

What is Gephi used for?

Networks, and why we visualize them

Mathieu Jacomy
Aalborg University TANTLab

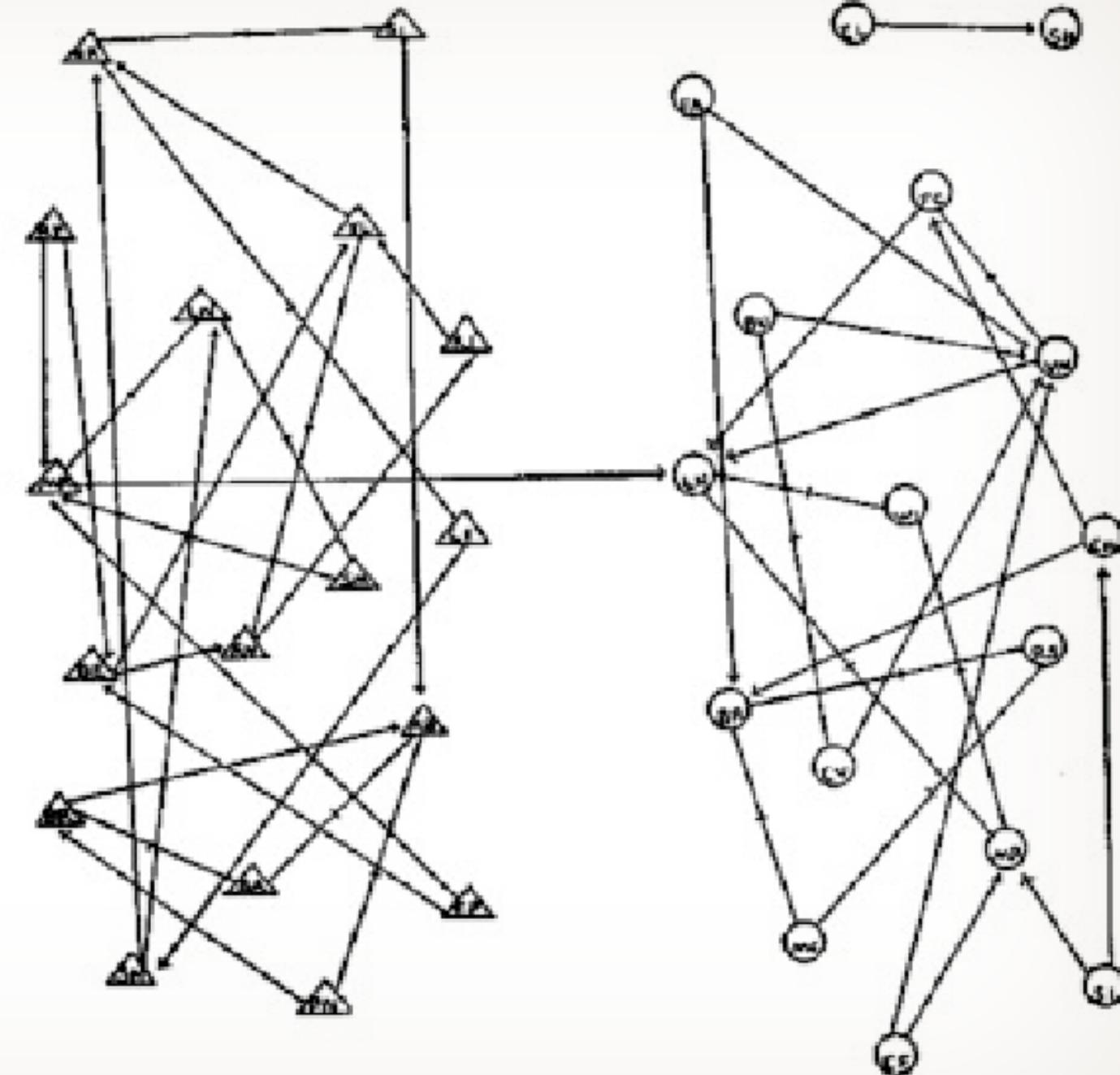
I.

