Visualising Social Networks with Cytoscape

12 October, 2020
Institute for Transnational & Spatial History - Institute Mondays

Konrad M. Lawson - kml8@st-andrews.ac.uk

Cytoscape tutorial overview posted here:
https://kmlawson.github.io/dh-tutorials/cytoscape-tutorial.html
Glossary, Links, and other Resources have been posted here:
https://kmlawson.github.io/dh-tutorials/cytoscape.html
These slides are available here:
https://kmlawson.github.io/dh-tutorials/cytoscape-slides.pdf
Cytoscape is available for download here:
https://cytoscape.org/

Overview

1. “Social Network Analysis” vs Visualising Social Networks
2. What is it for? When is it worth it?
3. Some basic SNA concepts for visualisation
4. How to collect and prepare data for visualisation
5. Getting your network into Cytoscape and creating network graphs.
6. Resources

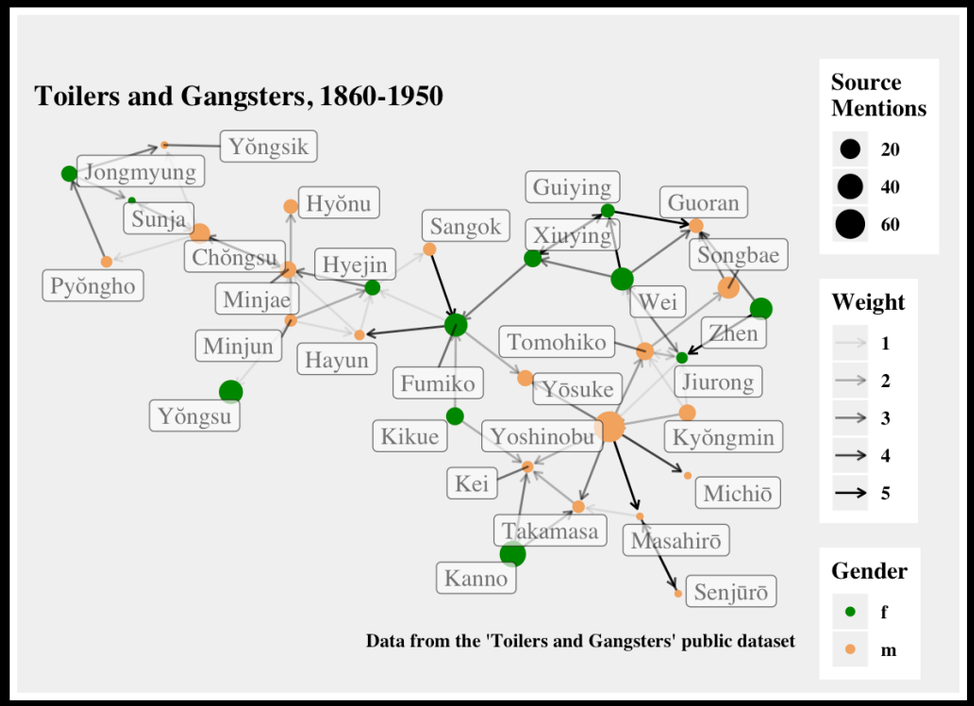
Social Network Analysis

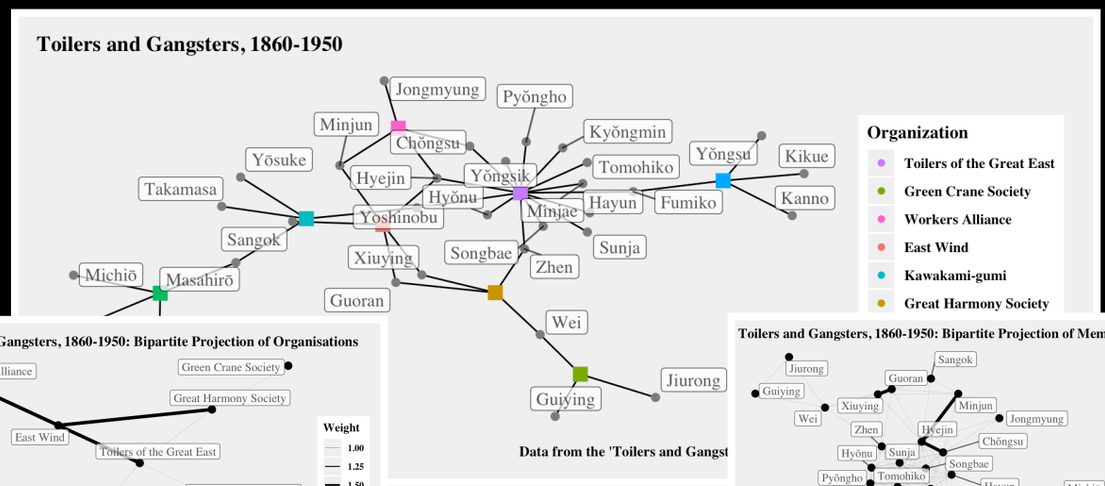
Sociological analysis that aims to describe and explore the patterns in social relationships. It assumes actors are interdependent, relations channel information, structures of relations constrain and facilitate action, and that patterns of relations define economic, political and social structures. Formal network methods employ statistics and the mathematics of graph theory to test hypotheses.

