CONCENTRATION OF MULTIVARIATE POLYNOMIALS AND ITS APPLICATIONS

Jeong Han Kim
Microsoft Research
One Microsoft Way
Redmond WA98052

Van H. Vu
Microsoft Research
One Microsoft Way
Redmond WA98052

Abstract. Suppose $t_1, ..., t_n$ are independent random variables which take value either 0 or 1, and Y is a multi-variable polynomial in t_i's with positive coefficients. We give a condition which guarantees that Y concentrates strongly around its mean even when several variables could have a large effect on Y. Some applications will be discussed.

§1 INTRODUCTION

§1.1 The problem

In this paper, we consider independent random variables $t_1, t_2, \dots, t_n$ which can have two values 0 and 1, and a polynomial $Y = \sum_{e\in\mathbf{E}} w_e \prod_{i\in e} t_i$, where w_e are positive coefficients, and $\mathbf{E}$ is a collection of subsets of $\{1, 2, \dots, n\}$. If the size of a largest subset in $\mathbf{E}$ is k, Y is called *a positive polynomial of degree k*. Positive polynomials naturally arise in many probabilistic combinatorics problems (see §4) and, in most cases, they are expected to be concentrated near their means. This paper is intended to give easy-to-check conditions which yield strong concentration. Several examples are presented to demonstrate the power of the result and the comfort of using these conditions.

If Y has degree 1, i.e., $Y = \sum_{i=1}^n w_i t_i$, then the well-known Chernoff's bound gives that

$$Pr(|Y - E(Y)| > \lambda) < 2\exp\big(-\lambda^2/(2\sum_{i=1}^n w_i^2)\big).$$

This bound is generalized by Azuma [Az].

Theorem 1.1. *Let $E_j(Y) = E(Y|t_1, \dots, t_j)$ and $d_j = d_j(t_1, \dots, t_j) := E_j(Y) - E_{j-1}(Y)$. Then*

$$Pr(|Y - E(Y)| > \lambda) < 2\exp\big(-\lambda^2/2\sum_{i=1}^n \|d_i\|_\infty^2\big),$$

where $\|d_i\|_\infty$ is the maximum of d_i over all possible $t_1, \dots, t_j$.

The quantity d_i is usually referred to as the *effect* of the i^{th} variable. Azuma's theorem roughly says that if the sum of squares of the maximum effects is small, then the objective function is strongly concentrated. Although a powerful tool and frequently used (see, for instance [AS, McD]) Azuma's theorem has its shortcoming, namely, that one needs to consider the maximum effects. If the effects are too large with very small probability and small enough otherwise, we still expect a similar concentration result. However, in this case Azuma's theorem cannot be applied. Here is an example.

Example. A random graph $G(n,p)$ on the vertex set $\{1,2,\dots,n\}$ is obtained by choosing every possible edge ij independently with probability p, where p is a function of n (see e.g. [Bol]). In this case, our $\{0,1\}$ random variables are t_{ij}, for all pairs $1 \le i < j \le n$. Let Y be the number of triangle in $G(n,p)$. Obviously,

$$Y = \sum_{1 \le i < j < k \le n} t_{ij} t_{ik} t_{jk}.$$

The mean of Y is of order n^3p^3, and the maximum effect d_{last} of the last variable (regardless ordering) is $(1-p)(n-2)$. So if p is much less than $n^{-2/3}$, the maximum effect would be larger than the mean, so Azuma's inequality is not useful.

Notice that in the above example, d_{last} is typically very small since the number triangles containing a fixed edge has mean $\Theta(np^2)$ and is concentrated near its mean. A couple of Azuma type inequalities ([Kim, Gra]) are invented to overcome these situations and used to solve graph coloring and hypergraph matching problems. The primary goal of this paper is to give a new concentration result which could be applied even when the maximal effect is large.

§1.2 **Main result**

In this subsection, we describe our main result. By technical reasons, we will allow some variables t_i to be constants. The case of interest, when all t_i are $\{0,1\}$ variables, can be seen as the special case of the statement. To state the theorem, we first need to introduce some technical terms.

Let H be a (weighted) hypergraph with $V(H) = \{1,2,...,n\}$. We allow H to have empty edges. Each edge e has some weight $w(e)$. We assume that each edge has at most k vertices. Suppose t_i, $i = 1,,2...,n$ are independent random variables, and t_i could be one of the two following types

- t_i is a $\{0,1\}$ random variable with expected value p_i.
- $t_i = p_i$ with probability 1.

Consider a polynomial

$$Y_H = \sum_{e \in \mathcal{E}(H)} w(e) \prod_{s \in e} t_s.$$

We call H the *supporting hypergraph* of Y. If e is the empty set, then by convention $\prod_{s\in e} t_s = 1$. In the whole paper, logarithms have natural base.

Example 1. If $V(H) = \{1,2,3\}$ and $\mathcal{E}(H) = \{\{1,2\},\{3\},\emptyset\}$ with weights $2, 0.2, 1$, respectively then:

$$Y_H = 2t_1t_2 + 0.2t_3 + 1.$$

Example 2. In the example of the previous subsection, $Y = Y_H$, where H is the 3-uniform hypergraph constructed by the triangles. The edge set of H contains all triples $\{ij, jk, ki\}$ for all $1 \le i < j < k \le n$ and all edges have weight 1.

