What is a Social Network?

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

Keywords
Social Networks; Network Analysis; Undergraduate; Introduction

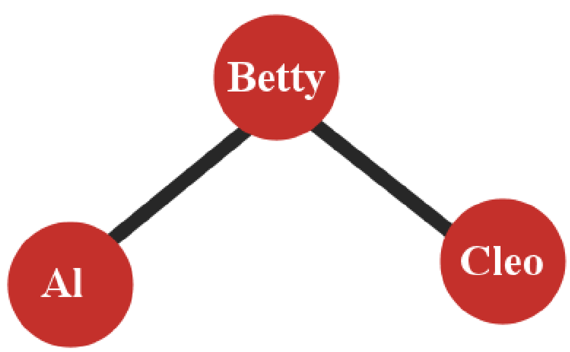
1. INTRODUCTION: WHY SOCIAL NETWORKS?

Sociology, the study of society, can be complicated. If you look at an introductory sociology textbook, you'll find references to hundreds of terms spanning dozens of theories in multiple paradigms. The sprawling complexity of sociology can be difficult to master.

The social network offers a simpler path to understanding society, so simple that the term "social network" can be defined in six words: *a structure of ties between nodes*. In turn, *node* can be defined as an *entity that can form relations with other entities*, *tie* can be defined as *the existence of a relation between nodes*, and *structure* can be defined as the *pattern of nodes and ties* in a network. *Nodes* and *ties* in a *structure* are all that is required to begin a social network analysis.

Another reason to study social networks is that there are clear, well-developed rules for describing them. Social networks can be depicted in many ways, but the most common methods are to draw *graphs* or to entered data into *matrices*. Consider, for instance, the social network graph depicted below in which circles labeled "Al," "Betty" and "Cleo" represent nodes. The two lines in the graph, between Al and Betty and between Betty and Cleo, represent ties. The relation represented in these ties is friendship: Al is friends with Betty and Betty is friends with Cleo. There is no line between Al and Cleo, meaning Al and Cleo are not tied. They are not friends. The separation between Al and Cleo creates a pattern, a network structure.

Social Network (Relation: Friendship)



Graph theory and matrix theory are branches of mathematics with already-identified techniques for describing the structures to be found in social networks. The mathematical techniques of network analysis are often surprisingly easy (counting and simple arithmetic usually suffice). When the math is not so simple, computer programs exist to do the heavy lifting.

A final reason to study social networks is substantive: society is full of all sorts of relations between all sorts of nodes: circles of friends, rumor mills, conduits of disease in communities, routes of trade between nations, terrorist networks, and favor-trading in politics are just a few examples of social networks in real life that are interesting and consequential. As our society moves ever more strongly into the digital age, and as the online platforms called "social networks" more strongly shape our friendships, our dating, and our careers, the tools of social network analysis become ever more important for those who wish to understand and master them.

2. BUILDING A NETWORK

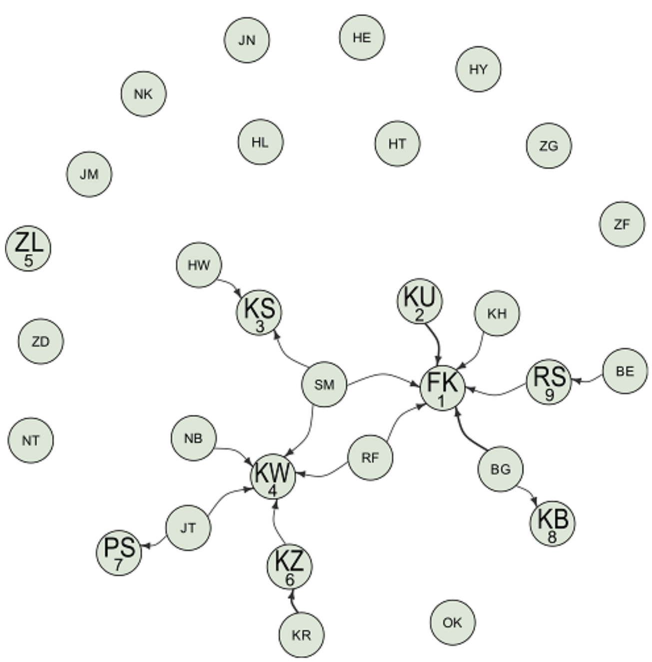
Nodes and *ties* are where social network analysis begin, so the beginning of constructing a network is clearly describing what qualifies as a node and what qualifies as a tie in your network.

