\mathcal{Q}} \sum_{k\in Q} \gamma_k^{-1} = \sum_{k\in[R]} \gamma_k^{-1}$$

holds (for $\tilde{q}_k = k$ the estimate (2.7) stays unchanged).

The remaining details for establishing Theorem 1.7 and 1.8 are rather straightforward: when invoking the martingale estimates we simply take the 'worst case' bounds for S, V and C over all possible sets of queried indices $Q \in \mathcal{Q}$ (where Q and $\mathcal{Q}$ are as defined in Section 1.1.3); for example, using $S = \max_{Q\in\mathcal{Q}} \sum_{k\in Q} (2\Delta_k)^2$ for Theorem 1.2. It is this last step where the accumulative random bounds in Lemmas 2.1 and 2.2 are crucial (the behaviour of each individual ΔM_k may vary significantly for different sample points due to the dynamic order in which the variables are queried).

2.2.4. Variants using the general Lipschitz condition. Finally, we discuss how to modify the proofs in Section 2.2.1 when the independence assumption is replaced by (GL). We first claim that $\rho_k = \rho_k(\Sigma_a, \Sigma_b) : \Sigma_a \to \Sigma_b$ is a bijection with equality in (1.15), that is, it satisfies

$$\mathbb{P}(X = x \mid X \in \Sigma_a) = \mathbb{P}(X = \rho_k(x) \mid X \in \Sigma_b). \tag{2.21}$$

Indeed, using (1.15) and that ρ_k is injective, it follows that

$$\sum_{x\in\Sigma_a} \mathbb{P}(X = x \mid X \in \Sigma_a) \leqslant \sum_{x\in\Sigma_a} \mathbb{P}(X = \rho_k(x) \mid X \in \Sigma_b) \leqslant \sum_{x\in\Sigma_b} \mathbb{P}(X = x \mid X \in \Sigma_b).$$

Noting that

$$\sum_{x\in\Sigma_z} \mathbb{P}(X = x \mid X \in \Sigma_z) = 1$$

for $z \in \{a, b\}$ with $|\Sigma_z| > 0$, we infer that all inequalities are in fact satisfied as equations, which establishes (2.21). Since every $x \in \Sigma_z$ satisfies $\mathbb{P}(X = x) > 0$, it also follows that ρ_k must be a bijection, as claimed.

In preparation for our forthcoming arguments we now relate the definitions used in the proofs of Section 2.2.1 with those occurring in (GL). Analogous to Σ_z, given any possible sequence of outcomes $a_1, \dots, a_{k-1}$ of $X_1, \dots, X_{k-1}$, define Σ as the set of all

$$x = (a_1, \dots, a_{k-1}, x_k, \dots, x_N) \in \prod_{j\in[N]} \Lambda_j$$

with $\mathbb{P}(X = x) > 0$. Recall that $\mathcal{F}_k$ is the increasing sequence of sub-σ-fields generated by $X_1, \dots, X_k$. The key point is that $(\mathcal{F}_k)_{0\leqslant k\leqslant N}$ naturally corresponds to an increasing sequence of partitions of the sample space, where two points belong to the same part if and only if they agree on the first k coordinates. For example, for $\omega \in \Sigma$ we have

$$\mathbb{E}(\cdot \mid \mathcal{F}_{k-1}, X_k = z)(\omega) = \mathbb{E}(\cdot \mid X \in \Sigma_z) \quad \text{and} \quad \mathbb{P}(\cdot \mid \mathcal{F}_{k-1}, X_k = z)(\omega) = \mathbb{P}(\cdot \mid X \in \Sigma_z). \tag{2.22}$$

Proof of Theorem 1.9. We modify the proof of Theorem 1.2, where independence is only used to establish (2.9). Using (2.22) and that the bijection $\rho_k : \Sigma_a \to \Sigma_b$ satisfies (2.21), we