Truncated subhypergraphs. For each subset A of $V(H)$, H_A (the A-truncated subhypergraph of H) is defined as follows:

$V(H_A) = V(H) \backslash A$.
$\mathcal{E}(H_A) = \{B \subset V(H_A), B \cup A \in E(H)\}$.
If $B \in \mathcal{E}(H_A)$ then $w(B) = w(B \cup A)$.

Formally, we can write

$$Y_{H_A} = \sum_{e, A \subset e} w_e \prod_{i \in e \backslash A} t_i.$$

For instance, if we consider the hypergraph in Example 2, and let $A = \{12\}$ (the set consists the edge 12). Then, $Y_{H_A} = \sum_{n \ge k \ge 3} t_{1k} t_{2k}$.

Our intuition behind the formalization of Main Theorem is that if the expectations of all partial derivatives (of any order) of a positive polynomial Y are significantly smaller than the expectation of Y, then Y is strongly concentrated. This gives an explanation to the introduction of truncated subhypergraphs. Under the given circumstances, the partial derivative of Y_H with respected to $\{t_i : i \in A\}$ is exactly Y_{H_A}.

Now let $E_i(H) = max_{A \subset V(H), |A| = i} E(Y_{H_A})$, this quantity could be interpreted as the maximal *average* effect of a group of i random variables. Note that $E_0(H)$ is simply the expected value of Y_H. Finally let $E(H) = \max_{i \ge 0} E_i(H)$ and $E'(H) = \max_{i \ge 1} E_i(H)$.

Main Theorem. *In this setting*

$$Pr(|Y_H - E_0(H)| > a_k (E(H) E'(H))^{1/2} \lambda^k) = O(exp(-\lambda + (k-1) \log n)),$$

for any positive number $\lambda > 1$ *and* $a_k = 8^k k!^{1/2}$.

The moral of this theorem is that if the **average** effect of any group of at most k random variables is considerable smaller than the expectation of Y, then Y is strongly concentrated. The power of this theorem resides in the magic word "average". In many cases, when the maximal effect is too large for us to apply Azuma's theorem, the average effects are still sufficiently small to allow us to use Main Theorem. To illustrate this idea, let us consider a quick and instructive application, which also shows that the concept of positive polynomials arises very naturally in probabilistic combinatorics.

§1.3 A sample application

Let us reconsider the problem of triangle counting. Let $G(n, p)$ be a random graph on n vertices with edge probabilities $p = n^{\epsilon - 1}$, for some positive $\epsilon < 1/3$. We are interested in Y, the number of triangles in $G(n, p)$. As discussed in §1.1, in this range of p, Azuma's theorem does not give any information about the concentration of Y. We will now show that our theorem yields a considerably strong concentration result. Recall that

$$Y = \sum_{1 \le i < j < k \le n} t_{ij} t_{ik} t_{jk}.$$

It is straightforward to check that $E_0(Y) = \binom{n}{3}p^3, E_2(Y) = p,\ E_3(Y) = 1$ and crucially $E_1(Y) = (n-2)p^2$. The real difference is made here, because instead of the maximal effect of a variable, which is $(n-2)(1-p)$ (and is larger than the expectation of Y), we only need to take into account the average effect $E_1(Y)$. This quantity is much smaller than the expected value of Y and this makes it possible to derive a concentration result.
Observe that except $E_0(Y) = \Omega(n^{3\epsilon})$, all other $E_i(Y)$ are at most 1. Applying Main Theorem we obtain

$$Pr(|Y - E_0| > a_3 E_0(Y)^{1/2}\lambda^3) = O(exp(-\lambda + 2\log n)) \tag{1.4.1}$$

Choosing $\lambda = \omega(\log n)$ one can see that (1.4.1) provides a fairly strong concentration result, where the tail is slightly larger than the square root of the expected value, and the bound is superpolynomially small.

The rest of the paper is organized as follows. In the next section, we introduce a computational model and few important lemmas. Main Theorem will be proved in Section 3. Section 4 is devoted to applications, found in various areas of probabilistic combinatorics and random graphs. In particular, we will give a short proof of a strengthened version of a theorem of Spencer from [Spe].

§2 THE COMPUTATIONAL MODEL

§2.1 The model

Let us introduce our computational model, which provides the underlying idea for the proof of Main Theorem. We consider the probability space generated by n independent $\{0,1\}$ random variables $t_1, ..., t_n$, where $E(t_i) = p_i$. This space has 2^n vectors and the weight (or probability) of a vector $v = (v_1, .., v_n)$ is $\prod_{i=1}^n p_i(v_i)$ (where $p_i(1) = p_i$ and $p_i(0) = 1 - p_i = q_i$). For instance, the weight of the vector $(1, 1, ..., 1)$ is $\prod_{i=1}^n p_i$.

In this section, we consider a general function Y, which is not necessarily a polynomial. Given a function $Y = Y(t_1, \dots, t_n)$, one can evaluate Y using a decision tree. We consider a decision tree of depth n, and at a node at level i, we ask the question "what is the value of v_i". If the answer is 1, then we go to the right hand side child of the recent node, if the answer is 0 then we go to the left and continue until we reach a leaf. There are 2^n leaves representing the vectors in the space. In general a node at level i will be labeled by a $0,1$ vector of length i, which is the sequence of answer leading to this node. We label the root by the empty set and at each leaf v we write the corresponding value of $Y(v)$.