2.1 Boundaries: What are the Nodes in the Network?

In a world with billions of people, hundreds of nations, and tens of thousands of corporations, it's impossible to study everyone and everything. The first step of building a social network is therefore to *specify the network's boundary.* Just as the boundary of a nation is meant to clearly mark off the land that belongs to the nation from the land that doesn't, so the boundary of a network clearly describes which nodes belong in the network and which nodes don't.

The first part of the boundary is to describe what kind of entity makes up a node. Perhaps because we are human ourselves, we often think primarily of our nodes as individual human beings, but there is no reason that nodes must be humans. Social network analysis has been used to study systems in which nodes are computer systems (Dasgupta and Biswas 2011), protein interactions (Xia et al. 2015), corporations (Mizruchi 1996), countries (Kim and Shin 2002), chimpanzees (Hobaiter et al. 2014), and wild crows (Rutz 2012).

Nearly every rule in social network analysis has been bent or broken at some point, but a good starting rule for building a social network is to choose a single kind of entity to be a node in one's social network. The more specific one can be, the better: "elementary school students" is preferable to "people" because it limits the number of observations to be made and because it makes clear what kind of social environment is to be studied.



Social interaction in a troop of chimpanzees. Source: Hobaiter et al. 2014.

The second part of the boundary to describe is the setting within which nodes will be observed. To continue the example from above, one might be interested broadly in "elementary school students." It is technically impossible, however, to observe the social connections between *all* elementary school students. There are just too many elementary school students in the world. When specifying your boundary, be sure to name the actual space or virtual place that you intend to observe. For example, you might say that your network will consist of "all elementary school students in the second-grade classrooms in East Rutherford, New Jersey." Be careful when choosing a space to observe: in network analysis, it is important to observe ALL nodes that sit within the network boundary you specify. Can you really observe *all* second graders in East Rutherford? If not, it might be time to draw the boundary a bit tighter.

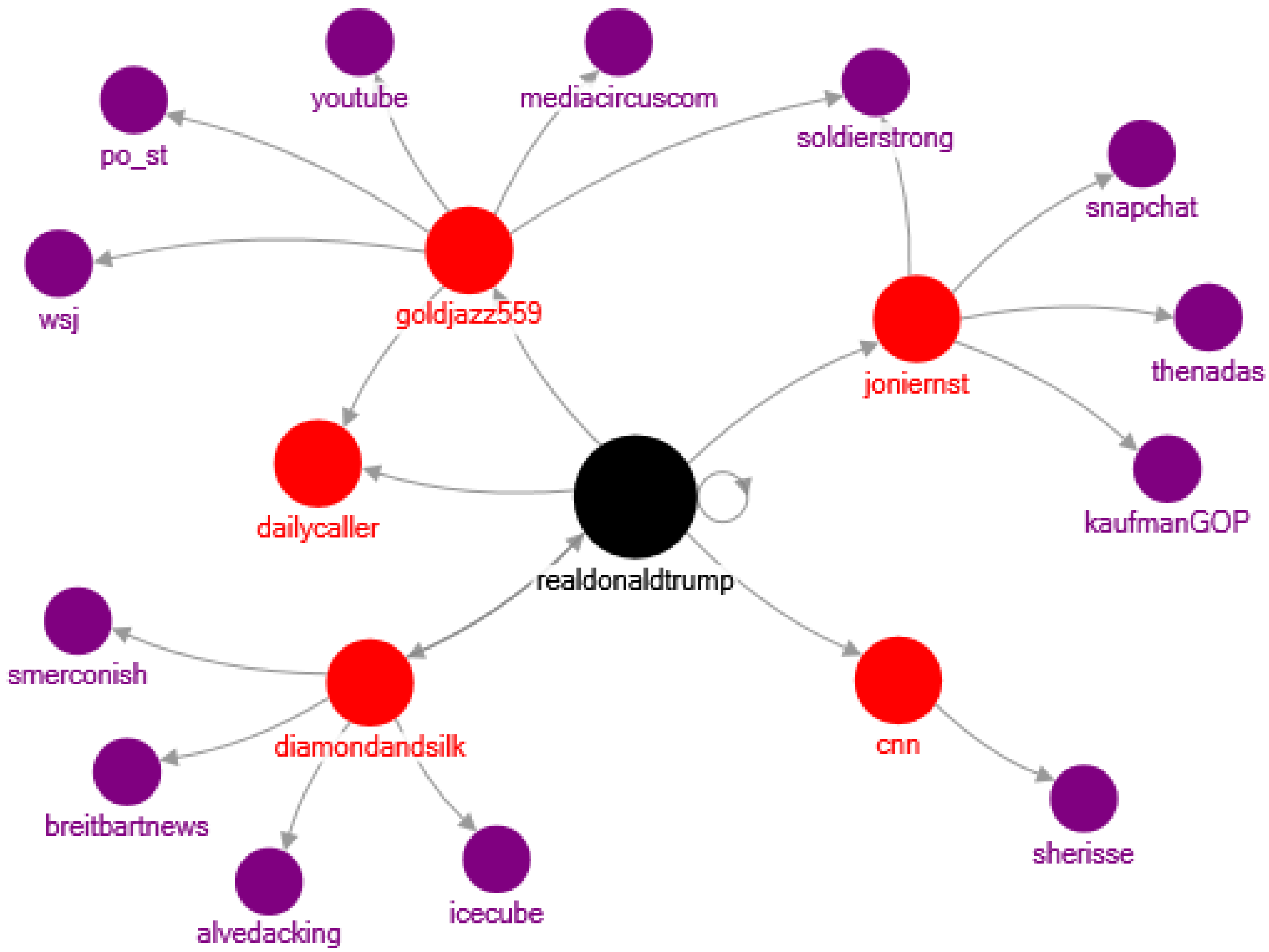
2.2 Relations: What Connects Nodes?

