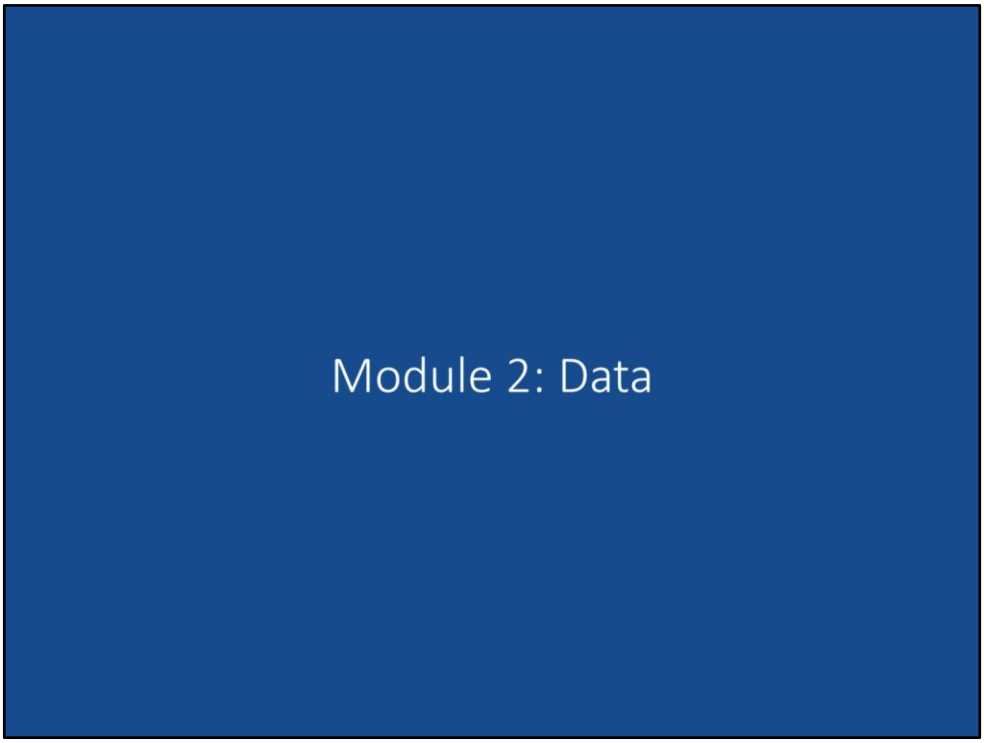
Yale University

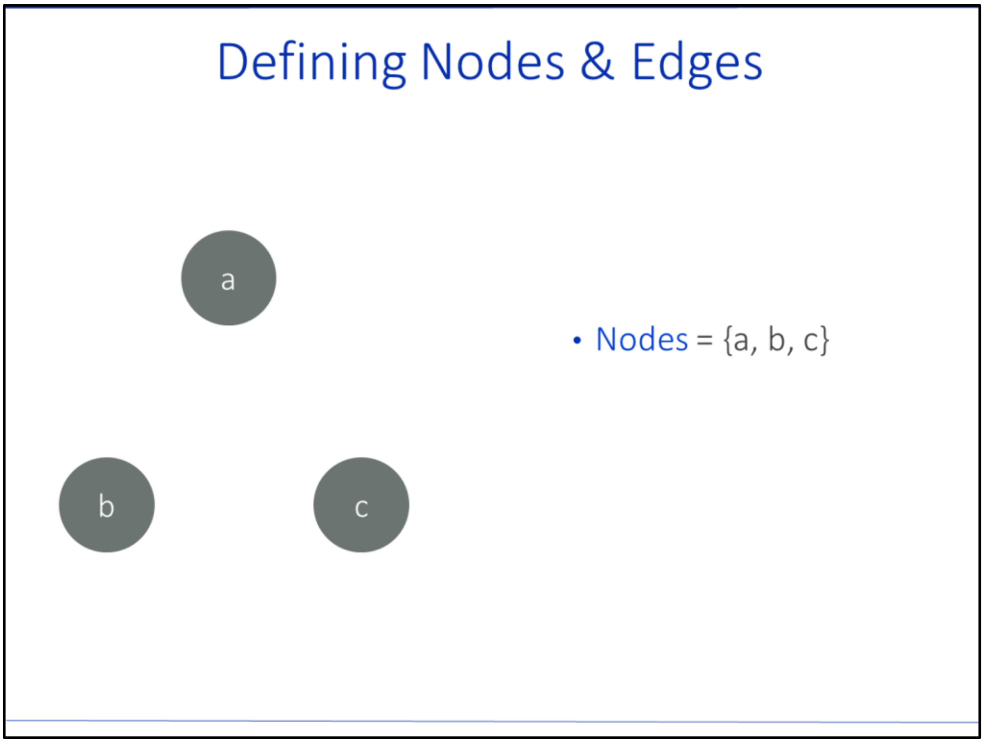
Social Network Analysis for Criminal Justice Practitioners and Analysts

Module 2: Data

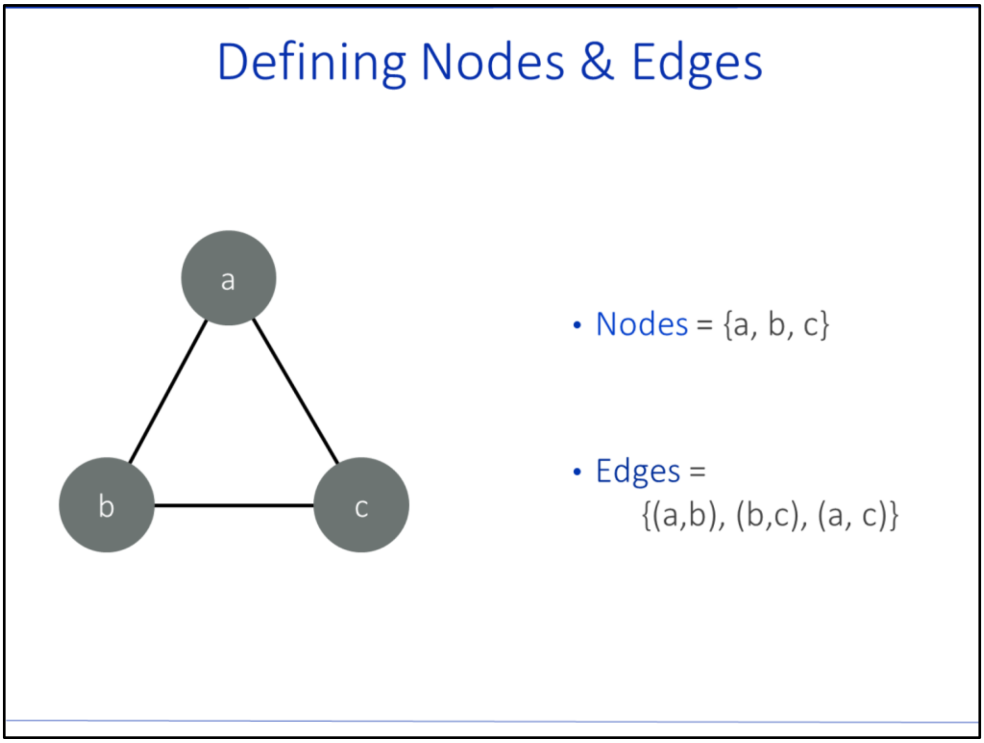
Andrew V. Papachristos



Welcome to Module 2: Data. The goal of this module is to expose participants to the logic and organization of social network data. Module 2: Data includes sections on basics of social networks, directed and undirected ties, the data structure of sociomatrices, the data structure of edgelists, attribute files, and the two-mode data of events or affiliations. This module concludes with a variety of ideas for where to find relational data.



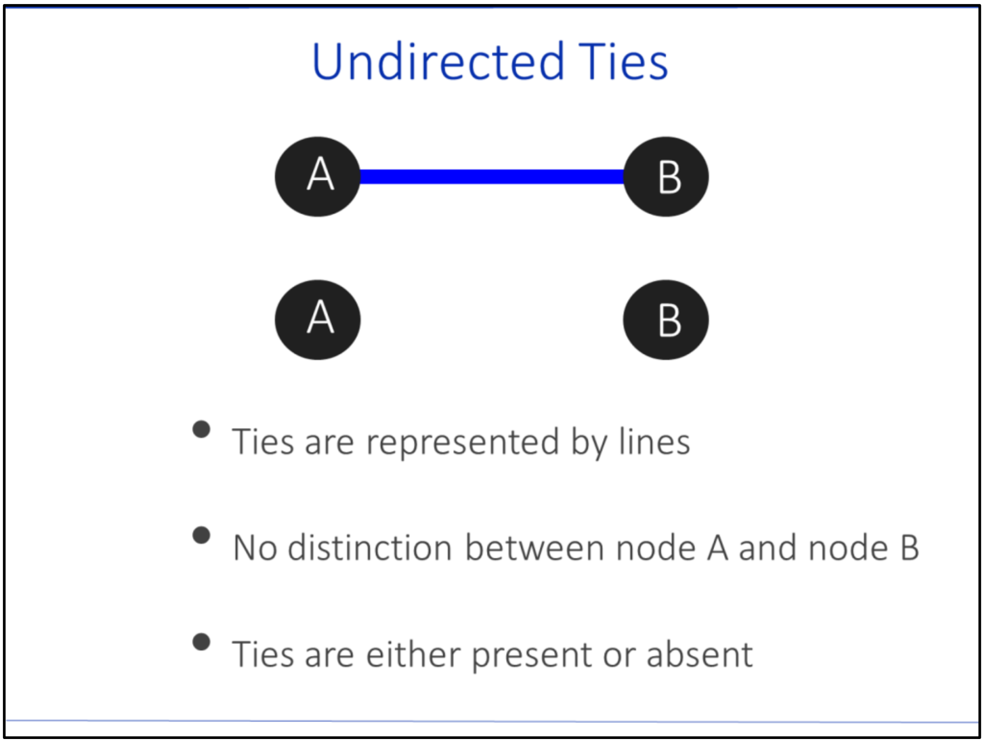
As defined in the first module, every social network requires two pieces of information, and these two pieces of information are required in the different data structures that can be used in social network analysis. The first piece of information is on the nodes or the dots representing actors. Here are three nodes labeled with the letters a, b, and c.



