Social Network Analysis

Seminar in Criminology, Research and Analysis– Crim 7301

Week 10, 10/27/16

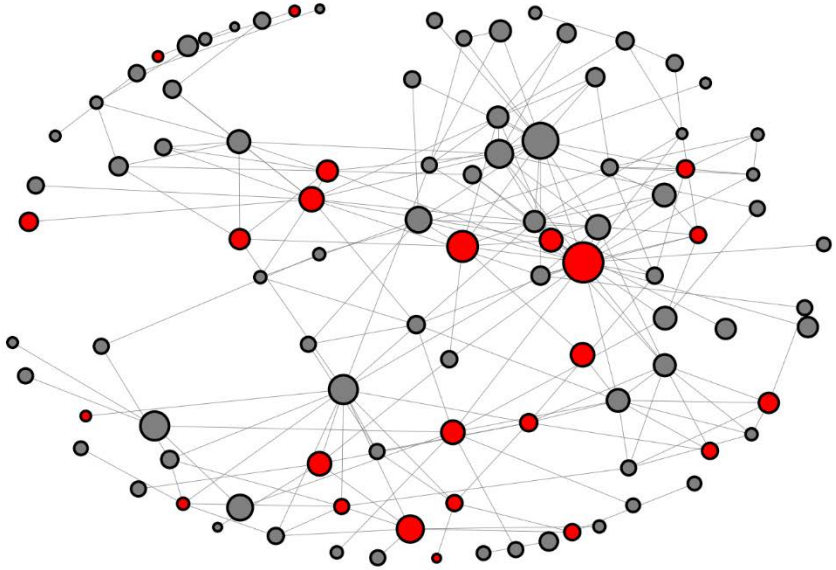
Andrew Wheeler

Class Overview

- Three general types of questions
  - What is the shape of the network
  - Why are certain nodes connected to other nodes
  - What effect does one node have on other connected nodes
- Centrality Measures
- Any dataset can be changed into a network
- Drawing and Software

What is the shape of the network

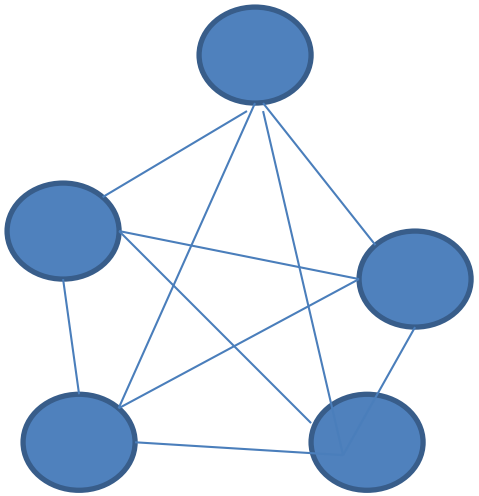
- Networks are collections of *nodes* that are connected by *edges*