If every tie represents a connection, ask yourself what *kind* of connection you wish to study in your network. That kind of connection is called a *relation*. Just as there are many kinds of nodes that could possibly be involved in social networks, so there are many kinds of relations that could possibly connect those nodes. But just as most social networks contain just one kind of node, so most social networks also contain ties expressing just one kind of relation. The next step in building a social network is therefore to decide what kind of relation you wish to depict between nodes.

The difference between a relation and ties is subtle but important. A relation is a general kind of connection that might or might not occur; it's a potential, hypothetical idea. A tie, on the other hand, is the actual occurrence of an actual connection. Sometimes actual ties occur between pairs of nodes, and sometimes they don't.

Before we proceed, let's clarify what we know so far through an example. Consider the social network displayed in graph form on the top of page 3, drawn from the social media platform Twitter[1]. This network was obtained from Twitter within a specific network boundary: <u>the set of nodes are those Twitter accounts that were mentioned on August 27, 2016 by Twitter users who themselves in turn were referred to by Donald Trump's Twitter account on the same day</u>. That's a two-step boundary standard; let's follow the color code in the graph to help more make the boundary

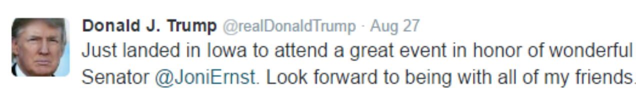
[1] If you're unfamiliar with Twitter, learn more at https://support.twitter.com/articles/13920.



A Twitter-Based Social Network: @Mentions by Twitter users who were themselves mentioned by @realdonaldtrump, August 27, 2016

