Dynamic and Social Network Analysis

Lecture-2

Miray Kas
Bilkent University
Computer Engineering Department

Introduction to Social Network Analysis Software

Announcements

- TA for the class:
  - Gozde Yazici
  - Office Hours: Monday 12pm-2pm
  - gozde.yazici@bilkent.edu.tr
  - Contact online for reaching out
- Enter your details in the project partner sign up sheet
- First homework will be available later in the week (You will have a week to complete)
  - Submissions on Moodle

Social Network Analysis Software

- Many tools are available online, mostly free for students (Links available in the syllabus)
  - You can choose/suggest another tool that is not in the syllabus.
- You are free to choose whichever tool works with your system.
- Make sure to try them out and choose one or two you like, you will need it starting Homework-1 all the way to your final project!

Social Network Analysis Software

- Some are totally visual, some are script based.
- Some work with Windows only, some need programming environment installed etc.
- Most tools have sample datasets available
  - They usually have their own file format.
  - Almost all of them accept **csv** format and convert from that.
- Some of them are open source and free, some are paid
  - Paid software usually gives long enough trial for students to complete a semester.

Social Network Analysis Software

- Wikipedia has a comprehensive list, but even that is not complete.
- https://en.wikipedia.org/wiki/Social_network_analysis_software

Specialization: VOSviewer

- Focused on scientific networks
- Works for Windows, Mac, Linux
- Data collection works while in Bilkent network/VPN due to access to the scientific databases
    - VPN Setup Instructions if you have not done it
- Download VOSviewer
- VOSviewer Manual

