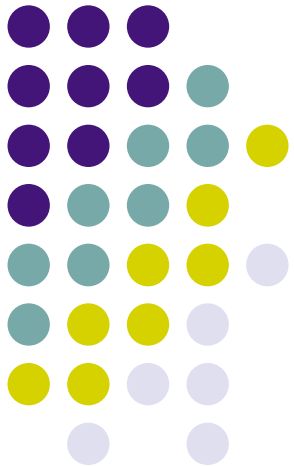
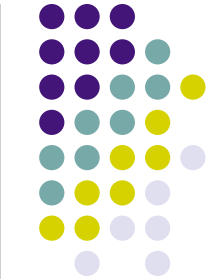
Social Networks and Graph Theory

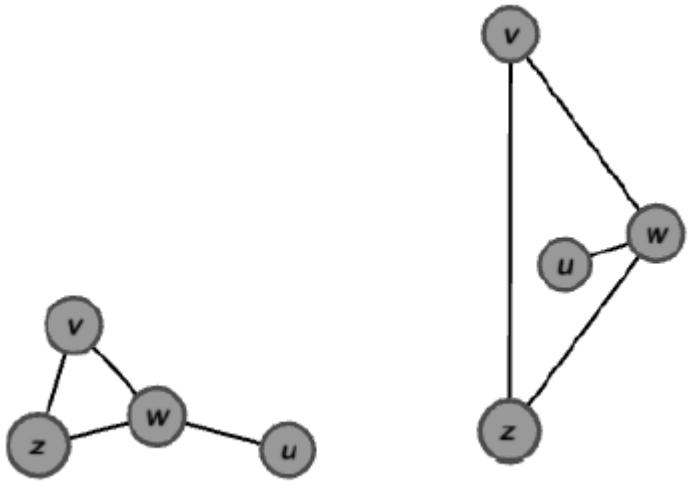
Holly Hirst
Mathematical Sciences
Appalachian State
NCCTM - November 2, 2017



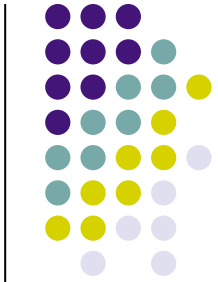
Graphs



- A ***graph*** is an ordered pair $G = (V;E)$, where
  - V is a set of ***vertices*** (aka nodes)
  - E is a set of ***edges*** (aka links) between the vertices in V.

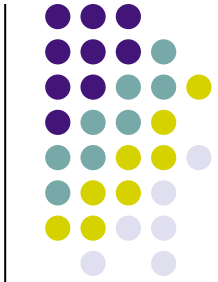


Social Networks



- Graphical Representation
  - *V* represents the set of all people/objects in the network
  - *E* represents the connections between people/objects
- Example: Ego Network Exploration
  - Track your electronic social contacts (texting, email, facebook, etc.) in one 24 hour period, and then ask each contact to confirm who on your list they also communicated with directly. Question: Who (besides you) is **central** to your network?
- These graphs will be ***simple, undirected*** graphs

Centrality

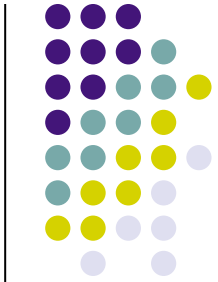


- We want to construct ***normalized*** numerical values in the interval [0,1] that measure how central a node is in our network.
  - 0 indicates not central at all
  - 1 indicates extremely central
- We will look at 3 today. There are others.
  - Degree Centrality
  - Betweenness Centrality
  - Closeness Centrality