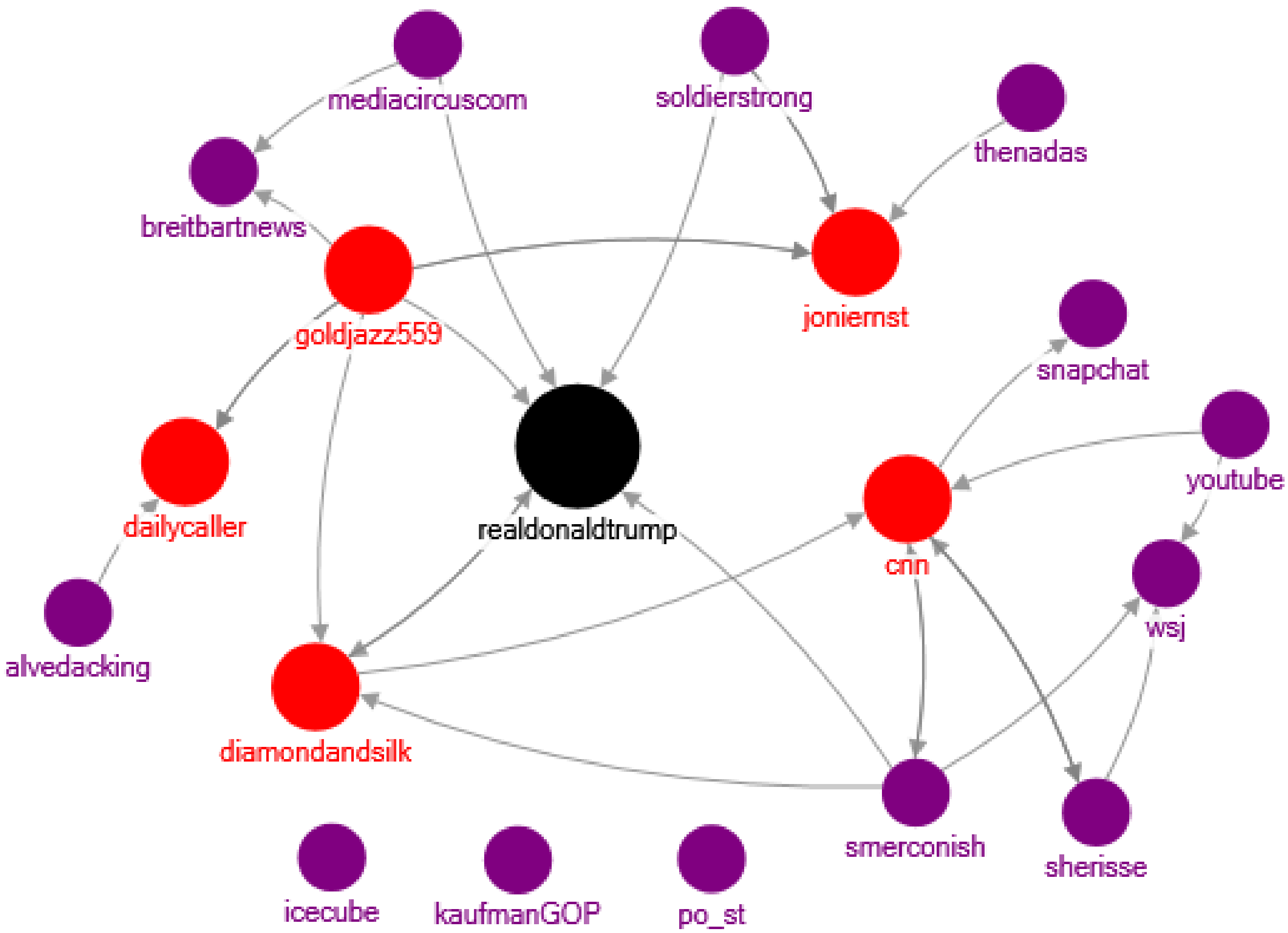
more clear. Let's start with Donald Trump's Twitter account (@realdonaldtrump), located in the center of the graph and colored black. Then, let's move outward to the other Twitter accounts that @realdonaldtrump mentioned on August 27, 2016; these accounts are in red -- @goldjazz559, @dailycaller, @diamondandsilk, @cnn, and @joniernst – and also @realdonaldtrump himself (the circular tie to the right of the @realdonaldtrump node indicates that he mentioned himself). The boundary for inclusion in the network consists of all the Twitter accounts that these red- and black-colored accounts themselves mentioned on August 27, 2016. That's why we, finally, include a third set of nodes, colored purple: they weren't directly mentioned by the account @realdonaldtrump, but they were mentioned by accounts that @realdonaldtrump mentioned.

If this is the boundary, what is the relation of the network? A "mention" on August 27. In Twitter, users can "mention" other users by adding the character "@" to the front of someone else's username. In this example from August 27 2016, for instance, Donald Trump uses his Twitter account to mention the Twitter account of Senator Joni Ernst of Iowa:



Because Donald Trump (@realdonaldtrump) mentioned Joni Ernst (@joniernst) on August 27, a tie exists between the two accounts. Donald Trump did not mention the account @thenadas on August 27, so there is no tie between @realdonaldtrump and @thenadas. However, Joni Ernst *did* mention @thenadas, so a tie between @joniernst and @thenadas is shown in the graph.

