Visualization and analysis of social media networks: an overview and use cases

Researcher Teemo Tebest
Hypermedialaboratory/TUT
11/11/11

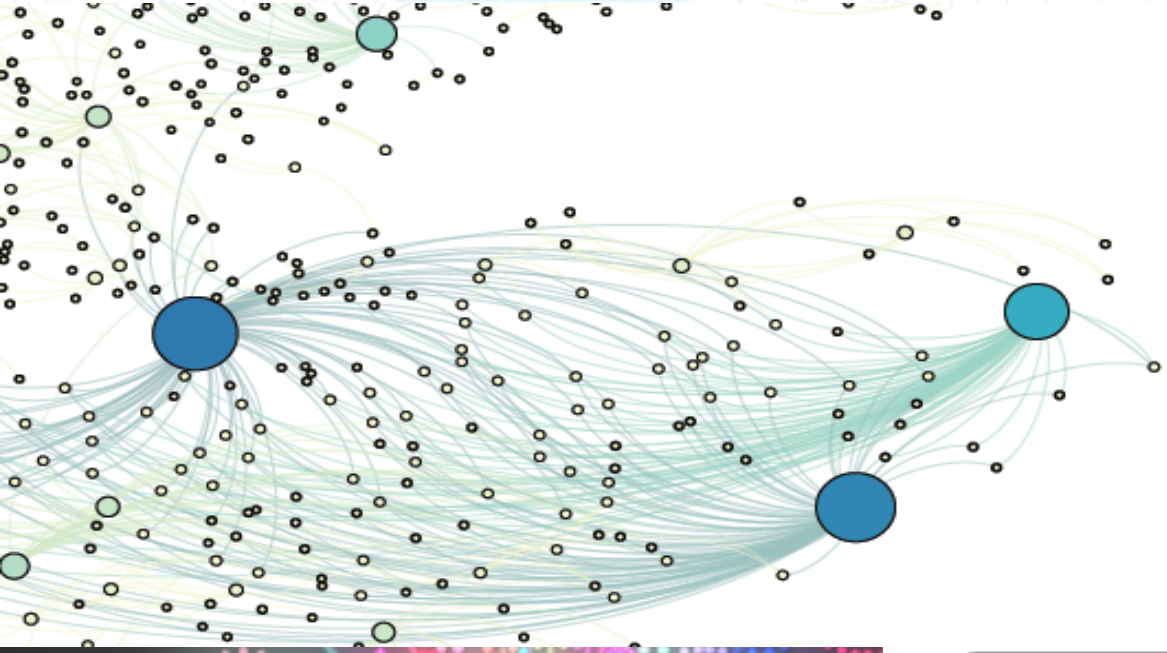
Contents

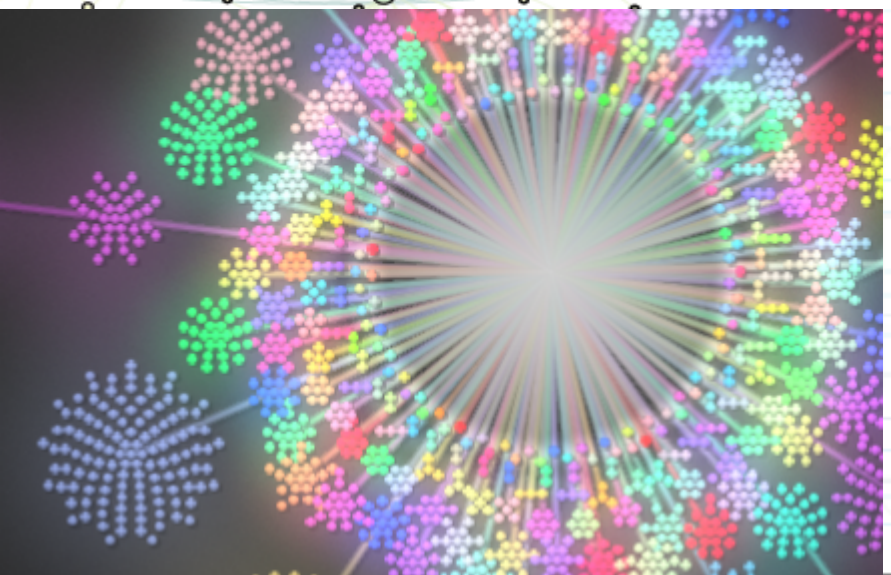
1) What is a visualization?
2) ...and what is social network analysis?
3) Where does the data come from?
4) How can we gather the data?
5) What kind of visualization tools can we use?
6) Use cases
7) Conclusions and next steps

