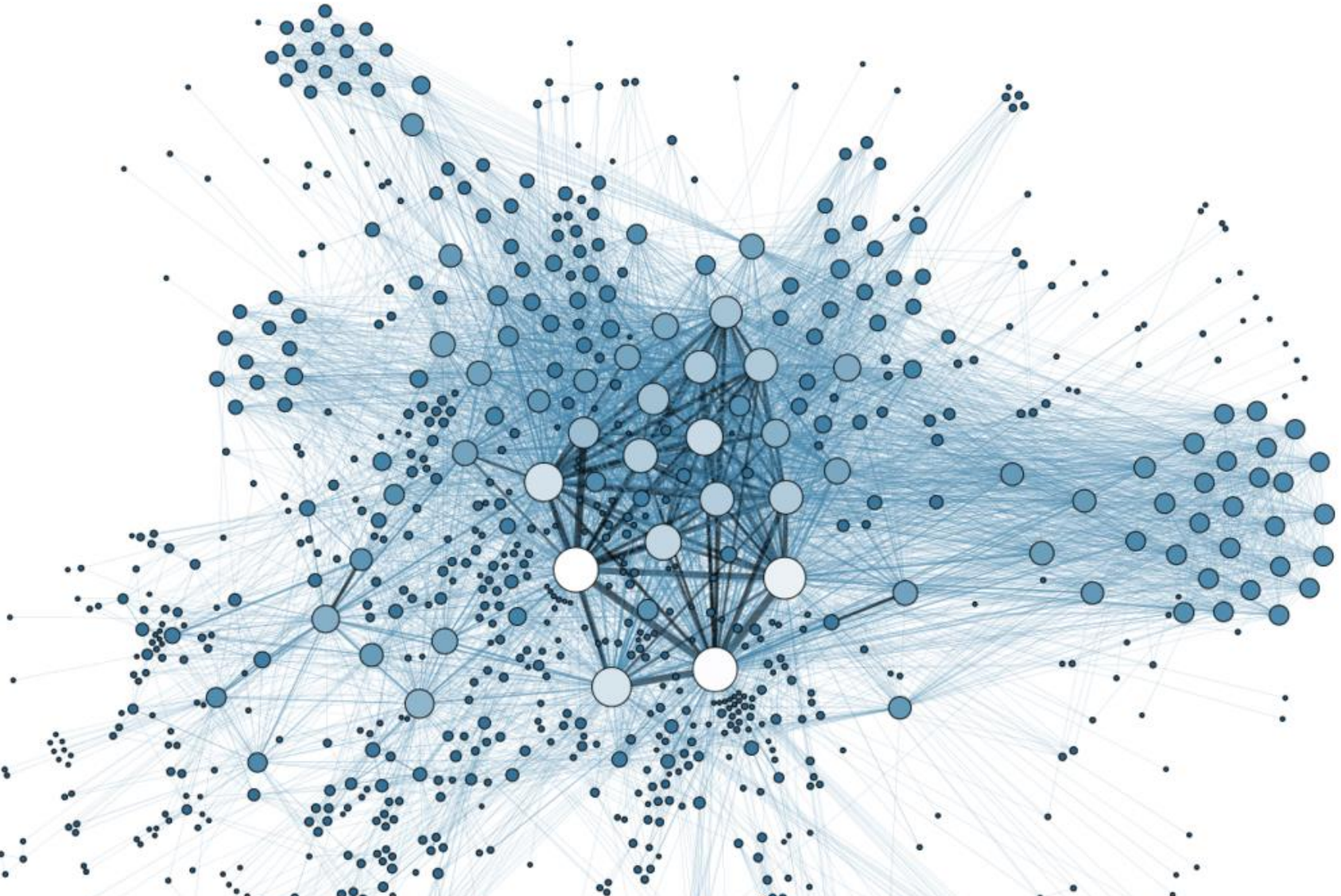
Network Analysis

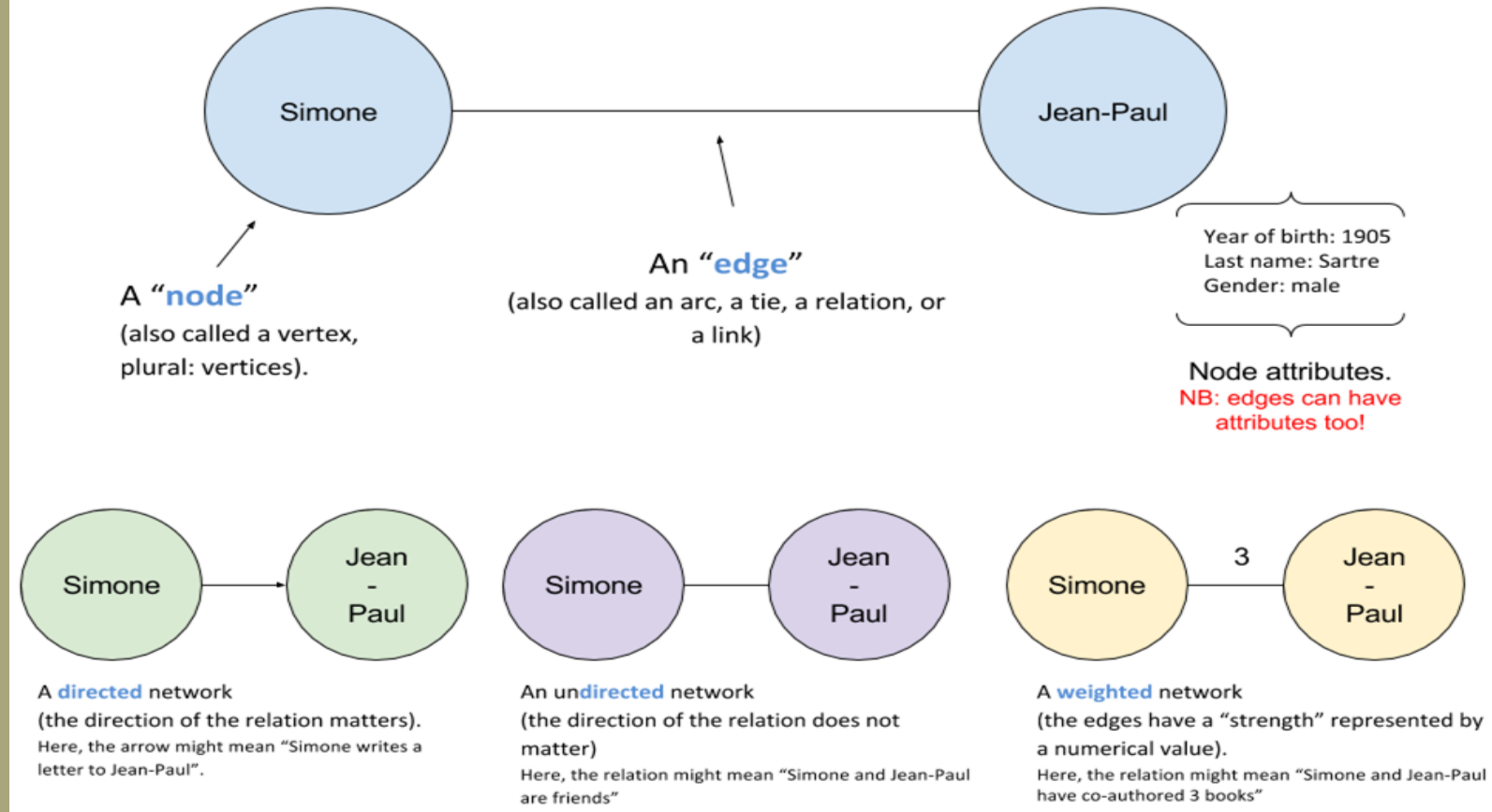
The Seven Bridges of Königsberg



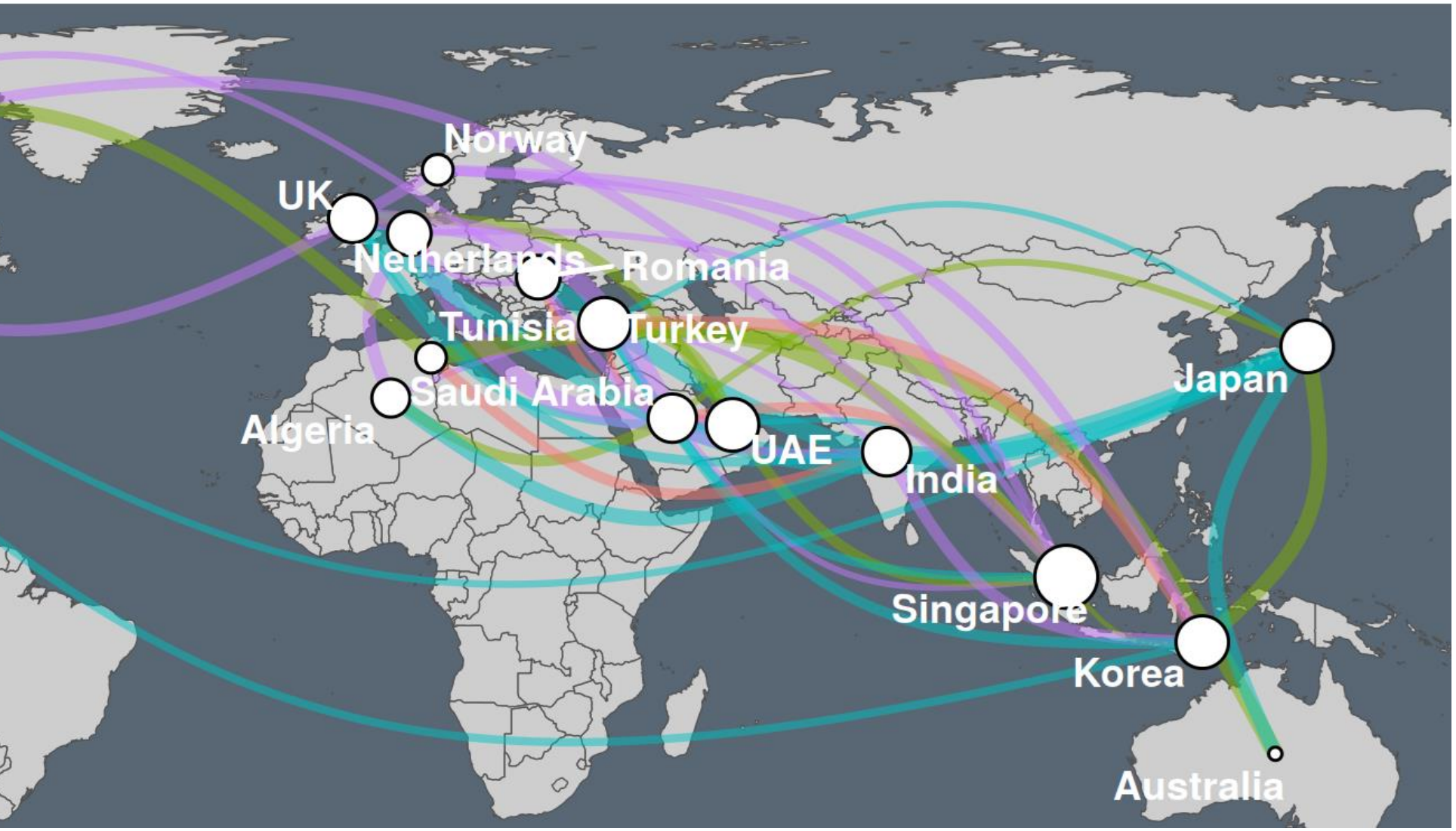
The problem:

- devise a walk that crosses each bridge once and only once
- …but the problem has no solution (Euler proved it)

Networks: Terminology



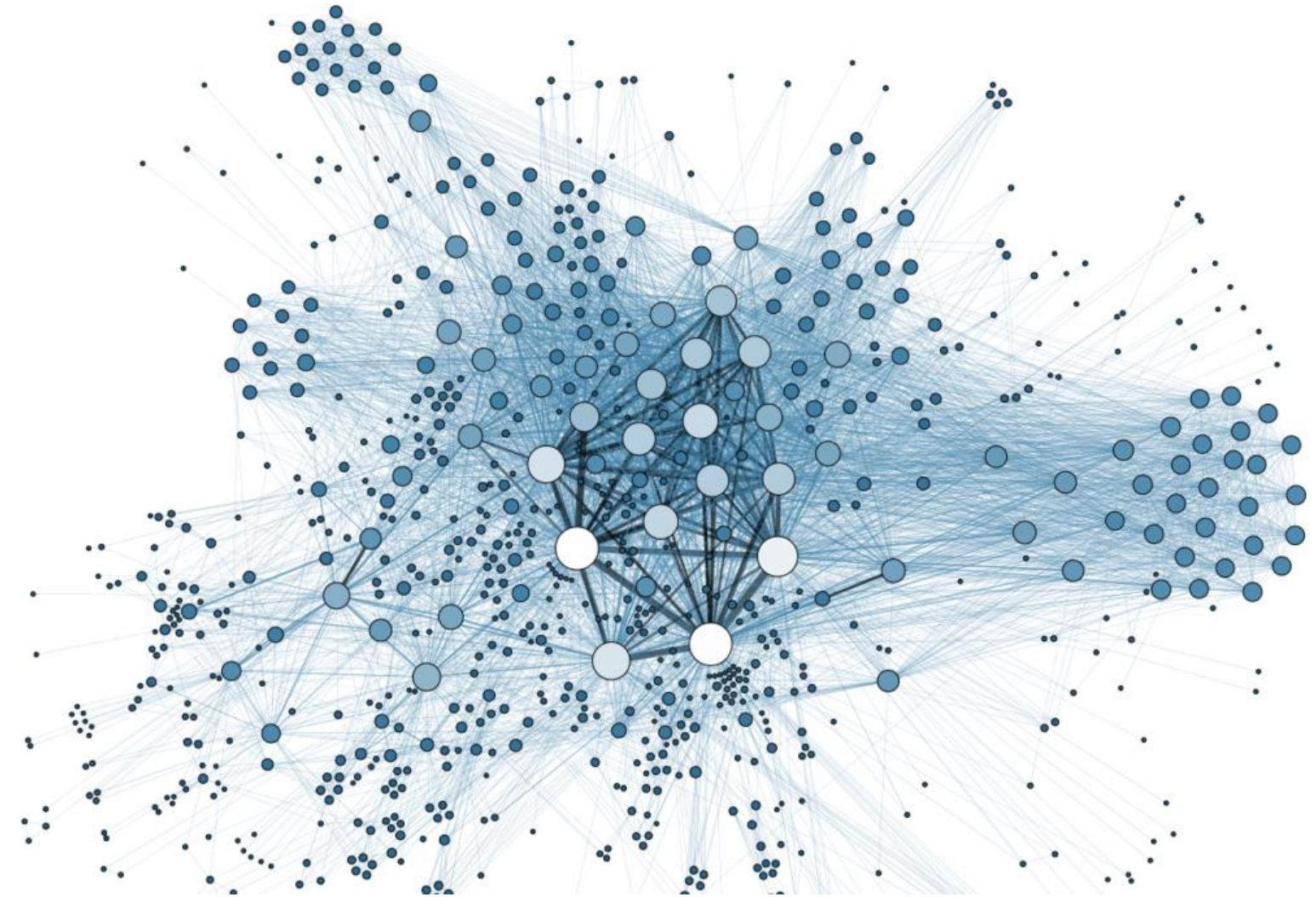
(Image taken from Clément Levallois, 2017)