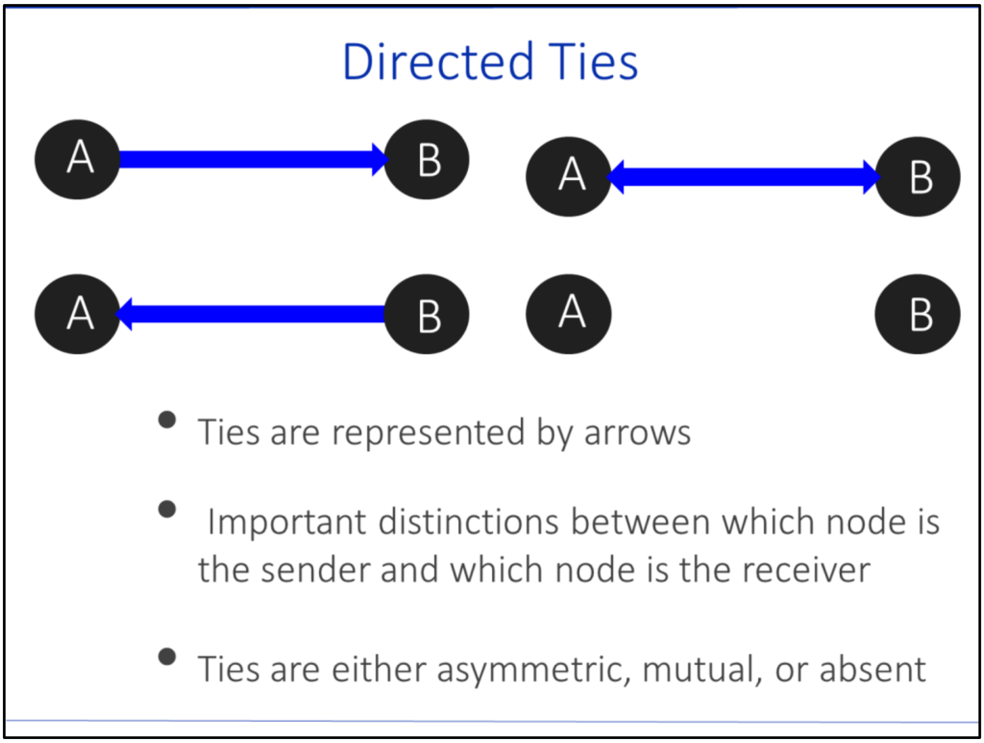
The second piece of information required is on the edges or the lines between the nodes representing relationships or ties. There are three edges in this example: the edge between nodes a and b (a, b), the edge between nodes b and c (b, c), and the edge between nodes a and c (a, c). Notice also that all of the edges in this example are the same length and thickness. More advanced networks could have the length of edges equal to distance in space or the width of edges equal to some varying values, but those applications are beyond the scope of this tutorial. For our purposes, all edges are equal.

Directed and Undirected Ties

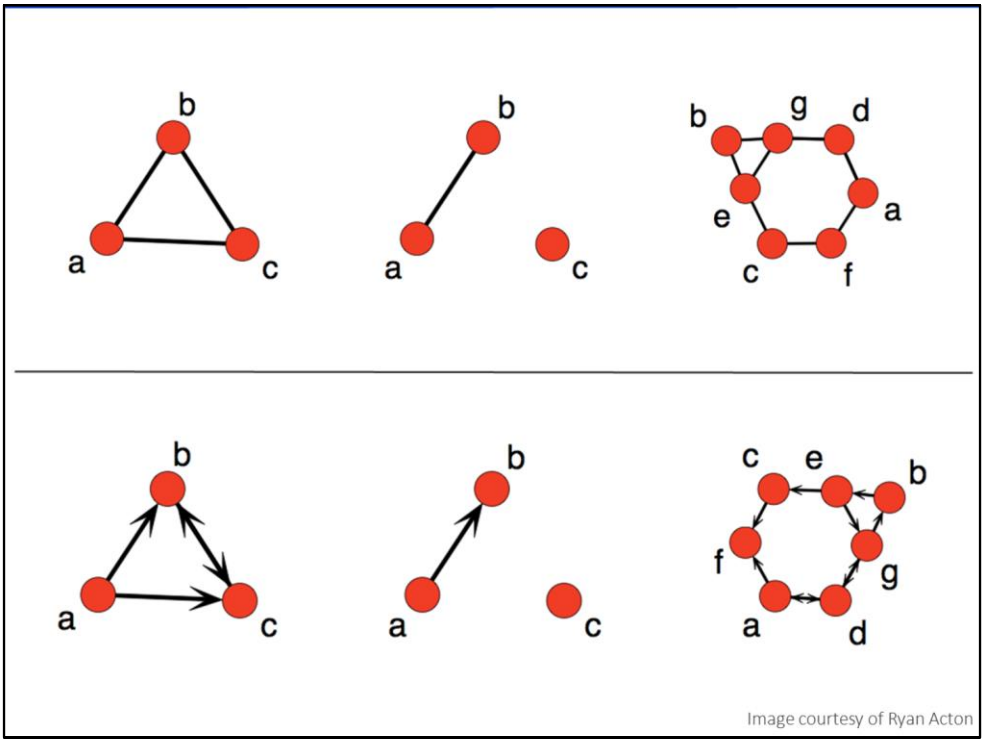
Ties in a network can be lines or arrows. This section introduces important distinctions between edges (lines) and arcs (lines with arrowheads) regarding direction in networks.



Undirected ties are represented by lines. In an undirected tie there is no distinction between node A and node B. A is to B what B is to A. In undirected networks there are only two possibilities between two nodes: (1) the presence of a tie or (2) the absence of a tie. Examples of ties that are undirected include friendships, associations, co-arrests, co-memberships, etc. These are symmetric or mutual relationships or associations. One node doesn’t give or receive a tie like friendships, co-arrests, or associations. These examples are ways that people are connected equally or mutually.



Directed edges are represented by arrows rather than by lines. Directed ties have an important distinction between node A and node B, or the sender and the receiver. In directed networks, there are four possibilities between two nodes: (1) an asymmetric tie from A to B, (2) an asymmetric tie from B to A, (3) a symmetric mutual tie (A to B and B to A), or (4) two absent ties (nothing from A to B and nothing from B to A). Asymmetric ties are ties in which one node sends the tie to a receiving node, but the receiving node does not return the tie to the sender—the relationship is asymmetric. Mutual ties are ties in which both nodes in a pair send the exact same tie to the other—even though the relationship is directed, it is a mutual relationship. Absent ties are instances when neither node of a pair sends a tie to the other. Examples of ties that are directed include nominations, financial or resource exchanges, disseminating information, offenders and victims, violence, etc. These examples of directed ties are all exchanges that are more interesting with more information. When someone pays someone $20, we want to know who paid the $20 and who received the $20, which requires more information than just $20 was exchanged.



The top row contains three undirected networks. Can you identify which edges are absent in the top row? The bottom row contains three directed networks. Can you identify which ties are symmetric, asymmetric, and absent in the bottom row?

Types of Data Structures

Three types of data structures are commonly used in social network analysis: sociomatrices, edgelists, and two-mode (a.k.a. affiliation) data. This section also introduces the structure of attribute files that accompany social network data.

Sociomatrix

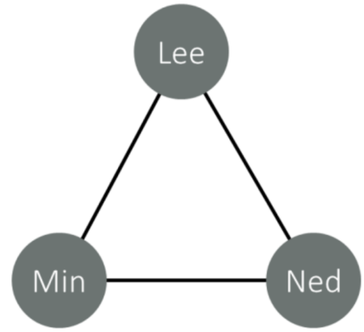
	Lee	Min	Ned
Lee	0	1	1
Min	1	0	1
Ned	1	1	0

- The matrix is a square with the order of nodes in the 1st row the same as the order of nodes in the 1st column
- Each cell indicates the presence of a tie (1s) or the absence of a tie (0s)
- The diagonal of 0s shows that nodes are not connected to themselves

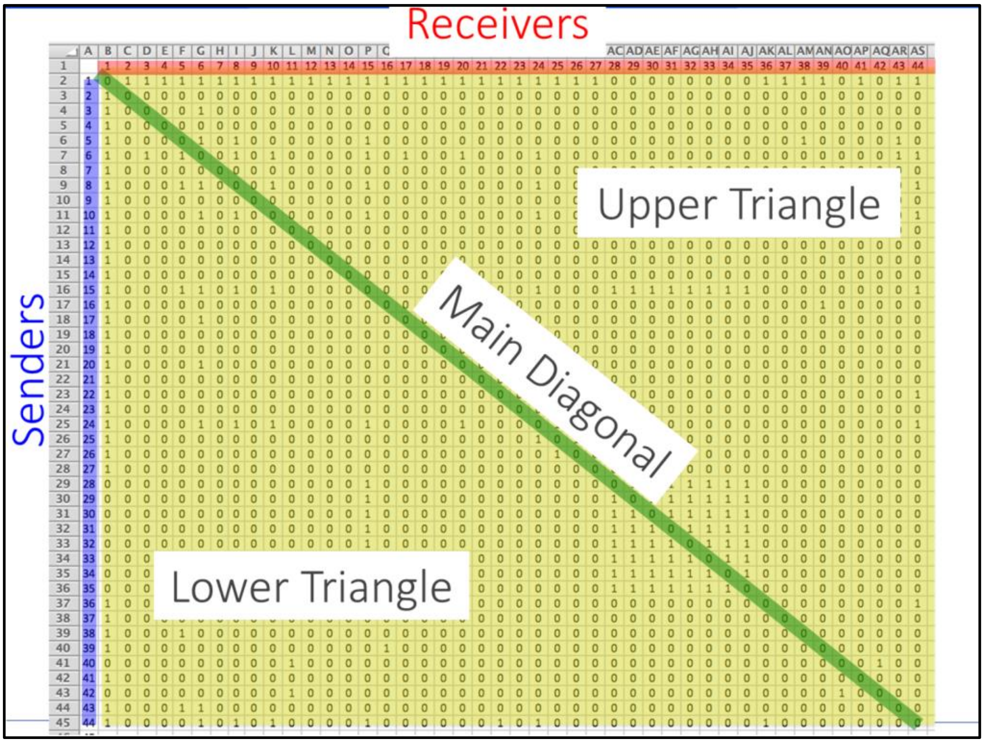
A sociomatrix is a format for organizing social network data. The matrix is a square that contains the same list of nodes along the top row and down the first column. Precise ordering is required when using matrices to organize social network data. Each cell in the middle of the sociomatrix adjoins a name in the top row with a name in the first column and shows information about the edge between those two names or nodes. The value in the cells indicates the presence (1), absence (0), or value (greater than or equal to 1) of edges or relationships. The diagonal shows all zeros because the nodes are not connected to themselves—i.e., there are no loops in this sociomatrix. One benefit of using social network data in sociomatrix form is that sociomatrices are complete because every possible connection between individuals (every cell in the sociomatrix) is accounted for. Another benefit is that the zeros are meaningful in sociomatrices because they clearly show that there is no tie between two nodes—the zeros clearly define absent ties.

