Graph Partitioning Using Betweenness Measures in Distributed Environment

Shiyang Cheng(SC57227), Cong Wang(CW37657),
Marina Thomas(MT34799), Kyle Sung(CS9893)

Why Partitioning Graph?

- Graph in real life
  - Graph consists of human: Facebook, Twitter
  - Graph consists of devices: Internet of Things

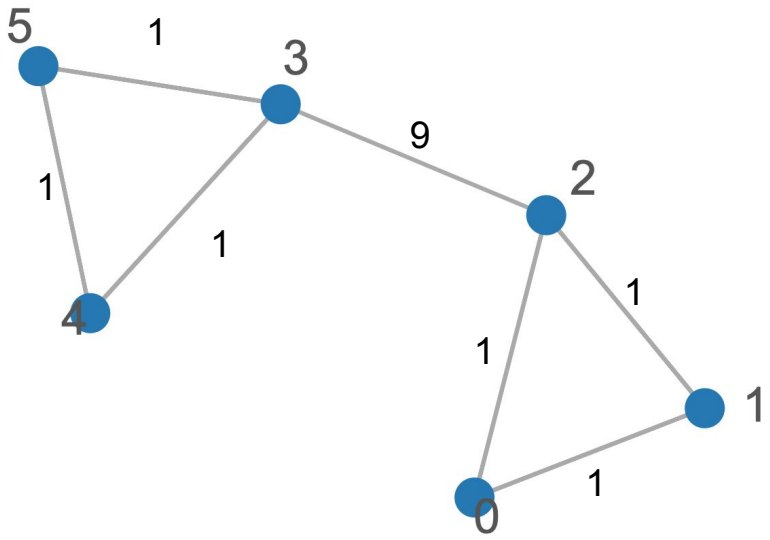
- Properties of graph in real life
  - A lot of vertices
  - Not every vertices connect to all other vertices

Original Girvan-Newman Algorithm

Key assumption: some vertices get high degree, some get low degree

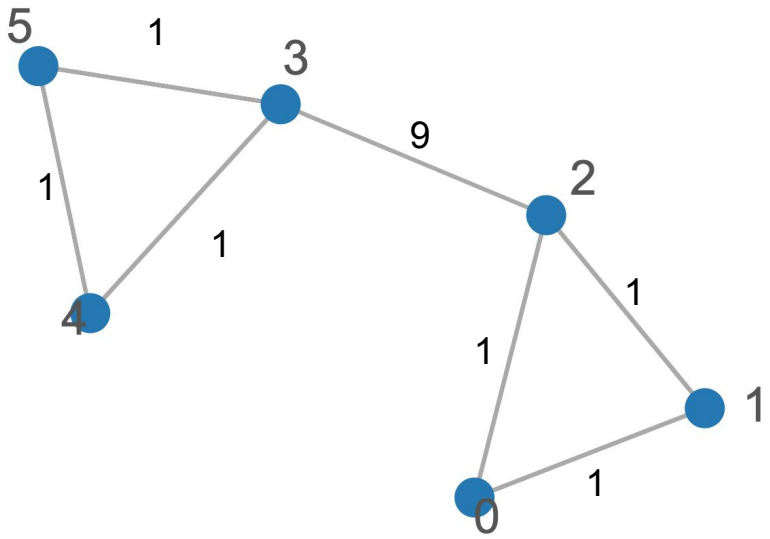
1. Calculate the betweenness for all edges in the network
2. Remove the edge with the highest betweenness
3. Recalculate betweenness
4. Repeat from step 2 until no edges remain

