Unexplained because it works

The unfound reasons for the practical effectiveness of force-directed network layouts and modularity clustering

Mathieu Jacomy
Aalborg University MASSHINE & Techno-Anthropology Lab
@jacomyma@mas.to

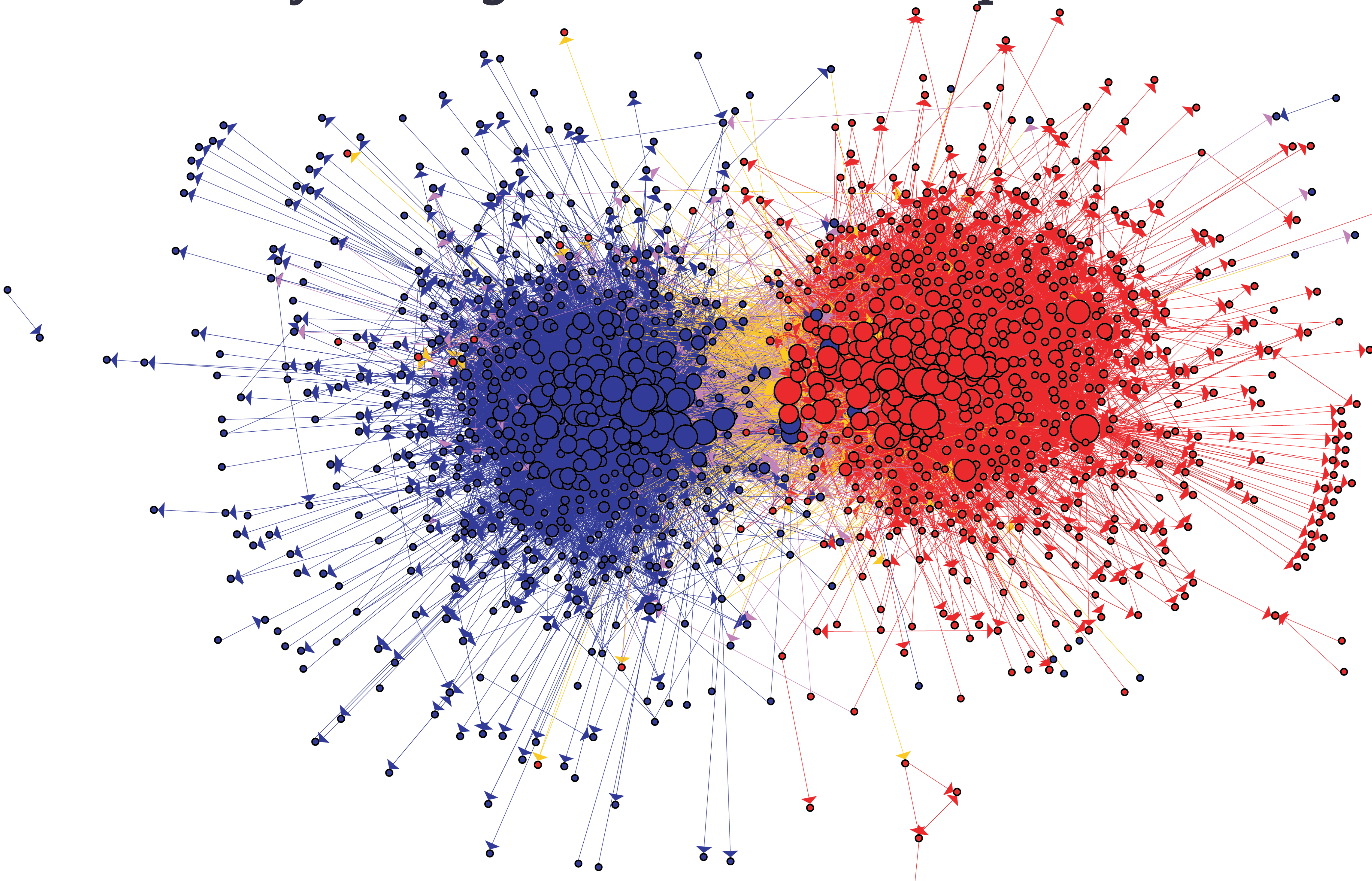
Case #1: Network layout algorithms

This case in short:

1. Those algos do work in practice
2. But we don't know why
3. Yet algo designers say we do
4. Yet we actually do not
5. But we don't care to not understand, because it works well enough.

Q: Should we care about understanding?
Why or why not?

1. Network layout algorithms work in practice



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

1. Network layout algorithms work in practice

- They are inspiring
- They are popular
- You don't have a choice
  (i.e. there are few other options)

That network visualization became so famous it ended up ALSO the poster boy of the programmatic paper of computational social science!

SOCIAL SCIENCE

Computational Social Science

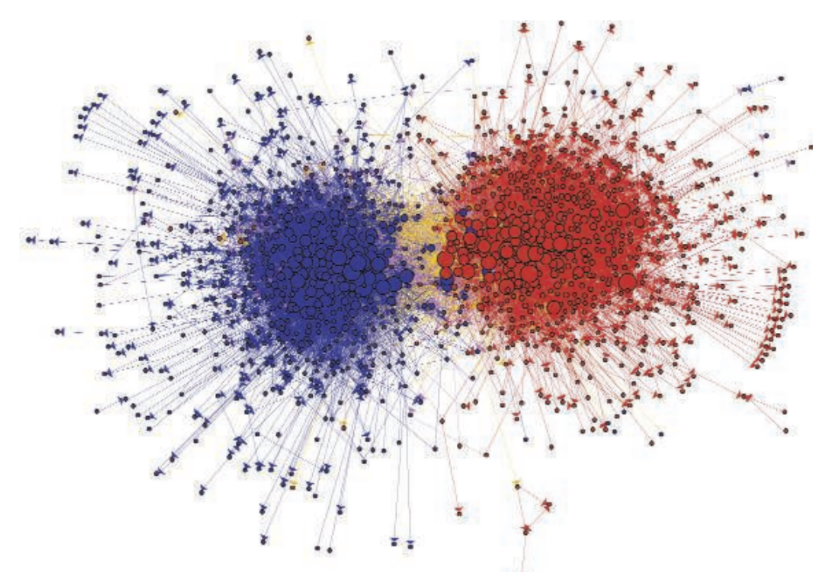
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



