NUMB3RS Activity: It's All Connected
Episode: "The O.G."

Topic: Social Network Analysis **Grade Level:** 9 - 12

Objective: To introduce networks and measures of centrality

Time: 25 minutes

Introduction

In "The O.G.," the FBI is investigating a series of gang killings. Don Eppes suspects that these killings are more than random acts of retaliation, so he asks for Charlie's assistance. Based on information about how the gang members are connected to each other, Charlie decides to create a mathematical network to model how information travels in the gang. From this model, he hopes to determine who may be the next victim and which gang member might be assisting the killers.

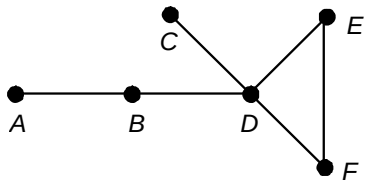
Social network analysis maps and measures the connectedness of members within groups that share common links. These links include relational connections, informational links, or physical proximities. A **network** is a map that consists of **nodes** (vertices) representing individuals within a group and **links** (edges) connecting the nodes to show how information flows within the group. From the network, it is possible to create a **shortest path matrix** illustrating the fewest number of links between any two connected nodes.

Analyzing the network involves examining how the individuals are connected and what type of role each individual plays within the network. Individual roles may or may not be consistent with any hierarchy or organization that exists within the group being analyzed. For example, an employee who has the trust and confidence of coworkers may be more influential than the manager of that particular group. This type of information can be determined using social network analysis.

Measuring the **centrality** (location) of a node within the network is one method for analyzing the network. We will examine two types of centrality, **degree centrality** which measures the direct connections within the network and **closeness centrality** which measures how quickly a node can connect to other nodes.

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

Consider the network shown.



Notice that in the network, each person (node) can be connected to 5 other people (nodes). To find the **degree centrality** of any one person, divide the number of direct connections (1-links) to other network members, by the number of possible connections.

For example, *A* has one direct link (to *B*). So, the degree centrality of *A* is $\frac{1}{5} = 0.2$.

To find **closeness centrality**, you will need to examine the shortest paths between members.

You can describe the shortest paths using a matrix. For example, there is one link in shortest the path from *A* to *B*, there are three links in the shortest path from *A* to *C*, and so on.

	A	*B*	*C*	*D*	*E*	*F*
A	0	1	3	2	3	3
B	1	0	2	1	2	2
C	3	2	0	1	2	2
D	2	1	1	0	1	1
E	3	2	2	1	0	1
F	3	2	2	1	1	0

To find the closeness centrality of *A*, use information from the matrix and the formula for closeness.

$$\text{Closeness} = \frac{n-1}{(l_1 \cdot 1) + (l_2 \cdot 2) + (l_3 \cdot 3) + \ldots}$$

In the formula, n is the number of people (nodes) in the network, l_1 is the number of 1-link paths, l_2 is the number of 2-link paths, l_3 is the number of 3-link paths, etc. Some students may notice that the denominator is also equivalent to the row sum in the shortest path matrix. The greater the value of the closeness, the better connected the person (node) is with the other members of the network (nodes).

This network has 6 people (nodes) in all, so $n = 6$. Notice that *A* has one 1-link path, one 2-link path, and three 3-link paths. So, the closeness centrality of *A* is

$$\frac{6-1}{(1 \cdot 1) + (1 \cdot 2) + (3 \cdot 3)} = \frac{5}{1+2+9} = \frac{5}{12} \approx 0.42.$$

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

The degree centrality and closeness centrality for each person are given in the table below.

Node	***A***	***B***	***C***	***D***	***E***	***F***
Degree Centrality	$\frac{1}{5}=0.2$	$\frac{2}{5}=0.4$	$\frac{1}{5}=0.2$	$\frac{4}{5}=0.8$	$\frac{2}{5}=0.4$	$\frac{2}{5}=0.4$
Closeness Centrality	$\frac{5}{12}\approx 0.42$	$\frac{5}{8}=0.625$	$\frac{5}{10}=0.5$	$\frac{5}{6}\approx 0.83$	$\frac{5}{9}\approx 0.56$	$\frac{5}{9}\approx 0.56$

Because *D* has the greatest degree centrality and the greatest closeness centrality, *D* is both the most active and the best connected person in this network.

