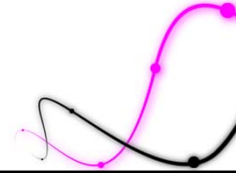


“NETWORK VISUALIZATIONS”

When do they make sense? Where to start?

Clement Levallois,
Erasmus University Rotterdam
Ehumanities group, Royal Netherlands Academy of Arts and Sciences (KNAW)

#cbs2013, May 24 2013



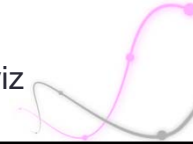
Bio notes

- Phd in History of economics
- Practice of digital methods of research since 2008
- Member of the Gephi Consortium, certified Gephi trainer.
- digital humanities, network analysis, natural language processing, data visualization, information retrieval, web applications ...

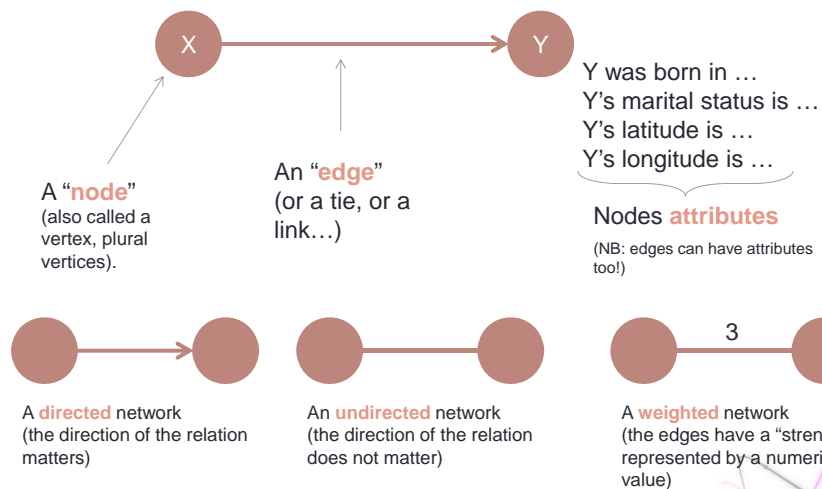


Workshop

1. **See** your datasets differently
2. Insights to be gained
3. **How to** work with connected data
4. Good practices and red flags
5. **Demo** – in 10 minutes from Excel to a online viz

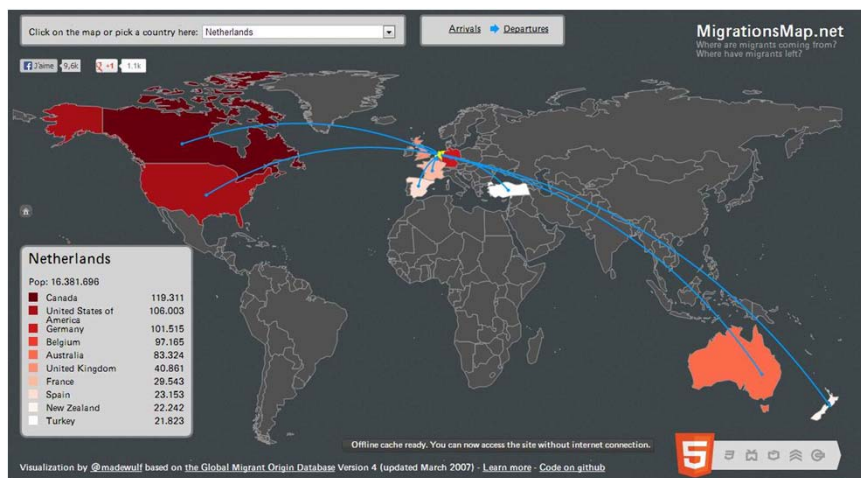


A note on the terminology



1. SEE YOUR DATASETS DIFFERENTLY

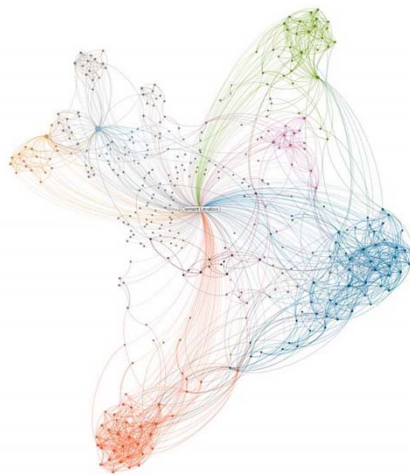
We are accustomed to geospatial data, and how to visualize them – on maps!