Examples of networks

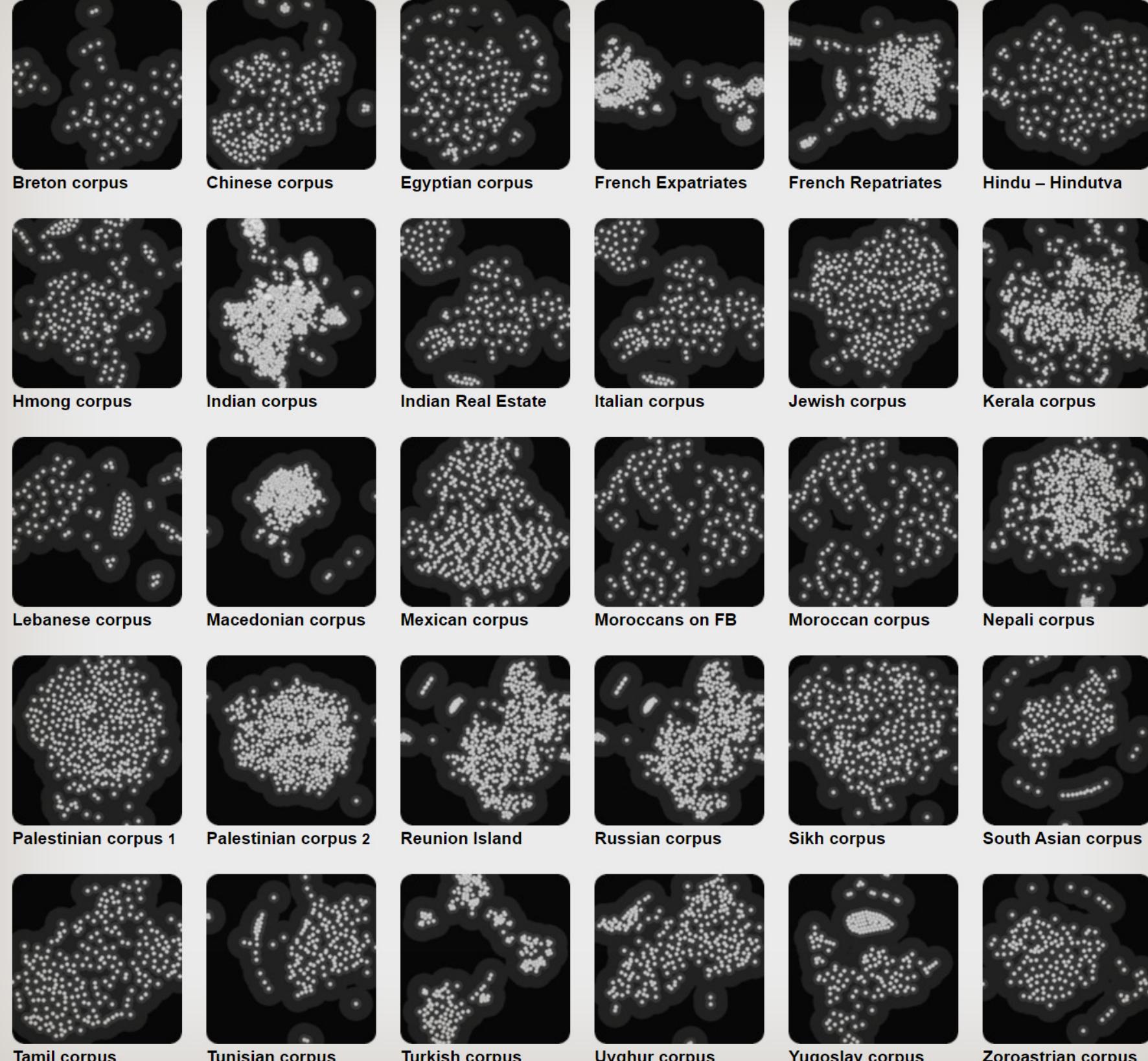
The sociogram

EMOTIONS MAPPED BY NEW GEOGRAPHY

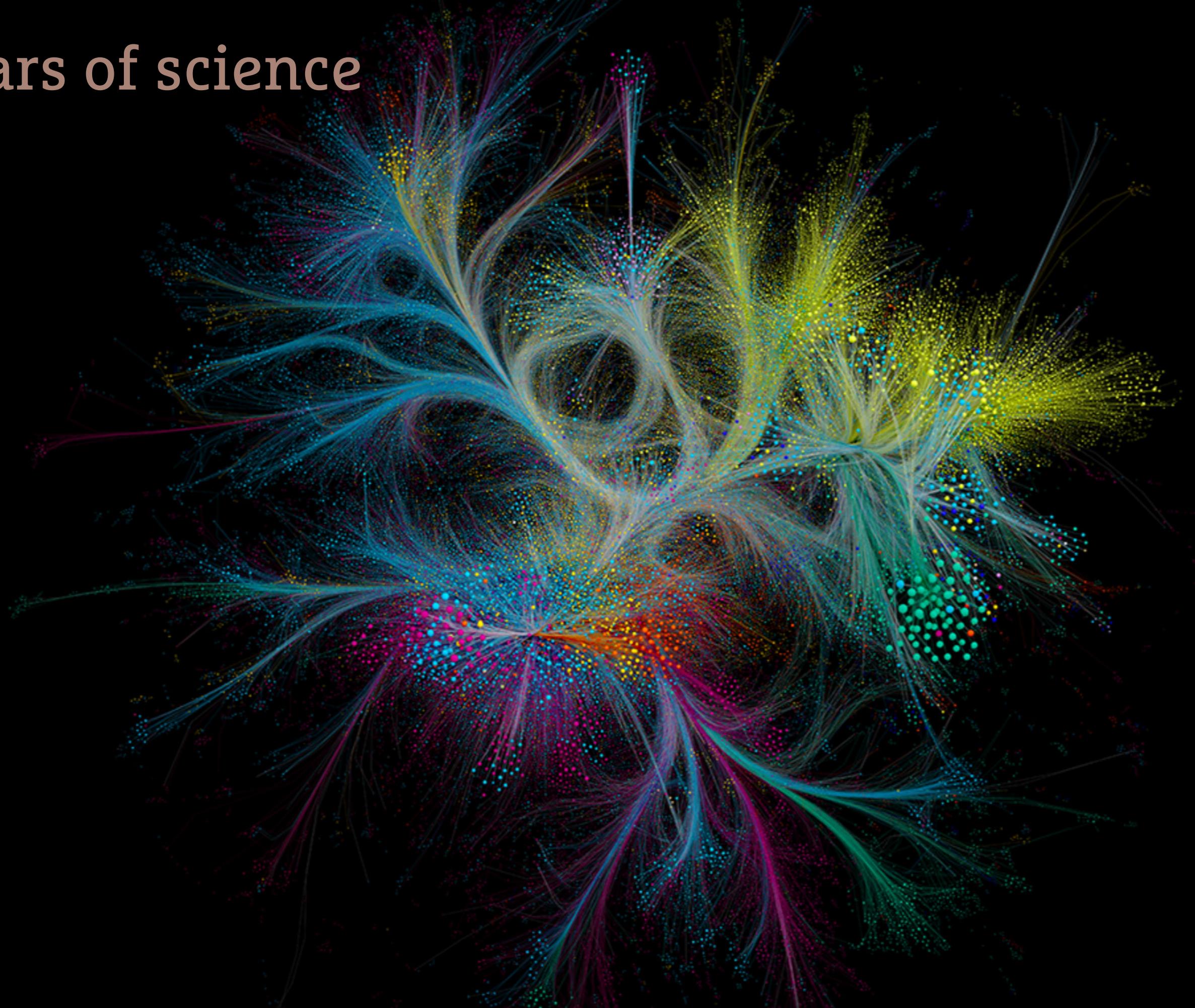
Charts Seek to Portray the Psychological Currents of Human Relationships.



