Social Network Analysis

MATTHEW DENNY

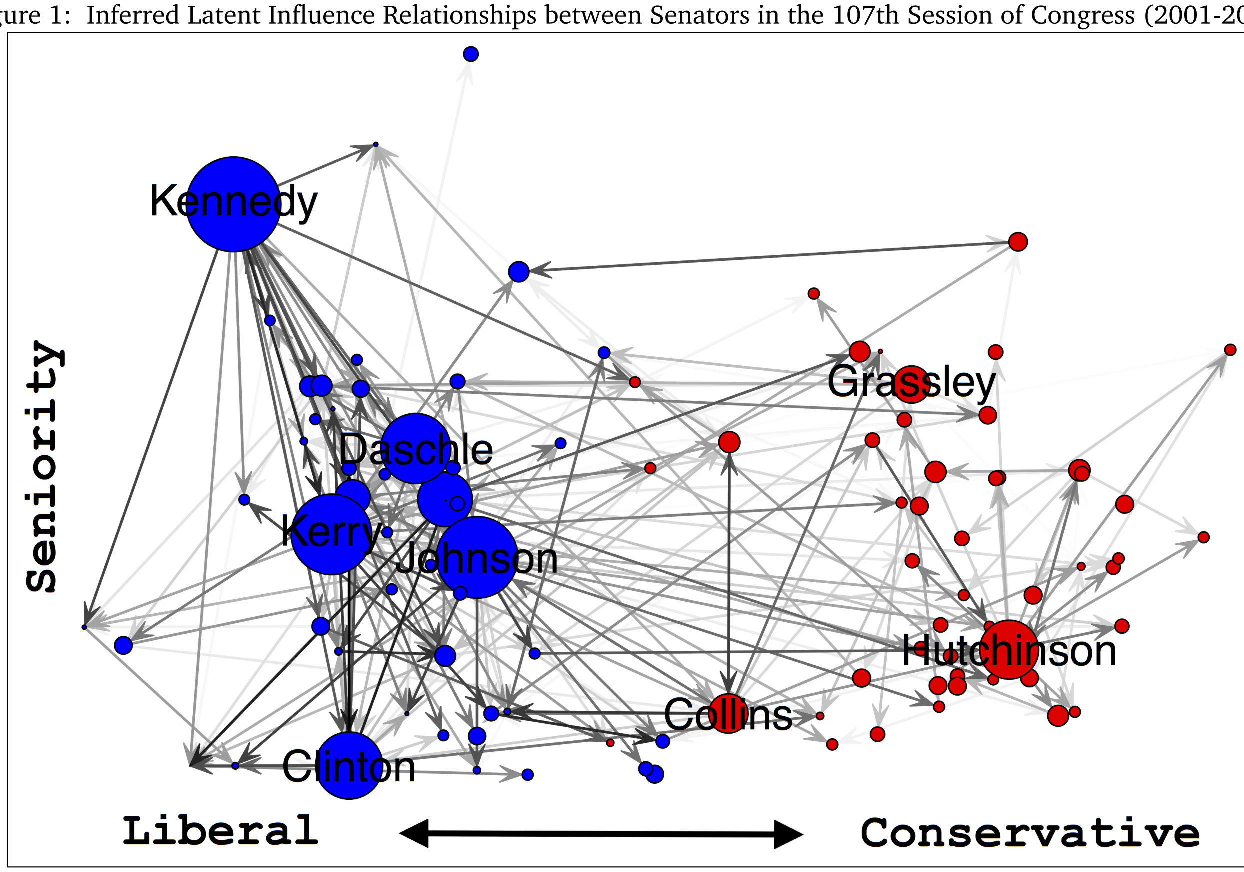
FRIDAY 26TH SEPTEMBER, 2014

INSTITUTE FOR SOCIAL SCIENCE RESEARCH
UNIVERSITY OF MASSACHUSETTS AMHERST

Welcome to this tutorial introducing Social Network Theory and Social Network Analysis (**SNA**) more generally. The study of networks is not restricted to sociology or even the social sciences; the relationships between entities can be gainfully studied in neuroscience (Neves et al., 2008), physics (Newman, 2003), political science (Fowler, 2006; Cranmer and Desmarais, 2011; Kirkland, 2011), economics (D'Exelle and Holvoet, 2011; Sundararajan et al., 2012; Jackson and López-Pintado, 2013), anthropology (Zachary, 1977), management science (Levin and Cross, 2004; Aral and Alstyne, 2011; Aral et al., 2013), statistics, (Hoff et al., 2002; Raftery et al., 2012), computer science (Gomez Rodriguez et al., 2010; Krafft et al., 2012), psychology (Moreno, 1934; Aral and Walker, 2012), engineering (Lubin et al., 2013) and of course sociology (Marwell et al., 1988; Stackman and Pinder, 1999; McPherson et al., 2001; Watts et al., 2002).

This tutorial will introduce a number of foundational concepts in network theory and analysis, with a focus on how a network perspective might be useful to the study of a wide range of phenomena. We will focus on general definitions and properties of individuals, their relationships and networks as a whole. Throughout this tutorial I will include citations to relevant articles and resources, please use them as a starting point for further exploration and realize that network theory is very often mis-applied, so take the time to really understand these concepts before applying them! Take the time to read as much as you can and if possible pursue further coursework in network analysis; this is just the beginning – you will have a lot of fun.

Figure 1: Inferred Latent Influence Relationships between Senators in the 107th Session of Congress (2001-2003)



Network analysis is still a growing field with a great deal of opportunity for new and transformative contributions, but its history goes back atleast 80 years (Moreno, 1934). It is important to build on this existing body of theory and empirical results before striking out on your own, even if you are working in an area where network analysis is just starting to catch on.

While social network theory can be readily applied in theoretical research and qualitative empirical studies, there is a general emphasis on the use of software to analyze and visualize network data once they have been collected. There are a number of different software packages available for this purpose, but two **R** packages (Statnet and iGraph) have become perhaps the most flexible and powerful tools for performing network analysis. As you considder getting into network analysis, please consider enrolling in a class or workshops that uses **R** as its primary computing language. If you put time and effort into gaining proficiency in **R** for data management it will pay huge dividends when you look to start doing more advanced network analysis using **R**. You can find out more about R and access a number of instructional materials at the following websites:

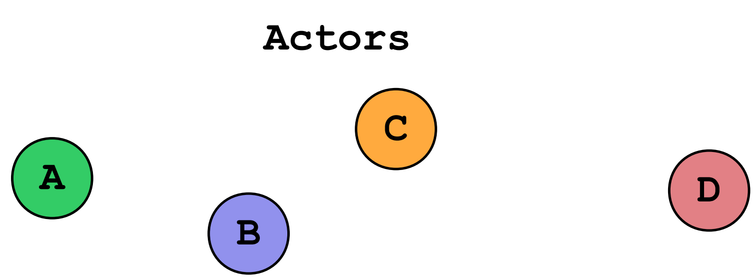
1. **R** is a free and open source statistical computing language with a vibrant community of contributors who are constantly updating its functionality through the creation of user defined add-on packages. To get started, check out this website: http://www.r-project.org/.
2. **Quick-R** is my favorite resources for learning basic **R** commands and is also available in book form. Check out the website: http://www.statmethods.net/
3. **RStudio** is an integrated development environment for **R** that provides a lot of useful features for beginner **R** users and powerful tools for advanced users as well (I primarily use it). Check out their website and download the program for free here: https://www.rstudio.com/

The rest of this tutorial will not focus on computing or statistic methods for network analysis so take what is provided above as a starting point. The rest of this tutorial will focus on an intuitive and visual introduction to social network theory.