What is the shape of the network

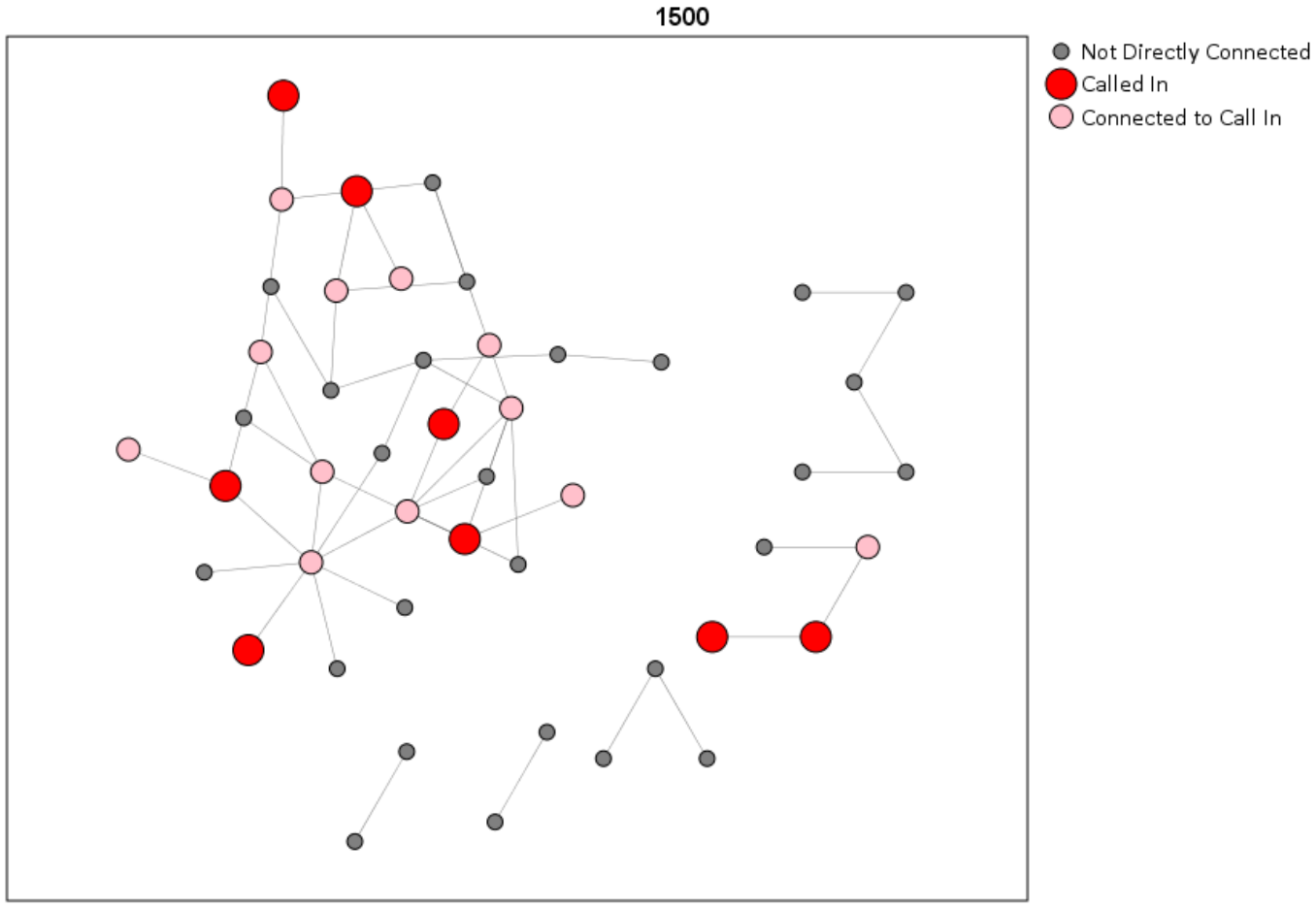
Some terminology

- *Clique:* A collection of nodes that are all connected to one another



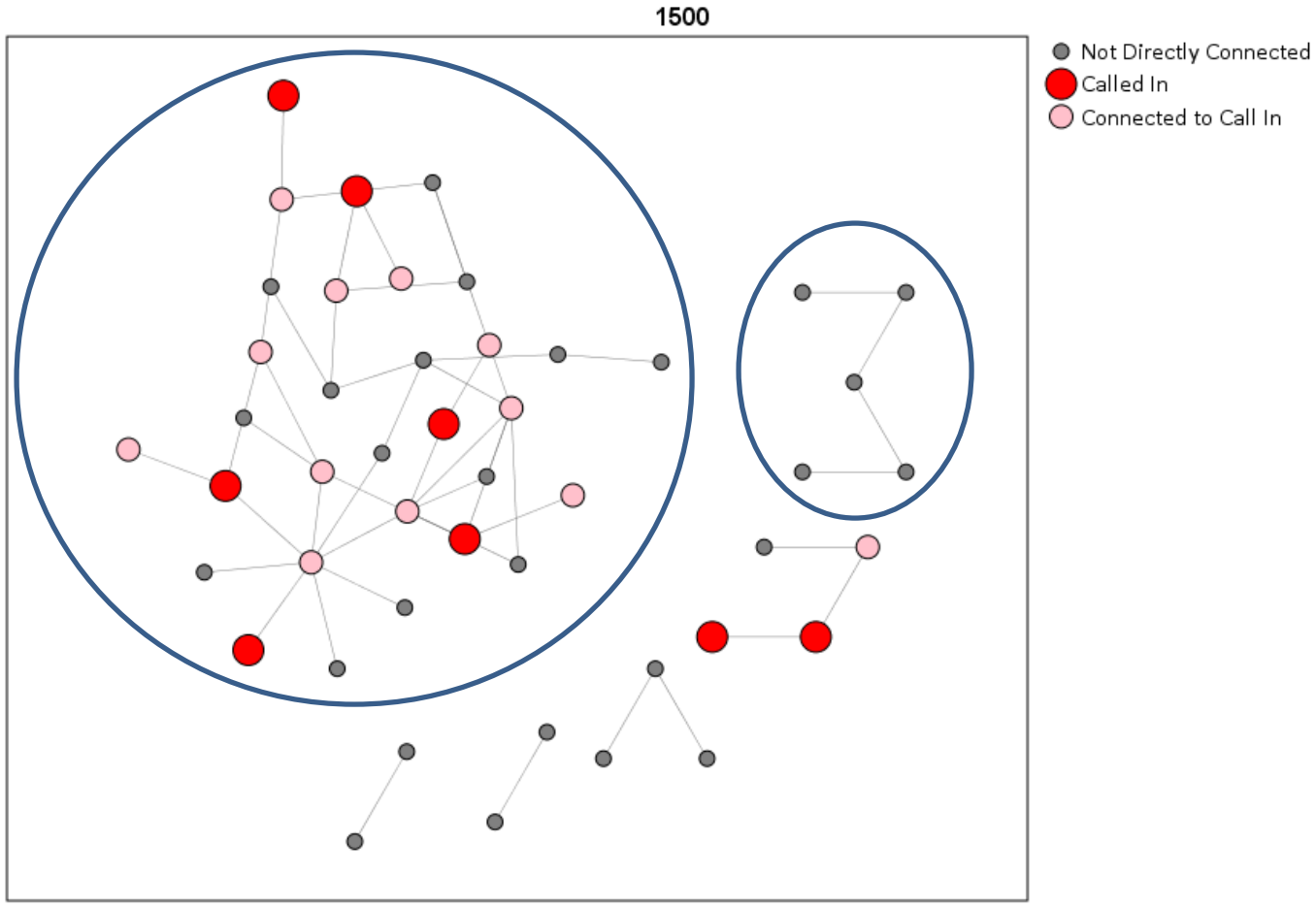
What is the shape of the network

- *Component*: A sub-set of a network that are not connected



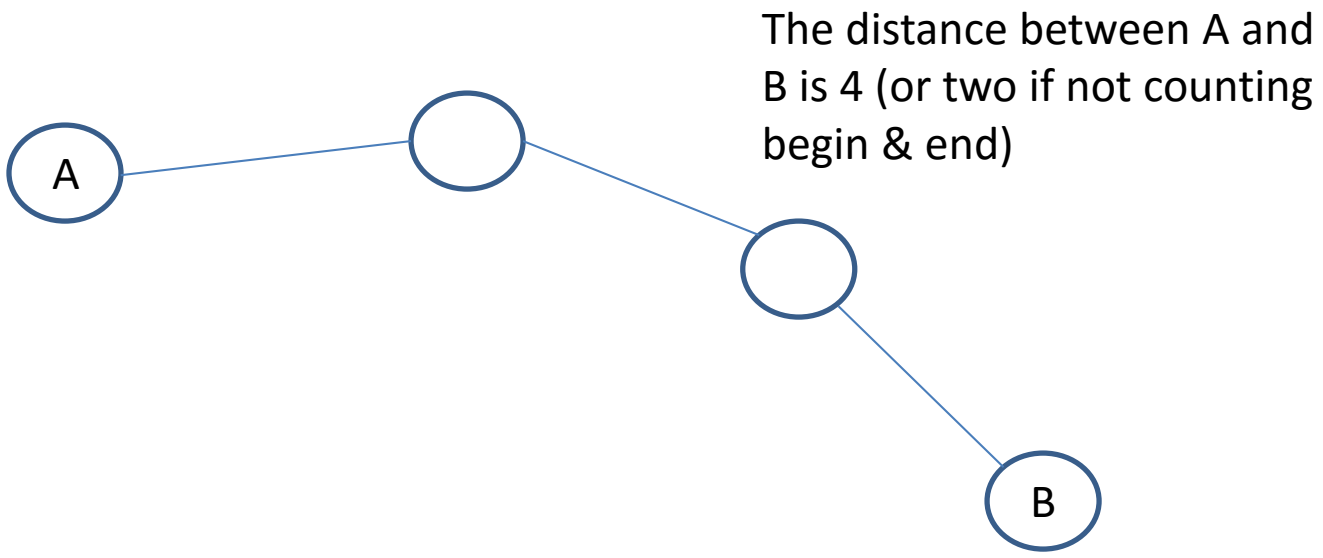
What is the shape of the network

- *Component*: A sub-set of a network that are not connected



What is the shape of the network

- *Geodesic:* A particular path between nodes



What is the shape of the network

