

COMP 543: Tools & Models for Data Science

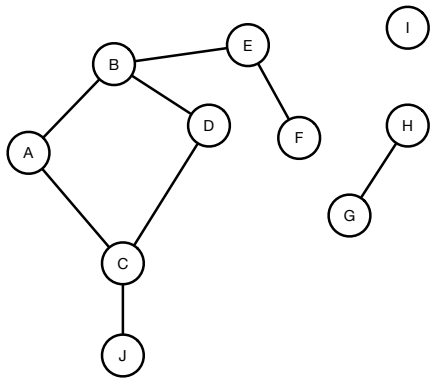
BFS & PageRank

Chris Jermaine & Risa Myers

Rice University



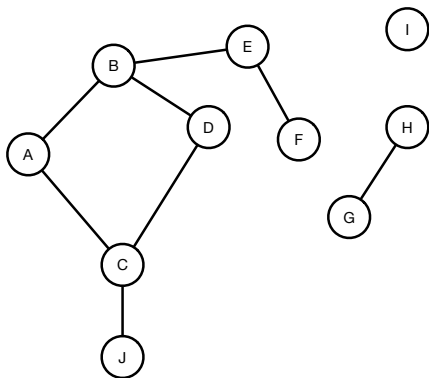
Connected Components



? How many connected components are there in this network?

Breadth First Search

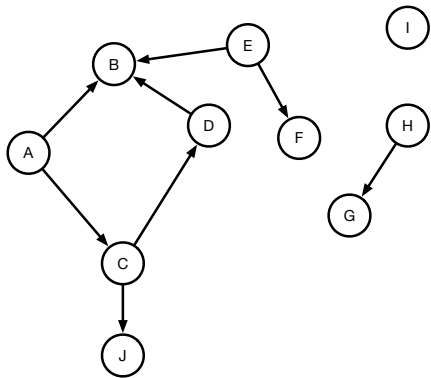
- 1 Pick a node at random
 - ? How do you pick a random node?
- 2 Search for all nodes reachable from that node
 - ? How do you keep track of this set of nodes?
- 3 Search for all nodes reachable from those nodes
- 4 Stop when no new nodes are found
- 5 Repeat with any nodes still in the original set
 - ? How do you keep track of which nodes are in which connected component?



Optimizing Breadth First Search

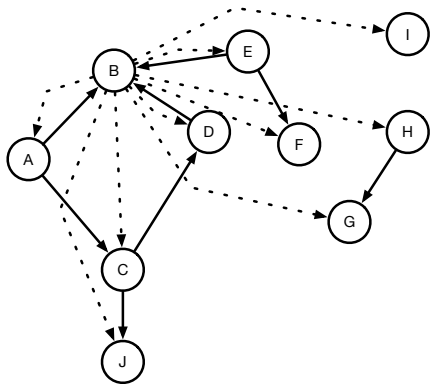
- ? What can be done declaratively?
- ? What has to be done imperatively?

PageRank network



? What's different here?

PageRank Sinks

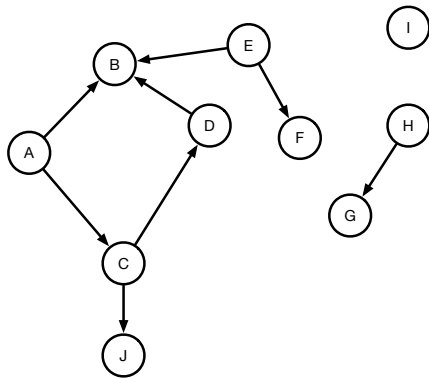


- Sinks point equally to all other nodes

- 1 Measures **relative** importance
- 2 Leverages network structure
- 3 Factors in the importance of the referencing nodes

PageRank

- 1 Start with all nodes having equal importance
- 2 Then iterate until PageRank values converge
- 3 For each node j in the network, update its PageRank based on
 - The chance that the browser teleports someplace else, $\frac{1-d}{n}$
 - The chance the user stays put: d times
 - The importance contribution of the nodes that point to node j



- ? What is the damping factor?
- ? What special cases do you have to consider?
- ? What can be done declaratively?
- ? What has to be done imperatively?

How do I know if I got the right answer?

- Try your algorithms on a (very) small network where you can hand calculate the answer
- You may compare ANSWERS (but not SOLUTIONS) with your classmates