Pattern Avoidance in Sequences

Shihan Kanungo
Mentor: Prof. Jesse Geneson (SJSU)

SJSU

PRIMES October Conference 2025

Outline

Outline

In the field of extremal combinatorics, we ask questions like

Extremal Question

What is the largest possible size of an object which avoids a given forbidden substructure?

For example:

- What is the largest possible graph on n vertices that does not contain K_3 as a subgraph? (Turan's Theorem & the forbidden subgraph problem)
- What is the largest possible group of people, such that for any set of k people, they are not all friends or not all strangers? (Ramsey Theory)
- What is the largest possible subset of $\{1, \ldots, n\}$ that does not contain a k-term arithmetic progression? (Szemeredi's Theorem)

Saturation

The saturation question is a bit more complicated.

Saturation Question

What is the SMALLEST possible structure that avoids a given forbidden substructure, BUT making it larger in any way induces a copy of the forbidden structure?

In other words: the *minimum* size of a *maximal* structure, rather than the *maximum* size.

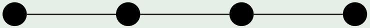
Saturation for Graphs

Definition

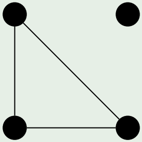
Let G and H be graphs. We say G is H-saturated if G avoids H as a subgraph, but adding any new edge to G induces a copy of H.

Example

Consider

$H =$ 

The graph



is H-saturated.

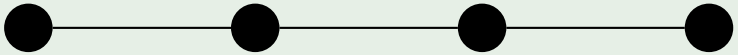
Saturation for Graphs

Definition

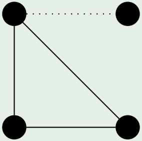
Let G and H be graphs. We say G is H-saturated if G avoids H as a subgraph, but adding any new edge to G induces a copy of H.

Example

Consider

$H =$ 

The graph



is H-saturated.

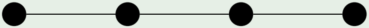
Saturation for Graphs

Definition

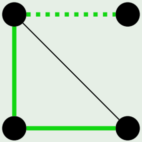
Let G and H be graphs. We say G is H-saturated if G avoids H as a subgraph, but adding any new edge to G induces a copy of H.

Example

Consider

$H =$ 

The graph



is H-saturated.

Dichotomy

Definition

The saturation function $\mathrm{Sat}(n, H)$ is the minimum number of edges in a H-saturated graph on n vertices.

$\mathrm{Sat}(n, H)$ exhibits an dichotomy:

Theorem (Kászonyi-Tuza, 1986)

We have $\mathrm{Sat}(n, H) = O(1)$ *or* $\mathrm{Sat}(n, H) = \Theta(n)$*.*

In many other settings, it has been seen that the saturation function exhibits the same dichotomy.

Outline

Sequence Saturation

We use "letters" to refer to the terms in a sequence.

Definition

Let s, u be two sequences. We say that s contains a copy of u if s has a subsequence that can be turned into u by a *one-to-one* renaming of letters.

Example

The sequence $s = \mathbf{1}, \mathbf{2}, 3, 2, 3, \mathbf{1}, \mathbf{2}$ contains a copy of $u = abab$:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 2 & 3 & 1 & 2 \\ a & b & & & & a & b \end{array}$$

However, $s = 1, 2, 3, 2, 1$ does not.

Definition

A sequence s is r-sparse if every consecutive r letters are pairwise distinct.

Example

The sequence $s = 1, 2, \mathbf{3}, \mathbf{2}, \mathbf{3}, 1, 2$ is 2-sparse, but not 3-sparse.

Definition

Let u be a sequence with r distinct letters. A sequence s is u-saturated if s avoids u, s is r-sparse, and inserting any new letter into s either induces a copy of u or violates r-sparsity.

If we dropped the r-sparsity condition, we would have arbitrarily long sequences like $1, 1, \cdots$ which avoid u.

Example

If $u = abca$ then the sequence $s = 1, 2, 3$ is u-saturated.

First, $r = 3$ since u has 3 distinct letters. Now we check:

s **avoids** u**:** Evident.

s **is 3-sparse:** Evident.

Saturation: Suppose we insert a 1 into s. The possibilities are:

$$\color{green}{\mathbf{1}}\color{black}\quad 1 \quad 2 \quad 3 \quad \text{(violates 3-sparsity)}$$

$$1 \quad \color{green}{\mathbf{1}}\color{black}\quad 2 \quad 3 \quad \text{(violates 3-sparsity)}$$

$$1 \quad 2 \quad \color{green}{\mathbf{1}}\color{black}\quad 3 \quad \text{(violates 3-sparsity)}$$

$$\begin{matrix} 1 & 2 & 3 & \color{green}{\mathbf{1}} \\ a & b & c & a \end{matrix} \quad \text{(has a copy of } u\text{)}$$

Similar checks for the other letters.

Definition

The saturation function $\mathrm{Sat}(n, u)$ is the length of the shortest u-saturated sequence with n distinct letters.

In 2021, Anand, Geneson, Kaustav, and Tsai conjectured

Conjecture

We have $\mathrm{Sat}(n, u) = O(n)$.