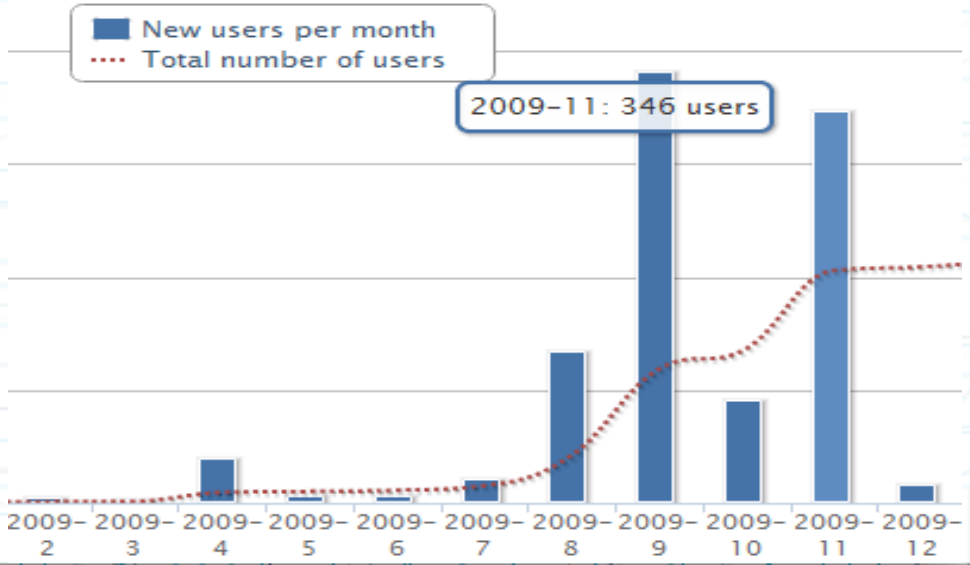
(Information) visualization

- Visualization is any technique for creating images, diagrams, or animations to communicate a message. (Wikipedia)
- Visualizations purpose is to represent existing and available data in a more human-readable form.
- Information visualization tries to give insight of the data that is being visualized.
- Visualization can be static or dynamic.