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

obtain

$$\begin{aligned}|\Delta Y_k(a,b)| &= |\mathbb{E}(f(X) \mid X \in \Sigma_a) - \mathbb{E}(f(X) \mid X \in \Sigma_b)| \\ &= \left| \sum_{x \in \Sigma_a} f(x)\mathbb{P}(X = x \mid X \in \Sigma_a) - \sum_{x \in \Sigma_b} f(x)\mathbb{P}(X = x \mid X \in \Sigma_b) \right| \\ &\leqslant \sum_{x \in \Sigma_a} |f(x) - f(\rho_k(x))| \mathbb{P}(X = x \mid X \in \Sigma_a),\end{aligned}$$

which is the natural analogue of (2.9). The remainder of the argument carries over with minor modifications. Indeed, proceeding as in (2.10) (applying (1.14) instead of (1.4)) and then appealing to (2.22), we infer

$$|\Delta Y_k(a,b)| \leqslant c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid X \in \mathcal{F}_{k-1}, X_k = a). \tag{2.23}$$

Now, by arguing as in (2.11), when $T \geqslant k$ holds we also have

$$|\Delta Y_k| \leqslant c_k + (d_k - c_k)\mathbb{P}(X \notin \Gamma \mid X \in \mathcal{F}_{k-1}) \leqslant \Delta_k, \tag{2.24}$$

completing the proof. □

Remark 9 follows by similar reasoning (noting that the proof of Remark 2 carries over and that (2.23) equals (1.16) after replacing $(d_k - c_k)$ with r_k).

Proof of Theorem 1.10. We modify the proof of Remark 6 by means of choosing (some) $\mathcal{F}_{k-1}$-measurable $\beta \in \Lambda_k$ maximizing $\mathbb{P}(X_k = \beta \mid \mathcal{F}_{k-1})$. By assumption we have $\mathbb{P}(X_k = \beta \mid \mathcal{F}_{k-1}) \geqslant 1 - p_k$, which in turn yields $\mathbb{P}(D_k \neq 0 \mid \mathcal{F}_{k-1}) \leqslant \mathbb{P}(X_k \neq \beta \mid \mathcal{F}_{k-1}) \leqslant p_k$. Noting that all remaining applications of independence are already covered by (2.23) and (2.24), this completes the proof. □

Proof of Theorem 1.11. With the above modifications in mind, the proof of Theorem 1.5 carries over word by word, which establishes the first part of the claim. Turning to the second part, our earlier discussion shows that for $\Sigma_z, \Sigma_\eta \subseteq \Sigma$ with $|\Sigma_z|, |\Sigma_\eta| > 0$ there is a bijection $\rho_k : \Sigma_z \to \Sigma_\eta$. So, since all possible outcomes occur with the same probability, we obtain

$$\mathbb{P}(X \in \Sigma_z) = \mathbb{P}(X \in \Sigma_\eta).$$

For $\eta \in \Lambda_k$ satisfying $|\Sigma_\eta| > 0$, it follows that

$$\frac{1}{\mathbb{P}(X_k = \eta \mid X \in \Sigma)} = \frac{\mathbb{P}(X \in \Sigma)}{\mathbb{P}(X \in \Sigma_\eta)} = \sum_{z \in \Lambda_k} \frac{\mathbb{P}(X \in \Sigma_z)}{\mathbb{P}(X \in \Sigma_\eta)} \leqslant |\Lambda_k|.$$

We deduce that

$$\min_{\eta \in \Lambda_k} \mathbb{P}(X_k = \eta \mid X_1, \ldots, X_{k-1}) \geqslant |\Lambda_k|^{-1},$$

so $q_k \leqslant |\Lambda_k|^{-1}$ suffices. □

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

3. Final number of edges in the reverse H-free process

In our analysis of the reverse H-free process we use several equivalent definitions (with respect to the final graph). Recall that, starting with the complete graph on vertex set $[n]$, in each step an edge is removed, chosen uniformly at random from all edges contained in a copy of H. As in [19, 33], a moment's thought reveals that we may instead traverse all $\binom{n}{2}$ edges in random order, each time removing the current edge if and only if it is contained in a copy of H in the evolving graph. As observed by Erdős, Suen and Winkler [19], after considering $e_{\binom{n}{2}}, \dots, e_{i+1}$ the decision whether e_i is removed depends *only* on the later edges $e_{i-1}, \dots, e_1$ (all other 'surviving' ones are by construction not contained in a copy of H). This allows us to consider the edges in reverse order, where e_i is added if and only if it does not complete a copy of H together with $e_1, \dots, e_{i-1}$ (it does not matter whether these were added or not). Given a random permutation, we denote the corresponding random graph process after i steps by $G_{n,i}(H) \subseteq G_{n,i}$, where $G_{n,i}$ is the uniform random graph with n vertices and i edges.

