Exploratory Social Network Analysis with Gephi

Thomas Ebermann
twitter.com/plotti
Code unter: bit.ly/zhsport

Process

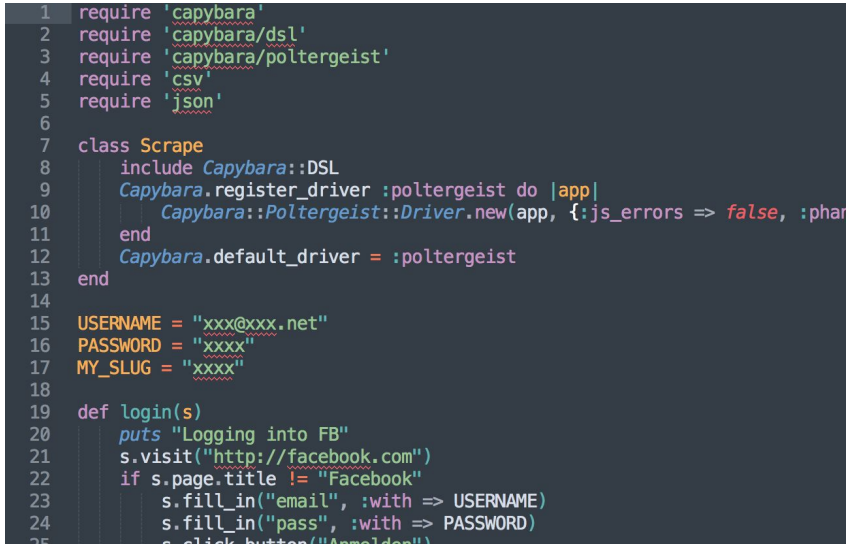
1. Import Data with Scraper
    a. Alternatively use the sample.gexf file
2. Gephi
    a. Open
    b. Layout
    c. Ranking (Degree)
    d. Statistics
    e. Ranking (Betweenness)
    f. Layout (Size Adjust)
    g. Labels
    h. Community detection
    i. Filter
    j. Label Adjust
    k. Preview
    l. Export

3. Use it out with your data or try these

- http://snap.stanford.edu/data/
- http://cdg.columbia.edu/cdg/datasets
- ...

Importing your Data with the scraper

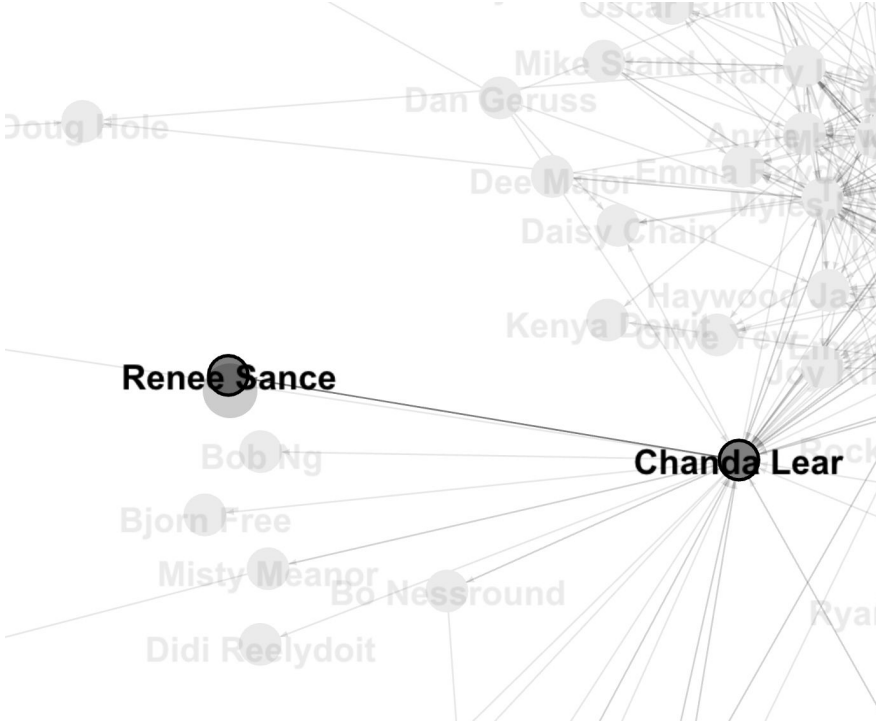
1. Go to https://github.com/plotti/gephi-workshop
2. git clone git@github.com:plotti/gephi-workshop.git
3. Install ruby (see README.md)
4. Install libraries (see README.md)



```
require 'capybara'
require 'capybara/dsl'
require 'capybara/poltergeist'
require 'csv'
require 'json'

class Scrape
    include Capybara::DSL
    Capybara.register_driver :poltergeist do |app|
        Capybara::Poltergeist::Driver.new(app, {:js_errors => false, :phar
    end
    Capybara.default_driver = :poltergeist
end

USERNAME = "xxx@xxx.net"
PASSWORD = "xxxx"
MY_SLUG = "xxxx"

def login(s)
    puts "Logging into FB"
    s.visit("http://facebook.com")
    if s.page.title != "Facebook"
        s.fill_in("email", :with => USERNAME)
        s.fill_in("pass", :with => PASSWORD)
```