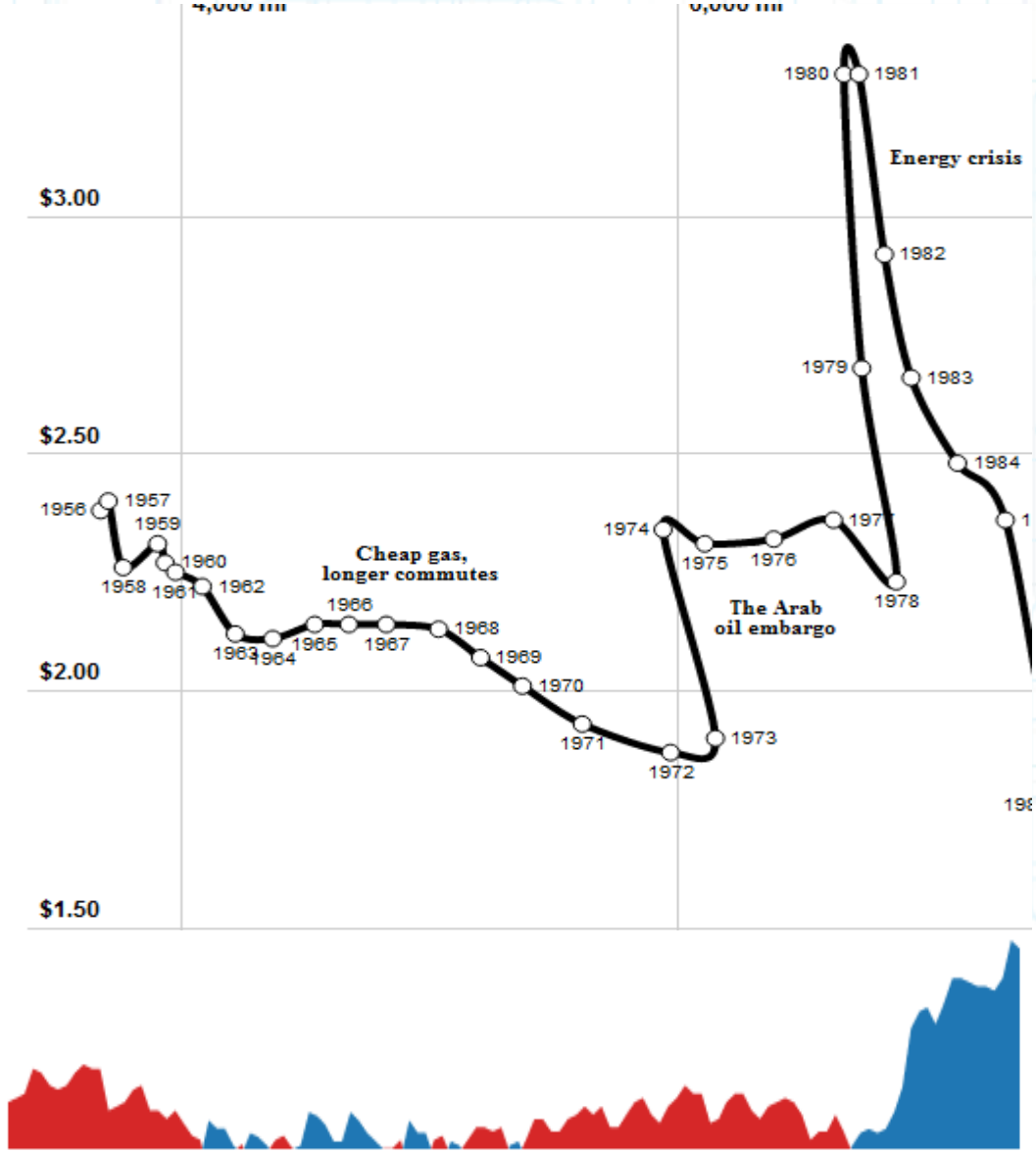
Visualizations

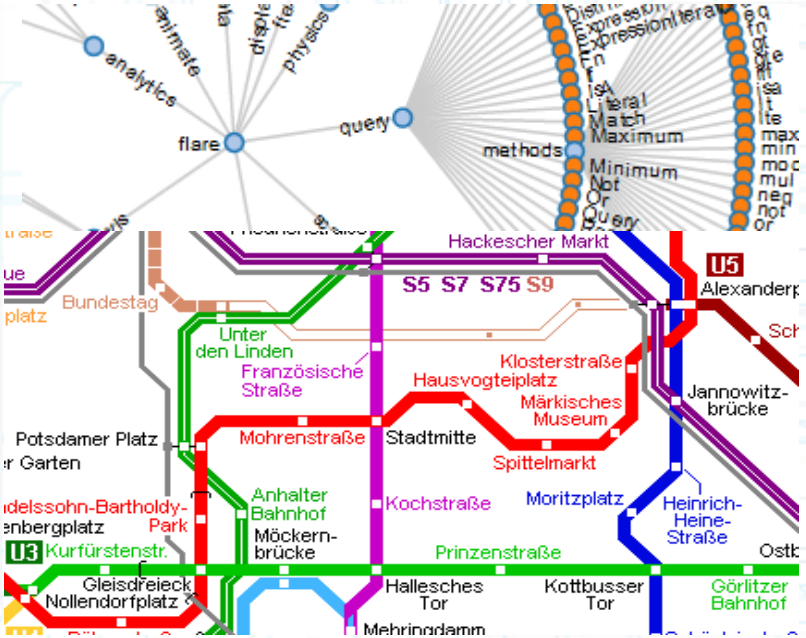






Visualizations

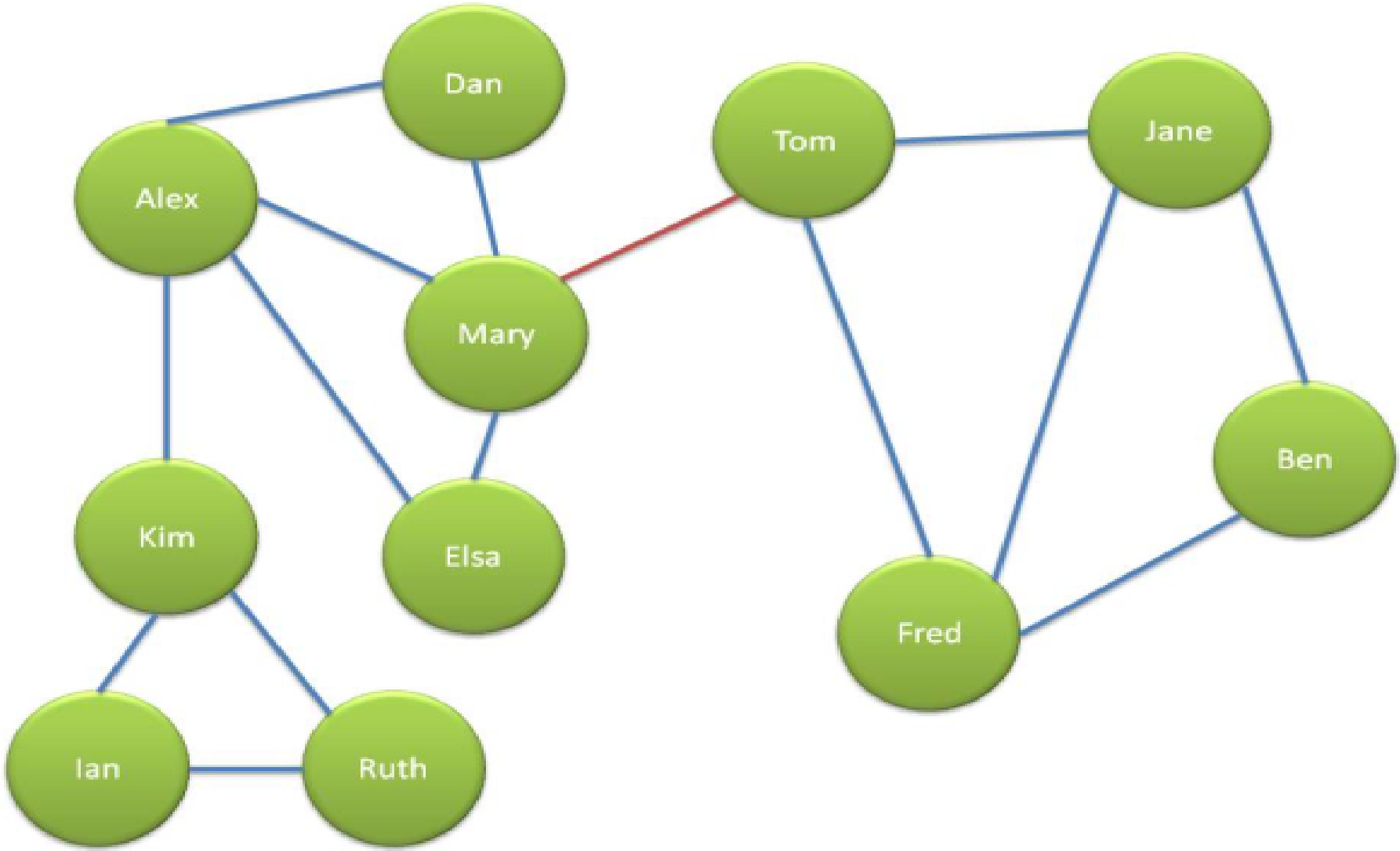




Social Network Analysis

- Social network analysis (SNA) views social relationships in terms of network theory consisting of nodes and edges.
- Nodes are the individual actors within the networks, and edges are the relationships between the actors.
- In its simplest form, a social network is a map of specified edges, such as friendship, between the nodes being studied. (Wikipedia)

Social Network Analysis



Transmission

- Well at least a short summary.
- Now we know what I mean by visualization.
- ...and we know the definition of SNA.

- Next we are looking on where does the data come from?