Combining Wasserstrom and Faust (1994) and Scott (2017)

Some terms and principles

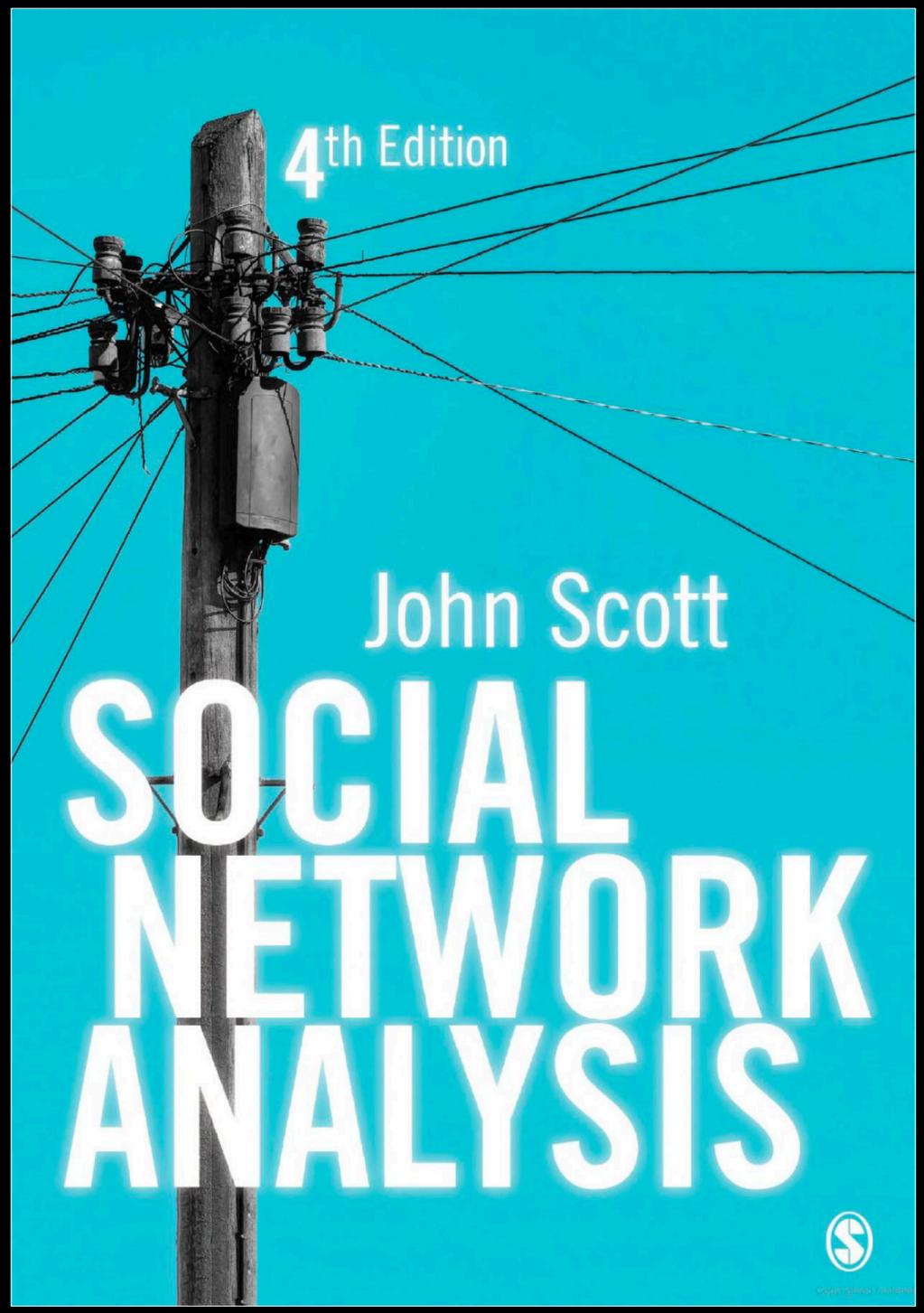
Graph
Nodes v Edges
Attributes
Whole Network vs
  Egocentric Network
Unimodal vs
  Bimodal (Bipartite) Networks
Directed vs
  Undirected Networks
Weighted vs Unweighted Networks
Centrality

