# Individuals Versus Networks

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

## Keywords

Data; Individualism; List; Matrix; Social Networks; Undergraduate;

## 1. RECAP: WHAT IS A SOCIAL NETWORK?

Our initial reading to introduce social network analysis introduced the following points:

- Many kinds of sociology can be complicated. Social networks, by contrast, are simple, consisting of a *structure of ties between nodes*.
- A *node* is an entity that can form relations with other entities. Nodes are often people, but not always.
- A *tie* forms when a particular kind of relation forms between two particular nodes.
- The relation defining a network may or may not involve direction.
- To create a network, you must describe what kind of nodes belong in the network (its *boundary*), and what (usually) single kind of relation describes the ties in the network. The more specifically you can do this, the better.
- For any set of nodes, there may be multiple kinds of possible relations between them. Because a network (usually) only depicts one kind of relation, each different kind of relation is defined in its own different network.
- The tendency of the same nodes to be involved in multiple different networks is called *multiplexity*.
- The *structure* of a network is the pattern of nodes and ties in that network. We will learn more about network structure in future readings.
- Networks can be drawn as pictures called *graphs*.
- Networks can be depicted as tables called *matrices* (singular: matrix).
- There are other ways to depict networks, too. We will learn more about them later in this chapter.

These points describe what a social network is. The distinctiveness of social networks becomes clearer when we compare the social network approach in social science to the more traditional individual-based approach.

## 2. INDIVIDUALS

### 2.1 The Logic of Studying Individuals

The word *individual* is many hundreds of years old, coming from the Latin *individuus*, a word referring to the ancient idea of some fundamental unit that cannot be divided without losing its essence.

Many scientific disciplines have some notion of the individual. For thousands of years, the “atom” of atomic theory has represented the tiniest possible bit of an element that still maintains the properties of that element. An atom of gold has all of the essential properties of a bar of gold, but if a scientist were to split that atom apart, the gold would lose its properties and cease to be gold. In evolutionary biology, an individual organism is the fundamental unit of the species because it has all of the properties of its species, such as behaviors, genes, and reproductive ability (Michod et al. 2003). The organism is individual because (with a few interesting exceptions) if you divide it, it can no longer carry out those vital functions. In medicine, the survival of the human individual is of central interest (Gringeri et al. 2008). Cut apart the individual and the patient dies; this act is prohibited before death by the physician’s oath to “first do no harm.” To treat the individual patient, medical practitioners administer tests to measure all sorts of characteristics regarding the individual’s health.

Like the natural sciences, the social sciences have tended to focus upon properties of individual units. In psychology the individual unit of study is the “person,” traits of which are referred to as "personality" and can be studied by assessing each individual person (Cattell and Schuerger 2003).

The social sciences are so named because each recognizes the existence of some important activity beyond the level of the individual person. Psychology recognizes that each person is surrounded by other people. Economics describes the behavior of markets. Political science describes policymaking and elections. Sociology has a particular emphasis on group and community. Yet because the individual-focused approach so strongly dominates science, social scientists still tend to measure think of the objects they study as individuals. For political science, the individual unit may be an election in a district, with district number and voter turnout being properties of that election. If sociologists study individual communities, the U.S. Census Bureau helps them by collecting information about the properties of those individual communities.

In sum, although the disciplines of the natural and social sciences are quite diverse in their subject matter, each has an individualist strain in which an individual unit is studied for its individual properties. The table below illustrates this common approach.

| Discipline | Individual | Individual Properties | |
|---|---|---|---|
| Physics | Atom | Atomic Number | Atomic Mass |
| Evolutionary Biology | Member of Species | Genotype | Phenotype |
| Medicine | Patient | Temperature | Blood Pressure |
| Psychology | Person | Impulsivity | Extroversion |
| Political Science | District Election | District Number | Voter Turnout |
| Sociology | Community | Median Income | Population Density |

**Individuals in the Natural and Social Sciences**

## 2.2 Individuals Recorded in Data

If individuals are the essential unit to study, and if each individual has properties, then when individuals are being studied, it makes sense to organize one's observations according to the individual and their properties. In individual-oriented science, each individual being observed is referred to as a *case*, and multiple cases are usually observed to track *variables*, properties of the individual that vary from case to case. When multiple cases are assembled together in an organized fashion, with observed values of multiple variables for each case, the result is called *data*.