For any leaf v, let v_i be its i^{th} coordinate, and v^i be the vector formed by the first i coordinates of v. If a is a vector of length i and b is a vector of length j, then $< a, b >$ denotes the vector of length $i + j$ obtained by writing b behind a. For a node v^i at level i, we let $E(v^i)$ denote the expected value of the leaves below v^i, namely:

$$E(v^i) = \sum_{u, u^i = v^i} Y(u) \prod_{j=i+1}^{n} p_j(u_j).$$

By definition, the value at the root is $E(\emptyset) = E(Y)$.

§2.2 Lemmas

First we introduce some new notations. For $z = 0$ or 1, let

$$\mu_{i,z}(v) = E(< v^{i-1}, z >) - E(v^{i-1})$$

It is easy to compute that

$$\mu_{i,1}(v) = q_i(E(< v^{i-1}, 1) - E(< v^{i-1}, 0 >))$$

and

$$\mu_{i,0}(v) = p_i(E(< v^{i-1}, 0 > - E(< v^{i-1}, 1 >))$$

Denote by $c(v)$ the maximum value of $\mu_{i,z}(v)$ over all possible choice of i and z. Let $c_Y = \max_v c(v)$. Set

$$V_i(v) = p_i\mu_{i,1}(v)^2 + q_i\mu_{i,0}(v)^2.$$

The previous computation yields

$$V_i(v) = p_iq_i(E(< v^{i-1}, 1) - E(< v^{i-1}, 0 >))^2 \leq p_iC_i^2,$$

where $C_i = |E(< v^{i-1}, 1 >) - E(< v^{i-1}, 0 >)|$.

It is apparent that $c(v) \leq \max_i C_i(v)$, and $c_Y \leq \max_v \max_i C_i(v)$. Furthermore, define

$$V(v) = \sum_{i=1}^{n} V_i(v) \text{ and } V_Y = \max_v V(v).$$

Theorem 2.2.1. *If $c' \geq c_Y$, $V' \geq V_Y$ and $0 < \lambda < V'/(c')^2$, then*

$$Pr(|Y - E(Y)| > (\lambda V')^{1/2}) < 2e^{-\lambda/4}.$$

Proof. The proof relies essentially on the following lemma, which is a special case of a lemma proved in [Gra].

Lemma. *If $x < 1/c_Y$ and $E(Y) = 0$ then $E(e^{xY}) < e^{x^2V_Y}$.*

Proof of Lemma. We use induction on n. The case $n = 0$ is trivial. Suppose $n \geq 1$. Notice that, by definition, $\mu_{1,0}(v)$ are the same for all v. So for all v, we have $\mu_{1,0}(v) = \mu_{1,0}$, where $\mu_{1,0}$ is some constant does not depend on v. Similarly, we can set $\mu_{1,1}(v) = \mu_{1,1}$ and $V_1(v) = V_1$ for all v.

Consider the function $Y - \mu_{1,0}$ assigned to the left subtree of depth $n-1$ of the original tree. This function has expected value 0, so the induction hypothesis gives

$$E(e^{x(Y-\mu_{1,0})}) < e^{x^2(V_Y - V_1)}.$$

A similar argument on the right subtree gives

$$E(e^{x(Y-\mu_{1,1})}) < e^{x^2(V_Y - V_1)}.$$

On the other hand,

$$E(e^{xY}) = p_1e^{x\mu_{1,1}}E(e^{x(Y-\mu_{1,1})}) + q_1e^{x\mu_{1,0}}E(e^{x(Y-\mu_{1,0})});$$

therefore,

$$E(e^{xY}) < p_1 e^{x\mu_{1,1}} e^{x^2(V_Y - V_1)} + q_1 e^{x\mu_{1,0}} e^{x^2(V_Y - V_1)}.$$

It remains to show

$$p_1 e^{x\mu_{1,1}} + q_1 e^{x\mu_{1,0}} \leq e^{x^2 V_1}. \tag{2.2.1}$$

Consider the Taylor expansion of the left hand side of (2.2.1)

$$p_1(1 + x\mu_{1,1} + x^2\mu_{1,1}^2/2 + ...) + q_1(1 + x\mu_{1,0} + x^2\mu_{1,0}^2/2 + ...)$$

$$= 1 + 0 + x^2(p_1\mu_{1,1}^2 + q_1\mu_{1,0}^2)/2 + ...$$

$$= 1 + V_1 x^2 \sum_{i=2}^{\infty} \frac{1}{i!}(q_1(x\mu_{1,1})^{i-2} + p_1(x\mu_{1,0})^{i-2}).$$

Since $x < 1/c_Y$, both $x\mu_{0,1}$ and $x\mu_{1,1}$ have absolute values less than 1. Thus

$$\sum_{i=2}^{\infty} \frac{1}{i!}(q_1(x\mu_{1,1})^{i-2} + p_1(x\mu_{1,0})^{i-2}) < \sum_{i=2} \frac{1}{i!}(p_1 + q_1) = \sum_{i=2} \frac{1}{i!} < 1.$$

