



NESSHI and GEPHI

Sociology of science as a breeding ground for tool building in the digital humanities

Clement Levallois

Meertens Institute, 28 November 2013

Bio notes

PhD
(2003-2008)



Post-Doc
(2008-2013)

- Member of the Gephi Consortium



- Starting at EMLyon in January 2014



Plan

1. NESSHI (2011-2014)
2. Building tools to reveal insights from data in the NESSHI project
3. A focus on GEPHI as a platform of distribution for data processing tools (not just networks!)

The take away

- NESSHI and other empirically oriented projects are very likely to lead to the creation of data processing tools to support the analysis.
- Instead of “wasting” these tools in scripts that will never be reused, one could look for platforms of distribution to release them as apps. Gephi is such a platform.

1. NESSHI

The Neuro-Turn in European Social Sciences and Humanities:
Impact of neurosciences on economics, marketing and
philosophy

NESSHI



The neuro-turn

- A focus on empirical investigations
- Questioning the turn itself:
 - do we find it?
 - is it unidirectional?
- A look at academia in relation to society

Challenges

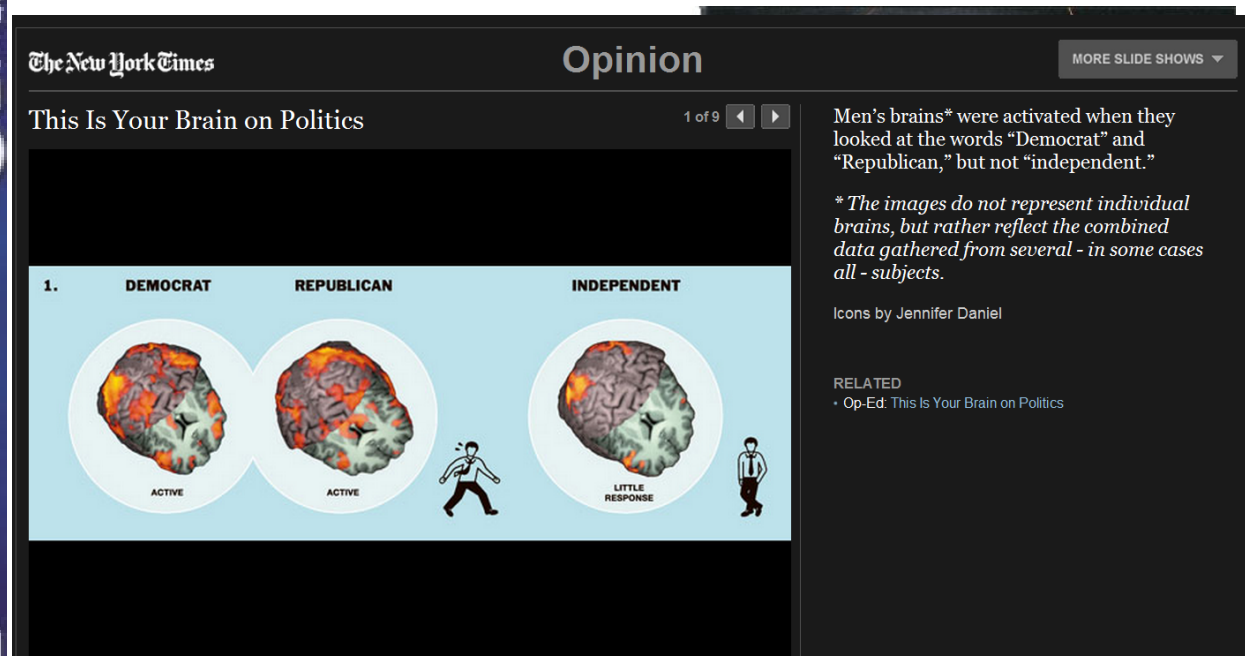
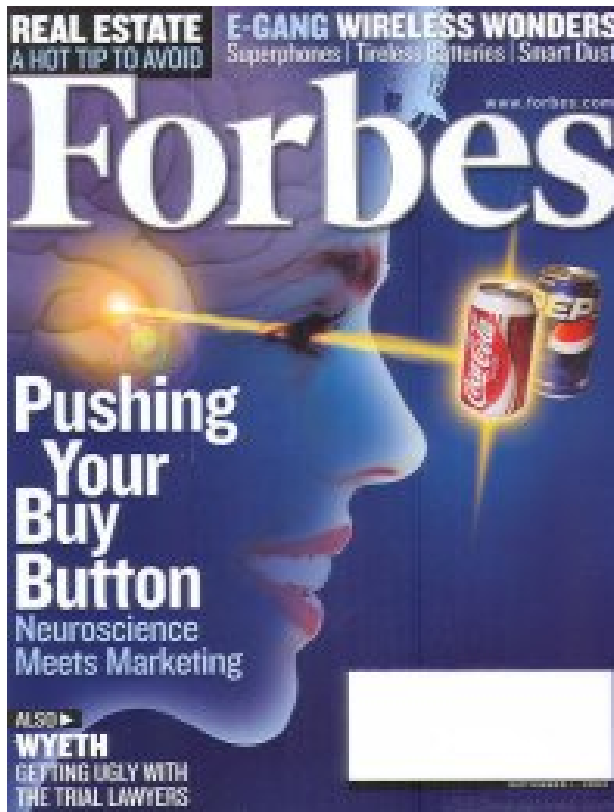
- We tried to track a broad, undefined intellectual movement through
 - Time (~1985-~2013?)
 - Scientific disciplines (all of them?)
 - Science/society spheres (beware dichotomies!)