Alternatively: Use the sample data

- I’ve prepared some anonymized sample data that you can use to follow along in the workshop.
- You won’t have the same kind of fun, when it is not your friends networks that you are studying.
- Open sample.gexf

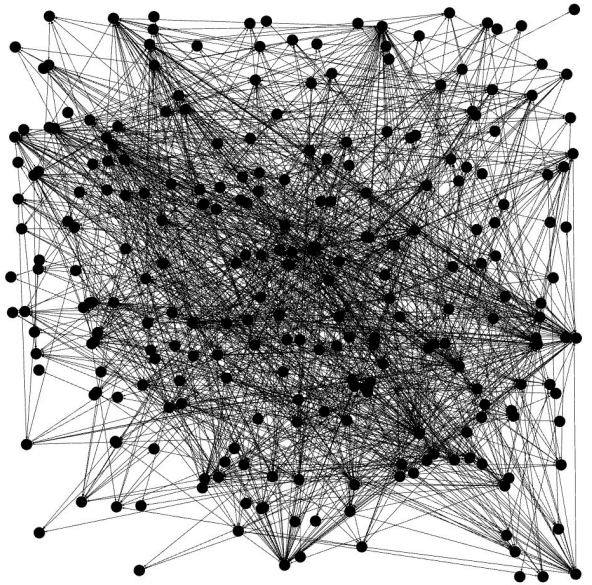


Gephi

- Gephi is an open-source network analysis and visualization software package.
- Envisioned as providing "easy and broad access to network data", it's advertised as being "Like Photoshop for graphs."
- Gephi has been used in a myriad of research projects in the university, journalism and elsewhere.

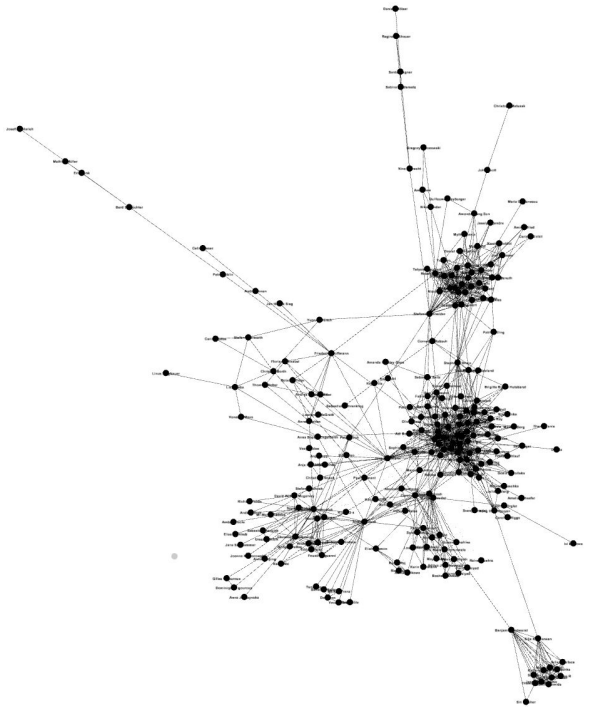
Gephi: 1. Open

- From File menu select Open and then select the .gexf file you saved from our Scraper or the sample.gexf file.
- At first it looks like a big hairball, so we'll change the layout to make some sense of the connections