We are accustomed to time series / economic data, and how to represent them – on charts!



We are accustomed to social network data, and how to represent them – as graphs!





second example of a social network representation that makes intuitive sense.

Network data visualization does not stop at social networks

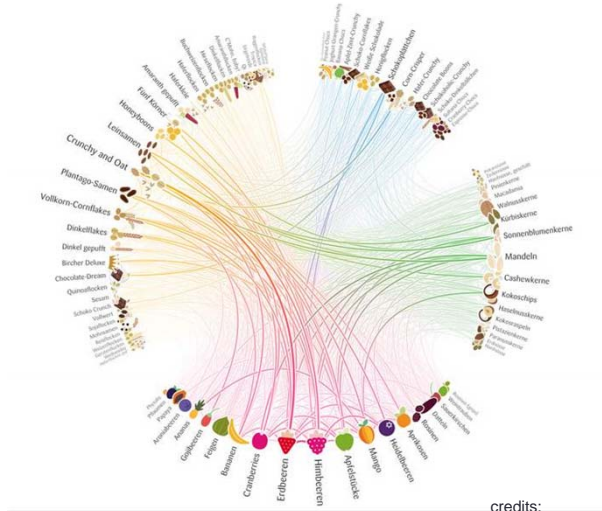
... or server logs.

When **any form of bond** exists between observations, these observations can be analyzed from the point of view of forming a network.

We have been so much trained to look at observations as independent entities, to be analyzed through aggregation...

... that we have lost sight that many datasets are actually rich of connections between individual data points, worth exploring.

Co-occurrences (Muesli ingredients)

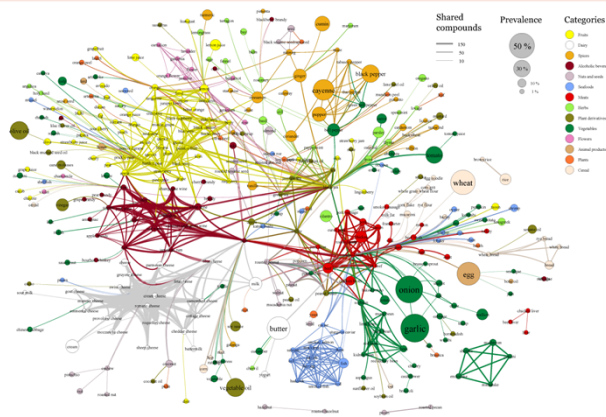


credits:
moritz.stefaner.eu – Müsli
Ingredient Network

Co-occurrences (all ingredients)

Flavor Network

Yong-Yeol Ahn, Sebastian Ahnert, James P. Bagrow, and A.-L. Barabási



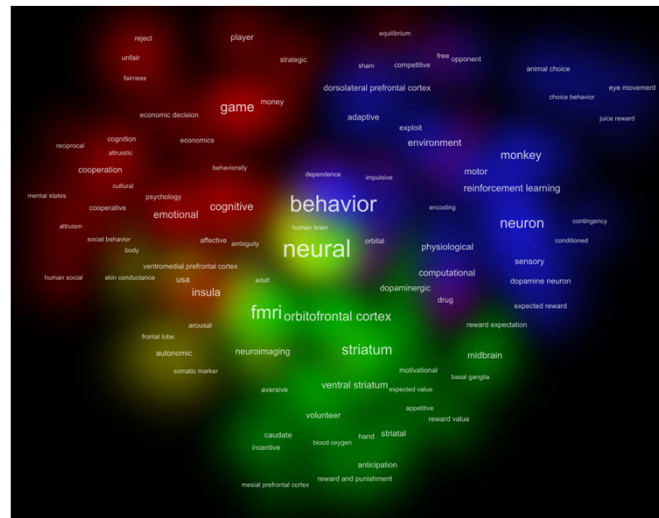
Map of flavors

2 flavors are connected

If they frequently appear in common recipes.