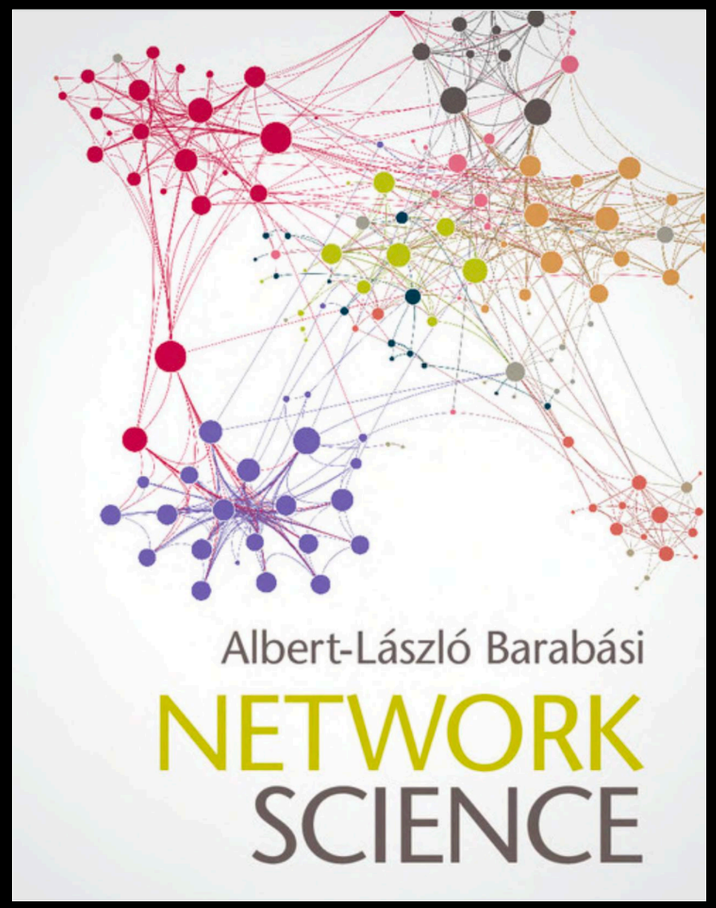


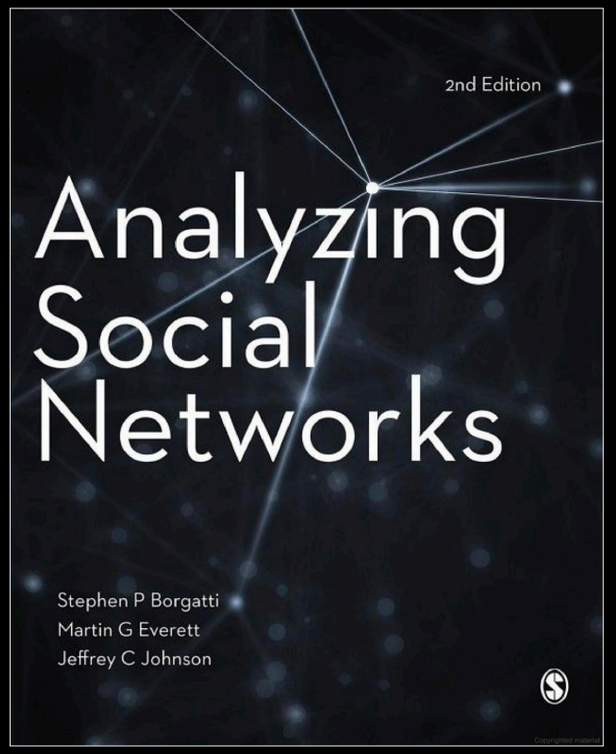


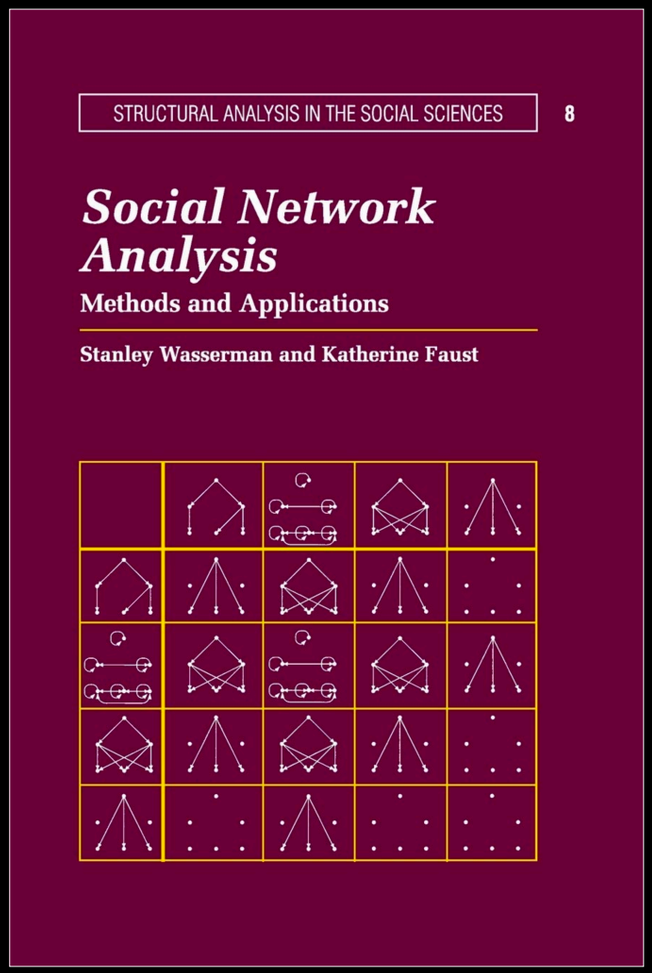
See: https://kmlawson.github.io/dh-tutorials/cytoscape.html

