Rainbow saturation and graph capacities

Dániel Korándi *

Abstract

The $t$-colored rainbow saturation number $\text{rsat}_t(n, F)$ is the minimum size of a $t$-edge-colored graph on $n$ vertices that contains no rainbow copy of $F$, but the addition of any missing edge in any color creates such a rainbow copy. Barrus, Ferrara, Vandenbussche and Wenger conjectured that $\text{rsat}_t(n, K_s) = \Theta(n \log n)$ for every $s \geq 3$ and $t \geq \binom{s}{2}$. In this short note we prove the conjecture in a strong sense, asymptotically determining the rainbow saturation number for triangles. Our lower bound is probabilistic in spirit, the upper bound is based on the Shannon capacity of a certain family of cliques.

1 Introduction

A graph $G$ is called $F$-saturated if it is a maximal $F$-free graph. The classic saturation problem, first studied by Zykov [14] and Erdős, Hajnal and Moon [4], asks for the minimum number of edges in an $F$-saturated graph (as opposed to the Turán problem, which asks for the maximum number of edges in such a graph). A rainbow analog of this problem was recently introduced by Barrus, Ferrara, Vandenbussche and Wenger [1], where a $t$-edge-colored graph is defined to be rainbow $F$-saturated if it contains no rainbow copy of $F$ (i.e., a copy of $F$ where all edges have different colors), but the addition of any missing edge in any color creates such a rainbow copy. Then the $t$-colored rainbow saturation number $\text{rsat}_t(n, F)$ is the minimum size of a $t$-edge-colored rainbow $F$-saturated graph.

Among other results, Barrus et al. showed that $\Omega \left( \frac{n \log n}{\log \log n} \right) \leq \text{rsat}_t(n, K_s) \leq O(n \log n)$ and conjectured that their upper bound is of the right order of magnitude:

Conjecture 1.1 ([1]). For $s \geq 3$ and $t \geq \binom{s}{2}$, $\text{rsat}_t(n, K_s) = \Theta(n \log n)$.

Here we prove this conjecture in a strong sense: we give a lower bound that is asymptotically tight for triangles.

Theorem 1.2. For $s \geq 3$ and $t \geq \binom{s}{2}$, we have

$$\text{rsat}_t(n, K_s) \geq \frac{t(1 + o(1))}{(t - s + 2) \log(t - s + 2)} n \log n$$

with equality for $s = 3$.

We should point out that Conjecture 1.1 was independently verified by Girão, Lewis and Popielarz [9] and by Ferrara et al. [5], but with somewhat weaker bounds. In fact, our result proves a conjecture in [9], establishing the stronger estimate $\text{rsat}_t(n, K_s) = \Theta_s \left( \frac{n \log n}{\log t} \right)$ with their upper bound.

*Institute of Mathematics, EPFL, Lausanne, Switzerland. Research supported in part by SNSF grants 200020-162884 and 200021-175977. Email: daniel.korandi@epfl.ch.
Our lower bound is probabilistic in spirit, using ideas of Katona and Szemerédi [10], and Füredi, Horak, Pareek and Zhu [6] (similar techniques were used in [12] 2 [11]). The upper bound for \( s = 3 \) is based on the following theorem that follows from a strong information-theoretic result of Gargano, Körner and Vaccaro [8] on the Shannon capacities of graph families.

**Theorem 1.3.** For every \( t \geq 3 \), there is a set \( X \subseteq [t]^k \) of \( m = (t - 1)^{(1 + o(1))} \) strings of length \( k \) from alphabet \( [t] = \{1, \ldots, t\} \) such that for any \( x, x' \in X \) and any \( a \in [t] \), there is a position \( i \) where \( x(i) \neq x'(i) \) and \( x(i), x'(i) \neq a \).

In the next section we derive Theorem 1.3 from results about the Shannon capacity of graph families. This is followed by the proof of Theorem 1.2 in Section 3.