The last inequality implies that the left hand side of (2.2.1) is at most $1 + V_1 x^2$. Since $1 + V_1 x^2 \leq e^{x^2 V_1}$, the proof of the lemma is complete. □

To prove the theorem, note that when we replace Y by $Y - E(Y)$ the parameters involved in the bound (such as μ, c and V) do not change. Therefore, without loss of generality, we can assume that Y has mean 0. Having done this, we can finish using the standard Markov inequality argument. Setting $x = \frac{1}{2}(\lambda/V')^{1/2}$; it is trivial that $x < 1/c' \leq 1/c$. Thus by the previous lemma and Markov's inequality we have

$$Pr(Y > (\lambda V')^{1/2}) = Pr(e^{xY} > e^{x(\lambda V')^{1/2}}) < E(e^{xY}) e^{-x(\lambda V')^{1/2}}$$

$$< e^{x^2 V_Y - x(\lambda V')^{1/2}} < e^{x^2 V' - x(\lambda V')^{1/2}} = e^{-\lambda/4}.$$

By symmetry, we obtain $Pr(|Y| > (\lambda V')^{1/2}) < 2e^{-\lambda/4}$ and the proof is finished. □

The following theorem is the key tool in the proof of Main Theorem.

Theorem 2.2.2. *Let V and c be two arbitrary positive numbers and $B = \{v | c(v) > c$ or $V(v) > V\}$. If $0 < \lambda < V/c^2$ then*

$$Pr(|Y - E(Y)| > (\lambda V)^{1/2}) < 2e^{-\lambda/4} + Pr(B).$$

Proof. Call a leaf v *bad* if it is in B. Consider the path from the root to a bad leaf v and let i be the first index so that either $\sum_{j=1}^{i} V_j(v) \geq V$ or $c_i(v) > c$. If i is such index, we call the node v^{i-1} *exceptional*. The key observation here is that all leaves below an exceptional node are bad. For every leaf u below an exceptional node v^{i-1} change the value of $Y(u)$ to $E_{i-1}(u)$. Let Y' be the new function. Since exceptional nodes are incomparable, Y'

is well-defined. Notice that in the new tree $V(v) < V$ and $c(v) < c$ for every leaf v. So Theorem 2.2.1 implies

$$Pr(|Y' - E(Y')| > (\lambda V)^{1/2}) < 2e^{-\lambda/4}.$$

To complete the proof, observe that $E(Y') = E(Y)$; moreover, $Y(v) = Y'(v)$ for all $v \notin B$. This yields

$$Pr(|Y - E(Y)| > (\lambda V)^{1/2}) < 2e^{-\lambda/4} + Pr(B),$$

completing the proof. □

Notice that Theorem 2.2.2. still holds if we allow some random variables be constants.

§3 PROOF OF THE MAIN THEOREM

The leading idea of the proof is to use induction on k and apply Theorem 2.2.2. Actually, the hardest part of the proof is to state a proper induction hypothesis, i.e., to come to the right inequality in the theorem. Once this has been found, it is not too hard to prove !

Let us start with a definition.

Shadow hypergraphs. The shadow H^* of a (weighted) hypergraph H on the vertex set $\{1, 2, \ldots, n\}$ is a (weighted) hypergraph on the same set of vertices, where the edge set of H^* contains of the "shadows" of edges in H. The formal description now follows.

For each edge $e = \{a_1, .., a_l\} \in \mathcal{E}(H)$, where $a_1 < a_2 < .. < a_l$, create l new edges $e_j = \{a_1, ..., a_j\}, j = 0, 1, 2..., l-1$ where $e_0 = \emptyset$; the set of new edges e_j will form the edge set of H^*. In H^* set the weight of e_j equal to $w(e)\prod_{s=j+1}^{l} p_{a_s}$, where p_i is the expectation of the random variable t_i. If an edge f appears in H^* many times (it could be a shadow of different edges in H), then we keep one copy of f and add up the weights.

Example. If H has three edges $\{1, 2, 3\}, \{1, 2\}, \{3\}$ with weights a, b and c, respectively, then H^*) has three edges $\{1\}, \{1, 2\}, \{\emptyset\}$ with weights $ap_2p_3+bp_2, ap_3, ap_1p_2p_3+bp_1p_2+cp_3$, respectively.

It is essential to note that if every edge of H has at most k vertices, and A is a nonempty sets of variables, then every edge of H_A and H^* has at most $k-1$ vertices. This observation will enable us to apply induction.

Proof of Main Theorem

Given a positive polynomial Y and its support hypergraph H, we will show by induction on k the following. For any $\lambda > 1$

$$Pr(|Y(H) - E(Y(H))| > c_k(E(H)E'(H))^{1/2}\lambda^k) < d_k exp(-\lambda/4 + (k-1)\log n),$$

where d_k is recursively defined as follows: $d_1 = 2$, $d_k = (1 + \frac{1}{n})d_{k-1} + \frac{2}{n^{k-1}}$ and $c_k = 2^{k-1}(k!)^{1/2}$. A straightforward calculation shows that $d_k < 2e^2$. Replacing $\lambda/4$ by λ we obtain Main Theorem. The constants (c_k and d_k) are not best possible, but we make no attempt to optimize them.

