When Do Random Subsets Decompose a Finite Group?

Ariel Yadin [*]

Abstract

Let A, B be two random subsets of a finite group G. We consider the event that the products of elements from A and B span the whole group; i.e. $\{AB \cup BA = G\}$. The study of this event gives rise to a group invariant we call $\Theta(G)$. $\Theta(G)$ is between $1/2$ and 1, and is 1 if and only if the group is abelian. We show that a phase transition occurs as the size of A and B passes $\sqrt{\Theta(G)|G|\log|G|}$; i.e. for any $\varepsilon > 0$, if the size of A and B is less than $(1-\varepsilon)\sqrt{\Theta(G)|G|\log|G|}$, then with high probability $AB \cup BA \neq G$. If A and B are larger than $(1+\varepsilon)\sqrt{\Theta(G)|G|\log|G|}$ then $AB \cup BA = G$ with high probability.

1 Introduction

Let G be a finite group. Two subsets $A, B \subset G$ are said to be a *decomposition* for G if $AB \cup BA = G$, where

$$AB \overset{\text{def}}{=} \left\{ ab \;\middle|\; a \in A, b \in B \right\}.$$

In [3], Kozma and Lev proved that for any finite group G, there always exists a decomposition A, B for G, such that $|A| \leq c\sqrt{|G|}$, $|B| \leq c\sqrt{|G|}$ (where $c > 0$ is some explicit constant, see [3] for details). We consider a similar question, but where the sets A, B are randomly chosen.

Let G be a finite group of size $n \geq 3$. Let $a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_k$ be $2k$ random elements (perhaps with repetitions) chosen independently from G, and let $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_k\}$. We investigate the event that A and B are a decomposition for G. Denote the probability of this event by

$$P(G, k) = \mathbb{P}\left[AB \cup BA = G\right].$$

[*] Department of Mathematics, The Weizmann Institute of Science, POB 26, Rehovot 76100, ISRAEL; ariel.yadin@weizmann.ac.il

Since $P(G,k)$ is monotone in k, it seems natural to ask whether a phase transition occurs, and if so, then what is the critical value. It turns out that there exists a group invariant $\Theta(G) \in (1/2, 1]$ (defined in Section 2 below), such that the critical value exists and is equal to $\sqrt{\Theta(G) n \log n}$, as stated in our main result:

Theorem 1. *Let G_n be a family of groups such that*

$$\lim_{n\to\infty} |G_n| = \infty.$$

For all n, let $C_n = \sqrt{\Theta(G_n)|G_n|\log|G_n|}$. Then for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(G_n, \lceil(1+\varepsilon)C_n\rceil) = 1,$$

and

$$\lim_{n\to\infty} P(G_n, \lfloor(1-\varepsilon)C_n\rfloor) = 0.$$

The proof of Theorem 1 follows from Lemmas 14 and 15. Actually, it can be seen from these lemmas that the window of the transition is smaller than stated by Theorem 1. Before we move to the proofs of these lemmas, we define the group invariant $\Theta(G)$, and elaborate on some of its properties.

2 A Group Invariant

For group theory background see [6].

Say we are interested in measuring how close a group is to being abelian. It seems reasonable to try and associate a number, say $\rho(G)$, to each group G, such that $\rho(G)$ has the following properties:

- $\rho(G) \in [0, 1]$.
- $\rho(G) = \rho(G')$, if G and G' are isomorphic as groups.
- $\rho(G) = 1$ if and only if G is abelian.

Perhaps the first "probabilistic" quantity that comes to mind is the probability that two randomly chosen elements commute. If a, b are two random independent uniformly chosen elements from a finite group G, then

$$\mathbb{P}[ab = ba] = \sum_{x\in G} \frac{|C(x)|}{|G|^2}, \tag{2.1}$$

where $C(x) = \{g \in G \ : \ gx = xg\}$ denotes the centralizer of x in G. If we view G as acting on itself by conjugation, then $C(x)$ is the set of all elements that fix x. Also, the number of different orbits is just the number of conjugacy classes of G. Thus, by Burnside's counting lemma (see [6] Chapter 3, page 58), $\mathbb{P}[ab = ba] = R(G)/|G|$, where $R(G)$ is the number of conjugacy classes in G. (An alternative proof can be given through character theory, using the Schur orthogonality relations.)

In this note, we define a different group invariant, $\Theta(G)$. As it turns out, $\Theta(G) \in (1/2, 1]$, and $\Theta(G) = 1$ if and only if G is abelian. $\Theta(G)$ arises naturally when considering the question that two random sets form a decomposition of a group G, as seen in Theorem 1.

We use the notation $C(x) = \big\{g \in G \bigm| gx = xg\big\}$ to denote the centralizer of $x \in G$. Note that $C(x)$ is a subgroup of G.