## 2 Graph capacities

Let \( \mathcal{G} = \{G_1, \ldots, G_r\} \) be a family of graphs on vertex set \( [t] \). Let \( N_k \) be the maximum size of a set \( X \subseteq [t]^k \) of strings of length \( k \) on alphabet \( [t] \) such that for any two strings \( x, x' \in X \) and any \( G_j \in \mathcal{G} \), there is a position \( i \in [k] \) such that \( x(i) = x'(i) \) is an edge in \( G_j \). The **Shannon capacity** of the family \( \mathcal{G} \) is defined as \( C(\mathcal{G}) = \lim_{k \to \infty} \frac{1}{k} \log N_k \) (see, e.g., [13, 3]). When \( \mathcal{G} = \{G\} \), we simply write \( C(G) \) for \( C(\mathcal{G}) \).

We need an analogous definition for strings where the occurrences of each \( a \in [t] \) are proportional to some probability measure \( P \) on \( [t] \). So let \( T^k(P, \varepsilon) \) be the set of all strings \( x \in [t]^k \) such that \( |\frac{1}{k} \#\{i : x(i) = a\} - P(a)| < \varepsilon \) for every \( a \in [t] \), and let \( M_{k, \varepsilon} \) be the maximum size of a set \( X \subseteq T^k(P, \varepsilon) \) such that for every \( x, x' \in X \) there is an \( i \) with \( x(i)x'(i) \in G \). The Shannon capacity within type \( P \) is \( C(G, P) = \lim_{\varepsilon \to 0} \limsup_{k \to \infty} \frac{1}{k} \log M_{k, \varepsilon} \). Using a clever construction, Gargano, Körner and Vaccaro [8] showed that \( C(\mathcal{G}) \) can be expressed in terms of the \( C(G_j, P) \):

**Theorem 2.1** ([8]). For a family of graphs \( \mathcal{G} = \{G_1, \ldots, G_r\} \) on vertex set \( [t] \), we have

\[
C(\mathcal{G}) = \max_{P} \min_{G_j \in \mathcal{G}} C(G_j, P).
\]

In fact, they proved a more general result for **Sperner capacities**, the analogous notion for directed graphs. What we need is a corollary that follows easily from this theorem using standard tools about graph entropy (see the survey of Simonyi [13] for more information). Here we give a self-contained argument that goes along the lines of a proof by Gargano, Körner and Vaccaro [7] of the case \( s = 2 \).

**Corollary 2.2.** Let \( 2 \leq s \leq t \) be an integer and let \( \mathcal{G} \) be the family of all \( s \)-cliques on \( [t] \) (each with \( t - s \) isolated vertices). Then \( C(\mathcal{G}) = \frac{s}{t} \log s \).

**Proof.** For the lower bound, we can take \( P \) to be the uniform measure on \( [t] \). Then by Theorem 2.1 it is enough to show that \( C(G, P) \geq \frac{s}{t} \log s \) where \( G \) is a clique on \( [s] \) with isolated vertices \( s + 1, \ldots, t \). Let \( X_k \subseteq T^k(P, \frac{1}{k}) \) be the set of all strings \( x \) of length \( k \) such that the first \( \lfloor sk/t \rfloor \)

\(^1\)The usual definition is with binary logarithm, but the base of our logarithms is unimportant for our purposes.
letters of $x$ contain $\lfloor k/t \rfloor$ or $\lceil k/t \rceil$ instances of each $a \in [s]$, and $x(i) = b$ for every $s + 1 \leq b \leq t$ and \( \frac{(b-1)k}{t} < i \leq \frac{bk}{t} \). Then

\[
C(G, P) \geq \lim_{k \to \infty} \frac{\log(X_k)}{k} = \lim_{k \to \infty} \frac{1}{k} \log \frac{(\frac{s^k}{t})^t}{(\frac{k}{t})^s} = \lim_{k \to \infty} \frac{1}{k} \log(s^{sk/t}) = \frac{s}{t} \log s.
\]

