Digital Humanities
UNIVERSITY OF GEORGIA
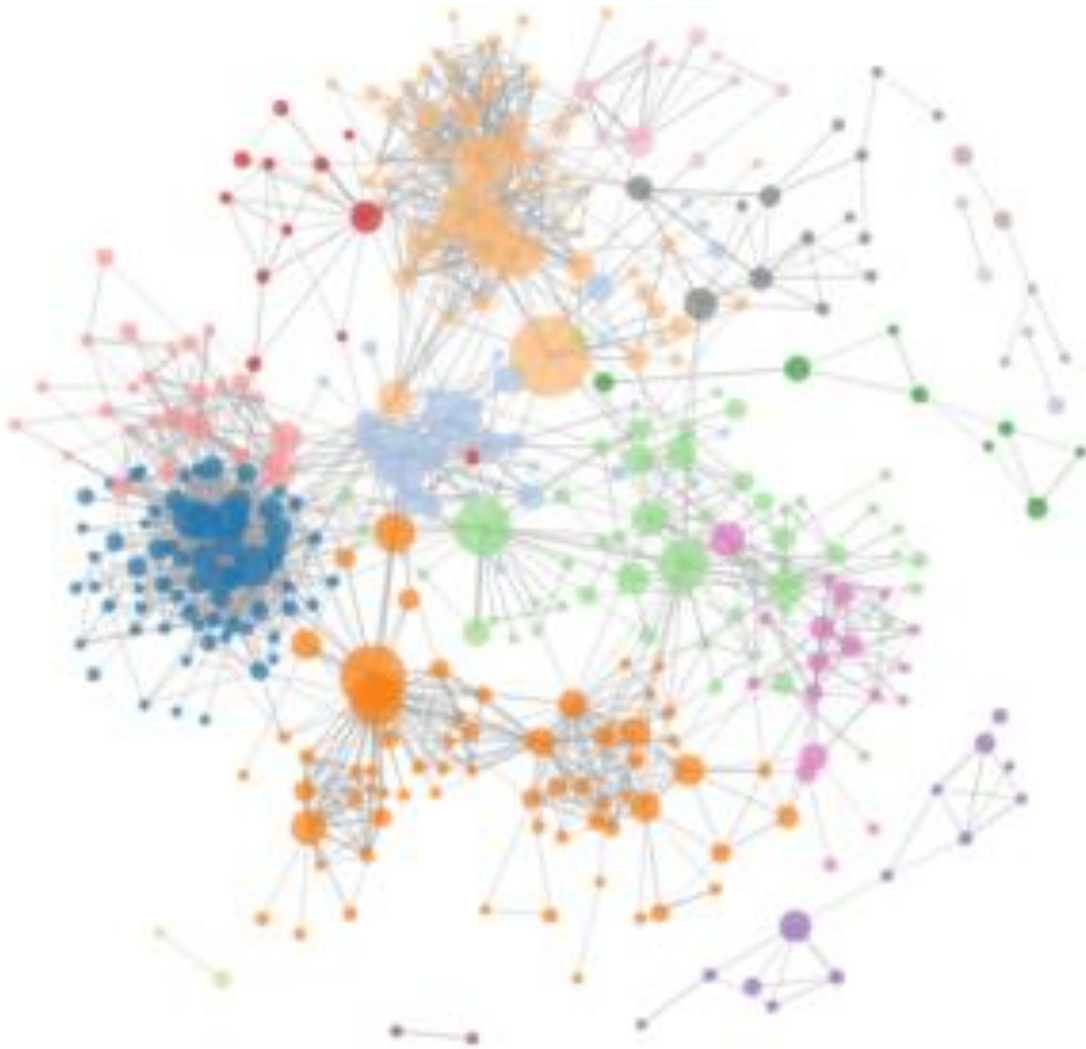
