# A Note on Coloring Random $k$-Sets

Noga Alon *    Joel H. Spencer †

In a recent elegant paper, Chvátal and Reed [2] consider the satisfiability of a conjunction of random $k$-clauses. A random $k$-clause is the disjunction $y_1 \vee \ldots \vee y_k$ where the $y_i$ are chosen uniformly and independently from atoms $x_1, \ldots, x_n$ and their negations $\overline{x}_1, \ldots, \overline{x}_n$. They set $C = C_1 \wedge \ldots \wedge C_m$ where each $C_i$ is an independently chosen random $k$-clause. Their most interesting result is with $k$ large and $n$ approaching infinity. They show that if $n < c\frac{2^k}{k}n$ then almost surely $C$ is satisfiable while if $n > c'2^k n$ then almost surely $C$ is not satisfiable. Here $c, c'$ are absolute constants and almost surely refers to asymptotics in $n$ for any fixed sufficiently large $k$.

We consider here the analogous problem for 2-colorability of random families. Let $A_1, \ldots, A_m$ be uniformly and independently chosen $k$-sets from the family of $k$-element subsets of $\Omega = \{1, \ldots, n\}$. Set $F = \{A_1, \ldots, A_m\}$. $F$ is 2-colorable (the term "has Property B" is equivalent) if there exists $\chi : \Omega \rightarrow \{Red, Blue\}$ so that no $A \in F$ is monochromatic. Note the analogy. True/False correspond to Red/Blue. Truth evaluation corresponds to 2-Coloring. $C_i = y_{i1} \vee \ldots \vee y_{ik}$ is satisfied if some $y_{ij} \leftarrow True$. $A_i = \{a_{i1}, \ldots, a_{ik}\}$ is nonmonochromatic if some $a_{ij} \leftarrow Red$ and some $a_{ij'} \leftarrow Blue$. $C$ is satisfied if all $C_i$ are; $F$ is 2-colored if all $A_i$ are nonmonochromatic.

Surprisingly, we have not been able to duplicate the Chvátal-Reed result. Again we think of $k$ large and $n \rightarrow \infty$. If $m > c'2^k n$ we show that $F$ is almost surely not 2-colorable. In the other direction we only have that if $m < c\frac{2^k}{k^2}n$ then $F$ is almost surely 2-colorable. The $c, c'$ are again explicit positive absolute constants, though not the same constants as in the Chvátal-Reed case.

*The Upper Bound.* The proof resembles the one in [3]. Suppose $m \sim c'2^k n$ with $c' > \frac{\ln 2}{2}$. There are $2^n$ potential colorings of $\Omega = \{1, \ldots, n\}$. Fix a coloring $\chi$ with $a$ Red and $b = n - a$ Blue points. The probability of a random $k$-set $A$ being monochromatic is then

$$\frac{\binom{a}{k} + \binom{b}{k}}{\binom{n}{k}} \geq \frac{2\binom{n/2}{k}}{\binom{n}{k}} \sim 2^{1-k}$$

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Research supported in part by a United States Israel BSF Grant

†Department of Mathematics and Computer Science, Courant Institute, NYU, NY, USA. Research supported in part by a United States Israel BSF Grant

The probability that none of the $A_1, \ldots, A_m$ are monochromatic is then

$$(1 - 2^{1-k} - o(1))^m = o(2^{-n})$$

by our choice of $m$. The expected number of valid 2-colorings is then $o(1)$ and so almost surely there aren't any of them. $\square$