Some Resources

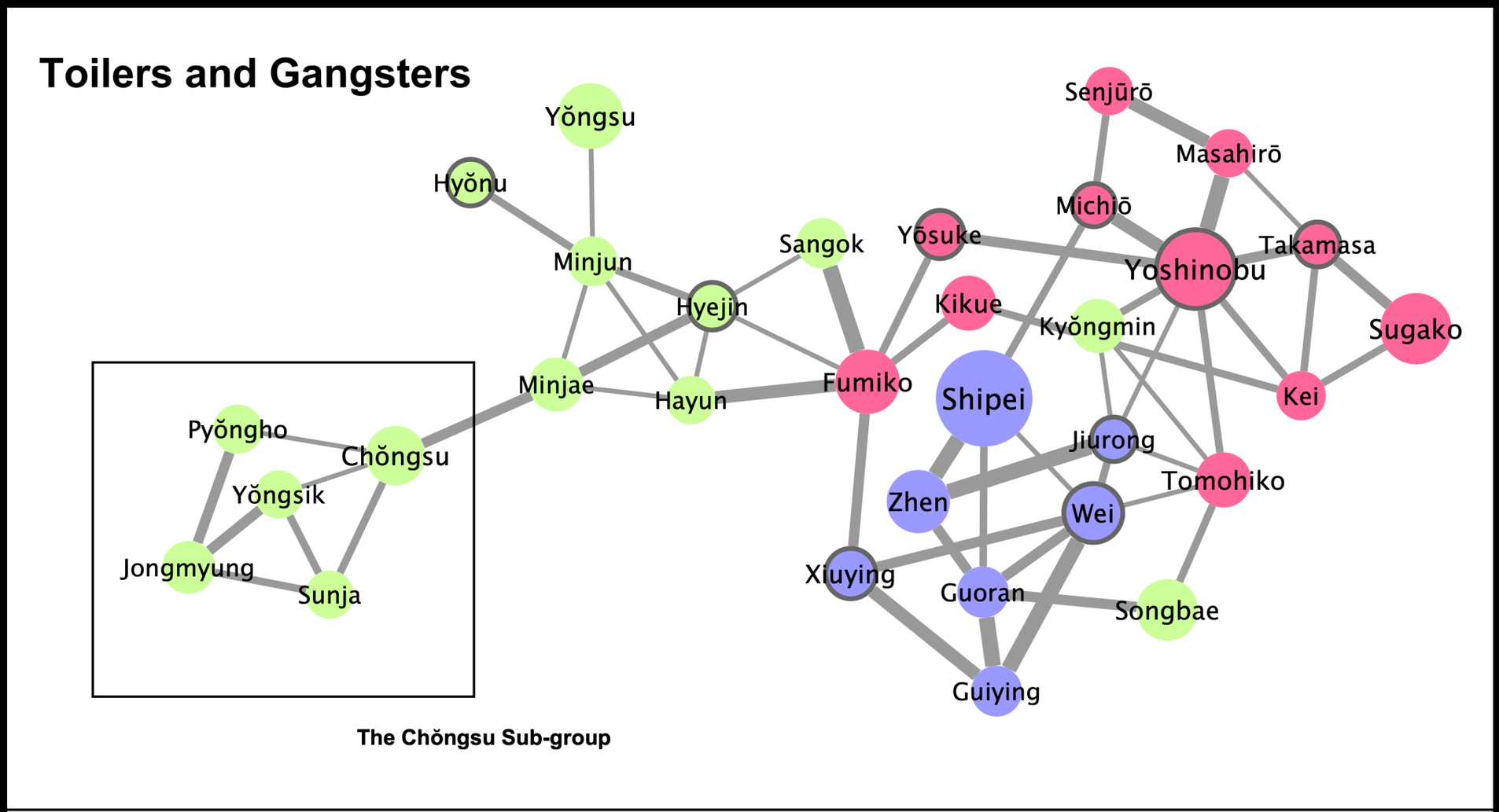








But learning how to visualise and explore a network does not require formal methods.



Many examples of network analysis *include no network graphs* among their outputs.

Many network graphs can serve a useful heuristic purpose during the course of exploring historical data, or as a helpful illustration in the final published work even *without formal statistical analysis.*

Unfortunately, many network graphs included in research outputs have *no heuristic value, and illustrate nothing* of much value.

When is it worth the time required?

- Visualising Networks Part 1: A Critique
- Should I do Social Network Analysis?
- If Everything is a Network, Nothing is a Network
- When Networks are Inappropriate