If $k = 1$ then all edges have at most one vertices, therefore

$$Y_H = \sum_{i=1}^{n} w_i t_i + w_0,$$

where w_i is the weight of t_i and w_0 is the weight of the (possible) empty edge. We can assume $w_0 = 0$. In this case $E_1(H) = max_i w_i$. For arbitrary $\lambda > 1$, set

$$V = max\{E_0(H), E_1(H)\}E_1(H)\lambda,$$

and

$$c = E_1(H).$$

To apply Theorem 2.2.2, notice that $C_i = w_i$. It is clear that our parameters satisfy the conditions of Theorem 2.2.2, namely

$$\sum V_i = \sum p_i w_i^2 \le (max w_i) \sum w_i p_i = E_1(H)E_0(H) < V,$$

and

$$\frac{V}{c^2} = \frac{V}{E_1^2(H)} \ge \lambda.$$

Hence by Theorem 2.2.2

$$Pr(|Y_H - E_0(H)| > V^{1/2}\lambda) \le Pr(|Y_H - E_0(H)| > (\lambda V)^{1/2}) < 2exp(-\lambda/4).$$

Now suppose the induction hypothesis holds for $k-1$, we prove it for k. Recall from Section 2.2 that $c(v) \le max_i C_i(v)$ and $V(v) \le \sum_{i=1}^n p_i C_i^2(v)$. So for arbitrary positive numbers c and V

$$Pr(c(v) > c \text{ or } V(v) > V) \le \sum_{i=1}^n Pr(C_i(v) > c) + Pr(\sum_{i=1}^n p_i C_i(v) > V/c). \tag{3.1}$$

We will bound each term in the right hand side of (3.1) by the induction hypothesis. Once these bounds are achieved, we use Theorem 2.2.2 to complete the proof. First we need the following elementary claims, whose proofs are omitted. The reader may verify these claims by a routine computation.

Claim 1. For any point $i \in V(H)$

$$E_j(H_{\{i\}}) \le E_{j+1}(H).$$

Claim 2.

$$E_j(H^*) \le kE_j(H).$$

The key idea in the proof is to notice that the effect $C_i(v)$ of the point i is again a positive polynomial, whose support hypergraph is the $\{i\}$-truncated hypergraph of H. By the definition of truncated hypergraphs, $C_i(v)$ has degree at most $k-1$. Since the variable v now plays no role, in the following we will write C_i instead of $C_i(v)$ and C instead of $C(v)$. We have that

$$C_i = \sum_{U \ni i} w(U) \prod_{s \in U \setminus \{i\}} t_s^i,$$

where $t_s^i = t_s$ if $s < i$ and $t_s^i = p_s$ if $s > i$. This corresponds to the fact that in the definition of μ_i (see the last section) we take the expectation after exposing the first i variables, so all variables with index larger than i become constants.

Now C_i is a polynomial defined on $H_{\{i\}}$ with a set of new atom variables t_s^i. Since the degree of C_i is at most $k-1$, we can apply the induction hypothesis and obtain that

$$Pr(|C_i - E(C_i)| > c_{k-1}(E(H_{\{i\}})E'(H_{\{i\}}))^{1/2}\lambda^{k-1}) < d_{k-1}exp(-\lambda/4 + (k-2)\log n). \tag{3.2}$$

Because $E(C_i)$, $E(H_{\{i\}})$ and $E'(H_{\{i\}})$ are at most $E'(H)$ by Claim 1, it follows that

$$Pr(C_i > 2c_{k-1}E'(H)\lambda^{k-1}) < d_{k-1}exp(-\lambda/4 + (k-2)\log n). \tag{3.3}$$

Now consider $C = \sum_{i=1}^{k} p_i C_i$. We can write C as a polynomial in the following way

$$C = \sum_{i \in V(H)} \sum_{E(H) \ni U \ni i} w(U) p_i \prod_{s \in U \setminus i} t_s^i = Y_{H^*}(t_1, \ldots, t_n).$$

One can verify that the support hypergraph of C is exactly the shadow hypergraph H^* of H. Since the degree of C is at most $k-1$, by the induction hypothesis we have

$$Pr(|C - E(C)| > c_{k-1}(E(H^*)E'(H^*))^{1/2}\lambda^{k-1}) < d_{k-1}exp(-\lambda/4 + (k-2)\log n). \tag{3.4}$$

By Claim 2, it follows that

$$Pr(C > 2c_{k-1}kE(H)\lambda^{k-1}) < d_{k-1}exp(-\lambda/4 + (k-2)\log n). \tag{3.5}$$