Data from the blogosphere. Shown is a 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. [Reproduced from (8) with permission from the Association for Computing Machinery]

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., … & Van Alstyne, M. (2009). Computational social science. Science, 323(5915), 721-723.

1. Network layout algorithms work in practice

- They are inspiring
- They are popular
- You don't have a choice
  (i.e. there are few

That network visualization became so famous it ended up ALSO the poster boy of the programmatic paper of computational social science!


SOCIAL SCIENCE

Computational Social Scien

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Ba
Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] I
Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in govern-

ment agencies such as the
ity Agency. Computationa
become the exclusive do
panies and government ag
there might emerge a pr
demic researchers presidi
from which they produce p

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

Data from the blogosphere. Shown is a 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. [Reproduced from (*8*) with permission from the Association for Computing Machinery]

n February 06, 2024

www.sciencemag.org SCIENCE VOL 323 6 FEBRUARY 2009 **721**
Published by AAAS

SO, IN OTHER WORDS:

My observation is that they work well-enough for the people who use them, because they keep using them.

You may think they still actually don't work for them. I guess you know better.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., … & Van Alstyne, M. (2009). Computational social science. Science, 323(5915), 721-723.

2. But we don't know why they work

The most common criteria
for a good graph drawing
are obsolete. For most people.

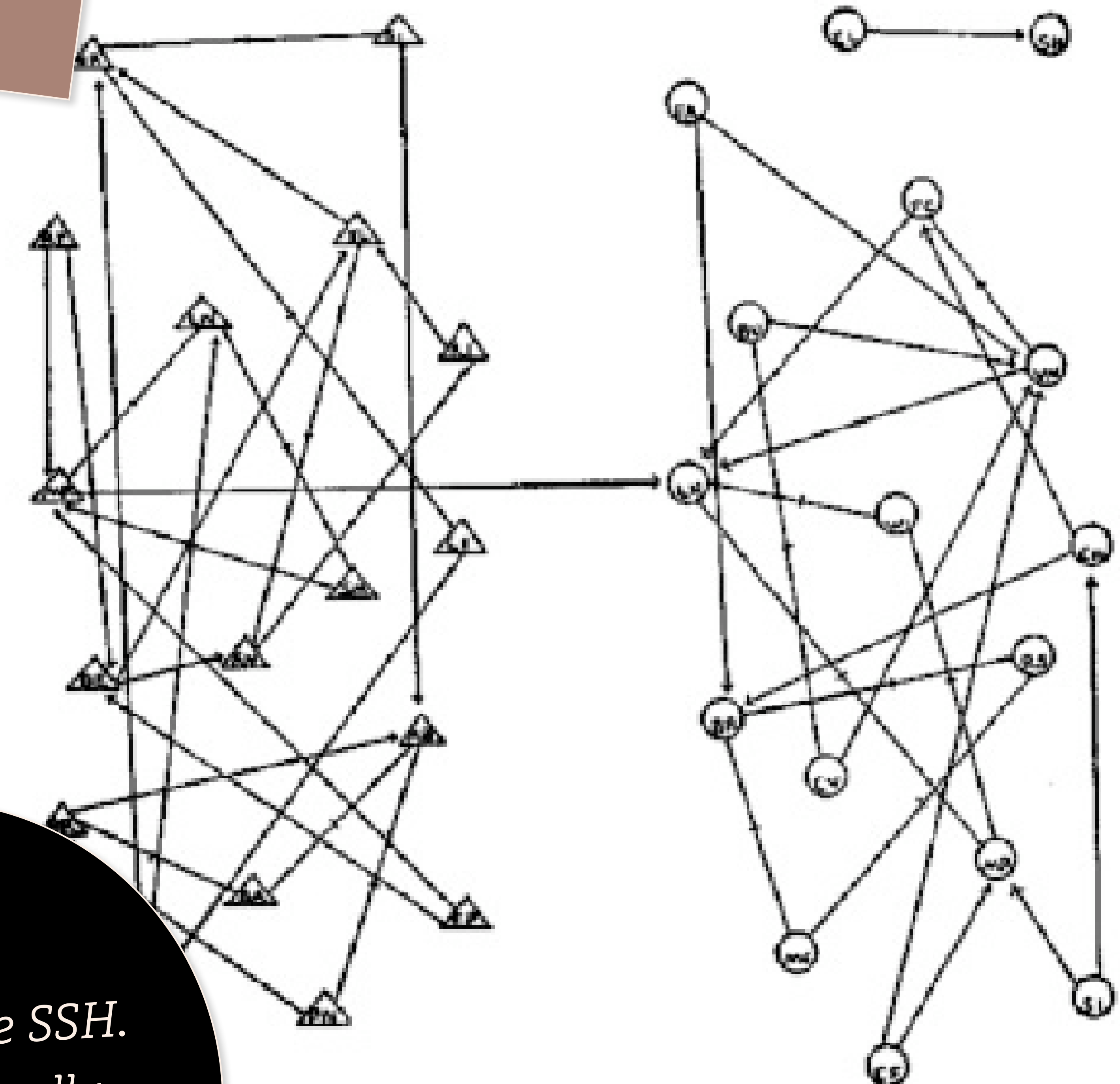
They have been for 15 years.
Arguably 30.