The position of the nodes is fixed on the map

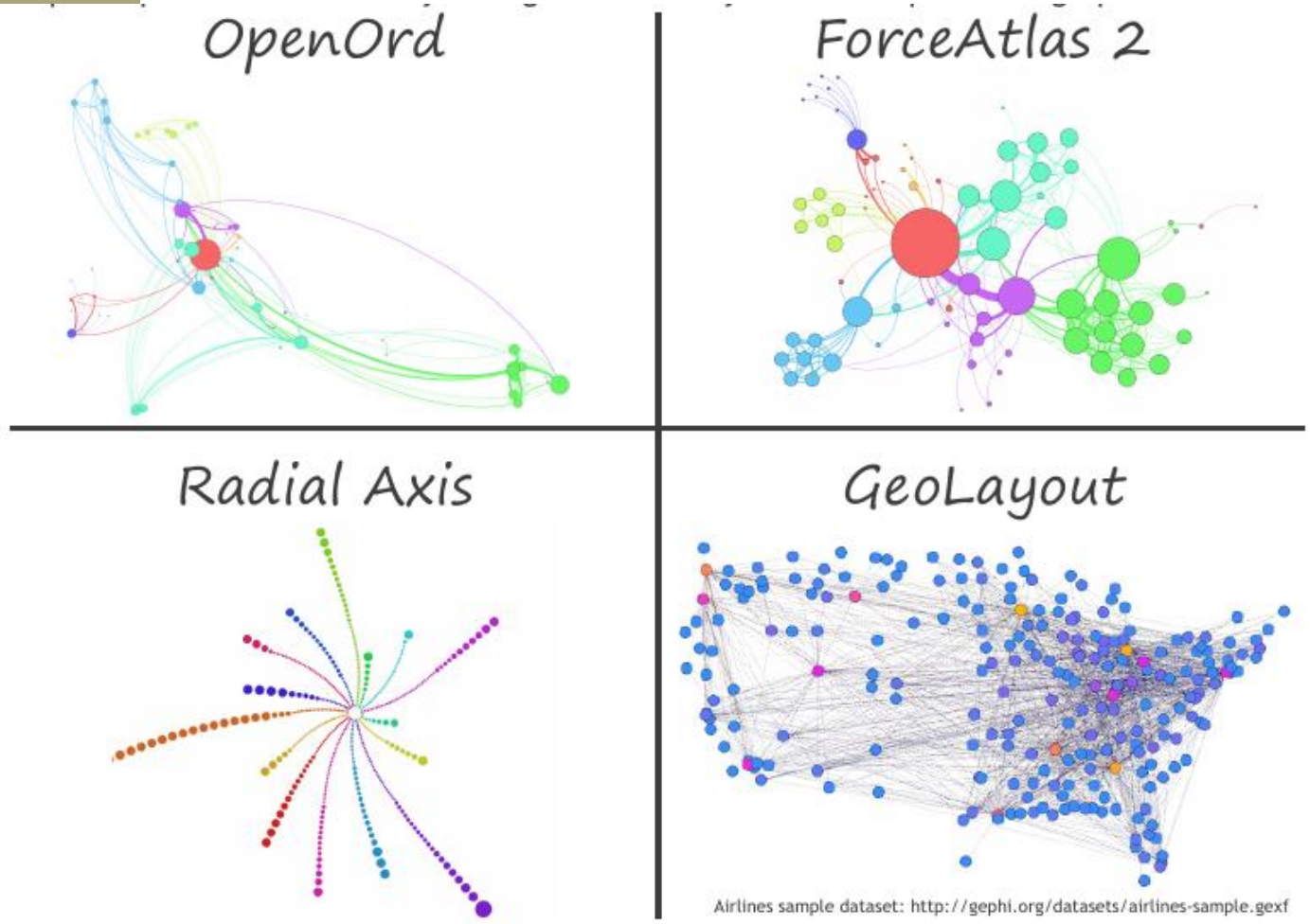
See for example:
https://koine82.carto.com/viz/5f6e08a5-ce1d-4227-b4b9-f29adab14023/public_map
(click on “Réseau”)

«Boot
Netwo



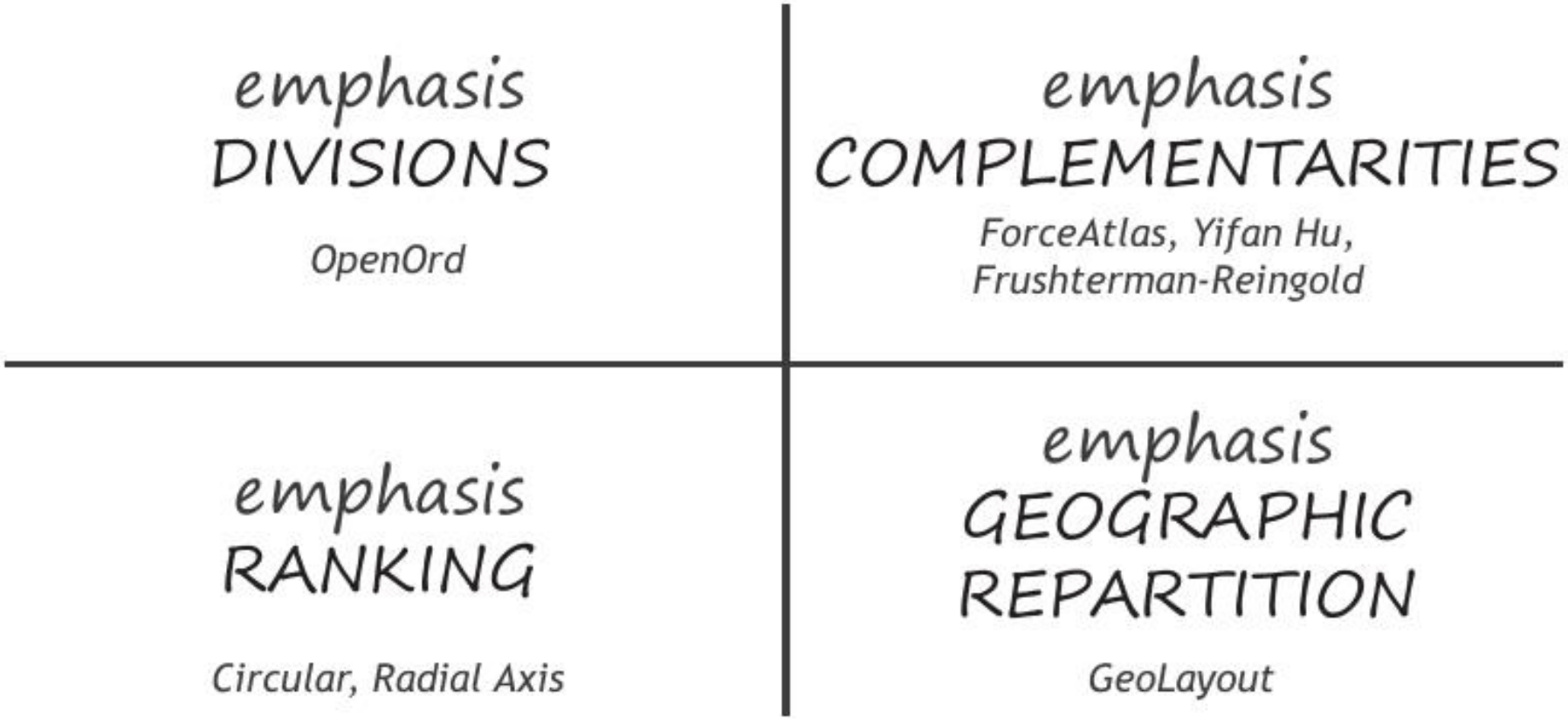
The position of the nodes is determined by the strength of their connections (i.e. by the edge's «weight»)