Now set $c = 2c_{k-1}E'(H)\lambda^{k-1}$ and $V = 4kc_{k-1}^2 E(H)E'(H)\lambda^{2k-1}$. Since $E(H) \geq E'(H)$, it is trivial that $V/c^2 \geq \lambda$. Moreover,

$$\sum_{i=1}^{n} Pr(C_i(v) > c) < nd_{k-1}exp(-\lambda/4 + (k-2)\log n) = d_{k-1}exp(-\lambda/4 + (k-1)\log n),$$

by (3.3). Furthermore, by (3.5) and the fact that $V/c = 2c_{k-1}kE(H)\lambda^k$ ($\lambda > 1$), we have

$$Pr(C > V/c) < d_{k-1}exp(-\lambda/4 + (k-2)\log n).$$

Now we are ready to apply Theorem 2.2.2, which implies

$$\begin{aligned} Pr(|Y_H - E(Y_H)| > V^{1/2}\lambda^{1/2}) &< 2exp(-\lambda/4) + (1 + \frac{1}{n})d_{k-1}exp(-\lambda/4 + (k-1)\log n)) \\ &= ((1 + \frac{1}{n})d_{k-1} + \frac{2}{n^{k-1}})(exp(-\lambda/4 + (k-1)\log n)) \\ &= d_k exp(-\lambda/4 + (k-1)\log n), \end{aligned}$$

by the definition of d_k. Since $(V\lambda)^{1/2} = 2c_{k-1}k^{1/2}(E(H)E'(H))^{1/2}\lambda^k$, we can complete the proof by setting $c_k = 2k^{1/2}c_{k-1}$. □

Remark. For the case $k = 1$, we need only $\lambda^{1/2}$ in the tail (instead of λ). Exploiting this, one can improve, for arbitrary k, the term λ^k to $\lambda^{k-1/2}$. However, in applications, this fact rarely matters.

§4 APPLICATIONS

§4.1 Main Theorem and the Semi-random method.

Main Theorem and Theorem 2.2.2 was originally found and proved in order to analyze an application of the semi-random method to solve Segre's long standing open problem in finite geometry (see [Seg], also [KV] and [Szo]). Consider a projective plane P of order q. A set A of vertices is an *arc* if no three points of A is on a line. An arc is complete if it cannot be extended by another point. Let $n(P)$ denote the minimum size of a complete arc in P. To determine the order of magnitude of $n(P)$ is among the central open questions in finite geometry. It was proven by Lunelli and Sce in the 50's that $n(P) = \Omega(q^{1/2})$, but no close upper bound has been known. For a Galois plane, which is a special projective plane, the best upper bound was $O(q^{3/4})$ proven by Szőnyi ([Szo]) by algebraic arguments.

Using the semi-random method combined with Main Theorem and Theorem 2.2.2, we succeeded to prove that the lower bound $\Omega(q^{1/2})$ is actually sharp up to a polylogarithmic term.

Theorem 4.1.1. [KV] *There is a constant c such that for any q and any plane P of order q*

$$n(P) < q^{1/2}\log^c q.$$

Later on, it has turned out that in many applications of the semi-random method, Main Theorem can be used in a very effective way to analyze the behavior of certain random processes. Since it may take a whole survey to discuss these techniques, we limit ourself to introducing one more result (Theorem 4.1.2), which was obtained by essentially combining the semi-random method (following [Joh]) with Main Theorem. This theorem involves the list chromatic number of a graph which is a natural generalization of the chromatic number. Given a graph G, the list chromatic number of G, $\chi_l(G)$, is the smallest number m such that if one assigns an arbitrary list of m colors to each vertex of G (the lists can be different on different vertices), then there is a proper coloring (in the classical sense) of the vertices so that each vertex receives a color from its list. The notion of list chromatic number was introduced by Erdős, Rubin and Taylor [ERT], and independently by Vizing [Viz] and is extensively studied in the last decade (see [Alo] for a survey). Theorem 4.1.2 below gives the best possible upper bound for the list chromatic number of locally sparse graphs, and improves several earlier results on the same topic (see [Kim, Joh, AKS, Vu1]).

Theorem 4.1.2. *(*[Vu2]*) Let G be a graph with maximal degree d. Suppose that for every vertex v of G, the neighborhood of v contains at most d^2/f edges, for some number $f > 2$. Then $\chi_l(G) = O(d/\log f)$.*

In the rest of this section, we focus on the applications of Main Theorem in the theory of random graphs. First, let us notice that Main Theorem implies the following

Corollary 4.1.3. *If there is a positive constant γ such that $E_i/E_0 = O(n^{-\gamma})$ for all $i > 0$, then there are positive constants ϵ and ϵ' such that $Pr(|Y - E(Y)| > n^{-\epsilon}E(Y)) < e^{-n^{\epsilon'}}$.*

§4.2 Number of rooted strictly balanced subgraphs in a random graph

Let H be a (small) graph with vertices labeled by $x_1, ...x_r, y_1, ..., y_{\mathbf{v}}$, where $R = (x_1, .., x_r)$ is a specified subset, called the roots. The pair (R, H) will be dubbed as *rooted* graph. Let G be a (big) graph on n vertices disjoint from H. Fix r points $x_1, .., x_r$ in G. We call an order $\mathbf{v}$-tuple $y_1, ..., y_{\mathbf{v}}$ an *extension* if the y_j are distinct from each other and from the x_i, moreover

$x_i \sim y_j$ in G whenever $x_i \sim y_j$ in H
$y_i \sim y_j$ in G whenever $y_i \sim y_j$ in H.

We denote by $N(x_1, ..., x_r)$ the number of extensions relative to a given pair (R, H) and a fixed set of vertices $x_1, .., x_r$.