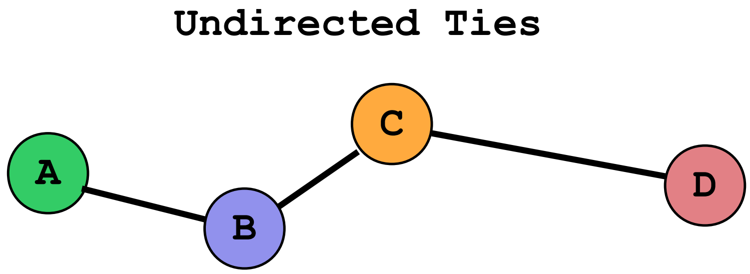
1 Some Definitions

Before we can get started we need to define some terminology so we can use a consistent language when talking about social networks:

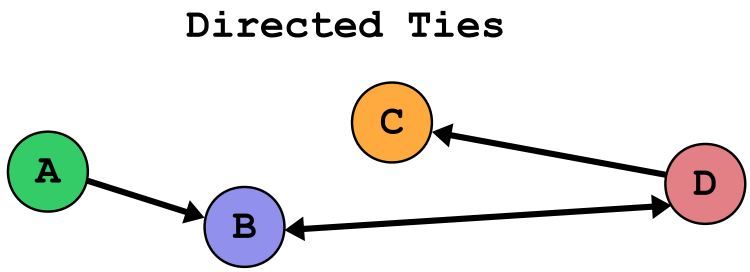
1. **Actor:** also called a **node** or a **vertex**, referrers to an individual hat can have relationships with other individuals and in this case, an individual or group of individuals we are choosing to study.



2. **Tie**: also called a **relation** or **edge**, describes a particular, well specified, relationship between two **Actors**. This could refer to a relationship like "went to the same school" or "likes potato chips" or something like "likes" or "trades with". Ties can be **un-directed** (like went to the same school), when the relationship means the same thing to both actors:

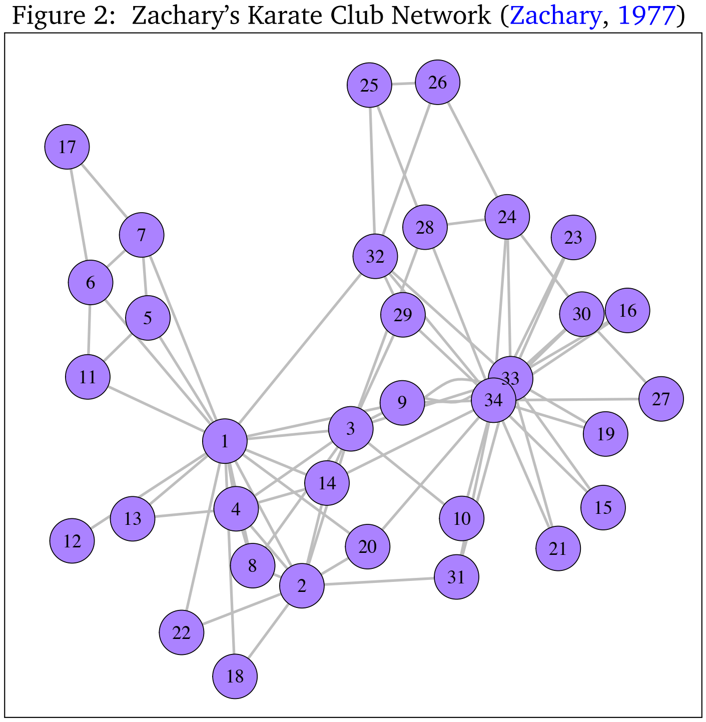


Ties can also be directed (such as “looks up to”) and either one directional or bidirectional:

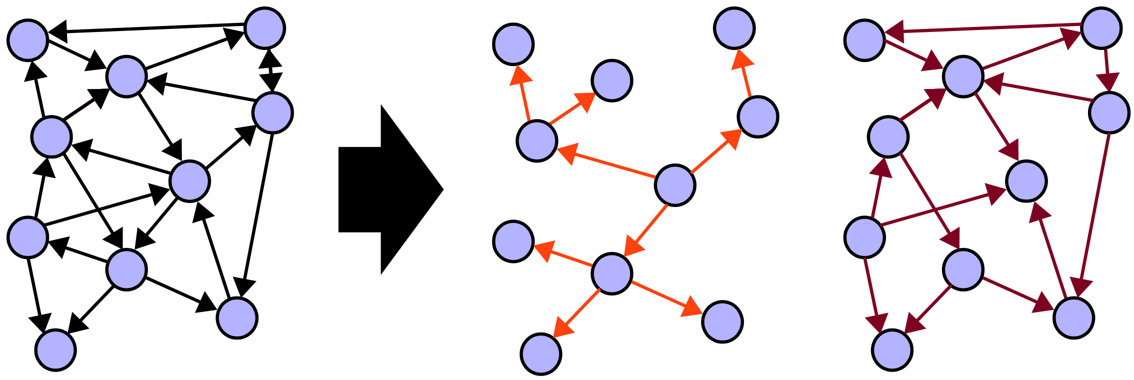


3. **Network**: also called a **Graph**, particularly in the physics and CS literature, referres to a collection of **Actors** and the **Ties** between them. Figure 2depicts a set of unidirected friendship relationships between members of a Karate club.

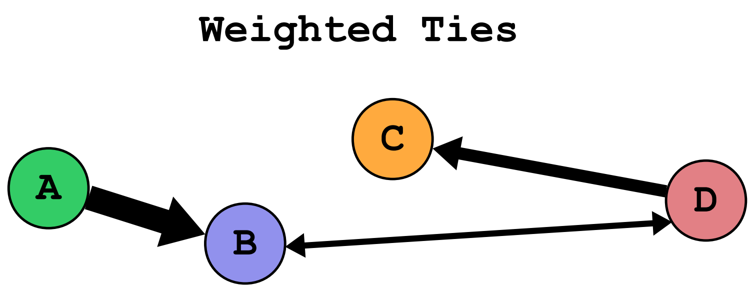
Figure 2: Zachary’s Karate Club Network (Zachary, 1977)



4. **Multiplex networks**: are networks where more than one kind of tie is present. For example, if we were to collect information about several different kinds of relationships between bank managers (goes to for advice, is friends with, works for, etc.) we essentially end up with a network containing multiple tie types between actors.

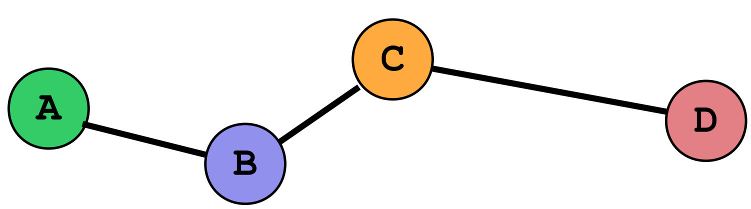


5. **Weighted Ties:** just as networks can contain multiple different kinds of edges between actors, they can also contain relationships of varying strength. For example **A** might like **B** a whole lot, but **B** and **C** only like each other moderately.



6. **Group:** A group in a network is just a subset of the actors which share some characteristic in common. If we were to look at an organizational network, one group could be made up of all actors that work in the human resources department. The definition of groups as commonality on some salient trait allows us to examine a number of network hypotheses and defined useful measures that are conditional on knowing the group membership of actors. For example we might want to test a hypothesis about the number of friendship ties between workers at a company who are part of different departments versus those in the same departments.

7. **Geodesic Distance:** is defined as the least number of connections (ties) that must be traversed to get between any two nodes. For example, in the network depicted below, the geodesic distance between actor A and actor D is 3, while the distance between actor B and C is only 1.



Social Network Data

There are two main kinds of social network data: **edge lists and sociomatricies**. Each of these data formats has its own advantages and weaknesses, mainly having to do with a trade off between ease of entering and storing the data and ease of using the data for analysis.

1. A **Sociomatrix (also known as an Adjacency Matrix):** is a way of representing directed or undirected ties between actors using a numerical matrix. There is one column for each actor and one

row for each actor. In general, the diagonal elements of this matrix (eg. second row, second column) are always equal, signaling that actors do not tie to themselves. To specify which entry in the matrix we are talking about we always use the same convention: **[row *i*, column *j*]** so that if we were to say the [3,5] entry in the sociomatrix we would be talking about the third row and fifth column.

Each row in the sociomatrix represents the ties that Actor *i* sends to all other actors (j's). As we notice in figure 3, manager one sends a directed friendship tie to manager two, as indicated by the value 1 in the **[1,2]** entry of the sociomatrix. The upside of taking this approach to storing data about a network is that it naturall encodes the fact that some actors may not send or receive any ties (something we call being a network *isolate*) and the format is very ready for many statistical analyses. The downside to using this data format is that it can take up a lot of space and be difficult to enter data into by hand.