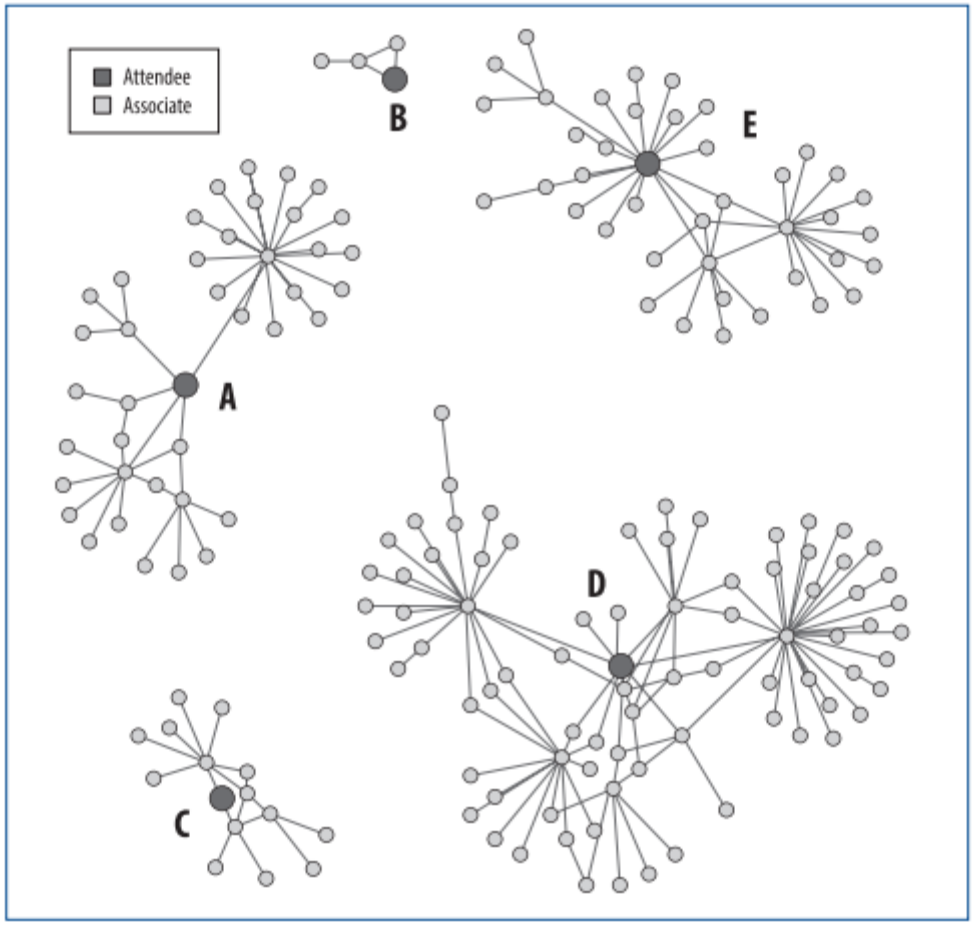
- *Density of a Network:* in an undirected network with N nodes, there are a total possible number of edges equal to [N*(N-1)]/2 nodes (or n choose 2). The density is the observed number of edges divided by this.

- *Length of a network*: the largest distance between any two nodes in a network (sometimes also called the *diameter*)

What is the shape of the network

The strength of weak ties (Granovetter, 1973)

Figure 2. Chicago's social network of call-in attendees, August 17, 2010



Source: Andrew Papachristos

Why are certain nodes connected to one another?

- *Homophily:* The tendency for connections to be between individuals with similar attributes

Makes it hard to tell the difference between correlation/causation – e.g. is obesity contagious?

Exponential random graph models: try to predict which nodes are connected to one another based on different node attributes

What effect does one node have on other nodes?

- Interventions can be either remove a node (which can change the entire network), or intervene with particular nodes and hope that intervention spreads