Let $\mathbf{e}$ be the number of edges of H, excluding edges between the roots. The ratio $\mathbf{e}/\mathbf{v}$ is the density of H. If H' is an induced subgraph of H and contains R, then we call the pair (R, H') a proper subextension of R. We denote by $max(R, H)$ the maximum density of a proper subextension of R.

Consider $G = G(n, p)$. As usual, our atom $\{0, 1\}$ random variables are $t_{ij}, 1 \leq i < j \leq n$. We say that a property Q holds *almost surely* in $G(n, p)$ if the probability that Q does not hold tends to 0 as n tends to infinity. We are interested in the concentration of the random variable $N(x_1, .., x_r)$ in certain range of p.

We say p is *safe* if $p = n^{-\alpha}$, where α is a positive constant smaller than $1/max(R, H)$. Our goal is to prove that if p is safe then $N(x_1, \ldots, x_r)$, for any choice of $x_1, \ldots, x_r$, concentrates strongly around its expected value.

The investigation of the function $N(x_1, \ldots, x_r)$ was motivated by a theorem of Shelah and Spencer on zero-one laws. In [SS] Shelah and Spencer proved the following:

Theorem 4.2.1. *If $p = n^{-\gamma}$ for γ irrational, then p satisfies a zero-one law.*

We omit the (rather involved) description of the zero-one law in concern and refer the reader to [SS]. Here we will focus on a concentration theorem, which is the key tool in the proof of Theorem 4.2.1. Let N be expectation of $N(x_1, \ldots, x_r)$; the value of N does not depend on the choice of $x_1, \ldots, x_r$.

Theorem 4.2.2. *Let p be safe, then there are positive constants c and c' such that almost surely*

$$N \log^{-c} n < N(x_1, \ldots, x_r) < c'N,$$

hold for all r-tuples $(x_1, \ldots, x_r)$.

It was conjectured in [SS] that one could replace the left hand side by cN. This was confirmed by Spencer in a later paper [Spe] (see also [AS]), in which he proved

Theorem 4.2.3. *If p is safe, then for any positive constant ϵ*

$$(1 - \epsilon)N < N(x_1, \ldots, x_r) < (1 + \epsilon)N,$$

almost surely.

The proof of the last theorem in Spencer's paper [Spe] is rather complicated and splits into two cases: N large and N small. A subtle and somewhat counter-intuitive point in

this proof (pointed out to us by Spencer), is that it was harder to prove the statement in the case N large. In order to handle this case, one first needs to prove the statement in the other case, then finish by some additional tricks.

In the following, we will give a short proof for a stronger version of Spencer's theorem, in which we allow ϵ be a negative power of n. The amazing thing about this proof is that it follows almost immediately from Corollary 4.3.1, and there is no need to distinguish cases based on the magnitude of N.

Theorem 4.2.4. *If p is safe, then there are positive constants ϵ and ϵ' such that for any $x_1, \dots, x_r$*

$$Pr(|N(x_1, \dots, x_r) - N| > Nn^{-\epsilon}) < e^{-n^{\epsilon'}}.$$

Consequently, almost surely $N(1 - n^{-\epsilon}) < N(x_1, \dots, x_r) < N(1 + n^{-\epsilon})$ hold for all r-tuples $(x_1, \dots, x_r)$.

Proof. Let us write $N(x_1, .., x_r)$ as a polynomial in the atom variables t_{ij}'s. It is clear that

$$Y = N(x_1, .., x_r) = \sum_{y_1, .., y_v} \prod_{x_i y_j \in H} t_{x_i y_j} \prod_{y_i y_j \in H} t_{y_i y_j}$$

where the sum is taken over all ordered v-tuples, and Y has degree $\mathbf{e}$. The expected value of Y is $N = E_0 = \Theta(n^{\mathbf{v}} p^{\mathbf{e}})$. All we need is to show that Y has strong concentration. To apply Corollary 4.1.3, we only need to compute $E_i(Y)$. From here the proof becomes a routine calculation.

For any number $i \le \mathbf{e}$, let $j(i)$ be the minimum number j such that there is a set W of j elements so that the subextension $(R, R \cup W)$ has at least i edges. It is not hard to verify that

$$E_i = O(n^{\mathbf{v} - j(i)} p^{\mathbf{e} - i}).$$

It follows that

$$E_0 / E_i = \Omega(n^{j(i)} p^i) = \Omega((np^{i/j(i)})^{j(i)}).$$

Notice that p is safe, thus by definition

$$np^{i/j(i)} > n^{1 - \alpha i/j(i)} > n^{1 - \alpha max(R,H)} > n^{\gamma}$$

for some positive constant γ. Therefore, the expectations $E_i(H)$ satisfy the condition of Corollary 4.1.3. So we may conclude that if p is safe, then there are positive constants ϵ and ϵ' such that

$$Pr(|Y - E(Y)| > E(Y)n^{-\epsilon}) < e^{-n^{\epsilon'}}.$$

Since the right hand side is superpolynomially small, we can conclude that the same result holds simultaneously for every r-tuples $x_1, .., x_r$.