VOSViewer Steps

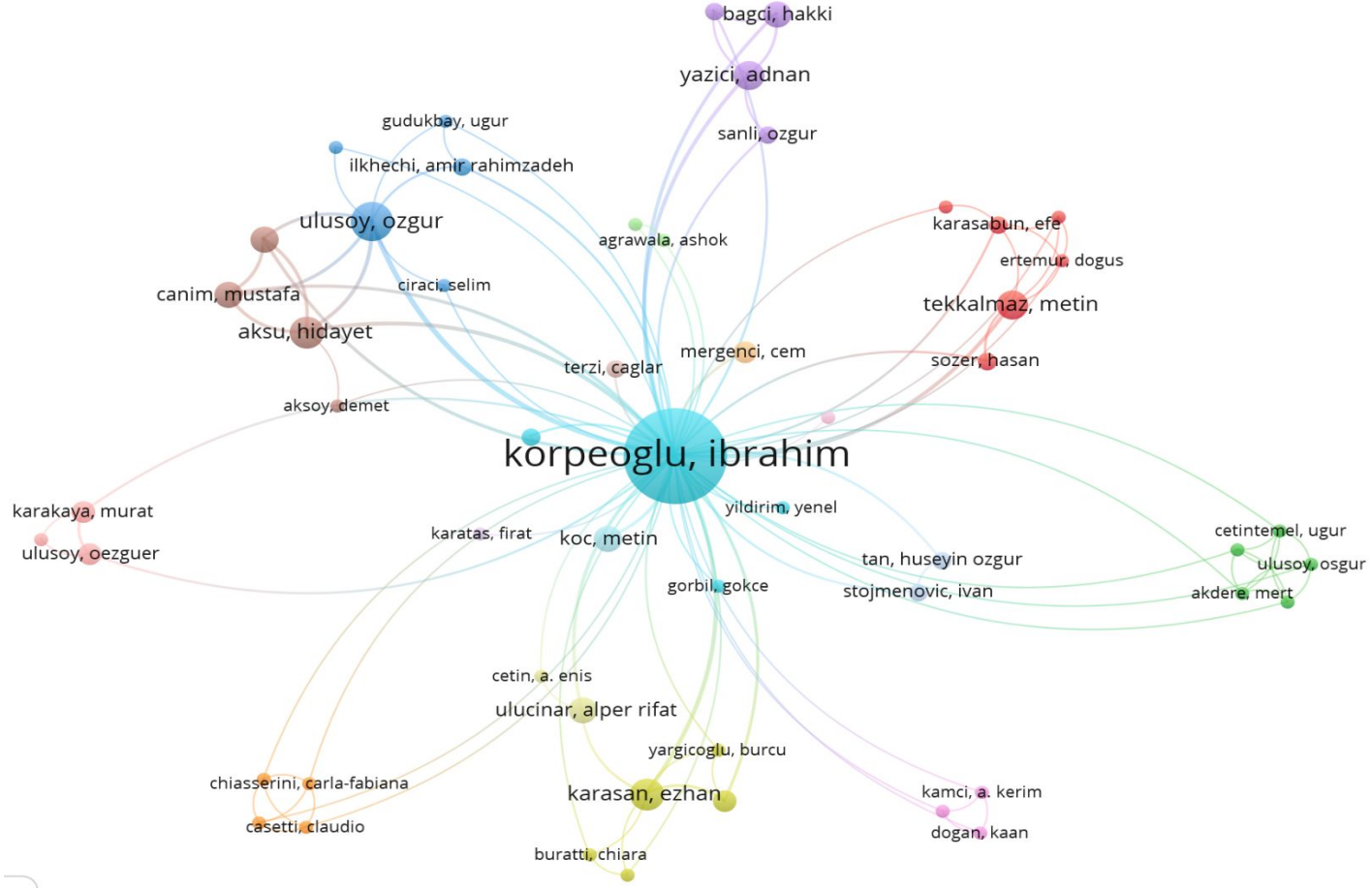
Collect Data from WebOfScience

1. Connect via Bilkent network
2. Go to http://www.webofscience.com/
3. Go to Authors, search Ibrahim Korpeoglu (Or anyone else you are curious about)
4. Export (500 is the upper limit at a time. If you have 2000 records, you will need to download 4 files)
5. Choose tab delimited, and full record and cited references options