For technical reasons it will be convenient to also consider a continuous variant of the above process, where each edge is independently assigned a uniform birth time $B_e \in [0,1]$; the edges are then traversed in ascending order of their birth times (which are all distinct with probability one). The resulting process that considers only those edges with $B_e \leqslant p$ is denoted by $G_{n,p}(H) \subseteq G_{n,p}$. So for $p = 1$ all edges are traversed in random order, and it follows that

$$G_{n,\binom{n}{2}}(H) = G_{n,1}(H). \tag{3.1}$$

Conditioned on $B_e = q$, the decision whether e is added only depends on the edges f with $B_f \leqslant q$, which have the same distribution as $G_{n,q}$. As noted by Makai [33], this allows for the use of classical random graph theory when estimating the probability that an edge is added to the evolving graph. Recall that $m_2(H) = d_2(H)$ for 2-balanced graphs H. For

$$m = n^{2-1/m_2(H)}(\log n)^2 \quad \text{and} \quad p = n^{-1/m_2(H)}(\log n)^2,$$

the next lemma follows from the results of Spencer [45] mentioned in Section 1.2.2. Note that in $G_{n,p}$ every pair of vertices is expected to have $\Theta((\log n)^{2(e_H-1)})$ 'extensions' to copies of H.

Lemma 3.1. *Let H be a 2-balanced graph. Let $\mathcal{D}$ (respectively, $\mathcal{I}$) denote the event that for every pair xy of vertices the following holds: after adding the edge xy there are at most $\Psi_H = (\log n)^{2e_H}$ copies (respectively, is at least one copy) of H containing the edge xy. For every $c > 0$ and $n \geqslant n_0(c,H)$ we have $\mathbb{P}(G_{n,m} \in \mathcal{D} \cap \mathcal{I}) \geqslant 1 - n^{-c}$.* □

The point is that whenever $\mathcal{I}$ holds no further edges are added. By combining Lemma 3.1 with (3.1) we thus arrive at the following lemma, which conveniently relates both variants of the reverse H-free process.

Lemma 3.2. *Let H be a 2-balanced graph. For every $c > 0$ and $n \geqslant n_0(c,H)$ we have*

$$\mathbb{P}\big(G_{n,m}(H) = G_{n,\binom{n}{2}}(H) = G_{n,1}(H)\big) \geqslant 1 - n^{-c}.$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Turning to the number of edges in $G_{n,m}(H)$, which we denote by $e(G_{n,m}(H))$, recall that each e_i is added if and only if it does not complete a copy of H together with $e_1,\dots,e_{i-1}$. So one edge can, in the worst case, influence the decisions of up to $O(\min\{m, n^{v_H-2}\})$ edges (whether they are added or not); however, on the 'typical' event $\mathcal{D}$ of Lemma 3.1 this is limited to at most $e_H \cdot \Psi_H = O((\log n)^{2e_H})$ edges. For this reason the standard bounded differences inequality *fails* to give useful bounds (due to large worst case c_k), whereas a *routine* application of the typical bounded differences inequality yields sharp concentration, illustrating its ease of use and effectiveness.

Theorem 3.3. *Let H be a 2-balanced graph. For every $c > 0$ and $n \geqslant n_0(c,H)$ we have*

$$\mathbb{P}\big(|e(G_{n,m}(H)) - \mathbb{E}e(G_{n,m}(H))| \geqslant \sqrt{m}(\log n)^{3e_H}\big) \leqslant n^{-c}. \tag{3.2}$$

