

PageRank

CS 111: Introduction to Computational Science

Spring 2019 Lecture #12

Ziad Matni, Ph.D.

Administrative

- Homework #5
 - Due **WEDNESDAY (5/15) @ 6:00 pm**
- Homework #6
 - Not this week...
- There will **NOT** be a lecture on
Thursday, May 16th
 - Section is STILL ON, however!

Reviewing Your Midterm #1 Exam

- Your grades are up on GauchoSpace
- To review your exams (optional, but not a bad idea), go to your TA's office hours:
 - Last name is A thru M See **Steven** (Tu. 1 – 3)
 - Last name is N thru Z See **Shiyu** (Fr. 10 – 12)
- When reviewing your exams:
 - Do not take pictures, do not copy the questions
 - TA cannot change your grade
 - If you have a legitimate case for grade change, the prof. will decide
 - Legitimate case = When we graded, we added the total points wrong
 - Not legitimate case = “Why did you take off N points on this question????”

Studying Networks

- Mathematically based on Graph Theory
- Describes systems using only nodes (vertices) and links
 - Nodes (the entities we're studying) connecting (linking) to one another
- Can describe a lot of naturally occurring phenomena
 - How certain cancer cell form
 - How people make social links with one another (i.e. social networks)
 - How words and their meanings are linked together
 - How documents end up being linked on the WWW

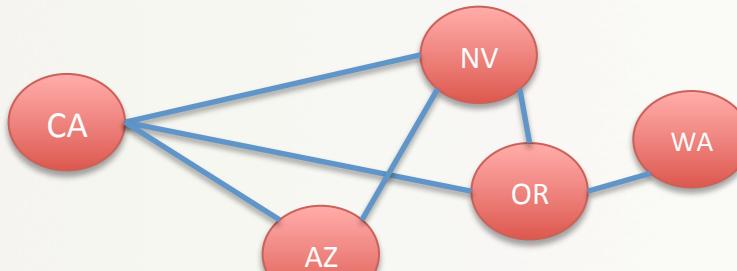
Who Studies Networks?

- Communication scholars
- Economists
- Anthropologists
- Sociologists
- Health care scholars
- Information specialists
- Physicists
- Computer Scientists and Engineers
- Biologists

and many, many others...

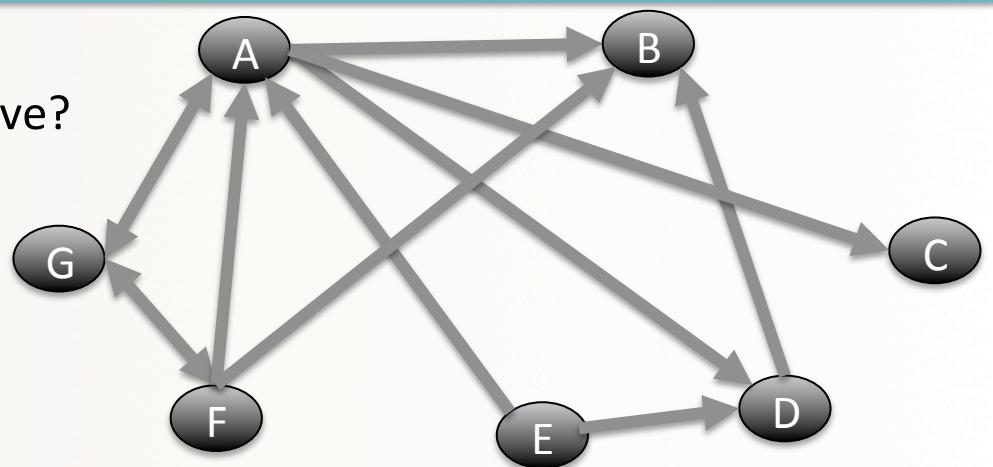
Nodes and Links

- The only 2 components of a network.
- Nodes (entities) are connected to each other via links (relationships)
 - **Nodes** can be: Web docs, people, organizations, words, etc...
 - **Links** can be: “has URL link to”, “knows”, “does business with”, etc...