Let G be a group of order n. Since for any $x \in G$, $2 \leq |C(x)| \leq n$, the function $f : [1/2, 1] \to \mathbb{R}$

$$f(\xi) = 2\xi \log n - \log \sum_{x \in G} \exp\left(\xi \log n \cdot \frac{|C(x)|}{n}\right)$$

is negative at $1/2$, non-negative at 1, and continuous monotone increasing on $[1/2, 1]$. Indeed,

$$f(1/2) = \log n - \log \sum_{x \in G} \exp\left(\frac{\log n}{2} \cdot \frac{|C(x)|}{n}\right) \leq \log n - \log\left(n \cdot e^{\log n/n}\right) = -\frac{\log n}{n}.$$

$$f(1) = 2\log n - \log \sum_{x \in G} \exp\left(\log n \cdot \frac{|C(x)|}{n}\right) \geq 2\log n - \log\left(n \cdot e^{\log n}\right) = 0.$$

$$f'(\xi) = 2\log n - \frac{\sum_{x \in G} \exp\left(\xi \log n \cdot \frac{|C(x)|}{n}\right) \cdot \frac{\log n |C(x)|}{n}}{\sum_{x \in G} \exp\left(\xi \log n \cdot \frac{|C(x)|}{n}\right)} \geq 2\log n - \log n > 0.$$

Thus, the following is well defined:

Definition 2. Let G be a finite group of order n. Define $\Theta = \Theta(G)$ to be the unique number in $[1/2, 1]$ satisfying:

$$2\Theta \log n = \log \sum_{x \in G} \exp\left(\Theta \log n \cdot \frac{|C(x)|}{n}\right). \tag{2.2}$$

Remark. $\Theta(G)$ is the solution of equation (2.2) If $x_1, x_2, \ldots, x_R$ are representatives of the conjugacy classes of G, the sum in the logarithm of the right hand side of (2.2) can be written as

$$\sum_{i=1}^{R} |[x_i]| \exp\left(\xi \log n \cdot \frac{1}{|[x_i]|}\right),$$

where $[x_i]$ is the conjugacy class of x_i. This sum may remind some readers of the "zeta function" studied by Liebeck and Shalev, see e.g. [4, 5]. Their zeta function is also used in the context of probabilistic group theory. We use the main result from [5] regarding this "zeta function" in Proposition 3 below.

The following proposition provides some properties of $\Theta(G)$. The proposition roughly shows that $\Theta(G)$ measures, in some sense, how "abelian" a group is. The properties of Θ are not essential to the proof of Theorem 1, and so some readers may wish to skip to Section 3.

Proposition 3. *Let G be a group of order n.*

(i). Let $Z(G)$ be the center of G; i.e. $Z(G) = \{g \in G \mid \forall\, x \in G \ : \ gx = xg\}$. Then,

$$\Theta(G) \geq \frac{\log|Z(G)|}{\log n},$$

and

$$\Theta(G) \leq \max\left\{\frac{2}{3}\left(1 + \frac{\log 2}{\log n}\right), \frac{\log|Z(G)| + \log 2}{\log n}\right\}.$$

(ii). G is abelian if and only if $\Theta(G) = 1$ (so the lower bound in (i) is tight).

(iii). Let $R = R(G)$ be the number of conjugacy classes of G (this is also the number of irreducible representations of G). Then,

$$\Theta(G) \geq \frac{1}{2 - R/n} > 1/2.$$

(iv). Let $G = D_{2m}$, the dihedral group of order $n = 2m$. Then,

$$\frac{2}{3} \cdot \left(1 - \frac{\log 2}{\log n}\right) \leq \Theta(D_{2m}) \leq \frac{2}{3} \cdot \left(1 + \frac{\log 2}{\log n}\right).$$

(This implies that the upper bound in (i) is tight.)

(v). Let $G = S_m$, the group of all permutations on m letters. So $n = m!$. Then,

$$\Theta(S_m) = \frac{1}{2} + o(1).$$

(vi). Let $1/2 \leq \alpha < 1$. Then, there exists a sequence of groups $\{G_n\}$, such that

$$\lim_{n\to\infty} \Theta(G_n) = \alpha.$$

(vii). Let G be a simple non-abelian group. Then,

$$\Theta(G) = \frac{1}{2} + o(1).$$

Proof. Let $\Theta = \Theta(G)$.

(i). For any $x \in Z(G)$, we have that $|C(x)| = n$. Thus,

$$2\Theta \log n \geq \log(|Z(G)|\, e^{\Theta \log n}) = \log |Z(G)| + \Theta \log n.$$

This proves the lower bound.

Note that since $C(x)$ is a subgroup, $|C(x)|$ must divide $|G|$. Thus, for any $x \notin Z(G)$, we have that $|C(x)| \leq n/2$. Thus,

$$n^{2\Theta} \leq |Z(G)| \cdot n^{\Theta} + (n - |Z(G)|) \cdot n^{\Theta/2} \leq n^{\Theta/2} \cdot 2 \max \left\{ |Z(G)|\, n^{\Theta/2}, n \right\}.$$

This proves the upper bound.

(ii). Assume towards a contradiction that $\Theta = 1$ and that G is not abelian. Then, there exists $x \in G$ such that $|C(x)| < n$. Since $|C(x)|$ must divide n, we get that $|C(x)| \leq n/2$. Thus, by the definition of Θ,

$$n^2 \leq (n-1)n + n^{1/2} = n^2 - n + n^{1/2},$$

a contradiction.

The other direction follows by (i), since if G is abelian, $|Z(G)| = n$.

(iii). By Burnside's Lemma, or by Schur's orthogonality relations, one can show that

$$\sum_{x \in G} |C(x)| = n \cdot R.$$

Using Jensen's inequality on the convex function $\exp\left(\frac{\Theta \log n}{n} \cdot \xi\right)$,

$$2\Theta \log n = \log \sum_{x \in G} \exp\left(\frac{\Theta \log n}{n} \cdot |C(x)|\right) \geq \log n \exp\left(\frac{\Theta \log n}{n} \cdot R\right) = \log n + \Theta \log n \cdot \frac{R}{n}.$$

The assertion follows.