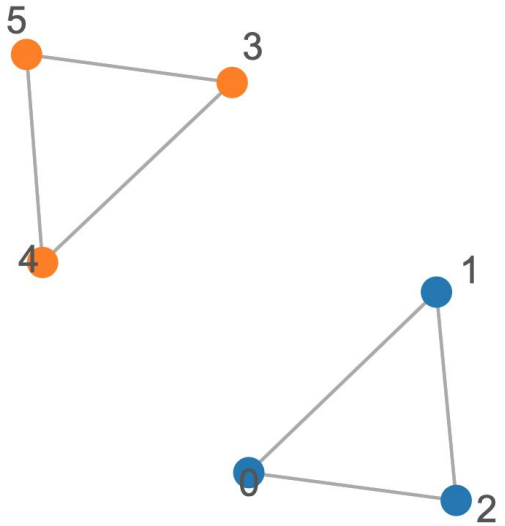
Review: Betweenness



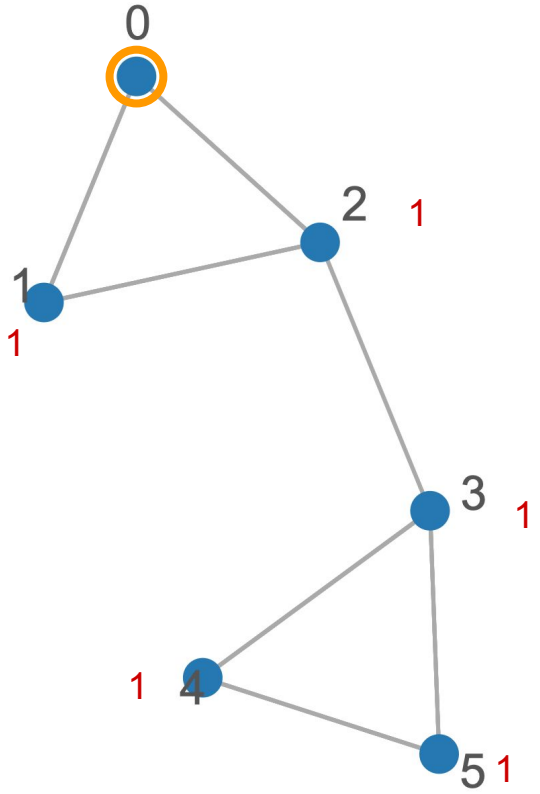
Edge betweenness: the number of shortest paths between all pairs of vertices that run along it.