- Archival work



NRC, 1996

- Archival work – on the web



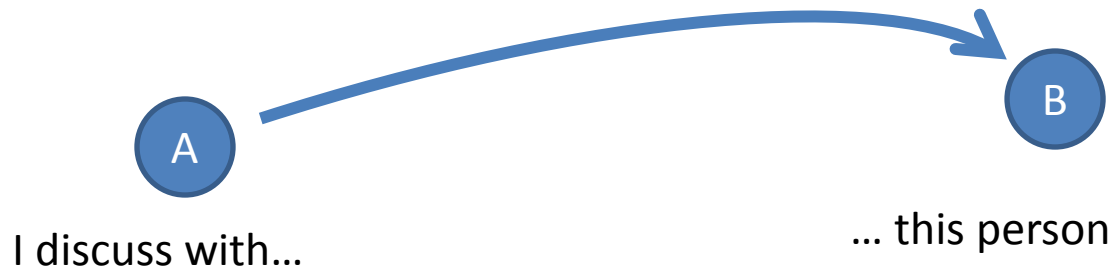
Neuromarketing in the press, 2003
(and 2007)

- Bibliographical data
 - Neuroeconomics papers overlaid on an ISI map
(Rafols, Porter and Leydesdorff overlay toolkit approach)
 - Co-author networks
 - Tried journal maps...

- Online surveys

- 630+ individual emails:

- “Who do you discuss with in neuroeconomics?”*



- Textual data
 - Web of Knowledge: cooccurrences in abstracts of papers in neuroeconomics
 - Mainz bibliography in neuroethics: clustering of papers in neuroethics based on a semi-manual classification scheme.

Research Group on Neuroethics/Neurophilosophy

arising from the rapid progress of the neurosciences have led to the development of a new scientific discipline: Neuroethics behavior as well as the applied ethics of neuroscience. The work of the Research Group on Neuroethics/Neurophilosophy for n-making and neuroscience & society.

Neuroethics Bibliography

y the Research Group on Neuroethics/Neurophilosophy aims at providing a complete and up-to-date overview of the publication i first starting point for researchers who want to engage in "Neuroethics". i about 3650 entries.

use click [here](#).



+

- Interviews (NL),
- Ethnographies (UK),
- Experimental work (France),
- Philosophical analysis (France, Germany)

Results (NL team)

- Neuroeconomics (in 2012): a complete survey of the field
- Neuromarketing (2001 -): a first history based on archives, in academia and industry
- Neuroethics (1995 – 2012): a definition based on the Mainz bibliography – collab with DE team.
- The neuro-turn society? Pictures of the neuro in Dutch newspapers (1987-2011) (on going)

2. DEVELOPMENT OF TOOLS FOR NESSHI

(note)

All software discussed here are available in open source with Apache License 2 or equivalent at:

<https://github.com/seinecle>

Why new tools?

