The Basics of Network Structure

James M. Cook
University of Maine at Augusta
james.m.cook@maine.edu

Keywords

Betweenness; Centrality; Closeness; Degree; Density; Diameter; Distance; Clique; Connected Component; Cut Point; Geodesic; Measurement; Path; Social Networks; Size; Social Structure; Undergraduate;

1. RECAP ON SOCIAL NETWORKS

Our initial readings have introduced the following points:

- Individual-level data in the social sciences measures attributes of individuals.
- In contrast, a social network is a *structure of ties between nodes*.
- A *node* is an entity that can form relations with other entities, and a *tie* is an existing relation between nodes.
- In standard form, every social network has one kind of entity as a node and one kind of relation forming ties. These form the network's *boundary*.
- The tendency of the same nodes to be involved in multiple different networks is called *multiplexity*.
- The relation defining a network may or may not involve direction.
- Networks can be drawn as pictures called *graphs*, as tables called *matrices*, as *edge lists* and as *adjacency lists.*
- The *structure* of a network is the pattern of nodes and ties in that network.

In this chapter, we delve into some important detail regarding that last point. If social network *structure* consists of patterns among nodes and ties, what kind of patterns are possible, and how can they be detected?

2. STRUCTURAL PROPERTIES OF TIES

2.1 Tie Strength

Up to this point in our learning about social networks, we have only considered whether a tie between two nodes exists or not. Such ties are called *dichotomous ties*, to represent the dichotomy between existence (1) and non-existence (0). But as Mark Granovetter pointed out a generation ago (Granovetter 1973), some kinds of network relations describe connections between nodes that involve notions of greater or lesser, more or less. *Tie strength* is the numerically-expressed amount or intensity of the relationship in a tie. It is a number that describes the extent of a tie, not just its presence.

For example, we might want to describe a relation "plays tennis with." As a *dichotomous* relation, we would only be able to indicate whether two people play tennis with one another at all. But if we measure *tie strength*, we can describe *how often* two people play tennis together. In another example, we could describe the relation "exports goods to" to describe trade between nations. Expressed dichotomously, "exports goods to" would only indicate whether Nation A exports any goods at all to Nation B. But if we measure *tie strength*, we can describe *how much* Nation A exports to Nation B.

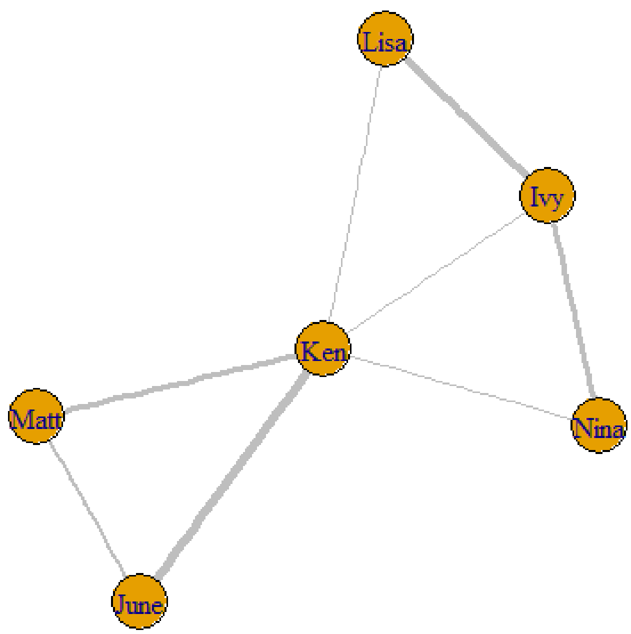
When we describe tie strength, we need to make sure to describe the units of strength so that the number associated with tie strength has some substantive meaning. For instance, we could measure a tennis relationship as "*number* of times per month two people play tennis together." We could measure export strength as "the annual dollar value of goods exported from one nation to another." The more specific the description of tie strength, the better.

