Module 4: Analytics Lab

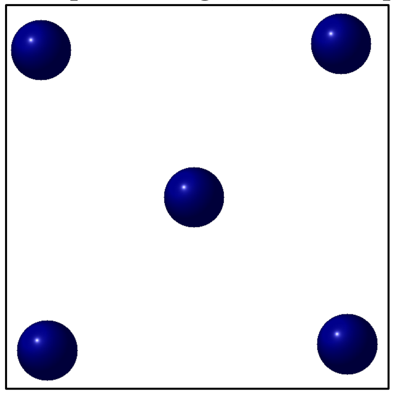
Goals for Module 4: Analytics Lab

~ Participants will practice the social network measures introduced in the tutorial by calculating the measures by hand based on a small practice network. By practicing the measures by hand, participants should develop a stronger understanding of some basic social network analytics.

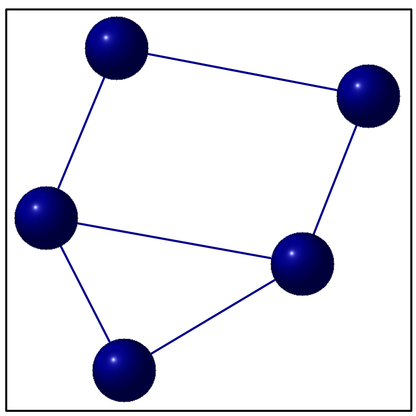
Density

Density is a single value for a network calculated as the ratio of edges present in the network to the total number of all edges possible in a network. Answer the questions about density using the social network images below the questions. The answers are available in the review section of this lab.

1. In any network of 5 nodes, how many possible edges are there?
2. What is the density of this network?
   Hint: Density = (the number of present edges)/(the total possible edges)