- Unstructured data brings specific needs
- Technologies evolve quickly – tools built 3 years ago don't exploit all possibilities available today
- From scripting to releasing:
 - A script can take days to create, why not add a few hours of work to make it useable by the larger public?
- Creating tools can be fun (actually, addictive).
- Is it worth always? Does it work always? No...

1. Calenicon (show demo at www.calenicon.org)

- Problem: how to edit / share / archive pictures in an online environment?
- Using the vocabulary of VRAcore for annotations
VRAcore: data standard for the description of works of visual culture as well as the images that document them. <http://www.loc.gov/standards/vracore/>
- Uses the API of www.datahub.io (CKAN), the repository of the Open Knowledge Foundation (Figshare and DANS were other candidates).
- Developed in Java EE.
- www.calenicon.org

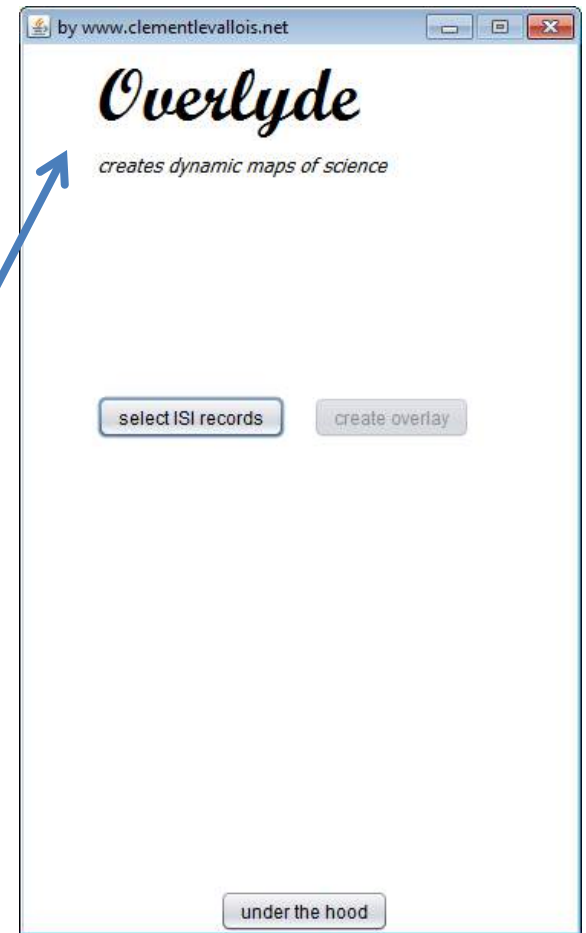
2. Alphabetical sorter

- Problem: how to get a neat alphabetical list of scientists in Inkscape or Illustrator?
- Developed in Java as a plugin in Gephi
- Trivial (40 lines of code).

(open pdf to show example of use)