For the upper bound, let $X \subseteq [t]^k$ be a maximum set of strings such that for any $x, x' \in X$ and for every $s$-clique $G \in \mathcal{G}$, there is an $i \in [k]$ such that $x(i)x'(i) \in E$. We set $m = |X|$ to be this maximum. We may assume that $\{1, \ldots, s\}$ are the $s$ least frequent elements appearing in the strings of $X$. Let $d_x$ be the number of elements in $x \in X$ that are not in $[s]$, so $\sum_{x \in X} d_x \geq \frac{k-2} kn$, and let $X_x$ be the set of strings obtained from $x$ by replacing these elements arbitrarily with numbers from $[s]$. Then $|X_x| = s^{d_x}$, and $X_x, X_{x'}$ are disjoint for distinct $x, x' \in X$ because any string from $X_x$ will differ from any string in $X_{x'}$ at the position $i$ where $x(i)x'(i)$ is an edge of the clique on $[s]$. Then using Jensen’s inequality we have

\[
s^k \geq \sum_{x \in X} s^{d_x} \geq m \cdot s^{\left(\sum_{x \in X} d_x\right)/m} \geq m \cdot s^{\left(\frac{t-s}{t}\right)^k},
\]

and hence $m \leq s^{sk/t}$, implying $C(\mathcal{G}) \leq \frac{1}{t} \log m \leq \frac{s}{t} \log s$. \(\square\)

Theorem 1.3 clearly follows from the case $s = t - 1$.

### 3 Rainbow saturation

**Proof of Theorem 1.4** For the lower bound, suppose $H$ is a $t$-edge-colored rainbow $K_s$-saturated graph, and split its vertices into two parts: let $A = \{a_1, \ldots, a_k\}$ be the set of vertices of degree at least $d = \log^3 n$, and $B$ be the rest. We may assume $|A| \leq \frac{n}{\log n}$ (otherwise $H$ has at least $\frac{1}{2} n \log^2 n$ edges), and thus $B$ contains $m \geq (1 - \frac{1}{\log n})n$ vertices. Now let us define a string $x_v \subseteq [t + 1]^k$ for every $v \in B$ that encodes the colors of the $A-B$ edges touching $v$ as follows: $x_v(i)$ is $t + 1$ if $a_iv$ is not an edge in $H$, otherwise it is the color of $a_iv$.

Assume, without loss of generality, that $t - s + 3, \ldots, t$ are the $s - 2$ most common colors among the $A-B$ edges. For $v \in B$, let $X_v \subseteq [t - s + 2]^k$ be the set of strings obtained from $x_v$ by replacing each $t - s + 3, \ldots, t + 1$ with an arbitrary number from $[t - s + 2]$. Then if $d_v$ denotes the number of $A-B$ edges in $H$ touching $v$ and $d''_v$ denotes the number of such edges of colors $t - s + 3, \ldots, t$, then $|X_v| = (t - s + 2)^{k-d_v+d''_v}$.

We claim that if $v, w \in B$ are non-adjacent with no common neighbor in $B$, then $X_v$ and $X_w$ have no string in common. Indeed, adding the edge $vw$ of color $t$ creates a rainbow $K_s$ with $s - 2$ vertices in $A$. So there must be an $a_i$ such that $a_iv$ and $a_iw$ have different colors, also differing from $t - s + 3, \ldots, t$. But then all the strings in $X_v$ have the color of $a_iw$ as their $i$’th letter, and all the strings in $X_w$ have the color of $a_iw$ as their $i$’th letter, so $X_v$ and $X_w$ are disjoint.

Since vertices in $B$ have degree at most $d$, each $v \in B$ has at most $d^2$ vertices $w \in B$ that are either adjacent to $v$ or have a common neighbor with $v$ in $B$. So each string in $[t - s + 2]^k$ can appear

3
in no more than $d^2 + 1$ collections $X_w$, and hence we get

$$(d^2 + 1)(t - s + 2)^k \geq \sum_{v \in B} |X_v| = \sum_{v \in B} (t - s + 2)^{k - d_v + d'_v}$$

$$d^2 + 1 \geq \sum_{v \in B} (t - s + 2)^{d'_v - d_v} \geq m \cdot (t - s + 2)^{1/m} (\sum_{v \in B} d'_v - \sum_{v \in B} d_v)$$

using Jensen’s inequality.