Proof. Lemma 3.1 implies that $G_{n,m} \in \mathcal{D}$ holds with probability at least $1 - n^{-(2c+6)}$. Note that the random sequence of edges $\underline{e} = (e_1,\dots,e_m)$ corresponds to the (uniform) random graph process $(G_{n,i})_{0\leqslant i\leqslant m}$ and uniquely determines $f(\underline{e}) = e(G_{n,m}(H))$. The crucial observation is that whenever $G_{n,m}, \tilde{G}_{n,m} \in \mathcal{D}$ have edge sequences $\underline{e}, \underline{\tilde{e}}$ that differ only in one edge (*i.e.*, $e_j \neq \tilde{e}_j$) or the order of two edges (*i.e.*, $e_j = \tilde{e}_k$ and $e_k = \tilde{e}_j$), then our earlier observations imply

$$|e(G_{n,m}(H)) - e(\tilde{G}_{n,m}(H))| \leqslant 2e_H(\log n)^{2e_H} = \Delta.$$

The point is that by the discussion of Section 1.1.4 this is exactly the condition that needs to be checked in order to apply Theorem 1.9 (with the two-sided Lipschitz condition (1.18) of Theorem 1.11) using $N = m$, the 'good' event $\Gamma = \mathcal{D}$, Lipschitz coefficients $c_k = \Delta$, $d_k = n^2$ and the 'two-sided parameter' $q_k = n^{-2}$. For the 'compensation factor' $\gamma_k = n^{-4}$ we have

$$e_k \leqslant 2\gamma_k d_k q_k^{-1} \leqslant 2, \quad c_k + e_k = \Theta((\log n)^{2e_H}) \quad \text{and} \quad \sum_k \gamma_k^{-1} \leqslant n^6.$$

So, using (1.6) and (1.5) we deduce that the left-hand side of (3.2) is at most

$$e^{-\Omega((\log n)^{2e_H})} + n^{-2c} \leqslant n^{-c}. \qquad \square$$

To establish Theorem 1.12 it remains to bound the expected final number of edges up to constant factors (note that all 2-balanced graphs H satisfy $e_H \geqslant 2$; see Section 1.2.2). Our argument is inspired by Makai [33], who proved asymptotically matching bounds in (3.3) for the class of strictly 2-balanced graphs (the case $H = K_3$ is due to Erdős, Suen and Winkler [19]). In fact, here we determine the correct order of magnitude for all graphs (since $e(G_{n,1}(H)) = 0$ when $e_H = 1$).

Theorem 3.4. *Let H be a graph with $e_H \geqslant 2$. There are $a, A > 0$ such that for $n \geqslant n_0(H)$ we have*

$$an^{2-1/m_2(H)} \leqslant \mathbb{E}e(G_{n,1}(H)) \leqslant An^{2-1/m_2(H)}. \tag{3.3}$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Proof. Define $\mathcal{Z}_e = \mathcal{Z}_e(H)$ as the event that the edge e is contained in $G_{n,1}(H)$. Let $Y_{e,H,q}$ count the number of copies of H in $G_{n,q} \cup e$ (the graph obtained by inserting e into $G_{n,q}$ if it is not already present). Recall that, conditioned on $B_e = q$, only edges f with $B_f \leqslant q$ are relevant for $\mathcal{Z}_e$; so e is added if and only if $Y_{e,H,q} = 0$. Hence for $q \in [0,1]$ we have

$$\mathbb{P}(\mathcal{Z}_e \mid B_e = q) = \mathbb{P}(Y_{e,H,q} = 0). \tag{3.4}$$

For the lower bound in (3.3) fix $F \subseteq H$ with $d_2(F) = m_2(H)$ and $e_F \geqslant 2$ (this choice is possible as $e_H \geqslant 2$). Given an edge e there are at most Dn^{v_F-2} extensions to F for some $D = D(F) > 0$, so whenever $q \leqslant n^{-1/m_2(H)}$ holds, monotonicity and Harris's inequality [23] yield

$$\mathbb{P}(Y_{e,H,q} = 0) \geqslant \mathbb{P}(Y_{e,F,q} = 0) \geqslant (1-q^{e_F-1})^{Dn^{v_F-2}} \geqslant e^{-2Dn^{v_F-2}q^{e_F-1}} \geqslant e^{-2D}. \tag{3.5}$$

Together with (3.4), we obtain

$$\mathbb{P}(\mathcal{Z}_e) = \mathbb{E}\,\mathbb{P}(\mathcal{Z}_e \mid B_e = q) \geqslant n^{-1/m_2(H)} \cdot e^{-2D},$$

and the lower bound in (3.3) now follows by linearity of expectation.