Flavor network. Colored ingredients link and their shared relationship are illustrated. The color of each ingredient represents the first category that the ingredient belongs to, and the size of an ingredient is proportional to the frequency it was collected from online recipe databases (prevalence, after recipe normalization). The values represent an estimate of their shared many-flavor compounds. We obtained the list of flavor compounds in each ingredient from the book "Flavor's Handbook of Flavor Ingredients" and then applied a hierarchical clustering analysis of 2000 top 1000 most statistically significant link between ingredients. The thickness of an edge represents the number of shared flavor compounds. To visualize this edge, we ranked based on the algorithm from [1] based on prevalence and distance.

Co-occurrences (concepts)



Co-location

Total Value: \$417,683,087,995



Map of product categories

2 products are connected

If they are frequently exported

by the same countries

Similarity

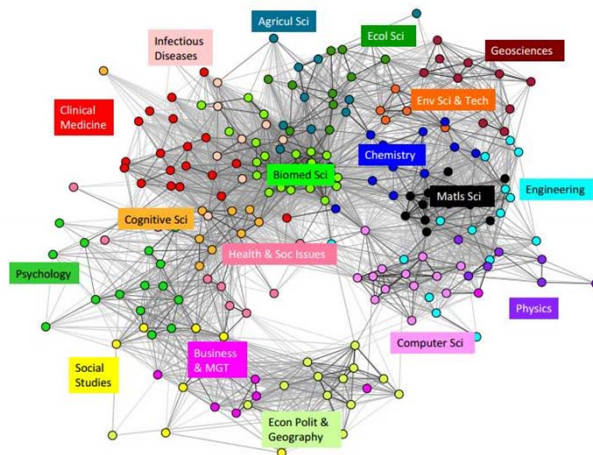


Figure 2. Global science map based on citing similarities among ISI Subject Categories (2007).

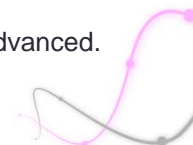
2 entities are connected if they ...

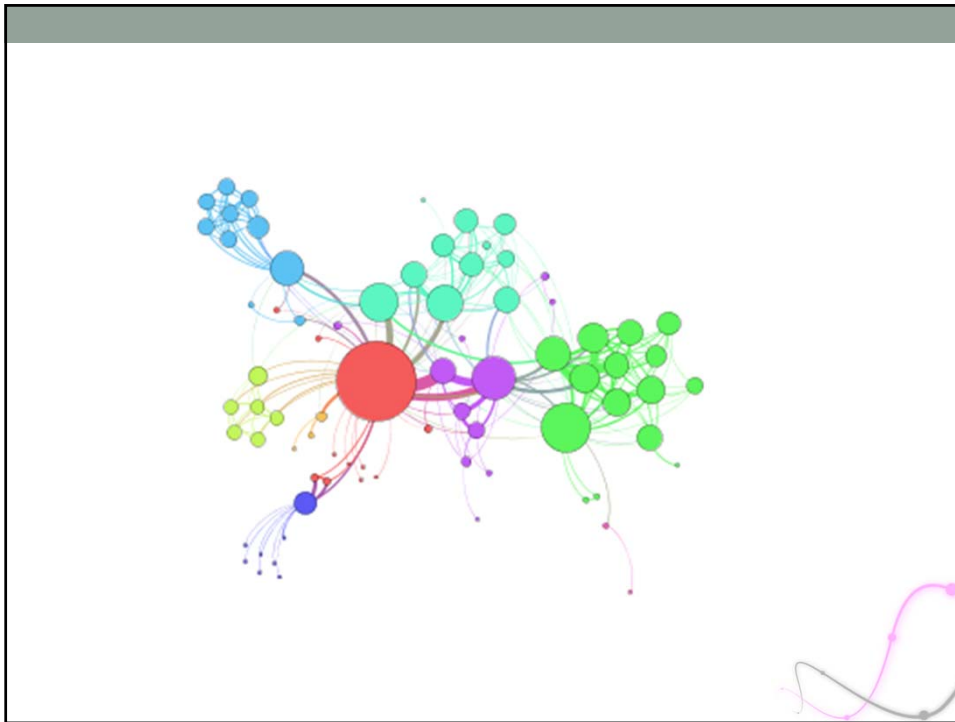
- ... they **sit** close geographically
- ... they **perform** the same kind of actions
- ... they **share** antecedents
- ... they **possess** common characteristics
- ... they **connect** to common entities
- ... they **transact** with each other
- ... they **are referred to** in pair
- ... they **refer to** common entities
- ... **or combinations of the above**