Data Visualization Series
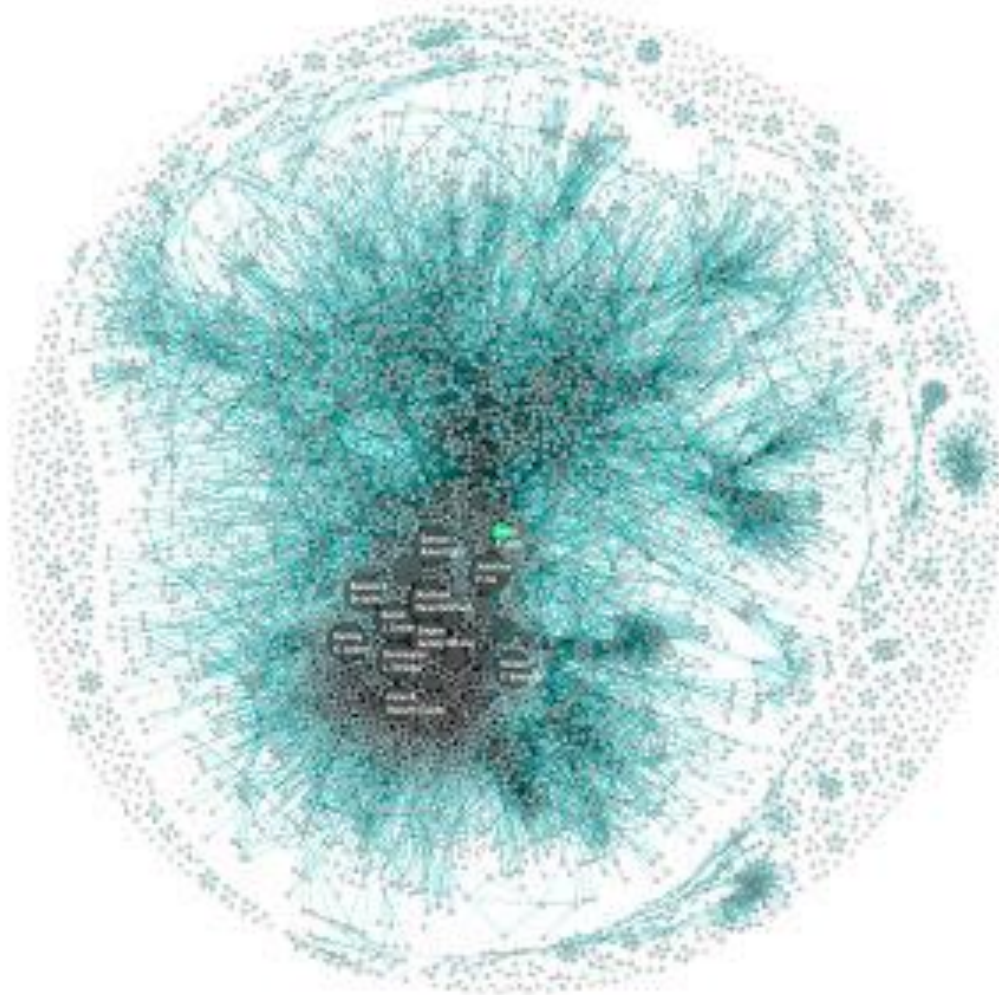
Network Analysis