*The Lower Bound.* Here we use the Lovász Local Lemma proved in [4] (see also, e.g., [1], Chapter 5), though it requires some interesting preparation. Suppose $m \sim c\frac{2^k}{k^2}n$ with $c < \frac{1}{4e}$. (Note that a set $A_i$ on average intersects $c2^k$ other sets. If every set had that number of intersections we could directly apply the Lovász Local Lemma even with $c < \frac{1}{2e}$.) Say $\epsilon > 0$ has $c(1+\epsilon) < \frac{1}{4e}$. Fix $\delta$ with $0 < \delta < \epsilon$. Call $x \in \Omega$ *special* if $deg(x) > c\frac{2^k}{k}(1+\delta)$, where here $deg(x)$ is the number of $A \in F$ with $x \in A$. Define $deg^+(x)$ to be $deg(x)$ if $x$ is special, otherwise zero. We know $deg(x)$ has Binomial distribution $B[c\frac{2^k}{k}n, \frac{k}{n}]$, asymptotically Poisson with mean $\mu = c2^k$. Large Deviation bounds (see, e.g., [1], Appendix A), give $E[deg^+(x)] = o(1)$, asymptotics in $k$. (In fact, that expectation is hyperexponentially small in $k$). Pick $k$ large enough (as a a function of $c$ and $\delta$ only) so that $E[deg^+(x)] \leq \frac{1}{10}$. Set $X = \sum_{x\in\Omega} deg^+(x)$ so that by Linearity of Expectation $E[X] \leq \frac{1}{10}n$. For $k$ fixed and $n \to \infty$ one can show $Var[X] = O(n) = o(n^2)$ as, basically, the $deg^+(x)$ have small, indeed negative, correlation. By Chebyschev's Inequality $X \leq \frac{1}{8}n$ almost surely. Call $A$ *special* if it contains any special $x$. The number of special $A$ is bounded by $X$ so almost surely there are fewer than $\frac{1}{8}n$ special $A$.

Now we claim that almost surely we can select $x_i, y_i \in A_i$ for each special $A_i \in F$ so that all elements selected are distinct. By Hall's Theorem it suffices to show that

$$|A_1 \cup \ldots \cup A_l| \geq 2l$$

for every subfamily $A_1, \ldots, A_l$ of special sets. But then it suffices to show the above holds for any $A_1, \ldots, A_l \in F$ with $1 \leq l \leq \frac{1}{8}n$. The probability of this failing is bounded by

$$\sum_{l=1}^{n/8} \binom{m}{l}\binom{n}{2l-1}\left(\frac{\binom{2l-1}{k}}{\binom{n}{k}}\right)^l$$

as we may fix $A_1, \ldots, A_l \in F$, $T \subset \Omega$ with $|T| = 2l-1$ and then the probability of $A_1 \cup \ldots \cup A_l \subseteq T$ is simply the $l$-th power of the probability that each $A_i \subseteq T$ which is precisely the expression in parenthesis. We replace $2l-1$ by $2l$ for convenience. Now we apply standard bounds

$$\binom{a}{b} \leq \left(\frac{ae}{b}\right)^b \text{ and } \frac{\binom{2l}{k}}{\binom{n}{k}} \leq \left(\frac{2l}{n}\right)^k$$

so that the above sum is at most

$$\sum_{l=1}^{n/8}\left(\frac{me}{l}\right)^{l}\left(\frac{ne}{2l}\right)^{2l}\left(\frac{2l}{n}\right)^{lk}$$

We select $k$ large enough so that $\frac{mn^2e^3}{4} < 2^{k-6}n^3$. Then the sum is at most

$$\sum_{l=1}^{n/8}\left[\left(\frac{4l}{n}\right)^{k-3}\right]^{l}$$

which is clearly $o(1)$.

Now consider an $F = \{A_1, \ldots, A_m\}$ with this property and fix the $x_i, y_i \in A_i$ for special $A_i$. Consider a random coloring of $\Omega$ in which $x_i, y_i$ are paired: $\chi(x_i)$ is chosen Red or Blue with probability .5 and then $\chi(y_i)$ is set equal to the other color. The remaining $z \in \Omega$ are independently colored Red or Blue with probability .5. In this probability space special $A_i$ are never monochromatic. For $A_i$ nonspecial let $B_i$ be the event that $A_i$ is monochromatic. Then either $\Pr[B_i] = 0$ (if $A_i$ happens to contain a pair $x_j, y_j$) or $\Pr[B_i] = 2^{1-k}$. Define a dependency graph on these $B_i$. $B_i, B_j$ are dependent if either $A_i \cap A_j \neq \emptyset$ or there is a pair $x_k, y_k$ with $x_k \in B_i$, $y_k \in B_j$. $A_i$ has $k$ elements and so at most $2k$ elements and members of pairs of elements. No element can be in more than $c\frac{2^k}{k}(1+\delta)$ nonspecial sets as otherwise those sets would be special. Thus the maximal degree $D$ of the dependency graph and the maximal probability $p$ of the $B_i$ satisfy

$$D + 1 \leq (2k)c\frac{2^k}{k}(1+\delta) = 2c(1+\delta)2^k \quad \text{and} \quad p \leq 2^{1-k}.$$