Sociomatrix

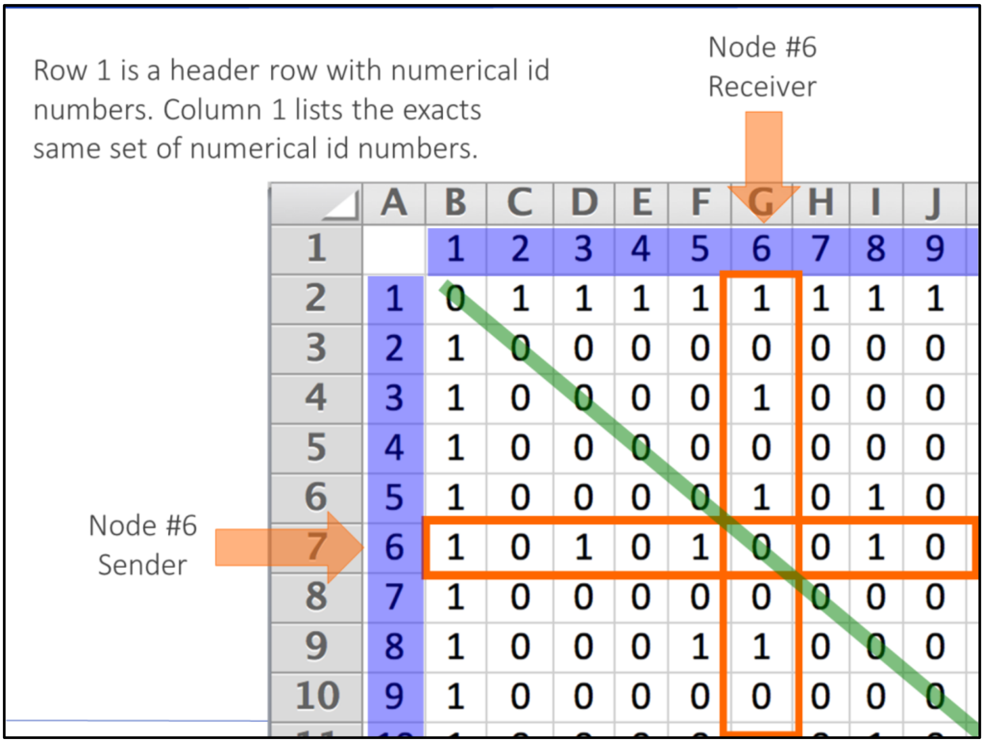
	Lee	Min	Ned
Lee	0	1	1
Min	1	0	1
Ned	1	1	0



The network of three nodes on the right was produced using the sociomatrix on the left. If we read across the rows in the sociomatrix above, we see that Lee has a tie to Min and to Ned, Min has a tie to Lee and to Ned, and Ned has a tie to Lee and Min. If we read down the columns we get the same results. The diagonal of zeros shows that there is no tie from Lee to Lee, no tie from Min to Min, and no tie from Ned to Ned—i.e., there are no loops in this network.



The example above shows a symmetric sociomatrix. This means that if you folded the matrix along the main diagonal of zeros, the upper triangle would mirror the lower triangle. In undirected networks the upper and lower triangles are symmetric meaning that there is no distinction between the senders and the receivers. Technically, you would only need one triangle of the sociomatrix to produce an undirected network since a 1 in the cell between node #1 and node #6 in the sociomatrix above is the same as a 1 in the cell between node #6 and node #1. If the ties in the network are directed, then the upper triangle and the lower triangles of the sociomatrix would not be the same—the triangles would by asymmetric. You would need both triangle to have the complete data for your social network.



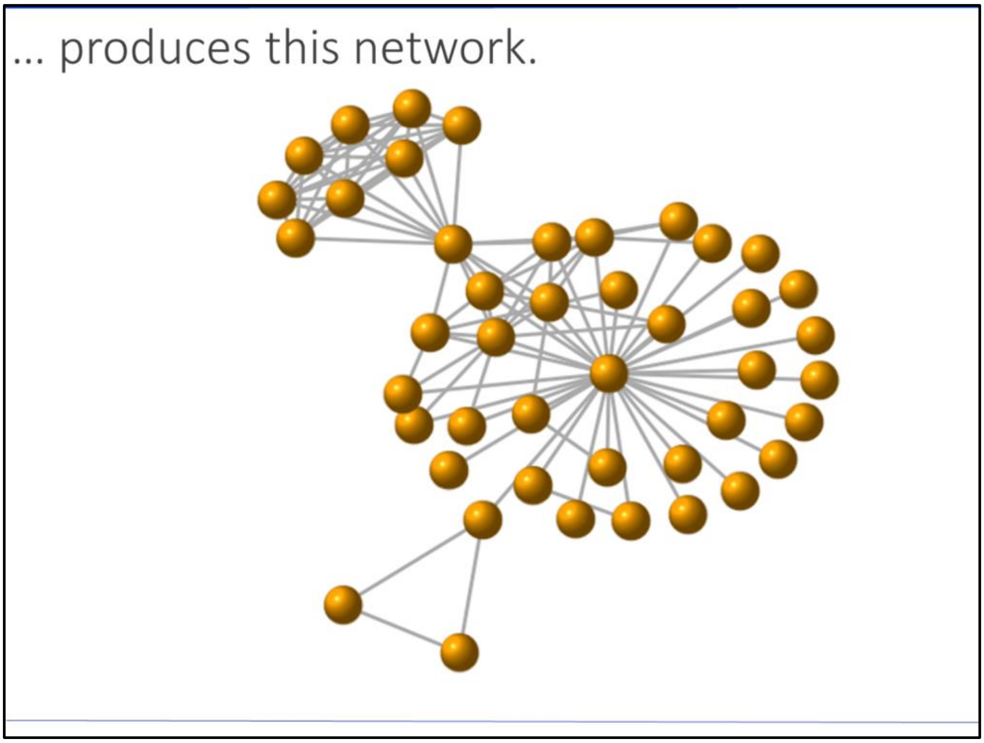
	A	B	C	D	E	F	G	H	I	J
1		1	2	3	4	5	6	7	8	9
2	1	0	1	1	1	1	1	1	1	1
3	2	1	0	0	0	0	0	0	0	0
4	3	1	0	0	0	0	1	0	0	0
5	4	1	0	0	0	0	0	0	0	0
6	5	1	0	0	0	0	1	0	1	0
7	6	1	0	1	0	1	0	0	1	0
8	7	1	0	0	0	0	0	0	0	0
9	8	1	0	0	0	1	1	0	0	0
10	9	1	0	0	0	0	0	0	0	0

In the case of directed networks in sociomatrices, the rows are the senders and the columns are the receivers. For example, reading across row 7 shows that node 6 (minus 1 because of the header row) sends a tie to node 1 (the 1 in column B), node 3 (the 1 in column D), node 5 (the 1 in column F), node 8 (the 1 in column I), etc. Node 6 does not send a tie to node 2 (the 0 in column C), node 4 (the 0 in column E), etc. To see which nodes send a tie to node 6 (when node 6 is the receiver), read down column G.

This sociomatrix …

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1
3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	5	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
7	6	1	0	1	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
8	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	8	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	10	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
12	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
13	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	15	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
17	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
18	17	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	20	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
24	23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	24	1	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26	25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	27	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
30	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
31	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
32	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0
33	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
34	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
35	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
36	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
37	36	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
38	37	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	38	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	39	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	40	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
42	41	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	42	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
44	43	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	44	1	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

We will return to this sociomatrix in the lab of Module 2. This is called the Siren Matrix.



The Siren sociomatrix on the previous slide produces this network image. The ties are undirected. Each 1 on either the upper or lower triangle of the Siren sociomatrix produced one of the edges in this network image. Each 0 in either the upper or lower triangle of the Siren sociomatrix resulted in absent edges or no relationship between those two nodes. Now let’s move to another data structure.

Edgelist

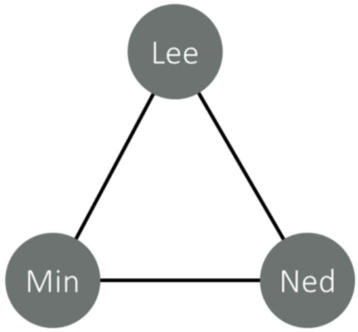
Sender	Receiver
Lee	Min
Lee	Ned
Ned	Min

- Each row in an edgelist represents one tie in the network
- Edgelists require at least 2 columns: column 1 is the sender and column 2 is the receiver
- The order of sender and receiver do not matter for undirected ties

An edgelist is another format for organizing social network data. Edgelists contain two columns listing nodes that are connected to each other and sometimes a third column containing more information about the edge. Each row of an edgelist represents one edge in the social network. The ordering of the rows does not matter in edgelists. The ordering of names in the columns only matters if the network is directed. It does not matter who you list as the sender and who you list as the receiver in an undirected network. In a directed network, however, the sender goes in the first column and the receiver goes in the second column. One benefit of edgelists is that they are easy to manage for large networks. Compared to sociomatrices, edgelists can feel incomplete because you don't include absent edges. This means that missing data and absent edges are treated the same in an edgelist.