(iv). The dihedral group of order $n = 2m$ is

$$D_{2m} = \left\langle x, y \ : \ x^m = y^2 = 1 \ , \ yxy = x^{-1} \right\rangle = \left\{ x^i, yx^i \ : \ i = 0, 1, \ldots, m-1 \right\}.$$

One can check that the following holds:

$$\begin{array}{cr} i \notin \{0, m/2\} & C(x^i) = \left\{1, x, \ldots, x^{m-1}\right\}, \\ i \in \{0, m/2\} & C(1) = C(x^{m/2}) = D_{2m}, \\ \text{if } m \text{ is even} & C(yx^i) = \left\{1, x^{m/2}, yx^i, yx^{i+m/2}\right\}, \\ \text{if } m \text{ is odd} & C(yx^i) = \left\{1, yx^i\right\}. \end{array}$$

Thus, $Z(D_{2m}) = \{1\}$ if m is odd, and $Z(D_{2m}) = \left\{1, x^{m/2}\right\}$, if m is even. So we get the upper bound by (i).

On the other hand, considering the elements $1, x, \ldots, x^{m-1}$, we have that

$$n^{2\Theta} \geq m \cdot n^{\Theta/2} = \frac{1}{2} n^{1+\Theta/2},$$

which implies the lower bound.

(v). We use the following notation: If $c = (i_1, i_2, \ldots, i_s) \in S_m$ is a cycle, and $\tau \in S_m$ is any permutation, then denote $c^\tau = (\tau(i_1), \tau(i_2), \ldots, \tau(i_s))$ (note that $c^\tau = \tau c \tau^{-1}$). For a permutation $\sigma \in S_m$ denote by $\mathrm{supp}(\sigma) = \{j \in [m] \ : \ \sigma(j) \neq j\}$ the support of σ. $|\sigma| = |\mathrm{supp}(\sigma)|$ denotes the size of the support.

Let $\sigma \in S_m$, and write $\sigma = c_1 c_2 \cdots c_\ell$, where c_i are cycles, ordered by their size from largest to smallest (i.e. $|c_i| \geq |c_j|$ for all $i \leq j$). Let $s = |c_1|$ be the size of the largest cycle in the decomposition. Let $r \geq 1$ be the index such that $|c_i| = s$ for all $1 \leq i \leq r$, and $|c_i| < s$ for all $i > r$.

Set

$$S = \bigcup_{i=1}^{r} \mathrm{supp}(c_i).$$

If $\tau \in C(\sigma)$, then $\tau\sigma\tau^{-1} = \sigma$. But it can easily be seen that

$$\tau\sigma\tau^{-1} = c_1^\tau c_2^\tau \cdots c_\ell^\tau.$$

Since $|c_i| = |c_i^\tau|$, we get that for any $j \in S$ we must have that $\tau(j) \in S$. Thus,

$$|C(\sigma)| \leq \left|\left\{\tau \in S_m \ \middle| \ \tau(S) = S\right\}\right| = |S|!(m - |S|)!.$$

If $r < \ell$, then since $|c_\ell| \geq 2$, we have that $|S| \leq m - 2$. Thus,

$$|C(\sigma)| \leq m! \cdot \frac{2}{m(m-1)} < m! \cdot \left(\frac{e}{m}\right)^2. \tag{2.3}$$

Assume that $r = \ell$. Then either σ is a cycle of length m, or a cycle of length $m-1$, or σ is the product of cycles of equal length.

If σ is the identity, then $|C(\sigma)| = m!$. If σ is a cycle of length m then $|C(\sigma)| = m$. If σ is a cycle of length $m-1$ then $|C(\sigma)| = m - 1$.

So we are left with the case where $\sigma = c_1 c_2 \cdots c_r$ and $|c_i| = m/r$ for all $1 \leq i \leq r$. Note that in this case,

$$C(\sigma) \subseteq \left\{c_1' c_2' \cdots c_r' \ \middle| \ \text{all } c_i' \text{ are cycles of length } m/r\right\}.$$

Thus,

$$|C(\sigma)| = \frac{m!}{\left(\frac{m}{r}\right)^r r!} < m! \cdot \left(\frac{e}{m}\right)^2 . \tag{2.4}$$

Combining (2.3) and (2.4) we get (for $n = m!$),

$$\begin{aligned} n^{2\Theta} &\leq n^{\Theta} + (m-1)! n^{\Theta/(m-1)!} + m(m-2)! n^{\Theta/m(m-2)!} \\ &\quad + (m! - (m-1)! - m(m-2)! - 1) \cdot n^{\Theta e^2/m^2} \\ &\leq (n-1) \cdot (1 + o(1)) + n^{\Theta} \end{aligned}$$

which shows that $\Theta \leq \frac{1}{2} + o(1)$ (as m tends to infinity).

(vi). Let $\alpha \in [1/2, 1)$. For all integers m, let $n_m = m!$ and $k_m = \lfloor n_m^{\alpha/(1-\alpha)} \rfloor$. So,

$$n_m^{\alpha/(1-\alpha)} \cdot (1 - 1/n_m) \leq k_m \leq (k_m n_m)^{\alpha}.$$

Let $G_m = C_{k_m} \times S_m$, where C_{k_m} is the cyclic group of order k_m. Note that for $c \in C_{k_m}$ and $\sigma \in S_m$ the centralizer of (c, σ) in G_m is the set $C_{k_m} \times C(\sigma)$. Thus, using the calculations for S_m in the previous proof, for $\Theta = \Theta(G_m)$,

$$\begin{aligned} |G_m|^{2\Theta} &\leq k_m \cdot \big(|G_m|^{\Theta} + (m-1)! |G_m|^{\Theta/(m-1)!} + m(m-2)! |G_m|^{\Theta/m(m-2)!} \\ &\quad + (n_m - 1 - (m-1)! - m(m-2)!) |G_m|^{\Theta e^2/m^2} \big) \\ &\leq k_m |G_m|^{\Theta} + (1 + o(1)) k_m n_m. \end{aligned}$$

Since $\alpha + \Theta > 1$, we get that $|G_m|^{2\Theta} \leq (2 + o(1))(k_m n_m)^{\Theta + \alpha}$, which implies that $\Theta(G_m) \leq \alpha + o(1)$.