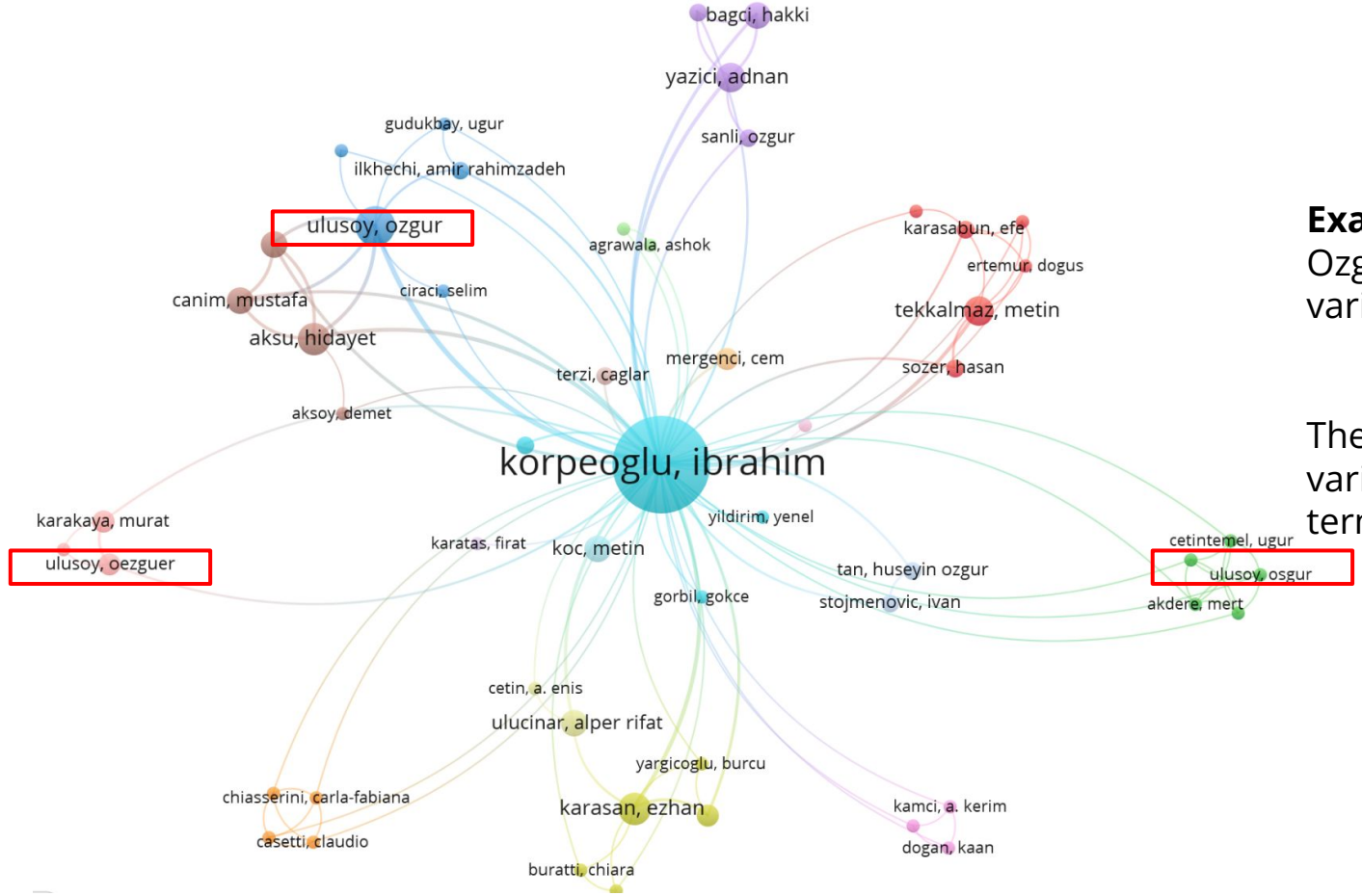
Use the data collected for analysis

1. Click on file
2. Click on Create
3. Choose “Create map based on bibliographic data”
4. Choose “Read data from bibliographic database files”
5. Choose WebOfScience option and point to your data file

Can you spot what's wrong with this?



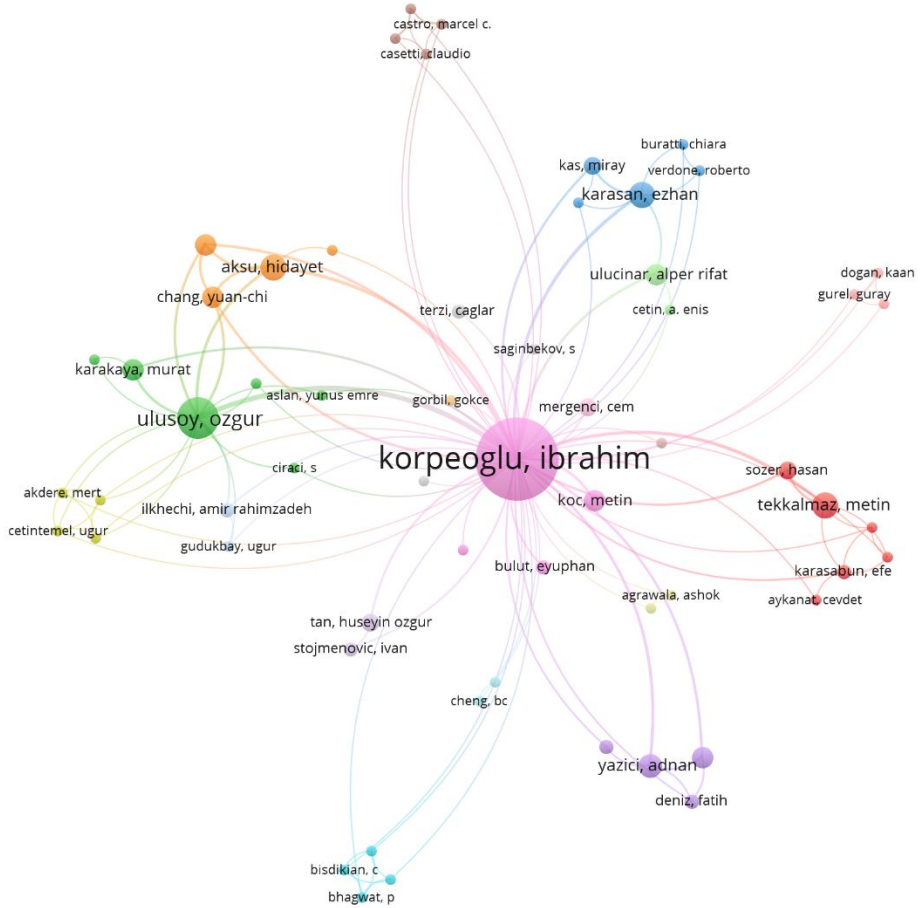
Can you spot what's wrong with this?