e-Diasporas Atlas



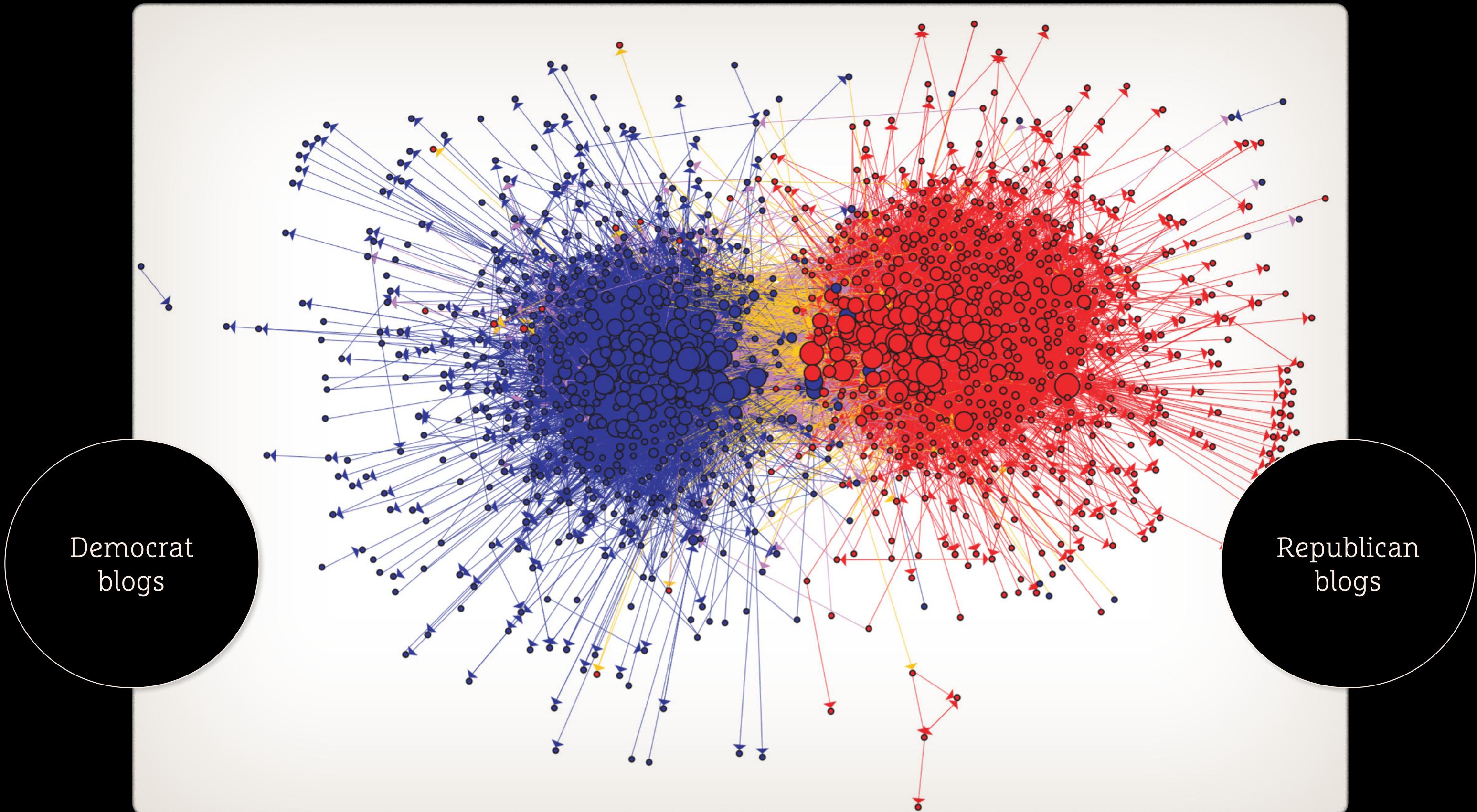
150 years of science



BarabasiLab for Nature, 2019

<https://www.nature.com/immersive/d41586-019-03165-4/index.html>

“Divided they Blog.” The most famous network?



Adamic, L. A., & Glance, N. (2005, August). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (pp. 36-43). ACM.

“Divided they Blog.” The most famous network?

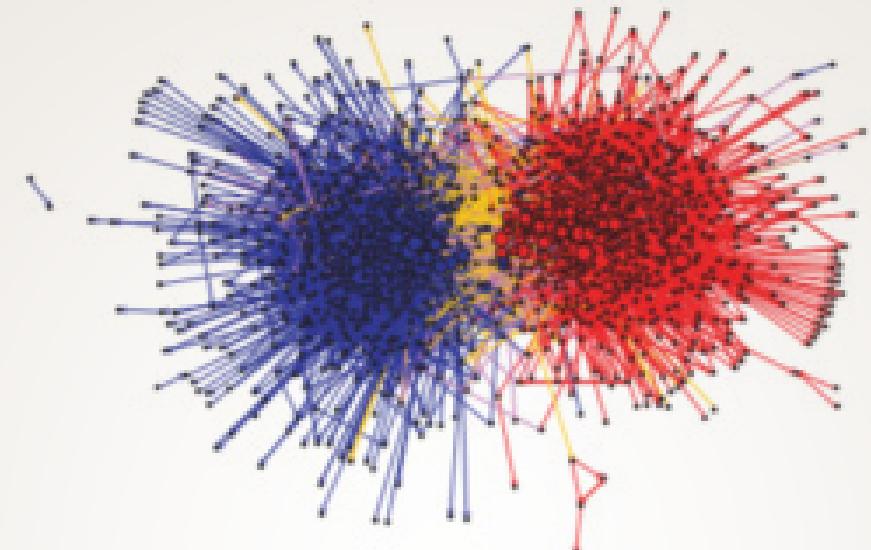


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

longer existed, or had moved to a different location. When looking at the front page of a blog we did not make a distinction between blog references made in blogrolls (blogroll links) from those made in posts (post citations). This had the disadvantage of not differentiating between blogs that were actively mentioned in a post on that day, from blogroll links that remain static over many weeks [10]. Since posts usually contain sparse references to other blogs, and blogrolls usually contain dozens of blogs, we assumed that the network obtained by crawling the front page of each blog would strongly reflect blogroll links. 47% blogs had blogrolls through blogrolling.com, while many others simply maintained a list of links to their favorite blogs. We did not include blogrolls placed on a secondary page.

We constructed a citation network by identifying whether a URL present on the page of one blog references another political blog. We called a link found anywhere on a blog's page, a “page link” to distinguish it from a “post citation”, a link to another blog that occurs strictly within a post. Figure 1 shows the unavoidable division between the liberal and conservative political (blog)ospheres. In fact, 91% of the links originating within either the conservative or liberal communities stay within that community. An effect that may not be as apparent from the visualization is that even though we started with a balanced set of blogs, conservative blogs show a greater tendency to link. 84% of conservative blogs link to at least one other blog, and 62% receive a link. In contrast, 71% of liberal blogs link to another blog, while only 67% are linked to by another blog. So overall, we see a slightly higher tendency for conservative blogs to link. Liberal blogs linked to 13.6 blogs on average, while conservative blogs linked to an average of 15.1, and this difference is almost entirely due to the higher proportion of liberal blogs with no links at all.

Although liberal blogs may not link as generously on average, the most popular liberal blog, Daily Kos and Eschaton (atrias.blogspot.com), had 328 and 264 links from our single-day snapshot

“Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout in the GUESS visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.”

Adamic, L. A., & Glance, N. (2005, August). The political blogosphere and the 2004 US election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (pp. 36-43). ACM.

“Divided they Blog.” The most famous network?

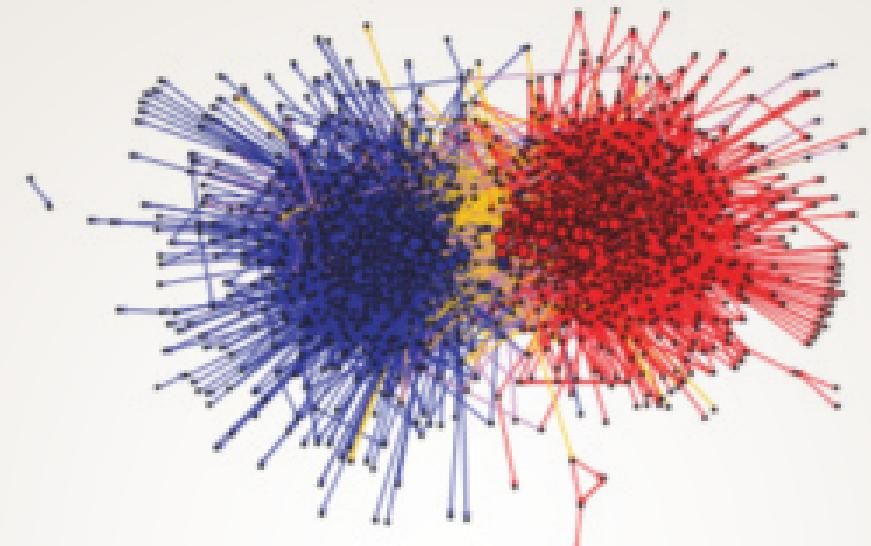


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

longer existed, or had moved to a different location. When looking at the front page of a blog we did not make a distinction between blog references made in blogrolls (blogroll links) from those made in posts (post citations). This had the disadvantage of not differentiating between blogs that were actively mentioned in a post on that day, from blogroll links that remain static over many weeks [10]. Since posts usually contain sparse references to other blogs, and blogrolls usually contain dozens of blogs, we assumed that the network obtained by crawling the front page of each blog would strongly reflect blogroll links. 47% blogs had blogrolls through blogrolling.com, while many others simply maintained a list of links to their favorite blogs. We did not include blogrolls placed on a secondary page.