Introducing:
A history of graph drawing in 7 min.
I tried 5 but I can't.

1934

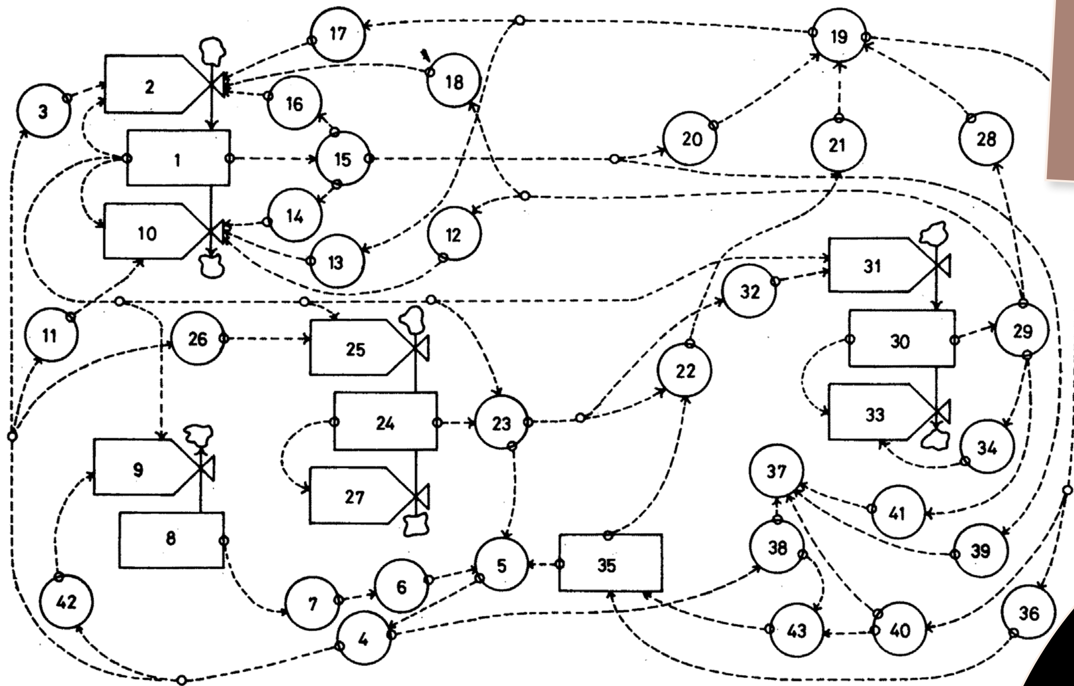

EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the Psychological Currents of Human Relationships.



It starts in the SSH.
Drawn manually.
Empirical data.

Moreno, J.L. (1933) 'Emotions mapped by new geography', New York Times, 3, p. 17.
Moreno, J.L. (1934) Who shall survive?: A new approach to the problem of human interrelations.



(a)