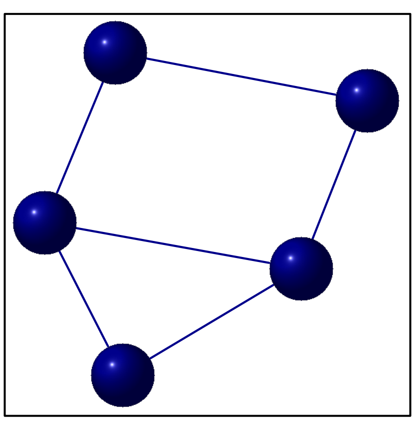


3. What is the density of this network?

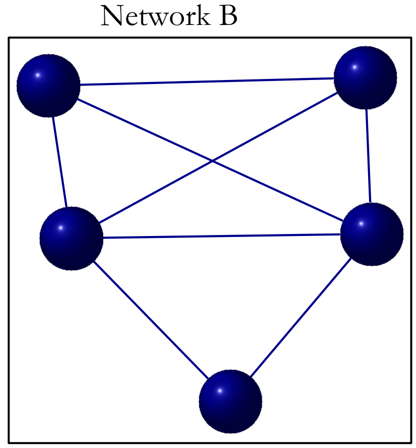


4. Which network is denser, Network A or Network B?

Network A



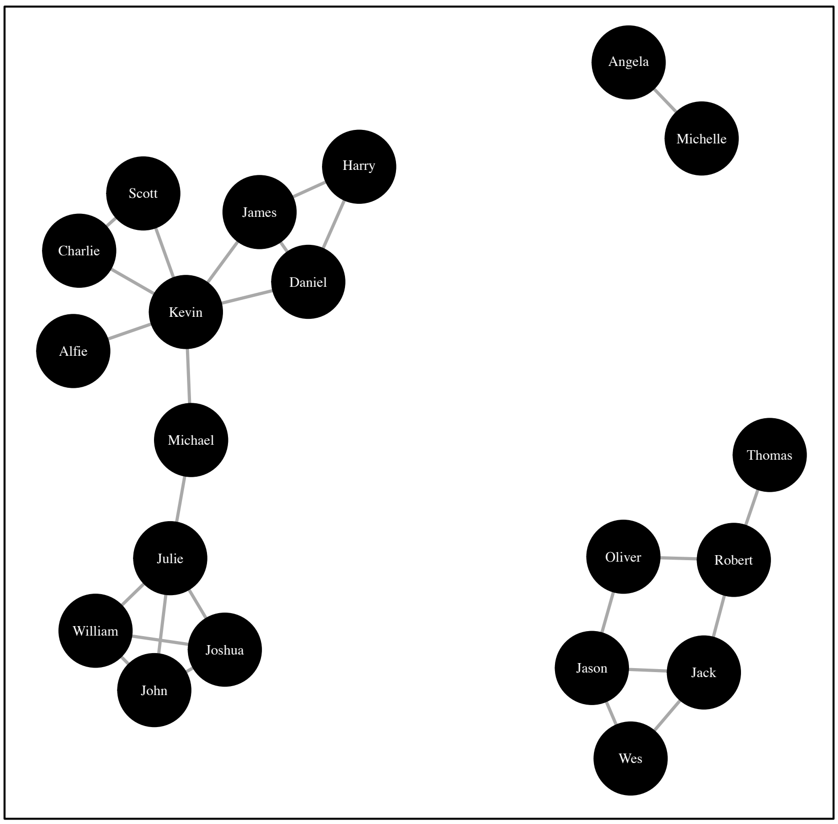
Network B



Components

Components measure the different unconnected pieces of a social network. Answer the questions about components using the social network image below. The answers are available in the review section of this lab.

1. How many components are in this network?
2. Are there any isolates in this network?
3. How large is the largest component?
4. What is the size of the smallest component?



Degree

Degree counts the number of ties for each node. Answer the questions about degree using the social network image from the previous page. The answers are available in the review section of this lab.

1. Below is a table of all of the degree scores from the network on the previous page. Complete the table for the missing degree scores.

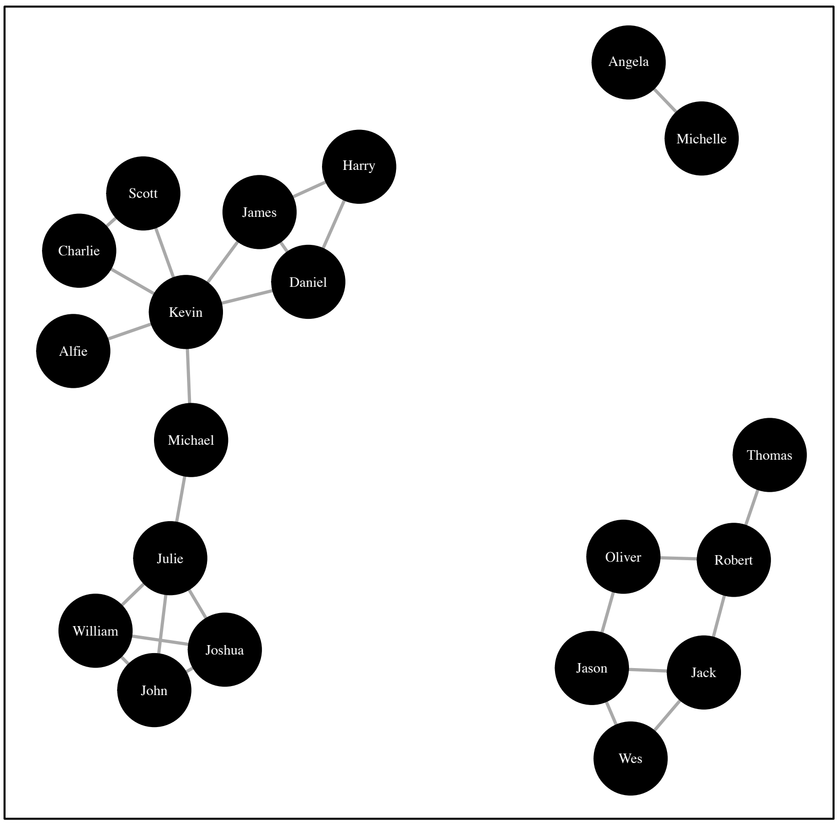
Node	Degree
Jason	3
Jack	3
Oliver	2
Wes	
Thomas	1
Robert	3
Angela	
Michelle	
John	3
Joshua	3
William	3
Julie	
Scott	2
Charlie	2
Daniel	3
Harry	2
Kevin	
James	3
Alfie	1
Michael	2

2. Which node has the highest degree in this network?
3. What is the lowest degree score in this network?
4. How many nodes have the lowest degree score in this network?
5. What do you think the average degree score is for this network?

Cut-points

Cut-points are nodes that when removed from the network increase the number of components in the network. Answer the question about cut-points using the social network image below. The answer is available in the review section of this lab.