Gephi: 2. Layout

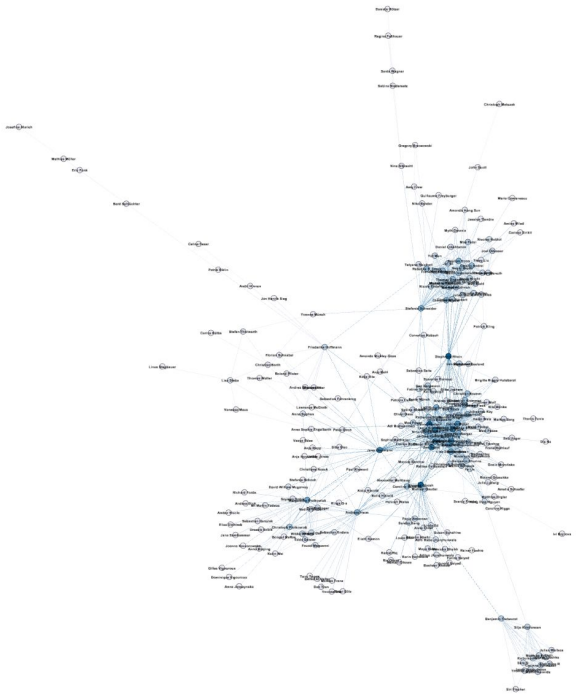
- From the Layout module on the left side chose Force Atlas* from the Dropdown Menu, then click run
- Force atlas makes connected nodes attract each other, while unconnected nodes are pushed towards the periphery
- Click stop when it seems that the layout has converged towards a stable state



*For graphs with a large number of nodes or edges rather chose YifanHuLayout

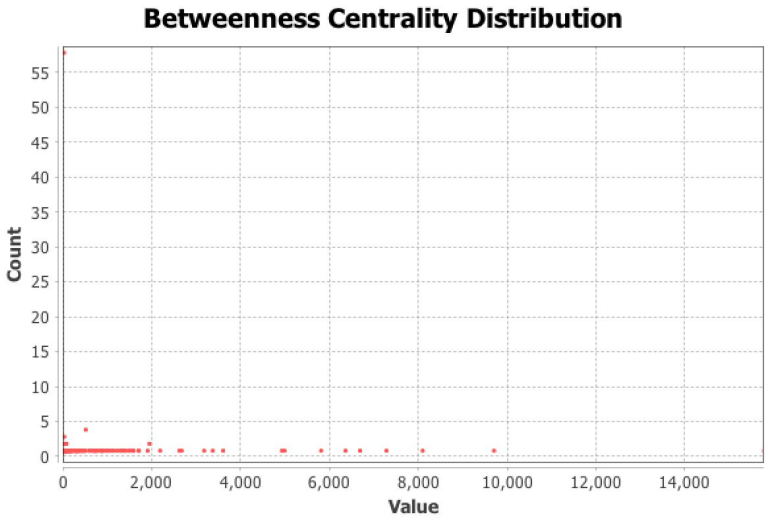
Gephi: 3. Ranking (Degree)

- Chose the Ranking-Nodes Tab in the top left module and chose Degree from the dropdown menu
- Degree = number of connections
- Hover your mouse over the gradient bar, then double click on each triangle to choose a color for each side of the range
- Try to use bright colors for the highest degree and dark for lowest
- Click apply