Data sources

- (In Internet) data can be found in all kind of systems.
  - Wikis, blogs, social media services, forums, databanks, logs, bonus card systems, bus route schedules, etc.
- In social media the data is created by the users.
  - Users interact within the system and the system used records user actions in some way (i.e. database).

Social Media Data Sources

Gathering data

- How can we gather the data from the system.
- There are two distinct ways to automatically gather data.
    1) Crawling from the frontend
    2) Saving in the backend

Gathering data

- There are two distinct ways to automatically gather data.
    1) Crawling from the frontend
    2) **Saving in the backend**
- It is crusial that all user actions are saved within the system.
    - These include page views, page edits, logins, content contributions, etc.
    - Some things are given but others must be considered while creating the system.

Gathering data to database

- Mostly the data is saved into a database.
  - MySQL, PostgreSQL, MongoDB, FS, etc.

	Field	Type	Collation	Attributes	Null	Default	Extra
☐	nid	int(10)		UNSIGNED	No		auto_increment
☐	vid	int(10)		UNSIGNED	No	0	
☐	type	varchar(32)	utf8_general_ci		No		
☐	language	varchar(12)	utf8_general_ci		No		
☐	title	varchar(255)	utf8_general_ci		No		
☐	uid	int(11)			No	0	
☐	status	int(11)			No	1	
☐	created	int(11)			No	0	
☐	changed	int(11)			No	0	
☐	comment	int(11)			No	0	
☐	promote	int(11)			No	0	
☐	moderate	int(11)			No	0	
☐	sticky	int(11)			No	0	
☐	tnid	int(10)		UNSIGNED	No	0	
☐	translate	int(11)			No	0	

Intermission

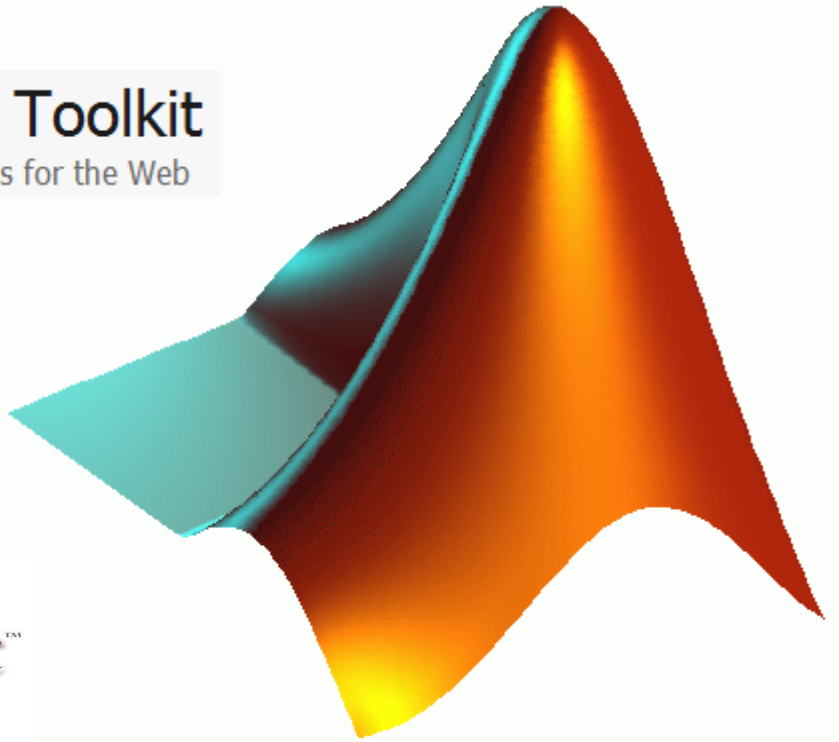
- Now we have the data collected, but it is located in some database in some raw format.
- This raw data gives very little insight on whats going on in the system.
- There are various tools available for visualizing data.
  - Commercial, free, open-source, etc.