Turning to the upper bound in (3.3), consider $q = \lambda n^{-1/m_2(H)}$ with $1 \leqslant \lambda \leqslant n^{1/m_2(H)}$. We apply Janson's inequality to $Y_{e,H,q}$, which intuitively counts 'extensions' of e to H (viewed as subgraphs, these do not contain the edge e). Note that $e_H \geqslant 2$, $\lambda \geqslant 1$ and $m_2(H) \geqslant (e_H-1)/(v_H-2)$ imply

$$\mu = \mathbb{E}Y_{e,H,q} = \Theta(n^{v_H-2}q^{e_H-1}) = n^{(v_H-2)-(e_H-1)/m_2(H)}\Theta(\lambda^{e_H-1}) = \Omega(\lambda).$$

Define $\mathcal{G}$ as the set of all proper subgraphs $G \subsetneq H$ with $e_G \geqslant 2$. Considering all possible 'overlaps' of extensions of e to H (note that only overlaps distinct from e are relevant, which explains the $e_G \geqslant 2$ condition), the Δ-term of Janson's inequality (see *e.g.* Theorem 2.18 in [26]) satisfies

$$\begin{aligned}\Delta &\leqslant O(n^{v_H-2}q^{e_H-1}) \cdot \sum_{G\in\mathcal{G}} O(n^{v_H-v_G}q^{e_H-e_G}) = O(\mu^2)\sum_{G\in\mathcal{G}} n^{-(v_G-2)}q^{-(e_G-1)} \\ &= O(\mu^2)\sum_{G\in\mathcal{G}} n^{-[(v_G-2)-(e_G-1)/m_2(H)]}\lambda^{-(e_G-1)} = O(\mu^2/\lambda),\end{aligned}$$

where the last inequality follows from $e_G \geqslant 2$, $\lambda \geqslant 1$ and $m_2(H) \geqslant (e_G-1)/(v_G-2)$. So, using $\mu/\lambda = \Omega(1)$ we infer $\mu + \Delta = O(\mu^2/\lambda)$ and thus

$$\mu^2/(\mu+2\Delta) = \Omega(\lambda) = \Omega(n^{1/m_2(H)}q).$$

Applying Janson's inequality we have

$$\mathbb{P}(Y_{e,H,q} = 0) \leqslant \exp(-Cn^{1/m_2(H)}q) \quad \text{for } C = C(H) > 0.$$

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Combining this with (3.4) when $q \geqslant n^{-1/m_2(H)}$ and the trivial bound $\mathbb{P}(\mathcal{Z}_e \mid B_e = q) \leqslant 1$ otherwise, for $A = 1 + e^{-C}/C$ we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{Z}_e) &= \mathbb{E}\,\mathbb{P}(\mathcal{Z}_e \mid B_e = q) \\ &\leqslant n^{-1/m_2(H)} + \int_{n^{-1/m_2(H)}}^{1} \exp(-Cn^{1/m_2(H)}q)dq \leqslant An^{-1/m_2(H)}.\end{aligned} \tag{3.6}$$

Linearity of expectation now yields the upper bound in (3.3). □

Our arguments partially generalize to arbitrary graphs with $e_H \geqslant 2$, which we shall now briefly discuss. In this case Lemma 3.1 remains true if we modify $\mathcal{D}$ to at most, say, $\Psi_H = (\log n)n^{v_H-2}p^{e_H-1}$ copies, and so the argument leading to Lemma 3.2 carries over (it only uses $\mathcal{I}$). With (3.1) in mind, Theorem 3.4 shows that the expected final number of edges is $\mu = \Theta(n^{2-1/m_2(H)})$. Adjusting the proof of Theorem 3.3 with $c_k = 2e_H\Psi_H$, a short calculation shows that we obtain concentration on an interval of length $\mu n^{-\gamma}$ with $\gamma = \gamma(H) > 0$ whenever

$$v_H \geqslant 4 \text{ and } m_2(H) < (2e_H - 3)/(2v_H - 6) \quad \text{or} \quad v_H = 3 \text{ and } e_H \geqslant 2. \tag{3.7}$$

Perhaps surprisingly, this condition is satisfied by standard examples of 'unbalanced' graphs such as a clique K_r with an extra edge hanging off.