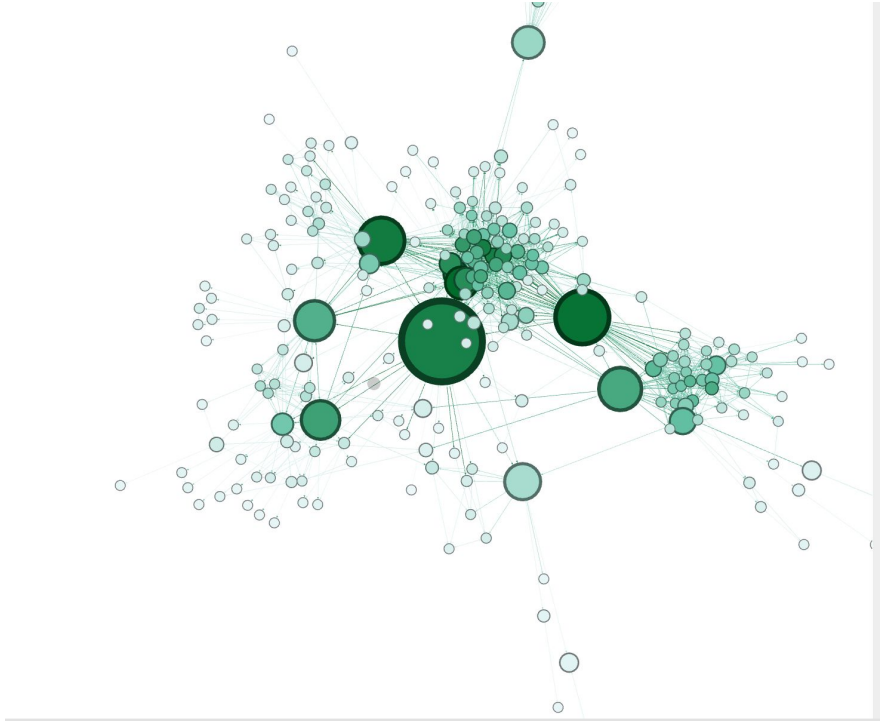
Gephi: 4 Statistics

- Click the Statistics tab in the top right module
- Click Run next to Average path length
- Chose directed from Popup Menu
- Click close when the graph reports shows up



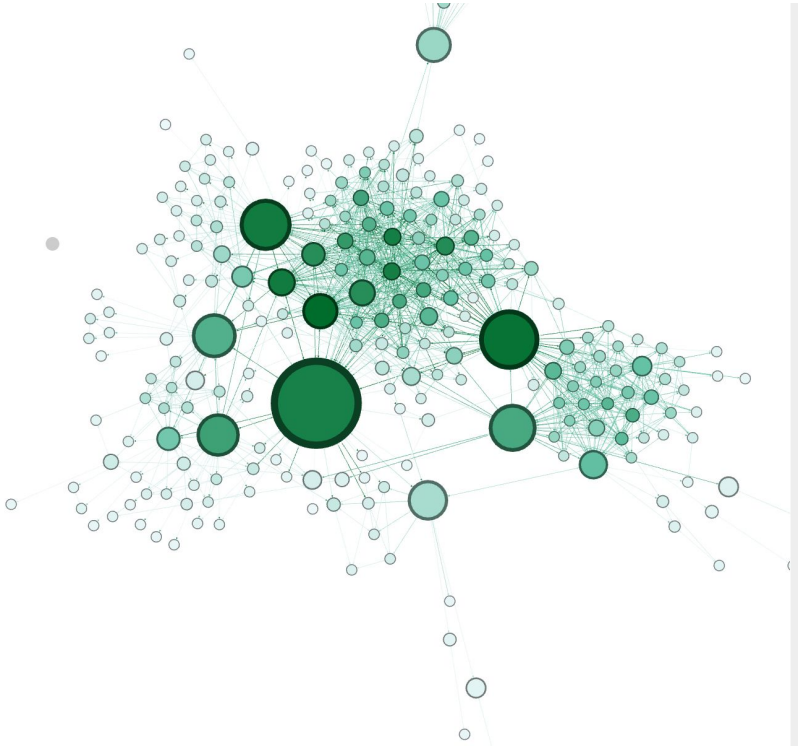
Gephi: 5. Rank (Betweenness)

- Return to Ranking in the top left module and click Chose a rank parameter from the dropdown
- Chose Betweeness Centrality from the dropdown menu
- Click on the icon for size, instead of color
- Set min size to 10 and max size to 80 (experiment a little)
- Click Apply



