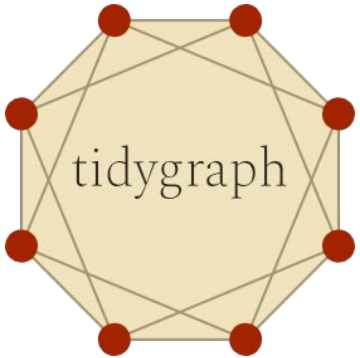
Intro to Network Analysis

Tony WT Hung (DAD, SBE)
For RUG @ UM

Contents

The igraph Package



Link: https://r.igraph.org/

One of the most used packages for network analysis.

Other network packages: sna, network, ergm… etc.

Of course, there's a 'tidy' approach: tidygraph and ggnetwork

Datasets + Links

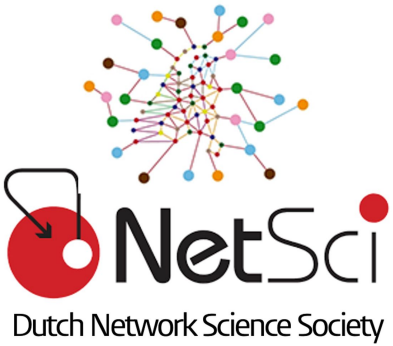
Big network datasets: https://snap.stanford.edu/data/

A wonderful repository: https://networkrepository.com/

GUI to analyze networks: NetLogo, JASP

Awesome List:
https://github.com/briatte/awesome-network-analysis

Network Science Dutch Chapter: https://www.netsci.nl/

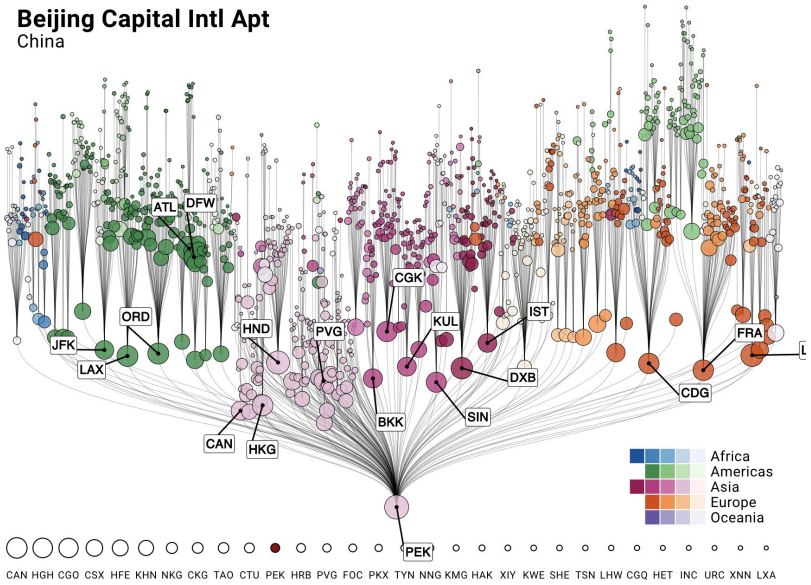


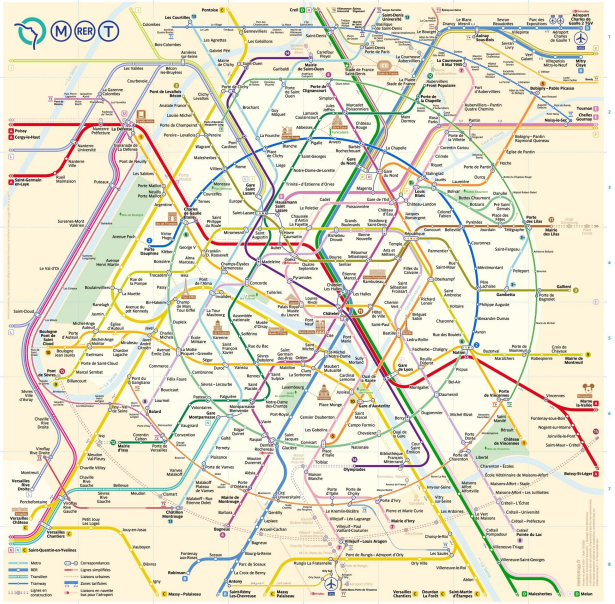
What and Why?