Visualization tools

d3.js

Gephi

JavaScript InfoVis Toolkit
Create Interactive Data Visualizations for the Web



Highcharts
Interactive JavaScript charts for your web pages

vizster

Google

Visualization tools

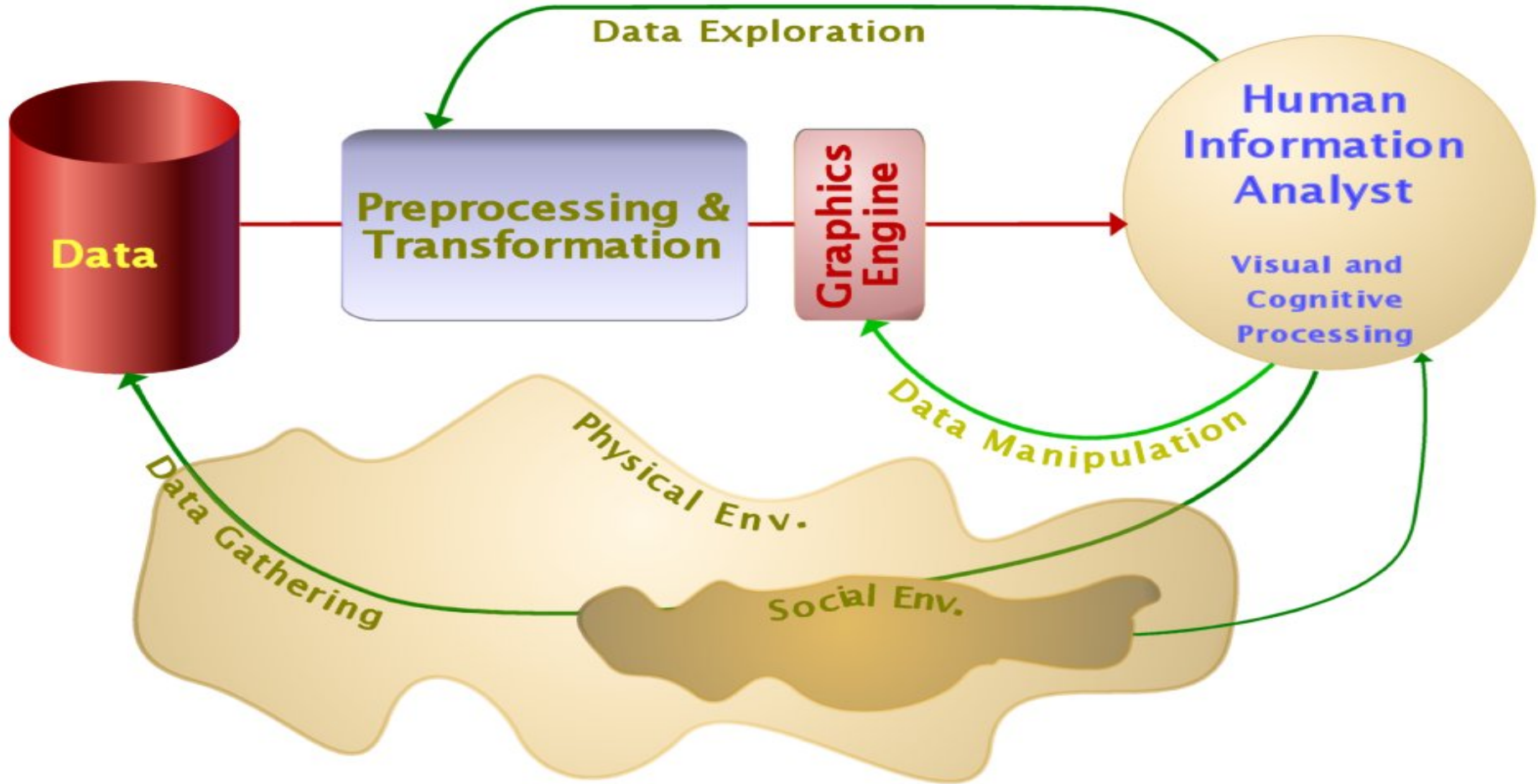
- Gathered data must be often converted to the format that is used in the desired visualization tool.
- Data formats include:
  - CSV (Excelsheets)
  - JSON
  - XML (i.e. .gexf)
  - Tailored arrays

A short overview of file formats

- Script for converting data to desired file format.
- .gexf – file format example.
- .csv – file format example.
- .gource custom file format example.

To summarize what we have gone through

- We are talking about information visualization process (Ware, 2004)



Intermission

- It's crusial to understand how information visualization (infovis) differs from scientific information visualization (scivis).
- There is a need for both ad its wrong to assume that only scivis is meaningfull and infovis has no value.

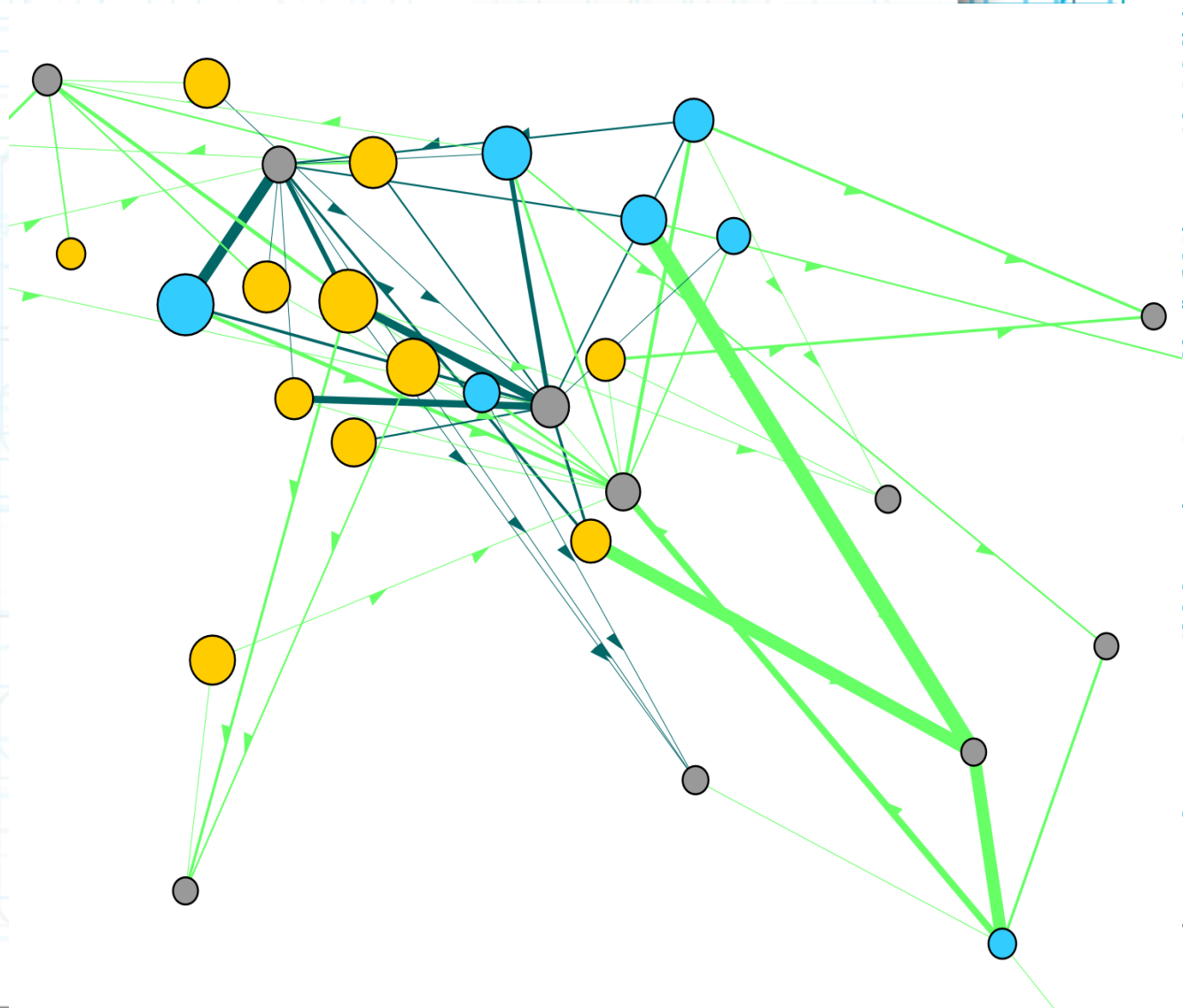
- Ok! Time to see some use cases.