Beware of the hairballs…

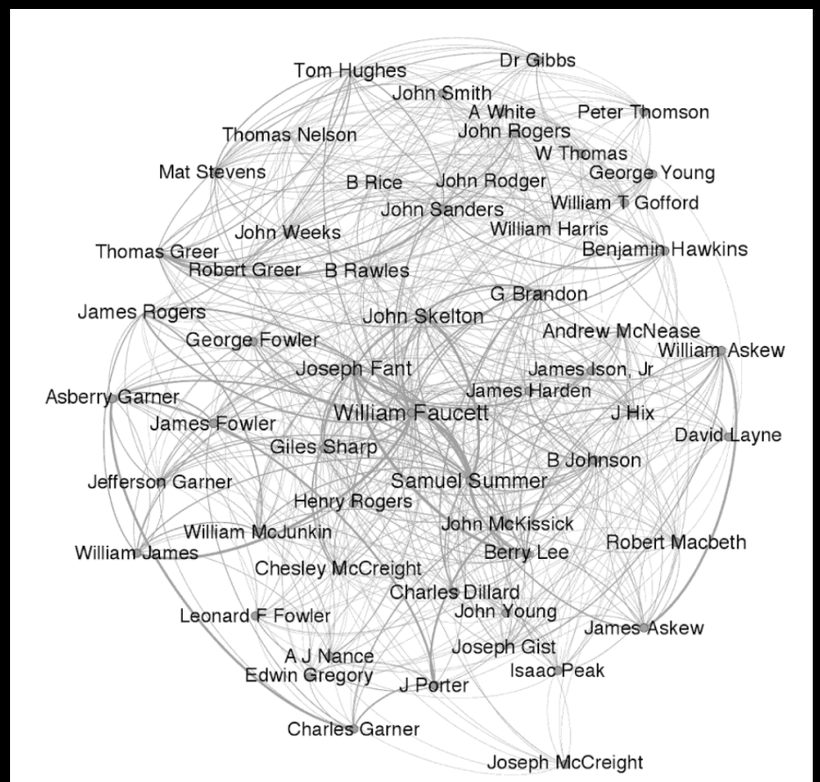


FIGURE 2. A visualization of the connections among William Faucett and his closest associates in Union, S.C. It is a dense subnetwork of all individuals who co-occur with at least thirty-four other members of the network. The k-core was derived in Pajek, and the visualization was rendered in Gephi, using the Fruchterman-Reingold algorithm followed by the Label Adjust algorithm for clarity.

When is it useful in historical research?

- Correspondence networks
- Institution Memberships
- Event participation
- Kinship structures
- Patronage networks
- Mentions in diaries or
  personal documents (egocentric network)
- Citation networks
- Using TEI encoded data

When can be it be used in historical research?