To review, a node is an entity that can be connected to other nodes. A boundary describes which nodes belong



Another Twitter-Based Social Network: Who Follows Whom?

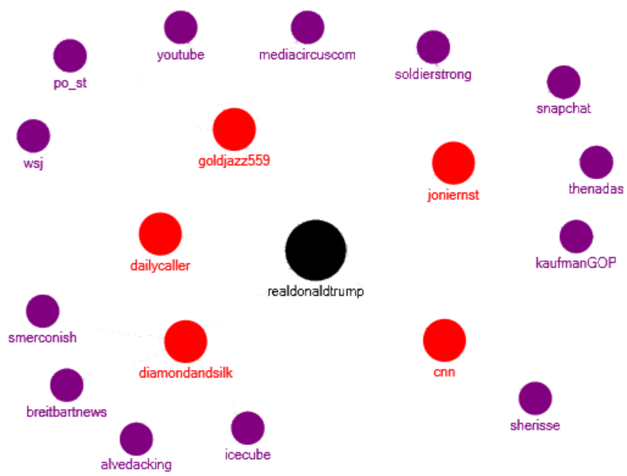
in a network. A relation is a kind of potential connection, and there's only one relation per network. A tie occurs between a pair of nodes when that pair actually is connected in the way described by the network's relation, and a network can have many ties.

At this point, you may have an objection: *why must* a network only display ties for *one* kind of relation? I can share two answers with you. The first reason is stylistic: the human eye can only interpret so much complexity. Although occasionally you can find a network graph out there in the world that displays ties for two different kinds of relation, most often network graphs avoid this complication because the graphs get messy. The second reason has to do with the other form networks take, the form in which network data is stored: a *matrix*. We'll discuss matrices in detail below; for now, it's enough to say that a matrix only has enough room to hold information about one kind of relationship.

These may be technically good reasons, but you may still feel dissatisfied because in real life the same set of nodes can and often do share more than one kind of relation between them. A wife and husband, for example, may be married (relation type #1), be friends (relation type #2), discuss the news with one another (relation type #3), collaborate as parents (relation type #4), and work together at a family business (relation type #5). How can these multiple kinds of relations, resulting in multiple ties between the same pair of nodes, be recognized if there's a rule forbidding them to appear in the same network?

Fortunately, there is an answer to the dilemma: create a different network for each different type of relation. These different networks will contain the same nodes, but they will have different sets of ties between them, reflecting the different kind of relation being highlighted. To continue the current example, this page features another Twitter-based network, using the same boundary as the first and containing the same nodes, but featuring a different relation: *following*. On Twitter, one account can *follow* another, which means that the posts made by the followed account appear in the newsfeed of the following account. Who follows whom in this network? As you can see in the graph on the previous page, resulting structure of following ties looks quite different than that the structure of mentioning ties.

We might consider yet another relation: marriage. Which of the Twitter account holders within this boundary are married to one another? Although the relation has potential, the resulting social network of actual ties is entirely empty.



Same Nodes, Another Network, for the Relation "Married to"

We refer to this circumstance – in which the same set of nodes are connected by multiple, different relations, leading to different patterns of ties, each relation described by a different network – as the *multiplexity* of networks. Far from being a worrisome bug of social network analysis, it is actually one of social networks' most interesting properties.

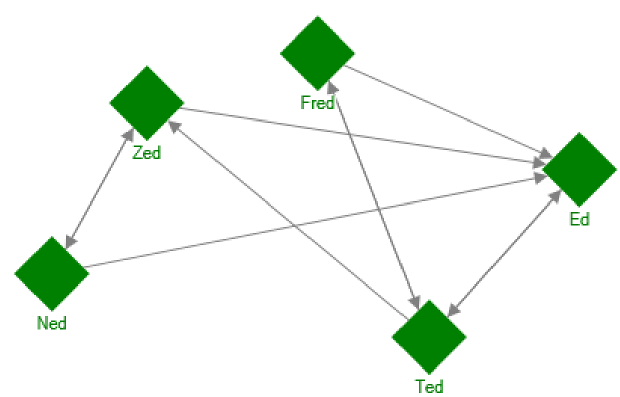
3. DEPICTING NETWORKS

3.1 Networks as Graphs

There are many ways to depict social networks, but the two most common are the *graph* and the *matrix*. You've already encountered network graphs in this chapter. A graph is a drawing of a social network in which nodes are depicted as shapes and ties are depicted as lines.