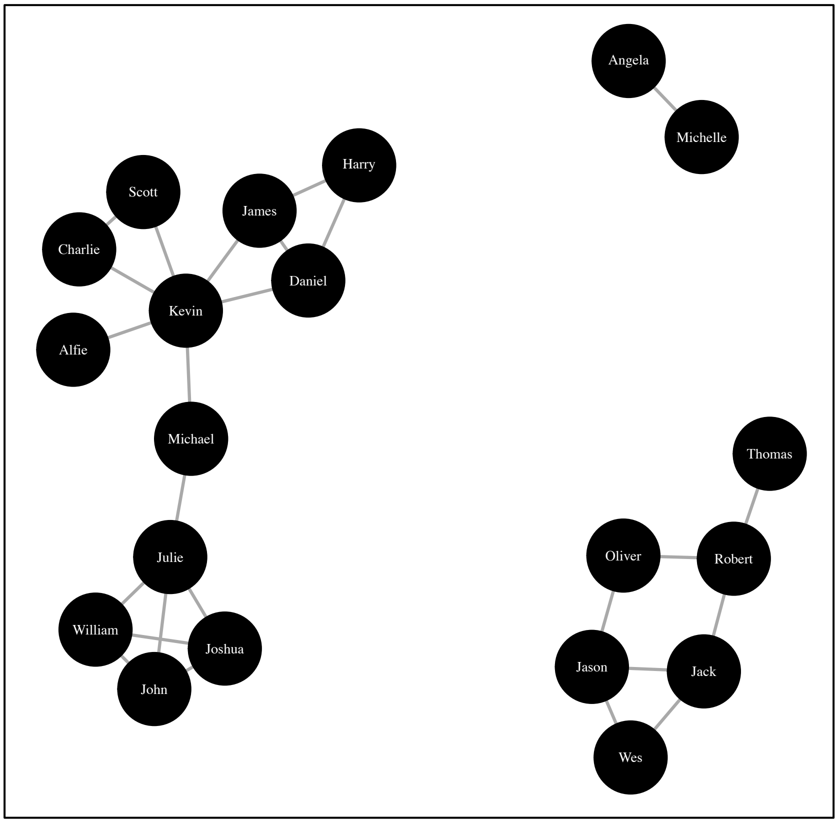
1. There are 4 nodes that are cut-points in this network. Can you identify them? Hint: Nodes Angela and Michelle are not cut-points because removing either one of them would not increase the number of components in the network. Their dyad counts as 1 component, and if you removed 1 of them, the remaining isolate would still only count as 1 component.



K-core

K-core measures connected areas of the network with the same degree often identifying subgroups in the network. Answer the question about k-core using the social network image below. The answer is available in the review section of this lab.

1. The majority of nodes in this example network below (12 to be exact) have a k-core of 2, meaning they belong in areas of the network with a degree of at least 2. Can you identify the 4 nodes that have a k-core of 1 and the 4 nodes that have a k-core of 3?

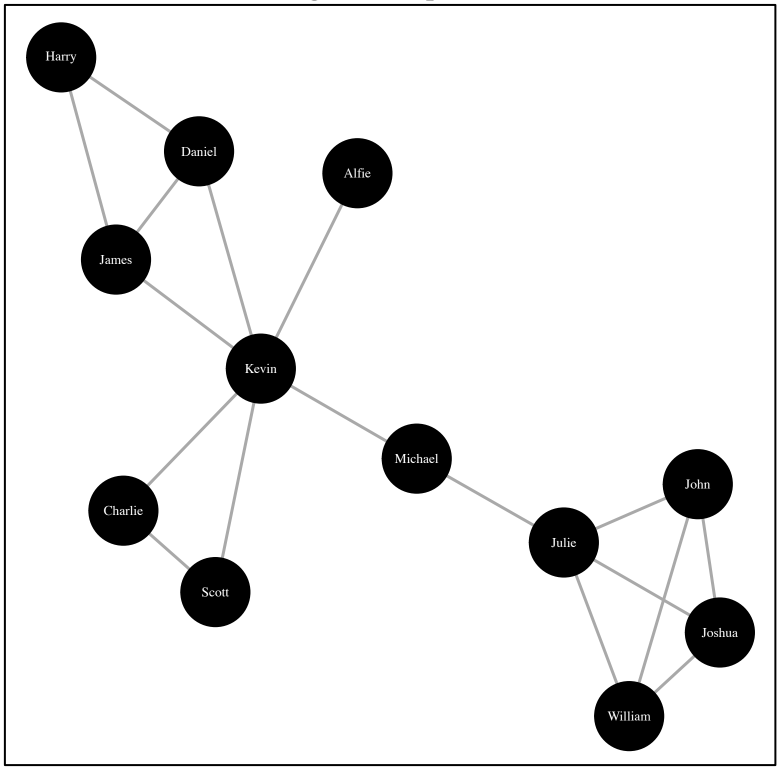


Distance

Geodesic distance measures the shortest path between two nodes within the same component. For this section, our example network includes only the largest component. Answer the questions about distance using the social network image on the next page. The answers are available in the review section of this lab.

1. Below the network image of the largest component on the next page is half of a matrix showing the geodesic distances between each pair of nodes in the largest component. Complete the geodesic distance table for node Michael.
2. What is the diameter of this network?
3. Which nodes are the farthest from each other in this component?
4. Just by looking at the matrix, what would you guess the average geodesic distance to be in this component?

