CPSC 421: Introduction to Theory of Computing **Winter Term 1 2018-19**

Lecture 5: September 14

Lecturer: Nicholas Harvey *Scribes: Kaitian Xie*

5.1 Regular Expressions

We will "show": A language is a regular iff it can be described by a regular expression.

R:

- 0^*1^*
- $\Sigma^*001\Sigma^*$
- $(\Sigma\Sigma)^*$
- $\Sigma^*1\Sigma^*1\Sigma^*$

Shorthand: $\Sigma = (a_1|a_2|\cdots|a_k)$ where $\Sigma = a_1a_2\cdots a_k$

$L(R)$: the set of strings generated by regular expression R.

- $\{0^i1^j : i, j \geq 0\}$
- {all strings containing 001 as a substring}
- {all strings of even length}
- $\{w : w$ contains at least three 1s$\}$

$\Leftarrow$: Any regular expression can be converted into an equivalent NFA.

$\Rightarrow$: Idea: generalize NFAs by allowing arbitrary R.Es as labels on their transitions. Keep simplifying.

Limits to the power of Finite Automata

Can any language be described by a F.A?

Example 1: Suppose L is a finite language (i.e. L is finite). Is L regular? Yes

Why? Suppose $L = \{x_1, \cdots, x_n\}$.

We can define $L_i = \{x_i\}$. This is regular. We know $L_1 \cup L_2$ is regular, so $L_1 \cup L_2 \cup \cdots \cup L_n$ is regular.

This fails if L is infinite because we could get an "Infinite Automaton".

Example 2: Suppose L is regular. Define $L^2 = \{xy : x \in L \text{ and } y \in L\}$

This is regular: it is $L \circ L$.

Example 3: $L^{dup} = \{xx : x \in L\}$

Is this regular? It depends $\cdots$

If L is finite, L^{dup} is finite, hence regular.

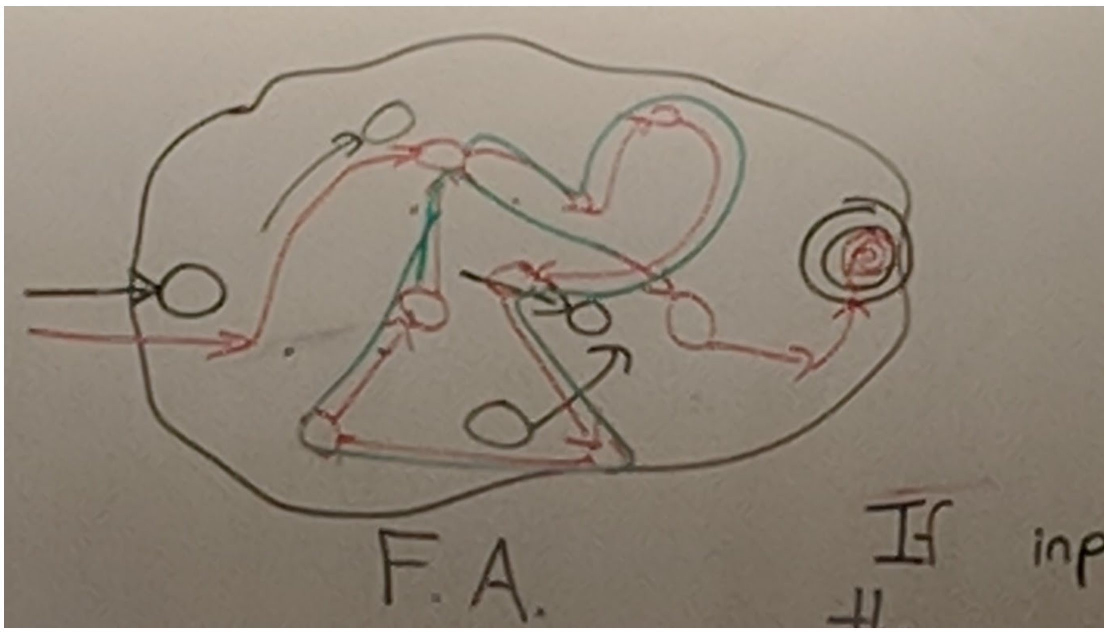
If $L = L(1^*)$, then $L^{dup} = \{1^{2i} : i \geq 0\} = L((11)^*)$

If $L = \Sigma^*$, $L^{dup} = \{xx : x \in \Sigma^*\}$. Is this regular?

Intuitive argument: $\{xx : x \in \Sigma^*\}$ is not regular.

- With a F.A. its states are its "memory".
- Any F.A. accepting L^{dup} "must remember" the first half to compare to the second half.
- So F.A. must remember arbitrary large information, but it can't.

Idea about finite automaton:



- $\rightarrow$ "computation path", transitions followed when processing input

If input is very long, this computation path must have a cycle.