Visualizatio Algorithms



(Image taken from Gephi, 2011)

Visuali
Algorit



(Image taken from Gephi, 2011)

Network Analysis

- What are networks? Networks (graphs) are set of nodes (verticies) connected by edges (links, ties, arcs) Additional details Whole vs. ego: whole networks have all nodes within a natural boundary (platform, organization, etc.).
- An ego network has one node and all of its immediate neighbors. Edges can be directed or undirected and weighted or unweighted Additionally, networks may be multilayer and/or multimodal.

Network Analysis

- Why? Characterize network structure How far apart / well-connected are nodes? Are some nodes at more important positions?
- Is the network composed of communities? How does network structure affect processes?

Information diffusion Coordination/cooperation Resilience to failure/attack

- A network: First questions when approaching a network
- What are edges? What are nodes? What kind of network? Inclusion/exclusion criteria

Network analysis

- 6. Network data repositories

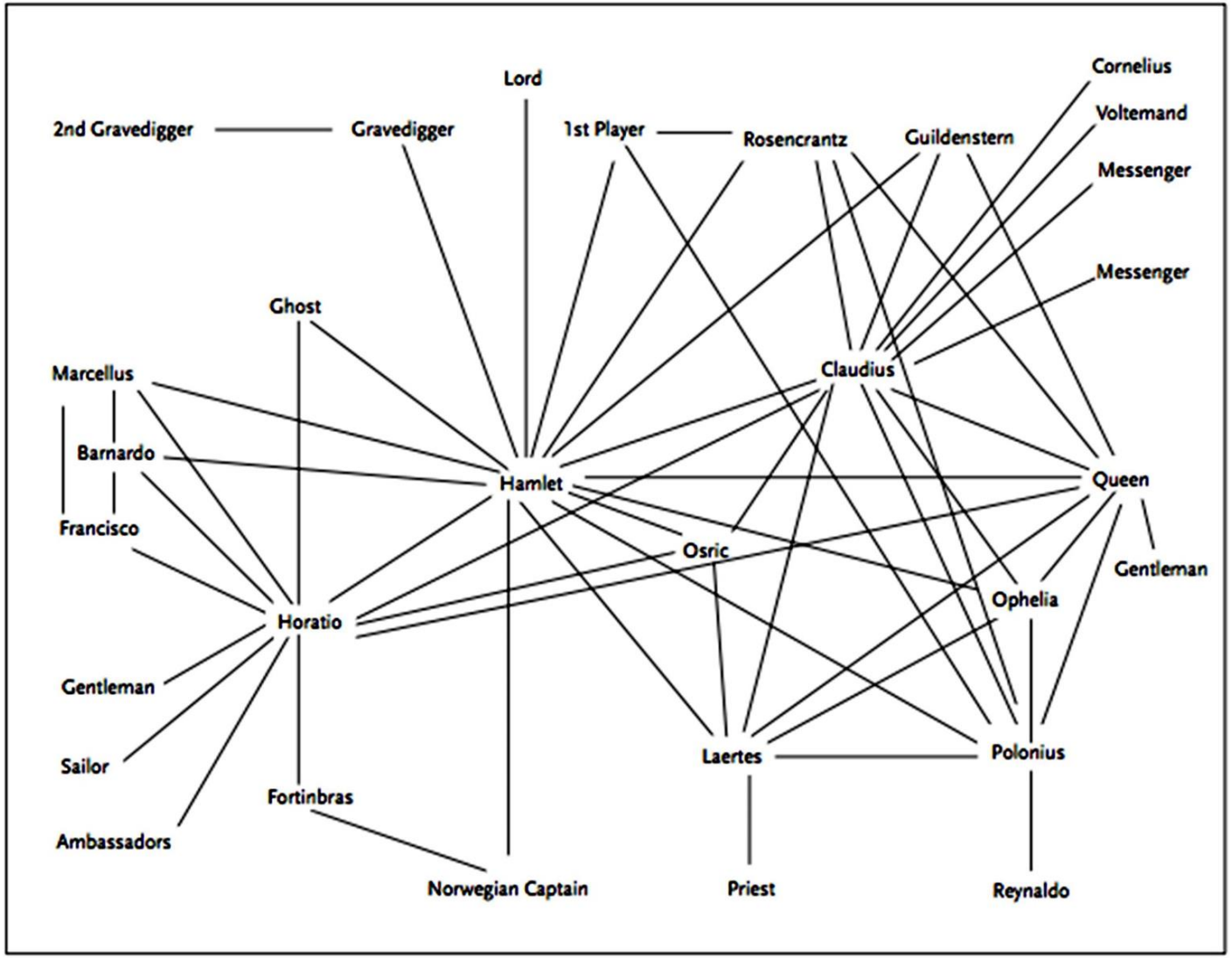
http://datamob.org

http://snap.stanford.edu/data

http://www-personal.umich.edu/~mejn/netdata

- 7. Python resources tweepy:
- Package for Twitter stream and search APIs (only python 2.7 at the moment) search and stream API example code along with code to create mentions/retweet network at
- https://github.com/computermacgyver/twitter-python Python two versions: 2.7.x –
- many packages, issues with non-English scripts 3.x – less packages, but excellent handling of international scripts (unicode)