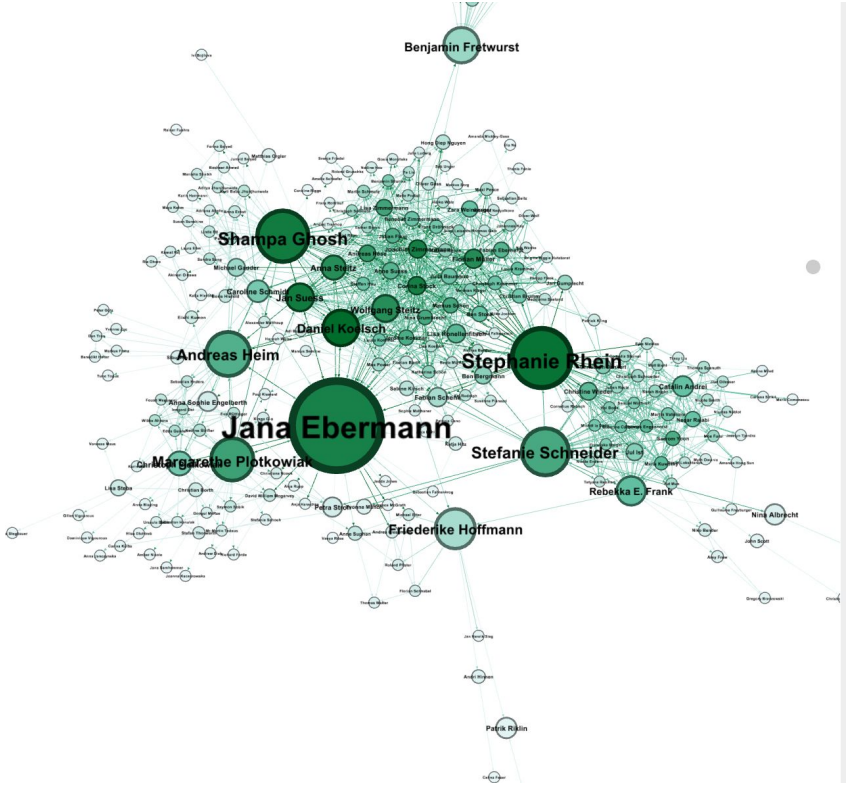
Gephi: 6. Layout

- To keep the larger nodes from overlapping smaller ones, go to the Layout tab and check the Adjust by sizes box
- Click Run and then Stop



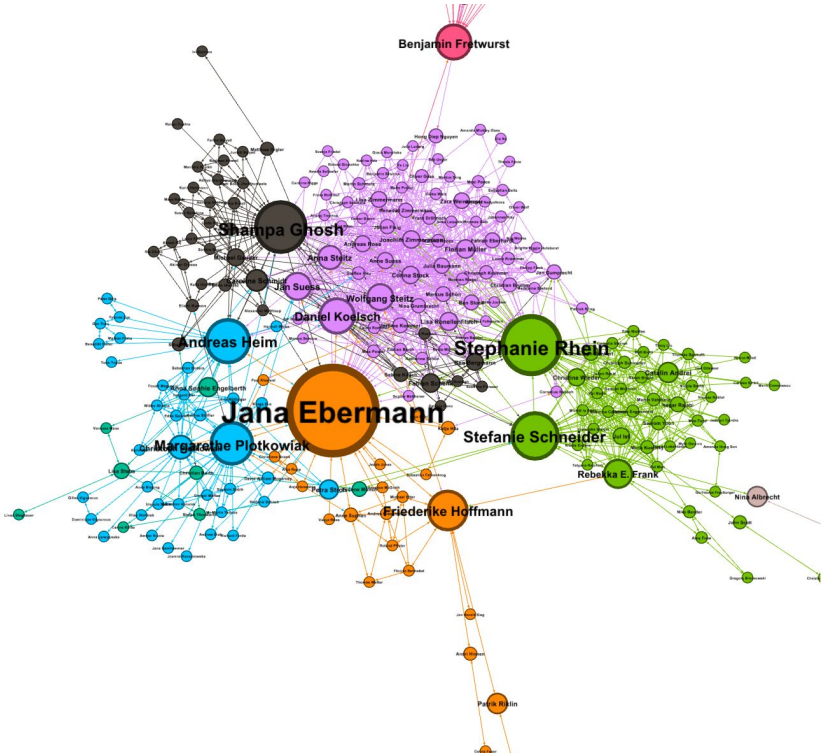
Gephi: 7. Labels

- Click the bold black T in the toolbar at the bottom of the window to turn labels on
- Click the black letter A in the same toolbar to select the Size Mode for the labels, and choose the node size option
- Use the slider on the right to adjust the size
- You can also change the font style by clicking next to the slider