Use cases

- Courses:
  - Developing an Online Publication 2011
  - Programming Hypermedia 2011
  - Usefulness of Web-Based Services 2011
- Data journalism and open data:
  - Tampere Data Journalism Day 29.9.2011
  - Varsova Open Government Data Camp 2011

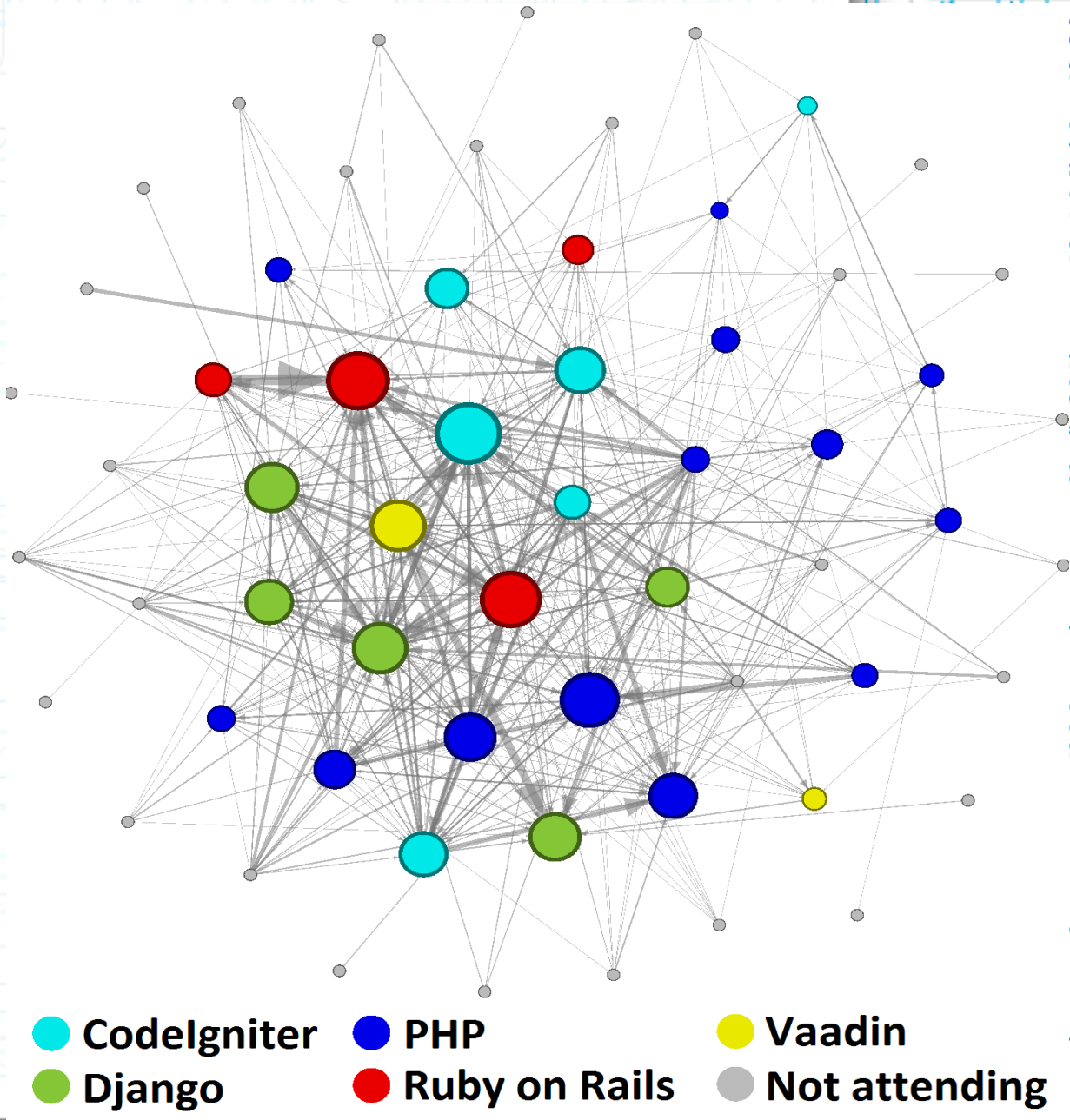
Developing an Online Publication

- Academic tribes.
- Yellow nodes are UTA students and cyan are TUT students. Grey nodes are wiki pages and edges represents edits.