Edgelist

Sender	Receiver
Lee	Min
Lee	Ned
Ned	Min

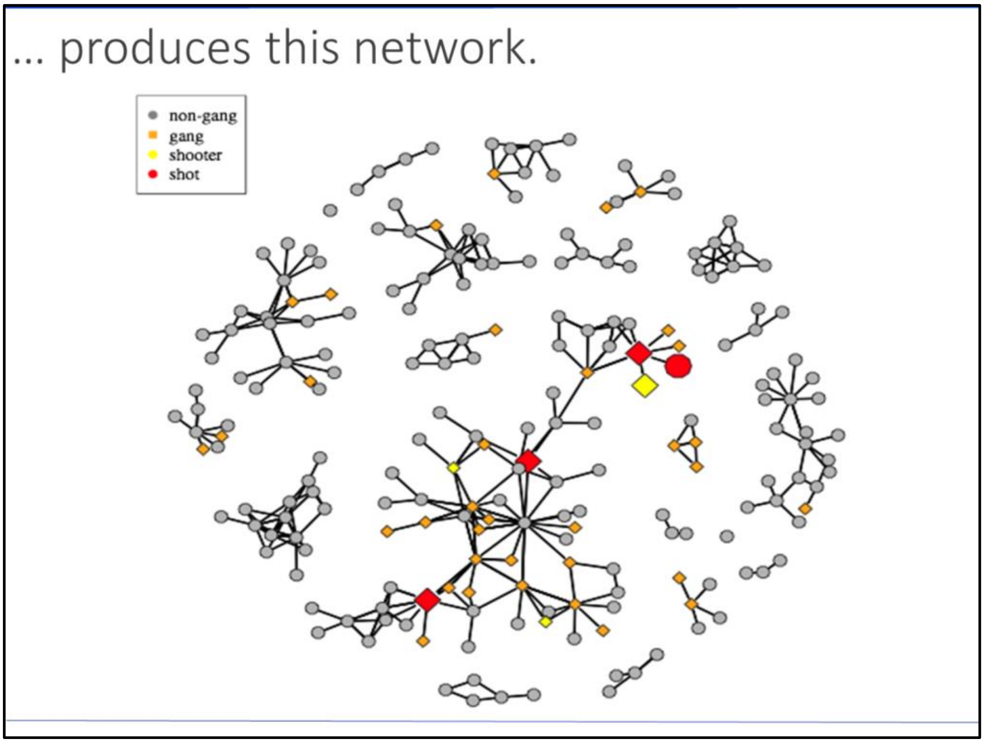


The undirected network of three nodes on the right was produced using the edgelist on the left. Reading across the three rows, we see that Lee has a tie to Min, Lee has a tie to Ned, and Ned has a tie to Min. Because this network is undirected, the reverse is also true: Min has a tie to Lee, Ned has a tie to Lee, and Min has a tie to Ned.

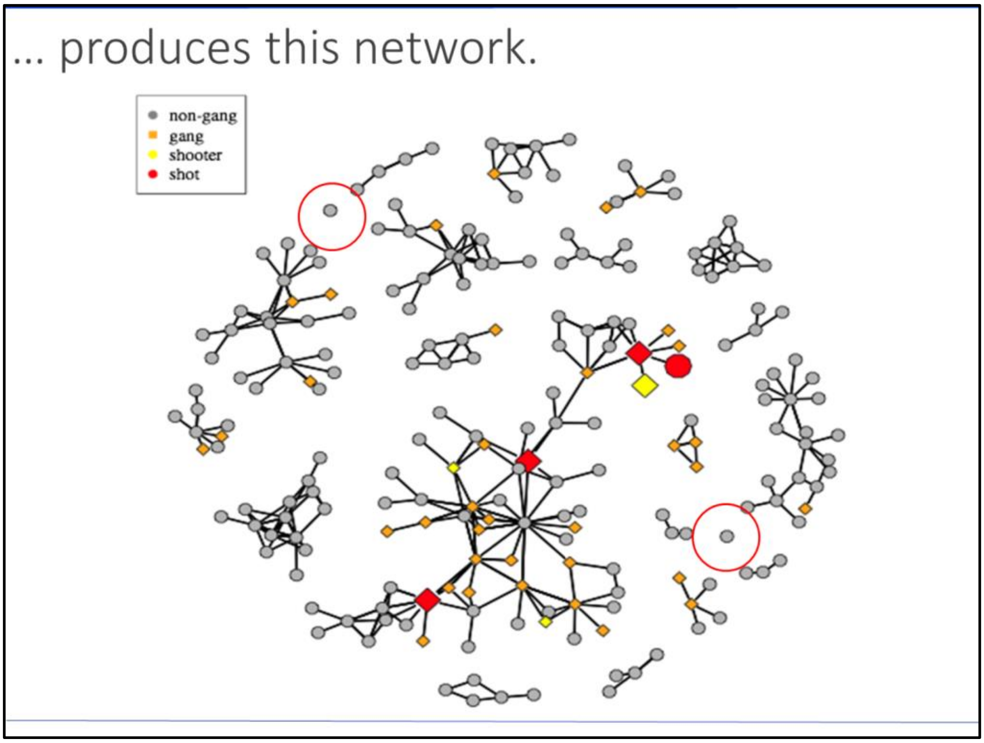
This edgelist …

V1	V2
1	221
1	238
1	254
1	217
1	234
1	3
1	151
1	6
1	12
1	108
1	328
1	4
1	184
1	265
1	324
1	67
1	7
1	193
1	2
1	5
1	17
1	133
1	135
1	339
1	231

Above is the beginning of a lengthy edgelist. In the first column under V1, we see a list of 1s. The beginning of this edgelist connects node #1 to 25 other individuals. The second column (V2) is the list of all of the receivers or other nodes that node #1 is connected to. Note that the receivers in column V2 do not appear to be in any order. Just so long as each edge is accounted for in a row, the order of the rows does not matter. Also this is an undirected network, so we could reverse columns 1 and 2 and the results would be the same.



Though you only saw a partial image of the edgelist on the previous slide, when used in its entirety it produces this network. Each tie in this network required one row in the edgelist that connected two nodes.



Notice that there are two isolates in this network. These nodes were not included in any of the rows in the edgelist.

Attributes

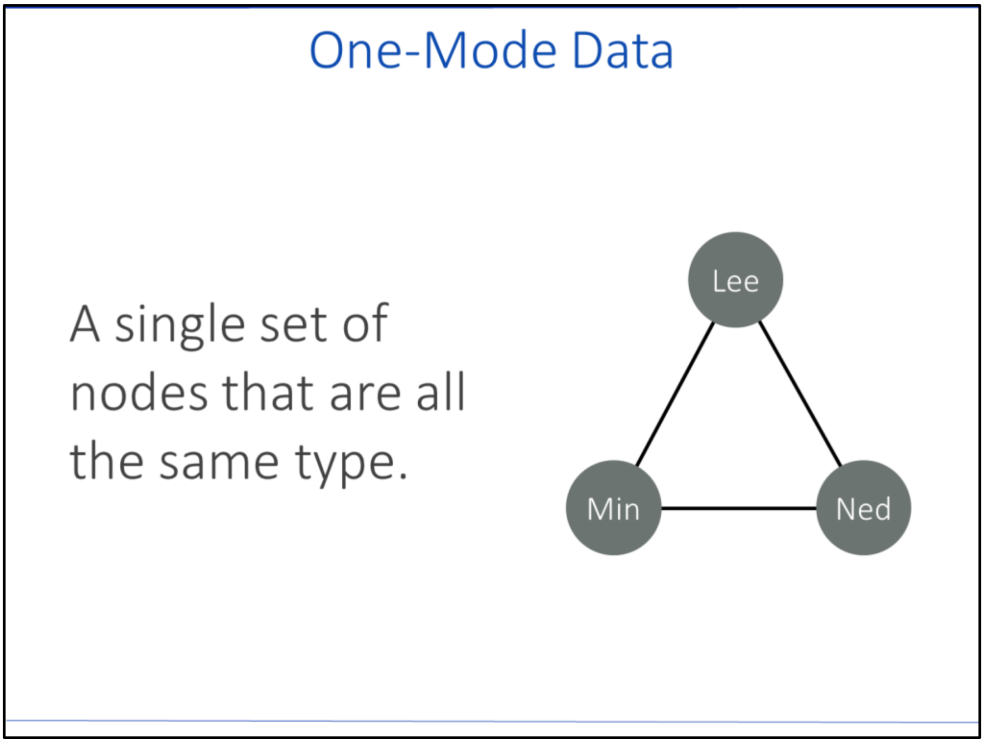
sender	receiver
1	2
1	3
1	4
2	3
2	4
3	4
4	5
5	6
6	7
7	8
7	9
10	11
12	13
12	14

edgelist.csv

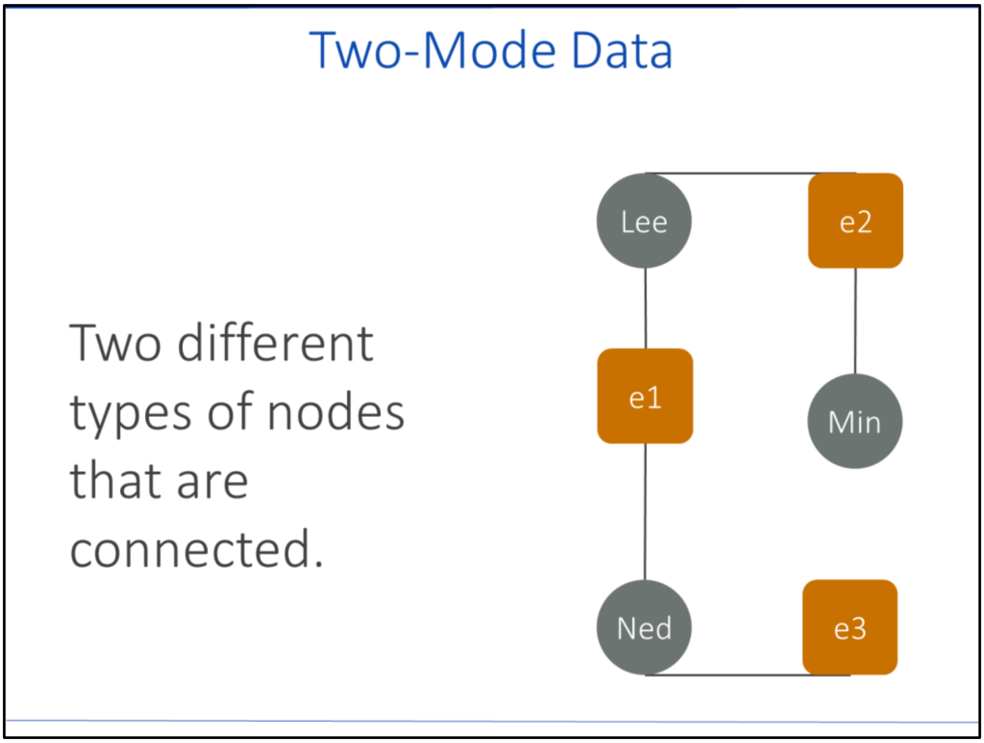
ID Number	Name	Sex	GPA
1	Joe	male	3.2
2	Sandy	female	3.5
3	Tom	male	2.5
4	Alicia	female	4
5	Andy	male	2.9
6	Don	male	3
7	Jen	female	3.1
8	Tina	female	3.5
9	Brad	male	2.8
10	Moreno	male	3
11	Joe	male	3.8
12	Hector	male	2.9
13	MJ	female	3
14	Joan	female	3