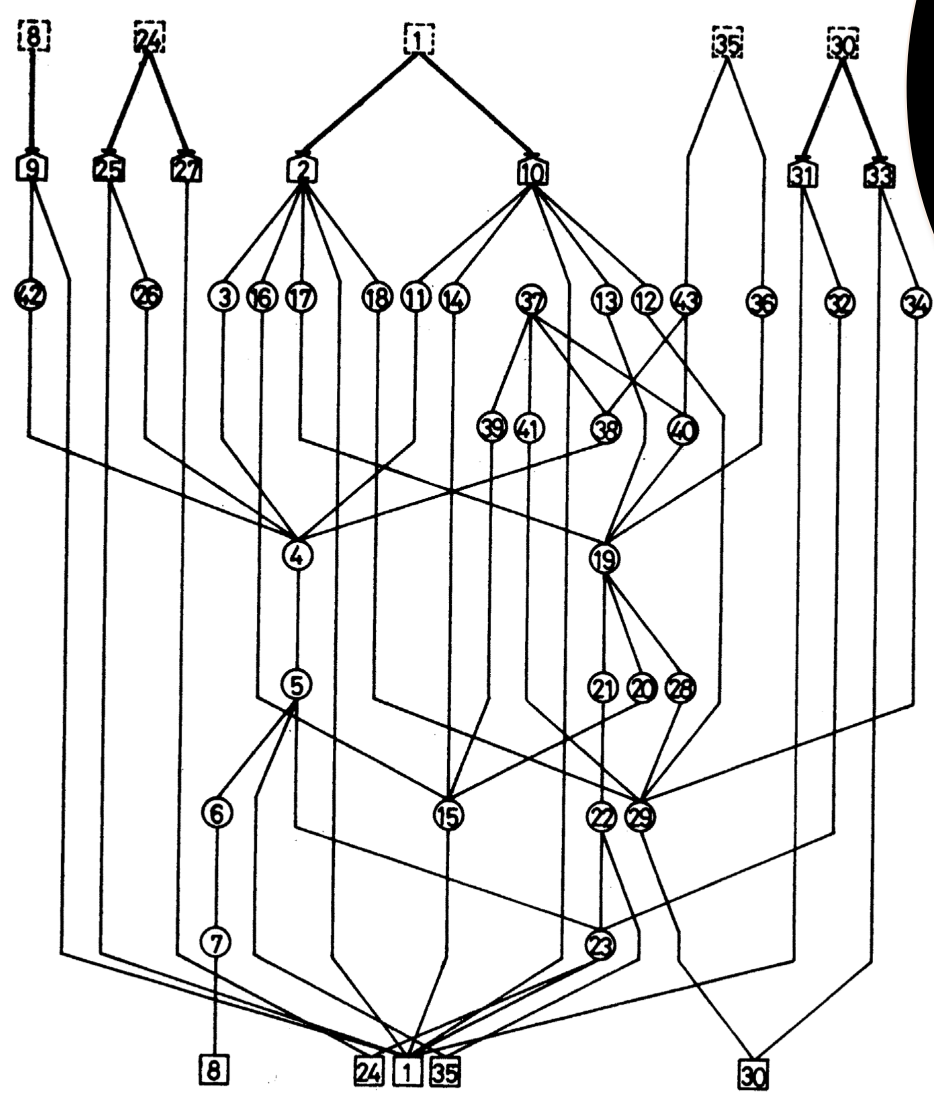
1981

Criteria extracted from practices. Diagrams: many semiotic features.

AREA	minimize the area occupied by the drawing.
BALAN	balance the diagram with respect to the vertical or horizontal axis.
BENDS	minimize the number of bends along the edges.
CONVEX	maximize the number of faces drawn as convex polygons.
CROSS	minimize the number of crossings between edges.
DEGREE	place nodes with high degree in the center of the drawing.
DIM	minimize differences among nodes' dimensions.
LENGTH	minimize the global length of edges.
MAXCON	minimize length of the longest edge.
SYMM	symmetry of sons in hierarchies.
UNIDEN	uniform density of nodes in the drawing.
VERT	verticality of hierarchical structures.

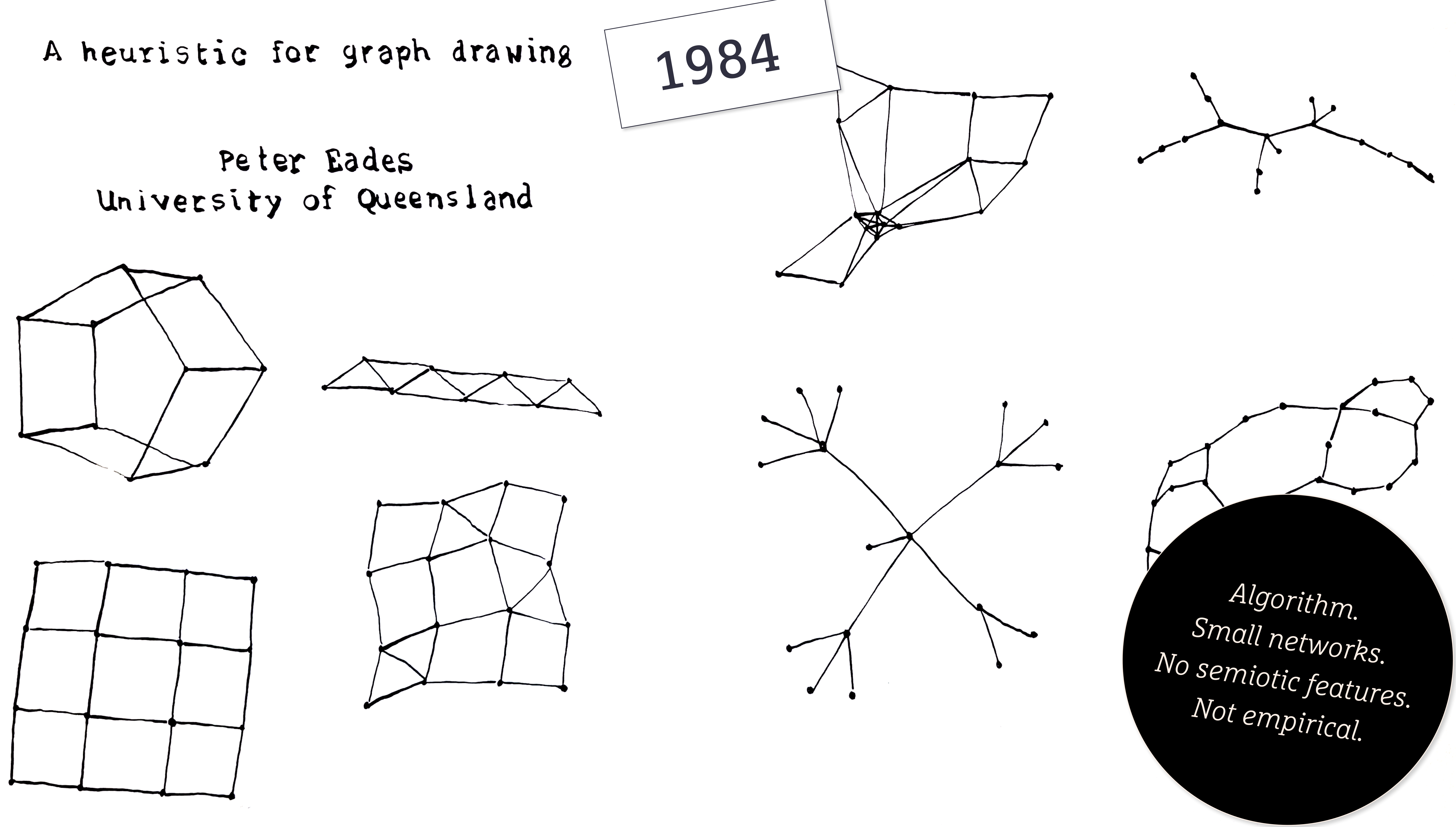
Sindre, G., Gulla, B. and Jokstad, H.G. (1993) 'Onion graphs: Aesthetics and layout', in Proceedings – 1993 IEEE Symposium on Visual Languages, VL 1993, Bergen, Norway, 24–27 August 1993. Institute of Electrical and Electronics Engineers Inc., pp. 287–291. doi:10.1109/ VL.1993.269613.