Figure 3: Sociomatrix of Directed Network of Friendship Ties Between Managers (Krackhardt, 1987)

Manager	
1	0 1 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0
2	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1
3	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0
4	1 1 0 0 0 0 0 1 0 0 0 1 0 0 0 1 1 0 0 0 0
5	0 1 0 0 0 0 0 0 1 0 1 0 0 1 0 0 1 0 1 0 1
6	0 1 0 0 0 0 1 0 1 0 0 1 0 0 0 0 1 0 0 0 1
7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8	0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
9	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
10	0 0 1 0 1 0 0 1 1 0 0 1 0 0 0 1 0 0 0 1 0
11	1 1 1 1 1 0 0 1 1 0 0 1 1 0 1 0 1 1 1 0 0
12	1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1
13	0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
14	0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0
15	1 0 1 0 1 1 0 0 1 0 1 0 0 1 0 0 0 0 1 0 0
16	1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
17	1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 1 1
18	0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
19	1 1 1 0 1 0 0 0 0 0 1 1 0 1 1 0 0 0 0 1 0
20	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
21	0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0

2. An **Edgelist** is the other primary form of data storage for social network analysis. This only captures information about existing ties so it needs to be supplemented with knowledge of the total number of actors in the network (even if they do not have any ties). In the example edgelist in Figure 4, directed friendship ties for the network shown in Figure 2 are presented in edgelist form where the first number on each line denotes the actors sending a tie to the second actor in the row.

   This form of data entry is best for storing information about data that are collected by hand as it is very efficient to store and relatively easy to enter, but one must be careful to use a common naming system and keep track of any nodes that do not have any ties to them.

Figure 4: Zachary's Karate Club Network Edgelist Representation (Zachary, 1977)

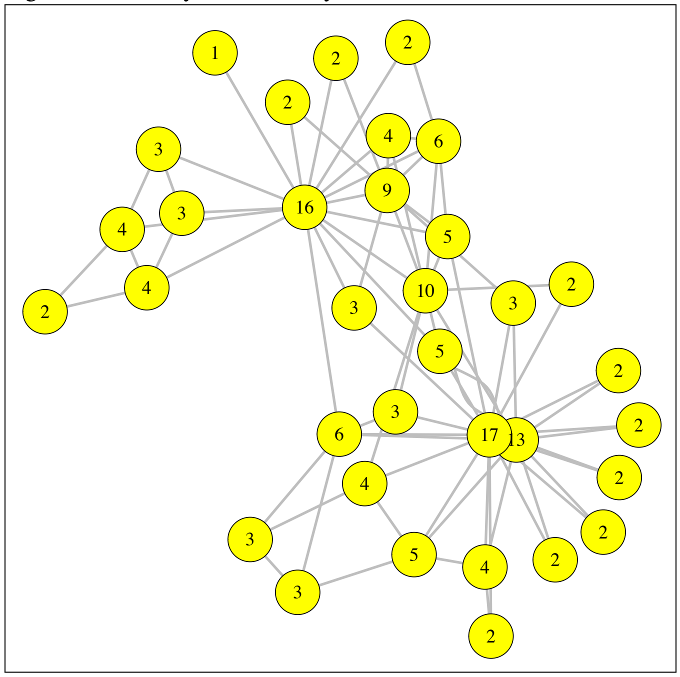
1 32	1 4	3 10	6 7	20 34	26 32
1 22	1 3	3 33	7 17	21 34	27 34
1 20	1 2	3 29	9 34	21 33	27 30
1 18	2 31	3 28	9 33	23 34	28 34
1 14	2 22	3 8	9 33	23 33	29 34
1 13	2 20	3 4	10 34	24 30	29 32
1 12	2 18	4 14	14 34	24 34	30 34
1 11	2 14	4 13	15 34	24 33	30 33
1 9	2 8	4 8	15 33	24 28	31 34
1 8	2 4	5 11	16 34	24 26	31 33
1 7	2 3	5 7	16 33	25 32	32 34
1 6	3 14	6 17	19 34	25 28	32 33
1 5	3 9	6 11	19 33	25 26	33 34

2 Properties of Nodes

Now that we have some basic terminology down, we can get into the heart of actor level properties that serve as the language for social network analysis. I am going to spend a majority of my time in this section explaining how to conceptualize social phenomena and hypotheses in a networks framework without going into too much detail on substantive theories of relational phenomena. The goal is to help you be literate enough to interface with and understand theories posed in the literature using a social networks/ relational framework.

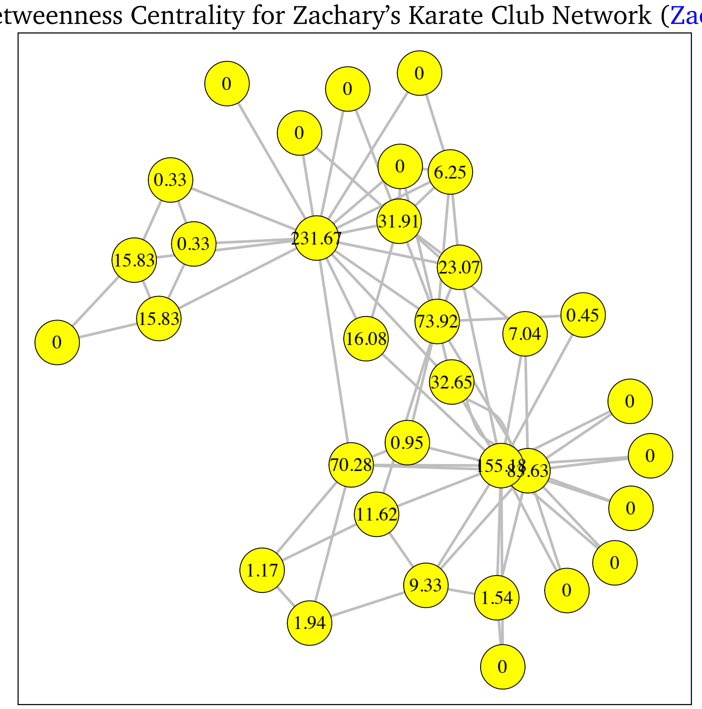
1. **Degree Centrality:** is the most basic network measure and captures the number of ties to a given actor. For **un-directed** ties this is simply a count of the number of ties for every actor. For directed networks, actors can have both **indegree** and **outdegree** centrality scores. As the name implies, centrality measures how central or well connected an actor is in a network. This theoretically signals importance or power and increased access to information or just general activity level and high degree centrality is generally considered to be an asset to an actor. Degree centrality is depicted for the Karate club network in Figure 5 where each actor is now labeled with their undirected degree centrality score.

Figure 5: Degree Centrality for Zachary's Karate Club Network (Zachary, 1977)



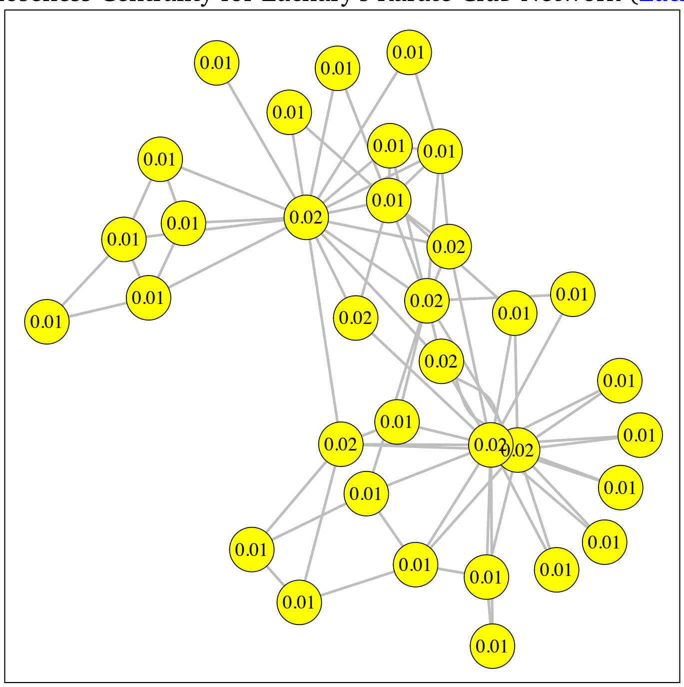
2. **Betweenness Centrality:** is roughly defined as the number of shortest paths between alters that go through a particular actor. More precisely, it is the sum of [the shortest path lengths between every set of alters where the path goes through the actor we are calculating the measure for divided by the shortest path lengths (not necessarily through the target actor) between those actors]. This intuitively measures the degree to which information or relationships have to flow through a particular actor and their relative importance as an intermediary in the network. Betweenness scores for Zachary's Karate club network are displayed in figure 6.