This implies the dichotomy $\mathrm{Sat}(n, u) = O(1)$ or $\Theta(n)$. They proved

Theorem (Anand-Geneson-Kaustav-Tsai, 2021)

We have $\mathrm{Sat}(n, u) = O(n)$ *for all sequences* u *with two distinct letters.*

However, the cases for u having ≥ 3 distinct letters remained completely open.

Outline

Algorithm

Consider the following algorithm:

Input: Alphabet $A = \{1, \ldots, n\}$, forbidden sequence u
Output: u-saturated sequence
Initialize the sequence: $s \leftarrow 1, 2, \ldots, r-1$ $\triangleright$ Initial sequence avoids u
while it is possible to extend the sequence **do**
  for each letter $x \in A$ **do**
    if x can be *properly inserted* into s **then**
      Insert x appropriately into s to form s' $\triangleright$ Smallest x, leftmost position
      Update $s \leftarrow s'$ $\triangleright$ New sequence
      break $\triangleright$ Exit loop after the first valid insertion
    end if
  end for
end while
Return s $\triangleright$ Final sequence is u-saturated

Algorithm

The output of the algorithm:

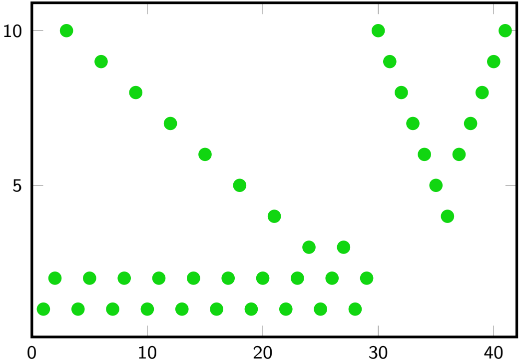


Figure: Algorithm on $u = abcacbc$.

Here, we represent $s = s_1 \cdots s_\ell$ by plotting the points (i, s_i). Using this pattern, we get $\mathrm{Sat}(n, abcacbc) = O(n)$!

Some more pictures:

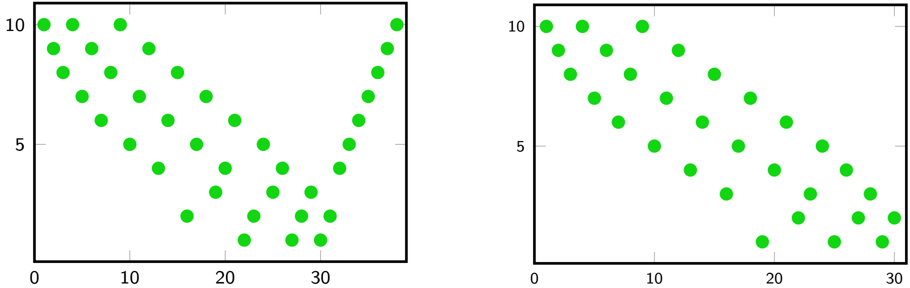


Figure: Algorithm on $u = abbacac$ (left) and $abcacba$ (right).

And even more:

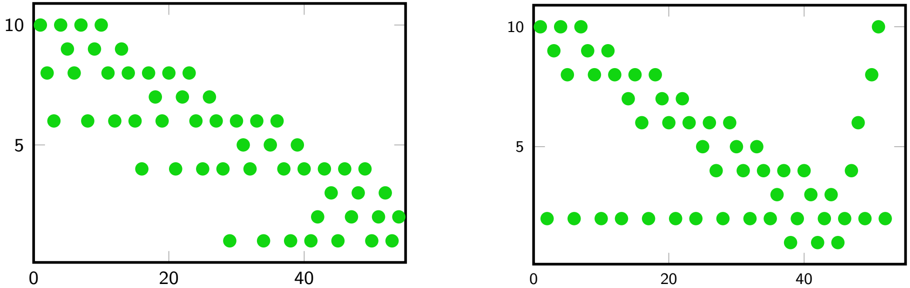


Figure: Algorithm on $u = abcacbacb$ (left) and $abcbacbac$ (right).

This lets us resolve the conjecture for many specific sequences u.

Say u is irreducible if u cannot be decomposed into sequences $u = u_1 u_2$ such that u_1 and u_2 have no letters in common.

Theorem (Kanungo, 2025+)

If u is irreducible and of the form $aa \cdots bb$, then $\mathrm{Sat}(n, u) = O(n)$.

Suppose u is a sequence on 3 letters, and $u = abc \cdots xyz$ where a, b, c are distinct. Define

$$f_0(u) = \#\{\,\text{consecutive pairs of the form } ab, bc, ca\,\},$$
$$f_1(u) = \#\{\,\text{consecutive pairs of the form } ac, ba, cb\,\}.$$

Theorem (Kanungo, 2025+)

Let $u = abc \ldots xyz$ be a three-letter sequence with a, b, c distinct. Suppose

$$xyz \in \{abc, bca, cab\}, \qquad f_0(u) \geq f_1(u) + 5.$$

Then $\mathrm{Sat}(n, u) = O(n)$.

Corollary

For any sequence u on 3 letters, $\mathrm{Sat}(n, (abc)u(abc)^t) = O(n)$ *for large enough t.*

Acknowledgments

- My mentor, Prof. Jesse Geneson, for his guidance and support.
- The PRIMES organizers, for making this experience possible.
- My parents, without whom none of this would be possible.

Questions?

Thank You!