Networks for Plot Analysis

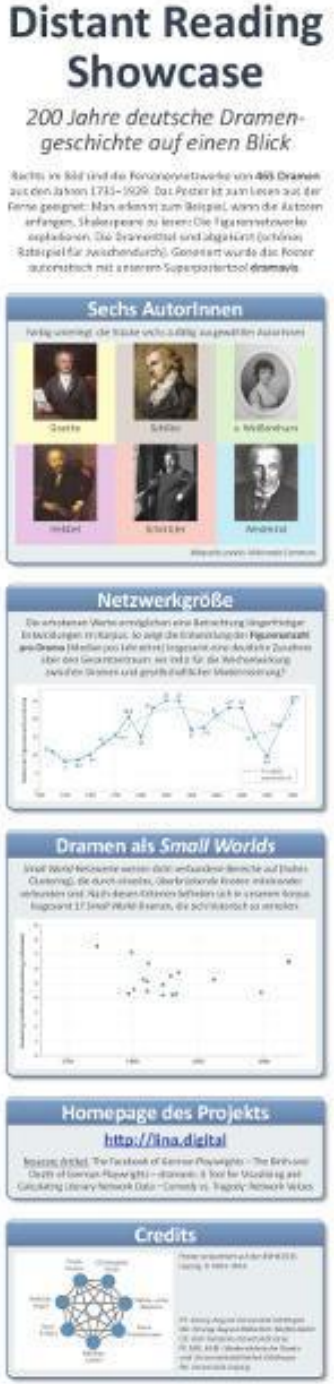


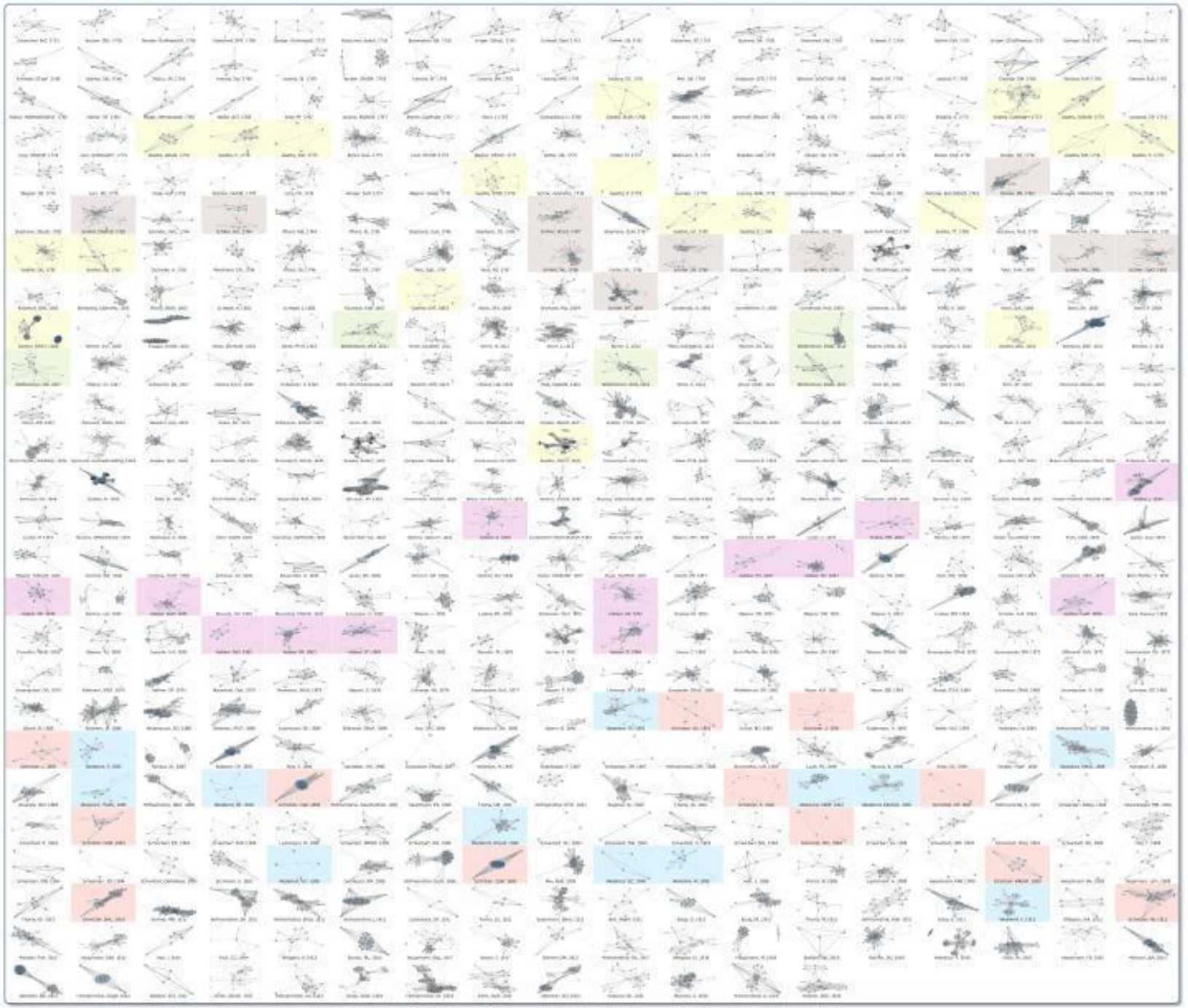
(see Moretti 2011)

Network Analysis of 200 years of German theater

(Fischer et al. 2016)

Download here: https://figshare.com/articles/Distant_Reading_Showcase_465_German_Dramas_DHd2016_Poster_pdf/3101203/1






Wedekind, FE, 1891
Schnitzler, AG, 1891
Schlaf, MO, 1892
Schnitzler, A, 1893
Sudermann, H, 1893
Wette, HuG, 1893
cheerbart, DR, 1897
Hofmannsthal, DFiF, 1898
Blumenthal, IwR, 1898
Laufs, PS, 1898
Panizza, N, 1898
Rilke, OG, 1898
Rosenow, KL, 1902
Thoma, DL, 1902
Schnitzler, R, 1902
Wedekind, DBdP, 1902
Wedekind, KNoSidL, 1902
Schnitzler, DP, 1903
heerbart, DW, 1904
Scheerbart, O, 1904
Scheerbart, DdL, 1904
Schnitzler, DtC, 1904
Scheerbart, LG, 1904
Scheerbart, HKK, 1904
Wedekind, DZ, 1908
Wedekind, M, 1908
Holz, S, 1908
Thoma, M, 1908
Lautensack, H, 1908
Hauptmann, FNB, 190

1.189 x 841 mm

Network Analysis of 200 years of German theater

"Daniel Keim uttered some criticism [...] broaching the problems of spring-embedder algorithms.* And we couldn't agree more: Spring embedders have "an undeniable aesthetic appeal, [...] yet a random layout is nearly always the default". One side effect of this is that graphs always look a tad different when generating them anew. Thus, similar graphs don't always look similar."