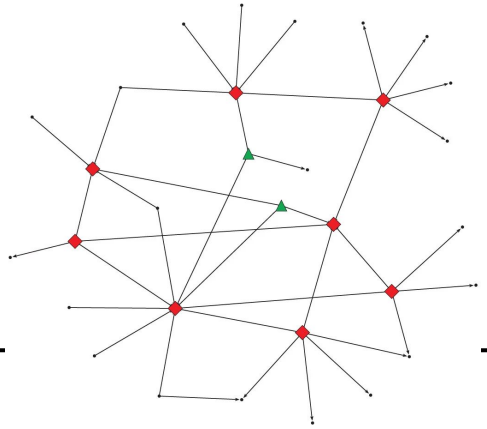
—

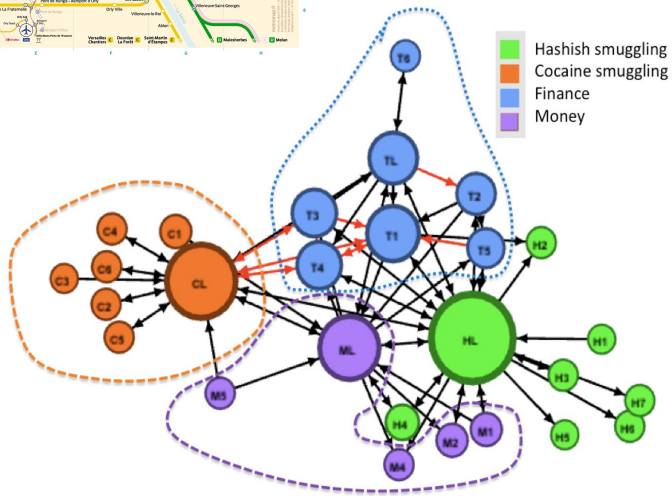
What and Why?

- Network analysis is perfect for relational data..
  - A lot of data that we know are rather relational (transactions, calling people…)
- Common networks: friendship, disease transmission, transportation, trade…
- ‘Revealed’ networks: financial transactions, co-signing of legislature, links between blogs…









Some Pictures

The Most Basic igraph Data Structures

	Mark	Jane	Peter
Mark	0	1	0
Jane	1	0	1
Peter	1	1	0

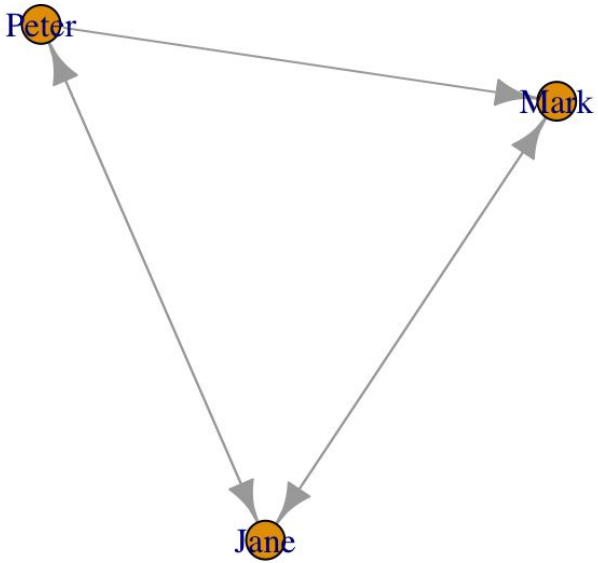
Adjacency Matrix

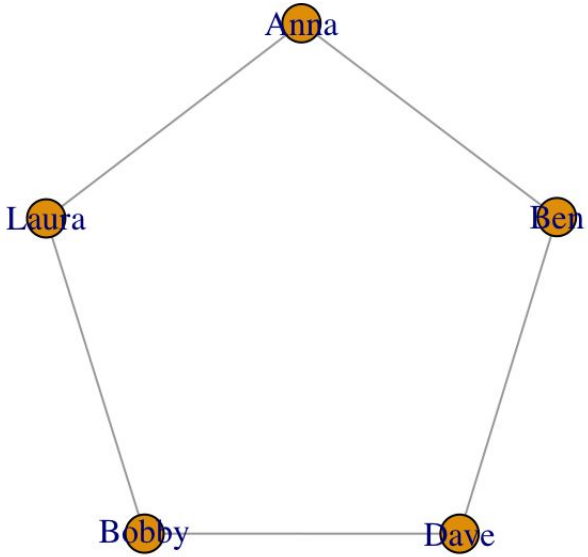
	V1	V2
1	Dave	Ben
2	Anna	Ben
3	Dave	Bobby
4	Laura	Bobby
5	Laura	Anna