Figure 6: Betweenness Centrality for Zachary's Karate Club Network (Zachary, 1977)



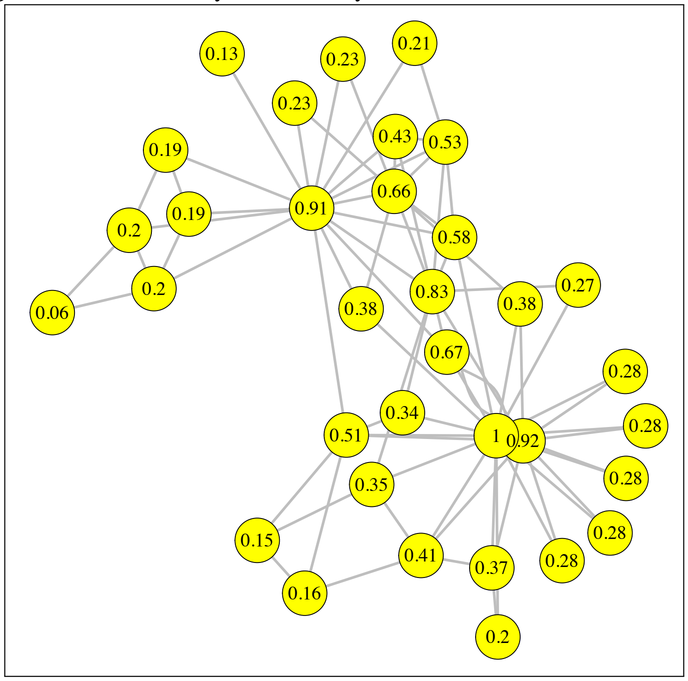
3. **Closeness centrality:** measures how many steps (ties) are required for a particular actor to access every other actor in the network. This is measured as 1 divided by the sum of geodesic distances from an actor to all alters in the network. The measure will reach its maximum for a given network size when an actor is directly connected to all others in the network and its minimum when an actor is not connected to any others. This captures the intuition that short path lengths between actors signal that they are closer to each other. Note that this measure is sensitive to network size and is decreasing in the number of actors in the network. This makes intuitive sense in many situations because it gets more difficult to maintain close relationships with all members of the network as the network grows but can also be corrected for by multiplying by the number of actors in the network. Closeness scores for Zachary's Karate club network are displayed in figure 7.

Figure 7: Closeness Centrality for Zachary's Karate Club Network (Zachary, 1977)



4. **Eigenvector centrality:** measures the degree to which an actor is connected to other well connected actors. It takes advantage of a mathematical property of networks (represented as adjacency matricies) that allows for the easy calculation of how well connected an actor is to other well connected actors. While we will not get into the details of its calculation, this measure captures the value of having a lot of friends in high places. Eigenvector scores for Zachary's Karate club network are displayed in figure 8.

Figure 8: Eigenvector Centrality for Zachary's Karate Club Network (Zachary, 1977)



5. **Brokerage:** describes the position of actors such that they occupy an advantageous position where they can broker interactions between other actors in the network. **Brokerage Centrality** is then a measure of the degree to which an actor occupies a brokerage position across all pairs of alters. It is meant to capture the intuition that a broker serves as a go-between and thus can gain benefits from their position as an intermediary. There are five kinds of brokerage relationships, each of which we will discuss briefly below:

   (a) A **Coordinator** is an Actor in the same group as two alters who connects the two nodes. An example might be a graduate student who makes sure that all of the rest of their cohort is made aware of parties being hosted by anyone in their cohort.

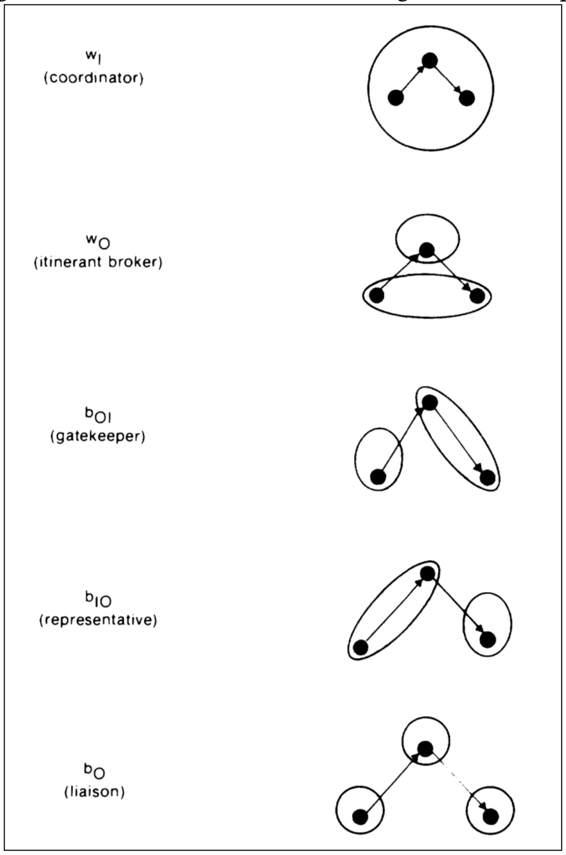
   (b) An **Itinerant** broker is a member of an outside group that connects two others who share group membership.

   (c) A **Gatekeeper** is a member of the same group as the target a member of another group hopes to connect with that can control whether or not that outside actor is able to gain access to the in group member. An example might be a secretary or office manager.

   (d) A **Representative** is a member of the same group as an Actor that wishes to connect with an actor outside of the group but has to go through an intermediary. An example is an Ambassador for a country.

   (e) A **Liason** is a member of a group that is distance from two actors that wish to connect but do not share group membership themselves. A delivery truck driver is a good example.

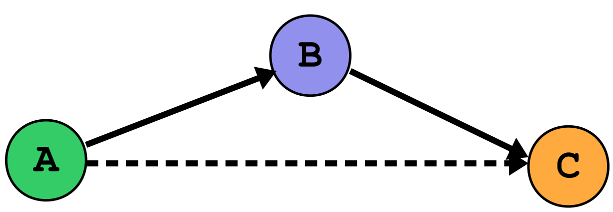
Figure 9: Different Kinds of Brokerage Relationships



3 Network Relationships and Structures

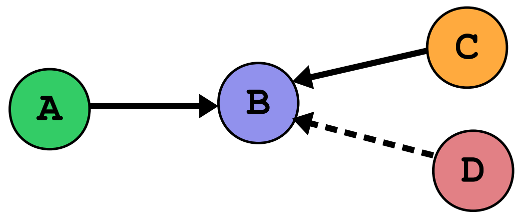
1. **Reciprocity:** is the tendency for directed ties from actor i to actor j be be reciprocated and sent back from actor j to actor i. This captures the classic finding that feeling and actions tend to be reciprocated.

2. **Transitivity:** is the tendency for friends of friends to be friends and enemies of enemies to be enemies. More generally a transitive relationship is one where two nodes being connected to a third increases the likelihood that they will connect themselves (Hoff et al., 2002; Carpenter et al., 2004).

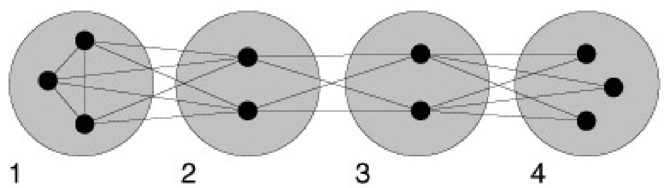


3. **Preferential Attachment (Popularity):** expresses the tendency for nodes that are already central to gain more connections at a greater rate than those who are not already central. This is often the

case in academia where as a researcher becomes more active and collaborates more in publishing, they are more likely to attract new collaborators who want to work with them.