In an adjacency matrix, the strength of a tie is expressed by placing a numeric value in the appropriate cell describing the relationship between nodes. The following adjacency matrix, for example, describes how often four people played tennis with one another in the month of June 2006:

	Ivy	June	Ken	Lisa	Matt	Nina
Ivy		0	1	4	0	3
June	0		5	0	2	0
Ken	1	5		1	3	1
Lisa	4	0	1		0	0
Matt	0	2	3	0		0
Nina	3	0	1	0	0	

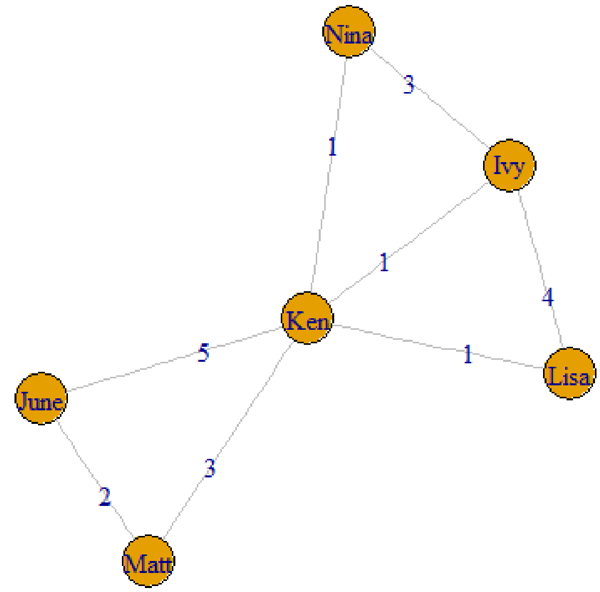
Note that "playing with" is a relation without direction, so this is an undirected network.

In a network graph, tie strength can be expressed by varying the width of a tie, with wider ties indicating stronger ties. The graph below corresponds to the adjacency matrix above, with tie strength indicated through tie width:



Tie strength expressed through tie width

This form of expressing tie strength is acceptable when there are only a few gradations of tie strength, More commonly, in order to be more exact, tie strength is indicated by placing a relevant number next to each tie. The following graph communicates the same information, but with tie strength indicated through number rather than through tie width:



Tie strength expressed through numbers

Tie strength can be expressed in an edge list by adding a third column to the original two, with the third column expressing a tie's strength. Here is such an edge list for the tennis network:

Ivy, Ken, 1

Ivy, Lisa, 4

Ivy, Nina, 3

June, Ken, 5

June, Matt, 2

Ken, Lisa, 1

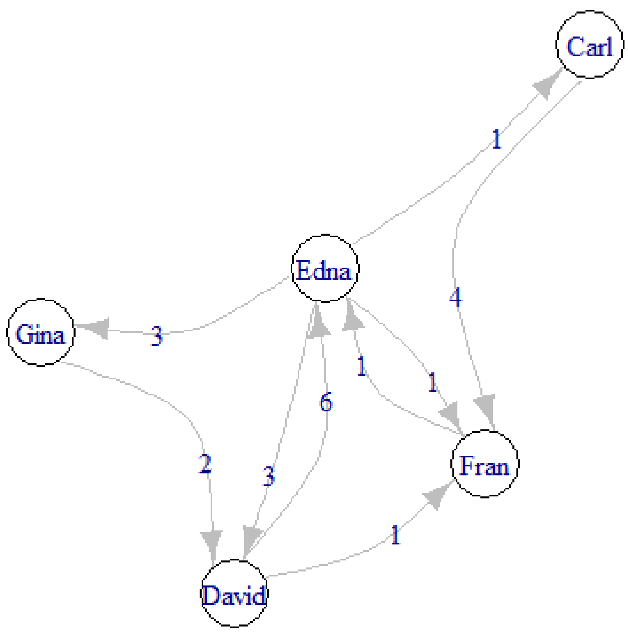
Ken, Matt, 3

Ken, Nina, 1

Tie strength can also be expressed for relations with direction. Consider the following directed network as an adjacency matrix, in which the relation is "number of letters written by ___ and sent to ___":

	Carl	David	Edna	Fran	Gina
Carl	0	0	0	4	0
David	0	0	6	1	0
Edna	1	3	0	1	3
Fran	0	0	1	0	0
Gina	0	2	0	0	0

The adjacency matrix can be expressed as a network graph with tie strength information, but when ties exist in *both* directions, we can no longer depict the two ties as a single double-headed line because each of the two ties may have its own distinct strength. The solution is to draw each directional tie as its own, adding a bit of a curve to ties so they can be distinguished from one another. This network graph with curved ties corresponds to the adjacency matrix for the network of letter-sending:



Tie strength expressed through numbers in a digraph

Edge lists for directed networks with tie strength are as straight-forward as ever. Just remember that in a directed network, the sender of a tie goes first in each row of the list. The edge list for the directed network of letter-sending looks like this:

Carl, Fran, 4

David, Edna, 6

David, Fran, 1

Edna, Carl, 1

Edna, David, 3

Edna, Fran, 1

Edna, Gina, 3

Fran, Edna, 1

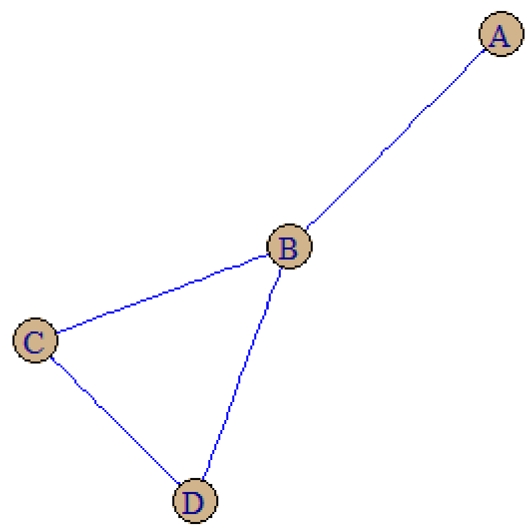
Gina, David, 2

You may have a question on the tip of your tongue: what about using *adjacency lists* to express tie strength? Unfortunately, adjacency lists simply aren't capable of communicating tie strength, at least not easily. The space-saving feature of adjacency lists – shoving all ties with one node into one line together – means that there isn't room to add tie strength in an easily-read way.

2.2 Paths, Geodesics and Distance

In a social network, any two nodes may or may not be connected by a tie. But networks are more than just pairs of nodes. Across multiple nodes, ties form *paths*. A *path* is a set of ties that connect nodes in a sequence. Just as a path through the woods may be described according to the order of the places where one walks, so network paths are described in the order of ties that take one from an initial origin node to a final destination node. The *length* of a path is simply the number of ties that the path contains.

Consider the following undirected network graph:



Simple undirected network

In this network, here are some of the paths we could identify:

A-B (length 1)

A-B-C (length 2)

A-B-C-D (length 3)

A-B-D (length 2)

A-B-D-C (length 3)

B-C (length 1)

B-D (length 1)

B-C-D (length 2)

C-D (length 2)

B-C-D-B-A (length 4)

Some of these paths describe a longer and more circuitous route than is necessary to get from one node to another. Instead of traveling from B to C to D to B to A, for instance (length 4), we could simply take a path from B to A. A special set of paths, called *geodesics*, are the paths with the shortest possible length between two nodes. Finally, the *distance* between two nodes is the length of the geodesic path from one of the nodes to the other.

In the network described above, A-B-D is a geodesic of length 2, but A-B-C-D is not a geodesic because with a length of 3 it is not the shortest path between A and D. The distance between A and D is 2 because that is the length of the shortest path between A and D.

In undirected networks, the distance between Node A and Node B is the same as the distance between Node B and Node A because it is possible to travel in any order along a path in an undirected network. This is not necessarily the case in networks with direction. In networks with direction, paths must follow the direction of ties.

Consider the following network, described as an edge list (try constructing it as a graph or a matrix):

A->B

B->C

C->A

The shortest path from C to A is simply C->A, meaning that the distance from C to A is 1. However, the shortest path from A to C is A->B->C, meaning that the distance from A to C is 2.