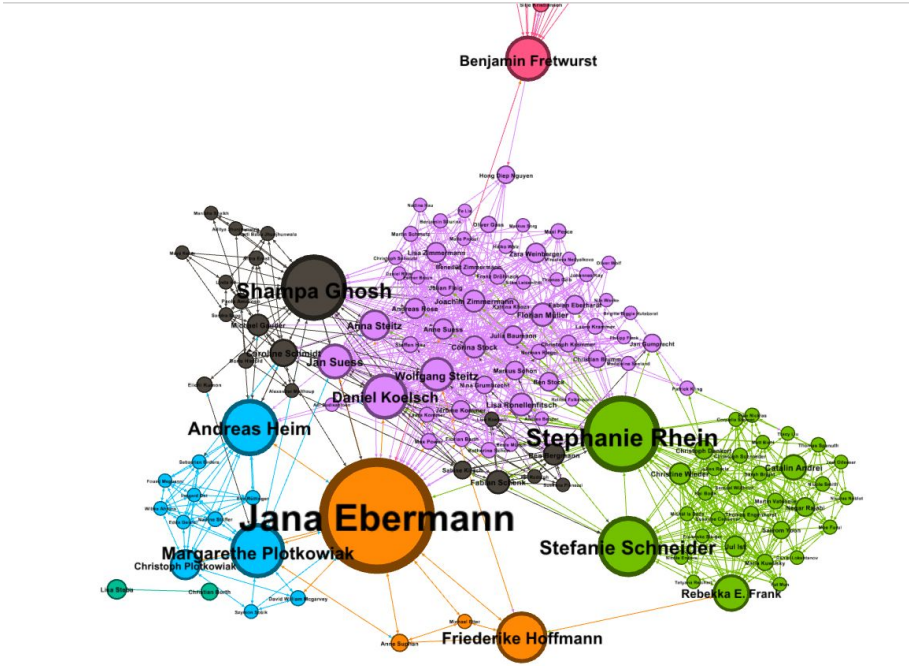
Gephi: 8. Community Detection

- Go back to the statistics tab on the right and click Run next to Modularity
- Check randomize and click OK
- Go to the partition tab in the top left module and click the refresh arrow
- Choose modularity class from the dropdown menu
- Right click to randomize colors
- Click Apply



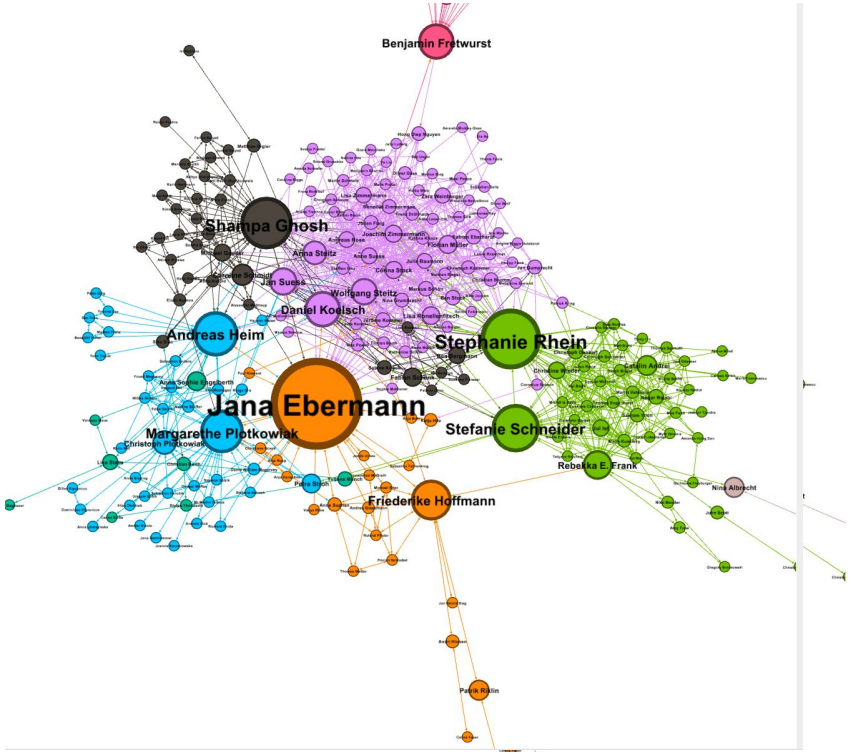
Gephi: 9. Filter

- Go to Filters in the top right module and open the Topology Folder
- Drag the degree range to the box below ("Drag filter here")
- Click on Degree Range to open the Parameters
- Click on the "0" and change it to a slightly higher value
- This removes the nodes that are not connected to many other nodes
- Click Filter