§4.3 Number of small subgraphs in a random graph

Fix a small graph H with $\mathbf{v}$ vertices and $\mathbf{e}$ edges, we are interested in X_H, the number of subgraphs of $G(n, p)$ isomorphic to H. This problem can be seen as a special case of the

previous application, because every graph is an extension of the empty set. Therefore, if δ is the largest density of a subgraph of H, and p is safe, namely, $p = n^{-\alpha}$, for some $\alpha < 1/\delta$, then a statement similar to the statement of Theorem 4.2.4 holds. Let us notice that we count the number of copies of H up to automorphism, therefore $E(X(H)) = n_{\mathbf{v}}p^{\mathbf{e}}/K$, where K is the size of the automorphism group of H. Since K is a constant, its presence does not change the content of the theorem.

Theorem 4.3.1. *If p is safe then there are positive constants ϵ and ϵ' such that*

$$Pr(|X_H - E(X_H)| > E(X_H)n^{-\epsilon}) < e^{-n^{\epsilon'}}$$

In particular, when H is strictly balanced (i.e, the density of H is larger than the density of any of its induced subgraphs), we have

Theorem 4.3.2. *If $\alpha\mathbf{e} < \mathbf{v}$ and $p = n^{-\alpha}$ then there are positive constants ϵ and ϵ' such that*

$$Pr(|X_H - E(X_H)| > E(X_H)n^{-\epsilon}) < e^{-n^{\epsilon'}}$$

Theorem 4.3.2 implies an exponential bound on the probability that a random graph does not contain a copy of a small fixed graph. This bound is weaker than a well known result of Janson, Łuczak and Rucinski [JLR], but is in the same spirit.

Corollary 4.3.3. *Under the condition of the previous theorem, the probability that $G(n,p)$ does not contain a copy of H is exponentially small (to be precise $e^{-n^{\epsilon}}$), for some positive constant ϵ).*

Theorem 4.3.2 still holds if we only require H to be balanced, that is, the density of H is not smaller than any of its subgraph. All theorems proved in this and the previous subsection can be generalized for random graphs with non-uniform edge probability (Main Theorem does not require the atom variables to be i.i.d.). Another direction to strengthen these applications is to allow the size of H be a function of n tending slowly to infinity. One can show that the results derived in the last two subsections still hold if $V(H) = (\log n)^{1-\epsilon}$ for any positive ϵ. We omit the details.

Added in proof. Main Theorem is not too effective when applied for functions with small expectations (of order $O(\log n)$, say). Recently, this case was investigated (motivated by applications in number theory [Vu5]) , and a concentration result on polynomials with small expectations (order $O(\text{polylog } n)$) was proven in [Vu3]. A generalized and strengthened version of Main Theorem along with several other applications will appear in a new paper [Vu4].

Aknowledgement. Part of the work of the second author was done while he was at IAS and supported by a grand from NEC Research and the State of New Jersey

References

[AKS] N. Alon, M. Krivelevich and Sudakov, Coloring graphs with sparse neighborhoods, submitted.

[AS] N. Alon and J. Spencer, **The Probabilistic Method**, Wiley, New York, 1992.

[Bol] B. Bollobás, **Random Graphs**, Academic Press, London, 1985.

[ERT] P. Erdős, A. L. Rubin and H. Taylor, Choosability in graphs, *Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium* XXVI (1979) 125-157.

[Gra] D. Grable, A large deviation inequality for functions of independent, multi-way choices, *Combinatorics, probability and Computing* (1998) 7, 57-63.

[Joh] A. Johansson, Asymptotic choice number for triangle free graphs, *DIMACS Technical Report* (1996).

[JLR] S. Janson, T. Łuczak and A. Rucinski, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, in: M. Karonski et al. eds., Random graphs 87 (Wiley , New York, 1990) 73-87.

[Kim] J. H. Kim, On Brooks' theorem for sparse graphs, *Combinatorics, Probability and Computing* 4 (1995), 97–132.

[KV] J.H. Kim and V.H. Vu, Small complete arcs on finite projective planes, *submitted*.

[Segr] B. Segre, Le geometrie di Galois, *Ann. Mat. Pura Appl.* (1959) 48, 1-97.

[Segr2] B. Segre, Introduction to Galos geometries (ed. J.W.P Hirschfeld) *Mem. Accad. Naz. Lincei* (1967) 8, 133-263.

[SS] S. Shelah and J. Spencer, Zero-one laws for sparse random graphs, *J. Amer. Math. Soc.* (1988) 1, 97-115.

[Spe] J. Spencer, Counting extensions, *Journal of Combinatorial Theory, Series A* (1990) 55, 247-255.

[Szo], T. Szőnyi, Complete arcs in $PG(2, q)$, a survey in *Quad. del Sem. Geom. Comb. Univ. di Roma ("La Sapienza")* (1989) 94.

[Vu1] V. H. Vu, On some simple degree conditions which guarantee the upper bound on the chromatic (choice) number of random graphs, *Journal of Graph Theory, 31 (1999), no. 3, 201–226.*

[Vu2] V. H. Vu, On the list chromatic number of locally sparse graphs, *preprint*.

[Vu3] V. H. Vu, On the concentration of multi-variate polynomials with small expectation, *submitted*.

[Vu4] V. H. Vu, Some new concentration results and applications, *preprint*.

[Vu5] V.H. Vu, On a thin Waring problem, *submitted*.