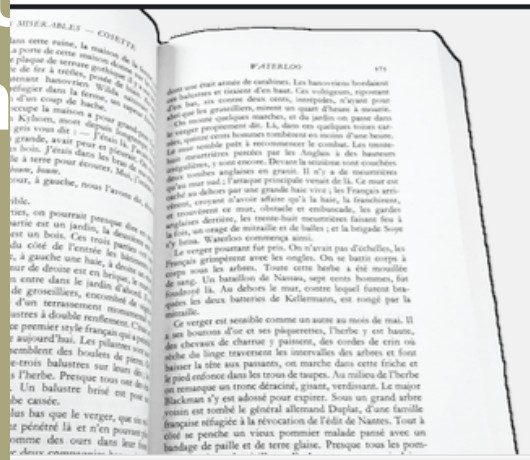
(https://dlina.github.io/Distant-Reading-Showcase-Poster-DHd2016-Leipzig/)

*spring embedders (also known as force-directed) are the algorithm that generate «bootstrap networks» (that change randomly)

Clu
Ch

https://lesmiserables.mla.hcommons.org/

Visualizing *Les Misérables*

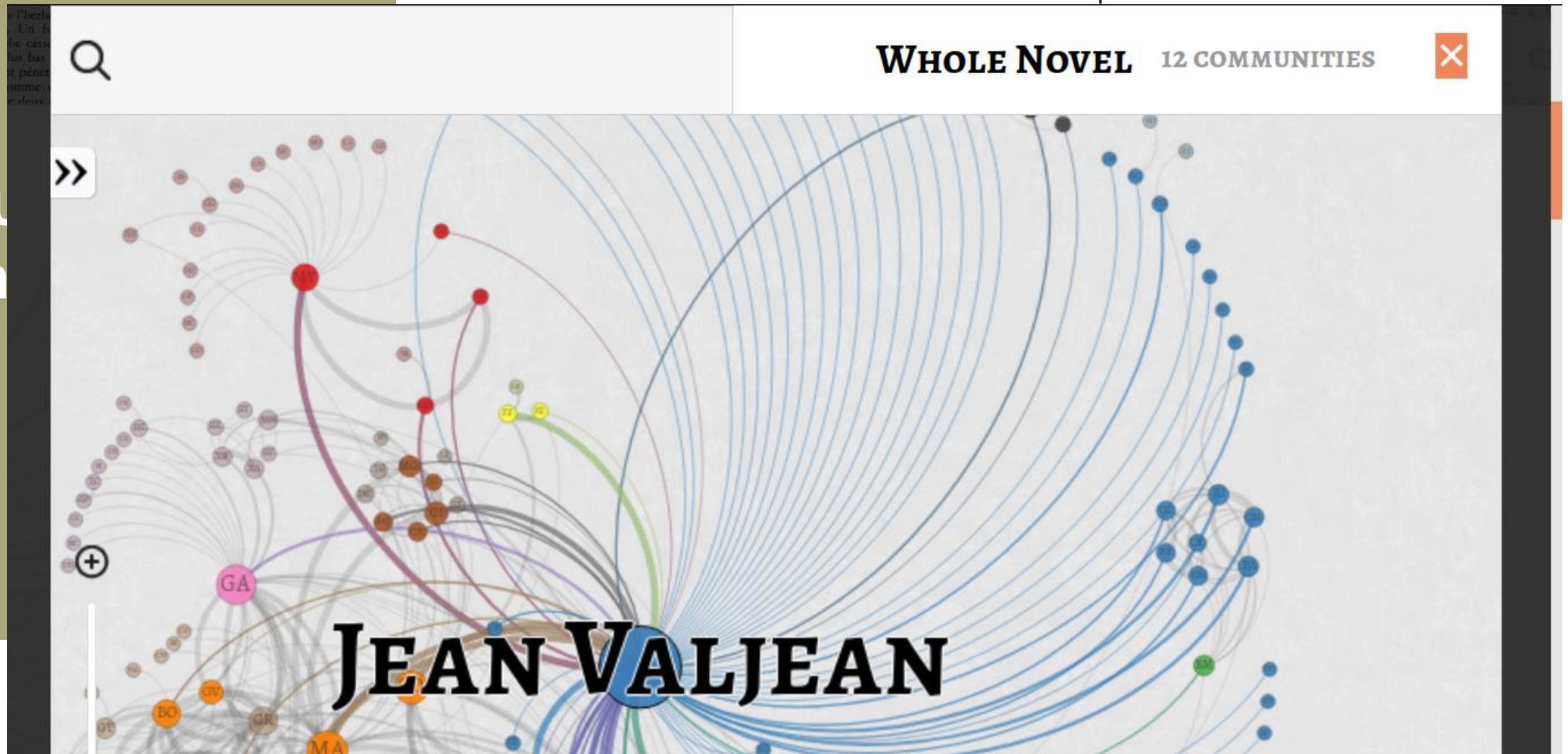


About the Project

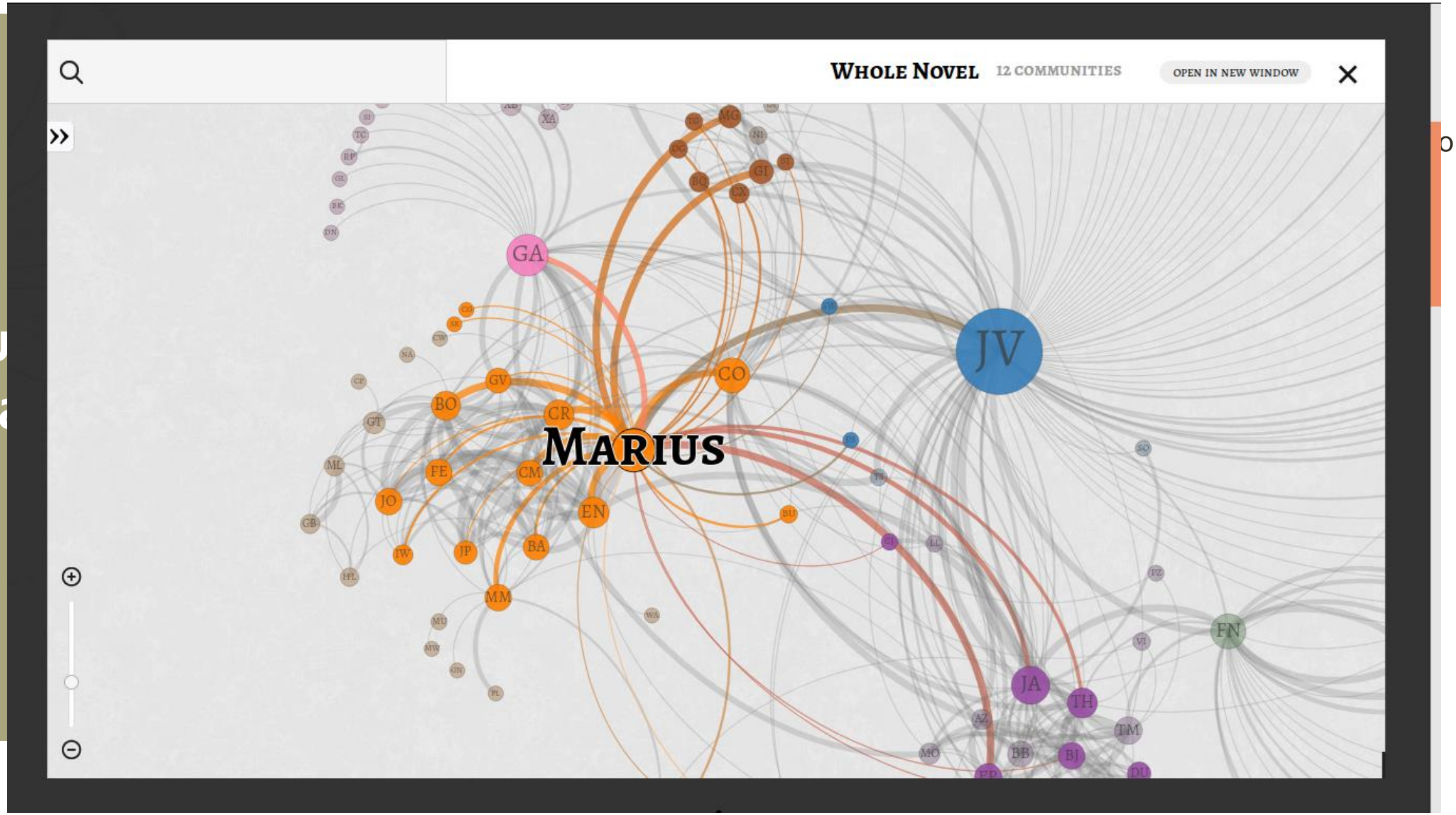
Characters of *Les Misérables*

Clu
Ch

https://lesmiserables.mla.hcommons.org



Clu
Cha


WHOLE NOVEL
12 COMMUNITIES
OPEN IN NEW WINDOW
MARIUS
GA
JV
CO
GV
BO
CR
GT
FE
CM
JO
EN
BA
IP
IW
MM
JA
TH
TM
FN
BB
BJ

org

Clustering Methods

"The modularity of a partition is a scalar value between -1 and1 that measures the density of links inside communities as compared to links between communities" (Blondel et al. 2008)

The algorithms we considered are:

Edge betweenness (Newman-Girvan). M.E.J. Newman and M. G
community structure in networks," *Physical Review* E 69, 026113 (2

Fast greedy optimization of modularity. A. Clauset, M.E.J. Newman, C. Moore: "Finding community structure in very large networks," http://www.arxiv.org/abs/cond-mat/0408187

Short random walks. Pascal Pons, Matthieu Latapy: "Computing communities in large networks using random walks," http://arxiv.org/abs/physics/0512106

Spin-glass model and simulated annealing. J. Reichardt and S. Bornholdt: "Statistical Mechanics of Community Detection, *Physical Review* E, 74, 016110 (2006), http://arxiv.org/abs/cond-mat/0603718