The proofs in this section also extend with minor modifications to the more general reverse $\mathcal{H}$-free process considered in Theorem 1.13. In this case the 'inverted' processes $\mathcal{G}_{n,m}(\mathcal{H})$ and $\mathcal{G}_{n,p}(\mathcal{H})$ are defined in analogous ways, where an edge is added only when it closes no copy of some $F \in \mathcal{H}$. We need to modify $\mathcal{D}$ of Lemma 3.1 so that for all $F \in \mathcal{H}$ it ensures at most $\Psi_F = \max\{(\log n)n^{v_F-2}p^{e_F-1}, (\log n)^{2e_F}\}$ copies, whereas the corresponding $\mathcal{I}$ only applies to the distinguished graph H with $m_2(H) = d_2(H) = \min_{F\in\mathcal{H}} d_2(F)$. As before, once $\mathcal{I}$ holds no more edges are added. With this in mind, Lemma 3.2 as well as the concentration result of Theorem 3.3 carry over in a straightforward way (noting that $d_2(H) \leqslant d_2(F)$ implies $\Psi_F \leqslant (\log n)^{2e_F}$ for all $F \in \mathcal{H}$). Turning to the expected final number of edges, for the lower bound of Theorem 3.4 we avoid all $F \in \mathcal{H}$ simultaneously. The resulting modification of (3.5) works for $q \leqslant n^{-1/m_2(H)}$ since $d_2(H) \leqslant d_2(F)$ implies $n^{v_F-2}q^{e_F-1} \leqslant 1$. For the upper bound it suffices to just avoid the distinguished 2-balanced graph H, so we may re-use the estimates of (3.6) to establish Theorem 1.13.

Finally, note that every edge added by $G_{n,m}(H)$ is also added by the H-free process defined in Section 1.2.3 (where e_i is added if and only if it does not complete a copy of H together with the *added* edges among $e_1, \dots, e_{i-1}$). It follows from Theorem 3.4 that the expected final number of edges in the H-free process is at least $\Omega(n^{2-1/m_2(H)})$ for any graph H with $e_H \geqslant 2$, which improves the $\Omega(n^{2-1/d_2(H)})$ bound resulting from the deletion argument of Osthus and Taraz [38]. In fact, if the technical conditions in (3.7) are satisfied, our earlier discussion implies that this lower bound also holds with probability tending to one (not only in expectation), which for 'unbalanced' graphs with $m_2(H) > d_2(H)$ does not follow from Theorem 1 in [38].

https://doi.org/10.1017/S0963548315000103 Published online by Cambridge University Press

Acknowledgements

I am grateful to Oliver Riordan and the referees for a very careful reading of the paper, and for many helpful suggestions concerning the presentation. I would also like to thank Tamás Makai for sending me a preprint of [33], Matas Šileikis for useful remarks, and Colin McDiarmid for asking whether Theorem 1.2 extends to random permutations.

Appendix

It is natural to ask whether Theorem 1.1 and conditioning (or a union bound) might not yield a simpler form of Theorem 1.2. For concreteness, suppose that Γ implies (L) with $c_k = c$, that is, that $|f(x) - f(\tilde{x})| \leqslant c$ whenever $x, \tilde{x} \in \Gamma$. Let $Y = f(X)$ and $\mu = \mathbb{E}Y$. One might hope that the bounded differences inequality applies with Lipschitz coefficient c unless Γ fails, that is,

$$\Pr(Y \leqslant \mu - t) \leqslant \Phi(c, t) + \Pr(X \notin \Gamma)$$

with $\Phi(c, t) = \exp(-2t^2/(Nc^2))$. Similarly,

$$\Pr(Y \leqslant \mu - t \mid \Gamma) \leqslant \Phi(c, t) \quad \text{or} \quad \Pr(Y \leqslant \mu - t \text{ and } \Gamma) \leqslant \Phi(c, t)$$