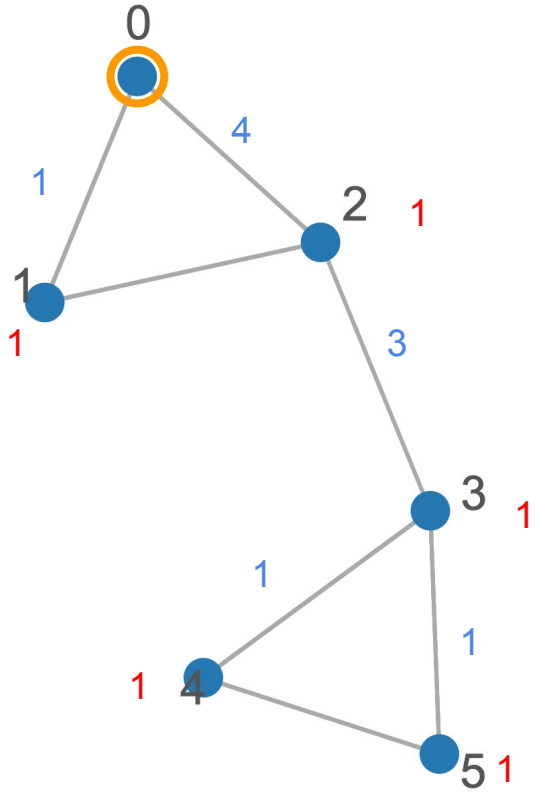
Review: Partitioning

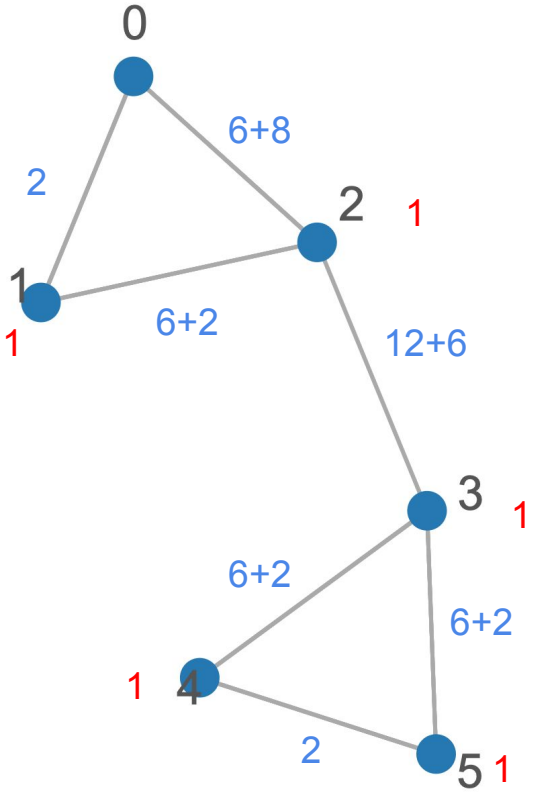




Review: Girvan-Newman method







Distributed Environment

Network nodes now exist on individual processes that can run on different servers;

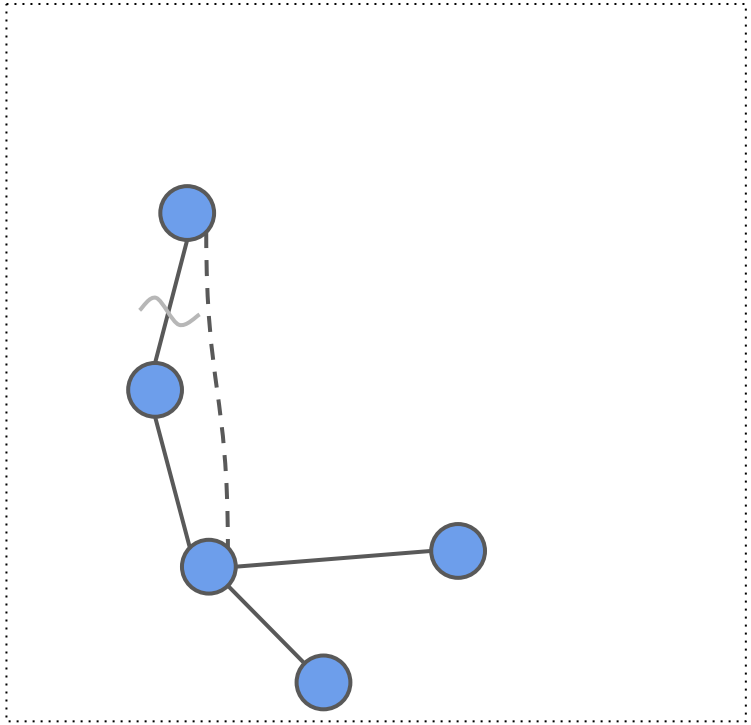
No shared memory,

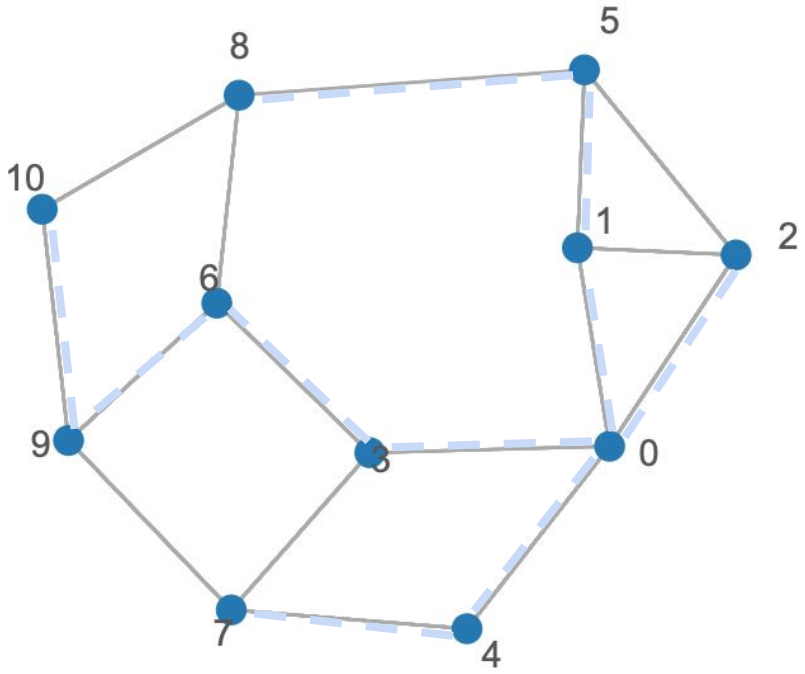
No global clock, messages can arrive with time delay

Each node initially has knowledge of their immediate neighbors only.