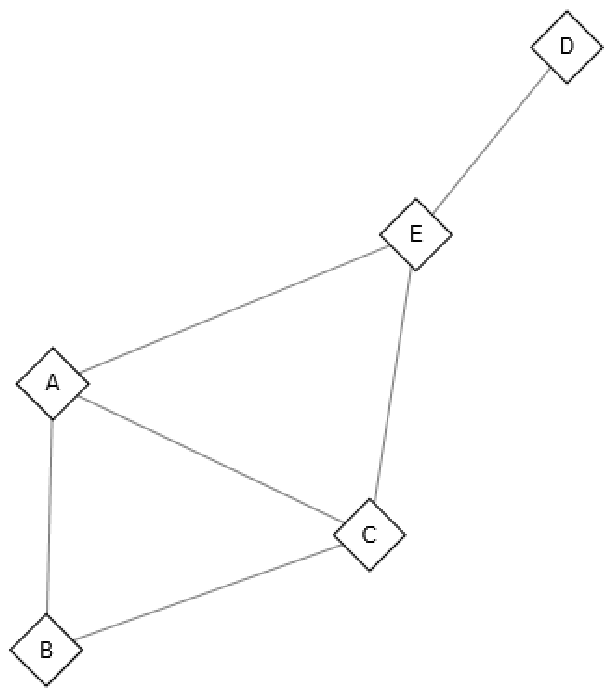
3. STRUCTURAL PROPERTIES OF NODES

The *structural* properties of nodes described below are not characteristics of nodes standing out there by themselves as individuals. Rather, they describe the context of the network surrounding a node. The ideas described in this section were introduced to sociologists by Linton C. Freeman (1979) a generation ago.

3.1 Degree, Indegree and Outdegree

Simply knowing how well some nodes are connected to others can reveal a great deal. In undirected networks (that is, networks in which the relation of interest implies no direction to ties), the *degree* of a node is defined as the number of ties involving that node. In directed networks, the *indegree* of a node is defined as the number of relations directed **toward** that node, while the *outdegree* of a node is defined as the number of relations that node directs **to others**. In friendship networks, indegree is a measure of popularity, while outdegree is a measure of gregariousness.

Let's take a look at a pair of networks for some examples of how degree, indegree and outdegree are measured. The network graph below represents an undirected network, as you can tell by the lack of arrowheads on ties. In this network, nodes A, C and E each have a degree of 3. Node B has a degree of 2, and node D has a degree of 1. It makes no sense to speak of indegree or outdegree for the network below, since there is no direction to any of the ties.

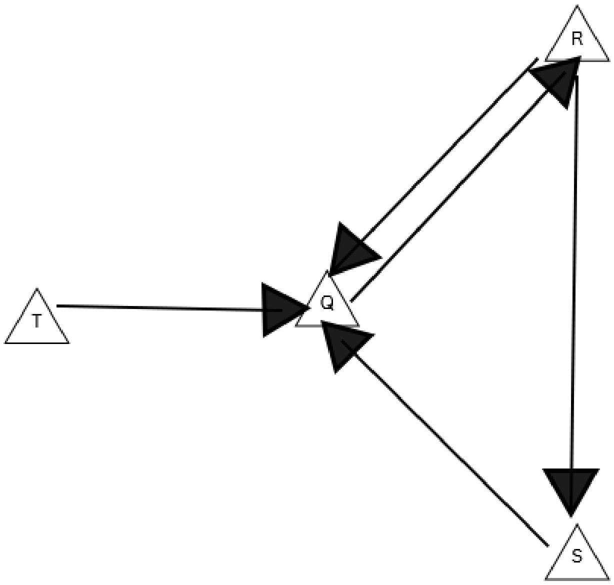


Undirected Network with Node Degrees Ranging from 1 to 3

In contrast, in the network graph representing a directed network on the next page, it makes no sense to discuss the degree of nodes. Instead, we should measure each node's indegree and outdegree.

The indegree and outdegree of each node are:

Node	Indegree	Outdegree
R	1	2
S	1	1
Q	3	1
T	0	1

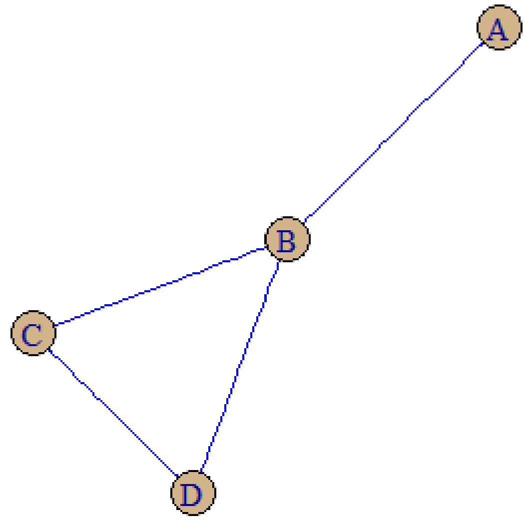


A directed network graph ("digraph") in which measurement of indegree and outdegree is most appropriate

3.2 Closeness Centrality

An important idea regarding social networks is that some nodes are in the "center" of the network (in the midst of the important communication going on), while other nodes are on the "edge" or "periphery" (and therefore not key players in the network). But what is really meant by the "center" of a network? Who is "central"? It turns out (Freeman 1979) that depending on how you think of and therefore measure centrality, different nodes may be thought of as central or peripheral.

