21 Feb 2018

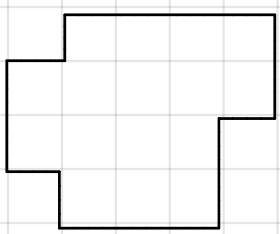
Announcements

[1] Solution sets 2, 3 posted on CMS.
[2] Prelim review session Friday morning in class. (Chapter 5 after that.)
[3] Prelim practice problems coming soon.
(Focus will be on dynamic programming.)
[4] No homework this coming week, due to prelim.

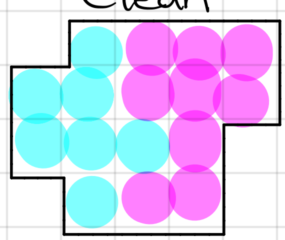
TROMINO TILING reduced to GRAPH SEARCH.

Given subset of $k \times n$ rectangle, can it be tiled with L-shaped trominoes?

<u>Def.</u> If A is a subset of $k \times n$ grid, a "clean partition" of A is a partition into two subsets, A_L and A_R, such that A_L is contained in columns $1, \dots, j$ of the grid, A_R is contained in $j, \dots, n$, for some value of j.

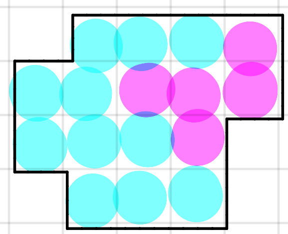
<u>E.g.</u> $A =$ 

Clean



$j = 3$

Not clean

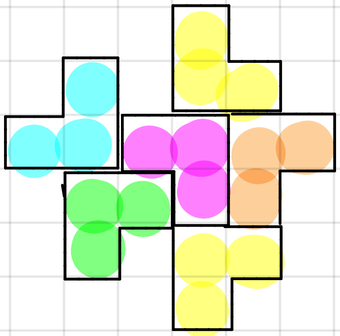


$j=3$, $j=4$ are both mixed

If $A_L \subseteq \{\text{columns } 1, \dots, j\}$, $A_R \subseteq \{\text{columns } j, \dots, n\}$ we call j the <u>level</u> of the partition.

<u>Observation.</u> Every tiling of A by trominoes can be built up "one level at a time." In other words, there is a sequence of clean partitions $(A_L^0, A_R^0), \dots, (A_L^n, A_R^n)$ such that

(1) $A_L^0 = \emptyset$, $A_L^n = A$.
(2) A_L^j, A_R^j can be tiled by trom's.
(3) (A_L^j, A_R^j) is clean at level j.
(4) $A_L^{j+1} \supseteq A_L^j$ $\forall j$.



A

$A_L^0 = \emptyset, \quad A_R^0 = A.$

$A_L^1 = \text{blue} \quad A_R^1 = A - A_L^1$

$A_L^2 = \text{blue} + \text{green}$

$A_L^3 = \text{blue, green, pink}$

$A_L^4 = \text{blue, green, pink, yellow}$

$A_L^5 = A$

$A_L^j =$ union of trominoes that are contained in columns $1, \ldots, j$

$A_R^j =$ union of trominoes in columns $j, \ldots, n.$

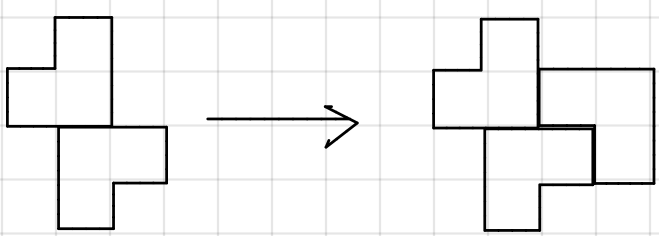
Create a graph with vertex set

$$V = V^0 \cup V^1 \cup V^2 \cup \cdots \cup V^n$$

$V^j =$ subsets of A that are contained in columns $1, \ldots, j$ and whose complement is contained " " $j, \ldots, n.$

Edge $A_L^j \rightarrow A_L^{j+1}$ exists if we can partition A_L^{j+1} into A_L^j and a width-2 set that can be tiled by trominoes.

E.g.



