# How to Compress the Solution

Samuel Epstein[*]

September 23, 2022


## Abstract

We provide an upper bound on the compression size of solutions to the graph coloring problem. In general, if solutions to a combinatorial problem exist with high probability and the probability is simple, then there exists a simple solution to the problem. Otherwise the problem instance has high mutual information with the halting problem.


## Results

$\mathbf{K}(x|y)$ is the conditional prefix Kolmogorov complexity. Algorithmic probability is $\mathbf{m}(x) = \{2^{-\|p\|} : U(p) = x\}$, where $U$ is the universal Turing machine. For set $D \subseteq \{0,1\}^*$, computable probability $P$, $O(1)\mathbf{m}(D) > 2^{-\mathbf{K}(P)}P(D)$. $\mathbf{I}(a;\mathcal{H}) = \mathbf{K}(a) - \mathbf{K}(a|\mathcal{H})$, where $\mathcal{H}$ is the halting sequence. $<^+ f$ is $<f+O(1)$ and $<^{\log} f$ is $<f+O(\log(f+1))$.

**Lemma 1 ([Eps22])** *For partial computable $f : \mathbb{N} \to \mathbb{N}$, for all $a \in \mathbb{N}$, $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.*

**Theorem 1 ([Lev16, Eps19])**
*For finite $D \subset \{0,1\}^*$, $-\log \max_{x\in D} \mathbf{m}(x) <^{\log} -\log \sum_{x\in D} \mathbf{m}(x) + \mathbf{I}(D;\mathcal{H})$.*

For graph $G = (V,E)$, with undirected edges, a $k$-coloring is a function $f : V \to \{1,\ldots,k\}$ such that if $(v,u) \in E$, then $f(v) \neq f(u)$.

**Theorem 2** *For graph $G = (V,E)$, $|V| = n$ with max degree $d$, there is a $k$ coloring $f$, where for small enough $d/k$, with $\mathbf{K}(f) <^{\log} \mathbf{K}(n,k) - n\log(1-d/k) + \mathbf{I}((G,k);\mathcal{H})$.*

**Proof.** Let us say we randomly assign a color to each vertex. The probability that the color of the $i$th vertex does not conflict with the previous coloring is at least $(k-d)/k$. Thus the probability of a proper coloring is $\geq ((k-d)/k)^n$. Let $D \subseteq \{0,1\}^{n\lceil \log k\rceil}$ be all encoded proper $k$ colorings of $G$. $\mathbf{K}(D|G,k) = O(1)$. Let $P : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ be a probability measure that is the uniform distribution over all possible color assignments. Thus

$$-\log P(D) < -n\log(1-d/k).$$

Thus by Theorem 1 and Lemma 1, there is a coloring $f \in D$ with

$$\begin{aligned}\mathbf{K}(f) &<^{\log} -\log \mathbf{m}(D) + \mathbf{I}(D;\mathcal{H})\\ &<^{\log} \mathbf{K}(P) - \log P(D) + \mathbf{I}(D;\mathcal{H})\\ &<^{\log} \mathbf{K}(n,k) - n\log(1-d/k) + \mathbf{I}((G,k);\mathcal{H}).\end{aligned}$$

□

## References


[Eps19] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps22] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016.


[*]JP Theory Group. samepst@jptheorygroup.org