- Beyond Humans:
  - Publication Networks (Cooccurence networks, dependency networks)
  - Text Networks (between terms in texts)
  - Networks between places (with mutual ties to organisations or people and their flows)
  - Networks between organisations (with mutual membership)

Tools for Network Visualization

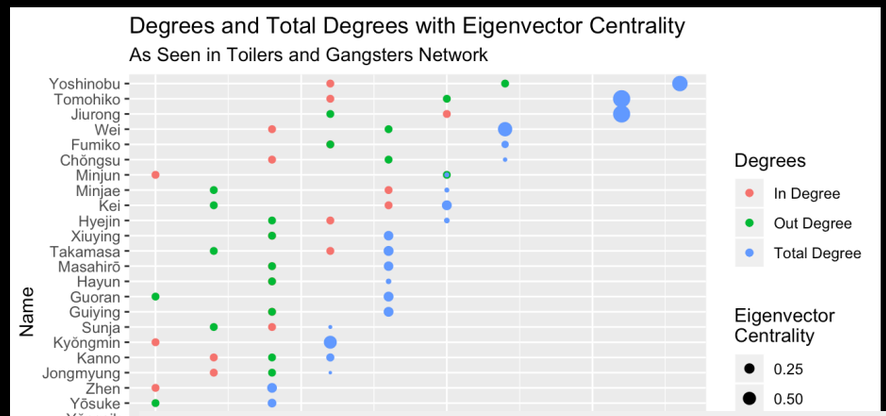
- Cytoscape (free, multi-platform, active development)
- Gephi (free, multi-platform, last update 2017)
- Pajek (free, Windows, active development)

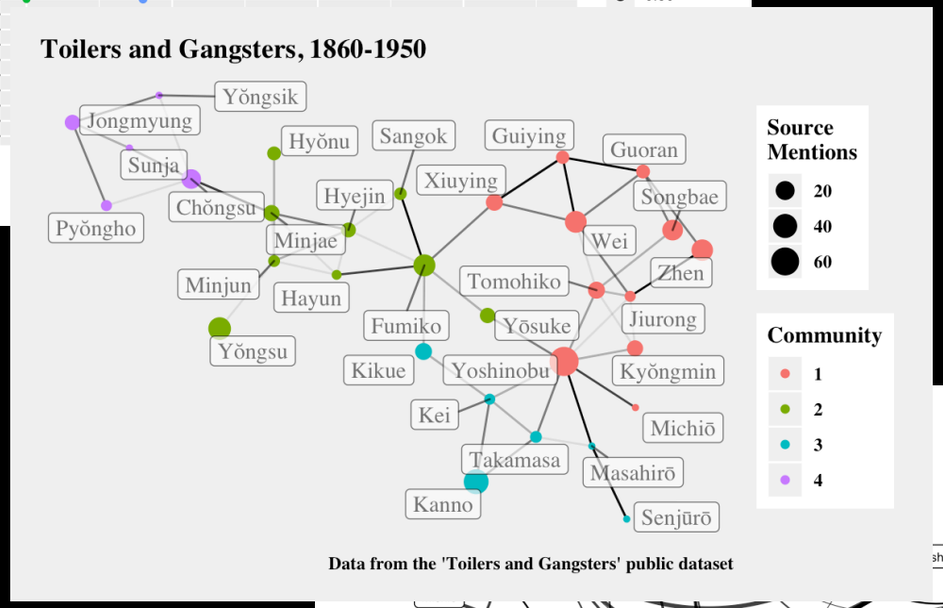
- Visone (free, multi-platform, active development)
- Palladio (web based, no longer actively developed)
- Orange (free, multi-platform, active development)
- VennMaker (free, multi-platform, last release 2018?)
- Socnetv (free, multi-platform, last release 2019)

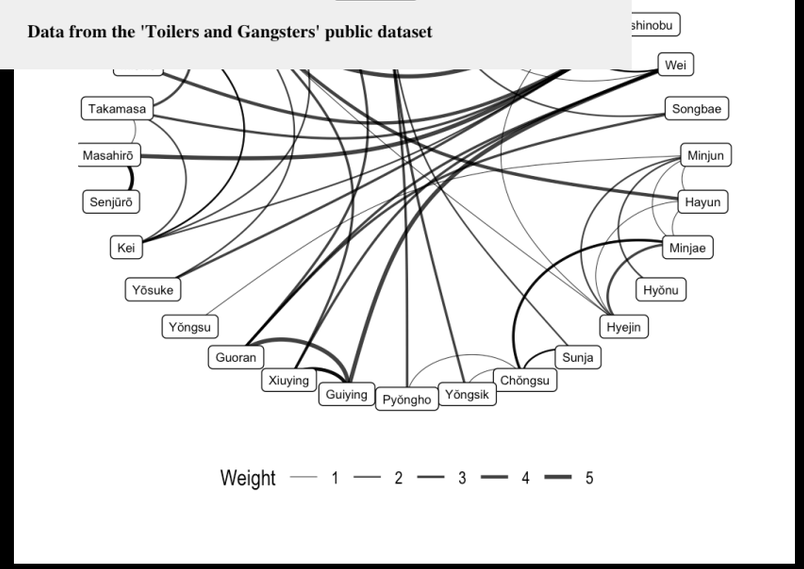
Toilers and Gangsters

Simple Network Visualization with R for Historians

Using techniques of the **ggraph**, **tidygraph**, **igraph**, and **visNetwork** packages.