attributes.csv

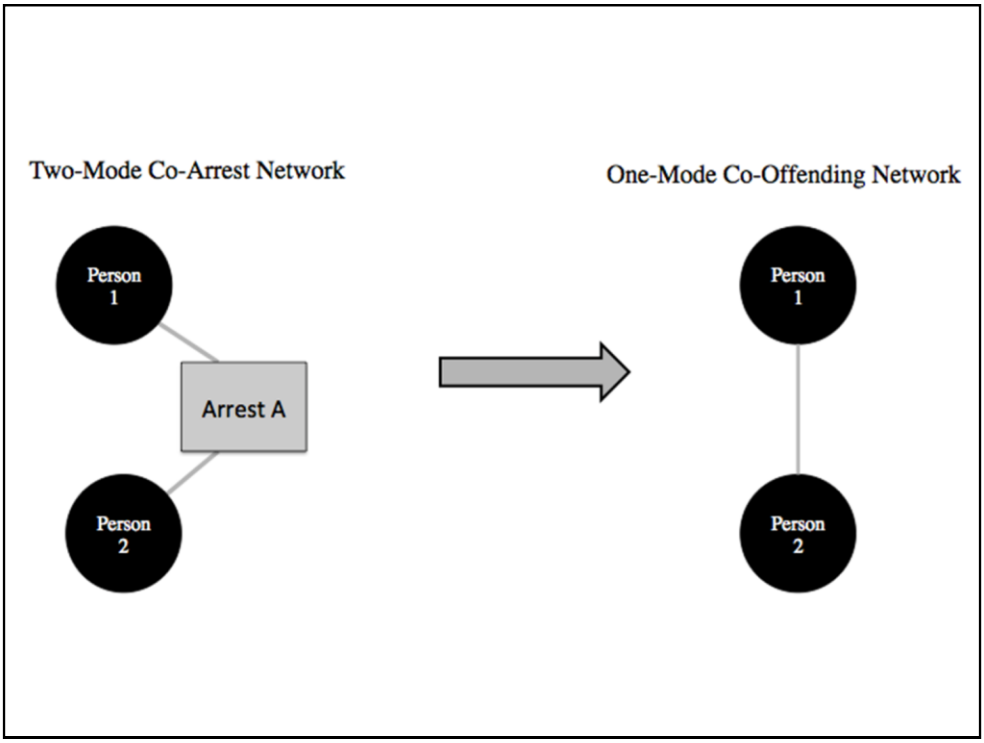
An attribute file can accompany any sociomatrix or edgelist. As you may have noticed in the examples above, sociomatrices and edgelists contain a lot of information on relationships but no information on individual nodes. Node-level information requires a separate file. In the example above, there is an edgelist on the left in which the nodes are identified by a number. On the right is a spreadsheet of attributes. The individual-level information matches each node through the id number. The id numbers in the attribute file on the right match the node numbers in the edgelist on the left. The attribute file allows you to vary the display of the nodes in a network. For example, you could add the names as labels to each node, color code the nodes by gender so that nodes for males were green and nodes for females were red, and you could vary the size of the nodes by grade point average (gpa) so that nodes with higher GPAs appear larger and nodes with smaller GPAs appear smaller. If you work with spreadsheets, the organization of the attribute file on the right probably looks familiar.



Now let's distinguish between one-mode data and two-mode data. To this point in Module 2, all of the examples have used one-mode data. One-mode data is a dataset in which all of the nodes measure the same thing—e.g., all the nodes are individual people, or discrete gangs, countries, or organizations—and we list or plot the relationships between the nodes. One-mode data only require either a single sociomatrix or a single edgelist. We are familiar with this triangle network containing Lee, Min, and Ned, which is a one-mode network.



Two-mode data contain two different types of nodes. These are also called bipartite networks or affiliation networks. In the example above, the two types of nodes are people and events. The people are the gray circles: Lee, Min, and Ned. The events are the orange blocks: event 1, event 2, and event 3. We can read this network through events or through people. For example, Lee and Min are connected through event 2, and Lee and Ned are connected through event 1. Only Ned is connected to event 3. Alternatively, event 1 and event 2 are connected through Lee, and event 1 and event 3 are connected through Ned. Imagine these events were meetings at work. We would know who connected during those meetings and which meetings were attended by the same people. Two-mode data could include a network of people and organizations, a network of vendors and buyers, a network of students and classes, a network of people and households, or a network of people and arrests.



Relevant to criminal justice applications is the conversion of two-mode co-arrest networks into one-mode co-offending networks. On the left is a two-mode co-arrest network. Person 1 and person 2 are connected through arrest A. We can convert this into a one-mode network by assuming that individuals involved in the same arrest events are connected through co-offenses, criminal associations, knowing each other, spending time with each other, getting involved in suspicious activities together, etc.

Classic Arrest Data

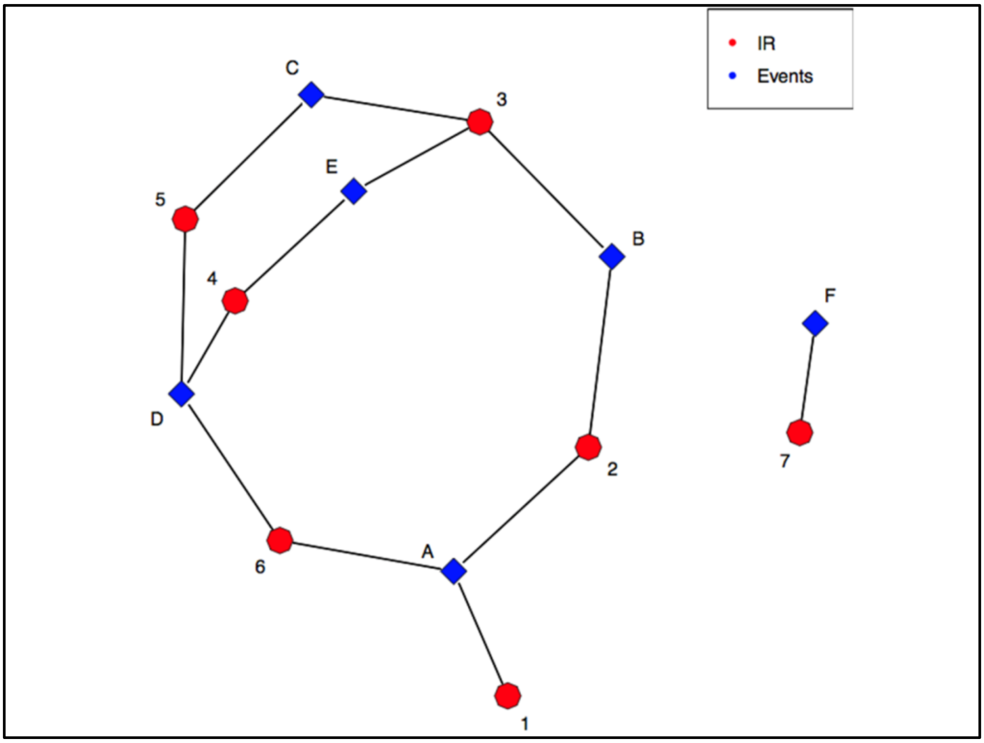
DATE	ARREST_NO	ID_NO
6/1/15 12:10 AM	ARN 105711	2055047
6/1/15 12:12 AM	ARN 105723	2386552
6/1/15 12:15 AM	ARN 105725	2966102
6/1/15 12:15 AM	ARN 105774	2843250
6/1/15 12:18 AM	ARN 105773	2420701
6/1/15 12:18 AM	ARN 105773	2534758
6/1/15 12:20 AM	ARN 105746	3190981
6/1/15 12:20 AM	ARN 105761	2916244
6/1/15 12:20 AM	ARN 105761	3190906
6/1/15 12:22 AM	ARN 105714	2888633
6/1/15 12:22 AM	ARN 105731	2248610

Here is a subset of some fake arrest data. This little arrest dataset includes the date and time of an arrest, and arrest identification number, and an identification number for each person arrested. In classic arrest data, the arrest id numbers can serve as identifiers for event nodes in two-mode data, and the individual arrestee id numbers can serve as identifiers for the person nodes in the two-mode data. In the example above, the “ARREST_NO” is the arrest identifier and the “ID_NO” identifies individuals.

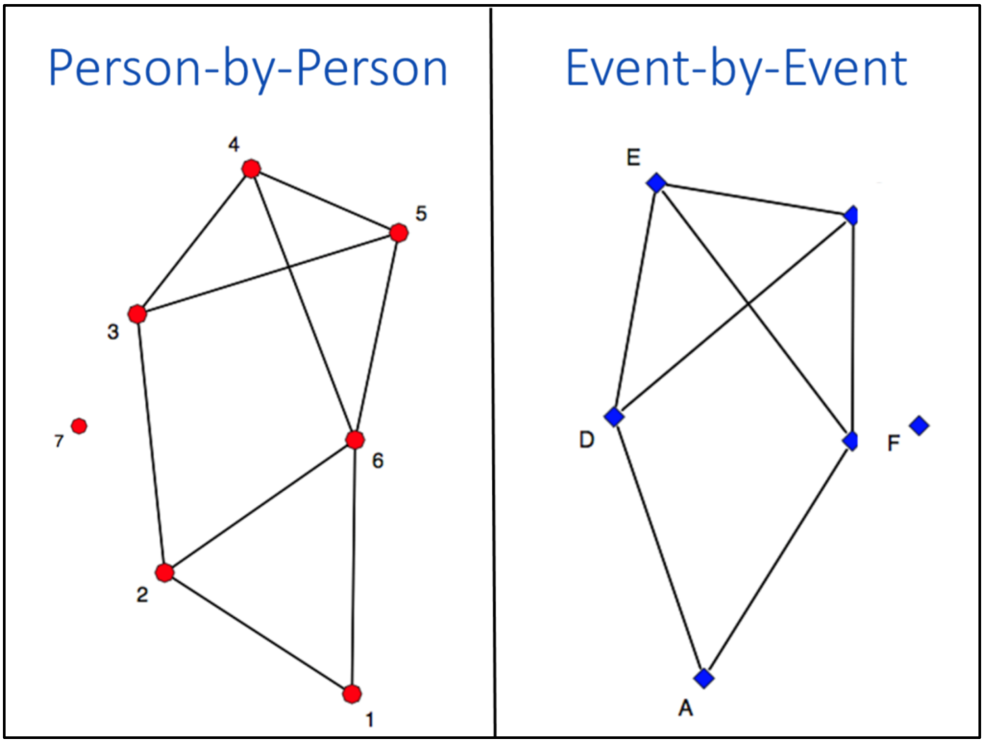
Classic Arrest Data

DATE	ARREST_NO	ID_NO
6/1/15 12:10 AM	ARN 105711	2055047
6/1/15 12:12 AM	ARN 105723	2386552
6/1/15 12:15 AM	ARN 105725	2966102
6/1/15 12:15 AM	ARN 105774	2843250
6/1/15 12:18 AM	ARN 105773	2420701
6/1/15 12:18 AM	ARN 105773	2534758
6/1/15 12:20 AM	ARN 105746	3190981
6/1/15 12:20 AM	ARN 105761	2916244
6/1/15 12:20 AM	ARN 105761	3190906
6/1/15 12:22 AM	ARN 105714	2888633
6/1/15 12:22 AM	ARN 105731	2248610