Now $t - s + 3, \ldots, t$ were the $s - 2$ most common colors, so we also have $\sum_{v \in B} d'_v \geq \frac{s^2 - 2 - t}{t} \sum_{v \in B} d_v$. Taking logs, we obtain

$$\sum_{v \in B} d_v \geq \frac{t}{t - s + 2} m \left( \log_{t - s + 2} m - \log_{t - s + 2} (d^2 + 1) \right).$$

As the left-hand side is a lower bound on the number of edges in $H$, this establishes the desired lower bound (using $d = \log^3 n$ and $m = n + o(n)$).

For the upper bound in the case of triangles, let $k$ be large enough, and take a set $X$ of size $m$ as provided by Theorem 1.3. Consider a $k$-by-$m$ complete bipartite graph $G_0$ with parts $A$ and $B$, where $A = \{a_1, \ldots, a_k\}$, and $B$ corresponds to the strings in $X$. For every vertex $v \in B$, we look at the corresponding string $x \in X$, and color each edge $va_i$ by the color $x(i)$. $G_0$ is clearly (rainbow) triangle-free, and by the definition of $X$, adding an edge to $G_0$ between two vertices of $B$ in any color $a \in [t]$ creates a rainbow triangle.

Now let $G$ be a maximal rainbow triangle-free supergraph of $G_0$. Then $G$ is rainbow triangle-saturated by definition, and compared to $G_0$, it only has new edges induced by $A$, thus it has at most $km + \binom{k}{2}$ edges. Here $n = k + m$ and $k = \frac{t(1 + o(1))}{(t-1) \log(t-1)} \log m$, implying the required upper bound. \[\Box\]

For $s > 3$ our lower bound is probably not tight. It would be interesting to determine the asymptotics of $\text{rsat}_t(n, K_s)$ for general $s$.

Acknowledgements. I thank Shagnik Das for finding [7] for me, and Gábor Simonyi for some clarifications about capacities.

References

[1] M. D. Barrus, M. Ferrara, J. Vandenbussche and P. S. Wenger, Colored saturation parameters for rainbow subgraphs, J. Graph Theory, 86 (2017), 375-386.

[2] B. Bollobás and A. Scott, Separating systems and oriented graphs of diameter two, J. Combin. Theory Ser. B 97 (2007), 193-203.

[3] I. Csiszár and J. Körner, Information Theory, 2nd edition, Cambridge University Press, 2011.

[4] P. Erdős, A. Hajnal and J.W. Moon, A problem in graph theory, Amer. Math. Monthly, 71 (1964), 1107-1110.
[5] M. Ferrara, D. Johnston, S. Loeb, F. Pfender, A. Schulte, H. C. Smith, E. Sullivan, M. Tait and C. Tompkins, On edge-colored saturation problems, arXiv:1712.00163 preprint

[6] Z. Füredi, P. Horak, C. M. Pareek and X. Zhu, Minimal oriented graphs of diameter 2, *Graphs Combin.*, 14 (1998), 345-350.

[7] L. Gargano, J. Körner and U. Vaccaro, Sperner capacities, *Graphs Combin.*, 9 (1993), 31-46.

[8] L. Gargano, J. Körner and U. Vaccaro, Capacities: from information theory to extremal set theory, *J. Combin. Theory Ser. A*, 68 (1994), 296-316.

[9] A. Girão, D. Lewis and K. Popielarz, Rainbow saturation of graphs, arXiv:1710.08025 preprint

[10] G. Katona and E. Szemerédi, On a problem of graph theory, *Studia Sci. Math. Hungar.*, 2 (1967), 23-28.

[11] D. Korándi and B. Sudakov, Saturation in random graphs, *Random Structures Algorithms* 51 (2017), 169-181.

[12] A. V. Kostochka, T. Łuczak, G. Simonyi and E. Sopena, On the minimum number of edges giving maximum oriented chromatic number, in: Contemporary Trends in Discrete Mathematics, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol 49. (1999), 179-182.

[13] G. Simonyi, Perfect graphs and graph entropy. An updated survey, in: Perfect Graphs, Wiley (2001), 293-328.

[14] A. Zykov, On some properties of linear complexes (in Russian), *Mat. Sbornik N. S.*, 24 (1949), 163-188.