On the other hand $|G_m|^{2\Theta} \geq k_m |G_m|^{\Theta} \geq (1 - 1/n_m) \cdot n_m^{\alpha/(1-\alpha)} |G_m|^{\Theta}$. Hence $\Theta(G_m) \geq \alpha - o(1)$.

(vii). Let G be a finite simple non-abelian group. Let $\mathcal{M}$ be the set of all maximal subgroups of G. Consider the following "zeta function" (defined in [2], and studied further in [4, 5]):

$$\zeta_G(s) = \sum_{M \in \mathcal{M}} [G : M]^{-s}.$$

Theorem 1.1 of [5] states that for any $s > 1$,

$$\zeta_G(s) \longrightarrow 0 \quad \text{as} \quad |G| \to \infty.$$

Since G is simple non-abelian, if x is not the identity in G, then $C(x)$ is a proper subgroup. Since any proper subgroup of G is contained in a maximal subgroup, we get that

$$\frac{|C(x)|}{|G|} = \sqrt{[G : C(x)]^{-2}} \leq \sqrt{\zeta_G(2)}.$$

So for all $x \neq 1$ in G we get that $\frac{|C(x)|}{|G|} = o(1)$. Plugging this into the definition of $\Theta = \Theta(G)$ we get that

$$n^{2\Theta} \leq n^{\Theta} + (n-1) \cdot n^{\Theta o(1)} \leq n^{\Theta} + n^{1+o(1)},$$

which implies that $\Theta \leq \frac{1}{2} + o(1)$. □

In the proof of Proposition 3, (vi), we use the product of a cyclic group with the symmetric group to obtain different values of Θ. This idea raises the following

Open Problem. Show that for any *abelian* group H and any group G, $\Theta(H \times G) \geq \Theta(G)$.

3 Suen's inequality

One of the main tools we use to prove our results is a correlation inequality by Suen (see Theorem 5 below).

Graph Notation. For a graph $\Gamma = (V, E)$ write $v \sim u$ if $\{v, u\}$ is an edge. For subsets $S, T \subseteq V$, write $S \sim T$ if there exists an edge between S and T. Thus, $S \not\sim T$ means that there is no edge between S and T. $v \sim S$ means there is an edge between v and some element of S.

Definition 4. Let $\{X_i\}_{i=1}^N$ be a collection of random variables. A graph $\Gamma = (V, E)$ is a *dependency graph* of $\{X_i\}_{i=1}^N$ if: $\Gamma = (V, E)$ is an undirected graph on the vertex set $V = \{1, \ldots, N\}$ such that for any two disjoint subsets $S, T \subset V$, if $S \not\sim T$ then the two families $\{X_i\}_{i \in S}$ and $\{X_i\}_{i \in T}$ are independent of each other. (In some texts Γ is called a superdependency digraph.)

The following is a result of Suen, slightly improved by Janson (see [1, 7]).

Theorem 5 (Suen's inequality)**.** *Let $X_1, \ldots, X_N$ be N Bernoulli random variables, and let $S_N = \sum_{i=1}^N X_i$. Let Γ be a* dependency graph *of $\{X_i\}_{i=1}^N$.*

Define

$$\Delta = \Delta\left(\Gamma, \{X_i\}_{i=1}^N\right) = \frac{1}{2} \sum_{i=1}^N \sum_{j \sim i} \mathbb{E}\left[X_i X_j\right] \prod_{k \sim \{i,j\}} \left(1 - \mathbb{E}\left[X_k\right]\right)^{-1},$$

and

$$\Delta^* = \Delta^*\left(\Gamma, \{X_i\}_{i=1}^N\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\sim i}\mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\prod_{k\sim\{i,j\}}\left(1-\mathbb{E}\left[X_k\right]\right)^{-1}.$$

Then,

$$\mathbb{P}\left[S_N = 0\right] \leq e^{\Delta}\prod_{i=1}^{N}\left(1-\mathbb{E}\left[X_i\right]\right),$$

$$\mathbb{P}\left[S_N = 0\right] \geq \left(1-\Delta^* e^{\Delta}\right)\prod_{i=1}^{N}\left(1-\mathbb{E}\left[X_i\right]\right).$$

4 Preliminaries

Let G be a finite group of size $n \geq 3$. Let $a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_k$ be $2k$ random elements chosen independently from G, and let $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_k\}$.

We use the notation $[k] = \{1, 2, \ldots, k\}$.