What do data look like in individual-centered research? Typically, data is preserved in a table, each case is assigned a horizontal row, and each variable is assigned a vertical column. Every combination of row and column in the data table is a space in which information about the value of a particular variable in a particular case may be entered.

### 2.2.1 *Example 1: Personality Inventory*

Let's look at one example of an individual-centered dataset. The following is a psychology dataset containing anonymous answers provided by twenty people to an online version of Cattell's 16 Personality Factors Test at the website [http://personality-testing.info/_rawdata/](http://personality-testing.info/_rawdata/) (Cattell & Schuerger 2003):

| *ID#* | *"I know how to comfort others"* | *"I enjoy bringing people together"* | *"I learn quickly"* | *"I have frequent mood swings"* |
|---|---|---|---|---|
| 1 | 4 | 4 | 4 | 2 |
| 2 | 5 | 4 | 2 | 1 |
| 3 | 4 | 4 | 3 | 2 |
| 4 | 2 | 3 | 4 | 3 |
| 5 | 4 | 5 | 2 | 4 |
| 6 | 4 | 5 | 5 | 4 |
| 7 | 3 | 4 | 5 | 3 |
| 8 | 4 | 5 | 5 | 1 |
| 9 | 4 | 4 | 4 | 4 |
| 10 | 4 | 4 | 4 | 4 |
| 11 | 5 | 5 | 5 | 1 |
| 12 | 4 | 4 | 3 | 2 |
| 13 | 1 | 3 | 4 | 3 |
| 14 | 4 | 4 | 4 | 3 |
| 15 | 3 | 3 | 4 | 3 |
| 16 | 5 | 4 | 3 | 3 |
| 17 | 1 | 3 | 5 | 3 |
| 18 | 4 | 4 | 4 | 2 |
| 19 | 2 | 3 | 4 | 2 |
| 20 | 5 | 4 | 5 | 1 |

**Data: Personality Factor Test Results**

In this table, the first row contains column headers that describe each of five variables:

*Variable 1*: ID#, the anonymized identification number for a person used instead of that person's name.

*Variable 2*: "I know how to comfort others," a statement to which the person completing the survey responded by circling one of the following numbers:

1. strongly disagree
2. disagree
3. neither agree nor disagree
4. agree
5. strongly agree

*Variable 3*: "I enjoy bringing people together," a statement with the same numbered options as in Variable 2.

*Variable 4*: "I learn quickly," a statement with the same numbered options as in Variable 2.

*Variable 5*: "I have frequent mood swings," a statement with the same numbered options as in Variable 2.

We could interpret the results for each person studied by reading across each row. For instance, we know that person #1 "agrees" that "I know how to comfort others," "agrees" that "I enjoy bringing people together," "agrees" that "I learn quickly," but "disagrees" that "I have frequent mood swings." Person #17, in contrast, "strongly disagrees" that "I know how to comfort others," "neither agrees nor disagrees" that "I enjoy bringing people together," "strongly agrees" that "I learn quickly," and "neither agrees nor disagrees" that "I have frequent mood swings."

We could also study and compare the tendencies for people in general in this individual-level dataset. 70% (14 out of 20) of the respondents agree or strongly agree that they know how to comfort others. In contrast, only 20% (4 out of 20) of the respondents agree or strongly agree that they have frequent mood swings. Finally, it appears that orientations toward comforting and bringing people together are related: every single one of the people who agree or strongly agree that they "know how to comfort others" also agrees or strongly agrees that they "enjoy bringing people together."

These results are interesting, but they have limits. They do not report the extent to which these individual people *actually* comfort others or bring people together; they only report what people *say* they know or enjoy about such social activities. The data do not report who in particular these people report, or who they bring together. This information is missing at the individual level.

### 2.2.2 Example 2: Election Results

For another example of an individual-level data set, let's turn to political science. The following is a dataset of 2008 U.S. presidential election results in individual Alaska state legislative districts, obtained from https://catalog.data.gov/dataset/2008-presidential-general-election-state-results-direct-download (National Atlas of the United States 2009):

