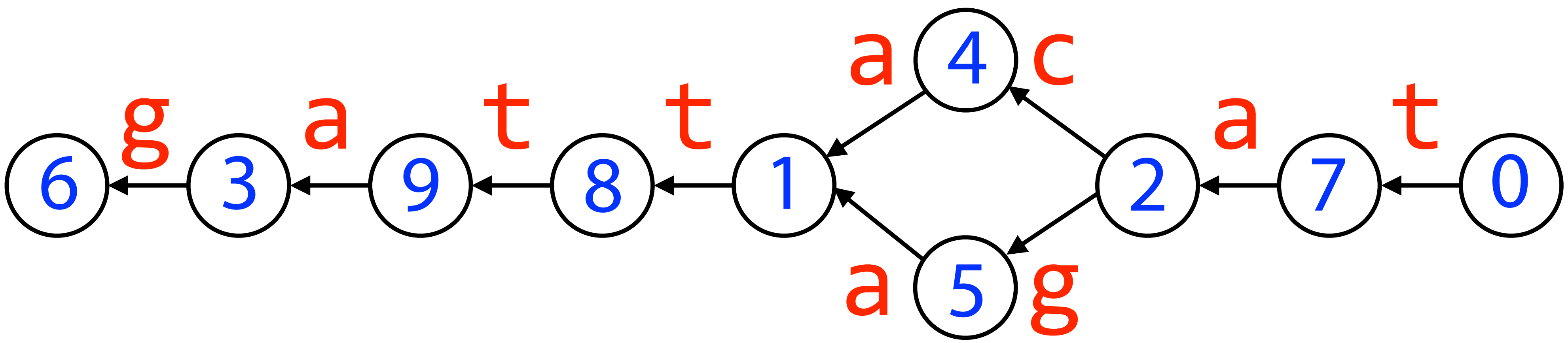
Wheeler graphs, part 4: Consecutivity

Ben Langmead

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Department of Computer Science

Wheeler graphs

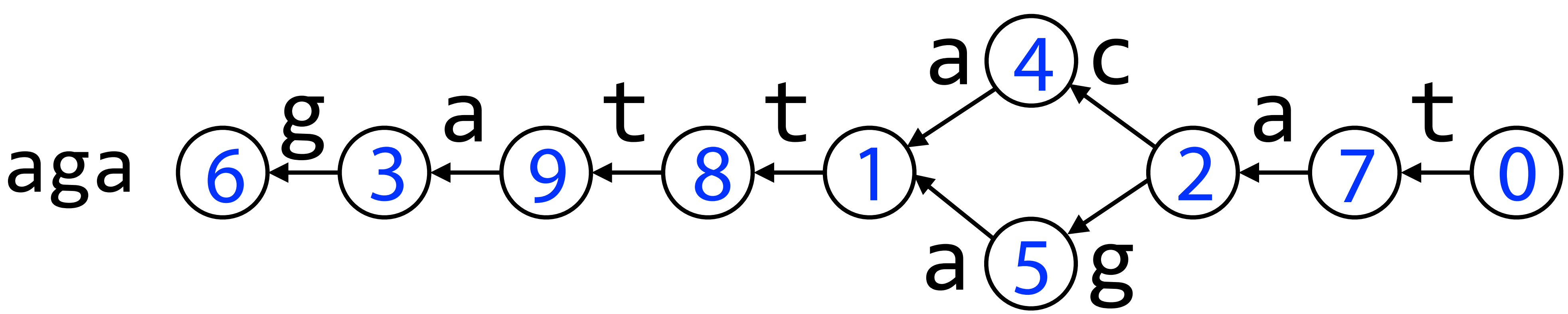


Do Wheeler Graphs have the kind of ***consecutivity*** that enables FM-Index-like matching?