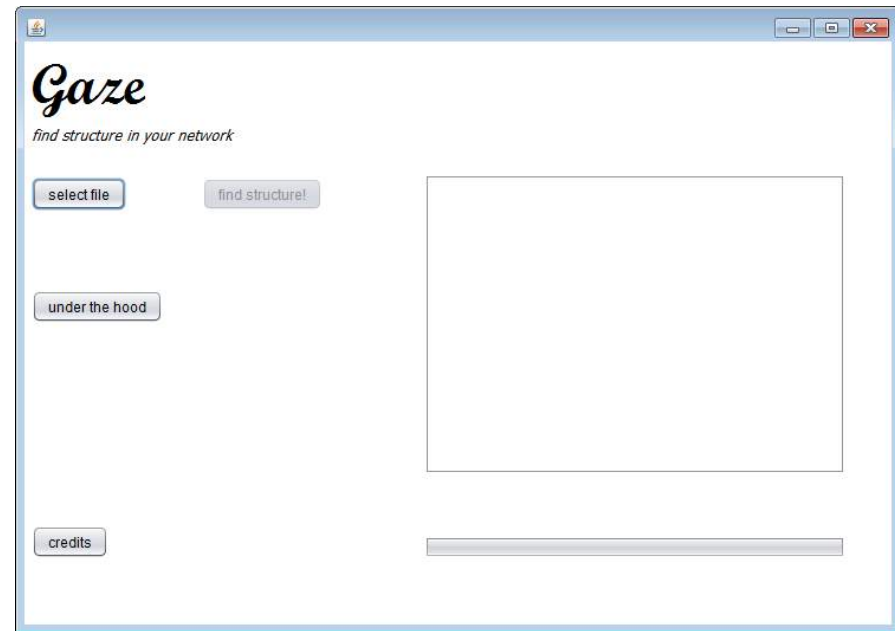
3. Overlay toolkit

- Problem: exploratory work for the history of neuroethics project
 - can we see an evolution of a neuroethics in terms of ISI categories?
- Overlay toolkit by Rafols, Porter and Leydesdorff (2010)
 - Originally a series of command-line utilities available from Loet Leydesdorff website, with an output to Pajek.
- Excel macro to enable an export to Gephi (2011)
- Overlyde (2012): stand alone Java program to create dynamic viz in Gephi.
- Now a Gephi plugin (2013)
 - Allows for dynamic datasets, filtering, relative views...



4. Gaze

- Problem: how to create networks of agents, based on attributes they share?
(attempt to create a map of journals)
(launch gephi file to illustrate)
- Gaze computes similarities between agents and creates the resulting network



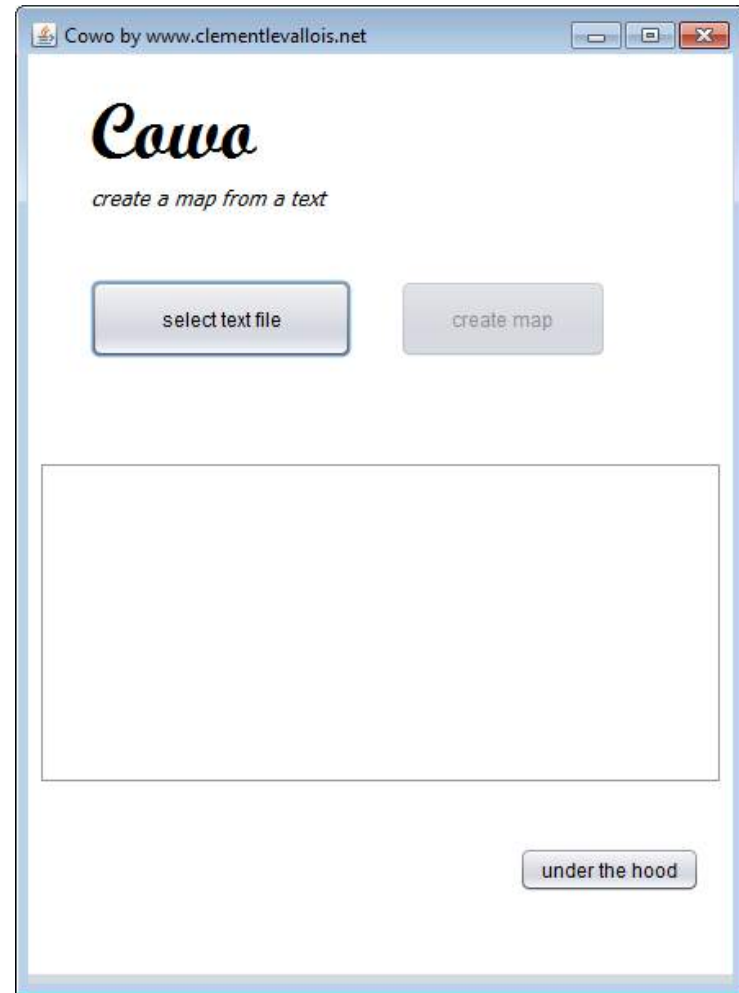
5. Cowo

- Problem:

For the map of cooccurences in neuroeconomics abstracts I need:

- “functional magnetic resonance imaging” (4-gram) in VosViewer!
- And adjectives too. And no stemming. And no td-idf weighting. Etc.

- Cowo provides an n-gram approach instead of a POS tagging approach for text analysis.