KEIKO BRIDWELL

OCTOBER 20, 2021
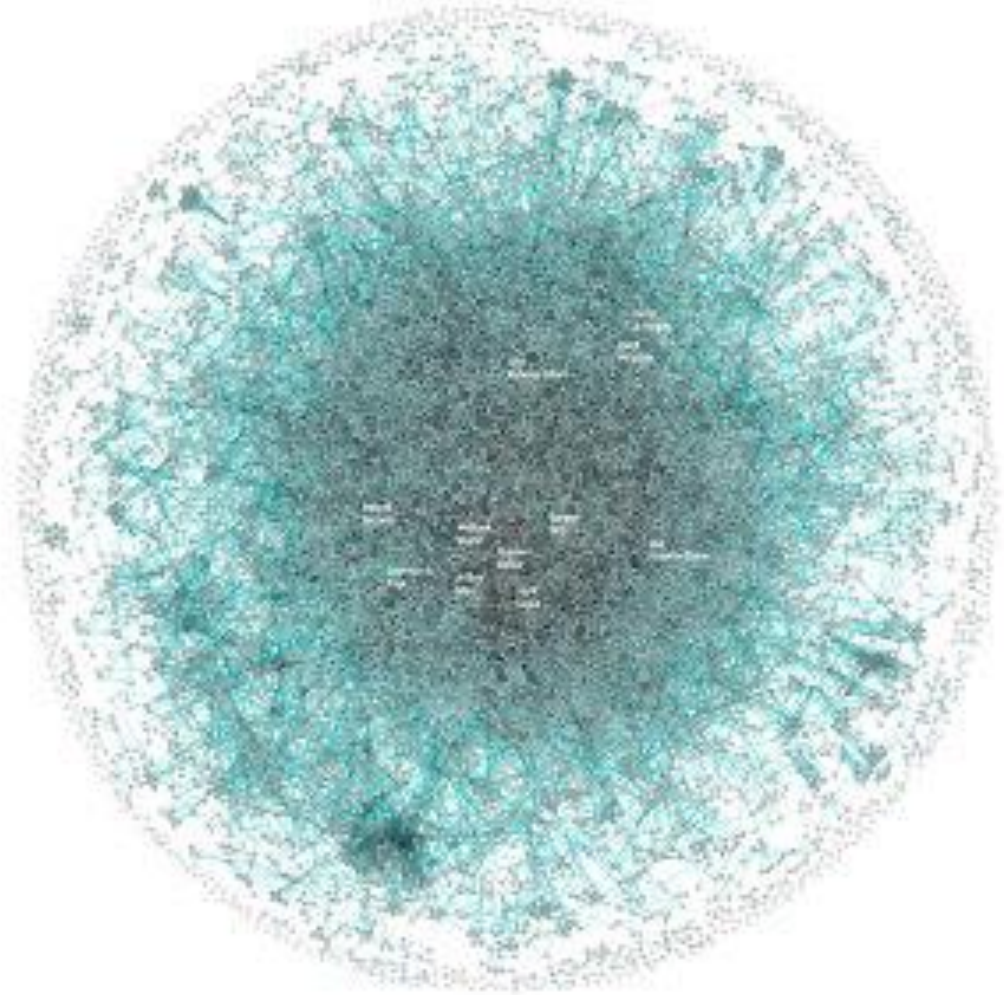
What is network analysis?

- Network = a map of relationships

- Composed of two elements
  - Nodes = objects of interest
  - Edges = connections between nodes