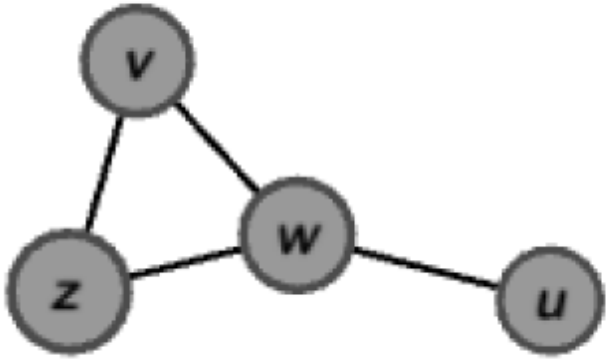
Degree Centrality

- ***A measure of the direct connections for a node.***
- A node's ***degree*** is the number of edges ***incident*** to it, which is also the number of nodes ***adjacent*** to it. To create a number in [0,1], divide by the number of **other** nodes in the network.
- ***So what?*** The node with the most connections has the most ability to influence others directly and immediately, which is important to know if we are trying to stop someone with a communicable disease from infecting others nearby.

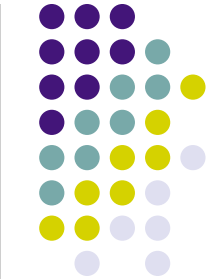
Degree Centrality



- Normalized number of incident edges
- *w:* 3/3 = 1
- u: 1/3
- v: 2/3
- z: 2/3

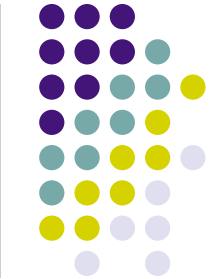


Betweenness Centrality



- ***A measure of how often a node appears in the interior of shortest paths in the network.***
- For each pair of other nodes, divide the number of ***shortest paths*** that pass **through** a node by the number of all shortest paths; sum and divide by the number of pairs of other nodes. (ACK!)
- ***So what?*** Betweenness measures how critical a node is for flow through an entire network, especially if our goal is interrupting flow of cash or information in a terrorist network.

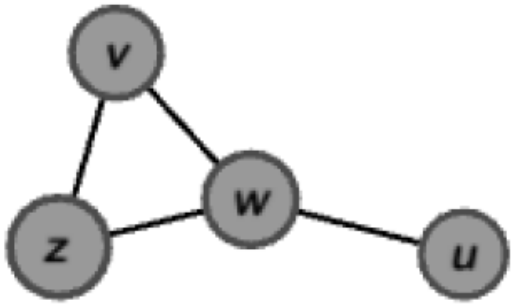
Betweenness Centrality



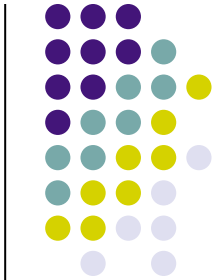
- Shortest paths:

(u,w):	1;	(u,w,v):	2
(u,w,z):	2;	(w, v):	1
(w,z):	1;	(v,z):	1

- w: (1/1 + 1/1 + 0/1)/3 since w is in the interior of u-z and u-v, but not v-z.
- u, v, z: 0



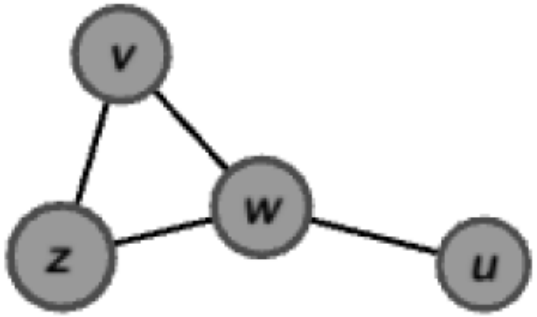
Closeness Centrality



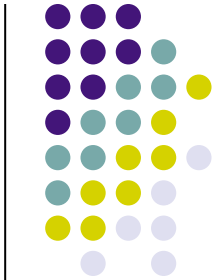
- ***A measure of the average distance from a node to all other nodes in the network.***
- The average ***length*** of the shortest paths between the node and all other nodes in the network is informative, but we take the reciprocal so that our number is in [0,1], with 1 being “best.”
- ***So what?*** Closeness identifies nodes that can get information out most efficiently to the entire network.