Did you notice that two of the three network graphs centered around @realdonaldtrump contained lines with arrowheads? Those arrows point ties in a particular direction. Why? Because the relation in two of the three Twitter networks contains direction. Think about mentioning. It is possible in social media for one user to mention another but not to be mentioned in return. The same goes (on Twitter, at least) for following: The account @smerconish (held by talk host Michael Smerconish) follows Donald Trump's @realdonaldtrump account, but @realdonaldtrump doesn't follow @smericonish back. Sometimes the arrows in networks with direction *do* point both ways, @realdonaldtrump follows @diamondandsilk, and @diamondandsilk follows @realdonaldtrump right back. In this case, a single line is drawn with arrowheads pointing in two directions, signaling the existence of *two ties* – one line for a following tie going one direction, and another line for a following tie going the other direction.

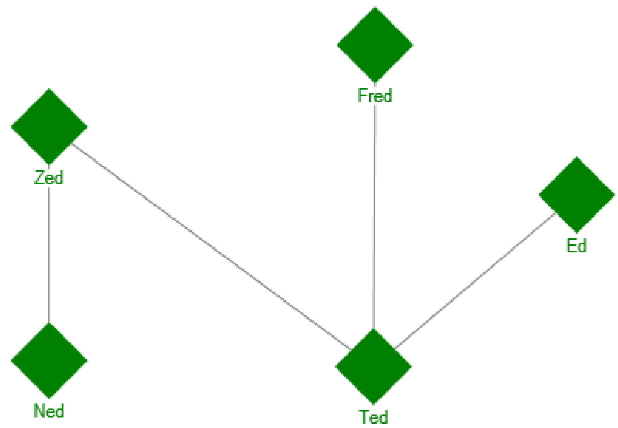
When a relation indicates direction, all ties in the associated network graph should have arrows indicating direction too. Let's consider another example of a network graph for a relation with direction – *known as a "digraph"* (short for directed graph). The network digraph you see below describes the the practice by which nodes Ed, Fred, Ted, Zed and Ned write letters to one another. "Writing letters to" is a directional relation, because it is possible for someone to write a letter and not receive one in return. Indeed, in this network Zed writes letters to Ed, but Ed does not write back to Zed. Ned and Zed, on the other hand, each write letters to one another, resulting in a double-headed arrow signifying two directional ties.



Directed Graph, or "Digraph."

Relation: ___ Writing Letters To ___

Ed, Fred, Ted, Zed and Ned are involved in a multiplexity of networks. The network graph you see below does *not contain arrows*. This is because the relation of the graph described below, "talking on the phone with," is by nature without direction. It takes two people to talk on the phone with one another; it's

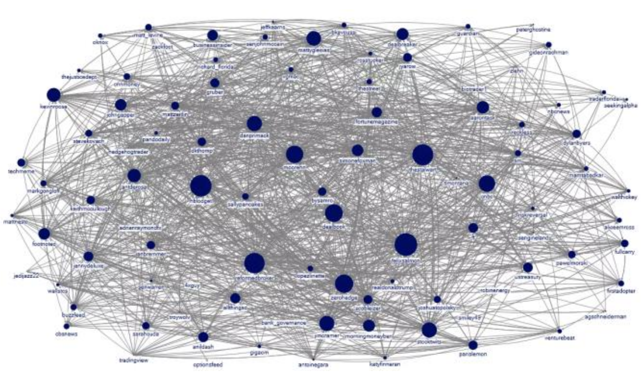


Graph. Relation: Talks on the Phone With

a joint activity. Indeed, it makes no sense to say that Fred was talking on the phone with Ted but that Ted was *not* talking on the phone with Fred. Such a directional claim makes no sense. Other kinds of social relations that must by their definition be without direction include "is married to," "is in a romantic relationship with," "plays tennis with," and "goes to the same school as." For networks depicting these kinds of relations, we draw graphs without arrows, to signify that there is no direction to the underlying relationship being represented. We simply refer to drawn networks without direction as "graphs" (or, if we want to be perfectly clear, "undirected graphs").