- Network analysis takes these maps to inform conclusions about the data
  - Visual patterns
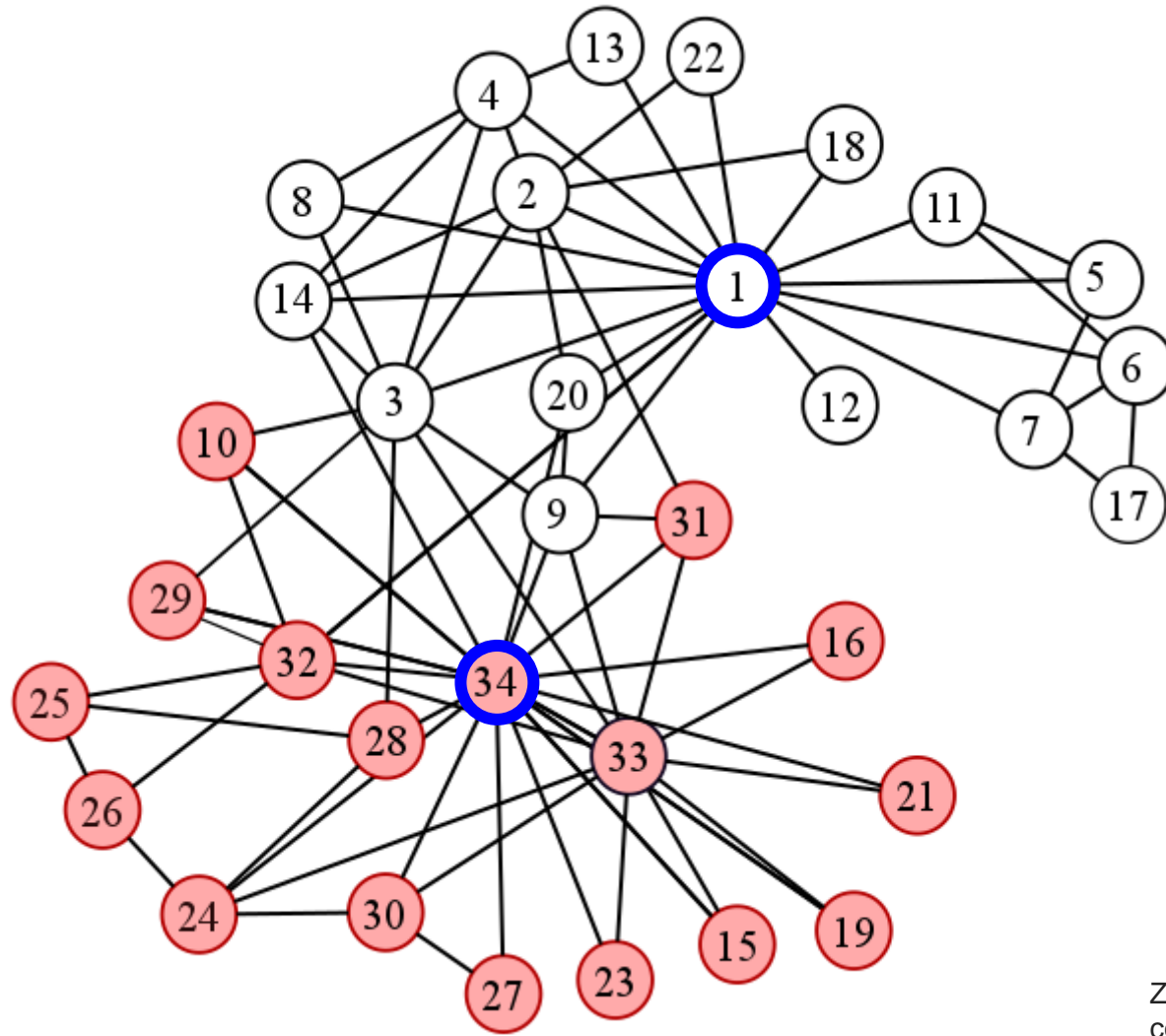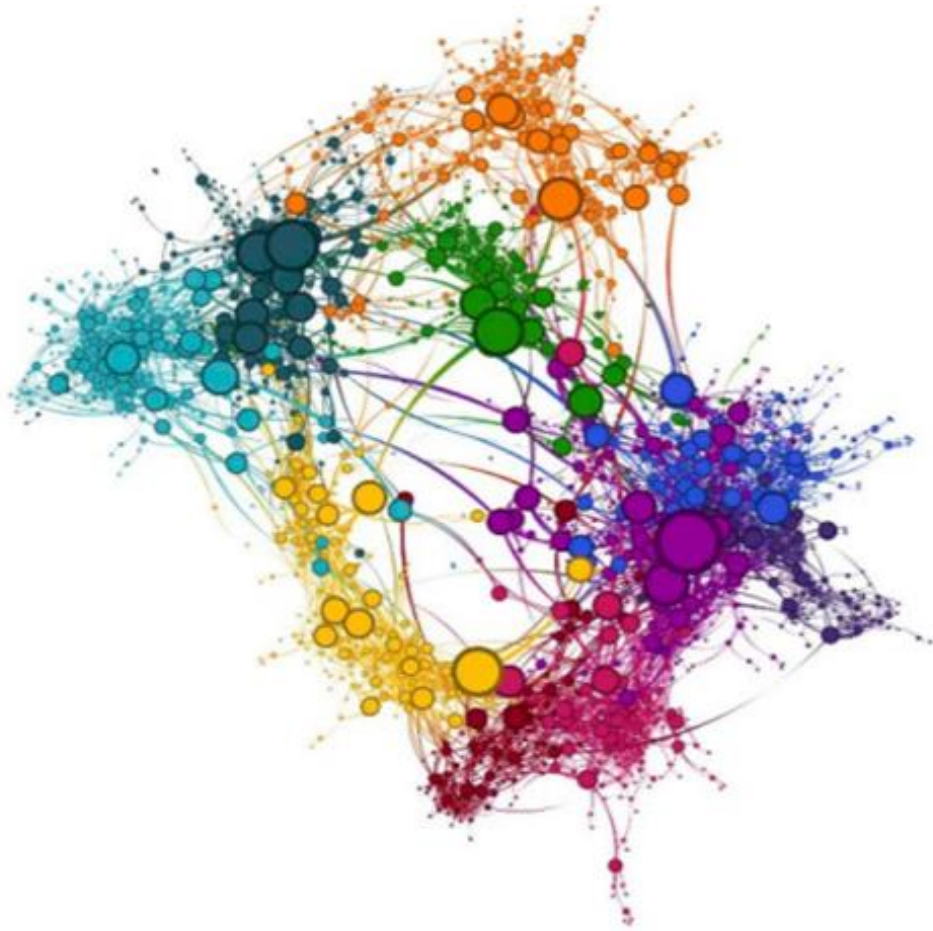 - Quantitative metrics