Let $V = [k] \times [k]$. For $(i, j) \in V$, define

$$I_{(i,j)}(x) = \mathbf{1}_{\{x = a_i \cdot b_j \quad \text{or} \quad x = b_j \cdot a_i\}}.$$

Definition 6. Define a graph $\Gamma = (V, E)$ on the vertex set V, by the edge relation

$$(i,j) \sim (\ell, m) \quad \iff \quad (i = \ell \text{ and } j \neq m) \text{ or } (i \neq \ell \text{ and } j = m).$$

Proposition 7. *For any $x \in G$, Γ is a dependency graph for $\{I_v(x)\}_{v\in V}$.*

Proof. Let S, T be disjoint subsets of V such that $S \not\sim T$. Note that the values of $\{I_v(x)\}_{v\in S}$ are completely determined by $\left\{a_i, b_j \mid (i,j) \in S\right\}$, and the values of $\{I_v(x)\}_{v\in T}$ are completely determined by $\left\{a_i, b_j \mid (i,j) \in T\right\}$. Since $S \not\sim T$ and $S \cap T = \emptyset$, by definition, for any $(i, j) \in S$ and $(\ell, m) \in T$, we have that $i \neq j$ and $j \neq m$. Thus, $\left\{a_i, b_j \mid (i,j) \in S\right\}$ and $\left\{a_\ell, b_m \mid (\ell, m) \in T\right\}$ are independent. So, the families $\{I_v(x)\}_{v\in S}$ and $\{I_v(x)\}_{v\in T}$ are independent. □

Definition 8. Let $x \neq y \in G$. Define $V(x, y) = V \times \{x, y\}$. Let $\Gamma_{x,y} = (V(x, y), E_{x,y})$ be the graph defined by the edge relations

$$(v, z) \sim (u, z') \quad \iff \quad \{v, u\} \in E,$$

for all $v, u \in V$ and $z, z' \in \{x, y\}$.

For $(v,z) \in V(x,y)$, define $J(v,z) = I_v(z)$. The following is very similar to Proposition 7, so we omit the proof.

Proposition 9. *For any $x \neq y \in G$, $\Gamma_{x,y}$ is a dependency graph for $\{I_v(x), I_v(y)\}_{v \in V} = \{J(v,z)\}_{(v,z) \in V(x,y)}$.*

The following Propositions prove to be useful in calculating the moments of $|AB \cup BA|$.

Proposition 10. *Let $x \in G$. Let $v \in V$. Then,*

$$\mathbb{E}\left[I_v(x)\right] = \frac{2}{n}\left(1 - \frac{1}{2} \cdot \frac{|C(x)|}{n}\right).$$

Proof. Let $v = (i,j) \in V$. Since a_i and b_j are independent, by the inclusion-exclusion principle,

$$\begin{aligned} \mathbb{E}\left[I_v(x)\right] &= \mathbb{P}\left[a_i = xb_j^{-1}\right] + \mathbb{P}\left[a_i = b_j^{-1}x\right] - \mathbb{P}\left[a_i = xb_j^{-1} = b_j^{-1}x\right] \\ &= \frac{1}{n} + \frac{1}{n} - \frac{1}{n}\mathbb{P}\left[b_j^{-1} \in C(x)\right]. \qquad \square \end{aligned}$$

Proposition 11. *Let $x, y \in G$. Let $v \in V$ and let $u \sim v$. Then,*

$$\mathbb{E}\left[I_v(x)I_u(y)\right] = \frac{4}{n^2}\left(1 - \frac{|C(x)| + |C(y)|}{2n} + \frac{|C(x) \cap C(y)|}{4n}\right).$$

Proof. Assume that $v = (i,j)$ and $u = (i,\ell)$ for $\ell \neq j$. Conditioning on $a_i = g$,

$$\begin{aligned} \mathbb{E}\left[I_v(x)I_u(y)\right] &= \mathbb{P}\left[(x = a_i \cdot b_j \text{ or } x = b_j \cdot a_i) \text{ and } (y = a_i \cdot b_\ell \text{ or } y = b_\ell \cdot a_i)\right] \\ &= \frac{1}{n}\sum_{g \in G}\mathbb{P}\left[b_j = xg^{-1} \text{ or } b_j = g^{-1}x\right]\mathbb{P}\left[b_\ell = yg^{-1} \text{ or } b_\ell = g^{-1}y\right]. \end{aligned}$$

Considering the four cases: $g^{-1} \in C(x) \cap C(y)$, $g^{-1} \in C(x) \setminus C(y)$, $g^{-1} \in C(y) \setminus C(x)$, $g^{-1} \notin C(x) \cup C(y)$, we get that

$$\begin{aligned} \mathbb{E}\left[I_v(x)I_u(y)\right] &= \frac{1}{n^3} \cdot \left(|C(x) \cap C(y)| + 4(n - |C(x) \cup C(y)|) + 2\,|C(x) \setminus C(y)| + 2\,|C(y) \setminus C(x)|\right) \\ &= \frac{1}{n^3}\left(4n - 2(|C(x)| + |C(y)|) + |C(x) \cap C(y)|\right) \\ &= \frac{4}{n^2}\left(1 - \frac{|C(x)| + |C(y)|}{2n} + \frac{|C(x) \cap C(y)|}{4n}\right). \end{aligned}$$

The case $u = (\ell, j)$ for $\ell \neq i$ is very similar (condition on $b_j = g$). $\square$

Proposition 12. *Let $v \in V$ and let $u \sim v$. Then,*

$$|\{w \in V \;:\; w \sim v\}| = 2(k-1),$$

$$|\{w \in V \;:\; w \sim \{v,u\}\}| = 3(k-1)+1.$$

Proof. Assume that $v = (i,j)$. The first assertion follows from

$$\{w \in V \;:\; w \sim v\} = \{(i,\ell) \;:\; \ell \neq j\} \cup \{(\ell,j) \;:\; \ell \neq i\},$$

since the above union is disjoint.