3.2 Networks as Matrices

Graphs are visual objects, but social network information does not need to be visually represented. Graphs are appealing to the human eye and, if they are thoughtfully presented in appropriate cases, can even highlight important and consequential patterns. However, when the number of nodes and ties in a network increases beyond a certain limit, our brains have a hard time making a sense of the mess.



In 2013, financial journalist Len Costa shared a network graph of the connections on Twitter between accounts deemed by billionaire Michael Bloomberg to be "very important" in sharing information on financial markets (Costa 2013). The graph in blue on this page is Costa's graph of the network of "following" relationships on Twitter for 100 of these VIP accounts.

What do you think of that graph? I can't make head nor tail of it. There simply too many overlapping lines in the graph for me to follow. The farthest I can get is, "wow, there sure are a lot of ties there." Any patterns that might be in the network are lost in the fog of sheer volume. Keep in mind that this messiness occurs in a network of only a hundred nodes. Think about most of the communities in which you live – even small universities and small towns tend to have thousands of inhabitants, not a mere hundred. Clearly, if we want to understand networks for even small communities, we need another way of working with them.

This is where matrices come in. They're not immediately as appealing as network graphs, but what they lack in beauty they gain in precision and power. Consider the digraph and undirected graph for Ed, Fred, Ted, Zed and Ned that we just reviewed. The ties can be represented in an *adjacency matrix*. "Adjacency" simply means "tied to," and an *adjacency matrix* is a table with ties. In an adjacency matrix, each node is assigned both a column and a row, and each combination of column and row is called a cell. A cell in an adjacency matrix contains a 0 if its combination of row node and column node do *not have a* tie. Where there is a tie, the corresponding cell is given a number greater than zero – usually just a 1 to indicate a tie does exist (although in future weeks we'll talk about the use of other numbers for other purposes).

Ed, Fred, Ted, Zed and Ned's adjacency matrices look like this:

Adjacency Matrix: "Talks on the Phone With"

	Ed	Ned	Ted	Fred	Zed
Ed	--	0	1	0	0
Ned	0	--	0	0	1
Ted	1	0	--	1	1
Fred	0	0	1	--	0
Zed	0	1	1	0	--

Adjacency Matrix: "Writes Letters To"

	Ed	Ned	Ted	Fred	Zed
Ed	--	0	1	0	0
Ned	1	--	0	0	1
Ted	1	0	--	1	1
Fred	1	0	1	--	0
Zed	1	1	0	0	--

The red-outlined cells are row-cell combinations in which nodes refer to themselves. Those cells are called the "diagonal," and often they are meaningless (can you talk to yourself on the phone?). In such cases they're left alone.

Looking at the remaining cells, it's always been the convention that for relations with direction (such as "writing letters to"), the sender is placed in rows and the receiver is placed in columns. Interpreting the second adjacency matrix, therefore, we can see that Fred writes letters to Ed (the cell Fred->Ed has a 1 in it), but that Ed does not write letters to Fred (the cell Ed->Fred has a 0 in it).

Looking at these matrices, you should be able to see why each network involves only one kind of relation: there is only one space for information about the relation between two nodes to be entered in a matrix, so it's literally impossible to discuss two relations in one matrix. Just as two different relations should lead to two different network *graphs* containing the same nodes but different ties, so two different relations should lead to two different *matrices* containing the same nodes but different ties.

A computer can't easily read and make sense of a network graph, which is all right, because after all graphs are drawn for humans. On the other hand, computers are experts at reading matrices. This becomes especially handy when you have a large network with thousands, tens of thousands or even millions of nodes on your hands. No human can interpret such a network (just look back at that messy Bloomberg graph of a page ago). But when expressed in matrix form, computers find matrices a cinch to understand.

Better yet, if you know how to ask a computer nicely, you can get it to tell you very interesting things about a network in matrix form. Add all the values of a row up, for instance, and you'll know how may ties a node sends out (for example, Ed has written 1 letter). Add all the values in a column together, and you'll know how many ties a node has sent to them (Ed has received 4 letters). Such counts are just the beginning: with the right network techniques applied to matrices, inquisitive souls can find out who's in the middle of a gossip network, who's likely to catch a deadly disease next, and who is most likely to hold a seat of decisive power.

4. LOOKING AHEAD: ASKING QUESTIONS ABOUT NETWORKS

This chapter is only a brief introduction to the essence of social networks, and yet it contains all the necessary elements upon which a fuller understanding of social networks can be built. The remaining chapters of this book will be devoted to four projects.