Wheeler graphs

A graph is ***path coherent*** if nodes can be ordered such that:

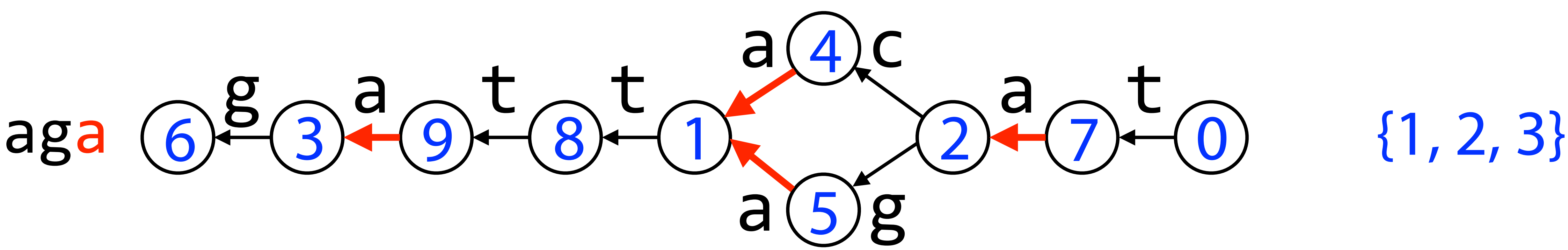
For any consecutive range $[i, j]$ of nodes and character c, the nodes reached by following edges matching c also form a consecutive range

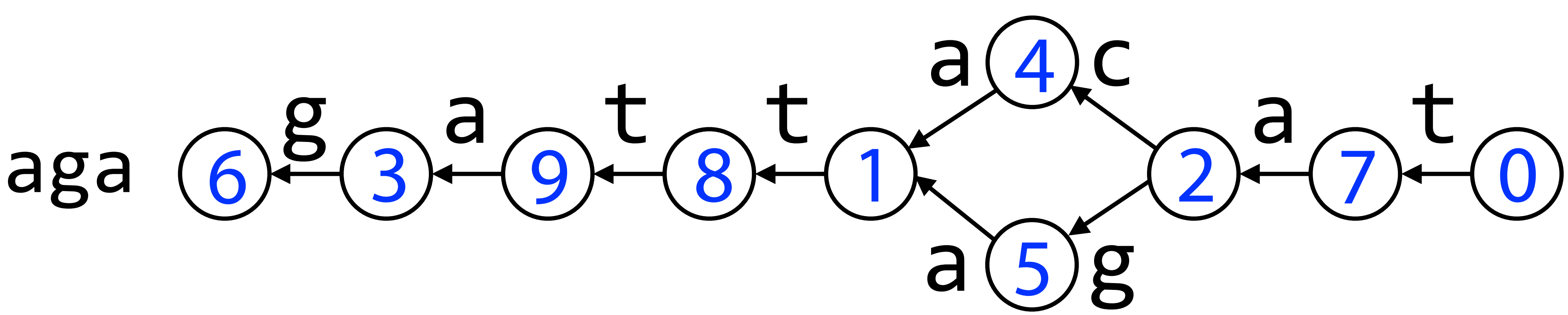


Wheeler graphs

A graph is ***path coherent*** if nodes can be ordered such that:

For any consecutive range $[i, j]$ of nodes and character c, the nodes reached by following edges matching c also form a consecutive range