Apple

Google

Zachary's karate club (1977)

Zachary, Wayne W. "An information flow model for conflict and fission in small groups." *Journal of anthropological research* 33.4 (1977): 452-473.

Centrality

- An important metric in network analysis, especially for humanists

- Betweenness centrality
  - Measures how quickly information can flow between nodes
  - If a piece of information needs to flow through the network, how quickly does it get to everyone?
  - Centrality measures how much influence a particular point has on the network based on how quickly information can flow from that point to the rest of the network

Topics for network analysis

- Interpersonal relationships
  - https://mina-loy.com/maps/loys-social-network/

- Citations

- Spread of disease

- Engagement on social media

- Modeling brain activity

How to structure network data

- Two spreadsheets:
  - List of edges (required) – see Edges.csv
  - List of nodes (optional) – see Nodes.csv

- Edge list
  - Provides a **source** node and **target** node (ex. A → B)
  - Can also include **interaction type** (ex. family member, coworker, friend) or **weight** (ex. number of interactions, distance apart, etc.)

- Node list
  - Contains the names/IDs of all the nodes
  - Extra columns for other attributes (ex. gender, species, genre, etc.)

- Dataset: Toy Story (1995) character interactions from R package *networkdata*

Network graph resources

- Browser-based resources: ConnectTheDots, Palladio
  - Free online resources
  - Only need edge list: paste in data or upload CSV

- ConnectTheDots
  - Pros: automatically color-codes based on natural grouping of data, provides centrality scores
  - Cons: interactive element will not be present once downloaded (meaning node labels will not be visible), no additional information about nodes or interactions

- Palladio
  - Pros: displays label names directly on the graph, can size nodes based on how frequently they occur in the edge list, can drag nodes around for better positioning
  - Cons: grayscale, no information about nodes or interactions

Network graph resources

- (Free) downloadable software: <u>Cytoscape</u>, <u>Gephi</u>
  - ◦ Allow for much more customization!