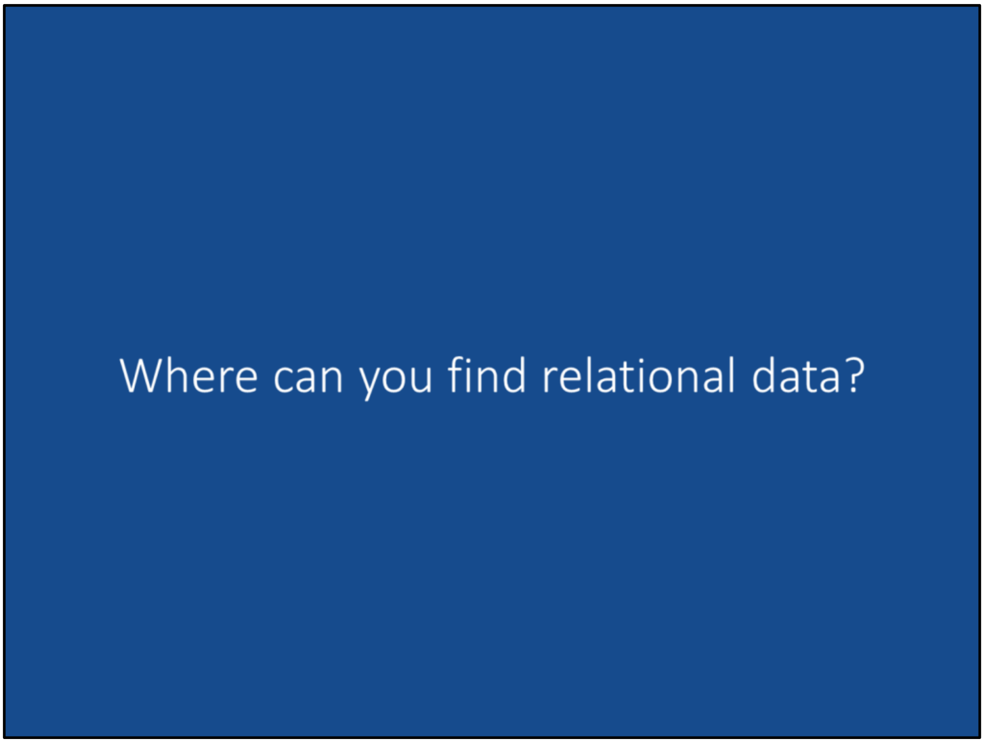
Notice that there are two co-arrests in this subset. Arrest #ARN 105773 links individuals #2420701 and #2534758, and arrest #ARN 105761 links individuals #2916244 and #3190906. Based on these co-arrests, we could produce a one-mode co-offending network assuming that the co-arrest linked co-offenders.



Here is an example of a two-mode arrest network. The blue diamond nodes are arrest events and the red octagons are the people arrested. We can convert two-mode networks into two different one-mode networks.

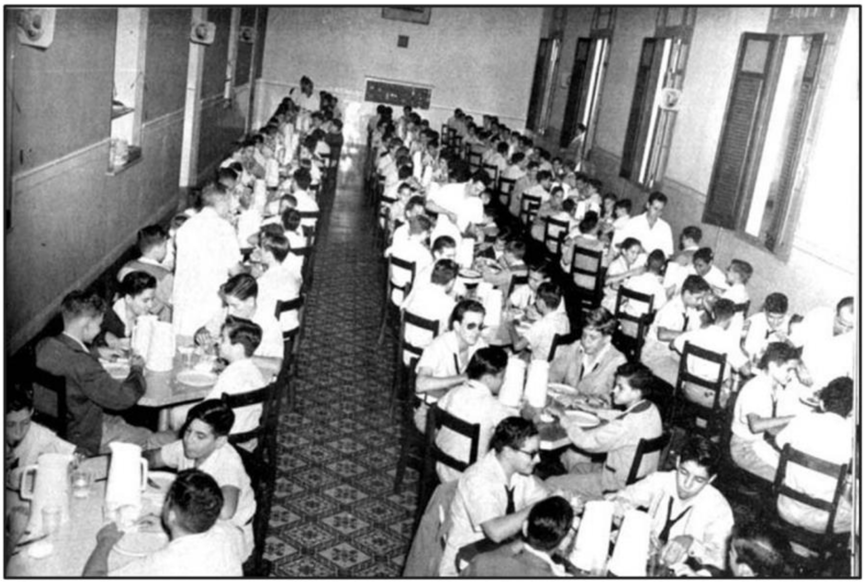


Here the two-mode network from the previous slide is converted into two one-mode networks. These are called one-mode projections. The person-by-person network and the event-by-event network both contain two components like the two-mode network from the previous slide. We no longer see which events connect which people and vice versa, but we are able to see the a larger co-offending network and a larger events network. When we remove the arrest events that connect the people, then we can more clearly see the structure of the ties between the people. When we remove the people that connect the arrest events, then we can more clearly see the structure of the ties between the arrests.



Now let’s think about where researchers can collect social network data.

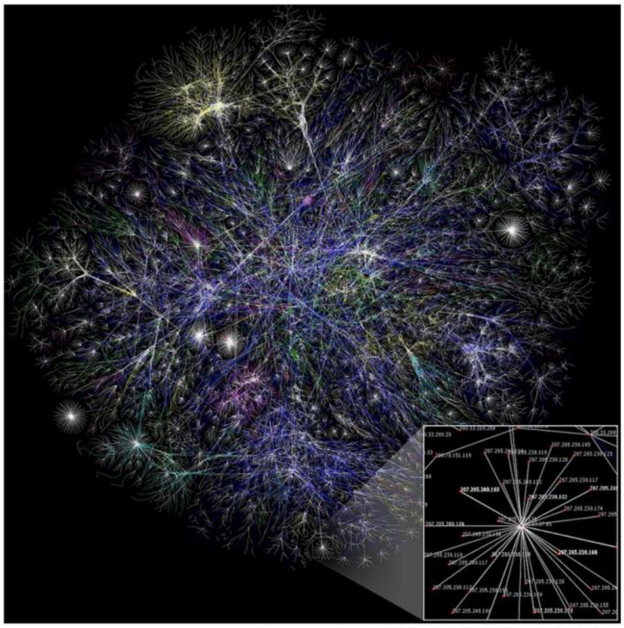
Observations



"Belen School cafeteria Havana" is public domain.

A systematic recording of observations could produce relational data for social network analysis. We could observe who sits together in a cafeteria, who plays together on the playground, or who hangs out together on the street. Which data structure might you use to turn this observation of a school cafeteria into data for a social network? There is no one answer, but you could use each lunch table as an event or affiliation and create a two-mode dataset. You could create a sociomatrix of the entire cafeteria and fill in 1s for people eating together and 0s for people not eating together. Or you could create an edgelist that would connect each pair sitting at the same table. (Public domain image available at: http://commons.wikimedia.org/wiki/File:Belen_School_cafeteria-_Havana.jpg#globalusage)

The Internet



"Internet Map 1024" (2006) by Matt Britt is licensed under CC BY 2.5.

Examples of social networks abound on the Internet, especially social networking websites, but even the Internet itself is one giant network as shown in this 2006 image above. Nodes are websites and edges are the links between websites. (Image available through Creative Commons: http://en.wikipedia.org/wiki/File:Internet_map_1024.jpg)

Surveys

Please think about the relations between the people you just mentioned. Some of them may be total strangers in the sense that they wouldn't recognize each other if they bumped into each other on the street. Others may be espe cially close, as close or closer to each other as they are to you.

First, think about NAME 1 and NAME 2.
ASK CLOSE1 FOR FIRST PAIR.
A. Are __________ and __________ total strangers?

IF YES.......(ASK NAME1, NAME2, NAME3, NAME4, NAME5 FOR NEXT PAIR DOWN)
IF NO........(ASK CLOSE12, CLOSE13, CLOSE14, CLOSE15, CLOSE 23, CLOSE24, CLOSE25, CLOSE34, CLOSE35, CLOSE45)

B. Are they especially close? PROBE: As close or closer to each other as they are to you?

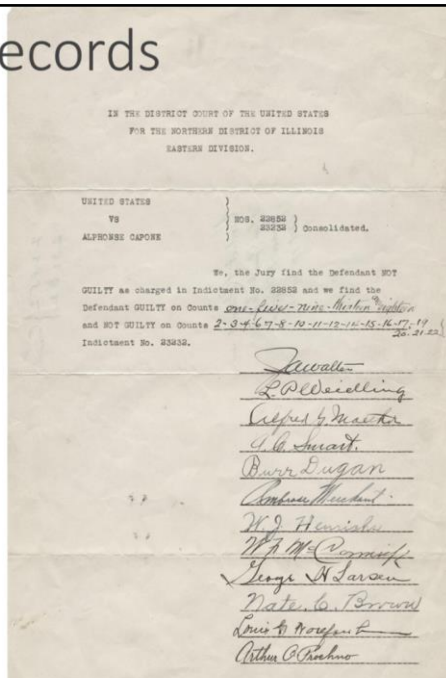
IF YES.......(ASK NAME1, NAME2, NAME3, NAME4, NAME5 FOR NEXT PAIR DOWN)
IF NO........(ASK NAME1, NAME2, NAME3, NAME4, NAME5 FOR NEXT PAIR DOWN)

PERSON	NAME1	NAME2	NAME3	NAME4	NAME5
NAME 2	A. YES....1 NO..... 2 B. YES...1 NO....2				
NAME 3	A. YES...1 NO.... 2 B. YES....1 NO.....2	A. YES...1 NO....2 B. YES...1 NO....2			
NAME 4	A. YES...1 NO.... 2 B. YES....1 NO.....2	A. YES...1 NO....2 B. YES...1 NO....2	A. YES...1 NO....2 B. YES...1 NO....2		
NAME 5	A. YES....1 NO..... 2 B. YES....1 NO.....2	A. YES...1 NO....2 B. YES...1 NO....2	A. YES...1 NO....2 B. YES...1 NO....2	A. YES...1 NO....2 B. YES...1 NO....2	

General Social Survey

Social scientists have designed surveys to collect social network data. These can get rather complex depending on the size. The United States General Social Survey includes a topic module that asks social network questions to a random sample of the U.S. population and has them list up to five people with whom they have discussed important matters in the past 6 months. Then there is a series of questions on if and how the people in one's social network are connected. This survey creates small networks for each person completing the survey, called ego-networks.