The idea of degree, as discussed in the previous section of this chapter, is that some nodes are possibly more important because they have more ties associated with them. But that is not the only important idea of importance in a network. Consider *closeness*: perhaps a node is more important in a network if its distance to other nodes is short. If a node is not too far away from other nodes, its ability to influence others may be strong.



A simple undirected network again

To calculate *closeness centrality* for a node, we first calculate the *farness* of that node. The *farness* of a node is simply the sum of all the distances from that node to all other nodes in the network. For the four-node network you see above as a graph, the farness of each node is:

Node	Distances to other Nodes	Farness
A	1,2,2	5
B	1,1,1	3
C	2,1,1	4
D	2,1,1	4

The less far a node is from other nodes, the closer it is. To make that idea work, closeness is measured as (1/farness):

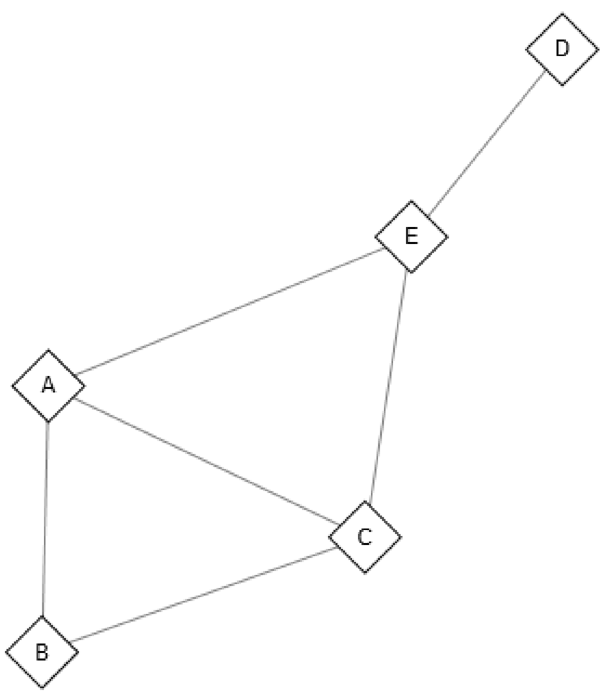
Node	Farness	Closeness
A	5	1/5, or 0.2
B	3	1/3, or 0.333
C	4	1/4, or 0.25
D	4	1/4, or 0.25

In principle, calculating closeness centrality is very easy, involving one step of simple addition and another step of simple division. In practice, the measurement of closeness centrality gets more difficult as the number of nodes in a network increases.

3.3 Betweenness Centrality

According to the idea of degree, nodes are important in a network if they directly interact with many other nodes. According to the idea of closeness centrality, nodes are important in a network if they tend not to be far from other nodes in terms of network distance.

Betweenness centrality thinks of importance in a network differently. As a measure of centrality, betweenness is based on how often a node tends to be in the middle of the shortest paths between other nodes. Betweenness centrality for a node we're interested in (let's call it "*Node X")* is calculated in three steps. In the first step, all geodesic paths between all nodes in a network are identified, including ties (paths between nodes that are equally short). In the second step, for each pair of nodes in the network, the fraction of geodesics in which *Node X* is in the middle is noted. To be in the middle of the geodesic means a node isn't at either end. In the third step, all these fractions are added up.



an undirected network, reprised

Let's calculate betweenness centrality for three nodes in the network you see above. Let's start with node D. Step 1 in calculating betweenness centrality of node D is to find all geodesic paths between all pairs of nodes in the network. Let's list them:

Pair of nodes	Geodesic(s)
A,B	A-B
A,C	A-C
A,D	A-E-D
A,E	A-E
B,C	B-C
B,D	B-A-E-D, B-C-E-D
B,E	B-A-E, B-C-E
C,D	C-E-D
C,E	C-E
D,E	D-E

Notice that for two pairs of nodes (B and D and B and E), there are two geodesics that are tied for the shortest path.

Step 2 in calculating betweenness centrality for node D is, for each pair, to find the fraction of geodesics in which D is in the middle. For D, that's easy: the fraction is always 0, because D is *never* in the middle of any geodesics.