For the second assertion, assume that $u = (i,\ell)$ for $\ell \neq j$ (the proof for $u = (\ell,j)$ for $\ell \neq i$ is very similar).

$$|\{w \in V \;:\; w \sim \{v,u\}\}| = |\{w \sim v\}| + |\{w \sim u\}| - |\{w \sim u \text{ and } w \sim v\}|.$$

Since

$$\{w \in V \;:\; w \sim u \text{ and } w \sim v\} = \{(i,m) \;:\; m \neq j \text{ and } m \neq \ell\},$$

we get that

$$|\{w \in V \;:\; w \sim \{v,u\}\}| = 4(k-1) - (k-2) = 3(k-1)+1. \qquad \square$$

4.1 Δ and Δ^*

In order to apply Suen's inequality (Theorem 5), we need to calculate Δ and Δ^* as in Theorem 5, for the families of indicators $\{I_v(x)\}_{v\in V}$ and $\{J(v,z)\}_{(v,z)\in V(x,y)}$.

Lemma 13. *Let $x \neq y \in G$.*

(i). Let $\Delta_I(x) = \Delta(\Gamma, \{I_v(x)\}_{v\in V})$ and $\Delta_I^*(x) = \Delta^*(\Gamma, \{I_v(x)\}_{v\in V})$ as in the statement of Theorem 5. Then, $\Delta_I(x)$ and $\Delta_I^*(x)$ are both not larger than $4 \cdot \frac{k^3}{n^2} \cdot \exp\left(\frac{6k}{n-2}\right)$.

(ii). Let $\Delta_J(x,y) = \Delta(\Gamma_{x,y}, \{J(v,z)\}_{(v,z)\in V(x,y)})$ and $\Delta_J^*(x,y) = \Delta^*(\Gamma_{x,y}, \{J(v,z)\}_{(v,z)\in V(x,y)})$. Then, $\Delta_J(x,y)$ and $\Delta_J^*(x,y)$ are both not larger than $16 \cdot \frac{k^3}{n^2} \cdot \exp\left(\frac{12k}{n-2}\right)$.

Proof. By Propositions 10 and 11, for any $v \sim u$, the quantities $\mathbb{E}[I_v(x)I_u(x)]$ and $\mathbb{E}[I_v(x)]\,\mathbb{E}[I_u(x)]$ are bounded by $\frac{4}{n^2}$. By Proposition 12,

$$\prod_{w\sim\{v,u\}} (1 - \mathbb{E}[I_w(x)])^{-1} \leq \left(1 - \frac{2}{n}\right)^{-3k} \leq \exp\left(\frac{6k}{n-2}\right),$$

where we have used the inequality $(1-\frac{1}{\xi})^{-1} \leq \exp\left(\frac{1}{\xi-1}\right)$, valid for any $\xi > 1$.

Plugging this into the definitions of $\Delta_I(x)$ and $\Delta_I^*(x)$ proves the first assertion.

Note that

$$\Delta_J(x,y) = \frac{1}{2}\sum_{v\in V}\sum_{u\sim v}\sum_{z,z'\in\{x,y\}} \mathbb{E}\left[I_v(z)I_u(z')\right] \prod_{w\sim\{v,u\}} \left(1-\mathbb{E}\left[I_w(x)\right]\right)^{-1}\left(1-\mathbb{E}\left[I_w(y)\right]\right)^{-1},$$

and

$$\Delta_J^*(x,y) = \frac{1}{2}\sum_{v\in V}\sum_{u\sim v}\sum_{z,z'\in\{x,y\}} \mathbb{E}\left[I_v(z)\right]\mathbb{E}\left[I_u(z')\right] \prod_{w\sim\{v,u\}} \left(1-\mathbb{E}\left[I_w(x)\right]\right)^{-1}\left(1-\mathbb{E}\left[I_w(y)\right]\right)^{-1}.$$

So, as above, the second assertion follows from

$$\sum_{z,z'\in\{x,y\}} \mathbb{E}\left[I_v(z)I_u(z')\right] \leq \frac{16}{n^2} \quad \text{and} \quad \sum_{z,z'\in\{x,y\}} \mathbb{E}\left[I_v(z)\right]\mathbb{E}\left[I_u(z')\right] \leq \frac{16}{n^2}. \qquad \square$$

5 Bounds on $|AB \cup BA|$

In this section we provide bounds on the probability of the event that $\{AB \cup BA = G\}$, i.e. that A and B are a decomposition of G. Let $S = G \setminus AB \cup BA$. Thus, $AB \cup BA = G$ if and only if $|S| = 0$. To bound the required probabilities, we bound the first and second moments of $|S|$.

Lemma 14. *Let $0 \leq \psi < \log n$, and let $k \geq \sqrt{\Theta(G)n(\log n + \psi)}$. Then,*

$$\mathbb{P}\left[AB \cup BA \neq G\right] \leq (1+o(1)) \cdot e^{-\Theta(G)\psi}.$$

Proof. Since, by Markov's inequality,

$$\mathbb{P}\left[AB \cup BA \neq G\right] = \mathbb{P}\left[|S| \geq 1\right] \leq \mathbb{E}\left[|S|\right],$$

it suffices to bound $\mathbb{E}\left[|S|\right]$.

Note that the event $\mathbb{P}\left[AB \cup BA = G\right]$ is monotone non-decreasing with k, so we can assume that $k = \lceil\sqrt{\Theta n(\log n + \psi)}\rceil$, where $\Theta = \Theta(G)$.