Archival Records



IN THE DISTRICT COURT OF THE UNITED STATES
FOR THE NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION.

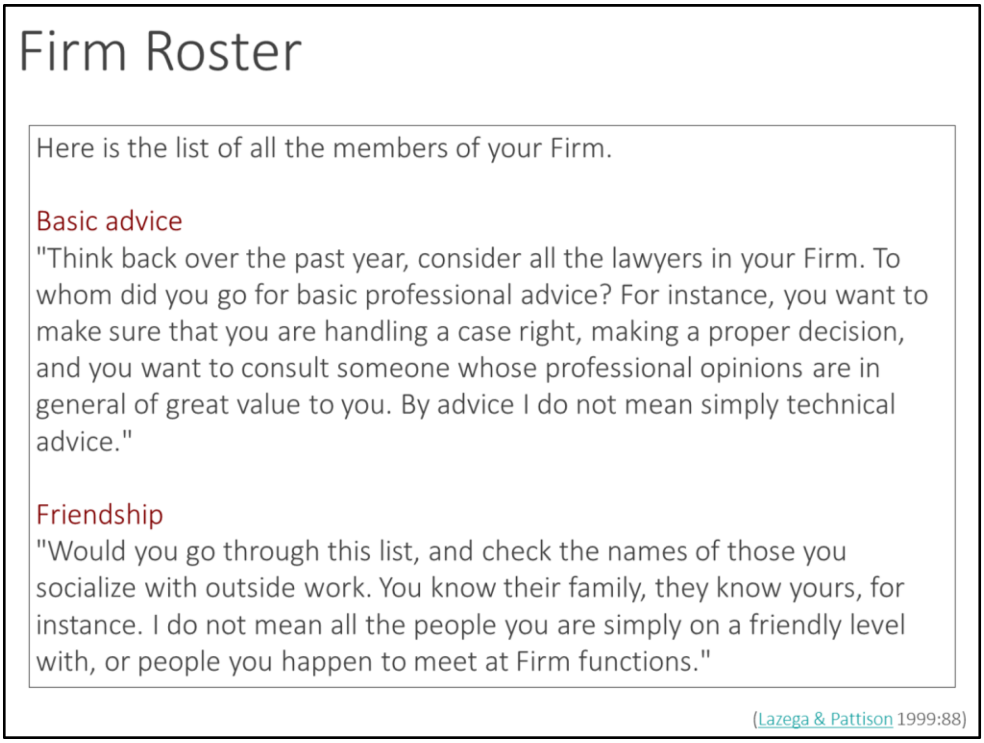
UNITED STATES
VS
ALPHONSE CAPONE

NOS. 22852 } 23232 } Consolidated.

We, the Jury find the Defendant NOT GUILTY as charged in Indictment No. 22852 and we find the Defendant GUILTY on Counts one-five-nine-thirteen-eighteen and NOT GUILTY on Counts 2-3-4-6-7-8-10-11-12-14-15-16-17-19-20-21-22 Indictment No. 23232.

National Archives Today's Document. Verdict in United States of America v. Alphonse Capone, October 17, 1931; Records of the U.S. District Court for the Northern District of Illinois

Archival documents contain information on events and associations that often link individuals. Here we could make a network of the jurors who found Al Capone guilty of tax fraud in 1931. (Image available at: http://todaysdocument.tumblr.com/search/al+capone)



Social scientists have also gone into firms and collected complete organizational networks on advice, friendship, collaboration, etc. between all of the members of that firm. For example, a study by Emmanuel Lazega and Philippa Pattison in 1999 used the questions above in a study of networks within a law firm. (This article is available at: http://www.sciencedirect.com/science/article/pii/S0378873399000027)

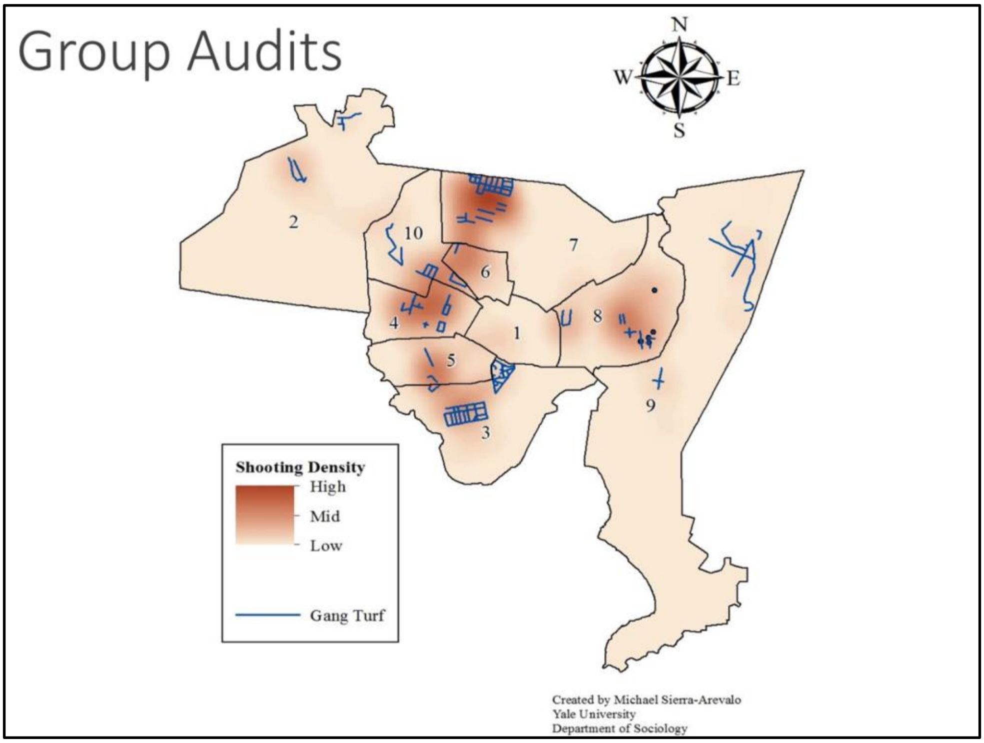
Official Records

Providence Police Department

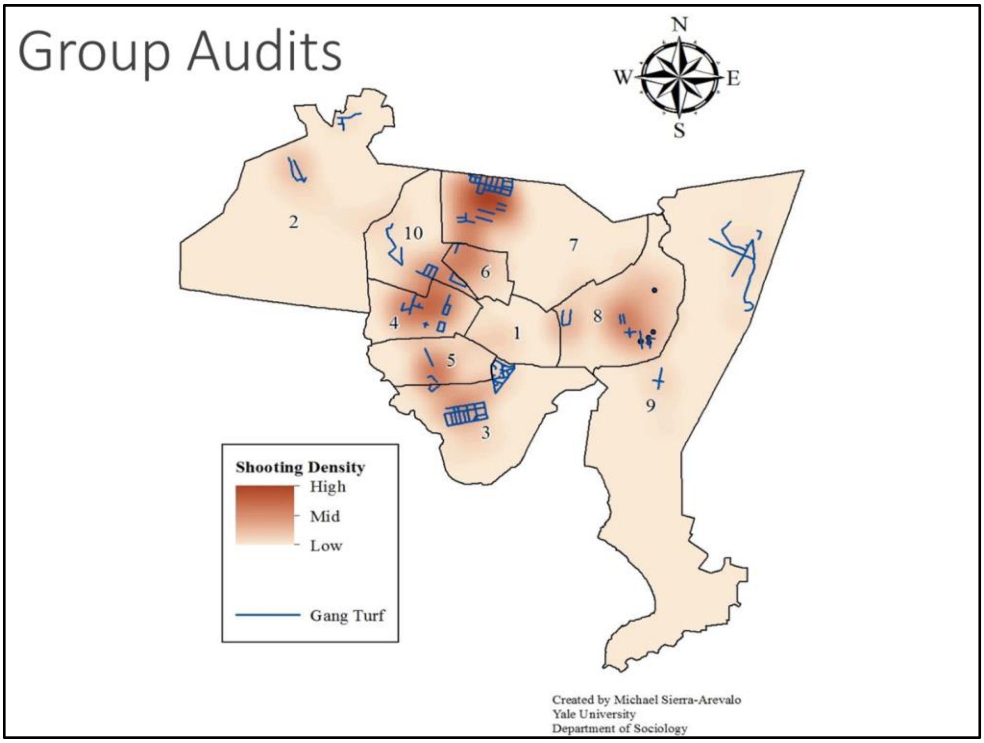
Arrest Charge Statute Description	First Name	Last Name	Age	Arrest Date And Time	Arrestee City	Arrestee State	Case Number
Murder-1St Degree; Loitering For Indecent Purposes Prostitution - Prostitution; Failure To Appear/Answer Summons;	Jovani	Torres	31	10/12/2012 00:00:00	Providence	RI	2012-00050110
Loitering For Indecent Purposes Prostitution - Prostitution; Failure To Appear/Answer Summons;	Sandra	Rivera	48	10/12/2012 00:00:00	Providence	RI	2012-00098812
Failure To Appear/Answer Summons;	Lillian	Rivera	51	10/12/2012 00:00:00	Providence	RI	2012-00098812
Loitering For Indecent Purposes Prostitution - Prostitution;	Hye Min	Jung	37	10/12/2012 01:01:00	Whitestone	NY	2012-00099112
Assault W/ Intent To Commit A Felony - Minor Injury;	Francisco	Depina	20	10/12/2012 01:10:00	Providence	RI	2012-00099092
Resisting Legal Or Illegal Arrest; Disorderly Conduct;	Kyle	Mcneil	19	10/12/2012 01:19:00	Warwick	RI	2012-00099116
Loitering For Indecent Purposes Prostitution - Assist/Promot;	Mi Jang	Park	46	10/12/2012 01:48:00	Milford	MA	2012-00099109
Loitering For Indecent Purposes Prostitution - Assist/Promot;	Ji Yu	Jin	43	10/12/2012 01:49:00	Flushing	NY	2012-00099109