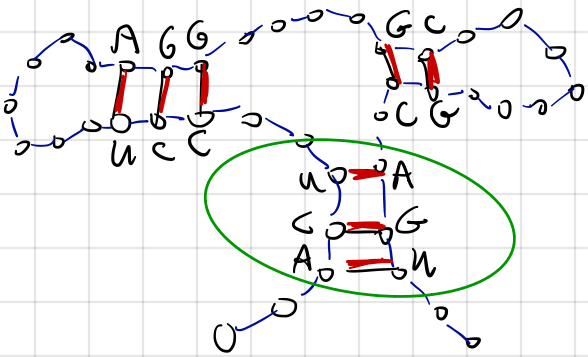
Search this graph using DFS to see if it contains a path from $s = \emptyset$ to $t = A$.

RNA Secondary Structure Prediction (§6. whatever)

RNA is a single-stranded molecule made up of $\{A, C, G, U\}$.
Certain of these can pair with each other namely:

A – U   C – G.

This enables it to form structures e.g.

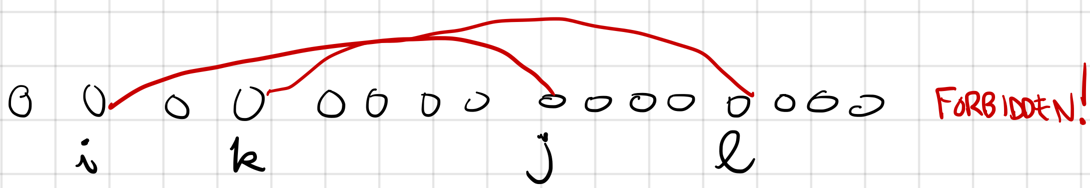


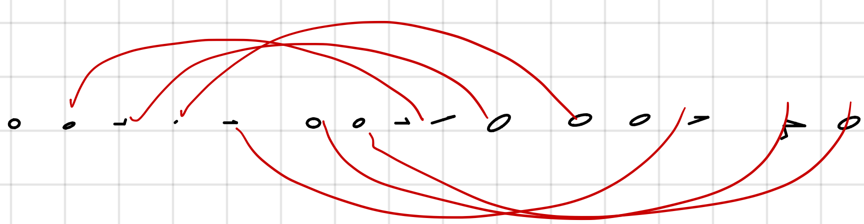
The "rules" for base pairing are:

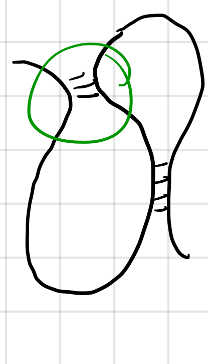
[1] A can pair only with U, C only with G,
U only with A, G only with C.

[2] ("No sharp turns") If the i^{th} and j^{th} nucleotides are paired then $|i-j| > 4$.

[3] ("Non-crossing") If (i,j) and (k,l) are paired then it is <u>not</u> the case that $i < k < j < l$.







Problem. Given a sequence representing an RNA molecule, find the maximum # of base pairs that it can form.

No pair

$$\boxed{A \quad G} \quad \underbrace{C \quad \boxed{U \quad A \quad A \quad U \quad G \quad C \quad U \quad U \quad A} \quad G}$$

What's the "last part" of the solution? What are the options for that "last part"?

What do the potential optimal solutions look like, for each of these possible choices?

- If n^{th} nucleotide is unpaired, choose maximum pairing of $1, \dots, n-1$.
- If (j, n) are paired, choose maximum pairings of $1, \dots, j-1$ and of $j+1, \dots, n-1$.

Two disjoint subproblems!

Define $T[j,k] :=$ max # of pairs than can be formed on subsequence $j, \dots, k$

$$(*) \qquad T[j,k] = \max\Big\{ T[j,k-1],\ \max\big\{ T[j,i-1] + T[i+1,k-1] + 1 \ \big|\ i \in P(k) \big\} \Big\}$$

$$P(k) = \big\{ i \in \{j, \dots, k-5\} \ \big|\ i \,\&\, k \text{ are a matched pair of nucleotides, e.g. A-U or C-G} \big\}$$

Fill in DP table in increasing order of $k-j$.