As

$$ep(D+1) \leq 2^{2-k}ce(1+\delta)2^k < \frac{1+\delta}{1+\epsilon} < 1$$

the Lovász Local Lemma gives $\wedge\overline{B}_i \neq \emptyset$, hence there is a two-coloring in the probability space for which no $A \in F$ is monochromatic. □

Where lies the truth? As with the random clause situation we do not have a concentration result. Analogous to it we conjecture that for each $k$ there exists a $c_k$ so that for any $\epsilon > 0$ if $m > c_k n(1+\epsilon)$ then almost surely $F$ is not 2-colorable while if $m < c_k n(1-\epsilon)$ then almost surely $F$ is 2-colorable. Even without this we can define $c_k^-$ as the supremum of those $c$ so that if $m \leq cn$ then almost surely $F$ is 2-colorable and define $c_k^+$ as the infimum of those $c$ so that if $m \geq cn$ then almost surely $F$ is not 2-colorable. Our results give

$$\Omega\left(\frac{2^k}{k^2}\right) = c_k^- \leq c_k^+ = O(2^k)$$

The correct order of magnitude remains elusive, though it seems plausible to suspect that $c_k^- = c_k^+ = \Theta(2^k)$.

Working by analogy to the Chvátal-Reed result we offer a randomized algorithm and a conjecture which, if true, would give $c_k^- = \Omega(\frac{2^k}{k})$.

ALGORITHM CR

*Input:* Sets $A_1, \dots, A_m \subseteq \Omega = \{1, \dots, n\}$, all of size $k$.

*Output:* Either a 2-coloring $\chi$ of $\Omega$ with no $A_i$ monochromatic or failure.

*Description:* The algorithm has $n = |\Omega|$ rounds. In each round one $x \in \Omega$ is colored. At each round the algorithm does the following.

(0) If some $A \in F$ is already monochromatic return failure.

(1) If not (0), but some $A \in F$ has $k-1$ points of one color, say Red, and its last point $x$ uncolored then the algorithm colors $x$ the "other" color, Blue. If there are several such $A$ the algorithm selects one at random and colors its $x$ accordingly. Note: if the $x$ so colored lies in one $A$ that is otherwise Red and one that is otherwise Blue then this will lead to (0) in the next round.

(2) If not (0) and not (1) but some $A \in F$ has $k-2$ points of one color, say Red, and its two other points $x, y$ uncolored then the algorithm selects $x$ or $y$ randomly and colors it in the "other" color Blue. If there are several such $A$ it first picks one such $A$ at random and then acts as above. Note: if the $x$ so colored is in two $A$, one nearly Red and the other nearly Blue, this leads to (1) in the next round.

(3) If not (0), not (1) and not (2) then pick randomly an uncolored point $x \in \Omega$ and color it randomly.

It is not difficult to see that this algorithm runs in polynomial time. Indeed, with the right data structure it runs in linear time in $n$ for fixed $c, k$. The question remains: how often does it lead to failure?

*Conjecture:* There exists $c > 0$ and $k_0$ so that for $k > k_0$ the following holds: For $m \sim c\frac{2^k}{k}n$ and for $F = \{A_1, \dots, A_m\}$ with $A_i$ chosen uniformly and independently from the $k$-sets of $\Omega = \{1, \dots, n\}$ the algorithm CR almost surely produces a coloring $\chi$ of $\Omega$ with no $A \in F$ monochromatic.

# References


[1] N. Alon and J. H. Spencer, **The Probabilistic Method**, Wiley, 1991.

[2] V. Chvátal and B. Reed, *Mick gets some (the odds are on his side)*, Proc. $33^{rd}$ IEEE FOCS, IEEE, 1992, 620-627.

[3] P. Erdős, *On a combinatorial problem, II*, Acta Math. Acad. Sci. Hungar. 15 (1964), 445-447.

[4] P. Erdős and L. Lovász, *Problems and results on* 3*-chromatic hypergraphs and some related questions*, in:"Infinite and Finite Sets" (A. Hajnal et. al. eds), Colloq. Math. Soc. J. Bolyai 11, North Holland, Amsterdam, 1975, pp. 609-627.