Closeness Centrality

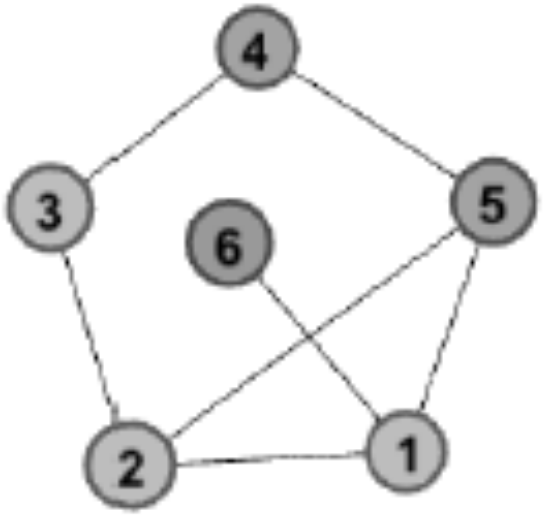
- w: 1 – the reciprocal of (1+1+1)/3 = 1 since the shortest path from w to each of the other nodes has length one, and there are three other nodes.
- v, z: 3/4 – the reciprocal of (2+1+1)/3 = 4/3
- u: 3/5 – the reciprocal of (1+2+2)/3 = 5/3



Another Example

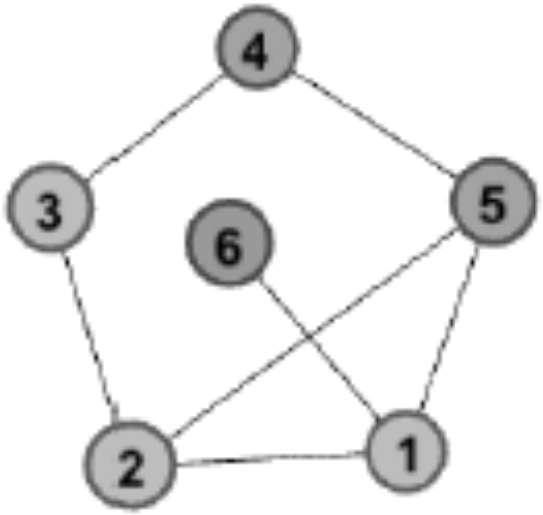


- Can you calculate Degree, Betweenness, and Closeness centrality?



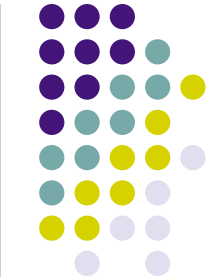
Another Example

- Can you calculate Degree, Betweenness, and Closeness centrality?



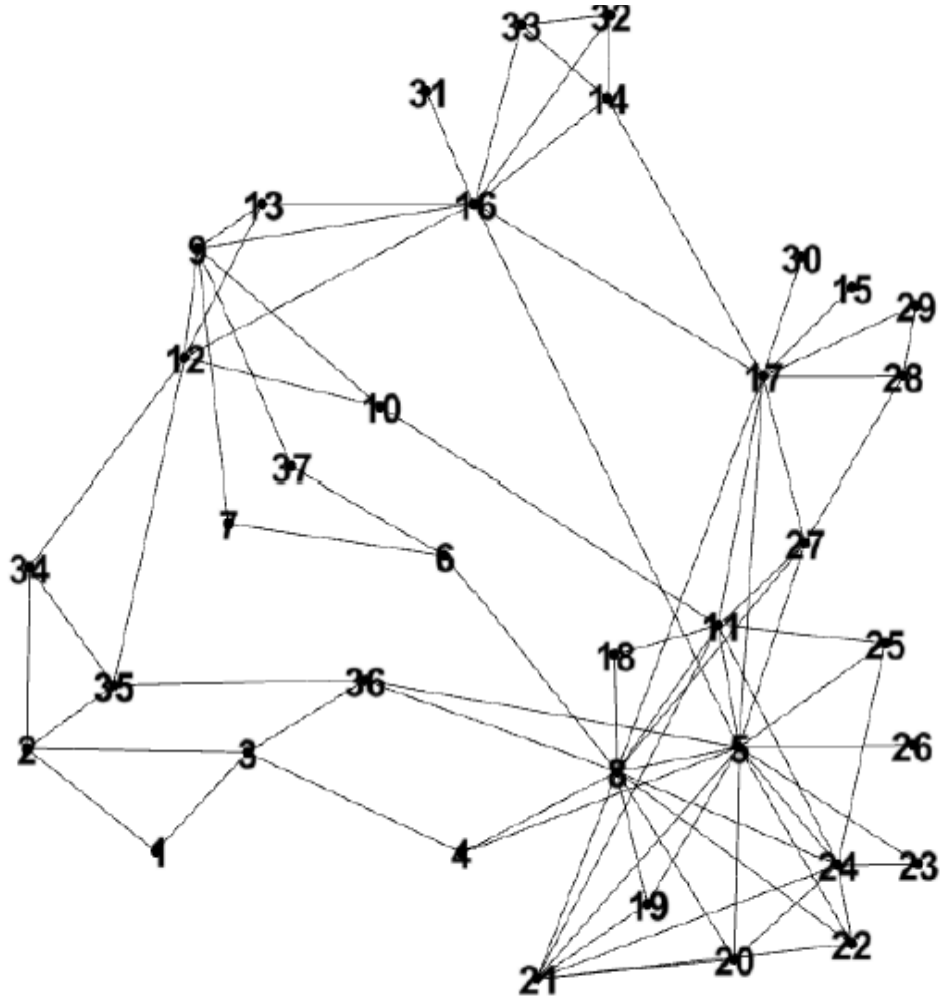
Node	Degree	Closeness	Betweenness
1	3/5	5/7	4/10
2	3/5	5/7	2.5/10
3	2/5	5/9	.5/10
4	2/5	5/9	.5/10
5	3/5	5/7	2.5/10
6	1/5	5/11	0/10