Pair of nodes	Geodesic(s)	Fraction (D in middle)
A,B	A-B	0
A,C	A-C	0
A,D	A-E-D	0
A,E	A-E	0
B,C	B-C	0
B,D	B-A-E-D, B-C-E-D	0
B,E	B-A-E, B-C-E	0
C,D	C-E-D	0
C,E	C-E	0
D,E	D-E	0

The third step for finding the betweenness centrality of node D is easy: simply add up all those zeroes. The betweenness centrality of Node D is 0, which makes sense if you just look at the network graph: D is out on the edge, never in between other nodes.

Let's calculate betweenness centrality again, this time for node A. Fortunately, we've already found all the geodesics in the network already, so we can skip ahead to steps 2 and 3:

Pair of nodes	Geodesic(s)	Fraction (A in middle)
A,B	A-B	0
A,C	A-C	0
A,D	A-E-D	0
A,E	A-E	0
B,C	B-C	0
B,D	B-A-E-D, B-C-E-D	1/2
B,E	B-A-E, B-C-E	1/2
C,D	C-E-D	0
C,E	C-E	0
D,E	D-E	0
	Betweenness centrality:	1

Node A is in the middle of two geodesics in the network: B-A-E-D for node pair B and D, and B-A-E for node pair B and E. But each geodesic is tied with another geodesic that doesn't have A in the middle, so for each of the two node pairs Node A earns a fraction of just 1/2. 1/2 + 1/2 =1, and so the betweenness centrality of Node A is 1.

Finally, let's calculate the betweenness centrality of Node E:

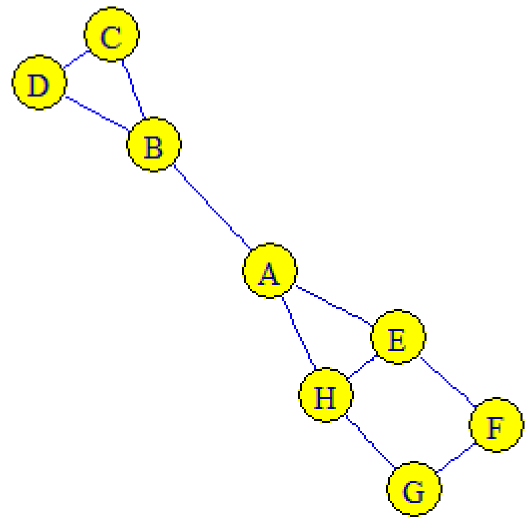
Pair of nodes	Geodesic(s)	Fraction (E in middle)
A,B	A-B	0
A,C	A-C	0
A,D	A-E-D	1
A,E	A-E	0
B,C	B-C	0
B,D	B-A-E-D, B-C-E-D	1
B,E	B-A-E, B-C-E	0
C,D	C-E-D	1
C,E	C-E	0
D,E	D-E	0
	Betweenness centrality:	3

Can you follow the calculation for Node E by yourself this time? If not, please get in touch and I'll help you through it.

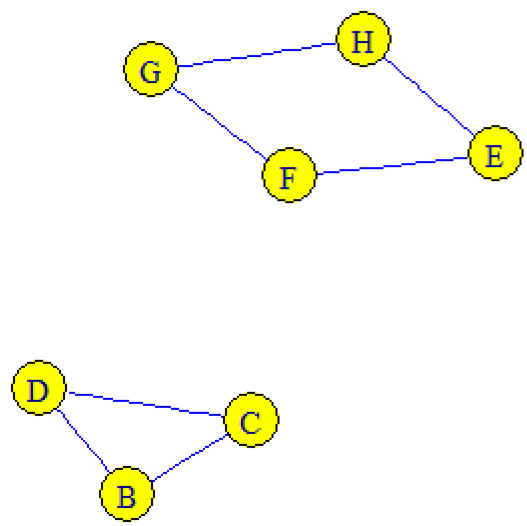
3.4 Cut Points

Sometimes, the presence of one node is all that keeps a network from falling apart. That node is called a *cut point*. A *connected component* of a social network is a portion of a network in which all nodes have some *path*, no matter how long, to all other nodes. A cut point is a node that, when removed, separates a single connected component into two or more connected components that no longer have a path to one another.