https://kmlawson.github.io/dh-tutorials/network.html

Preparing Your Data

Node Table

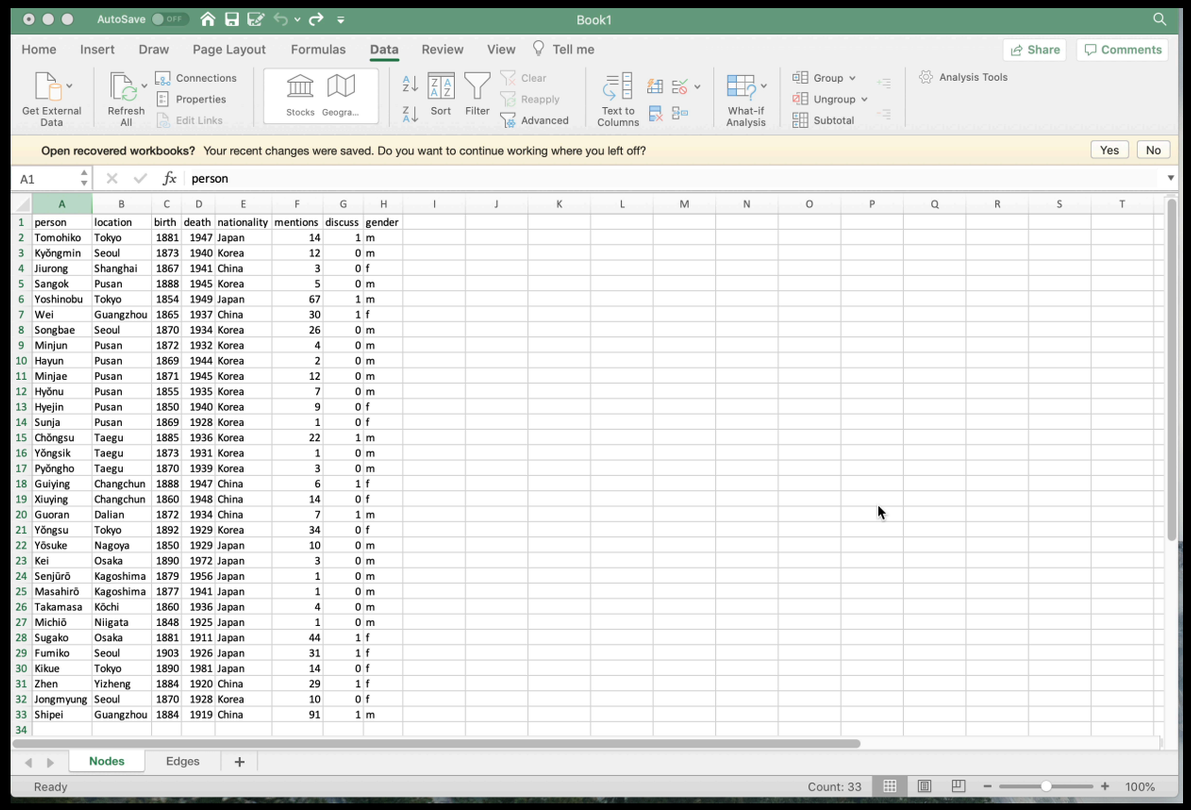
id	person	location	birth	death	nationality	mentions	discuss	gender
1	Chŏngsu	Taegu	1885	1936	Korea	22	1	m
2	Fumiko	Seoul	1903	1926	Japan	31	1	f
3	Guiying	Changchun	1888	1947	China	6	1	f
4	Guoran	Dalian	1872	1934	China	7	1	m
5	Hayun	Pusan	1869	1944	Korea	2	0	m
6	Hyejin	Pusan	1850	1940	Korea	9	0	f
7	Hyŏnu	Pusan	1855	1935	Korea	7	0	m
8	Jiurong	Shanghai	1867	1941	China	3	0	f
9	Jongmyung	Seoul	1870	1928	Korea	10	0	f
10	Kei	Osaka	1890	1972	Japan	3	0	m
11	Kikue	Tokyo	1890	1981	Japan	14	0	f
12	Kyŏngmin	Seoul	1873	1940	Korea	12	0	m
13	Masahirō	Kagoshima	1877	1941	Japan	1	0	m
14	Michiō	Niigata	1848	1925	Japan	1	0	m
15	Minjae	Pusan	1871	1945	Korea	12	0	m