9/11 Terrorist Network



- In his article “Uncloaking Terrorist Networks,” Valdis Krebs analyzes the contact information for the 9/11 Terrorists. He was looking at the flow of money, information, and expertise among more than 70 people identified after the attack as being part of the network of the 19 hijackers. Using a subset of the data that includes the 19 hijackers and their closest associates, use measures of centrality to name the top three contacts we should have targeted in a raid.
- (journals.uic.edu/ojs/index.php/fm/article/view/941/863)

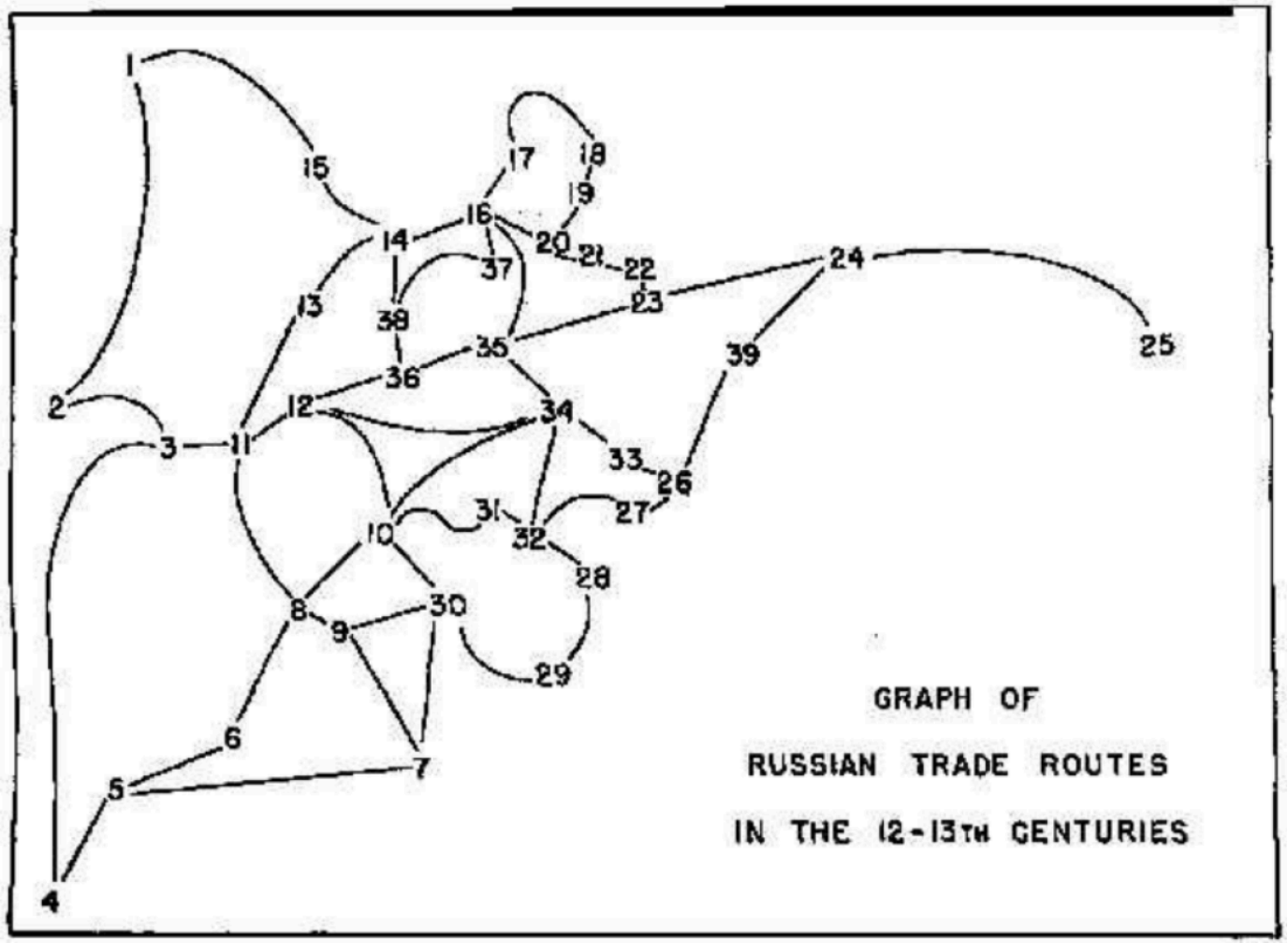
9/11 Terrorist Network



Russian Trade Routes

- In 1965, Forrest Pitts produced a graph of the major medieval urban locations along Russian river transportation routes (people and goods traveled by boat in summer and by sleigh on the frozen rivers in winter) in an effort to see if a graphical analysis could shed light on the situation. He used a self-designed analysis of shortest paths to show