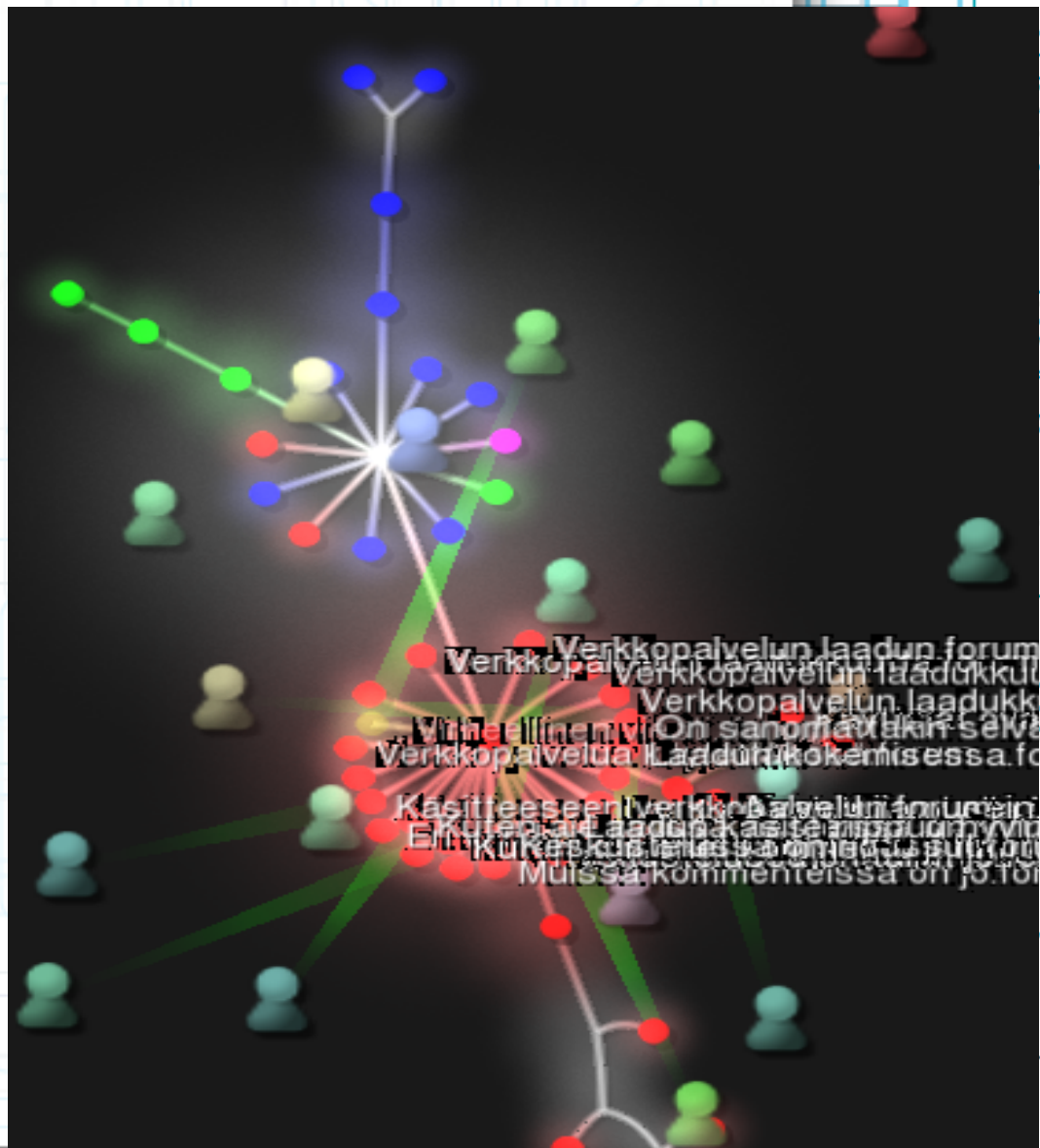
Programming Hypermedia

- Peer-learning.
- Nodes represent users reporting their work done in various different technologies and edges represent read counts.



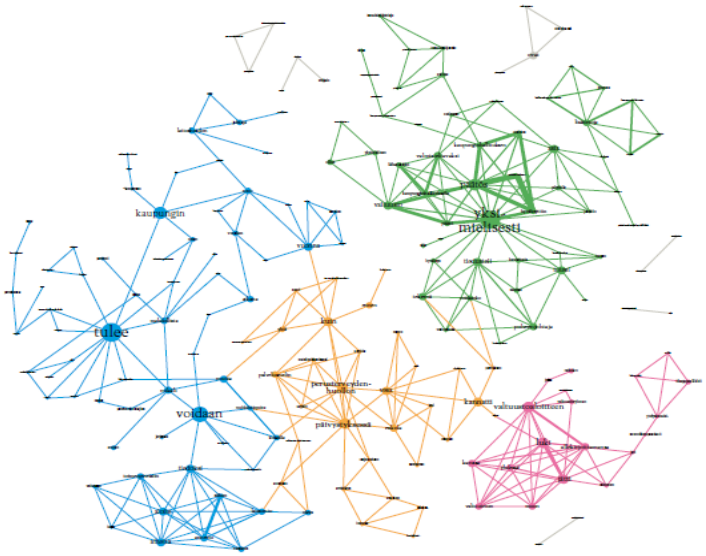
Usefulness of Web-Based Services

- Bursts and deadline culture. (Barabasi, 2010)
- The motivation to accomplish things increases towards the deadline.
- Also very little voluntary contribution were seen.