2. INSIGHTS TO BE GAINED

1. Community detection

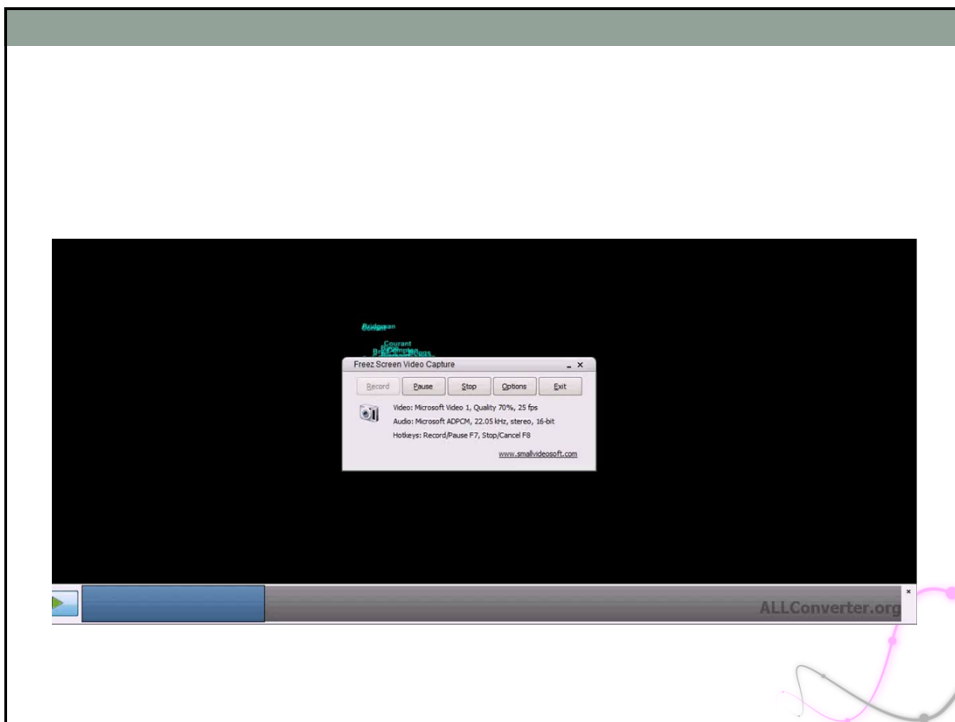
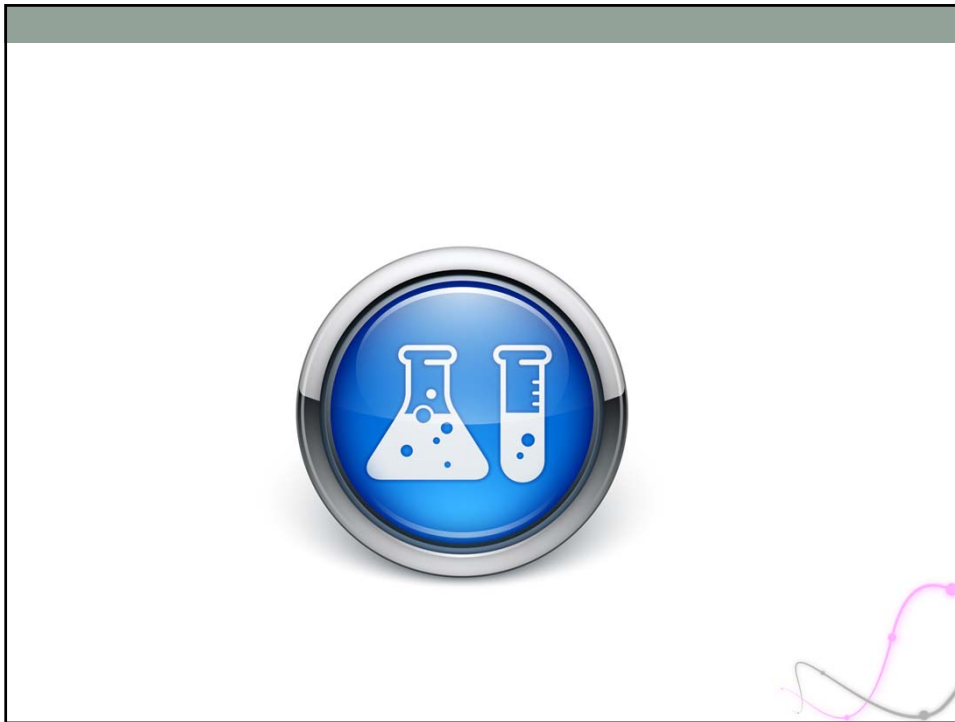
- A community / cluster is a group of nodes more densely connected with each other than with the rest of the network.
 - How to define “more densely connected”, and how to calculate it efficiently, is research in progress.
 - Two options:
 - Disjoint communities
 - Overlapping communities
- => Research on disjoint communities is currently more advanced.