Let's take a look at another network:

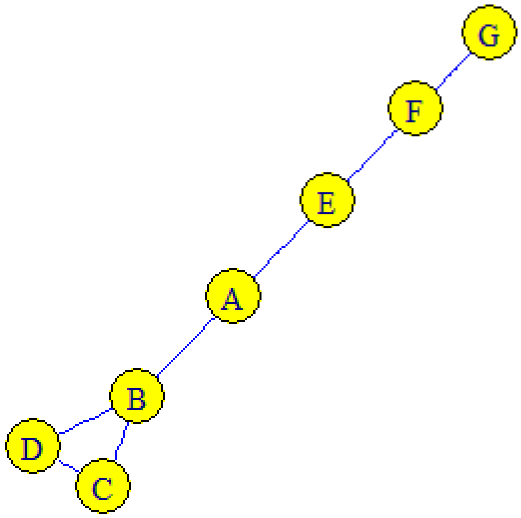


All the nodes in this network form one *connected component* because for every node there is some path connecting it to all other nodes. But what happens if we start removing some nodes from the network? Let's try removing Node A:



Node A certainly *is* a cut point, because without it, the network is broken into two pieces, two connected components that are utterly separated from one another.

Let's put A back in where it was and try taking out another node. How about Node H?



Without H, that network certainly becomes more strung out. The closeness centrality values for most of the nodes should increase. But regardless of that, the remaining network is still a single connected component. For that reason, Node H is *not* a cut point.

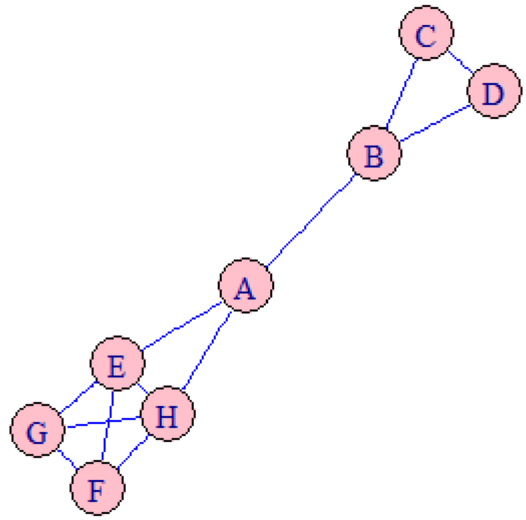
Adjacency lists can be a bit harder to read than edge lists, but they will take up many fewer rows, especially in larger networks.

4. STRUCTURE IN PARTS OF NETWORKS

Connected components are not just important because they help us find cut points. They are interesting on their own because they are one kind of pattern within a part of a social network.

There are other pieces of structure that you can find in a social network if you look carefully. Finding these without the help of a computer can be a bit like playing "Where's Waldo?"

One more kind of structure in a part of a network is the ***clique*,** a set of nodes that are all tied to one another.



In the above network, there are many cliques. The set {B,C,D} is a clique because every node in the set is tied to every other node in the set. Other cliques include {A,E,H}, {E,G,H}, {E,G,F}, {G,H,F}, {E,F,H}, and {E,F,G,H}.

5. STRUCTURAL PROPERTIES OF NETWORKS

So far in this chapter, we have defined structural properties of ties, of nodes and of parts of networks. But it is also possible to describe structure properties of an entire network. *Size* is perhaps the simplest of these. A network's size is simply the number of nodes in it. Network size is usually referred to by the letter *n*.

Network *density* is a bit more complicated, but not too complicated to master. The idea of density is simple: it expresses how full of ties a network is. A system's density is defined as the share of possible ties in a network that actually *do* exist. Expressed mathematically, it is a simple fraction:

Density = (# actual ties) / (# possible ties)

The lowest number of actual ties that can occur in a network is 0, and 0 divided by any number remains 0. Therefore, the lowest possible density is 0. The largest number of actual ties that can occur is the same as the number of possible ties, and any number divided by itself equals 1. Therefore, the highest possible density is 1. Multiply density by 100 and you get an easily-interpreted number: the percentage of possible ties that are actually there in a network.