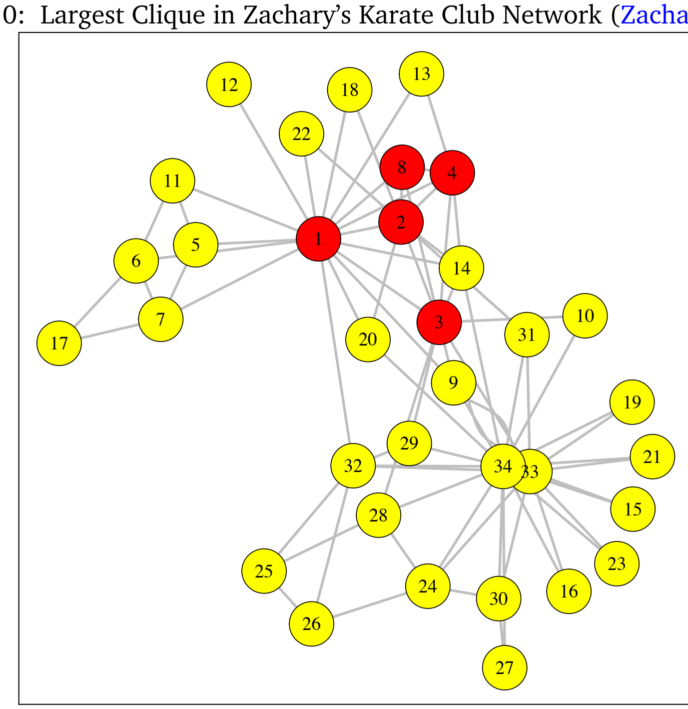


4. **Structural Equivalence:** is a concept that describes actors occupying the same position in the network relative to all other actors (Lorrain and White, 1971). In the example figure below, each grey circle contains a set of actors that are structurally equivalent to all others. This concept is important in making comparisons between nodes about their relative importance and position in a network. Check out the following web resources for more information: Robert A. Hanneman's Page on Structural Equivalence, Tom Schnijders Lecture on Structural Equivalence.

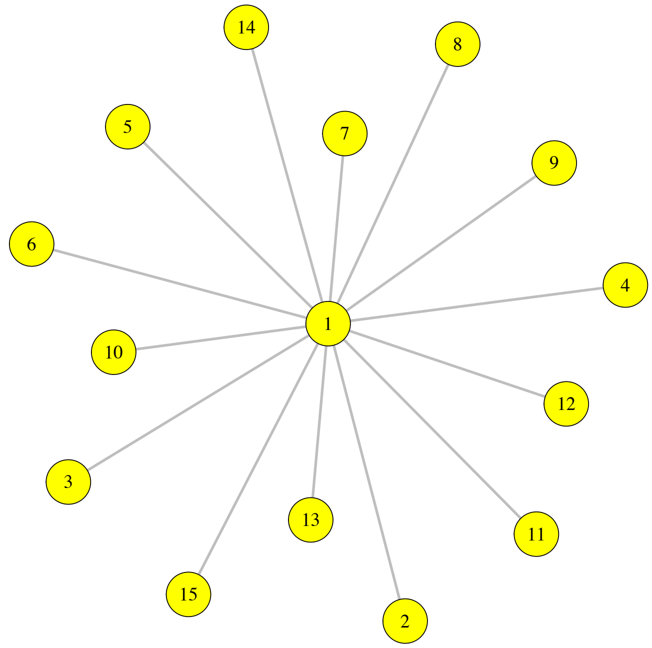


5. A **Clique:** is a subset of actors in a network such that every two actors in the subset are connected by a tie. This definition follows the common english language usage of the word meaning a densely connected group. A large example clique is colored red in Figure 10.

Figure 10: Largest Clique in Zachary's Karate Club Network (Zachary, 1977)



6. A **Star:** is a network structure where all ties connect to one central node, making the shape of a star.

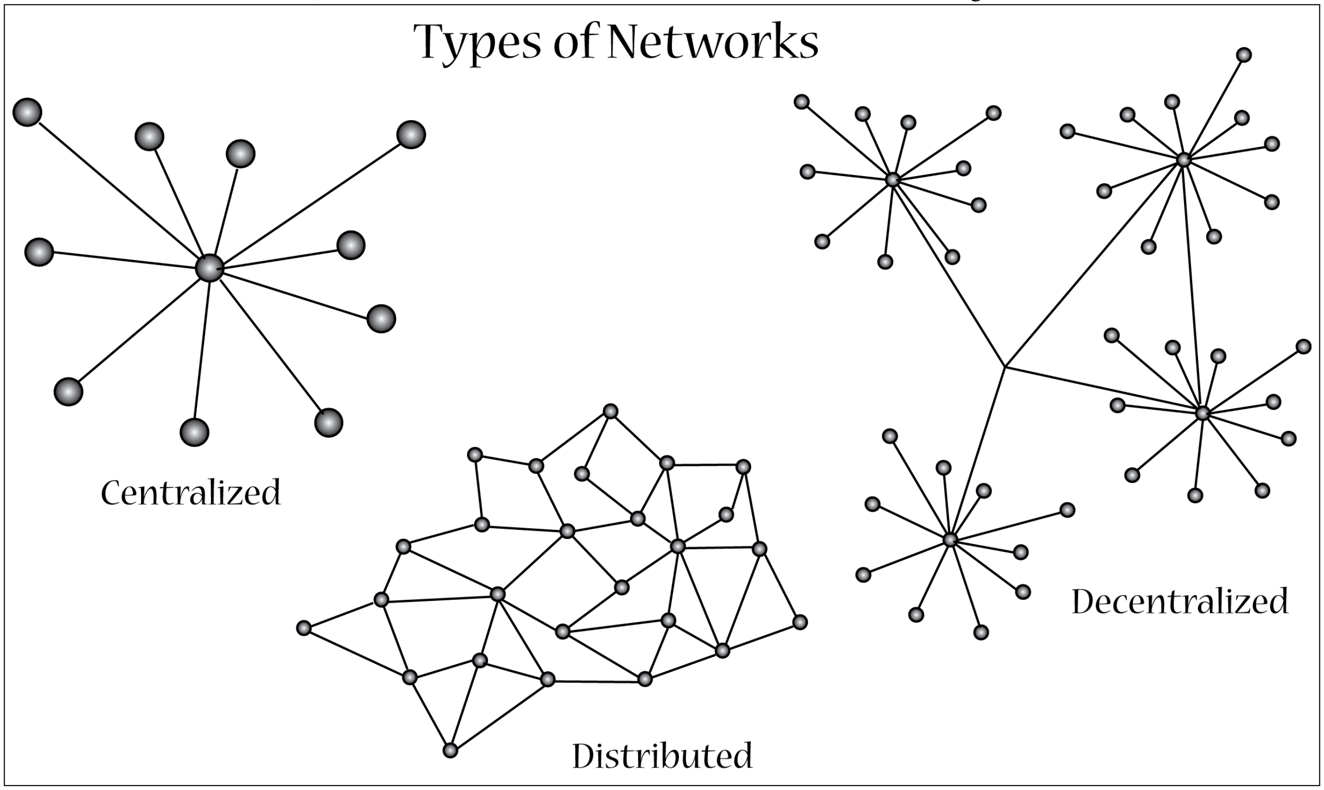


4 Network Properties

All of the properties discussed above refer to individual actors or subsets of actors in a network. While these are important characteristics to measure, we can also think about properties that a network as a whole exhibits. These properties are important because they impost structure on the entire space of interactions and relationships and can have profound aggregate effects on how actors in the network behave and function as a whole.