Largest Component



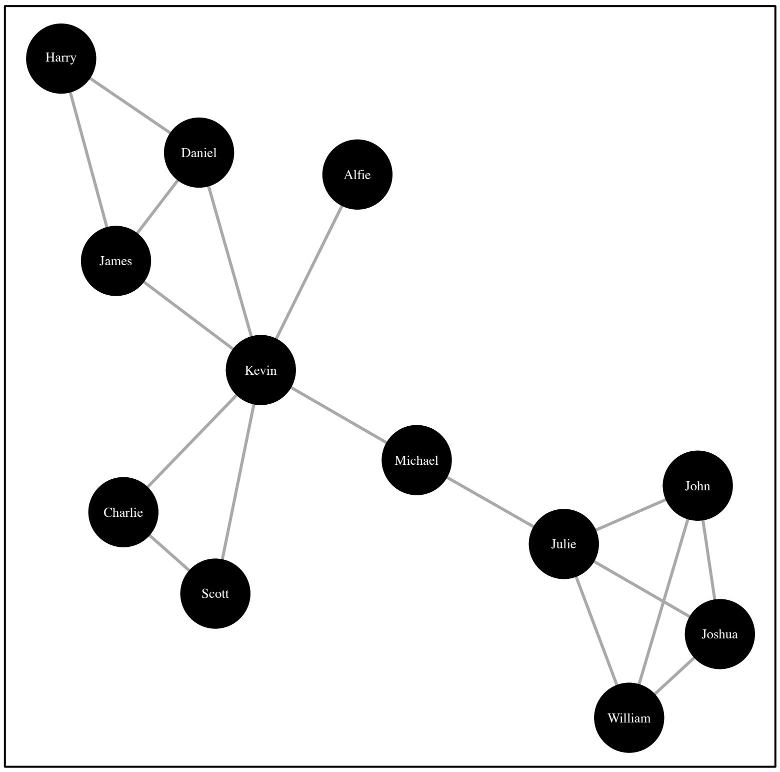
Geodesic Distances

	John	Joshua	William	Julie	Scott	Charlie	Daniel	Harry	Kevin	James	Alfie	Michael
John	--											
Joshua	1	--										
William	1	1	--									
Julie	1	1	1	--								
Scott	4	4	4	3	--							
Charlie	4	4	4	3	1	--						
Daniel	4	4	4	3	2	2	--					
Harry	5	5	5	4	3	3	1	--				
Kevin	3	3	3	2	1	1	1	2	--			
James	4	4	4	3	2	2	1	1	1	--		
Alfie	4	4	4	3	2	2	2	3	1	2	--	
Michael												--

Brokerage/ Betweenness

Brokers reside on many geodesic distances that connect other nodes. One measure of brokerage is called betweenness, which counts the number of geodesic paths that nodes are between. Similar to the section above on geodesic distance, our example network includes only the largest component. Answer the questions about brokerage/betweenness using the social network image below. The answers are available in the review section of this lab.

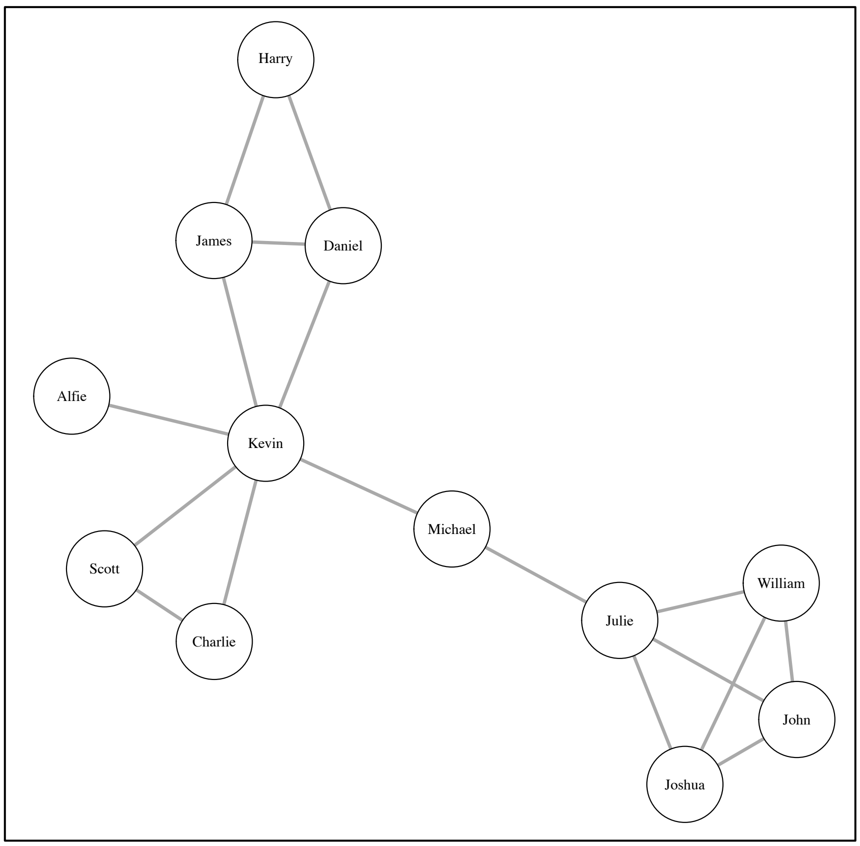
1. Which node is the most important broker in this network?
2. Which 7 nodes have a betweenness score of 0?
3. Who has a higher betweenness score, Kevin or Julie? Why?
4. Who has a higher betweenness score, Daniel or James? Why?