Discuss with Students

Start this activity with a simple exploratory exercise with students. Use the examples and questions below to define nodes, networks, and links as described.

1. Tom shares information with Carol, Wanda, Fred, and Sam. Sam shares information with Fred and Wanda.
   a. Describe how to create a diagram showing how the information is shared.
   b. What is this type of diagram called?

2. How could you use a table or matrix to show shortest paths between the people sharing information found in the diagram?

3. Using the diagram, how might you tell who the most important person in the group is?

Discuss with Students answers:
1a. *You could let a node represent a person and a link represent when two nodes share information.* **1b.** *This is called a network or a graph.* **2.** *Each row and column head could represent a person. The entries in each cell would represent the number of links found in the connection to each of the other nodes in the network.* **3.** *Sample response: The person with the most connections might be the most important.*

Student page answers:
1. *Sample matrix:*

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
A	0	1	1	1	2	1	2	2	3	4
B	1	0	2	1	1	2	1	2	3	4
C	1	2	0	1	2	1	2	2	3	4
D	1	1	1	0	1	1	2	2	3	4
E	2	1	2	1	0	2	1	2	3	4
F	1	2	1	1	2	0	1	1	2	3
G	2	1	2	2	1	1	0	1	2	3
H	2	2	2	2	2	1	1	0	1	2
I	3	3	3	3	3	2	2	1	0	1
J	4	4	4	4	4	3	3	2	1	0

2a.

Node	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***
Degree Centrality	0.44	0.44	0.33	0.56	0.33	0.56	0.44	0.33	0.22	0.11

2b. *The most active members (members with the greatest degree centrality) are Damion and Fredrico.*

3a.

Node	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***
Closeness Centrality	0.53	0.53	0.50	0.56	0.50	0.64	0.60	0.60	0.43	0.31

3b. *The best connected members (the member with the greatest closeness centrality) is Fredrico.*

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

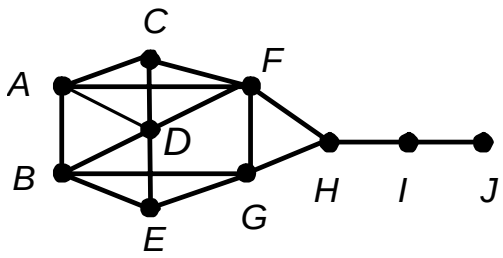
Name: _____ Date: _____

NUMB3RS Activity: It's All Connected

The FBI is investigating a series of killings of members of the East Side Gang. Agent Don Eppes suspects that these killings are more than random acts of retaliation, so he asks for Charlie's assistance. They want to determine who the next target might be, and who is providing a rival gang with information about who to kill to have the greatest impact on the East Side Gang. Charlie suggests that through **social network analysis**, he may be able to help.

Social network analysis uses a network to model how information flows within a group. A **network** is a map that consists of **nodes** (vertices) representing the individuals in a group and **links** (edges) connecting the nodes to show the flow of information.

Suppose that the members of the gang are Antony, Bernard, Carmon, Damion, Edwin, Fredrico, Garfield, Harry, Ivan, and Jake. The network below shows how information travels in the East Side Gang. (Each letter represents the gang member whose name begins with that letter. Each link shows which members share information, and the information can be shared in both directions.)



[Source: **http://www.orgnet.com/sna.html**]

1. A **shortest path matrix** is a matrix that shows the fewest number of links between connected people (nodes). Using the network above, complete the shortest path matrix for the East Side Gang. The row for Antony has been completed for you.