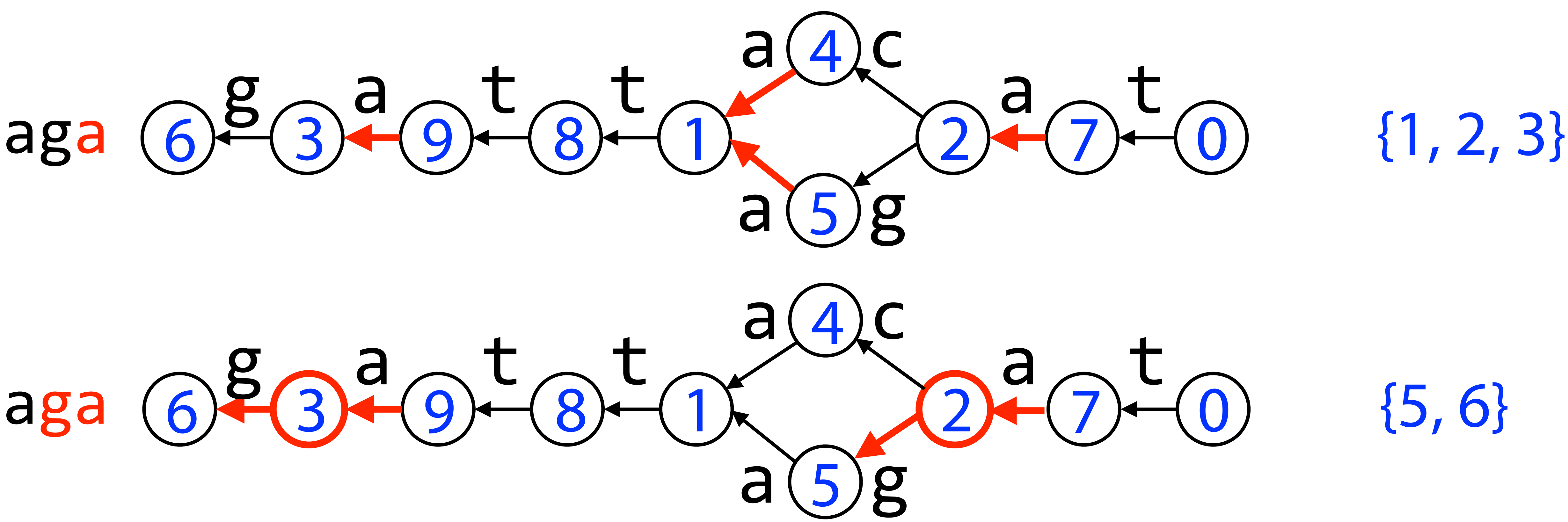




Wheeler graphs

A graph is ***path coherent*** if nodes can be ordered such that:

For any consecutive range $[i, j]$ of nodes and character c, the nodes reached by following edges matching c also form a consecutive range



Wheeler graphs

Consider one step:

Initial set of nodes are in consecutive range $[i, j]$

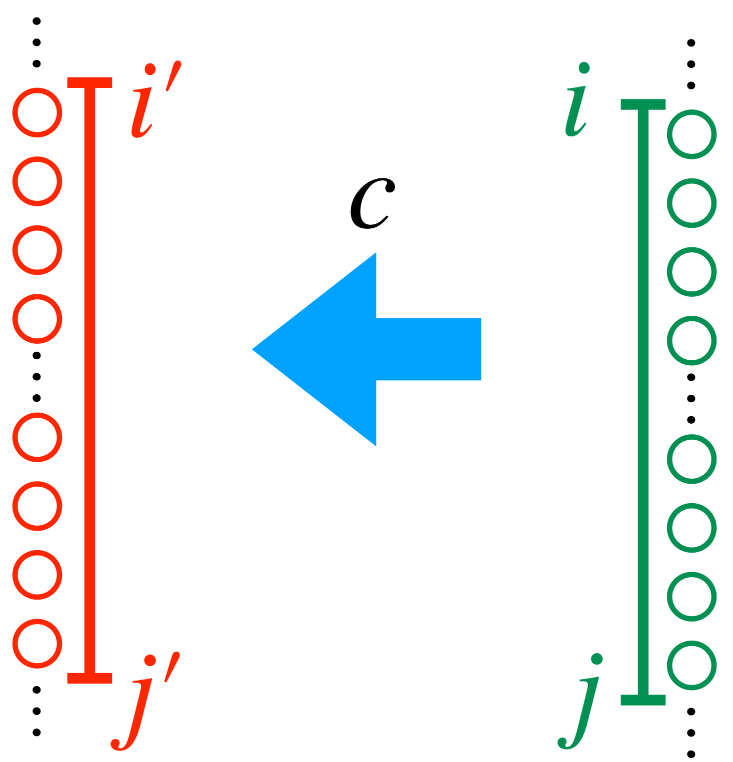
After advancing on character c, $[i', j']$ is the smallest range containing **next** set of nodes