1993



(b)

Sugiyama, K., Tagawa, S. and Toda, M. (1981) 'Methods for visual understanding of hierarchical system structures', IEEE Transactions on Systems, Man and Cybernetics, 11(2), pp. 109–125. doi:10.1109/TSMC.1981.4308636.

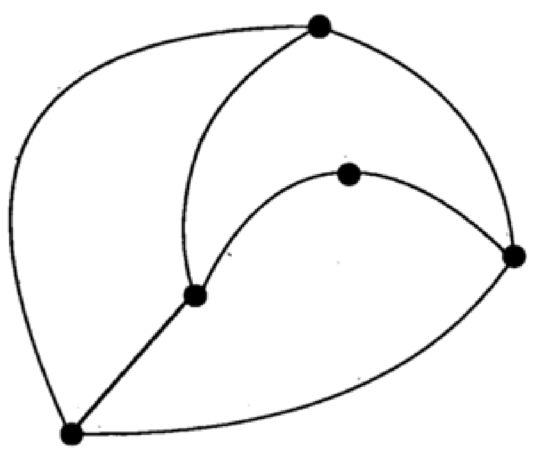


Eades, P. (1984) 'A heuristic for graph drawing', Congressus Numerantium, 42, pp. 149–160. Available at: https://ci.nii.ac.jp/naid/10000023432 (Accessed: 13 October 2020).

Automatic Graph Drawing and Readability of Diagrams

ROBERTO TAMASSIA, GIUSEPPE DI BATTISTA, AND CARLO BATINI

1988



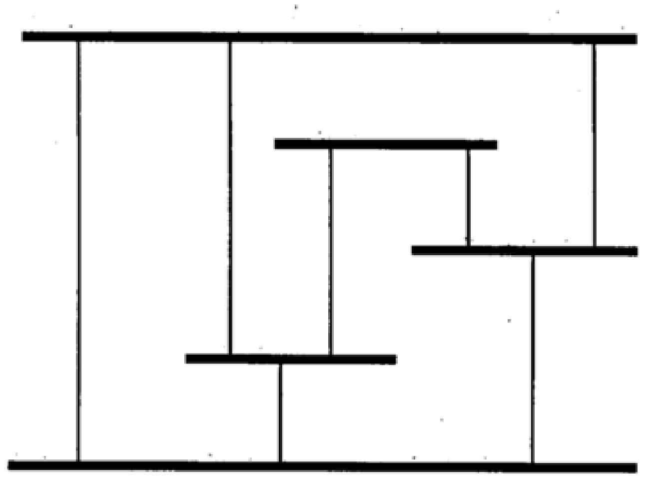
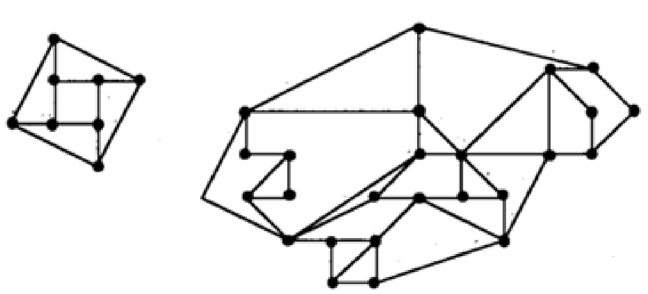


Fig. 8. Visibility representation.