We constructed a citation network by identifying whether a URL present on the page of one blog references another political blog. We called a link found anywhere on a blog's page, a “page link” to distinguish it from a “post citation”, a link to another blog that occurs strictly within a post. Figure 1 shows the unmistakable division between the liberal and conservative political (blogs)spheres. In fact, 91% of the links originating within either the conservative or liberal communities stay within that community. An effect that may not be as apparent from the visualization is that even though we started with a balanced set of blogs, conservative blogs show a greater tendency to link. 84% of conservative blogs link to at least one other blog, and 62% receive a link. In contrast, 71% of liberal blogs link to another blog, while only 67% are linked to by another blog. So overall, we see a slightly higher tendency for conservative blogs to link. Liberal blogs linked to 13.6 blogs on average, while conservative blogs linked to an average of 15.1, and this difference is almost entirely due to the higher proportion of liberal blogs with no links at all.

Although liberal blogs may not link as generously on average, the most popular liberal blog, Daily Kos and Eschaton (atrix.blogspot.com), had 328 and 264 links from our single-day snapshot

“What immediately stands out is the extreme separation between liberals and conservatives. ... Just like the real-world political networks ..., the online social network appears to be strongly homophilous and polarized.”

Christakis, N. A., & Fowler, J. H. (2009). Connected: The surprising power of our social networks and how they shape our lives. Little, Brown Spark.

“Divided they Blog.” The most famous network?

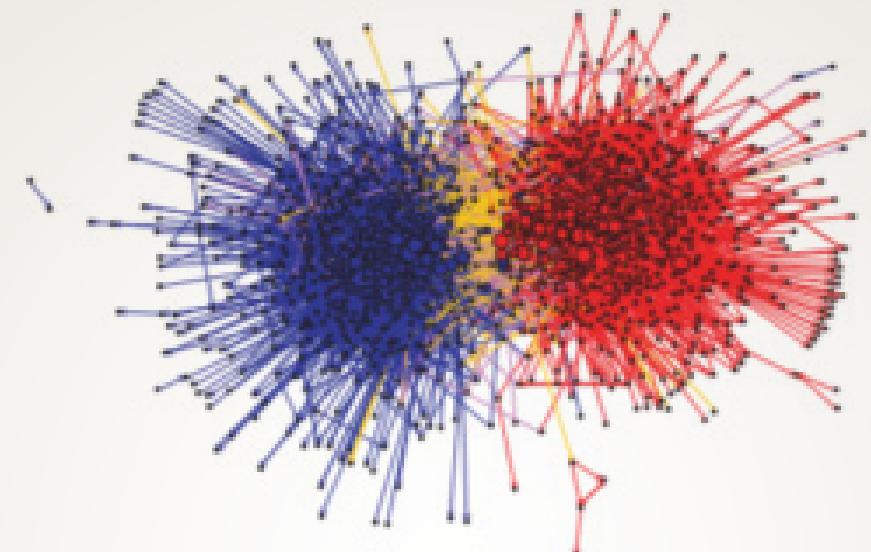


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

longer existed, or had moved to a different location. When looking at the front page of a blog we did not make a distinction between blog references made in blogrolls (blogroll links) from those made in posts (post citations). This had the disadvantage of not differentiating between blogs that were actively mentioned in a post on that day, from blogroll links that remain static over many weeks [10]. Since posts usually contain sparse references to other blogs, and blogrolls usually contain dozens of blogs, we assumed that the network obtained by crawling the front page of each blog would strongly reflect blogroll links. 47% blogs had blogrolls through blogrolling.com, while many others simply maintained a list of links to their favorite blogs. We did not include blogrolls placed on a secondary page.

We constructed a citation network by identifying whether a URL present on the page of one blog references another political blog. We called a link found anywhere on a blog's page, a “page link” to distinguish it from a “post citation”, a link to another blog that occurs strictly within a post. Figure 1 shows the unavoidable division between the liberal and conservative political (blogs)spheres. In fact, 91% of the links originating within either the conservative or liberal communities stay within that community. An effect that may not be as apparent from the visualization is that even though we started with a balanced set of blogs, conservative blogs show a greater tendency to link. 84% of conservative blogs link to at least one other blog, and 62% receive a link. In contrast, 71% of liberal blogs link to another blog, while only 67% are linked to by another blog. So overall, we see a slightly higher tendency for conservative blogs to link. Liberal blogs linked to 13.6 blogs on average, while conservative blogs linked to an average of 15.1, and this difference is almost entirely due to the higher proportion of liberal blogs with no links at all.

Although liberal blogs may not link as generously on average, the most popular liberal blog, Daily Kos and Eschaton (atrix.blogspot.com), had 328 and 264 links from our single-day snapshot

“Finally, and most critically, we call on computational social scientists, especially network scientists, to interrogate their own visualization practices. ... [C]onstructing network graphs remains as much an art as a science, with few conventions regarding the “right” way to represent node-link data.”

Foucault Welles, B., & Meirelles, I. (2015). Visualizing computational social science: The multiple lives of a complex image. Science Communication, 37(1), 34-58.

“Divided they Blog.” The most famous network?

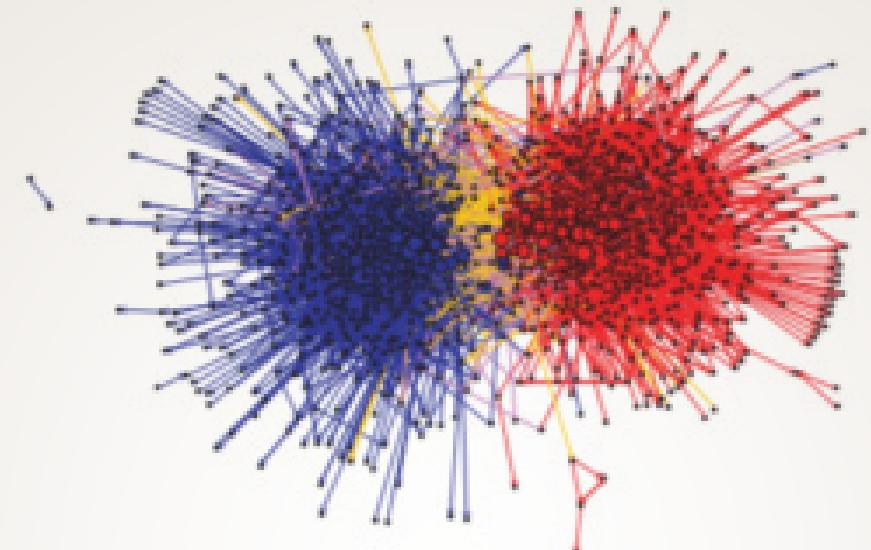


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

longer existed, or had moved to a different location. When looking at the front page of a blog we did not make a distinction between blog references made in blogrolls (blogroll links) from those made in posts (post citations). This had the disadvantage of not differentiating between blogs that were actively mentioned in a post on that day, from blogroll links that remain static over many weeks [10]. Since posts usually contain sparse references to other blogs, and blogrolls usually contain dozens of blogs, we assumed that the network obtained by crawling the front page of each blog would strongly reflect blogroll links. 47% blogs had blogrolls through blogrolling.com, while many others simply maintained a list of links to their favorite blogs. We did not include blogrolls placed on a secondary page.