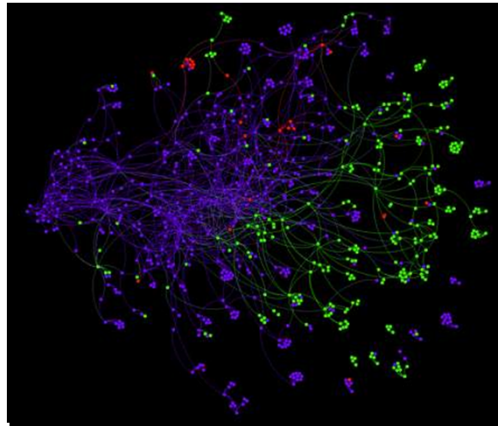


2. Sense of evolution, dynamics

- “Dynamics” in networks covers:
 - Topology
 - Creation and deletion of nodes and edges
 - Attributes
 - Change in the values of the quantitative and qualitative attributes of nodes and edges
- Network analysis is focusing on the topology
- Network visualization (with Gephi) accommodates both.



3. Overlay analysis



Social network of
scientists

Colored by they
research interests.

Colored by they
geographical
location

4. “The picture-examining eye is the best finder we have of the wholly unanticipated.”

- Exploratory analysis:

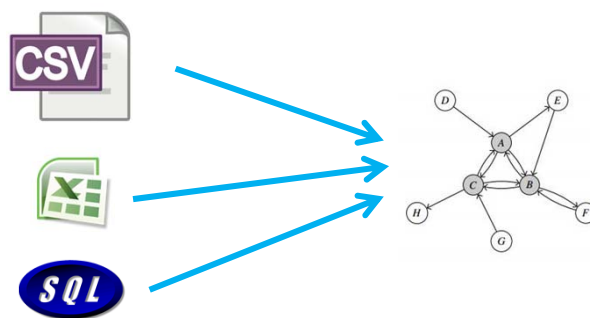
Visualizing the network stimulates the process of forming hypotheses.

- Needs confirmatory analysis at a later stage:

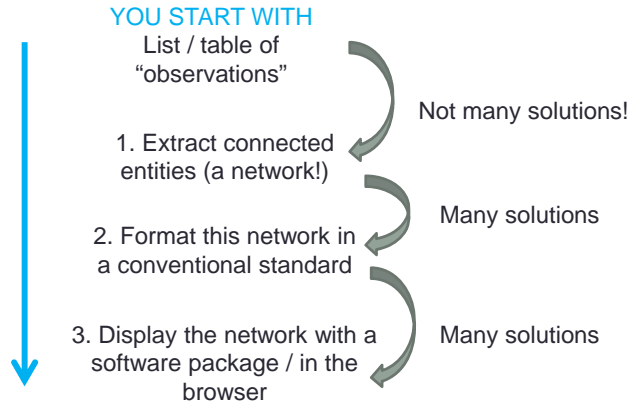
Either with network statistics, or with “traditional” forms of statistics / numerical analysis.

3. HOW TO WORK WITH CONNECTED DATA

The general idea



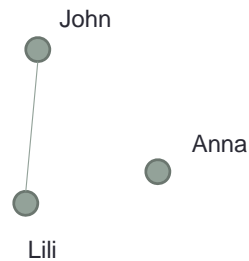
The steps



①

Extract connections through similarities

	A	B	C
1	John	coke	
2	John	red wine	
3	John	fanta	
4	Anna	wine	
5	Anna	orange juice	
6	Anna	milk	
7	Lili	coke	
8	Lili	fanta	
9	Lili	cocktails	



Gaze

Simple program that creates networks from similarities

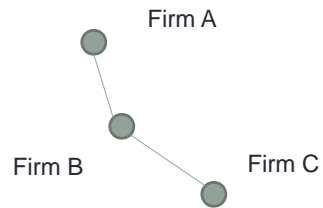
(available from www.clementvallois.net)