might seem plausible. However, a simple example shows that these naive guesses can fail badly. Let $Y = f(X) = \sum_{k \in [N]} X_k$, where $\Pr(X_k = 0) = 1 - 1/N$ with $X_N \in \{0, N^3\}$ and $X_k \in \{0, 1\}$ otherwise. Note that $\Gamma = \{X_N = 0\}$ implies (L) with $c_k = 1$. Since $\mu = \Theta(N^2)$, we see that $\Pr(Y \leqslant \mu/2 \mid \Gamma) = 1$ and $\Pr(Y \leqslant \mu/2 \text{ and } \Gamma) = 1/N$ are both substantially larger than $\Phi(1, \mu/2) = e^{-\Theta(N^3)}$, and that $\Pr(Y \leqslant \mu/2) = 1 - 1/n$ is also much larger than $\Phi(c, t) + \Pr(X \notin \Gamma) \approx \Pr(X \notin \Gamma) = 1/N$. (Here one issue is that Γ significantly shifts the expected value, *e.g.*, from $\mathbb{E}(Y) = \Theta(N^2)$ to $\mathbb{E}(Y \mid \Gamma) = \Theta(1)$.)

References

[1] Achlioptas, D. (2000) Setting 2 variables at a time yields a new lower bound for random 3-SAT. In *Proc. 32nd Annual ACM Symposium on Theory of Computing: STOC'00*, ACM, pp. 28–37.
[2] Alon, N., Kim, J. H. and Spencer, J. (1997) Nearly perfect matchings in regular simple hypergraphs. *Israel J. Math.* **100** 171–187.
[3] Azuma, K. (1967) Weighted sums of certain dependent random variables. *Tôhoku Math. J.* (2) **19** 357–367.
[4] Bapst, V., Coja-Oghlan, A., Hetterich, S., Rassmann, F. and Vilenchik, D. The condensation phase transition in random graph coloring. *Comm. Math. Phys.*, to appear (Proceedings version appeared in RANDOM 2014). `arXiv:1404.5513`.
[5] Bohman, T. (2009) The triangle-free process. *Adv. Math.* **221** 1653–1677.
[6] Bohman, T., Frieze, A. and Lubetzky, E. (2015) Random triangle removal. *Adv. Math.* **280** 379–438.
[7] Bohman, T. and Keevash, P. (2010) The early evolution of the H-free process. *Invent. Math.* **181** 291–336.
[8] Bollobás, B. (1988) The chromatic number of random graphs. *Combinatorica* **8** 49–55.
[9] Bollobás, B. (2012) Personal communication.
[10] Bollobás, B. and Riordan, O. (2015) An old approach to the giant component problem. *J. Combin. Theory Ser. B* **113** 236–260.

[11] Bollobás, B. and Brightwell, G. (1992) The height of a random partial order: Concentration of measure. *Ann. Appl. Probab.* **2** 1009–1018.
[12] Bollobás, B. and Riordan, O. (2009) Random graphs and branching processes. In *Handbook of Large-Scale Random Networks*, Vol. 18 of *Bolyai Society Mathematical Studies*, pp. 15–115.
[13] Borgs, C., Chayes, J. T., Mertens, S. and Pittel, B. (2004) Phase diagram for the constrained integer partitioning problem. *Random Struct. Alg.* **24** 315–380.
[14] Bushaw, N., Collares Neto, M., Morris, R. and Smith, P. (2015) The sharp threshold for maximum-size sum-free subsets in even-order abelian groups. *Combin. Probab. Comput.* **24** 609–640.
[15] Chalker, T. K., Godbole, A. P., Hitczenko, P., Radcliff, J. and Ruehr, O. G. (1999) On the size of a random sphere of influence graph. *Adv. Appl. Probab.* **31** 596–609.
[16] Chung, F. and Lu, L. (2006) Concentration inequalities and martingale inequalities: A survey. *Internet Math.* **3** 79–127.
[17] Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications*, Springer.
[18] Dubhashi, D. P. and Panconesi, A. (2009) *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press.
[19] Erdős, P., Suen, S. and Winkler, P. (1995) On the size of a random maximal graph. *Random Struct. Alg.* **6** 309–318.
[20] Freedman, D. A. (1975) On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.
[21] de Graaf, M. and Manthey, B. (2014) Probabilistic analysis of power assignments. In *Mathematical Foundations of Computer Science 2014*, Vol. 8635 of *Lecture Notes in Computer Science*, Springer, pp. 201–212.
[22] Grable, D. A. (1998) A large deviation inequality for functions of independent, multi-way choices. *Combin. Probab. Comput.* **7** 57–63.
[23] Harris, T. E. (1960) A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** 13–20.
[24] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
[25] Janson, S. (1990) Poisson approximation for large deviations. *Random Struct. Alg.* **1** 221–229.
[26] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience.
[27] Janson, S. and Ruciński, A. (2004) The deletion method for upper tail estimates. *Combinatorica* **24** 615–640.
[28] Janson, S. and Warnke, L. The lower tail: Poisson approximation revisited. *Random Struct. Alg.*, to appear. `arXiv:1406.1248`.
[29] Kim, J. H. (1995) On Brooks' theorem for sparse graphs. *Combin. Probab. Comput.* **4** 97–132.
[30] Kim, J. H. and Vu, V. H. (2000) Concentration of multivariate polynomials and its applications. *Combinatorica* **20** 417–434.
[31] Kim, J. H. and Vu, V. H. (2004) Divide and conquer martingales and the number of triangles in a random graph. *Random Struct. Alg.* **24** 166–174.
[32] Krivelevich, M., Lubetzky, E. and Sudakov, B. (2013) Longest cycles in sparse random digraphs. *Random Struct. Alg.* **43** 1–15.
[33] Makai, T. (2015) The reverse H-free process for strictly 2-balanced graphs. *J. Graph Theory* **79** 125–144.
[34] McDiarmid, C. (1989) On the method of bounded differences. In *Surveys in Combinatorics*, Vol. 141 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, pp. 48–188.
[35] McDiarmid, C. (1998) Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, Vol. 16 of *Algorithms and Combinatorics*, Springer, pp. 195–248.
[36] McKay, B. and Skerman, F. (2013) Degree sequences of random digraphs and bipartite graphs. `arXiv:1302.2446`.