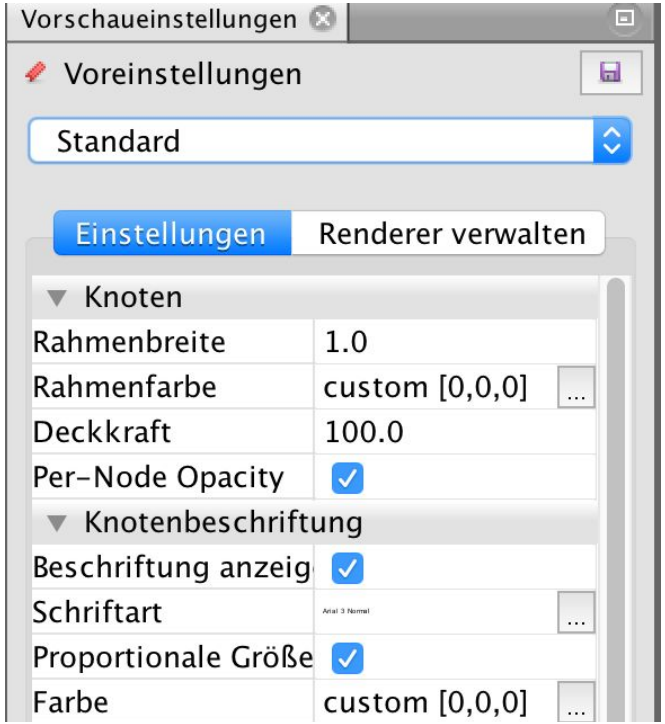
Gephi:10. Label Adjust

- Go to the Layout module on the left
- Chose label Adjust layout to make the labels not overlapping
- Click Run and then Stop



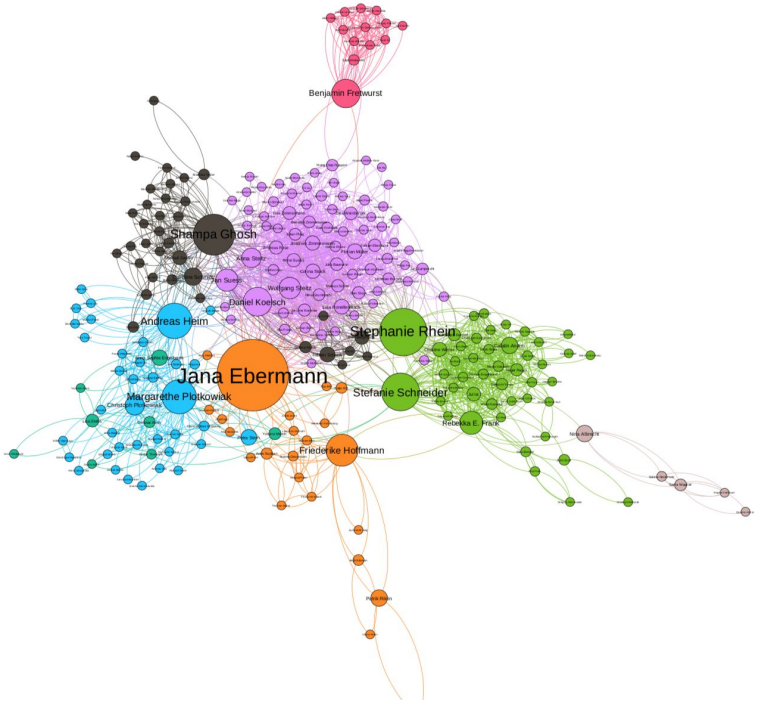
Gephi: 11. Preview

- At the very top click on the Preview tab
- Under Node, check the box "Show Labels"
- Click Refresh at the bottom, and choose your label font
- Play around with the options until you like your graph (Don't forget to click refresh every time)