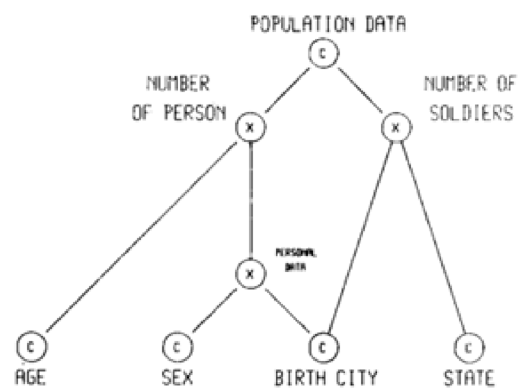
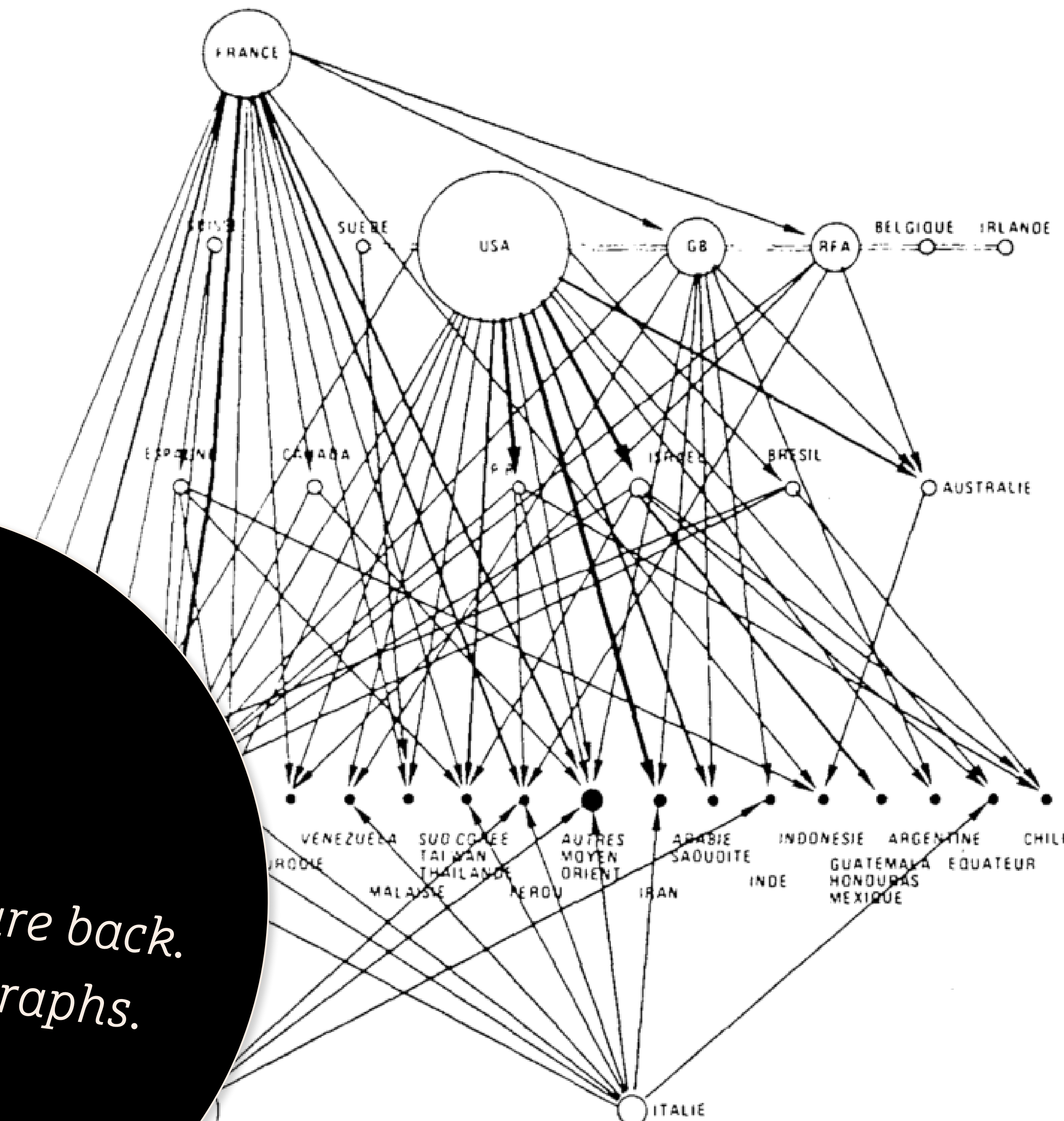


Fig. 10. Hierarchic graph used for statistical databas

Fig. 11. Proper *k*-layer graph.



12. Hierarchic graph drawn with Carpano algorithm (from [7]).

Algorithm.
Semiotic features are back.
Some mid-sized graphs.

Tamassia, R., Di Battista, G. and Batini, C. (1988) 'Automatic graph drawing and readability of diagrams', IEEE Transactions on Systems, Man and Cybernetics, 18(1), pp. 61–79. doi:10.1109/21.87055.

Graph Drawing by Force-directed Placement

1991

THOMAS M. J. FRUCHTERMAN* AND EDWARD M. REINGOLD
Department of Computer Science, University of Illinois at Urbana-Champaign, 1304 W. Springfield Avenue, Urbana, IL 61801-2987, U.S.A.