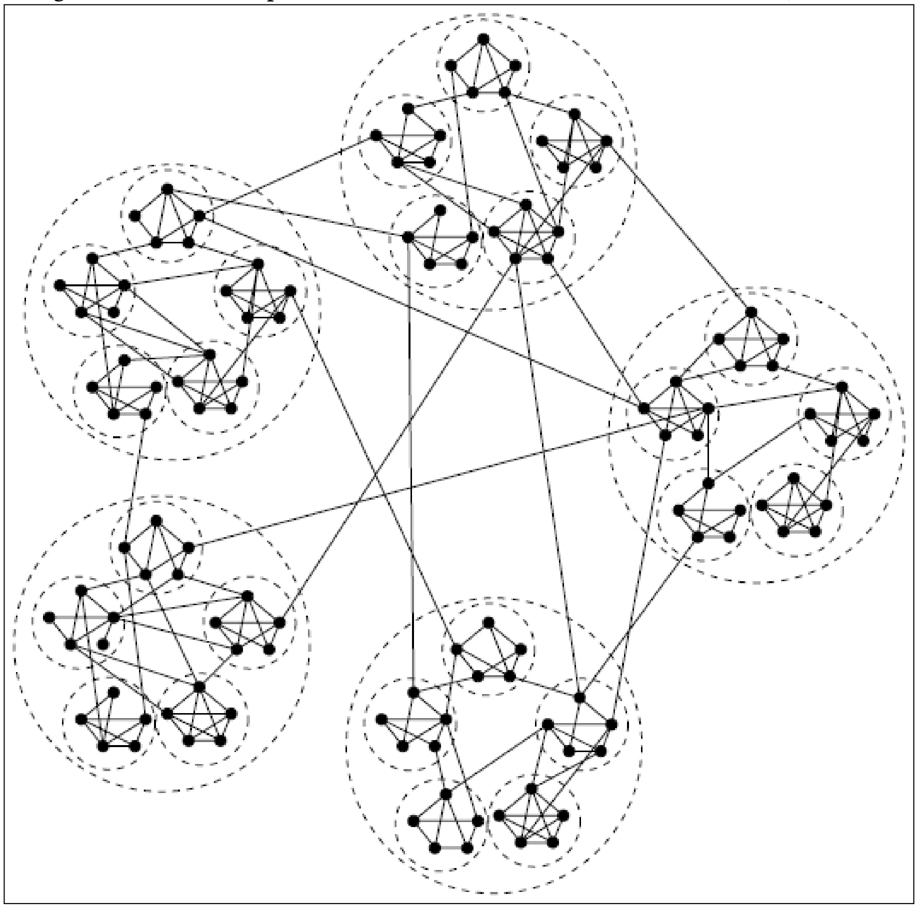
1. **Centralization (Degree, Betweenness, Closeness, Eigenvector, etc.):** is a measure of the unevenness of the centrality scores of actors in a network. It ranges from zero, when every actor is just as central for whatever score we are interested in, to 1, when one node is maximally central and all others are minimally central. This measure is a good way to express the idea that there are couple of very powerful or important actors in a network or that power/importance is spread out evenly in one simple measure (Ward et al., 2011).

Figure 11: An example of a highly **centralized** network and for comparison, a **decentralized** network (small centralized components that are connected), and a **distributed** network (actors all have a similar degree).



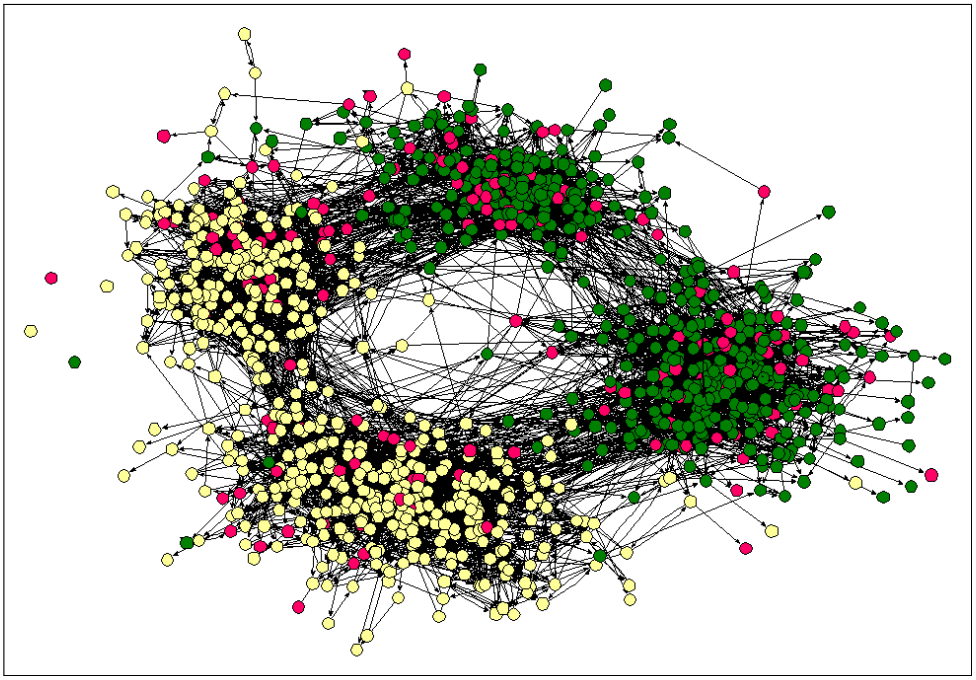
2. The network **Clustering Coefficient:** measures the degree to which actors form ties in in dense, relatively unconnected (between groups) groups. This measure is agnostic about why the network is clustered. The degree of clustering in a network is related to the efficiency with which information can diffuse over the network, as well as its robustness to disruption. (Latora and Marchiori, 2001; Newman, 2003; Suri and Watts, 2011; Mason and Watts, 2012)

Figure 12: An example of clusters within a network (Kaiser et al., 2007).



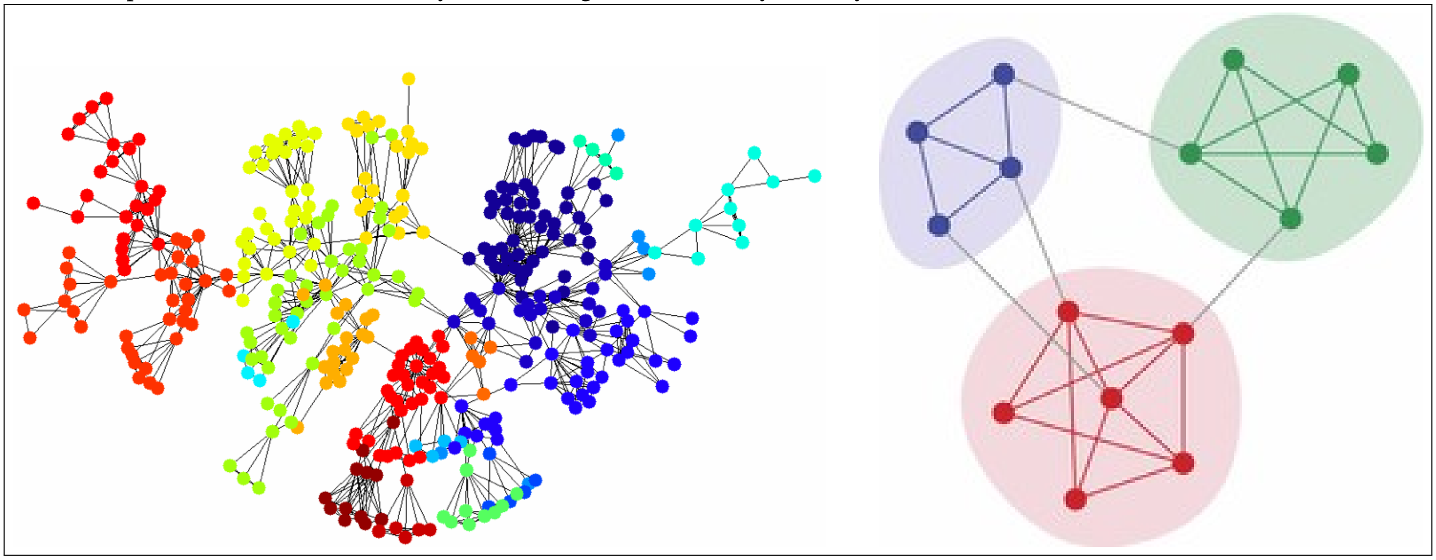
3. **Homophily:** is a process where actors who are similar on a particular trait are more likely to form ties. This has been confirmed in over 100 empirical studies, with a few examples including: (Ibarra, 1992; Straits, 1996; McPherson et al., 2001; Centola et al., 2007; Goodreau et al., 2009; Kossinets and Watts, 2009; McDonald, 2011). This process is the basis for the commonly used phrase “birds of a feather, flock together”. A classic example of a sociological study of homophily (by race) is provided in Figure 13. The opposite of homophily is **Heterophily**, which refers to a process whereby actors who are different from each other are more likely to form ties. An example of heterophily may be that of formal academic advising relationships, with students being more likely to form ties to faculty for advising than to other students.

Figure 13: Racial homophily in a high school friendship network. Nodes are connected if students are friends and colored by race with yellow and green nodes forming two distinguishable groups and even smaller minority students (red) in both main clusters (Moody, 2001).