Edge Table

idfrom	idto	from	to	kind	intensity	year_start	year_end
1	15	Chŏngsu	Minjae	3	3	1907	1921
5	15	Hayun	Minjae	3	1	1902	1943
12	8	Kyŏngmin	Jiurong	3	1	1895	1920
24	10	Takamasa	Kei	3	2	1910	1934
26	4	Wei	Guoran	3	3	1892	1920
30	25	Yoshinobu	Tomohiko	3	2	1898	1915
1	28	Chŏngsu	Yŏngsik	2	1	1919	1931
1	17	Chŏngsu	Pyŏngho	2	1	1901	1939
3	4	Guiying	Guoran	2	5	1905	1934
5	6	Hayun	Hyejin	2	1	1885	1940
6	15	Hyejin	Minjae	2	3	1885	1930
6	18	Hyejin	Sangok	2	1	1901	1945
8	30	Jiurong	Yoshinobu	2	1	1911	1928
12	25	Kyŏngmin	Tomohiko	2	1	1901	1925
13	24	Masahirō	Takamasa	2	1	1910	1918

Good Practice: Use a column with unique IDs for each agent as the ‘key’ when you use it in SNA software

If you haven’t been using ID numbers and want to add them after the fact, I’ve made a YouTube screencast showing you how to add them to your **node** and **edge tables** using the **LOOKUP()** function in Microsoft Excel.



	person	location	birth	death	nationality	mentions	discuss	gender
2	Tomohiko	Tokyo	1881	1947	Japan	14	1	m
3	Kyŏngmin	Seoul	1873	1940	Korea	12	0	m
4	Jiurong	Shanghai	1867	1941	China	3	0	f
5	Sangok	Pusan	1888	1945	Korea	5	0	m
6	Yoshinobu	Tokyo	1854	1949	Japan	67	1	m
7	Wei	Guangzhou	1865	1937	China	30	1	f
8	Songbae	Seoul	1870	1934	Korea	26	0	m
9	Minjun	Pusan	1872	1932	Korea	4	0	m
10	Hayun	Pusan	1869	1944	Korea	2	0	m
11	Minjae	Pusan	1871	1945	Korea	12	0	m
12	Hyŏnu	Pusan	1855	1935	Korea	7	0	m
13	Hyejin	Pusan	1850	1940	Korea	9	0	f
14	Sunja	Pusan	1869	1928	Korea	1	0	f
15	Chŏngsu	Taegu	1885	1936	Korea	22	1	m
16	Yŏngsik	Taegu	1873	1931	Korea	1	0	m
17	Pyŏngho	Taegu	1870	1939	Korea	3	0	m
18	Guiying	Changchun	1888	1947	China	6	1	f
19	Xiuying	Changchun	1860	1948	China	14	0	f
20	Guoran	Dalian	1872	1934	China	7	1	m
21	Yŏngsu	Tokyo	1892	1929	Korea	34	0	f
22	Yōsuke	Nagoya	1850	1929	Japan	10	0	m
23	Kei	Osaka	1890	1972	Japan	3	0	m
24	Senjūrō	Kagoshima	1879	1956	Japan	1	0	m
25	Masahirō	Kagoshima	1877	1941	Japan	1	0	m
26	Takamasa	Kōchi	1860	1936	Japan	4	0	m
27	Michiō	Niigata	1848	1925	Japan	1	0	m
28	Sugako	Osaka	1881	1911	Japan	44	1	f
29	Fumiko	Seoul	1903	1926	Japan	31	1	f
30	Kikue	Tokyo	1890	1981	Japan	14	0	f
31	Zhen	Yizheng	1884	1920	China	29	1	f
32	Jongmyung	Seoul	1870	1928	Korea	10	0	f
33	Shipei	Guangzhou	1884	1919	China	91	1	m

https://www.youtube.com/watch?v=qlqgFg-QzuQ

A typical workflow in Cytoscape and overview of the software…

Switching now to shared screen.
Simple version posted as screencast here:

https://www.youtube.com/watch?v=c2s578jEHHs

Visualising Social Networks with Cytoscape

12 October, 2020
Institute for Transnational & Spatial History - Institute Mondays

Konrad M. Lawson - kml8@st-andrews.ac.uk

Cytoscape tutorial overview posted here:
https://kmlawson.github.io/dh-tutorials/cytoscape-tutorial.html
Glossary, Links, and other Resources have been posted here:
https://kmlawson.github.io/dh-tutorials/cytoscape.html
These slides are available here:
https://kmlawson.github.io/dh-tutorials/cytoscape-slides.pdf
Cytoscape is available for download here:
https://cytoscape.org/