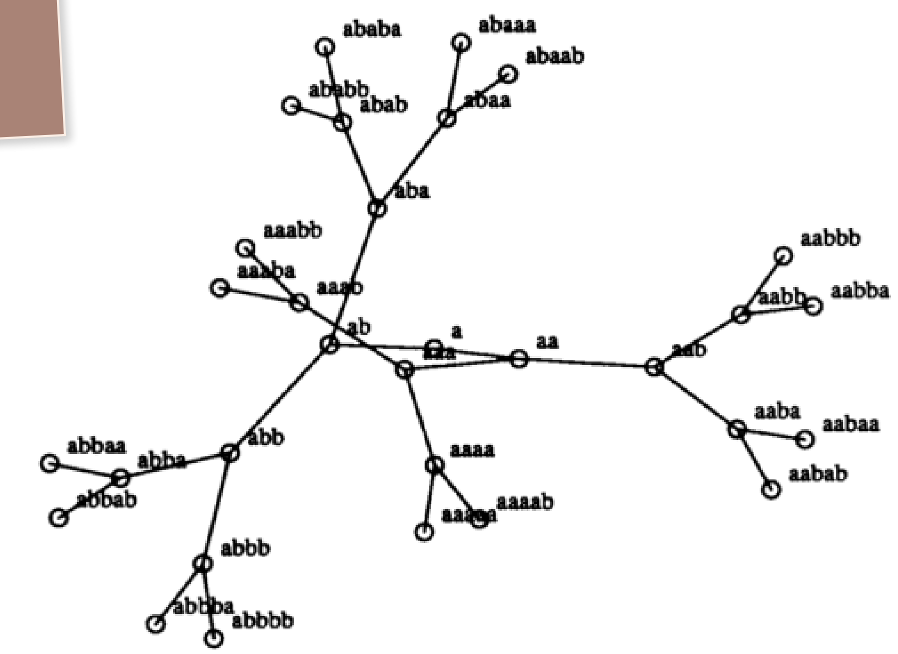


Figure 42. Example of a potential barrier

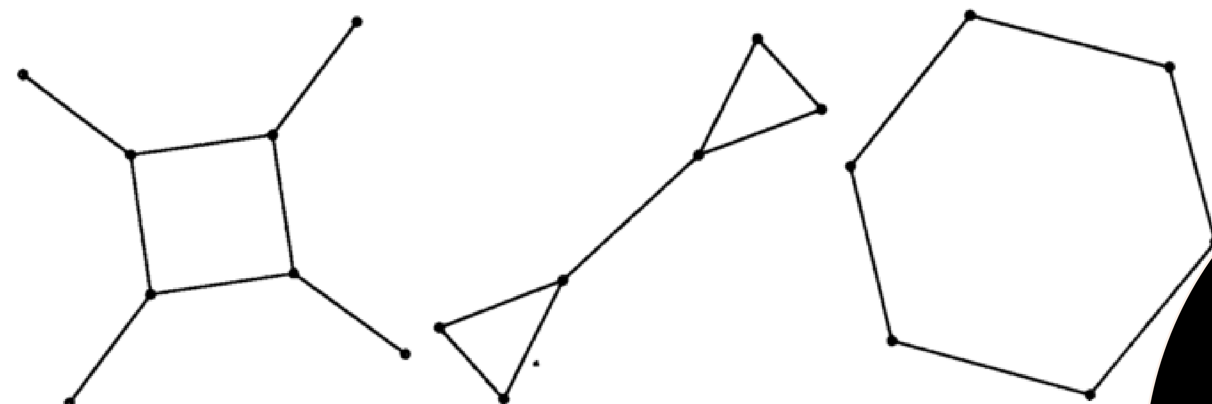


Figure 16. Graphs in Figures 6(a), 4, and 3, respectively, from Kamada and Kawai[8]

Algorithm.
Small networks.
No semiotic features.
Not empirical.

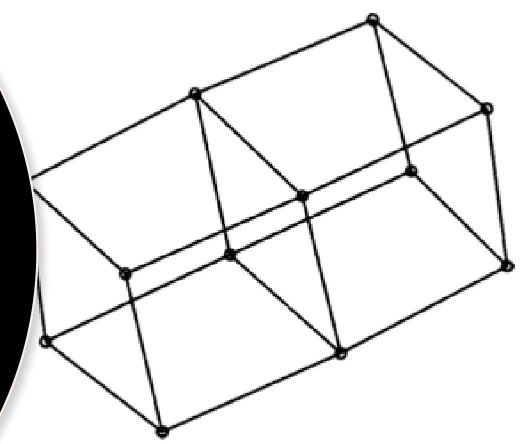


Figure 50. Twin cubes (graph in Figure 11 from Davidson and Harel[10]*)*

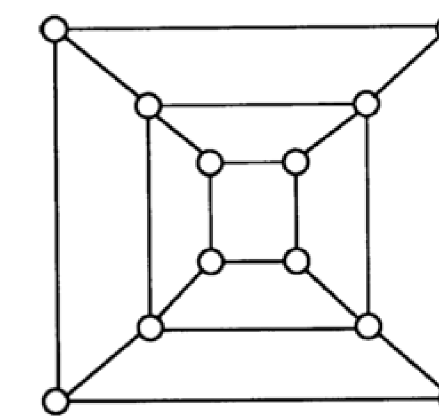


Figure 51. Figure 11(a) from Davidson and Harel as proposed by Davidson and Harel

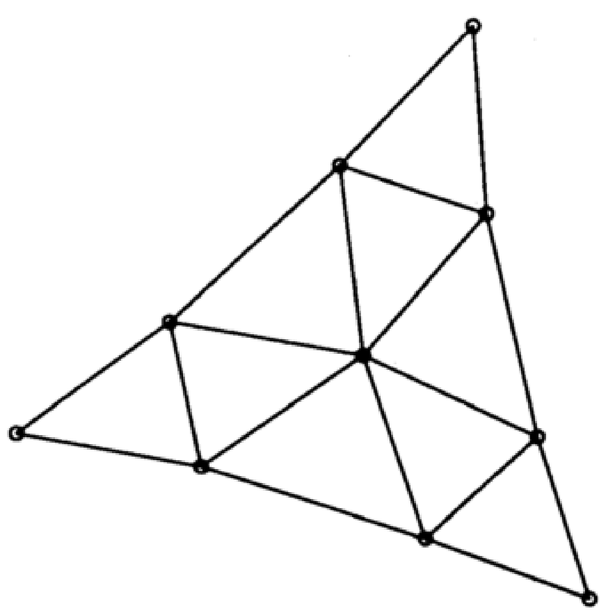


Figure 17. Triangulated triangle (graph in Figure 6(c) from Kamada and Kawai[8]*)*

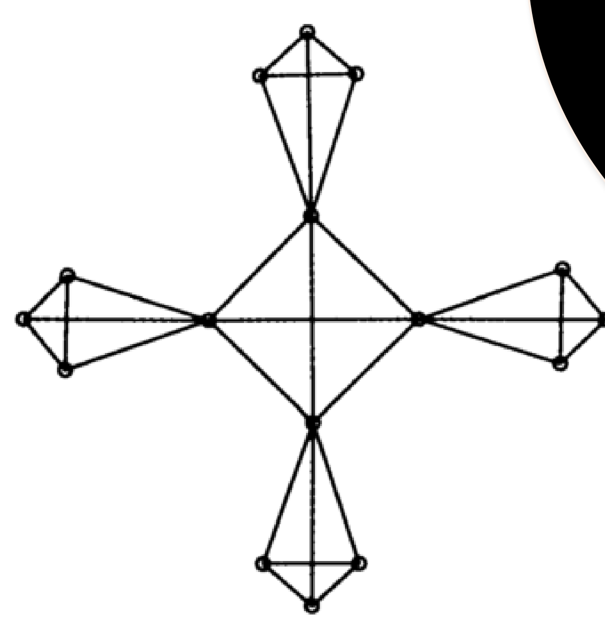


Figure 18. Graph in Figure 16 from Davidson and Harel[10]