①

Extract connections in transactional data

May-12-2012 May-13-2012 May-14-2012

	A	B	C	D
1	Firm A	Firm B	500 €	12-May-12
2	Firm A	Firm B	200 €	13-May-12
3	Firm C	Firm B	1,500 €	14-May-12



Eonydis

Simple program that creates dynamic networks from transactional data

(available from www.clementvallois.net)



①

(not tested)



Table 2 Net

Load your CSV table

It has to be **comma-separated** and the first row must be dedicated to **column names**.

No file chosen

Note: you can drag and drop a file

by Mathieu Jacomy.
<http://tools.medialab.sciences-po.fr/table2net/>



②

Format the network in a standard

- A simple list of edges in a csv file is enough:

A,B

B,C

C,A

....

- Programming packaging for richer network formats (that include attributes for nodes and edges, and dynamics)

R, Python, Java...



③

Display the network

in an Internet browser



More than 150 nodes?

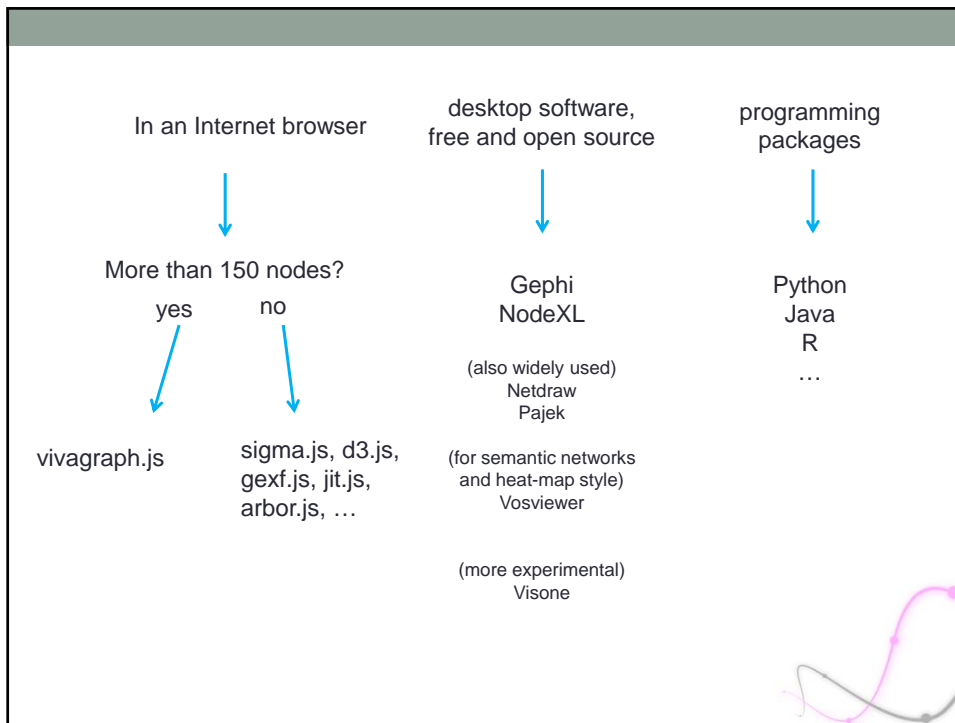
yes

no



vivagraph.js

sigma.js, d3.js,
gexf.js, jit.js,
arbor.js, ...with a desktop
softwareGephi
NodeXLCytoscape
Tulip
Guess
Netdraw
Pajek
Vosviewer
Visone
Tableauprogramming
packagesPython
Java
R
...



GOOD PRACTICES AND RED FLAGS

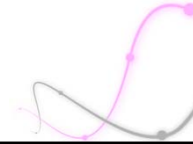
4 pieces of advice

① this is **visual** analytics - making a network readable is not a sin

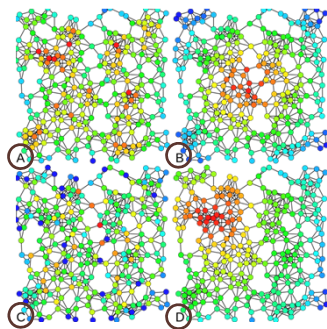
- add labels, captions
- curved versus straight edges
- color codes

this is not infographics – data are sacred

- size and position of nodes and edges are determined by a computational procedure
- choices in this domain depend on design (viz should be readable, engaging and insightful) but not at the expense of methodological soundness.