Wheeler graphs

Consider one step:

Initial set of nodes are in consecutive range $[i, j]$

After advancing on character c, $[i', j']$ is the smallest range containing **next** set of nodes

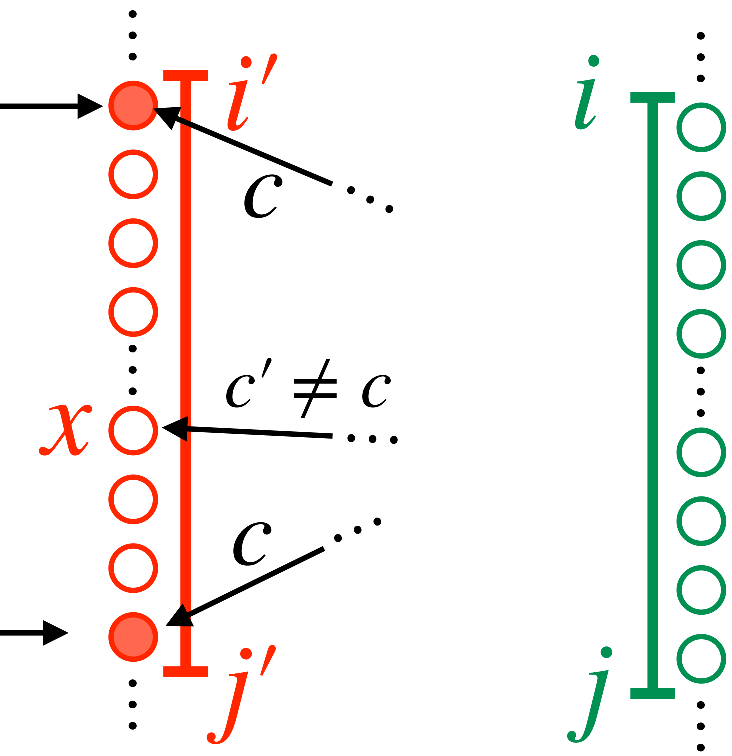


Do nodes in $[i', j']$ consist *only* of *the* c-successors of nodes in $[i, j]$?

Wheeler graphs

As defined, i' is reachable via an edge labeled c from a node in $[i, j]$

Same for j'



Consider node x, where $i' < x < j'$ with incoming edge labeled c'. Suppose $c' \neq c$.