| *State* | *District* | *Area* | *Vote Dem.* | *Vote Rep.* | *Vote Other* |
|---|---|---|---|---|---|
| AK | 8 | 7.010 | 4995 | 4983 | 342 |
| AK | 37 | 16.637 | 1868 | 2661 | 136 |
| AK | 12 | 12.091 | 1914 | 5467 | 208 |
| AK | 13 | 0.044 | 2800 | 8432 | 294 |
| AK | 14 | 0.026 | 2132 | 8108 | 216 |
| AK | 16 | 0.480 | 2636 | 7774 | 287 |
| AK | 15 | 5.947 | 2510 | 8227 | 349 |
| AK | 39 | 16.297 | 2695 | 2323 | 166 |
| AK | 40 | 73.347 | 2137 | 2686 | 177 |
| AK | 2 | 1.931 | 3468 | 4029 | 238 |
| AK | 3 | 0.838 | 5657 | 2829 | 281 |
| AK | 4 | 0.265 | 4161 | 4302 | 273 |
| AK | 6 | 104.121 | 2351 | 4234 | 239 |
| AK | 17 | 0.003 | 2645 | 6621 | 182 |
| AK | 31 | 0.004 | 3596 | 6419 | 205 |
| AK | 25 | 0.002 | 3233 | 3042 | 168 |
| AK | 18 | 0.034 | 2046 | 4252 | 113 |
| AK | 29 | 0.002 | 2684 | 4127 | 151 |
| AK | 30 | 0.006 | 3486 | 5500 | 197 |
| AK | 24 | 0.002 | 3380 | 4127 | 181 |

**Data: 2008 Presidential Vote by State House District**

In this table, the first row contains column headers that describe each of six variables:

*State*: Always "AK" for these rows, because the districts are all in Alaska.

*District:* the Alaska State House District from which vote totals were obtained.

*Area:* using units of decimal degrees, the size of the area of the district. For reference, Alaska House District 6's area of 104.121 decimal degrees is nearly 211,000 square miles. Alaska House District 24's area of 0.002 decimal degrees is less than 6 square miles.

*Vote Dem.:* the number of votes for the Democratic presidential candidate (Barack Obama).

*Vote Rep.:* the number of votes for the Republican presidential candidate (John McCain).

*Vote Other:* the number of votes for another presidential candidate.

There are a number of conclusions we can draw from this data. Among these districts, the Republican candidate tended to do well (the Democrat won only 4 of the 20 districts). There is a huge range in the amount of land area covered by these districts (reflecting the very low population in some areas of Alaska). If you add up all the vote counts, you'll find out that the bigger the land area of a district, the fewer the number of people who turned out to cast a vote.

Such patterns are quite interesting. Still, there are some kinds of patterns that can't be expressed in individual-oriented data. For instance, could it be that districts that trade economically with another district would be more likely to vote for the same candidate? Are districts that share a boundary more likely to have the same voting patterns? These are ideas about relations between pairs of districts, not ideas about individual districts. These ideas just won't fit in a column of individually-oriented data. As long as we stick to individual-level data, we won't be able to test these ideas. We'll be stuck in a world in which only individuals seem to matter because it's only individuals and their individual properties that we manage to measure.

# 3. NETWORKS

## 3.1 The Logic of Studying Networks

The failure of individual-driven social science to account for relationships highlights the potential for social network analysis to provide an alternative. Despite criticism that social network analysis is "only a method" (Howison et al 2011), it actually is more than that, more than a simple set of calculation techniques. Social networks are a way of seeing the world differently, and therefore of asking and answering different questions about why the world is as it is. This makes the social network perspective an important new theory for social scientists (Wellman 1997).

But if we are going to seriously study the patterns of relationships described in social networks, we'll have to learn how to construct them. Network data looks very different from individual data, and it from in that difference that new possibilities in studying the social world are born.

## 3.2 Networks Recorded in Data

There are four common ways of representing a social network. Two of these, the matrix and graph, we have already reviewed in a previous chapter; we'll review them briefly here. To them we will add two new ways of representing social networks: the *edge list* and the *adjacency list*.

### *3.2.1 Matrices*

The *adjacency matrix* is the way of representing a social network that looks most like an individual data set. Like individual data, it is presented in a table. Unlike individual data, the network data in an adjacency matrix includes a column *and* a row for each node. Every cell in an adjacency matrix represents a combination of two nodes: a row node and a column node.

Let's take a look at a sample adjacency matrix to see what network data looks like:

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | | 1 | 1 | 0 | 1 |
| B | 1 | | 1 | 0 | 0 |
| C | 1 | 1 | | 0 | 1 |
| D | 0 | 0 | 0 | | 1 |
| E | 1 | 0 | 1 | 1 | |

**An adjacency matrix without direction**

In this adjacency matrix, a value of "1" for a pair of nodes means that a tie exists between the nodes. A value of "0" for that pair of nodes means that there is no tie between the nodes. Using this information, we can tell that there is a tie between node A and node B, and also that there is *not* a tie between node A and node D.