We constructed a citation network by identifying whether a URL present on the page of one blog references another political blog. We called a link found anywhere on a blog's page, a “page link” to distinguish it from a “post citation”, a link to another blog that occurs strictly within a post. Figure 1 shows the unmistakable division between the liberal and conservative political (blogs)spheres. In fact, 91% of the links originating within either the conservative or liberal communities stay within that community. An effect that may not be as apparent from the visualization is that even though we started with a balanced set of blogs, conservative blogs show a greater tendency to link. 84% of conservative blogs link to at least one other blog, and 62% receive a link. In contrast, 71% of liberal blogs link to another blog, while only 67% are linked to by another blog. So overall, we see a slightly higher tendency for conservative blogs to link. Liberal blogs linked to 13.6 blogs on average, while conservative blogs linked to an average of 15.1, and this difference is almost entirely due to the higher proportion of liberal blogs with no links at all.

Although liberal blogs may not link as generously on average, the most popular liberal blog, Daily Kos and Eschaton (atrix.blogspot.com), had 328 and 264 links from our single-day snapshot

“This figure summarizes the link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.”

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Jebara, T. (2009). Computational social science. *Science*, 323(5915), 721-723.

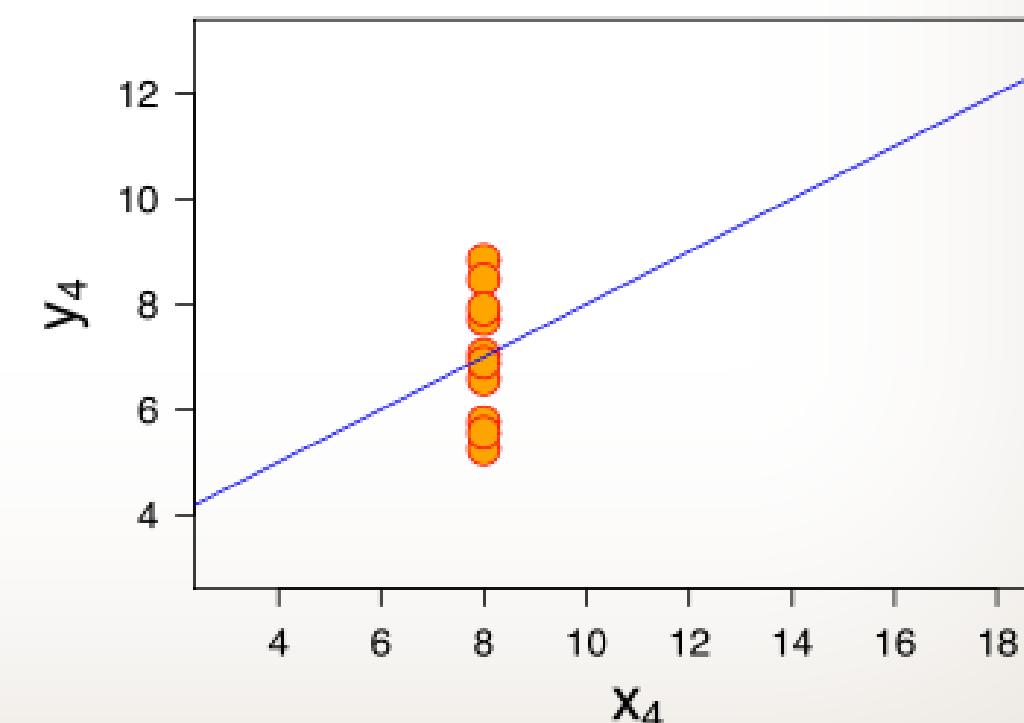
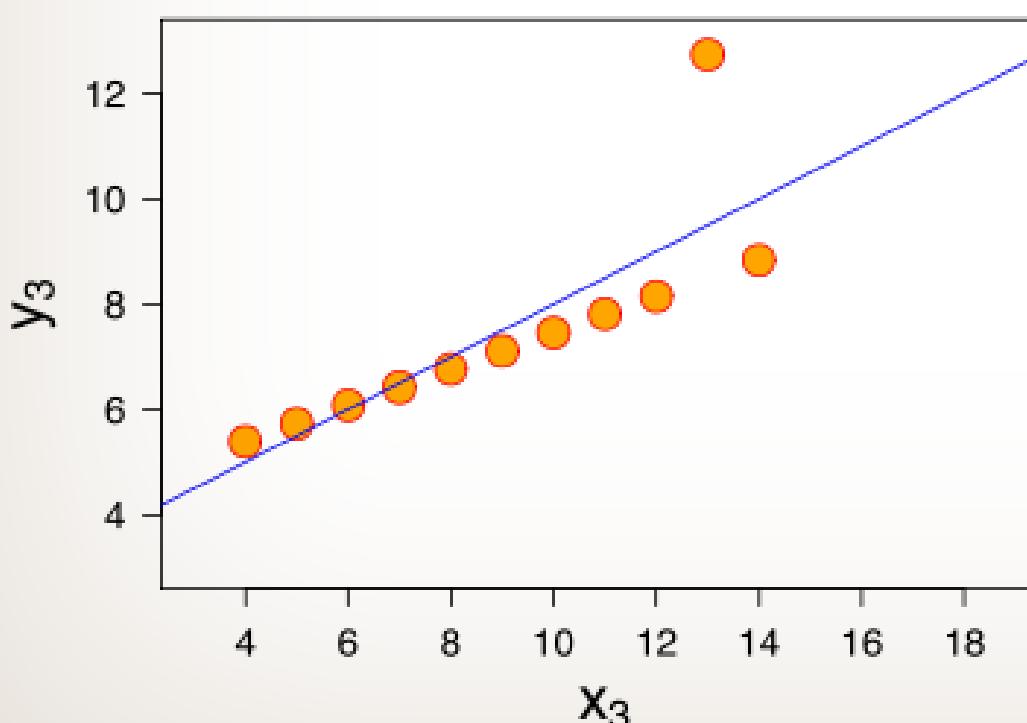
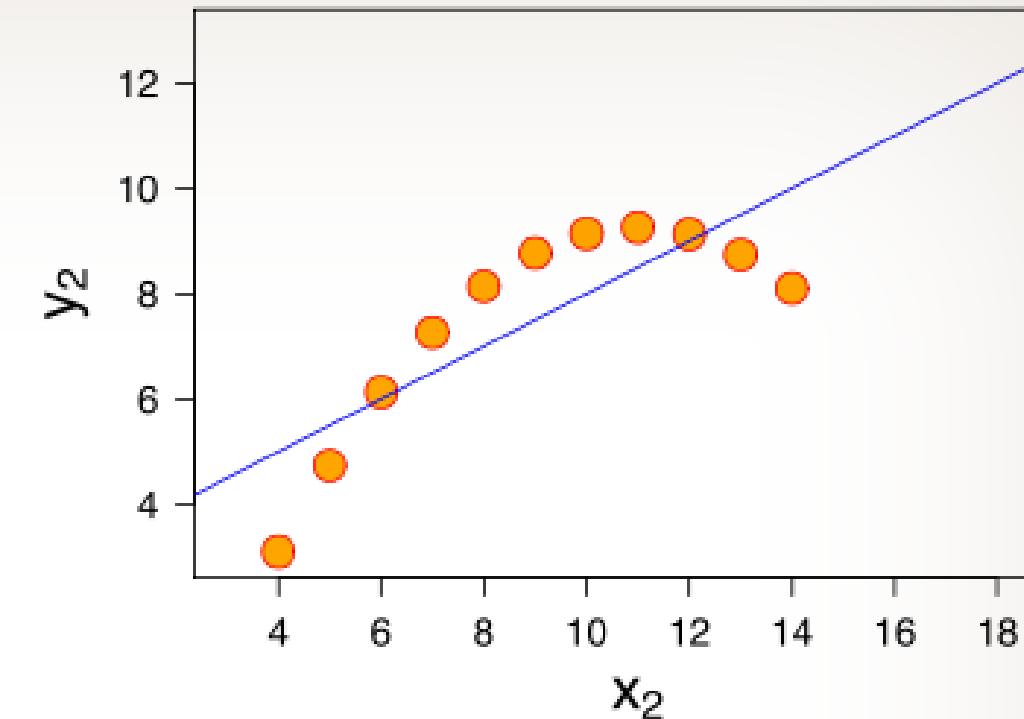
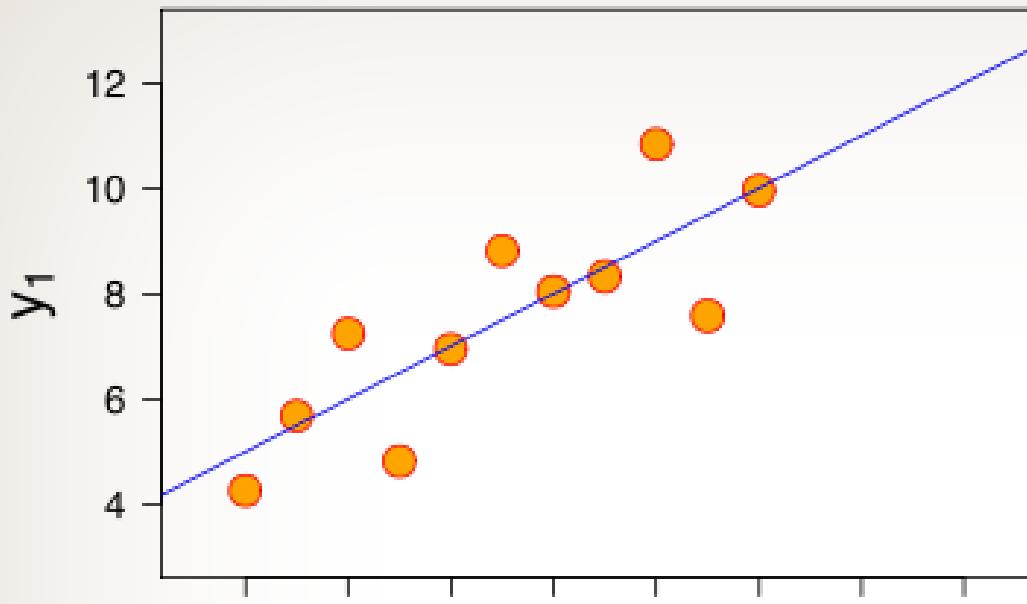
II.