Multi-level optimization of modularity. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre: "Fast unfolding of communities in large networks." J. Stat. Mech. (2008) P10008.

Max the modularity measure over all possible partitions. Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski: "On Modularity Clustering," *IEEE Transactions on Knowledge and Data Engineering* 20(2):172-188, 2008.

Leading eigenvector of the community matrix. M.E.J. Newman: 'Finding community structure using the eigenvectors of matrices," *Physical Review* E 74 036104, 2006.

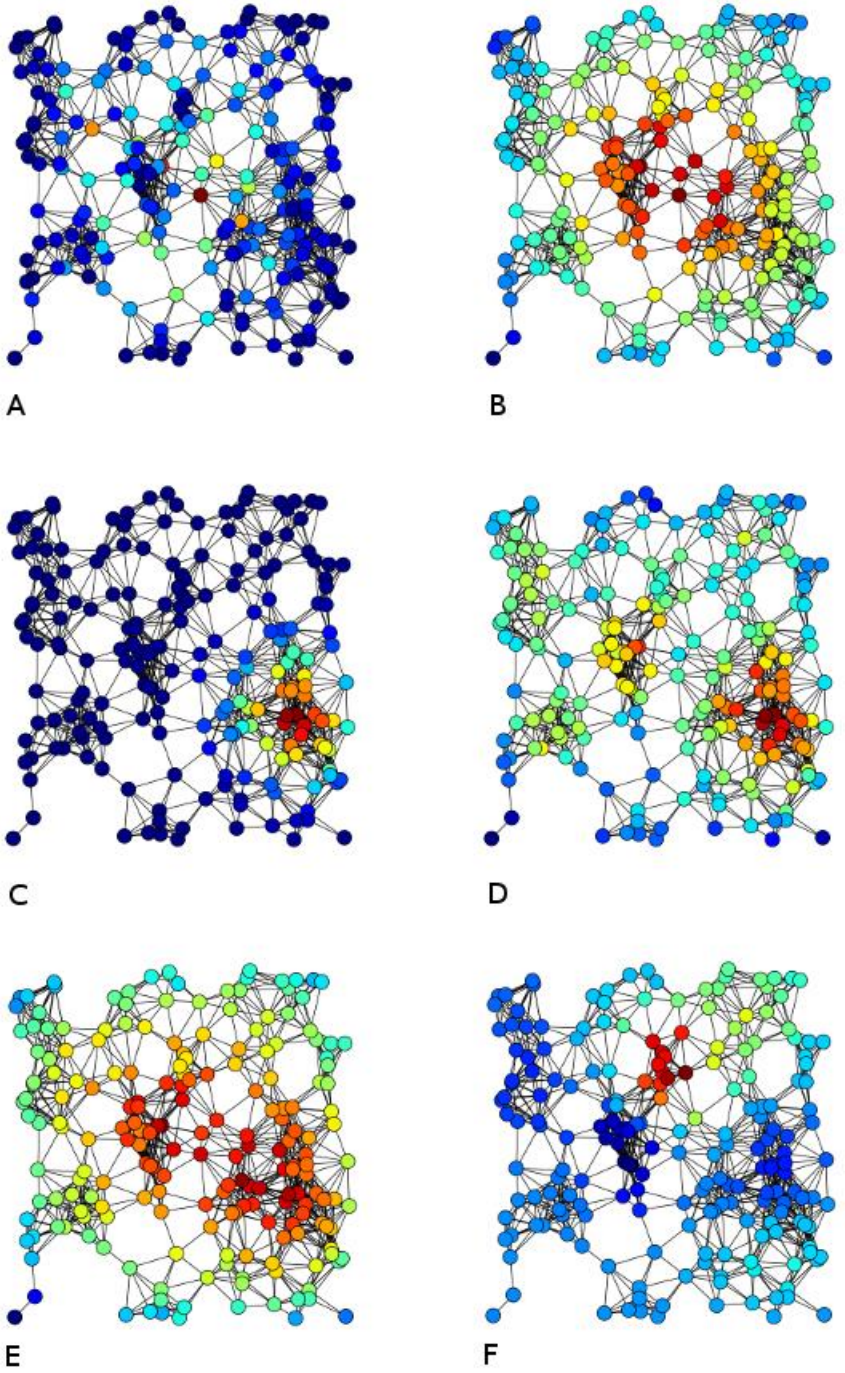
(https://lesmiserables.m

Detecting the Protagonist

Multiple Centrality measures:

A) Betweenness centrality
B) Closeness centrality
C) Eigenvector centrality
D) Degree centrality
E) Harmonic Centrality
F) Katz centrality

…which one shows the protagonist??



D
P

Detecting Protagonists in German Plays around 1800 as a Classification Task

Abstract

In this paper, we aim at identifying protagonists in plays automatically. To this end, we train a classifier using various features and investigate the importance of each feature. A challenging aspect here is that the number of spoken words for a character is a very strong baseline. We can show, however, that a) the stage presence of characters and b) topics used in their speech can help to detect protagonists even above the baseline.