Providence Police Department Daily Arrest Log October 12, 2012

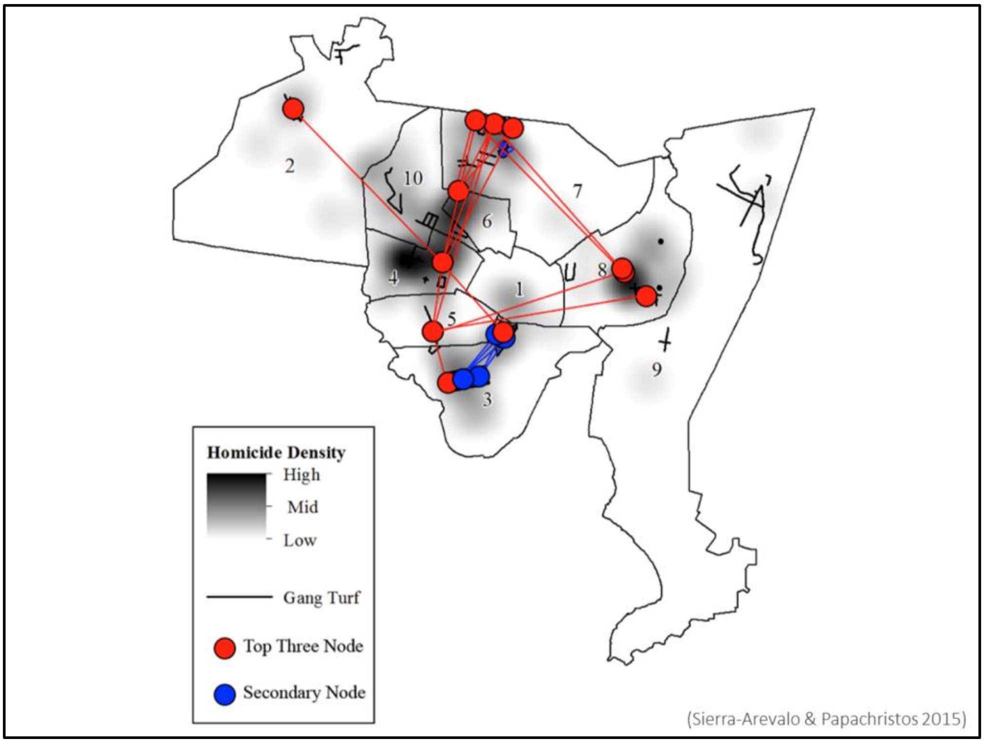
Official arrest records contain information to create two-mode networks that can be converted into one-mode networks. Some of these are publicly available, such as this example from the Providence Police Department Daily Arrest Log in 2012. People are connected through matching case numbers. (http://www.providenceri.com/efile/3346)



A group audit is a criminal justice tool that facilitates the collection of data on groups. Neighborhood patrol, homicide detectives, gang experts, officers from other specialized units, and probation and parole officers come together with a project manager to combine and map their knowledge and experience of the criminal groups in the area. Groups aren't necessarily just gangs; groups also include "any set, clique, or crew of individuals who run together" (NNSC 2015:36). Gathered around maps of the city or maps of specific areas, officers identify group names, territories on the map, number of members, influential or violent individuals in the group, illegal group activities, conflicts with other groups, alliances with other groups, level of violence, level of organization, and any larger affiliation (NNCS 2015:36). A detailed guide on conducting group audits, *Group Violence Intervention: An Implementation Guide*, by the National Network for Safe Communities is available through the U.S. Department of Justice, Office of Community Oriented Policing Services. The complete bibliographic information for this guide is included in the reference list at the end of this tutorial.

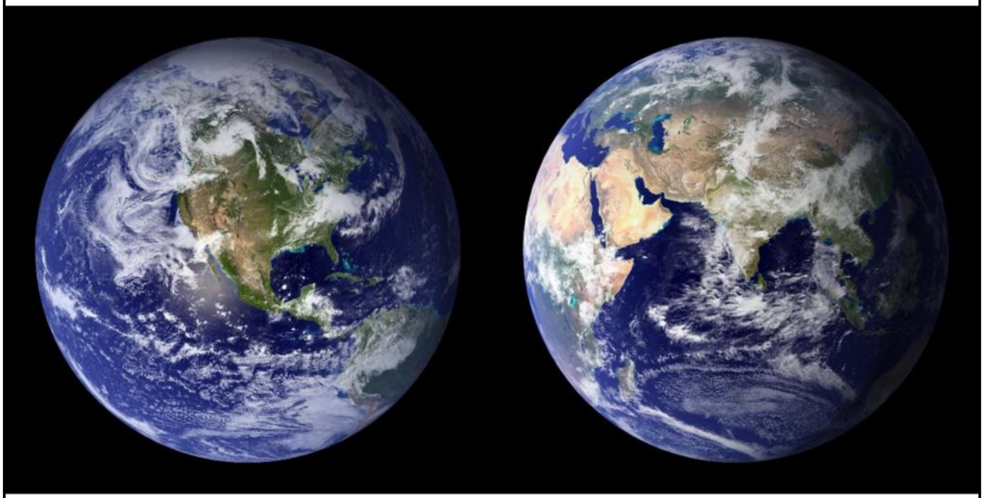


Near the end of 2012, New Haven became part of a statewide project to reduce gun violence by targeting specific populations in high-violence areas. This was a focused-deterrence program called Project Longevity. Group audits were one of the initial methods used to determine: (1) which groups were involved in New Haven's violence, (2) the membership and activities of identified groups, and (3) the patterns of relationships among these groups. The New Haven audit identified 52 active groups, with at least one group or gang identified in nine of the ten police districts. The map above shows blue lines and black points that identify streets and addresses that police identified as being group turf or hangouts. The numbers identify the police district. The darker areas on the map indicate areas with higher shooting density and also tend to be the areas with gang presence.



From group audits can come figures like the one here that shows how relational and geographic data can be combined. The nodes representing different gangs are plotted over their turf and hangout locations and the edges represent alliances or feuds between the groups.

Everywhere



"Blue Marble 2001-2002" is public domain.

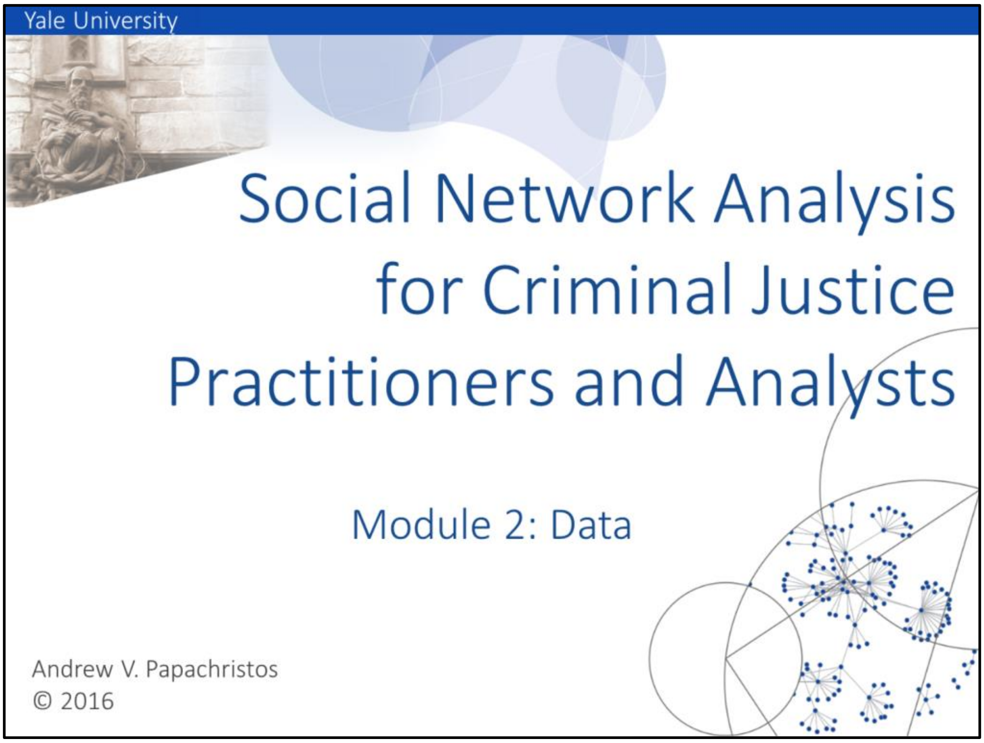
Relational data are available in more places than we realize. The trick is taking the time to turn these relationships into a dataset for social network analysis. (Blue Marble 2001-2002 is public domain available at: http://commons.wikimedia.org/wiki/File:BlueMarble-2001-2002.jpg#/media/File:BlueMarble-2001-2002.jpg)

REFERENCES

Lazega, Emmanuel and Philippa E. Pattison. 1999. "Multiplexity, Generalized Exchange and Cooperation in Organizations: A Case Study." Social Networks 21:67-90.

National Network for Safe Communities. 2013. Group Violence Intervention: An Implementation Guide. Washington DC: U.S. Department of Justice, Office of Community Oriented Policing Services. Retrieved July 15, 2015 (http://nnscommunities.org/uploads/GVI_Guide.pdf).

Sierra-Arevalo, Michael and Andrew V. Papachristos. 2015. "Applying Group Audits to Problem Oriented Policing in New Haven." In Preventing Crime with Network Analysis, edited by A. Malm and G. Bichler. Boulder, CO: Lynne Reinner.



Module 2: Data introduced participants to the logic and organization of social network data. Module 2: Data included sections on directed and undirected ties, the data structure of sociomatrices, the data structure of edgelists, the structure of attribute lists, the structure of two-mode data, how to convert two-mode networks into one-mode projections, and ideas for where to find relational data. Participants can review the different network data files in the Module 2: Data Lab and learn how to make data files compatible for analysis in NAVCAP or Rstudio.