Why visualizing?

Why visualizing?

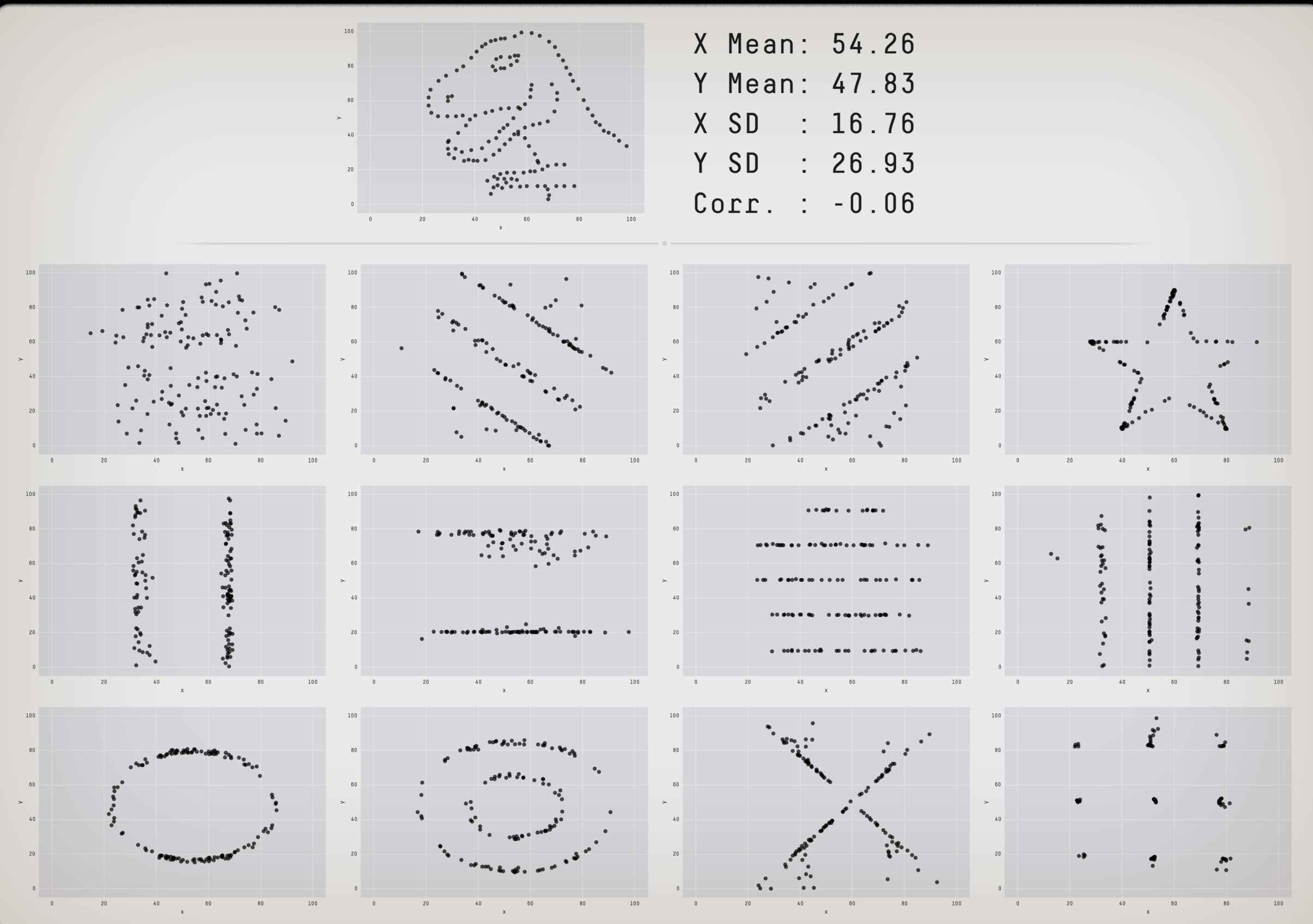
Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	± 0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places

Why visualizing?



