Wheeler graphs, part 4: Consecutivity

Ben Langmead

Department of Computer Science

Please sign guestbook (www.langmead-lab.org/teaching-materials) to tell me briefly how you are using the slides. For original Keynote files, email me (ben.langmead@gmail.com).
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

\[
\begin{align*}
\text{aga} & \quad \begin{array}{ccccccc}
6 & g & a & t & t & 1 & \text{aga} \\
3 & a & t & 8 & 5 & 2 & 7 & 0 \{1, 2, 3\}
\end{array}
\end{align*}
\]
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: \( a < a' \implies v < v' \)

Since \( x \not< i' \), we have \( c' \not< c \)

Since \( j' \not< x \), we have \( c \not< c' \)

We have \( c' \succeq c \), \( c \succeq c' \), 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:

$$(a = a') \land (u < u') \implies v \leq v'$$
Wheeler graphs

For consecutive range \([i, j]\) of nodes & string \(\alpha\), the nodes reached by matching \(\alpha\) also form a consecutive range

Proof idea: extend previous arguments to string \(\alpha\) inductively

Base case: initial range has all nodes

\(\{0, 1, ..., 9\}\)
Wheeler graphs: review

Definition of Wheeler graph

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?