[37] Milman, V. D. and Schechtman, G. (1986) *Asymptotic Theory of Finite-Dimensional Normed Spaces*, Vol. 1200 of *Lecture Notes in Mathematics*, Springer.
[38] Osthus, D. and Taraz, A. (2001) Random maximal H-free graphs. *Random Struct. Alg.* **18** 61–82.
[39] Riordan, O. and Warnke, L. (2015) The Janson inequalities for general up-sets. *Random Struct. Alg.* **46** 391–395.
[40] Riordan, O. and Warnke, L. (2015) The evolution of subcritical Achlioptas processes. *Random Struct. Alg.* **47** 174–203.
[41] Rödl, V. and Thoma, L. (1996) Asymptotic packing and the random greedy algorithm. *Random Struct. Alg.* **8** 161–177.
[42] Schudy, W. and Sviridenko, M. (2011) Bernstein-like concentration and moment inequalities for polynomials of independent random variables: Multilinear case. `arXiv:1109.5193`.
[43] Schudy, W. and Sviridenko, M. (2012) Concentration and moment inequalities for polynomials of independent random variables. In *Proc. 23rd Annual ACM–SIAM Symposium on Discrete Algorithms: SODA'12*, SIAM, pp. 437–446.
[44] Shamir, E. and Spencer, J. (1987) Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica* **7** 121–129.
[45] Spencer, J. (1990) Counting extensions. *J. Combin. Theory Ser. A* **55** 247–255.
[46] Spencer, J. (1995) Asymptotic packing via a branching process. *Random Struct. Alg.* **7** 167–172.
[47] Steiger, W. L. (1969) A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Statist.* **40** 764–769.
[48] Talagrand, M. (1996) A new look at independence. *Ann. Probab.* **24** 1–34.
[49] Vu, V. H. (2000) On the concentration of multivariate polynomials with small expectation. *Random Struct. Alg.* **16** 344–363.
[50] Vu, V. H. (2002) Concentration of non-Lipschitz functions and applications. *Random Struct. Alg.* **20** 262–316.
[51] Warnke, L. (2014) The C_ℓ-free process. *Random Struct. Alg.* **44** 490–526.
[52] Warnke, L. (2014) When does the K_4-free process stop? *Random Struct. Alg.* **44** 355–397.
[53] Wormald, N. C. (1999) Models of random regular graphs. In *Surveys in Combinatorics*, Vol. 267 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, pp. 239–298.