Neighborhood

Neighborhoods zoom in on subgroups within a certain distance from a focal node or ego. Answer the questions about neighborhood using the social network image below. The answers are available in the review section of this lab.

1. Let's assume that Daniel is our ego or node of interest. Color in the largest component below using red for the ego, orange for the 1st order neighborhood, and yellow for the nodes added in the 2nd order neighborhood. Leave all neighborhoods greater than 2 white.
2. How many nodes are in Daniel's 1st order neighborhood? Who are they?
3. What is the largest order neighborhood needed to include everyone in this component?
4. Which node is more likely to influence Daniel, Kevin or Michael?



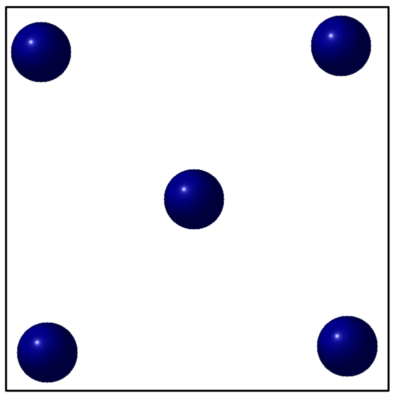
Review of Module 4: Analytics Lab

This lab asked some relatively simple calculation questions about example networks. The answers to the questions are here in the review section of the lab.

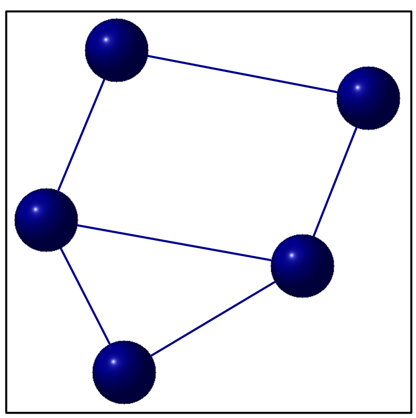
Answers to the Density section questions:

(1) In a network of 5 nodes, if every single node was connected to every other node there would be 10 edges, so 10 edges are possible among a network of 5 nodes.

(2) There are 0 edges in this network of 5 nodes. The total possible edges in a network of 5 nodes is 10. The formula to calculate this network's density is 0/10=0. This network has a density of 0.

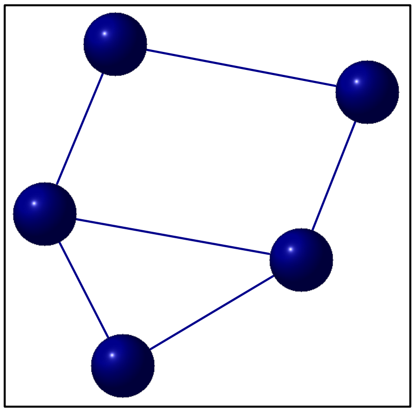


(3) There are 6 edges present in this network of 5 nodes. The total possible edges in a network of 5 nodes is 10. The formula to calculate this network's density is 6/10=0.6. The density of this network is 0.6, or we could say that 60% of the possible edges are present in this network.

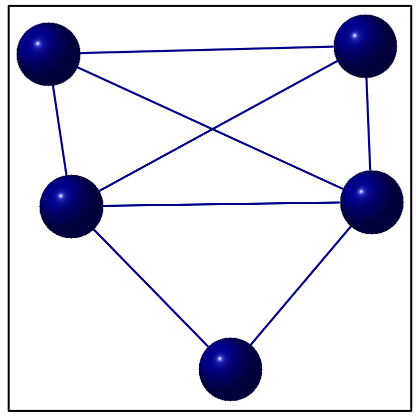


(4) Networks A and B both have 5 nodes, but Network B has 8 edges compared to Network A's 6 edges. More edges between the same number of nodes results in a denser network. Mathematically, we calculated the density for Network A to equal 0.6 in question 3 of this section. The density for Network B = 8/10, which is 0.8 or 80% of all possible ties are present in Network B. Since 0.8 is larger than 0.6, we know that Network B is denser than Network A, and Network A is more sparse than Network B.