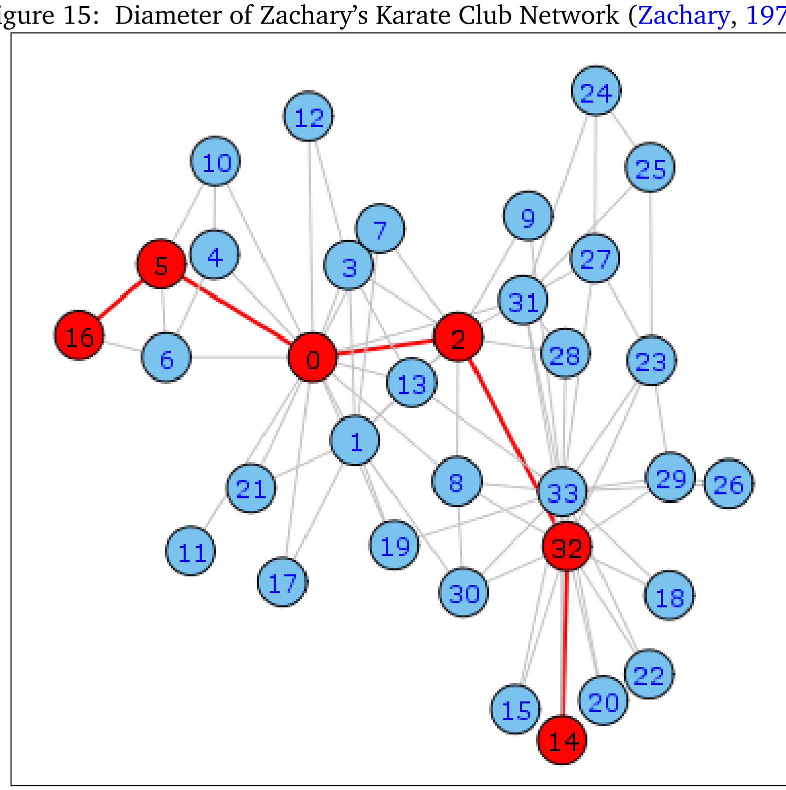
4. **Modularity:** is a measure of the degree to which a network displays **Community Structure**, with clusters that are not densely connected to others but densely connected within cluster. This measure is very difficult to calculate, but provides a way to identify community structure on a network where where one is unsure if such a structure exists (Newman, 2006; Zhang et al., 2008; Karrer and Newman, 2011). However, this measure is not consistent across networks of different size and group size. *Graph Compartmentalization* – a related measure – does allow for comparison between networks of arbitrary size and structure, but is not designed for detecting communities (Denny, 2014). An example of community structure between authors of papers about network analysis is presented in Figure 14.

Figure 14: The largest connected component of citation network for authors publishing on networks with actors colored by community membership is shown on the left (Porter et al., 2009). On the right we have a simple example of community membership in a network recovered by maximizing the modularity of a toy network.



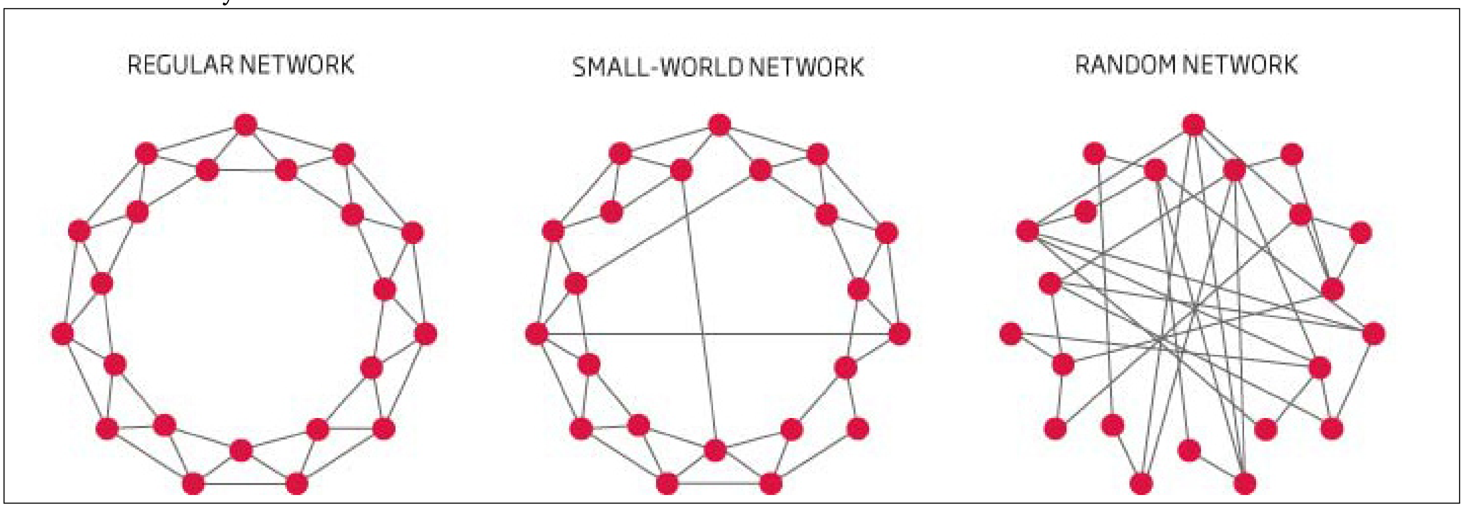
5. The **Diameter:** of a network is defined as the longest of all the calculated shortest paths between actors. Network diameter gives us an idea about how easily reachable Actors are on a network. A very large diameter means that even though there is theoretically a way for ties to connect any two actors through a series of intermediaries, there is no guarantee that they actually will be connected. Diameter is thus a signal about the ability for information or disease to diffuse on the network. The diameter of of Zachary's Karate club network is displayed graphically in Figure 15.

Figure 15: Diameter of Zachary's Karate Club Network (Zachary, 1977)



6. There are a number of classic **Network Types** that can be used to characterize the stereotypical social structure in different situations. Regular networks are characterized by all actors having the same degree and are often a starting point for simulation studies of networks (Centola et al., 2007). Small world networks are very efficient for information transfer in that most nodes are not connected (so a high degree of clustering) but also have a relatively short average path length between actors (Travers and Milgram, 1969; Watts and Strogatz, 1998). Random networks are very robust to disruptions (Latora and Marchiori, 2001; Callaway et al., 2000) but may be difficult for people to maintain, especially if ties are across long distances (Dodds et al., 2003; Aral et al., 2012). Examples of network types originally discussed in Watts and Strogatz (1998) are shown in Figure 16.

Figure 16: An example of a **Regular Network** (all actors have the same degree and are structurally equivalent to eachother), a **Small World Network** where dense clusters are connected by random and far reaching ties and a **Random Network**, where actors are randomly connected and there is no discernable structure.



5 Resources

What follows is a non-exhaustive list of my favorite reference materials on social network analysis and theory:

1. **The Bible:** *Social Network Analysis: Methods and Applications* by Wasserman and Faust (1994) is really the only book you should buy on social network analysis. It is used by everyone, it is **the** gold standard textbook on the subject and covers the theory and mathematical derivations behind a whole bunch of useful network properties.

2. *The structure and function of complex networks* (Newman, 2003) is a nice (technical) review of the literature on network analysis with 492 relevant citations to other articles.

3. **ICPSR** runs two 4 week summer sessions where they teach an introductory network analysis course and an advanced topics course. Check out their website for course offerings:
http://www.icpsr.umich.edu/icpsrweb/sumprog/index.jsp

4. Some interesting academic websites related to network analysis:

(a) Stanford SNAP Lab

(b) Lazer Lab at Northeastern

(c) Carter Butts' Research Page

References

Aral, Sinan and Marshall Van Alstyne. The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, 117(1):90–171, 2011. http://www.jstor.org/stable/10.1086/661238. 1

Aral, Sinan and Dylan Walker. Identifying influential and susceptible members of social networks. *Science (New York, N.Y.)*, 337(6092):337–41, July 2012. http://www.ncbi.nlm.nih.gov/pubmed/22722253. 1

Aral, Sinan, Erik Brynjolfsson, and Marshall Van Alstyne. Information, Technology, and Information Worker Productivity. *Information Systems Research*, 23(3-part-2):849–867, September 2012. http://pubsonline.informs.org/doi/abs/10.1287/isre.1110.0408. 17

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily. *Network Science*, pages 0–43, 2013. 1

Callaway, Duncan S., Mark E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: percolation on random graphs. *Physical review letters*, 85(25):5468–71, December 2000. http://www.ncbi.nlm.nih.gov/pubmed/11136023. 17