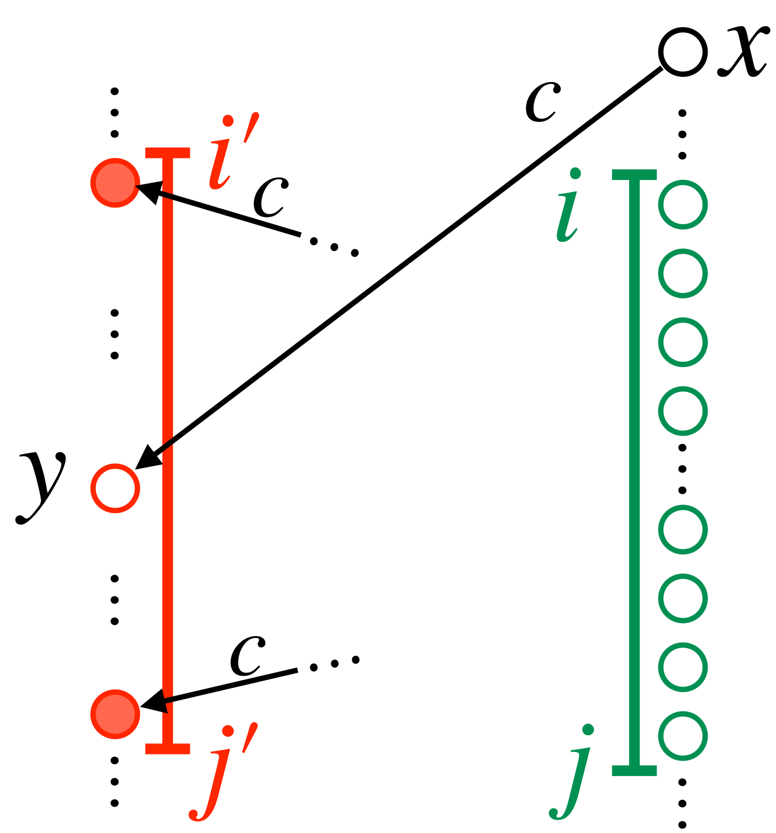
Recall: $a \prec a' \implies v < v'$

Since $x \nless i'$, we have $c' \nprec c$

Since $j' \nless x$, we have $c \nprec c'$

We have $c' \succcurlyeq c$, $c \succcurlyeq c'$, and $c' \neq c$, giving a contradiction

Wheeler graphs



Could node $x \notin [i, j]$ be a $c-$predecessor of a node y, $i' < y < j'$?

No. Proof idea: draw contradiction, similar to previous argument, but using rule 3:

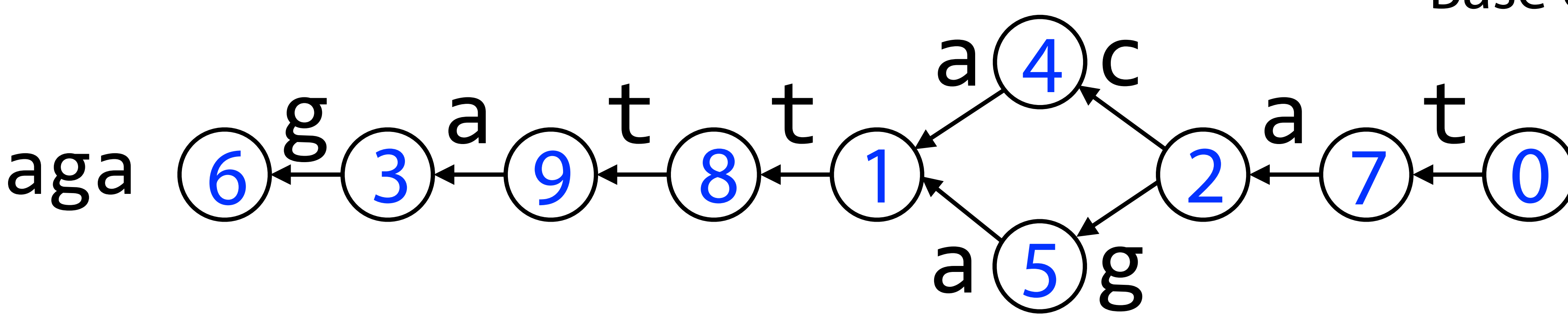
$$(a = a') \wedge (u < u') \implies v \leq v'$$