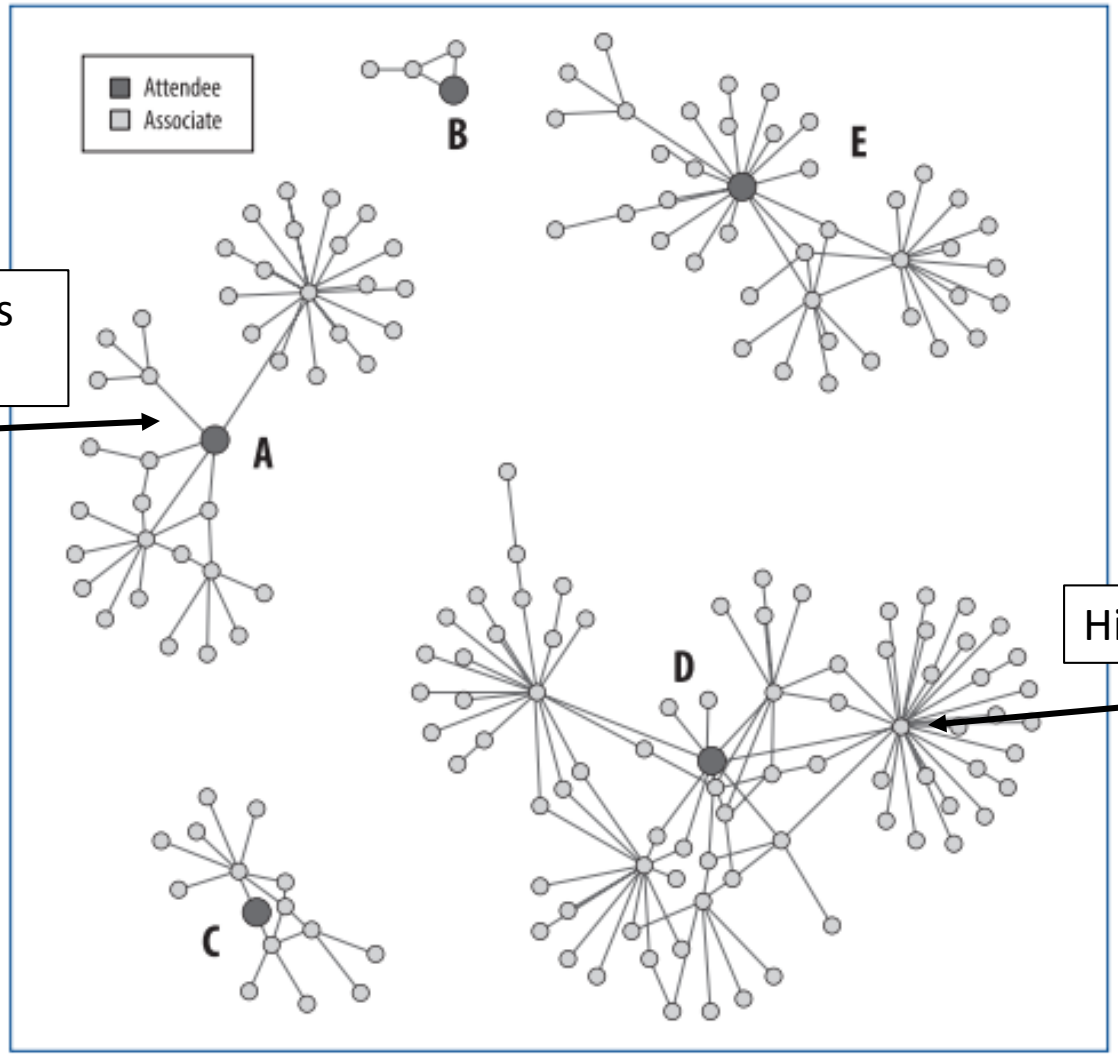
- can estimate network effects (spillovers) same as spatial effects, by incorporating network weight matrix (see *On the Large-Sample Estimation of Regression Models with Spatial- Or Network-Effects Terms: A Two-Stage Least Squares Approach* by Land and Deane)

Centrality Measures

- Edge centrality – the total number of edges connected to a node
- Betweenness Centrality – the number of shortest paths that pass through a particular node
- Many others – Eigenvector, Laplacian, Closeness, Bonachich

Centrality Measures

Figure 2. Chicago's social network of call-in attendees, August 17, 2010



Source: Andrew Papachristos

High betweenness low edge

High edge centrality

Any data can be network data

- Any table can be turned into a *bipartite* graph – a graph with two types of nodes

	V1	V2	V3
A	X	X	X
B	X	O	X
C	O	X	X

A	V1	X
A	V2	X
A	V3	X
B	V1	X
B	V2	O
B	V3	X
C	V1	O
C	V2	X
C	V3	X

Edge List format

Any data can be network data

- Any bipartite network can be *projected* into a single node network (see Breiger, 1974 – *The Duality of Persons and Groups*)