Measures of the Importance of a Node

- Degree Centrality
 - How many links does a node have?
- Example:
 - A: 3 in-degree
4 out-degree
- Pros: It's simple
- Cons: It's too simple



Eigenvector Centrality

- A more sophisticated version of degree centrality
 - Not all highly central nodes have the same importance!
 - *Central Idea:*
A node is important if it is linked to other important nodes
- Like degree centrality, E.C. calculates the number of links on a given node
 - But then each adjacent node is **weighted** by its own centrality
 - So, a link to a high-centrality node “**counts for**” **more**

Eigenvector Centrality

- Popular in measuring *influence* and *access* to influential nodes in a network
 - The Google search algorithm (PageRank) uses a variation on this measure to decide how “important” a webpage is

Betweenness Centrality

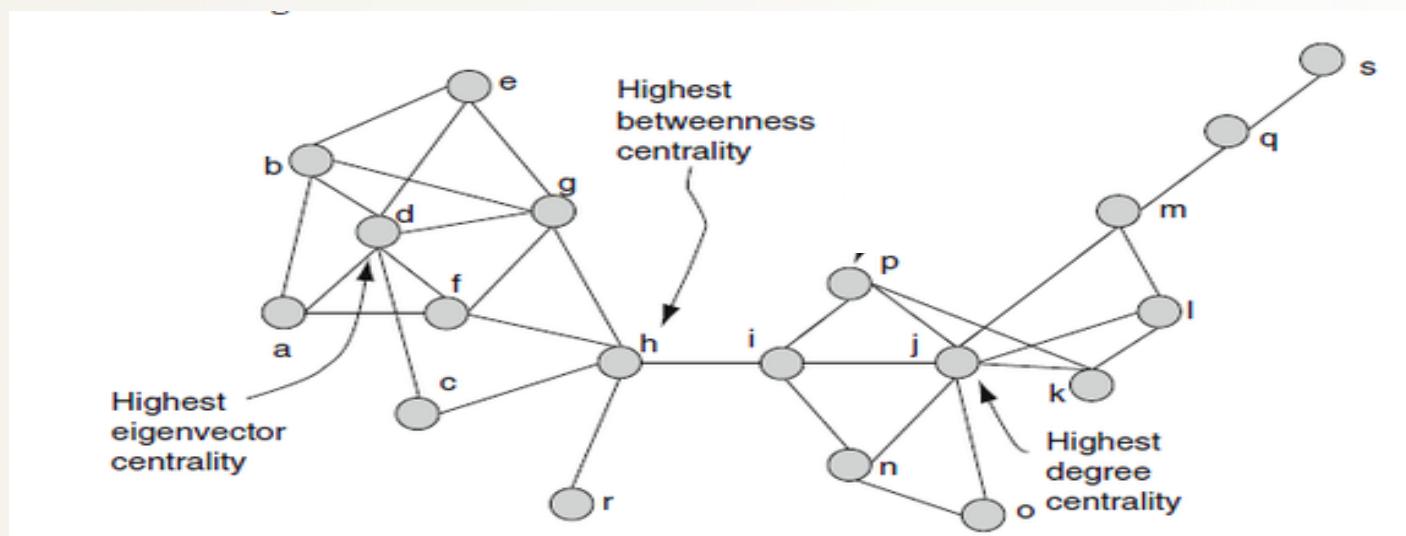
- A different focus on the idea of “importance”:
How “quickly” can a node get to any other node?
- Measures *how often* a given node falls along the **shortest path** between two other nodes
- *In other words:*
*How much of an **intermediary role** does a certain node play w/ respect to others?*

Betweenness Centrality

- If a node has high betweenness centrality
→ It has a *large potential for controlling flows through the network*
 - These nodes are prime candidates for elimination, if we want to disrupt the network

Why Different Centrality Measures?

- Different centrality measures answer different questions about the nodes in a network



Networks Can Be About... Anything

Network-centric views help us look at things from a certain perspective.

Example: **Network of Ideas**

[http://www.ted.com/talks/
eric_berlow_and_sean_gourley_mapping_ideas_worth_spreading](http://www.ted.com/talks/eric_berlow_and_sean_gourley_mapping_ideas_worth_spreading)
(7:51)

PageRank

- An algorithm famously used by Google Search to rank web pages in their search engine results.
- Looks at the **network** of documents on the Web and **analyzes** how they're connected and **ranks them** in importance to one another
- The idea is that this is the best way to ensure the **retrieval of relevant documents**
 - Proven to be quite superior to older methods
e.g. simply look for frequency of word occurrences in a document

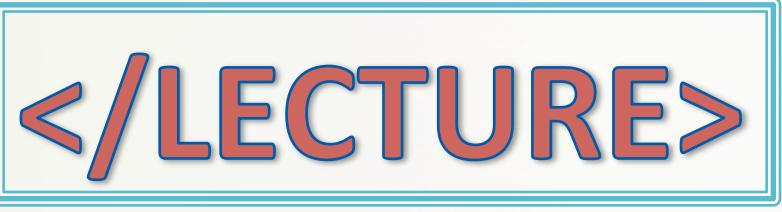
Examples – See Blackboard

Quick! To the Python-mobile!



Your To-Dos

- Homework 5 due WEDNESDAY
- Remember: NO LECTURE on Thursday



</LECTURE>