Assumes FIFO channel, reliable processes. Can broadcast to all nodes.

Diffusing Computation, Termination





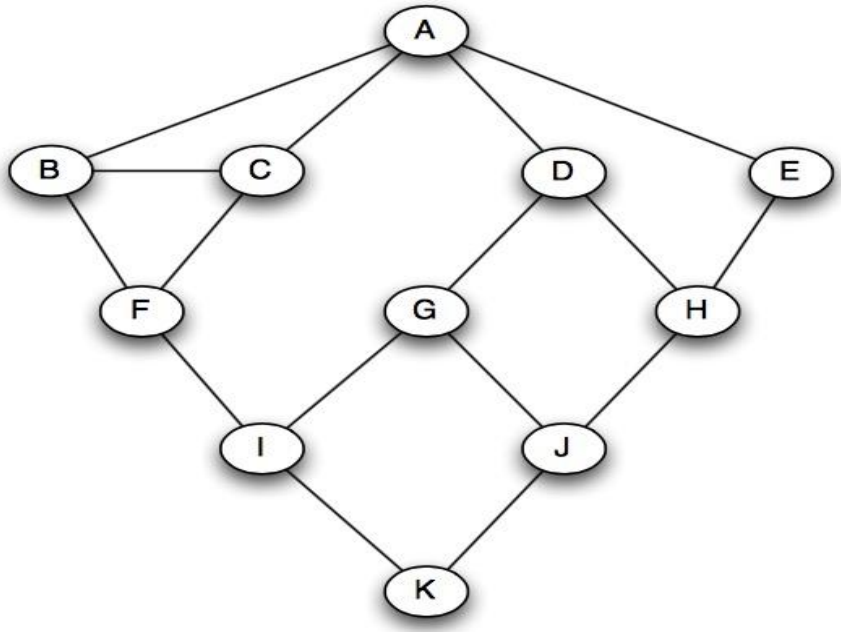
Algorithm - 3 Steps

1. Shortest path using diffusing computation
2. Find parent value
3. Find flow value

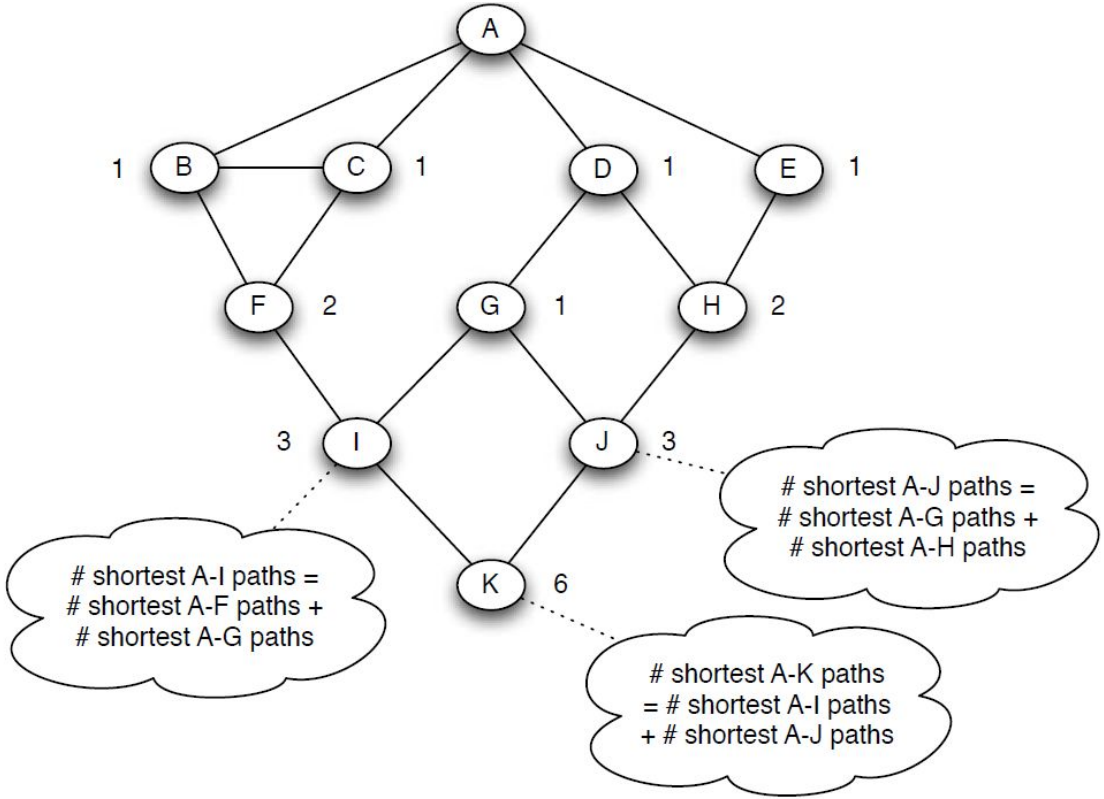
Shortest Path and Distance from farthest node