Example:
Ozgur Ulusoy has multiple variations

Thesaurus files can help fix variations of authors, terms etc.

Better/More Accurate Version



Example:
Closer collaborators are closer together

Before Exiting VoSViewer

- Get screenshots of your images/analysis
- Save your work as map and network files
- **Map file**
  - It has node information
  - Unique ID and Labels, and other information
- **Network file**
  - ID to ID connectivity information along with weight of the link
- These files can be slightly modified and used in other tools as needed
  - It usually needs minimal to no formatting

ORA

- 180-day trial version (enough for a semester)
- From Carnegie Mellon University, CASOS Group
- Runs on Windows
- Has very detailed metrics and reports
  - Stronger than Gephi
- Many datasets available: CASOS Tools: Network Analysis Data | CASOS
  - Classical benchmarks and CASOS group's own research
- Data Import Wizard
  - Many data types and APIs are supported

Getting raw data into ORA

Import Your network (We can continue with the data we generated from VoSViewer)

1. Go to Data Import Wizard from File menu at the top left
2. Choose “Import Excel or Text Delimited Files”
3. Choose table of network links, use your network file
4. Create a new MetaNetwork
5. Choose your file and Complete the steps

Import Your attributes (We can continue with the data we generated from VoSViewer)

1. Go to Your Metadata and Choose NodeSet > Editor > Attributes > Import Attributes
2. All columns will show up, check to include all attributes, choose correct types for them
3. Import
4. Go to Info tab, and change Display Nodes by Field to use labels

Customizing visualizations

Visualizations

- You can try different layouts (Pay attention to run, stop buttons)
- You can drag and customize the look to be more readable
- Links, nodes, colors are customizable
- You can color nodes based on clustering algorithms
    - For example: Node Appearance > Node Color > Newman Grouping
- From Actions menu, you can remove isolates and pendants

Generate Stylized Networks and Reports

Stylized Networks

- Artificial networks that follow particular network models
- Generate Networks > Create Stylized Networks > [Pick your network model]
  - Useful for people that do random graph and mathematical research
  - Useful if you want to compare your network's properties against well-known benchmarks

Report Generation

- You can analyze individual networks or combined networks
- You can generate reports in HTML format for easier reading, you can also get them presentation slides
- Click on your MetaNetwork
  - Click on Generate Reports
  - Check mark the features you would like to include in the reports
  - Point your reports to an existing folder
  - You can choose more than one type of output

Before you exit ORA

Save your work !!!

- Save your MetaNetwork file!
  - Do this even if you generated and saved your analysis reports
  - This will help you not to deal with import steps again
  - It will also save your corrections, your combined files, generated metrics, etc.
- Save your visualizations
  - You can spend a good amount of time prettifying visuals

SocNetV

- Runs on Windows, Mac OS X and Linux (You can download from Downloads - Social Network Analysis and Visualization Software)
- You can import from different data formats
  - Like most tools edge list works here too.
- If your network is small and you need pretty prints of adjacency matrix etc., this tool is helpful.
- Easy to create some of the simulated network models.
- Easy to create known datasets (You can export them as matrices)
  - These datasets refer to highly cited classical datasets
  - Many papers use them as benchmark.
  - For example, if you are proposing a partitioning algorithm and cannot show it to work on Karate club data, it is likely not good.

