CLOUD COMPUTING APPLICATIONS

MapReduce Example: Page Rank

Roy Campbell & Reza Farivar

PageRank Algorithm

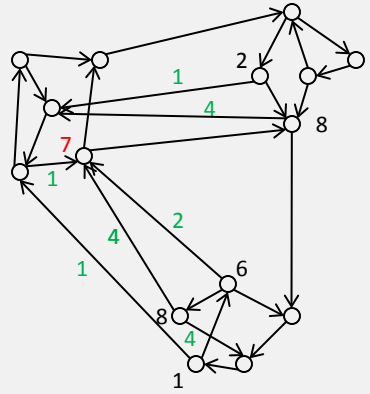
- Program implemented by Google to rank any type of recursive “documents” using MapReduce
- Initially developed at Stanford University by Google founders, Larry Page and Sergey Brin, in 1995
- Led to a functional prototype named Google in 1998

- PageRank value for a page *u* is dependent on the PageRank values for each page *v* out of the set B_u (all pages linking to page *u*), divided by the number *L(v)* of links from page *v*

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

PageRank

- Phase 1: Propagation
- Phase 2: Aggregation



- Input: A pool of objects, including both vertices and edges

PageRank: Propagation

- Map: for each object
  - If object is vertex, emit key=URL, value=object
  - If object is edge, emit key=source URL, value=object
- Reduce: (input is a web page and all the outgoing links)
  - Find the number of edge objects→outgoing links
  - Read the PageRank value from the vertex object
  - Assign PR(edges)=PR(vertex)/num_outgoing

PageRank: Aggregation

- Map: for each object
  - If object is vertex, emit key=URL, value=object
  - If object is edge, emit key=destination URL, value=object
- Reduce: (input is a web page and all the incoming links)
  - Add the PR value of all incoming links
  - Assign PR(vertex)=$\sum$PR(incoming links)