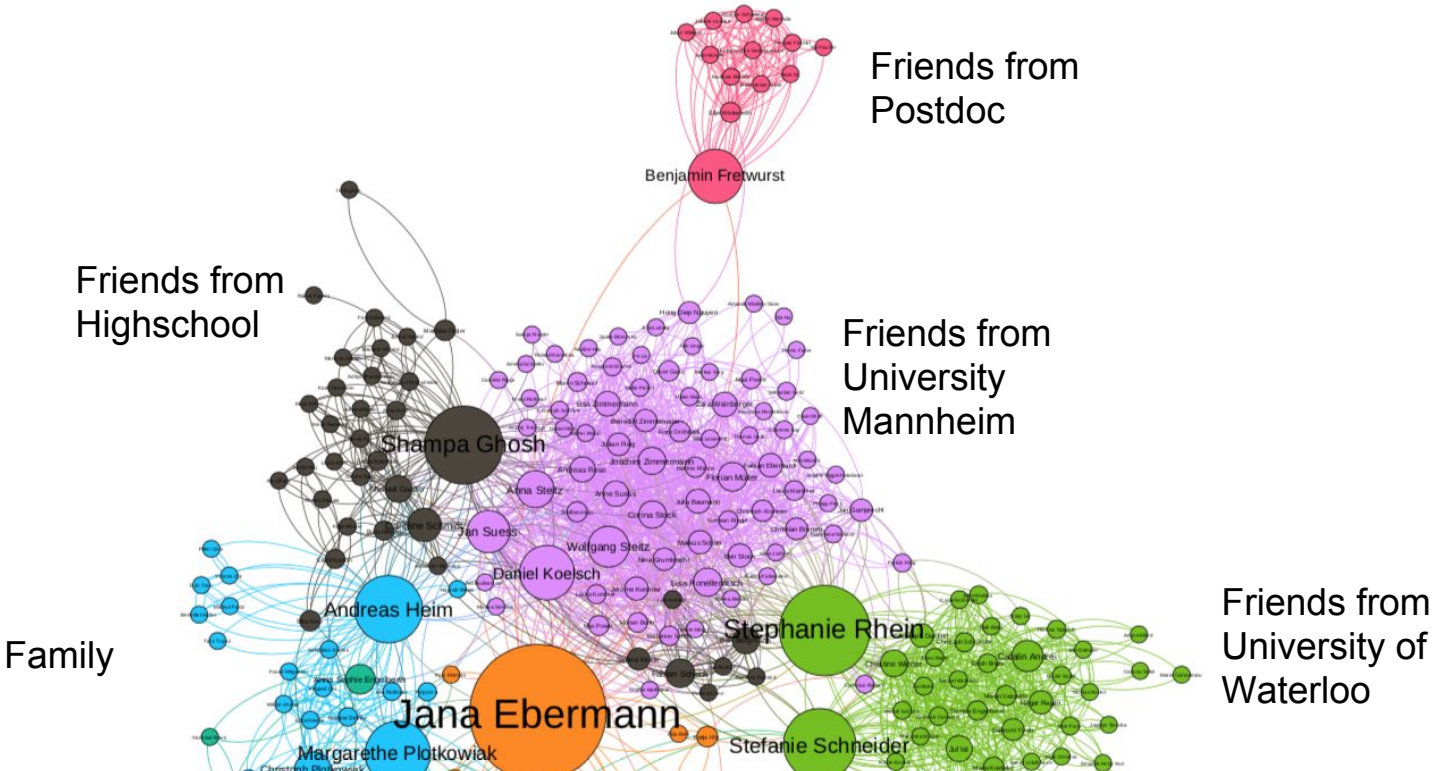


Gephi: 12. Export

- To Export your graph for publication in SVG or PDF click the Export button
- Save



Gephi: 13. Make Sense out of it



Create your own data

Try it out with these datasets:

- http://snap.stanford.edu/data/
- http://cdg.columbia.edu/cdg/datasets
- https://github.com/iKhaled/MCFCAnalytics
- ...

Create your own sport data:

- Soccer ball passes
- Soccer fouls between players
- Money Transfers between soccer clubs
- Waterpolo ball passes
- ...

Thanks, I hope you had some fun!