Carpenter, Daniel P., Kevin M. Esterling, and David M. J. Lazer. Friends, Brokers, and Transitivity: Who Informs Whom in Washington Politics? *Journal of Politics*, 66(1):224–246, February 2004. http://doi.wiley.com/10.1046/j.1468-2508.2004.00149.x. 10

Centola, D, J C González-Avella, V M Eguíluz, and M San Miguel. Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905, 2007. 14, 17

Cranmer, Skyler J. and Bruce A. Desmarais. Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66, 2011. http://people.cs.umass.edu/~wallach/courses/s11/cmpsci791ss/readings/cranmer11inferential.pdf. 1

Denny, Matthew J. Graph Compartmentalization. 2014. 15

D'Exelle, Ben and Nathalie Holvoet. Gender and Network Formation in Rural Nicaragua: A Village Case Study. *Feminist Economics*, 17(2):31–61, April 2011. http://www.tandfonline.com/doi/abs/10.1080/13545701.2011.573488. 1

Dodds, Peter Sheridan, Duncan J. Watts, and Charles F. Sabel. Information exchange and the robustness of organizational networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21):12516–21, October 2003. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=218789&tool=pmcentrez&rendertype=abstract. 17

Fowler, James H. Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, 14(4):456–487, March 2006. http://pan.oxfordjournals.org/cgi/doi/10.1093/pan/mpl002. 1

Gomez Rodriguez, Manuel, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, page 1019, 2010. http://dl.acm.org/citation.cfm?doid=1835804.1835933. 1

Goodreau, Steven M., James A. Kitts, and Martina Morris. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1):103–125, 2009. http://link.springer.com/article/10.1353/dem.0.0045. 14

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, December 2002. http://www.tandfonline.com/doi/abs/10.1198/016214502388618906. 1, 10

Ibarra, Herminia. Homophily and Differential Returns : Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Sceince Quarterly*, 37:422–447, 1992. http://www.jstor.org/stable/10.2307/2393451. 14

Jackson, Matthew O. and Dunia López-Pintado. Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(01):49–67, April 2013. http://www.journals.cambridge.org/abstract_S2050124212000070. 1

Kaiser, M, M Görner, and C C Hilgetag. Criticality of Spreading Dynamics in Hierarchical Cluster Networks Without Inhibition. *New Journal of Physics*, 9(5):110–110, May 2007. http://stacks.iop.org/1367-2630/9/i=5/a=110?key=crossref.dfed57894257e4d7543a2907e823ade6. 14

Karrer, Brian and Mark E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83 (1):1–11, 2011. http://pre.aps.org/abstract/PRE/v83/i1/e016107. 15

Kirkland, Justin H. The Relational Determinants of Legislative Outcomes: Strong and Weak Ties Between Legislators. *The Journal of Politics*, 73(03):887–898, August 2011. http://www.journals.cambridge.org/abstract_S0022381611000533. 1

Kossinets, Gueorgi and Duncan J. Watts. Origins of Homophily in an Evolving Social Network. *American Journal of Sociology*, 115(2):405–450, September 2009. http://www.jstor.org/stable/10.1086/599247. 14

Krackhardt, D. Cognitive social structures. *Social Networks*, 9(2):109–134, 1987. 5

Krafft, Peter, Juston Moore, Bruce A Desmarais, and Hanna Wallach. Topic-partitioned multinetwork embeddings. In *Advances in Neural Information Processing Systems Twenty-Five*, 2012. http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2012_1288.pdf. 1

Latora, Vito and Massimo Marchiori. Efficient Behavior of Small-World Networks. *Physical review letters*, 2001. http://prl.aps.org/abstract/PRL/v87/i19/e198701. 13, 17

Levin, Daniel Z. and Rob Cross. The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Management Science*, 50(11):1477–1490, November 2004. http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.1030.0136. 1

Lorrain, François and Harrison C. White. Structural Equivalence of Individuals in Social Networks. *The Journal of Mathematical Sociology*, 1(1):49–80, January 1971. http://www.tandfonline.com/doi/abs/10.1080/0022250X.1971.9989788. 11

Lubin, Ben, Jesse Shore, and Vatche Ishakian. Communication Network Design: Balancing Modularity and Mixing via Extremal Graph Spectra. 2013. http://mackinstitute.wharton.upenn.edu/wp-content/uploads/2013/07/Shore-Jesse_Communication-network-design.pdf. 1

Marwell, G, P E Oliver, and R Prahl. Social networks and collective action: A theory of the critical mass. III. *American Journal of Sociology*, pages 502–534, 1988. 1

Mason, Winter and Duncan J. Watts. Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, 109 (3):764–769, 2012. http://www.pnas.org/content/109/3/764.short. 13

McDonald, Steve. What's in the old boys network? Accessing social capital in gendered and racialized networks. *Social Networks*, 33(4):317–330, October 2011. http://linkinghub.elsevier.com/retrieve/pii/S0378873311000554. 14

McPherson, M, L Smith-Lovin, and J M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001. 1, 14

Moody, James. Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, 107(3): 679–716, 2001. http://www.jstor.org/stable/10.1086/338954. 15

Moreno, Jacob Levy. Who shall survive?: A new approach to the problem of human interrelations. 1934. 1, 2

Neves, Guilherme, Sam F Cooke, and Tim V P Bliss. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature reviews. Neuroscience*, 9(1):65–75, January 2008. http://www.ncbi.nlm.nih.gov/pubmed/18094707. 1

Newman, M E J. The structure and function of complex networks. *SIAM review*, pages 167–256, 2003. 1, 13, 17

Newman, Mark E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–82, June 2006. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1482622&tool=pmcentrez&rendertype=abstract. 15

Porter, Mason A., Jukka-Pekka Onnela, and Peter J. Mucha. Communities in Networks. *Notices of the AMS*, 56(9), 2009. http://www.ams.org/notices/200909/rtx090901082p.pdf. 16

Raftery, Adrian E., Xiaoyue Niu, Peter D. Hoff, and Ka Yee Yeung. Fast Inference for the Latent Space Network Model Using a Case-Control Approximate Likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919, 2012. http://www.tandfonline.com/doi/abs/10.1080/10618600.2012.679240. 1

Stackman, Richard W. and Craig C. Pinder. Context and Sex Effects on Personal Work Networks. *Journal of Social and Personal Relationships*, 16:39–64, 1999. http://spr.sagepub.com/content/16/1/39.short. 1

Straits, Bruce C. Ego-net diversity: Same- and cross-sex coworker ties. *Social networks*, 18:29–45, 1996. http://www.sciencedirect.com/science/article/pii/0378873395002545. 14

Sundararajan, Arun, Foster Provost, Gal Oestreicher-Singer, and Sinan Aral. Information in Digital, Economic and Social Networks. *SSRN Electronic Journal*, 3, 2012. http://www.ssrn.com/abstract=2141682. 1

Suri, Siddharth and Duncan J Watts. Cooperation and contagion in web-based, networked public goods experiments. *PloS one*, 6(3):e16836, January 2011. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3055889&tool=pmcentrez&rendertype=abstract. 13

Travers, J and S Milgram. An experimental study of the small world problem. *Sociometry*, pages 425–443, 1969. 17

Ward, Michael D., Katherine Stovel, and Audrey Sacks. Network Analysis and Political Science. *Annual Review of Political Science*, 14(1):245–264, June 2011. http://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.12.040907.115949. 12

Wasserman, Stanley and Katherine Faust. *Social Network Analysis: Methods and Applications*, volume 8 of *Structural analysis in the social sciences, 8*. Cambridge University Press, 1994. http://www.amazon.com/dp/0521387078. 17

Watts, Duncan J. and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, June 1998. http://www.ncbi.nlm.nih.gov/pubmed/9623998. 17, 17

Watts, Duncan J., Peter Sheridan Dodds, and Mark E .J. Newman. Identity and Search in Social Networks. *Science (New York, N.Y.)*, 296(5571):1302–5, May 2002. http://www.ncbi.nlm.nih.gov/pubmed/12016312. 1

Zachary, Wayne W. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 33(4):452–473, 1977. http://www.maths.tcd.ie/~mnl/store/Zachary1977a.pdf. 1, 3, 6, 7, 7, 8, 9, 11, 16

Zhang, Yan, A. J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler, and Peter J. Mucha. Community Structure in Congressional Cosponsorship Networks. *Physica A*, 387(7):1705–1712, 2008. http://www.sciencedirect.com/science/article/pii/S037843710701206X. 15