Now, $x \in S$ if and only if $\sum_{v\in V} I_v(x) = 0$. By Lemma 13,

$$\Delta_I = \Delta_I(x) = O\left(k^3/n^2\right) = o(1).$$

Thus, using Suen's inequality (Theorem 5), for any $x \in G$,

$$\begin{aligned} \mathbb{P}[x \in S] &\leq e^{\Delta_I} \cdot \left(1 - \frac{2}{n}\left(1 - \frac{1}{2} \cdot \frac{|C(x)|}{n}\right)\right)^{|V|} \\ &\leq (1+o(1)) \cdot \exp\left(-\frac{2k^2}{n} + \frac{k^2\,|C(x)|}{n^2}\right). \end{aligned}$$

Summing over all $x \in G$, we get

$$\begin{aligned} \mathbb{E}[|S|] &\leq (1+o(1)) \exp\left(-\frac{2k^2}{n}\right) \cdot \sum_{x \in G} \exp\left(\frac{k^2\,|C(x)|}{n^2}\right) \\ &\leq (1+o(1) \cdot \exp\left(-2\Theta\psi - 2\Theta\log n\right) \sum_{x \in G} \exp\left(\Theta \log n \frac{|C(x)|}{n}\right) \exp\left(\Theta\psi \frac{|C(x)|}{n}\right) \\ &\leq (1+o(1)) \cdot \exp\left(-\Theta\psi\right). \end{aligned}$$

□

Lemma 15. *Let* $0 \leq \psi < \log n$*, and let* $k \leq \sqrt{\Theta(G)n(\log n - \psi)}$*. Then,*

$$\mathbb{P}[AB \cup BA = G] \leq e^{-\Theta(G)\psi} + o(1).$$

Proof. As in the proof of Lemma 14, we can assume that $k = \lfloor\sqrt{\Theta n(\log n - \psi)}\rfloor$, for $\Theta = \Theta(G)$.

We can bound the moments of $|S|$ using Suen's inequality, as in the proof of Lemma 14. To simplify the notation we will use $p_x = (1 - \mathbb{E}[I_v(x)])^{k^2}$ (which does not depend on v, by Proposition 10). By our choice of k, since $\Delta_I(x) = o(1)$ and $\Delta_I^*(x) = o(1)$, $\mathbb{P}[x \in S] \geq (1 - o(1)) \cdot p_x$ and $\mathbb{P}[x \in S] \leq (1 + o(1)) \cdot p_x$. Thus,

$$\mathbb{E}[|S|] \geq (1 - o(1)) \cdot \sum_{x \in G} p_x.$$

Furthermore, note that for $x \neq y \in G$, since $\Delta_J(x, y) = o(1)$,

$$\mathbb{P}[x, y \in S] \quad \leq \quad (1+o(1)) \cdot \prod_{(v,z) \in V(x,y)} (1 - \mathbb{E}[J(v,z)]) = (1+o(1)) \cdot p_x p_y.$$

Hence,

$$\begin{aligned} \mathbb{E}\left[|S|^2\right] &= \sum_{x \neq y \in G} \mathbb{P}[x, y \in S] + \sum_{x \in G} \mathbb{P}[x \in S] \\ &\leq (1+o(1)) \cdot \sum_{x \neq y \in G} p_x p_y + (1+o(1)) \cdot \sum_{x \in G} p_x. \end{aligned}$$

Now we use the Paley-Zygmund inequality:

$$\begin{aligned}\mathbb{P}\left[AB \cup BA \neq G\right] &= \mathbb{P}\left[|S| > 0\right] \geq \frac{\left(\mathbb{E}\left[|S|\right]\right)^2}{\mathbb{E}\left[|S|^2\right]} \\ &\geq (1-o(1)) \cdot \frac{\left(\sum_x p_x\right)^2}{\sum_{x \neq y} p_x p_y + \sum_x p_x} = (1-o(1)) \cdot \left(1 - \frac{\sum_x p_x - \sum_x p_x^2}{\sum_{x \neq y} p_x p_y + \sum_x p_x}\right) \\ &= (1-o(1)) \cdot \left(1 - \frac{\sum_x p_x (1-p_x)}{\sum_x p_x \left(1 - p_x + \sum_y p_y\right)}\right)\end{aligned}$$

So it suffices to show that for all $x \in G$,

$$\frac{1-p_x}{1-p_x+\sum_y p_y} \leq (1+o(1)) \cdot e^{-\Theta\psi}.$$

But this follows immediately from

$$\frac{1-p_x}{1-p_x+\sum_y p_y} \leq \left(\sum_y p_y\right)^{-1},$$

and from the fact that

$$\begin{aligned}\sum_{y \in G} p_y &\geq \sum_{y \in G} \exp\left(-\frac{2k^2}{n-2} \cdot \left(1 - \frac{|C(y)|}{2n}\right)\right) \\ &= \sum_{y \in G} \exp\left(-\frac{2k^2}{n} \cdot \left(1 - \frac{|C(y)|}{2n}\right) - \frac{4k^2}{n(n-2)} \cdot \left(1 - \frac{|C(y)|}{2n}\right)\right) \\ &\geq (1-o(1)) \cdot \exp\left(2\Theta\psi - 2\Theta\log n\right) \cdot \sum_{y \in G} \exp\left(\Theta \log n \frac{|C(y)|}{n}\right) \exp\left(-\Theta\psi \frac{|C(y)|}{n}\right) \\ &\geq (1-o(1)) \cdot \exp\left(\Theta\psi\right),\end{aligned}$$

(where we have used the inequality $1 - \frac{1}{\xi} \geq \exp\left(-\frac{1}{\xi-1}\right)$, valid for any $\xi > 1$). □

6 Concluding Remarks and Open Problems

- One can ask whether a result similar to Theorem 1 holds if we only require that $AB = G$. It can be shown that for any finite group G of order $|G| = n$, if $k \geq (1+\varepsilon)\sqrt{n \log n}$ then $\mathbb{P}\left[AB = G\right] = 1 - o(1)$, and if $k \leq (1-\varepsilon)\sqrt{n \log n}$, then $\mathbb{P}\left[AB = G\right] = o(1)$. The proof of this is almost identical to the proof of Theorem 1.