Diffusing computation

- Var: ID, deficiencyMap, parents, shortestPathsDistance
- Send message
- Track messages
- When deficit 0, all nodes that were called has responded
- Caller gains knowledge



Parent Value



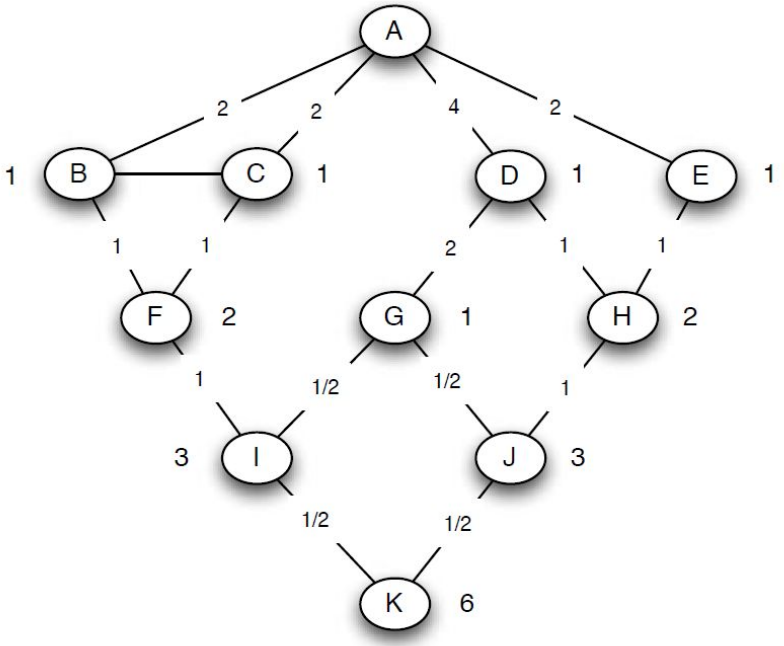
Flow value or The Betweenness

1. Betweenness captured in a matrix
2. Find leaf node
   a. IncomingValue for leaf node is 0
3. Find the edgeValue
   a. (Incoming value +1)PVi/PVj
   b. Where PVi is parent value of parent

      PVj is parent value of current node

4. Parent's incoming value = edge value from

child



Maximum Betweenness

- Servers send betweenness matrix as a string to neighbors and propagate common knowledge
- For each new string received, add it to current betweenness matrix
- Once received from all, matrix complete
- Find max value
  - If max value is from the edge from the current server, remove neighbor
- Initiate shortest path calculation again
  - At any time, if current node’s distance to farthest node increases, termination

Demo

Message Complexity

Graph	Nodes	Edges	Message Count
4-in-a-row	4	3	164
3-3 barbell	6	7	610
A-K example graph	11	16	2792