This adjacency matrix represents a relation *without* direction. We can tell this is true because in the matrix, there are two possible combinations for every node, and whenever a tie exists for one of those combinations, it exists for the other combination as well. For example, a tie for node combination [Row B, Column C] is present, and it is also present for node combination [Row C, Column B]. Conversely, there is *no* tie for node combination [Row D, Column C], and there is also no tie for node combination [Row C, Column D].

The adjacency matrix below represents a relation *with* direction, which you can tell by the way that a tie exists from node R to node S [Row R, Column S = 1] but not from node S to node R [Row S, Column R = 0].

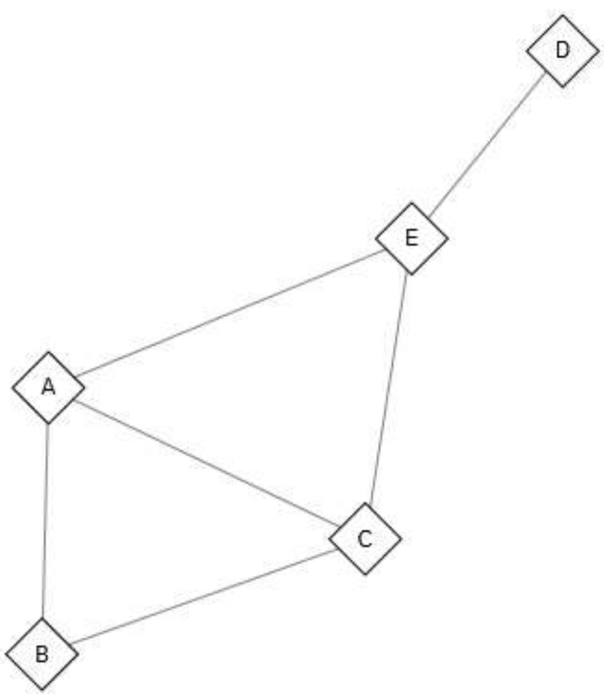
| | Q | R | S | T |
|---|---|---|---|---|
| Q | | 1 | 0 | 0 |
| R | 1 | | 1 | 0 |
| S | 1 | 0 | 0 | 0 |
| T | 1 | 0 | 0 | |

**An adjacency matrix with direction**

### *3.2.2 Graphs*

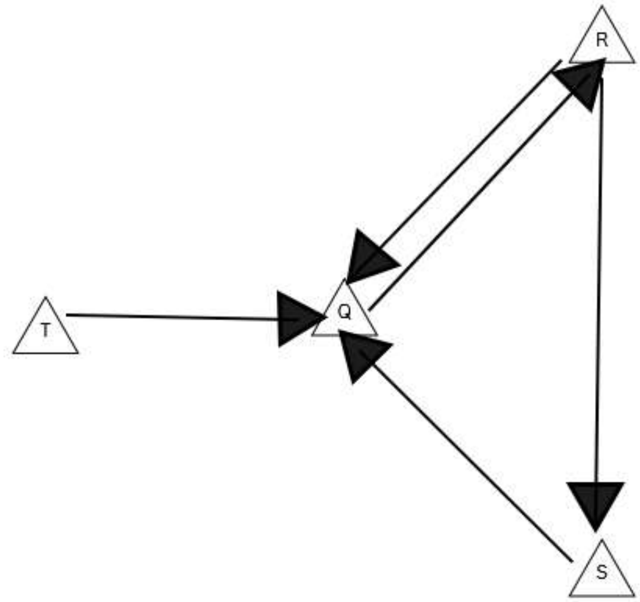
It is possible to describe every adjacency matrix as a graph. A graph is simply a drawing in which each tie is a line and each node is a shape. Where the underlying relation for a network indicates direction, ties should have arrowheads to point in the proper direction from one node to another.

The following network graph corresponds to the {A,B,C,D,E} adjacency matrix above:




**A network graph without direction**

The following network graph corresponds to the {Q,R,S,T} adjacency matrix of section 3.2.1:




**A network graph with direction ("digraph")**

### *3.2.3 Edge Lists*

Because social network analysis is still relatively new, the terms being used by social network analysts are still sometimes unusual. "Edge" is simply the word used by social network researchers for "tie," and an edge list is simply a list of ties.