Project 1: Depicting Social Networks

- ⇨ How are social networks depicted, and how does that differ from the way that most social science depicts information?
  - Individual-level data structure
  - Social network data structure
  - Social network data presentation
    - Graph
    - Matrix
    - Edge List
    - Adjacency List

Project 2: Technologies for Social Network Analysis

- ⇨ What are the available, most useful, and most accessible technologies for carrying out the analysis of social networks?
  - Pen and paper (surprise!)
  - Computer programs
    - R
      - free
      - command-based
      - extensive power
    - NodeXL
      - paid
      - menu-based
      - limited power
    - UCINET
      - paid
      - menu/cmd-based
      - extensive power

Project 3: How can information about a network be obtained?

- ⇨ Node, dyad, group, and network information
- ⇨ Expressing information about individuals in a social network
- ⇨ Finding social facts (homophily, density, class size, group size, social distance) in social networks

Project 4: How do social networks affect social outcomes we care about?

- ⇨ Politics
- ⇨ Online social networking
- ⇨ Disease
- ⇨ The surveillance society

This book is not intended for those who are already expert in the basics of social network analysis. Rather, it is meant for those would like to quickly learn the basics of social networks, without fuss, so they can quickly begin to make practical use of social network analysis.

There are many mathematically-intense, computationally-difficult applications of social network analysis. This book ignores them. They are not the first aspects of network analysis that a person should learn, and a great deal of insight into the social world can be obtained without them.

By the end of this brief and simply-worded book, you will not be a world-renowned expert in social network analysis, but you probably don't need to be. My more modest hope is that you will be confident in your understanding of essential social network principles and capable of performing foundational social network analysis. In a world increasingly organized around explicit network principles, network literacy will be necessary if you'd like to have a place at the table. Your place is reserved; take a seat and enjoy the ride.

5. GLOSSARY

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*

boundary: standard by which entities are considered to be either members of a network or excluded from a network

diagonal: set of cells in a matrix that refer to a node's relationship with itself. Often, the diagonal is meaningless.

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

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

multiplexity: a number of different kinds of relation, reflected in different sets of ties in different networks, connecting the same nodes to one another in different patterns

node: entity that can relate in some way to other entities

relation: manner or kind of connection in a network

social network: structure of ties between nodes

structure: arrangement or pattern of elements

tie: instance of a relation existing between two particular nodes

6. REFERENCES

Costa, Len. 2013. "Inside Bloomberg's Twitter A-List (Well, at Least a Fraction of It)." *Enterprising Investor* April 15: https://blogs.cfainstitute.org/investor/2013/04/15/inside-bloombergs-twitter-a-list-well-at-least-a-fraction-of-it/.

Dasgupta, Mou, and G. P. Biswas. 2011. "Relieving Critical Nodes and Designing of Fault Tolerant Data Networks." *International Journal of Engineering, Science and Technology* 3(1): 41-51.

Hobaiter, Catherine, Timothée Poisot, Klaus Zuberbühler, William Hoppitt, and Thibaud Gruber. 2014. "Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees." *PLoS Biology* 12(9): e1001960. DOI:

Jianguo Xia, Erin E. Gill, and Robert E.W. Hancock. 2015. "NetworkAnalyst for Statistical, Visual and Network-based Approaches for Meta-analysis of Expression Data." *Nature Protocols* 10: 823–844. DOI: http://dx.doi.org/10.1038/nprot.2015.052

Kim, Sangmoon, and Eui-Hang Shin. 2002. "A Longitudinal Analysis of Globalization and

Regionalization in International Trade: A Social Network Approach." *Social Forces* 81(2): 445-468. DOI: http://dx.doi.org/10.1353/sof.2003.0014

Mizruchi, Mark. 1996. "What do Interlocks do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates." *Annual Review of Sociology* 22: 271-298. DOI: http://dx.doi.org/10.1146/annurev.soc.22.1.271

Rutz, Christian, Zackory T. Burns, Richard James, Stefanie M.H. Ismar, John Burt, Brian Otis, Jayson Bowen and James J.H. St. Clair. 2012. "Automated Mapping of Social Networks in Wild Birds." *Current Biology* 22(17): R669-R671. DOI: http://dx.doi.org/10.1016/j.cub.2012.06.037