Shortest Path Matrix

	Antony	Bernard	Carmon	Damion	Edwin	Fredrico	Garfield	Harry	Ivan	Jake
Antony	0	1	1	1	2	1	2	2	3	4
Bernard										
Carmon										
Damion										
Edwin										
Fredrico										
Garfield										
Harry										
Ivan										
Jake										

Once Charlie has made the network and the shortest path matrix, he can analyze them to help determine who the most important members of the gang are, and who the next victim(s) might be.

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

2. One measure used to analyze a network is **degree centrality**. Degree centrality measures how active a person (node) is within the network. To find the degree centrality of a person or member, divide the number of direct connections (1-links) to other members by the total number of members he or she could be connected to.

In the network of the East Side Gang, each person can be connected to 9 other members. *A* has four 1-link connections, so the degree centrality of *A* is $\frac{4}{9} \approx 0.44$.

a. Use the shortest path matrix you created in question 1 to determine the degree centrality of each person.

Node	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***
Degree Centrality	0.44									

b. Which gang member(s) appear to be the most active? _____

3. Another measure that could be used to analyze a network is **closeness centrality (closeness)**. Closeness measures how easily a member connects with all other members in the network. Only links in the *shortest* path to each member are counted when finding closeness. The greater the closeness centrality, the more easily that person can interact with *all* the other members of the gang. To find the closeness of a member, use the formula

$$\text{Closeness} = \frac{n-1}{(l_1 \cdot 1) + (l_2 \cdot 2) + (l_3 \cdot 3) + \ldots}$$

where n is the number of members in the network, l_1 is the number of 1-link connections, l_2 is the number of 2-link connections, l_3 is the number of 3-link connections, etc.

In the network of the East Side Gang, there are 10 members. *A* has four 1-link connections, three 2-link connections, one 3-link connection, and one 4-link connection. So, the closeness of *A* is $\frac{10-1}{(4 \cdot 1) + (3 \cdot 2) + (1 \cdot 3) + (1 \cdot 4)} = \frac{9}{4 + 6 + 3 + 4} = \frac{9}{17} \approx 0.53$.

a. Use the shortest path matrix you created in question 1 to determine the closeness centrality of each member.

Node	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***
Closeness Centrality	0.53									

b. Which gang member(s) appear to be the best connected? _____

4. Charlie tells Don that he has a likely target based on his analysis of the network. Who do you think Charlie suggested? Why?

5. While some members of the gang may not appear to be well connected within the gang, they may get and share information with others outside the gang. They are referred to as **peripheral players**. Suppose that Jake was once a member of a rival gang. Charlie suspects that Jake is getting information from an East Side Gang member and leaking the information to the rival gang. Which East Side Gang member is the most likely to be giving this information to Jake? Why?

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT

The goal of this activity is to give your students a short and simple snapshot into a very extensive mathematical topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

In this activity, you have seen how Social Network Analysis can be used to draw conclusions based on the spread of information between members of a gang. In addition to finding degree centrality and closeness centrality, there are other ways to find information about the members of the gang.

- **Betweenness centrality (betweenness)** is a third measure often used to analyze a network. Betweenness takes into consideration which nodes lie along the shortest path between other nodes. The nodes with the greatest betweenness have the most influence over what information flows within the network.

  The betweenness of a node is based on the probability that the shortest paths between any two other nodes passes through that node. This probability is then divided by the total number of possible connections involving the other nodes.

  Determine which member(s) of the East Side Gang had the greatest influence of the flow of information in the gang.

  For more information on betweenness centrality, go to page 8 of **http://mrvar.fdv.uni-lj.si/sola/info4/uvod/part4.pdf**

- **Ego networks** are networks used to analyze one particular member of a group and how that member interacts with others. Identify the differences and similarities of ego network and the kind of networks you worked with in this activity. Then describe where you might use an ego network in a real-world setting.

  For more information on ego networks, visit the following Web sites:

  http://www.analytictech.com/networks/egonet.htm

  http://www.orgnet.com/booknet.html

© 2005 Texas Instruments Incorporated
Terry Souhrada, University of Montana, Missoula, MT