The Anscombe quartet

Why visualizing?



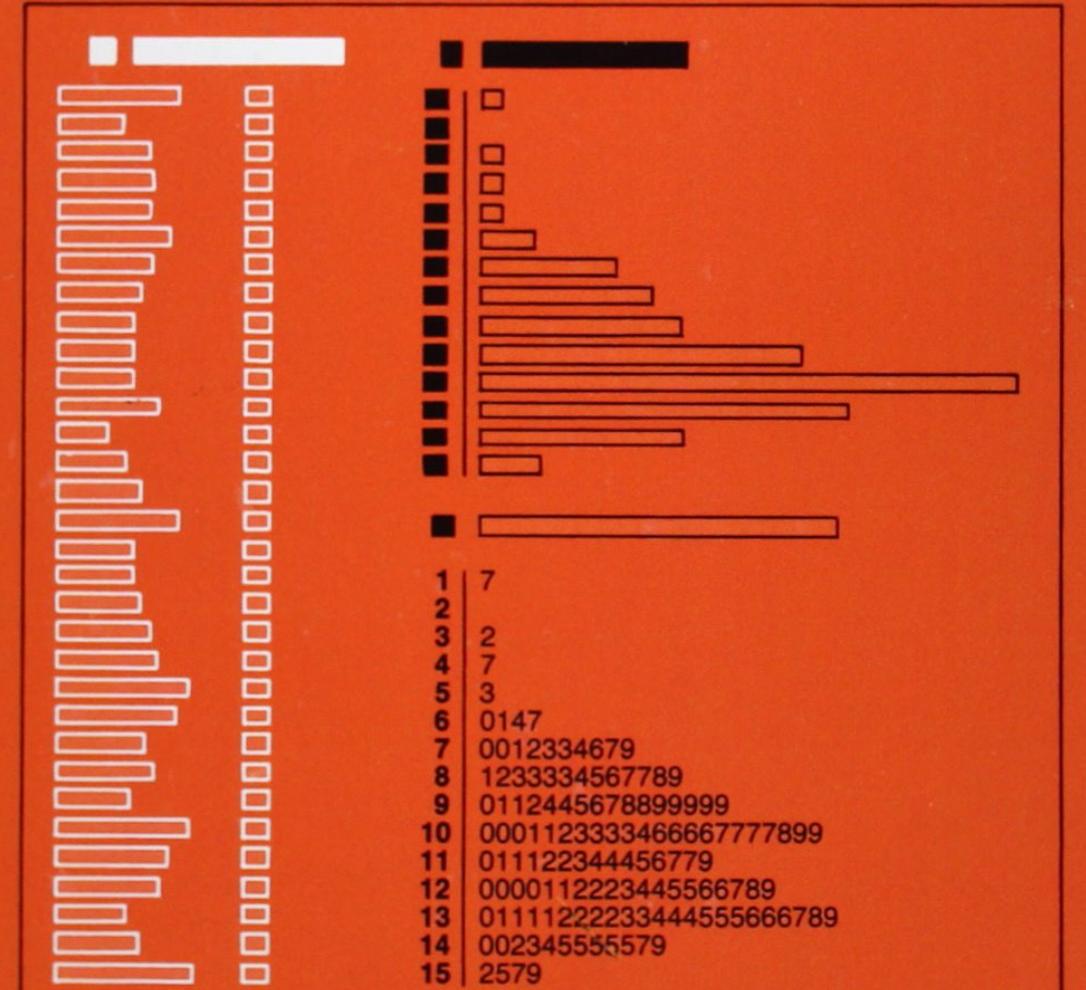
“Datasaurus”

<https://www.autodeskresearch.com/publications/samestats>

Exploratory Data Analysis (EDA)

John W. Tukey

EXPLORATORY DATA ANALYSIS



👉 John W. Tukey

“The greatest value of a picture is when it forces us to notice what we never expected to see.”

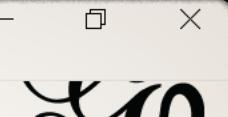
— John W. Tukey

“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.”

— John W. Tukey

III.

What is Gephi?



Workspace 1

Appearance

Nodes Edges



Unique Partition Ranking

#c0c0c0

Layout

ForceAtlas 2



Run

Threads

Threads number 7

Performance

Tolerance (speed) 1.0

Approximate Repulsion

Approximation 1.2

Tuning

Scaling 2.0

Stronger Gravity

Gravity 1.0

Behavior Alternatives

Dissuade Hubs

LinLog mode

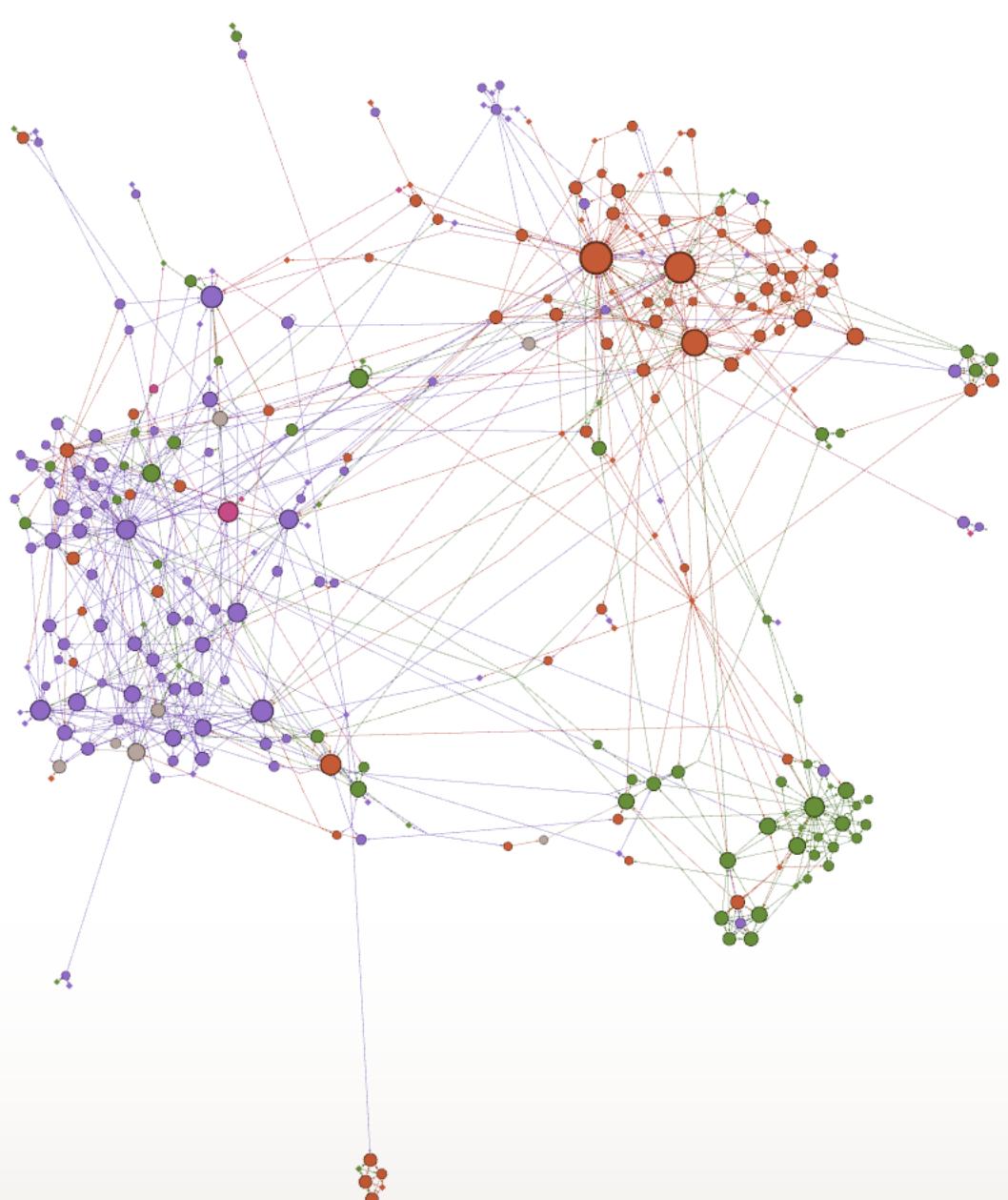
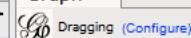
Prevent Overlap