In the list below, you'll see that pairs of nodes are placed next to one another, with one pair per row. The existence of a row with a pair of nodes in it indicates that those two nodes are tied to one another in the network. If a pair of nodes is not tied, the pair is not included in the edge list.

A,B
A,C
A,E
B,C
C,E
D,E

This edge list corresponds to the {A,B,C,D,E} network we've been describing above. Because the {A,B,C,D,E} network does not include direction, it is not necessary to include "B,A" in the edge list because "B,A" is already included and in a non-directed graph the two are the same.

In a "digraph" (directed network), the first node mentioned is the "sender" of the tie, and the second node mentioned is the "recipient" of the tie. Here is an edge list for the {Q,R,S,T} network we've described above:

Q,R

R,Q

R,S

S,Q

T,Q

Because this edge list features direction, order matters.

### *3.2.4 Adjacency Lists*

Finally, a network may be represented as an adjacency list. This format is less common than the other three, but it does pop up from time to time.

An adjacency list is a set of rows of text, just like an edge list. In a twist, the first node listed in each row is the node from which ties emanate, and all the other nodes listed in that row are tied to that first node.

Here is an adjacency list for the {A,B,C,D,E} network we've been working with above:

A,B,C,E

B,C

C,E

D,E

And here is an adjacency list for the {Q,R,S,T} network:

Q,R

R,Q,S

S,Q

T,Q

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. GLOSSARY

**adjacency list:** representation of a social network taking the form of a set of rows of text. The first node listed in each row of an adjacency list is the node from which ties emanate, and all the other nodes listed in that row are tied to that first node.

**adjacency matrix:** table containing social network data in which each node is assigned both a column and a row, in which cells contains a zero where ties *do not exist*, and in which cells contain a number greater than zero where ties *do exist.*

**case**: a particular instance of a phenomenon being studied. In individualist research, a case is a collection of data regarding a single individual and that single individual's properties. In social network research, a case may be an observation of a particular tie and that tie's properties, or a case may involve an observation of a set of ties and properties of the structure in that set of ties.

**data:** collected and organized information about the values of variables observed for a set of cases.

**diagonal**: set of cells in a matrix that refer to a node's relationship with itself. Often (but not always), the diagonal is meaningless.

**digraph**: graph of a social network in which arrowheads are added to ties to show direction.

**edge:** a synonym for tie.

**edge list:** a form of representing a social network in which each row of text indicates one pair of nodes. The existence of a row with a pair of nodes in it indicates that those two nodes are tied to one another in the network. If a pair of nodes is not tied, the pair is not included in the edge list. In a "digraph" (directed network), the first node mentioned is the "sender" of the tie, and the second node mentioned is the "recipient" of the tie.

**graph**: drawing of a social network in which nodes are depicted as shapes and ties are depicted as lines.

**method**: a set of tools and techniques by which data can be entered and studied in order to answer research questions.

**theory**: an carefully organized story about the way that the world works that allows for predictions to be made about future behavior.

**variable**: property of a case being studied that can take on multiple values

## 5. REFERENCES

Cattell, Heather EP, and James M. Schuerger. 2003. *Essentials of 16PF Assessment*, volume 45. Hoboken, NJ: John Wiley & Sons.

Gringeri, Alessandro, and Sylvia von Mackensen. 2008. "Quality of Life in Haemophilia." *Haemophilia* 14(3): 19-25.

Howison, James, Andrea Wiggins, and Kevin Crowston. 2011. "Validity Issues in the use of Social Network Analysis with Digital Trace Data." *Journal of the Association for Information Systems* 12(12): 767-797.

Michod, Richard E., and Aurora M. Nedelcu. 2003. "On the Reorganization of Fitness During Evolutionary Transitions in Individuality." *Integrative and Comparative Biology* 43(1): 64-73.

National Atlas of the United States. 2009. "2008 Presidential General Election, State Results - Direct Download." https://catalog.data.gov/dataset/2008-presidential-general-election-state-results-direct-download

Wellman, Barry. 1997. "Structural Analysis: From Method and Metaphor to Theory and Substance." In *Social Structures: A Network Approach*. *Contemporary Studies in Sociology* volume 15: 19-61.