Edge List

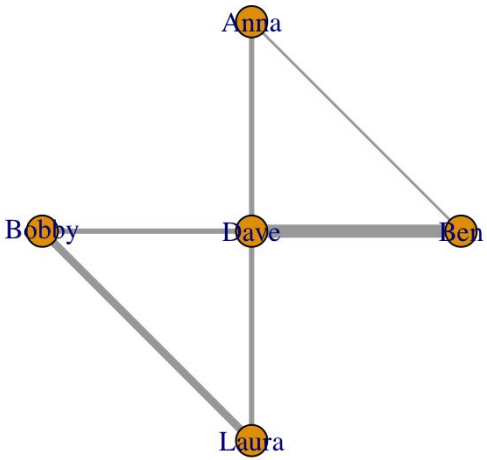
Network Types

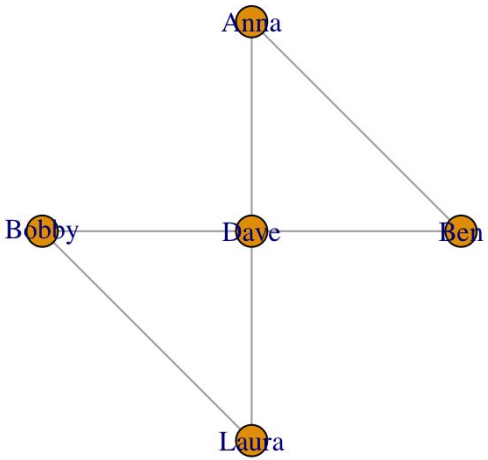
Directed / Undirected





Weighted / Unweighted





Look at the ‘Slide Script’

Describing a Network

Node(s)	Also called vertex or vertices → # nodes
Edge(s)	Also called link(s) or tie(s) → # edges
Network	Also called structure or graph
Density	The number of actual links divided by the number of potential links
Average Path Length	The average number of steps of all shortest (geodesic) paths

Common Network Measures (unweighted)

Degree centrality	The number of edges a node has. The higher, the more connections.
Betweenness centrality	Measures how often a node lies on the geodesic path between other nodes. The higher, the more in between.
Closeness centrality	Measures how close a node is from all other nodes, using the average distance. The higher, the closer.
Eigenvector centrality	Measure the type of connections you have. If you have a lot of powerful connections, then this mean that you are more important. The higher the better.
Assortativity / homophily	Measures how nodes will connect with each other based on a characteristic. Should interpret it somehow like 'correlation'.
Transitivity	Also means clustering or cliquishness. It measures whether a node is embedded in a tightly knit group of a clique, or other nodes. Basically clustering coefficient.

Renaissance in Florence, Italy

Let's go back in time...

- The Renaissance was a period of great human advancement, where they looked back at the classical antiquity, while inspiring novel creations.
- We know about the rich and powerful Medici family in the Renaissance. The Medici family basically 'funded' the Renaissance, being patrons of Da Vinci, Michelangelo, Machiavelli, and Galileo. Interestingly the Medici family also produced four popes.
- However, the Medici family used not only business connections, but also family connections for its rise to be the most influential family in Renaissance Italy...

Who was the most influential in Renaissance Italy? How?

Look at the Florentine script

Next Steps

More on network visualization

What we sort of did today, but with more explanation

Bipartite Graphs, Network Identification, Network Structures, ERGM

Sources + Attribution

Prof. E Power's Course on Social Network Analysis

http://rocs.hu-berlin.de/corona/

https://www.nature.com/articles/srep03711

https://cina.gmu.edu/projects/detecting-and-disrupting-transnational-criminal-organizations-analytics-for-interdependent-smuggling-and-money-laundering-networks/

https://metromap.fr/en