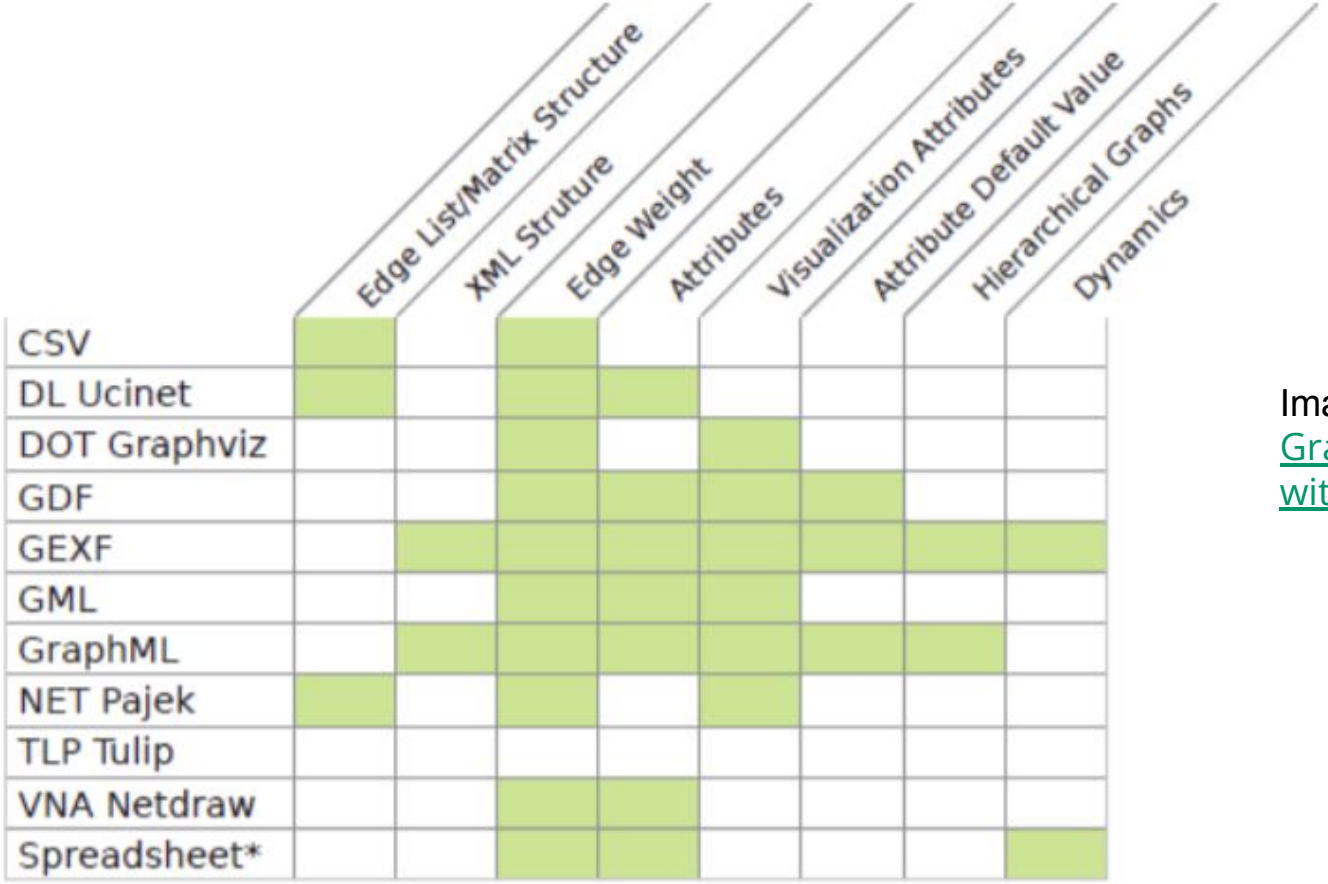
Specialization: SocioViz

- Very basic tool that allows queries on Twitter/Facebook
  - Good for marketing analysts, trend analysts, social media managers
  - Facebook features are experimental
- Create a free account (Caveat: Free access is limited)
  - More sophisticated queries can be run in paid form.
- SocioViz
  - You can run a query with #hashtag or from:username
  - Export all data and get visualizations
  - Gexf files generated by SocioViz can be used in Gephi
- You can download much more data from Twitter API directly for free.