Edge Weight Influence 1.0

ForceAtlas 2

Presets... Reset

Graph



Context

Nodes: 366

Edges: 1107

Directed Graph

Filters Statistics

Settings

Network Overview

Average Degree

Avg. Weighted Degree

Network Diameter

Graph Density

Bridging Centrality

HITS

Modularity

PageRank

Connected Components

Girvan-Newman Clustering

Leiden algorithm

Node Overview

Avg. Clustering Coefficient

Eigenvector Centrality

Edge Overview

Avg. Path Length

Dynamic

Nodes

Edges

Degree

Clustering Coefficient

What is Gephi?

A project ~10 years old

Free, libre, open source software

A project based on voluntary participation

A small team of designers and developers

Slow pace, but still maintained

A community

Many resources can be found online, but
there is also an active **Facebook group**:
<https://www.facebook.com/groups/gephi/>

A tool downloaded > 2 million times

Import and export networks

Apply **layout algorithms** to visualize your network, or use the **list view**

Apply **semiotic** treatment: set the size and color of nodes and edges

Compute statistical **metrics** and other algorithms like community detection

Filter the nodes and edges

Interact with your network

Good to know

1. Gephi requires installing **JAVA**.

JAVA is a software layer. Gephi works on top of it.

Installing it is innocuous if you mind to **uncheck the “parasites”** suggesting during the install process. (default search engine...)

2. Gephi works on Windows, Mac OS & Linux (thanks to Java)

3. It is possible to **report bugs** on GitHub

<https://github.com/gephi/gephi/issues>

Useful resources

Official website

<https://gephi.org/>

Gephi Cheat Sheets (C. Levallois)

http://www.clementlevallois.net/gephi/tuto/en/gephi_cheat%20sheets_en.pdf

Many available videos

https://www.youtube.com/results?search_query=gephi

Official tutorials

<https://gephi.org/users/>

Clément Levallois' tutorials

<http://www.clementlevallois.net/gephi.html>

Martin Grandjean's tutorials

<http://www.martingrandjean.ch/gephiintroduction/>

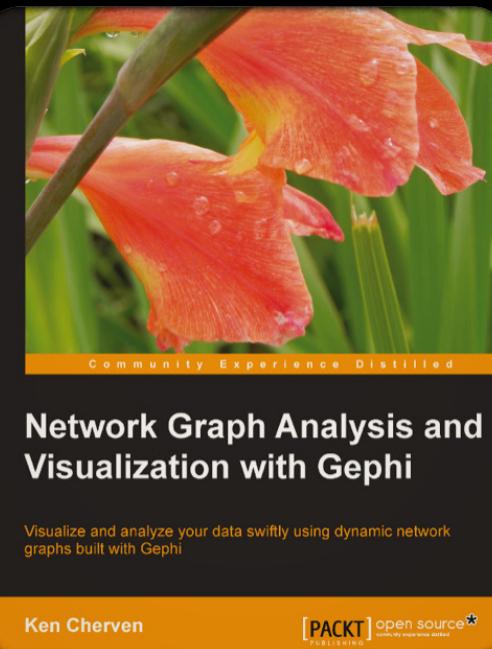
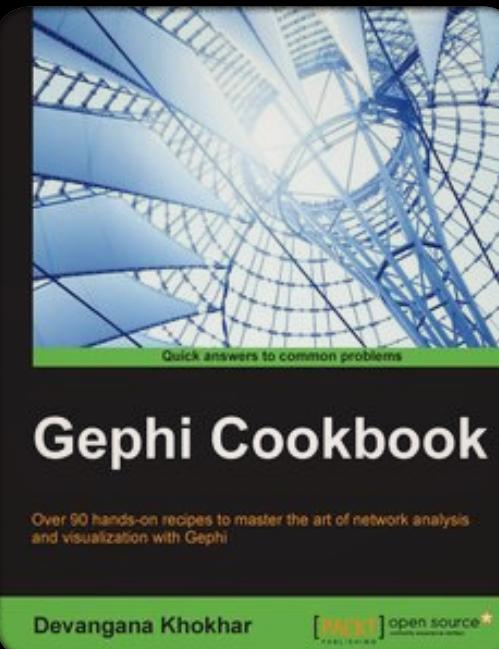
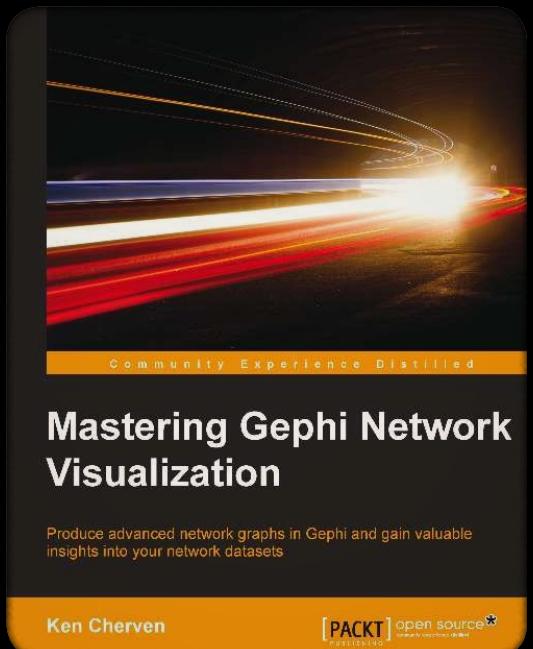
The “Gephi paper”

Bastian, M., Heymann, S., & Jacomy, M. (2009, March). *Gephi: an open source software for exploring and manipulating networks*. In *Third international AAAI conference on weblogs and social media*.

The “Force Atlas 2 paper”

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*. *PloS one*, 9(6), e98679.

There are also books on Gephi...



Thank you for your attention

*@jacomy
reticular.hypotheses.org
Mathieu.Jacomy@gmail.com*