Network A



Network B



Answers to the Components section questions:

(1) There are 3 components in this network. Three separate pieces of the network do not connect.

(2) There are no isolates in this network because every node in the network has at least 1 tie to another node, and isolates are defined as having 0 ties.

(3) The largest component contains 12 nodes.

(4) The smallest component contains 2 nodes: Angela and Michelle.

Answers to the Degree section questions:

(1) The table below includes all of the degree scores for the nodes in the example network.

Node	Degree
Jason	3
Jack	3
Oliver	2
Wes	2
Thomas	1
Robert	3
Angela	1
Michelle	1
John	3
Joshua	3
William	3
Julie	4
Scott	2
Charlie	2
Daniel	3
Harry	2
Kevin	6
James	3
Alfie	1
Michael	2

(2) Node Kevin has the highest degree score in this network with 6 ties.

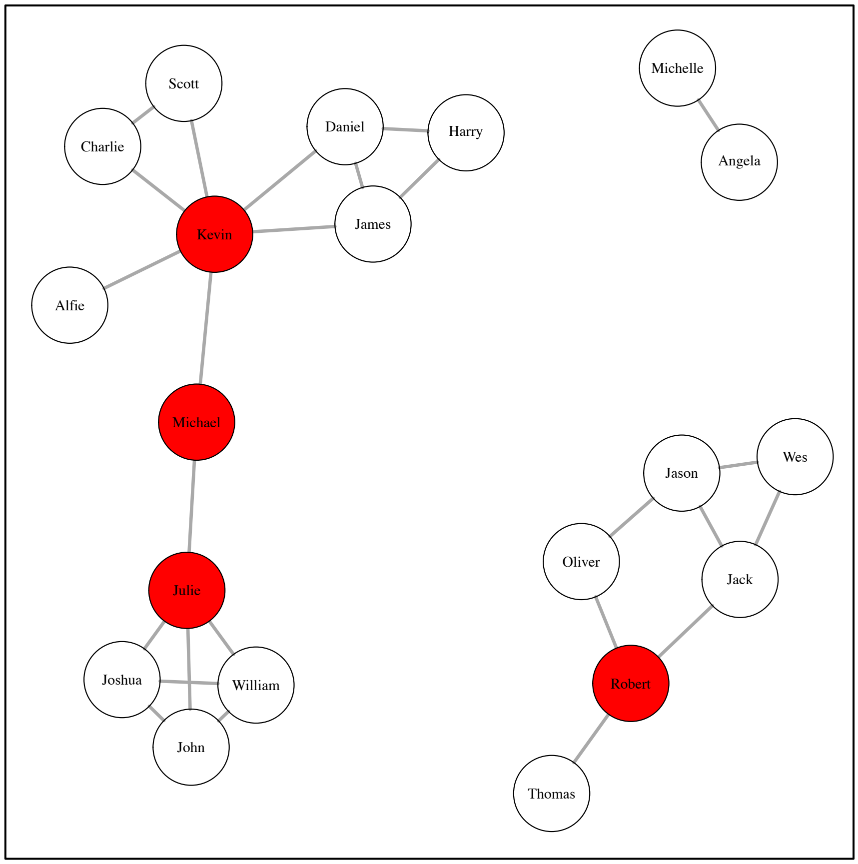
(3) The lowest degree score in this network is 1. These are the nodes that have only a single tie to one other node.

(4) 4 nodes have a degree score of 1: Thomas, Angela, Michelle, and Alfie.

(5) The average degree for this network is 2.5 ties.

Answer to the Cut-point section question:

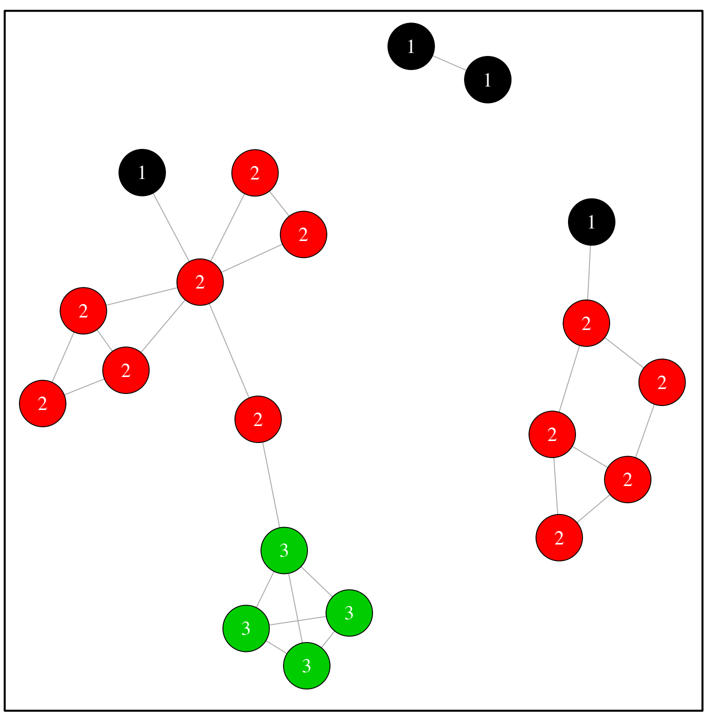
(1) The 4 cut-points in this example network are Robert, Kevin, Michael, and Julie. The cut-points are identified in the image below in red. They are the cut-points because the removal of any one of these nodes would break the network into additional components.



Answer to the K-core section question:

(1) The table below includes all of the k-core scores for the nodes in the example network, and the network plot below shows the example network with the k-core score replacing the name labels and the nodes colored by k-core score.

Node	K-core
Jason	2
Jack	2
Oliver	2
Wes	2
Thomas	1
Robert	2
Angela	1
Michelle	1
John	3
Joshua	3
William	3
Julie	3
Scott	2
Charlie	2
Daniel	2
Harry	2
Kevin	2
James	2
Alfie	1
Michael	2



Answers to the Distance section questions:

(1) The distance table below is complete and includes all of the distances between node Michael and the other 11 nodes.

	John	Joshua	William	Julie	Scott	Charlie	Daniel	Harry	Kevin	James	Alfie	Michael
John	--											
Joshua	1	--										
William	1	1	--									
Julie	1	1	1	--								
Scott	4	4	4	3	--							
Charlie	4	4	4	3	1	--						
Daniel	4	4	4	3	2	2	--					
Harry	5	5	5	4	3	3	1	--				
Kevin	3	3	3	2	1	1	1	2	--			
James	4	4	4	3	2	2	1	1	1	--		
Alfie	4	4	4	3	2	2	2	3	1	2	--	
Michael	2	2	2	1	2	2	2	3	1	2	2	--

(2) The diameter of a network is the longest geodesic distance. In this network, the diameter equals 5.

(3) There are 3 geodesic distances requiring 5 steps in this component. Node Harry is 5 steps from nodes John, William, and Joshua and vice versa.

(4) The mean of all 66 of the geodesic distances in this network is 2.545. On average, nodes in this component are 2.5 steps away from each other.

Answers to the Brokerage/ Betweenness section questions:
Below is a table of each node's betweenness score.

Node	Betweenness
John	0.0
Joshua	0.0
William	0.0
Julie	24.0
Scott	0.0
Charlie	0.0
Daniel	4.5
Harry	0.0
Kevin	41.0
James	4.5
Alfie	0.0
Michael	28.0

(1) Kevin is the most important broker in this network. His betweenness score is 41. In other words, Kevin resides of 41 of the 66 geodesic paths in this network or 62% of the total geodesic paths.