Wheeler graphs

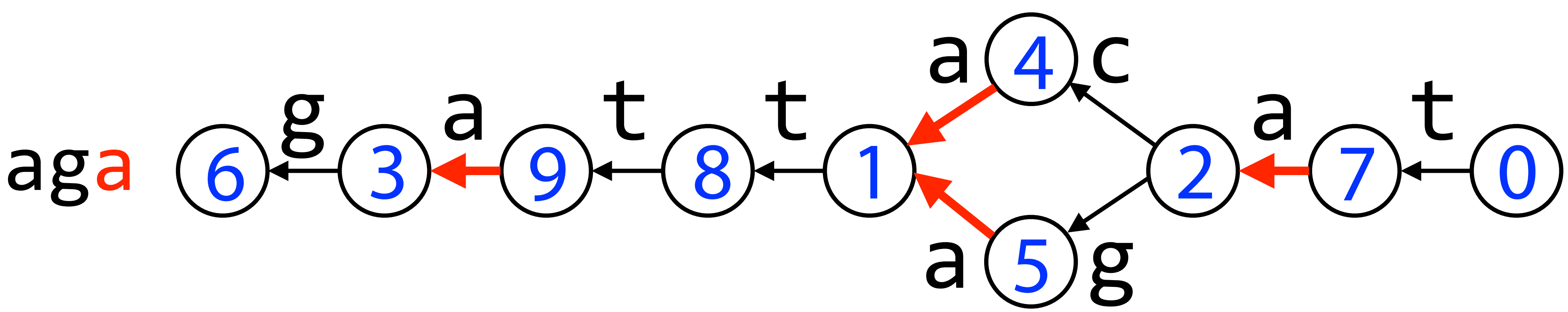
For consecutive range $[i, j]$ of nodes & string α, the nodes reached by matching α also form a consecutive range

Proof idea: extend previous arguments to string α inductively

Base case: initial range has **all nodes**



{0, 1, ..., 9}



{1, 2, 3}

Wheeler graphs: review

Definition of Wheeler graph

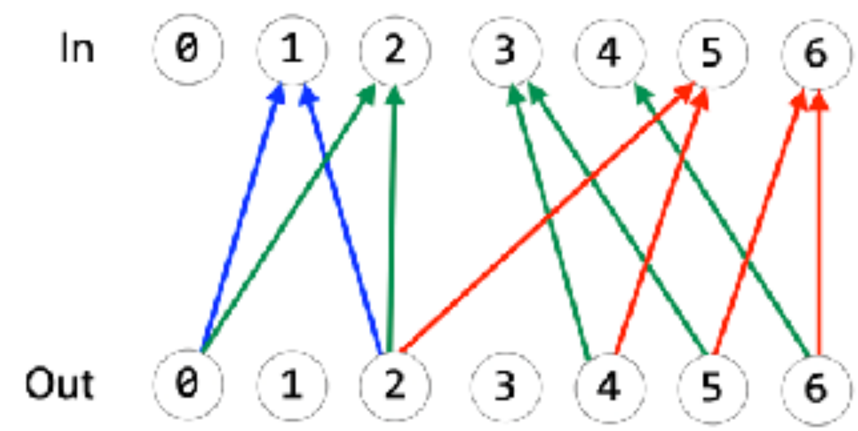
0 in-degree nodes come before others (1)

For all pairs of edges:

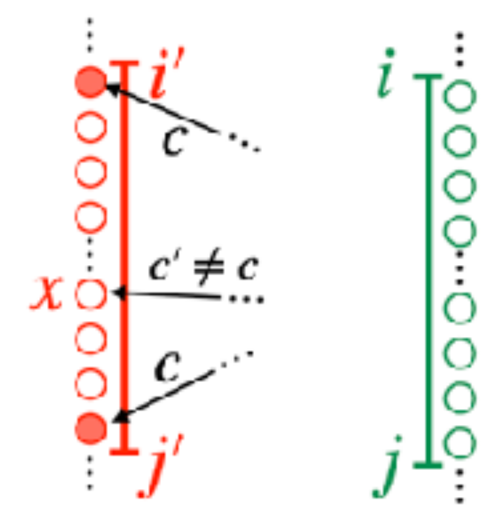
$$a \prec a' \implies v < v' \quad (2)$$

$$(a = a') \wedge (u < u') \implies v \leq v' \quad (3)$$

"Ordered destinations & no crossing" interpretation



Proved consecutivity property and, by extension, path coherence



Wheeler graphs: next

How do we represent & query a Wheeler graph?

Can we query with FM-Index-like ease & efficiency?