With small social networks, such as those you find in the pages of this chapter, the number of actual ties is determined by simply counting all existing lines. It is also possible (although difficult) to determine the number of possible ties by careful eyework. However, most social systems contain nodes by the hundreds, thousands or millions, which makes determining the number of possible ties an onerous and error-prone task. Fortunately, two simple formulas exist to determine the number of possible ties in a social network. If n equals the number of nodes in a network, then:

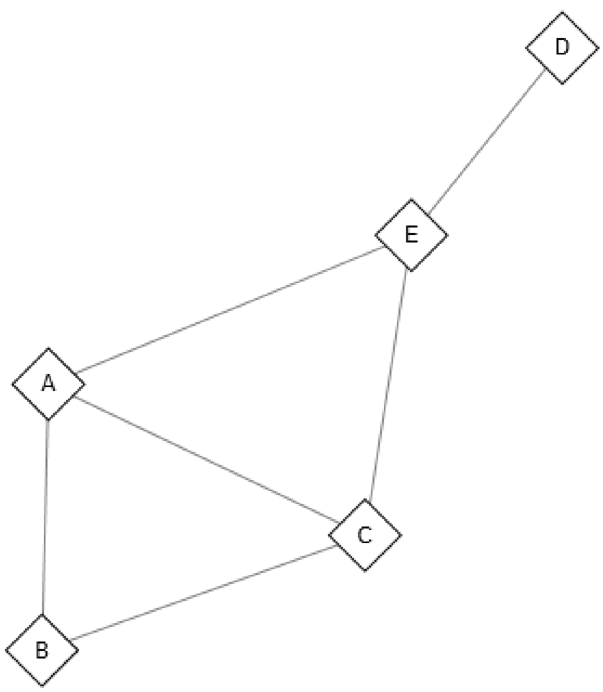
For an undirected network:

possible ties = (n*(n-1))/2.

For a directed network:

possible ties = n*(n-1)

Let's work through an example using this network again:



There are five nodes in this network. In other words, the network's size is 5; n=5. The number of actual ties in the network is 6; those are easy to count.

How many possible ties are there? Even for a network of only 5 nodes, it's a bit tricky to find them all. But we know that for an undirected network like this, the number of possible ties is (n*(n-1))/2. N=5, so that means the number of possible ties is (5*(5-1))/2, or (5*4)/2, or 20/2, or 10.

Don't stop there! Remember that density is a fraction:

Density = (# actual ties) / (# possible ties)

If the number of possible ties is 10, and the number of actual ties is 6, then the density of this network is 6/10. If we wanted, we could express this density as a decimal, 0.6, or as a simplified fraction, 3/5. The network is 3/5 full of ties. It is 60% full of ties.

As you work through examples of network measurements, you may begin to feel that making those measurements on your own starts to get a little bit complicated. For bigger networks, making such measurements can get to be a big headache. This is why most social network researchers use computer programs to analyze with social networks. Why not let a computer do most of the heavy lifting, then bask in the glow of accomplishment yourself?

6. GLOSSARY

betweenness: a measure of centrality based on how often a node tends to be in the middle of the shortest paths between other nodes. Betweenness centrality for *Node X* is calculated in three steps. In the first step, all geodesic paths between all nodes in a network are identified, including ties (paths between nodes that are equally short). In the second step, for each pair of nodes in the network, the fraction of geodesics in which Node X is in the middle is noted. In the third step, all these fractions are added up.

clique: a set of nodes that are all tied to one another.

closeness: a measure of centrality based on how close a node tends to be to other nodes. Calculated as (1/farness).

connected component: a portion of a network in which all nodes have some path, no matter how long, to all other nodes.

cut point: a node that, when removed, separates a single connected component into two or more connected components that no longer have a path to one another.

degree: a characteristic of a node in an undirected network, measured as the number of ties involving that node.

density: the share of possible ties in a network that actually *do* exist.

dichotomous tie: a tie that can either take the quality of existence (represented in a graph by a line and in an adjacency matrix by a 1) or non-existence (represented in a graph by the absence of a line and in an adjacency matrix by a 0).

distance: length of the geodesic path between two nodes.

farness: sum of distances from a node to all other nodes in a network.

geodesic: path of shortest possible length between two nodes.

in-degree: a characteristic of a node in an directed network, measured as the number of ties in which that node is the recipient (or target) of a tie.

length: number of ties contained in a path.

out-degree: a characteristic of a node in an directed network, measured as the number of ties in which that node is the sender (or originator) of a tie.

path: a set of ties that connect nodes in a sequence**.**

size: number of nodes in a network. Denoted by the letter *n*.

tie strength: the numerically-expressed amount or intensity of the relationship in a tie.

7. REFERENCES

Freeman, Linton C. 1979. “Centrality in Social Networks I: Conceptual Clarification.” *Social Networks* 1: 215-239.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78(6): 1360-1380.