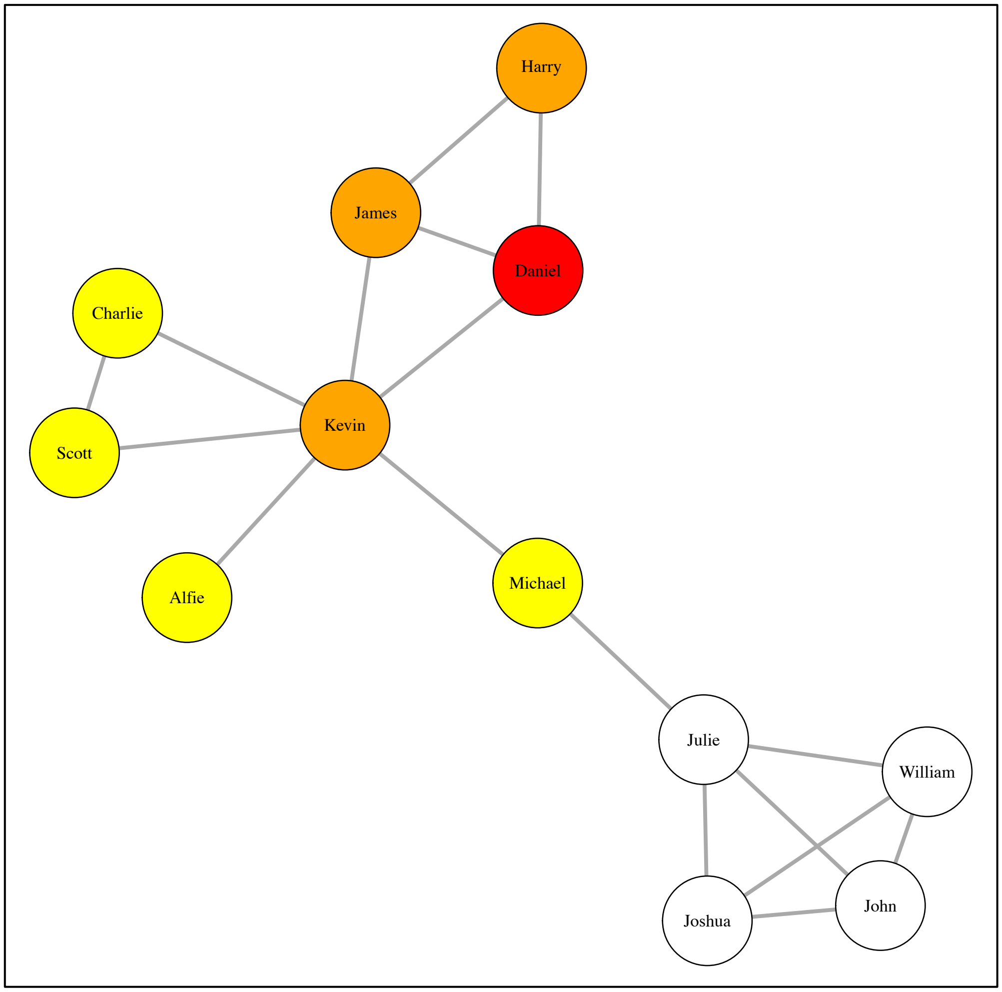
(2) 7 of the 12 nodes fail to broker any ties in this network and, thus, have a betweenness score of 0: John, Joshua, William, Scott, Charlie, Harry, and Alfie.

(3) Kevin has a higher betweenness score than Julie. Kevin resides between 41 geodesic distances and Julie resides on 24 geodesic distances. Both Kevin and Julie are important to brokering the far left and far right sections of the network to each other—all of the distances connected those on the far left and those on the far right go through Kevin, Michael, and Julie. However, also Kevin resides on geodesic distances between the nodes in the upper left section of the network. Julie's lower right section of the network is denser and all the nodes are already connected to each other without her brokering those paths. For example, Kevin brokers the distance for Charlie and Alfie, but Julie does not broker the distance between Joshua and William.

(4) Daniel and James have the same betweenness score of 4.5. They are tied for brokering all of Harry's geodesic distances to the rest of the network. Not counting Harry's ties to Daniel and James, there are 9 other nodes in this component that have geodesic distances to Harry. All 9 of these distances must pass through either Daniel or James since they are tied for providing Harry with the shortest paths. The betweenness score is split in 2 for the tie. Daniel resides on half of Harry's geodesic distances and James resides on the other half of Harry's geodesic distances. The formula is: 9 geodesic distances /shared by 2 brokers = a betweenness scores of 4.5.

Answers to the Neighborhood section questions:

(1) In the image below, Daniel, the ego, is red. Nodes that are added to Daniel in the 1st order neighborhood are orange, and these are all of the nodes directly connected to Daniel. Nodes that are added to the 1st order neighborhood when we move to the 2nd order neighborhood are yellow, and these are all of the nodes 2 handshakes away from Daniel. Neighborhoods of the 3rd order and greater are white.



(2) There are 4 nodes in Daniel's 1st order neighborhood: Daniel as the ego plus Harry, James, and Kevin, who are all directly connected to Daniel.

(3) It would a 4th order neighborhood to reach all nodes in the component. Julie would be the only node added in a 3rd order neighborhood, and Joshua, William, and John would be added in the 4th order neighborhood.

(4) Tobler's Law states that everything is related to everything else, but near things are more related than distant things. In social networks, closer neighbors are more likely to influence and have more influence than farther neighbors. Michael is a farther neighbor than Kevin, so Kevin is more likely to influence Daniel than Kevin.

That concludes this lab portion of Module 4: Analytics. For participants in the RStudio track of this training, proceed to the Module 4: Analytics RStudio to use software to produces some of these same calculations done here by hand.

The last and final module of this Social Network Analysis for Criminal Justice Practitioners and Analysts training series is Module 5: Direction. Module 5: Direction revisits previous modules on data, visualization, and analytics with necessary revisions for the analysis of social networks with directed ties.