that Kolomna (34 on the network) was the in fact the best connected place with Moscow (35) ranking second. Do our measures of centrality support this conclusion?
- www.analytictech.com/networks/pitts.htm

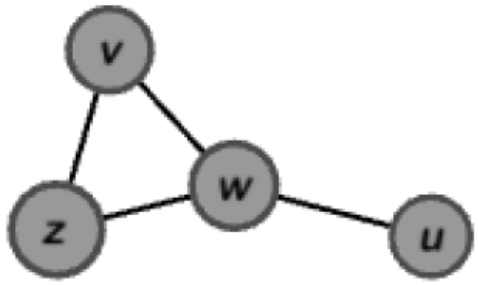
Russian Trade Routes



Gephi – Software for Graph Analysis

- Build a CSV file showing connections.

	A	B	C
1	u	w	
2	w	v	z
3	v	z	
4			
5			



- Open in Gephi and indicate an undirected graph.
- Drag nodes and use tools at the bottom of the Overview window.

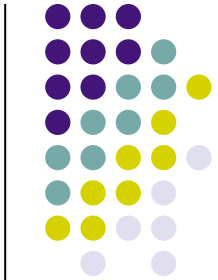
Gephi….

- Run the average degree and network diameter statistics
- In the Data Laboratory View:

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export tabl

Id	Label	Interval	Degree	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality
u	u		1	2.0	0.6	0.666667	0.0
w	w		3	1.0	1.0	1.0	0.666667
v	v		2	2.0	0.75	0.833333	0.0
z	z		2	2.0	0.75	0.833333	0.0

Contact Info



hirsthp@appstate.edu

appstate.edu/~hirsthp/

gephi.org/users/download/