- R packages
  - ◦ igraph: package specifically for creating network graphs
  - ◦ ggraph: extension of ggplot2 for network graphs
  - ◦ networkD3: JavaScript-based network graphs for online widgets (can be interactive!)
  - ◦ networkdata: contains hundreds of datasets of network data

Gephi

1. Download [here](#) (you will need Java installed!)

2. Open the nodes file first. Make sure it is identifying all the columns as strings/integers correctly

3. Next, open the edges file. For this, the ID columns MUST be named "Source" and "Target". Again, make sure all the columns are being correctly assigned as strings/integers. **Append to existing workspace!**
   ◦ Make sure to choose directed if you want arrows going from source to target, and undirected if you want no arrows.

4. Under "Layout" in the left panel, choose a layout from the drop-down list that works well with your data. (Force Atlas is usually most common!)
   ◦ Increase repulsion strength to make the graph larger
   ◦ Zoom in or out by clicking the arrow in the bottom right corner and going to the "Global" tab. Use the slider bar to zoom.

Gephi

5. Changing node **darkness** based on statistics:
   ◦ To make the nodes with the most connections darkest: Under "Ranking", choose "Degree".
   ◦ To make the nodes with the highest centrality darkest: Under the "Statistics" tab in the right panel, run the calculation for average path length. This will give centrality scores. Then, under "Ranking", choose "Betweenness centrality".
   ◦ Adjust the color scale as necessary so no nodes are too light or too pale

6. Changing node **size** based on statistics:
   ◦ Follow the above steps, but click in the top left panel, choose the icon with three circles instead of the one with the palette
   ◦ If you end up with a few giant nodes and lots of small ones, consider adjusting the scaling by clicking "Spline". The third option (smooth circular curve) tends to work well.
   ◦ If your nodes start to overlap, check the "Adjust by Sizes" box under Layout (for the Force Atlas layout).

Gephi

7. Color-coding the graph based on data clustering:
   ◦ In the "Statistics" panel, run "Modularity".
   ◦ Go to the color palette part of the "Nodes" tab under "Appearance" and choose "Partition".
   ◦ Choose "Modularity Class" from the drop-down menu.
   ◦ You can choose color schemes from "Palette" using "Generate", then click "Apply."

8. Color-coding the graph based on node characteristics:
   ◦ Go to the color palette part of the "Nodes" tab and choose "Partition".
   ◦ Choose the name of your characteristics column from the drop-down menu.

Gephi

9. Adding text to nodes:
   - Click the black "T" button at the bottom of the screen.
   - Resize text as necessary using the slider bar
   - Note: If the column with your labels is not already titled "Label", go to the arrow in the bottom right corner of the screen and go to the "Labels" tab. Click "Configure". Uncheck "Label" and choose the correct header from the list.

10. Sizing text based on node size:
    - Option 1: Using the drop-down menu at the bottom right corner, go to the "Labels" tab. For the "Size" drop-down menu, choose "Node size".
      - This option is less customizable.
    - Option 2: Go to the label size part of the "Nodes" tab under "Appearance" (the two T's) and choose "Ranking". Choose whichever metric you used to size your nodes.
      - Here, you can set minimum/maximum node sizes.

Gephi

11. To save the image, click the camera icon at the bottom of the screen, which will save the image as a png.

Resources

DigiLab Network Graphs webpage (with Cytoscape tutorial)

DigiLab Network Analysis Tutorial by Caleb Crumley

ggraph tutorial blog post

Gephi tutorials

Resources at UGA

- Linked-in learning tutorials

- Previous and upcoming tutorials at the DigiLab

- DigiLab Data Consultations

DigiLab Data Consultations

Photo by Isaac Smith on Unsplash

Consultations for Data Cleaning, Structuring, and Visualizing

Whether just starting your work, or trying to make sense of your research, schedule an appointment for our Data Consultations and bring your data (text, archival information, numerical data, etc.) for advice and guidance on your project. Expertise in corpus linguistics, Excel, and R, among other tools for data structuring and visualization.

To schedule an appointment visit:

DIGI.UGA.EDU/RESOURCES

Coming up next

Oct 27

Mapping
Geographical Data