- Another variant is to take A and B of different sizes. That is, let $a_1, \ldots, a_k$ and $b_1, \ldots, b_m$ be $k + m$ random elements of G, and let $A = \{a_1, \ldots, a_k\}$ and $B =$

$\{b_1, \ldots, b_m\}$. What can be said about the probability $\mathbb{P}[AB \cup BA = G]$? It turns out that if k and m are both not too large, then the threshold is identical to the case $m = k$. That is, provided that $\max\{k, m\} = o\left(\frac{|G|}{\log|G|}\right)$, we can prove that $\mathbb{P}[AB \cup BA = G] = 1 - o(1)$ if $k \cdot m \geq (1+\varepsilon)\Theta(G) n \log n$ and if $k \cdot m \leq (1-\varepsilon)\Theta(G) n \log n$, then $\mathbb{P}[AB \cup BA = G] = o(1)$. Again, the proof is the same as that of Theorem 1.

- We can also ask what is the probability of the event $AA = G$. In this case, our method breaks down for groups G such that $\Theta(G)$ is very small. That is, we can prove a phase transition in k for the event $\{AA = G\}$, but only for families of groups $\{G_n\}$, such that $\Theta(G_n) \geq \frac{1}{2} + \frac{\log\log|G_n|}{\log|G_n|}$. Note that in Section 2 it is shown that there are groups (e.g. the symmetric group) that do not have this property. The main problem in dealing with AA, is that one needs to control the size of the set $\left\{a^2 \ : \ a \in A\right\}$. This means controlling the probability $\mathbb{P}\left[a_i^2 = x\right]$ for all x. Thus, we have the following

  Open Problem. Prove or provide a counter-example:

  Let G_n be a family of groups such that
$$\lim_{n\to\infty} |G_n| = \infty.$$
  For all n, let $a_1, a_2, \ldots, a_k$ be k randomly chosen elements of G_n, and let $A = \{a_1, a_2, \ldots, a_k\}$. Let $P'(n,k) = \mathbb{P}[AA = G_n]$.

  For all n, let $C_n = \sqrt{2\Theta(G_n)|G_n|\log|G_n|}$. Then for any $\varepsilon > 0$,
$$\lim_{n\to\infty} P'(n, \lceil(1+\varepsilon)C_n\rceil) = 1 \ , \quad \lim_{n\to\infty} P'(n, \lfloor(1-\varepsilon)C_n\rfloor) = 0.$$

- Another interesting problem, is to determine what happens inside the transition window: As can be seen by Lemmas 14 and 15, if $\psi(n)$ is any function tending to infinity with n, then for $k \geq \sqrt{\Theta(G) n \log n} + \sqrt{n\psi(n)}$, with high probability $AB \cup BA = G$. For $k \leq \sqrt{\Theta(G) n \log n} - \sqrt{n\psi(n)}$, with high probability $AB \cup BA \neq G$.

  The question is, what happens for $\sqrt{\Theta(G) n \log n} - \sqrt{n} < k < \sqrt{\Theta(G) n \log n} + \sqrt{n}$? What can be said about the size of $AB \cup BA$ in this case?

- Here are some further open questions, proposed by Itai Benjamini:

  Let G be a finite group. Consider the family of subsets
$$\mathbb{S} = \left\{B \subset G \;\middle|\; \exists\ A \subset G \ : \ AA = B\right\}.$$
  (i). Determine the size of $\mathbb{S}$.

(ii). Sample $B \in \mathbb{S}$ from the uniform distribution.

(iii). Devise an (efficient) algorithm to decide whether a subset $A \subset G$ is in $\mathbb{S}$ or not.

(iv). Devise an (efficient) algorithm to decide whether $A \subset G$ is "almost" an element of $\mathbb{S}$; i.e. whether there exists $B \in \mathbb{S}$ such that $|A \triangle B| = o(|G|)$.

It will be interesting to solve some of these problems even with relaxed conditions, such as assuming that G is abelian or even cyclic.

Acknowledgements. I wish to thank Itai Benjamini for suggesting this problem and useful discussions. I thank Nir Avni for his comments on a preliminary version of this paper, his ideas regarding the group invariant Θ, and for bringing the papers by Liebeck and Shalev to my attention. I thank Gady Kozma for useful discussions and his insights into simplifying some of the proofs.

References

[1] S. Janson, New versions of Suen's correlation inequality. *Random Structures and Algorithms* **13** (1998), 467–483.

[2] W.M. Kantor, A. Lubotzky, The probability of generating a finite classical group. *Geom. Dedicata* **36** (1990), 67–87.

[3] G. Kozma, A. Lev, Bases and decomposition numbers of finite groups. *Arch. Math.* **58** (1992), 417–424.

[4] M.W. Liebeck, A. Shalev, Classical groups, probabilistic methods, and the (2,3)-generation problem. *Annals of Math.* **144** (1996), 77–125.

[5] M.W. Liebeck, A. Shalev, On conjugacy classes of maximal subgroups of finite simple groups, and a related zeta function. *Duke Math. J.* **128** (2005), 541–557.

[6] J.J Rotman, *An Introduction to the Theory of Groups*, fourth edition, Springer-Verlag, New York, (1995).

[7] W.C.S. Suen, A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures and Algorithms*, **1** (1990), 231–242.