② Same network, different centralities



Legend:



Ref: <http://en.wikipedia.org/wiki/File:Centrality.svg>

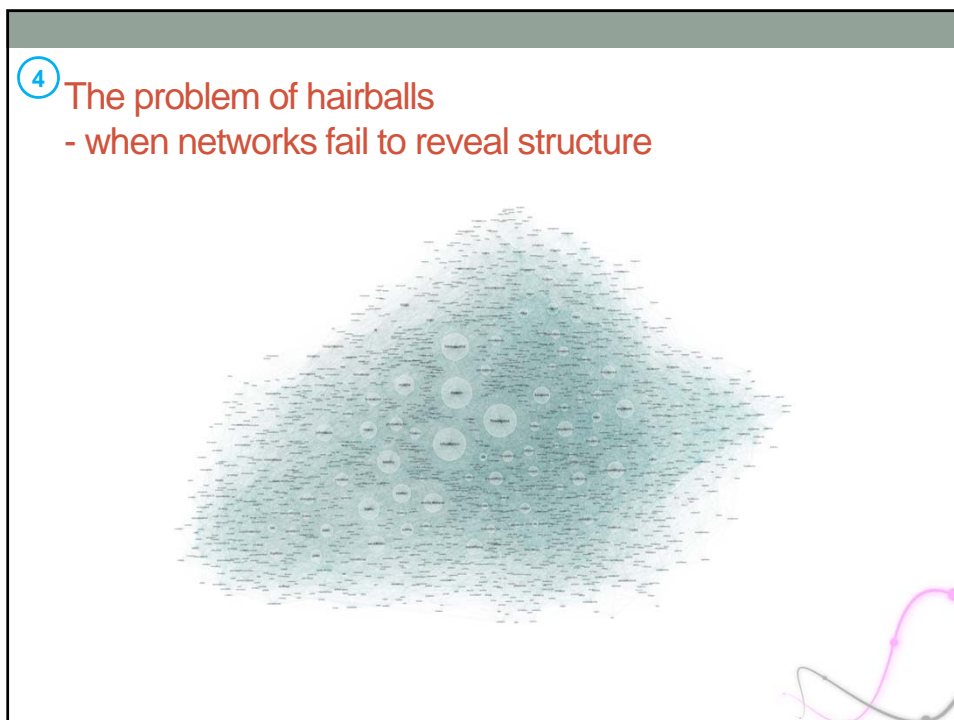
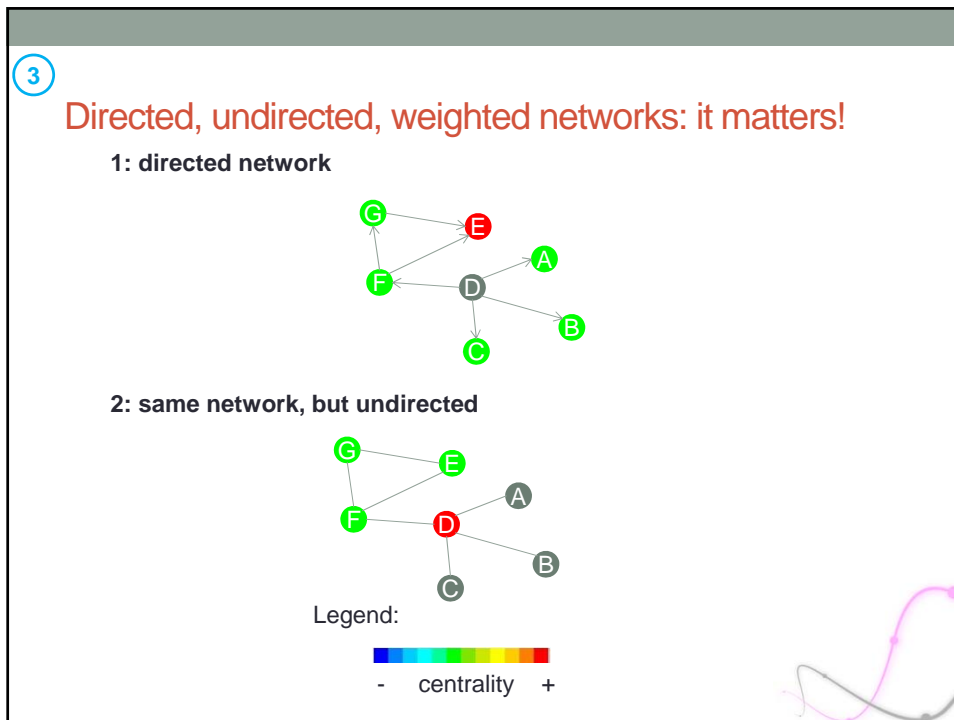
A: degree centrality = local connectivity
=> nodes having many edges are central