Gephi

- Open source and free
  - Runs on Windows, Mac OS X and Linux
  - You can download from https://gephi.org/
- Many plugins available (Gephi Plugins)
  - They have well published APIs, you can write your own plugin (Gephi developers)
  - They have an active developer facebook group
- So many ways of getting started with Gephi
  - Get Your Data into Gephi: A Quick and Basic Tutorial
  - Gephi Tutorial Quick Start
  - import data from CSV into Gephi

Gephi Data Types



	Edge List/Matrix Structure	XML Struture	Edge Weight	Attributes	Visualization Attributes	Attribute Default Value	Hierarchical Graphs	Dynamics
CSV	■		■					
DL Ucinet	■		■	■				
DOT Graphviz			■		■			
GDF			■	■	■	■		
GEXF		■	■	■	■	■	■	■
GML			■	■	■			
GraphML		■	■	■	■	■	■	
NET Pajek	■		■		■			
TLP Tulip								
VNA Netdraw			■	■				
Spreadsheet*			■	■				■

Image Source:
[Graph Visualization with Gephi]

LesMiserables Dataset: gefx and gml

- GEFX version - Sample data provided by Gephi [Gephi Datasets]
  - Character social network from the novel
  - We will also download the GML version to go through steps of how visualization works

- If you want to watch demo/tutorial videos
  - How to get the beautiful graphs generated [Tutorial]
  - A more detailed Gephi tutorial [Link]

Editing Visualization in the Overview Page

- **Appearance**
  - Size, color, label
  - For partition and ranking, initially you won't have many options
  - Run some statistics to get metrics generated.
- **Statistics**
  - Run modularity to get partition info generated.
  - Run network diameter to get betweenness and a couple of other centrality metrics generated.
  - **Caveat:** Menu namings are not clear, you need to know what is where.
- **Filters**
  - Drag your criteria to Query window
  - E.g. Filter > Topology > Degree Range ⇒ Query

Editing Visualization in the Overview Page

- **Layouts**
  - ForceAtlas is commonly used in demos
    - Make the repulsion parameter 10000 for a spread out graph
  - Each algorithm has its own parameters, you can adjust based on what you see
  - Stop the layout algorithm when it is in reasonable shape
  - **Caveat:** Some algorithms are computationally expensive
    - Gephi will ask for memory increase or may even freeze Gephi

Further Editing Visualization in the Preview Page

- Further editing such as making the edges curvy
- May take a while to load, reset zoom and refresh.
- Export and save as SVG/PDF/PNG

Importing Raw Data into Gephi

- As usual, pure edge list idea would work
- Create 2 files:
  - One for nodes, one for edges
- Pay attention to the column names (e.g. headers)!
  - If you import your own data, those headers are how Gephi understands
  - **Edges:** [Source, Target, Type, Weight]
  - **Nodes:** [Id, Label]

- Data laboratory tab has the detailed data both for nodes and edges
  - Data Laboratory > Nodes > Import Spreadsheet
  - Make sure to choose Node table when importing
  - Similar steps for Edge table

Gephi Twitter Streaming Importer

- Gephi also has a Twitter streaming functionality
  - It is a separate plugin you can choose to install [Tutorial]
  - If you did not install initially, Tools > Plugins > Available Plugins
- Get a free Twitter developer account for students
  - https://developer.twitter.com/
  - Latest Academic Research topics [*Twitter academic developers forum*]
  - Provide your keys/tokens
- Load your query from a file or enter comma separated criteria
- Choose Network type:
  - Full Twitter Network : This will represent all entities (User, Tweet, Hashtags, URL, Media, Symbol etc...) as a graph.
- Connect (at some point disconnect, you will see node/edge count updating on the top right)

Next Lecture:

Representing Social Network Data
(Graphs and Matrices)