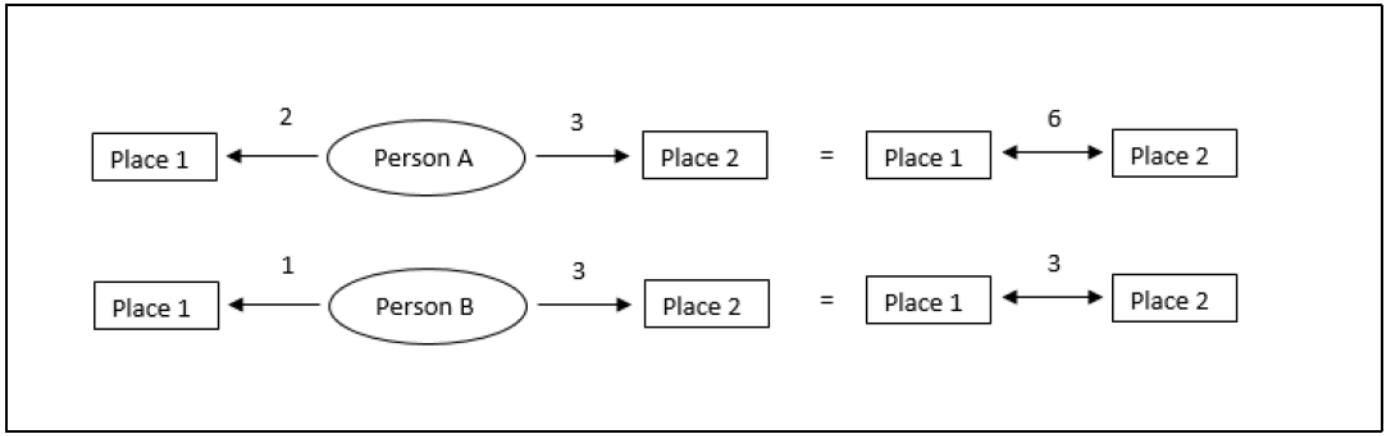
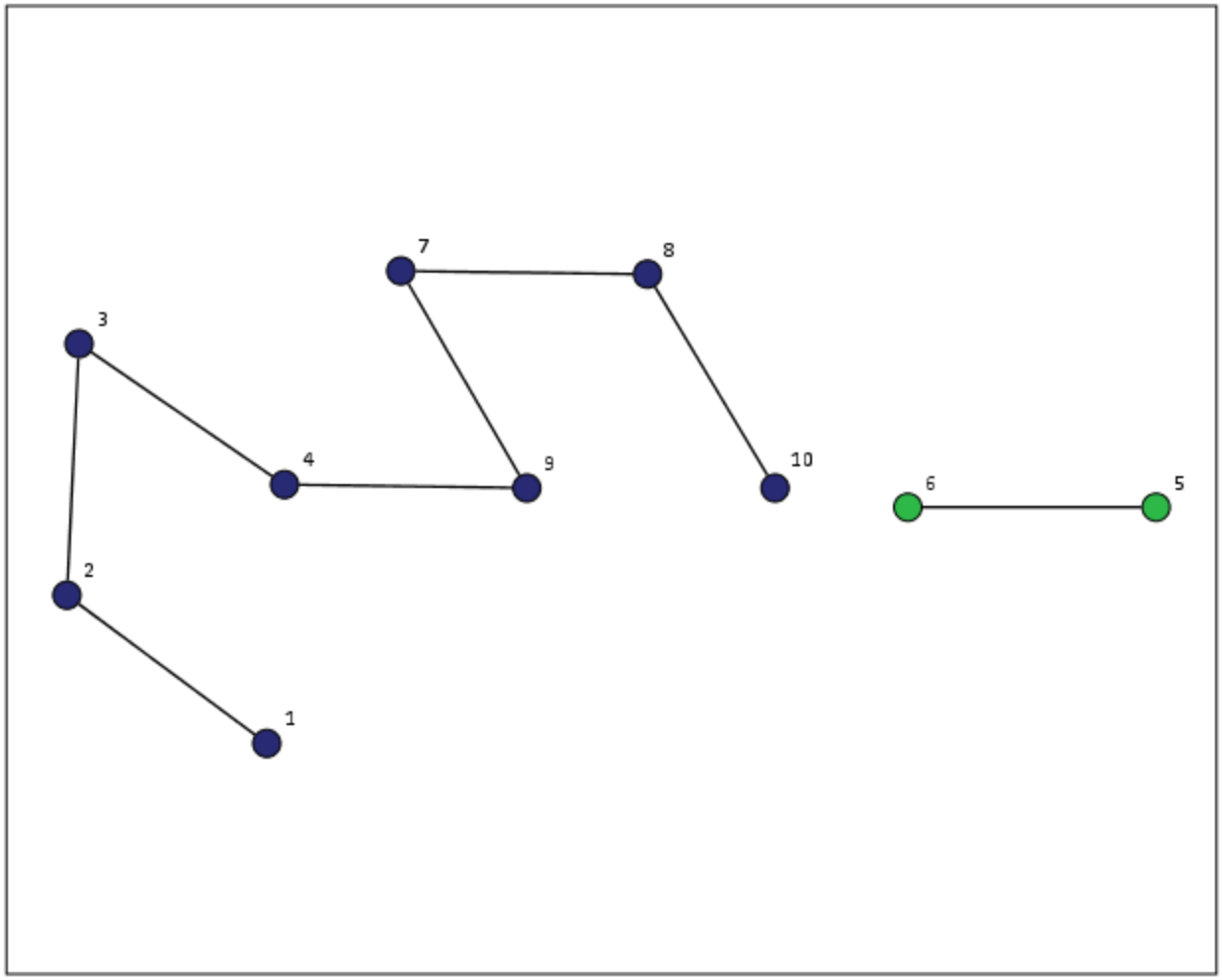