6. Force Atlas 3D

- Problem: can't we get more insight on the structure of a network with a layout in 3D space?
(GraphInsight provides 3D layouts, but not as convenient to use as in Gephi IMHO)
- Solution: plugin in Gephi
(built on top of Force Atlas 2)


3. TOOLS – LESSONS LEARNED

Some things fail

- Calenicon? Project members did not use it
- I am the only user, not a good effort/results ratio!
- Got no star or forks on Github 😞
- Time of development would be too heavy to make it useable by anybody with a www.datahub.io account.
- Not sure of the extent of the potential user base

Some things work but remain confidential

- Cowo: 6x
- Overlyde: 0x
- Gaze: 0x



Monthly downloads from
www.clementvallois.net
in the 3 months leading to Nov 2013

Some things work

- Alphabetical Sorter: 20x
- Force Atlas 3D: 120x

- Excel/csv importer: 300x
- Map of countries: 210x

2 general purpose plugins I released for the Gephi community

Monthly downloads from
<https://marketplace.gephi.org>
Since their release (all sometime in 2013)

Figures can be checked on the website
above.

(the big difference in downloads is
probably due to the distribution
channel)

4. HOW TO MAXIMIZE THE IMPACT OF TOOLS?

Choose the distribution channel wisely

DISCLAIMER: next slides might read like an advertisement for Gephi.

I am a member of the “Gephi Community” support team and I have no stake in the software – just sharing lessons learned here.

Several ingredients needed

1. Easy to use
2. Cross platform support
3. Easy to install
4. Can be found
5. Meets a need

(2. and 3. are necessary but not sufficient – see the low number of downloads on my apps reported above).

How to share tools?

- Desktop applications
 - Loet Leydesdorff command-line tools
 - VosViewer
 - Zotero desktop
- Web applications
 - Issue Crawler <https://www.issuecrawler.net/>
 - Sc-Po MediaLab <http://www.medialab.sciences-po.fr/tools/>
- Sharing the source code
 - Git (Github, Bitbucket, SourceForge...)
 - Run My Code (<http://www.runmycode.org/>)
- Via platforms
 - Firefox: offers plugins like Zotero
 - Gephi: can be extended with plugins
 - R: packages can be reviewed and shared widely

| | Can be discovered? | Runs on Mac, PC, Linux | Easy to install / user friendliness |
|-------------------|--------------------------------|------------------------|---|
| Desktop app | Depends on your own notoriety. | NO (except for Java) | NO, with exceptions (well designed UI and good doc. Example: Vosviewer) |
| Web app | Depends on your own notoriety. | YES | YES |
| Run my code / Git | YES – for techy audiences | Not guaranteed | NO, except for specialized audiences (example: an R script for R users) |
| ➡ Gephi | YES - wide audience | YES | YES |

Gephi is a platform because

- Extensible with plugins (your tool!) with extensive support to create them
 - <https://github.com/gephi/gephi-plugins-bootcamp>
- Functions for branding: your tool can have its own identity (not obliged to blend into Gephi)
- License: free including for commercial purposes. Your plugin can have a different license.
- OpenGL visualization engine
- GUI all managed by the underlying Netbeans platform
- All java libraries available
- Includes a “data laboratory” for spreadsheet-like operations
- And I/O, client-server communication, hardware acceleration, auto-update functionalities, etc.

Example: API client for Europeana

- Use case: creating an app to query Europeana
 - To retrieve meta-data, including pictures in some cases
- Desktop app or Europeana website?
 - Desktop app: how to make it cross platform? How to make this app discoverable by a large user base?
 - Europeana website: no possibility to export in a personalized way
- Create the Europeana as a Gephi plugin:
 - gives the flexibility of a rich client application (RCA),
 - without the pain to develop a RCA from scratch,
 - and provides a popular distribution channel for the app

The take away (again)

- NESSHI and other empirically oriented projects are very likely to lead to the creation of data processing tools to support the analysis.
- Instead of “wasting” these tools in scripts that will rarely be reused, one could look for platforms of distribution to release them as apps. Gephi is such a platform.

Thank you.



@seinecle

These slides are online at:

www.clementvallois.net