Features

Table 2: Features and their description

Feature	Info. source	Description
Tokens	Text amount	The normalized number of tokens a character speaks during the whole play
Degree	Social relations	The degree of the node representing a character in a co-presence network based on scenes
Active scenes	Stage presence	The number of scenes in which a character speaks, normalized for the play
Passive scenes	Stage presence	The number of scenes in which a character is mentioned by name, normalized
T1-T20	Speech content	The probability of topic 1-20 given the speech of a character

Degree Centrality

De
Pr

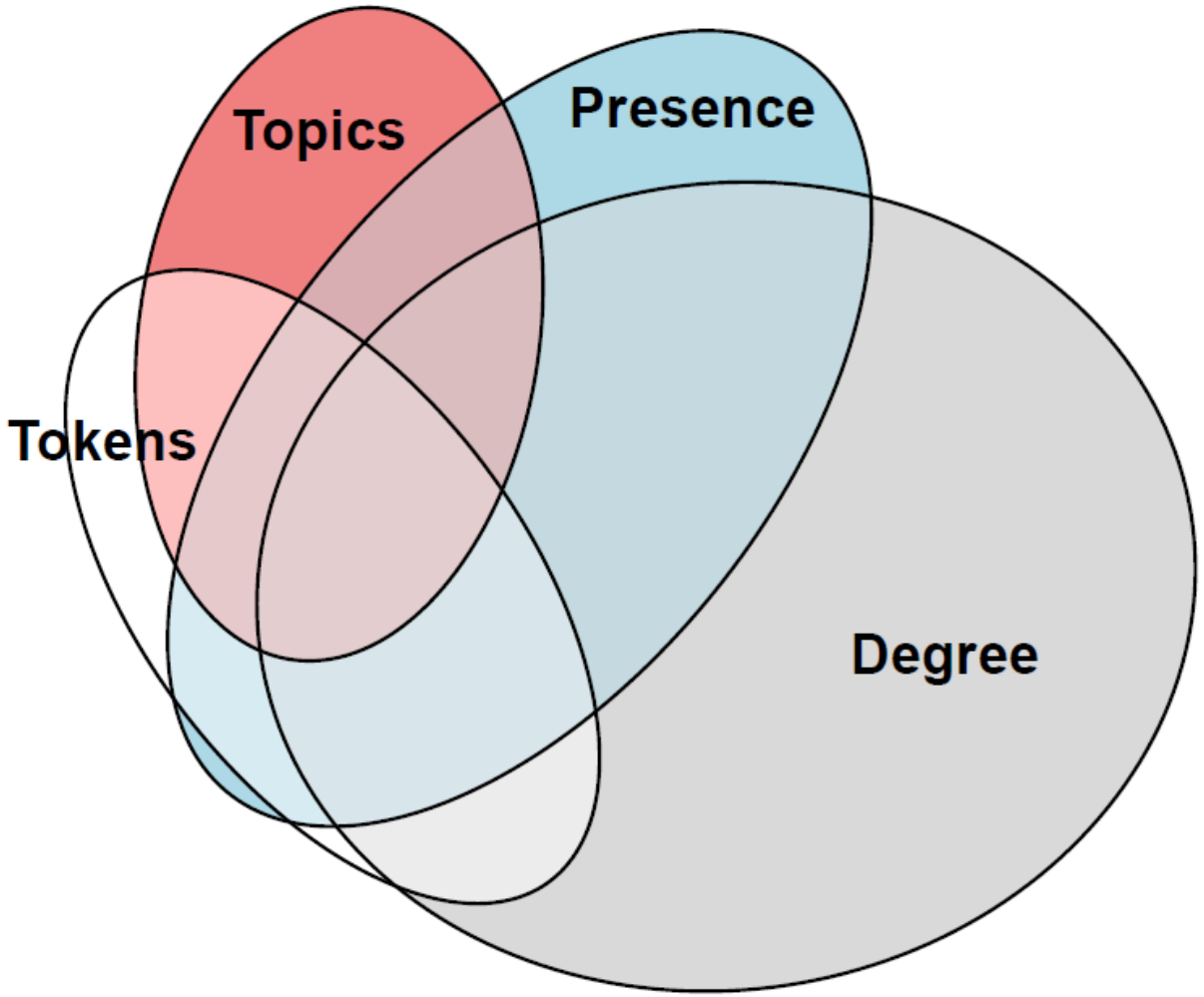


Figure 1: Classification errors by feature set as an Euler chart

*network centrality is the least accurate predictor for the main character!

(Reiter and Pagel 2018)

Description of the project

- This project is for complete beginners to Gephi. It supposes you have Gephi installed and running on your computer. That is all.
- When finishing this tutorial, you should be able to:
- be familiar with the vocabulary to discuss networks
- download a network file for this exercise
- description of the file / the network
- open a network file
- read the report after opening a file
- show the labels of the nodes
- layout the network
- visualize attributes of the network
- prettify the network for enhanced readability
- compute the centrality of the nodes in the network
- visualize attributes created by Gephi
- export a visualization as a picture or pdf
- export a web visualization of the network

What to do

- download a network file
- download this zip file and unzip it on your computer.
- You should find the file miserables.gexf in it.
- Save it in a folder you will remember (or create a folder specially for this small project).

description of the file / the network

- **description of the file / the network**
- This file contains a network representing "who appears next to whom" in the 19th century novel *Les Misérables* by Victor Hugo[1].
- A link between characters A and B means they appeared on the same page or paragraph in the novel.
- The file name ends with ".gexf", which just means this is a text file where the network information is stored (name of the characters, their relations, etc.), following some conventions.

Gephi

- open the network in Gephi
- open Gephi. On the Welcome screen that appears, click on Open Graph File
- find miserables.gexf on your computer and open it

Gephi

read the report after opening a file

- A report window will open, giving you basic info on the network you opened:
- This tells you that the network comprises 74 characters, connected by 248 links.
- Links are undirected, meaning that if A is connected to B, then it is the same as B connected to A.
- The report also tells us the graph is not dynamic: it means there is no evolution or chronology, it won’t "move in time".
- Click on OK to see the graph in Gephi.

- Gephi è “a visualization and exploration software for all kinds of
graphs and networks”

Gephi

Open Gephi (double click on the icon).

2. Click on New Project

3. Click on Import Spreadsheet

4. https://dracor.org/ita choose a csv file and select separator "comma" and import as "Edge Table"

6. Click Next and then click Finish.

Gephi may ask if you want to add the table to a new project or add it to the existing data table: choose this second option

7. In the Nodes table, at the bottom, click on "Copy data to other column," choose "Id" and copy to "Label" then to "ok": so we have the labels of the

Gephi

- 8. In the Nodes table, at the bottom, click on "Add Column" andadd "Gender" as the title: add the gender of thecharacters (M/F)
- 9. Click on Overview10. In the Layout section, choose "ForceAtlas2" and click on "Run"
- 10. Change the options "Scaling = 200," "Dissuade Hubs," and "PreventOverlap" and click on "Run"
- 11. Click on the "T" at the bottom below the network to bring up the labelsof the nodes
- 12. Click on Preview to see the network14. Go back to Overview: in the Appearance section (left)choose "Partition," under "Choose an Attribute" choose "Gender."then click on "Apply"