Figure 8: Description of the algorithm to turn a weighted bipartite graph into a weighted graph with a single type of node (Breiger, 1974). One multiples the weights along the path for each Place/Person/Place combination, and the final edge weight is the sum of those combinations. The final weight between Place 1 and Place 2 will be 6+3=9 in this example.

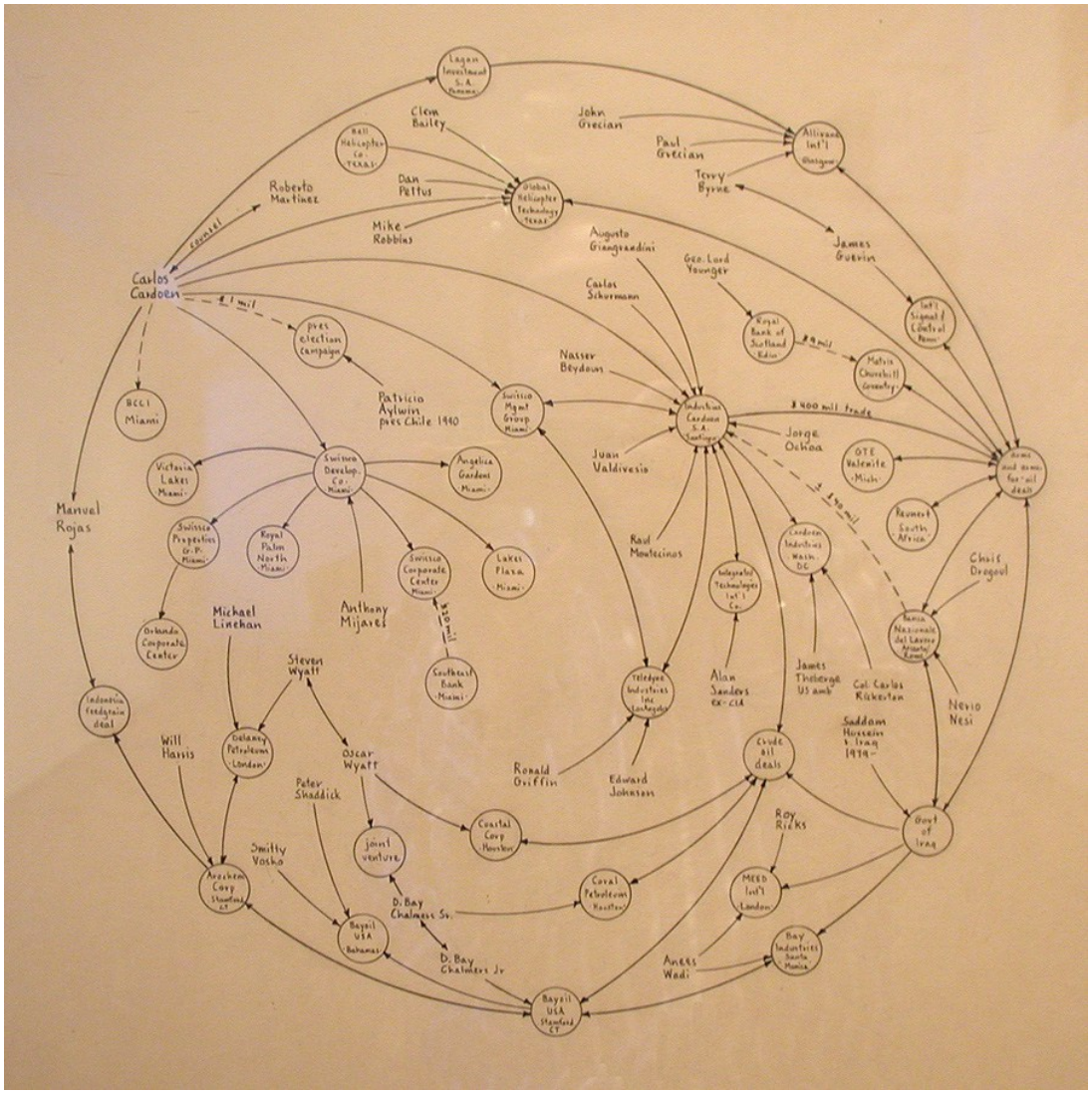
Drawing and Software

- Layout of networks – *planar* (lines do not cross) is better



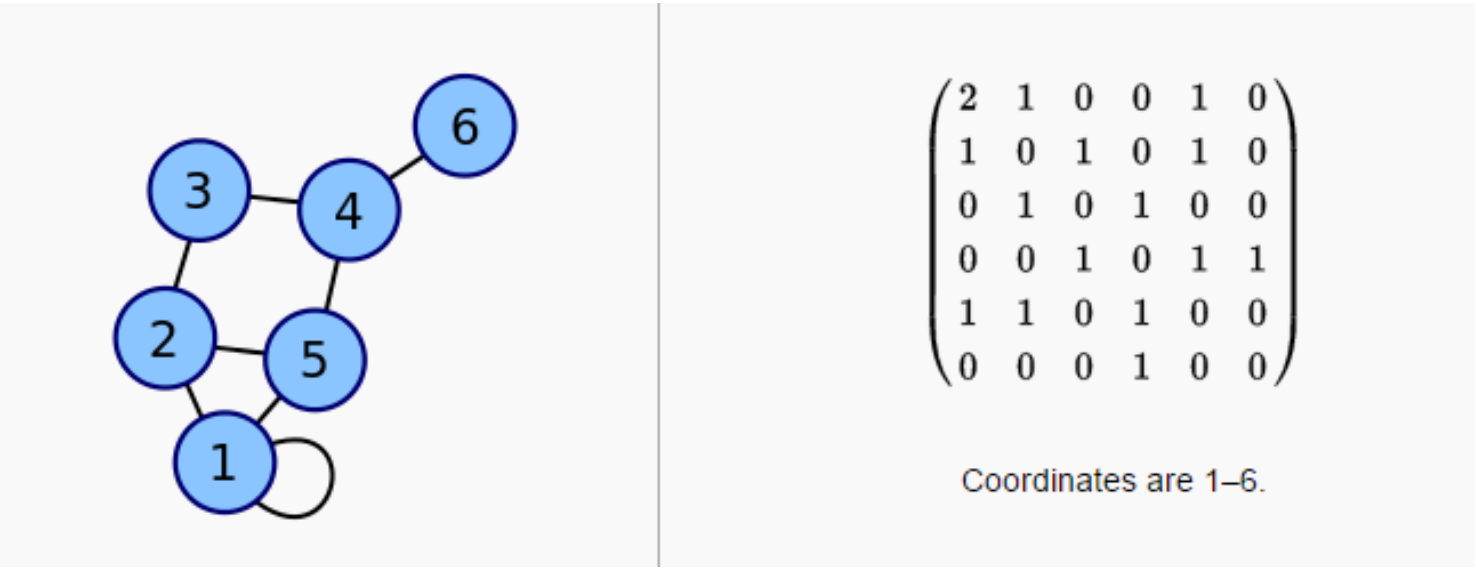
Drawing and Software

- Work of Mark Lombardi



Drawing and Software

- Adjacency Matrix (via Wikipedia, https://en.wikipedia.org/wiki/Adjacency_matrix)



Drawing and Software

- R has several packages
  - Igraph – good for larger networks, several clustering routines
  - Sna, network – good for smaller networks, exponential random graph models

- Python has networkx

- Gephi, https://gephi.org/, nice for drawing and interacting with graphs

Others:

- NodeXL (add on for excel)
- Software from Borgatti: UCINET, key-player
- PAJEK (for drawing large networks)

Homework & Next Weeks Class

Lab Assignment

Import, draw, and estimate different centrality metrics in R or python.

For Next Week – Machine Learning

- Berk, R. (2008). Forecasting methods in crime and justice. *Annual Review of Law and Social Science*, 4(1):219-238.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3): 289-310.
- Tollenaar, N. and van der Heijden, P. G. M. (2013). Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):565-584.
- Bushway, S. D. (2013). Is there any logic to using logit. *Criminology & Public Policy*, 12(3):563-567.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7):571-582.
- Ridgeway, G. (2013). The pitfalls of prediction. *NIJ Journal*, 271 February, 34-40.