B: closeness centrality = "geographic" middle
=> nodes that are close to all other nodes are central

C: betweenness centrality = connectivity
=> nodes that lay on many shortest paths are central
(a shortest path is the quickest way to go from one node to another)

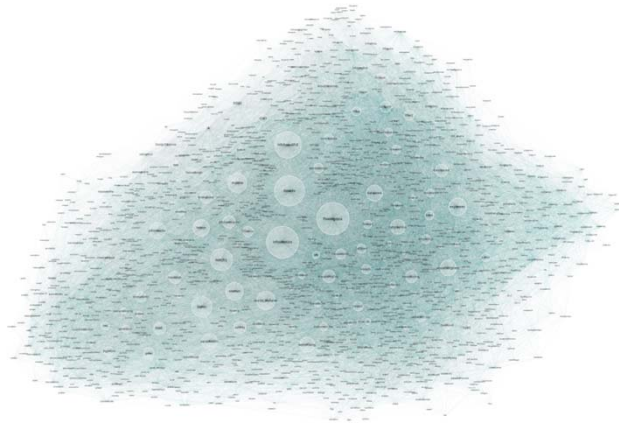
D: Eigenvector centrality = authority
=> nodes that are connected to highly ranked nodes are central (recursive approach).





5

Come with a purpose



- yes, the picture can yield unexpected insights
- does not mean that visualizations are a substitute for working at formulating a research question.



CONCLUSION

To go further

Data Visualization: a successful design process
Andy Kirk
PACKT

Beautiful Visualization
Looking at Data Through the Eyes of Experts
Edited by Julie Steele & Noah Ikeny
O'REILLY

Intentional Communication from Data to Display
Designing Data Visualizations
Noah Ikeny & Julie Steele
O'REILLY

NATHAN YAU
DATA POINTS
VISUALIZATION THAT MEANS SOMETHING
WILEY

Derek Hansen Ben Shneiderman Marc A. Smith
Analyzing Social Media Networks with NodeXL
Insights from a Connected World
MK

Gephi

 #dataviz, #datavis

5. DEMO

In 10 minutes from Excel to an online viz

Steps

- Import in Gephi
- Layout
- Community detection
- Partition
- Export to sigmaJs
- Upload to server



Thank you.



@seinecle

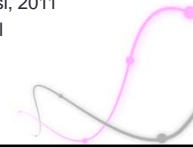
These slides will be online at:

www.clementvallois.net



Credits

- Migration map
 - Martin de Wulf, 2010
 - <http://migrationsmap.net>
- GapMinder
 - Gapminder.org
- LinkedIn Map
 - <http://inmaps.linkedinlabs.com>
- Facebook network of friendships
 - Paul Butler, 2010.
 - https://www.facebook.com/note.php?note_id=469716398919
- Musli ingredient network
 - Moritz Stefaner, 2012
 - <http://moritz.stefaner.eu/projects/musli-ingredient-network/>
- Flavor network
 - Yong-Yeol Ahn, Sebastian E. Ahnert, James P. Bagrow & Albert-László Barabási, 2011
 - http://www.nature.com/srep/2011/111215/srep00196/fig_tab/srep00196_F2.html



Credits

- Map of terms
 - Levallois, Clithero, Smidts, Wouters and Huettel (2012).
 - <http://www.nature.com/nrn/journal/v13/n11/full/nrn3354.html>
- Social network of scientists
 - Levallois (private data).
- Map of product categories
 - The Observatory of Economic Complexity, Alexander Simoes (2012).
 - <http://atlas.media.mit.edu/>
- Map of science
 - Rafols, Porter and Leydesdorff (2010).
 - <http://www.leydesdorff.net/overlaytoolkit/>