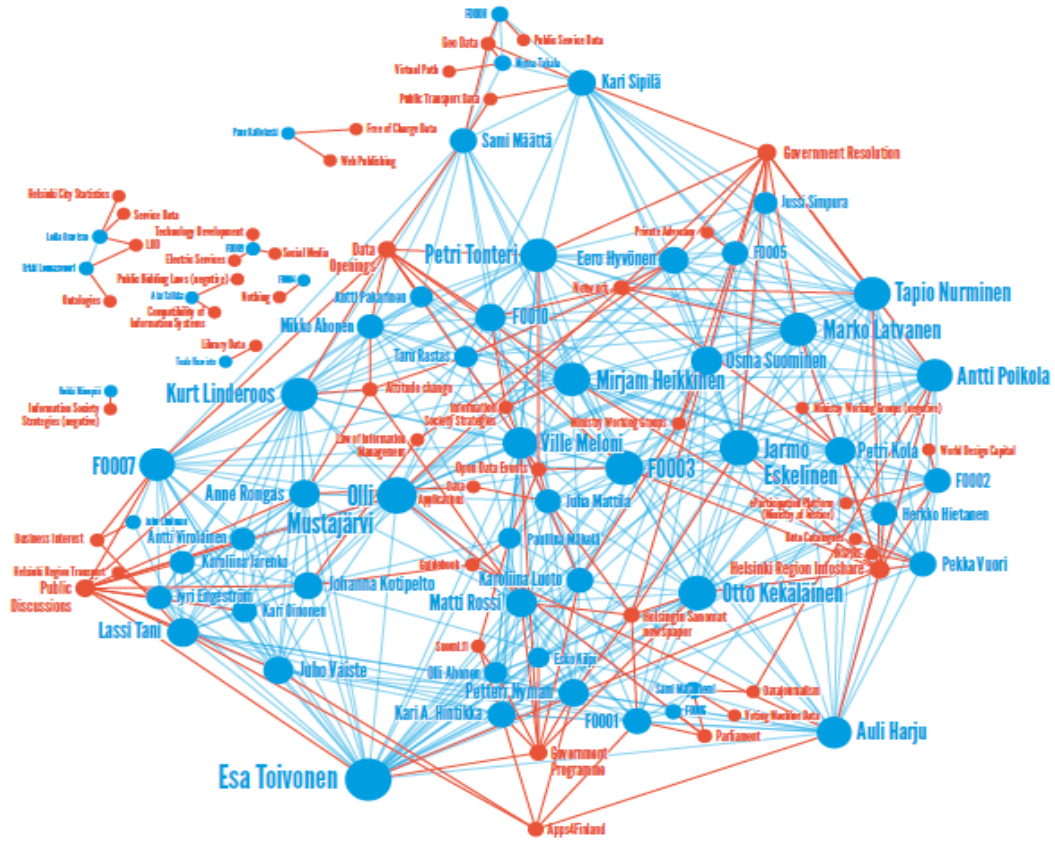
Tampere Data Journalism Day

- Collaboration workshop with journalists, graphic designers and engineers.
- Tampere city government decision-making data.
- All was done in one day.



Varsova Open Government Data Camp

- Finnish open-source actors and factors.
- Web survey: ”Mention 3 things in the area of Finnish open-source”



Conclusions and Next Steps

- Visualizations can reveal things that would otherwise remain hidden.
- Visualizations can be informatic and/or impressive.

- In the future visualizations should be more user-controlled (dynamic) and user-oriented.
- Ideally anyone should be able to be an amateur social network analysist.

Thanks to

- Jukka Huhtamäki, Kirsi Silius, Anne Tervakari, Meri Kailanto, Thumas Miilumäki and Jarno Marttila from Hypermedialaboratory/TUT.
- Antti Syrjä, Johan Laitinen and Raimo Muurinen from Avanto Insight Oy.
- Juuso Koponen from Informaatiomuotoilu.fi and Kirstu.
- Antti Poikola from Otavan Opisto.

Thanks for your interest

Feel free to contact me!

- https://twitter.com/teelmo
- http://fi.linkedin.com/in/teelmo

Further readings

- Barabasi A. L. (2010). Bursts: The Hidden Pattern Behind Everything We Do. Dutton Adult
- Ware C. (2004). Information Visualization: Perception for Design. Morgan Kaufmann
- Telea A. (2008). Data Visualization: Principles and Practice. AK Peters
- Keim D., Kohlhammer J., Ellis G., Mansman F. (Eds.) (2010). Mastering the Information Age: Solving problems with visual analytics. Eurographics Association
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Seppo, P. (2009, December). Social media enhanced studying and learning in higher education. IEEE EDUCON Education Engineering 2010 - The Future of Global Learning Engineering Education
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Seppo, P. (2010, January). Students' motivations for social media enhanced studying and learning. Knowledge Management & E-Learning: An International Journal (KM&EL)

Further readings

- Tebest, T. (2010, June). Web service monitoring and visualization of surveillance data [Master Thesis]. Tampere University of Technology
- To be published in Bali 2011:
  - Developing an Online Publication: Collaboration among Students in Different Disciplines
  - Programming of Hypermedia: Course Implementation in Social Media
- Web sites
  - http://datajournalismi.fi/
  - http://gephi.org/
  - http://code.google.com/p/gource/
  - http://hlab.ee.tut.fi/piiri/groups/hypermedian-ohjelmointi-2011
  - http://hlab.ee.tut.fi/piiri/groups/vpkk-2011
  - http://hlab.ee.tut.fi/piiri/groups/verkkojulkaisun-kehittaminen