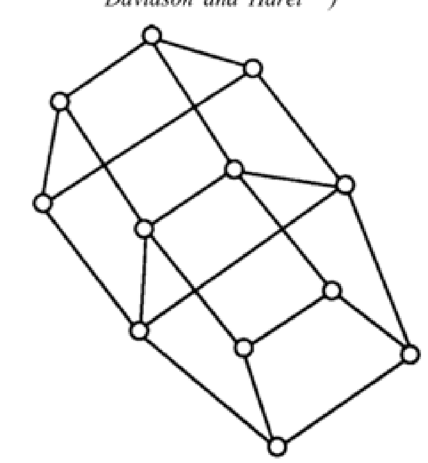


Figure 52. Figure 11(b) from Davidson and Harel[10] *as drawn by Davidson and Harel*

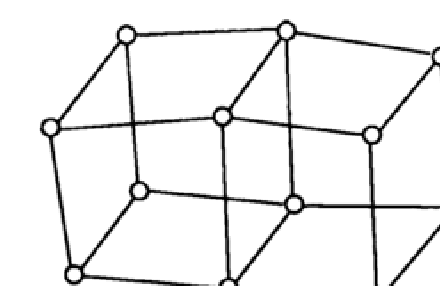
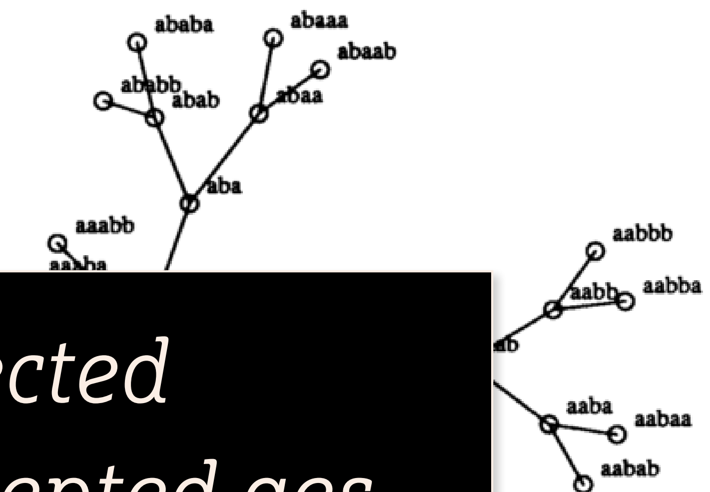


Figure 53. Figure 11(c) from Davidson and Harel[10] *as drawn by Davidson and Harel*

Früchterman, T.M.J. and Reingold, E.M. (1991) ‘Graph drawing by force-directed placement’, Software: Practice and Experience, 21(11), pp. 1129–1164. doi:10.1002/ spe.4380211102.

Graph Drawing by Force-directed Placement

THOMAS M. J. FRUCHTERMAN* AND EDWARD M. REINGOLD
Department of Computer Science,
Springfield Ave



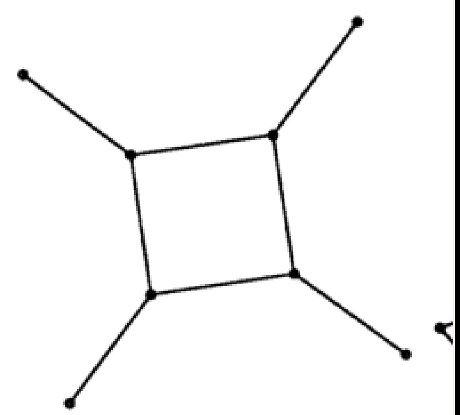


Figure 16. Graphs in Figure

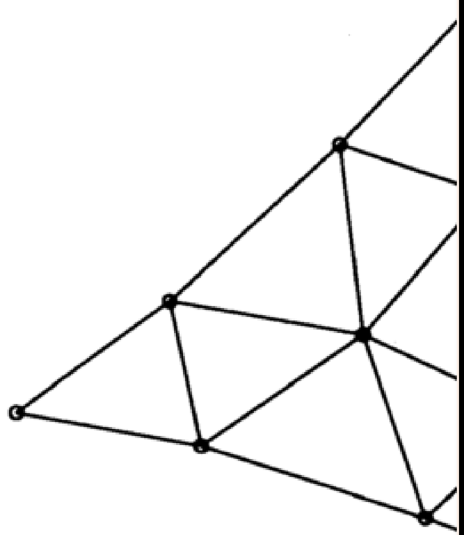
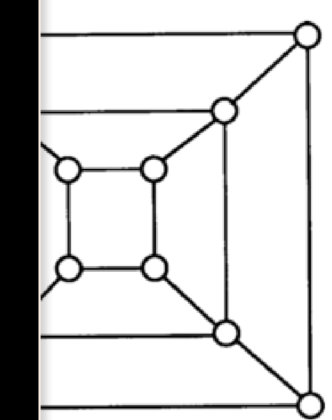


Figure 17. Triangulated triangle (gra
6(c) from Kamada and Kaw



ier

11(a) from Davidson and Harel
by Davidson and Harel

Harel " as drawn by Davidson and Harel

"We are concerned with drawing undirected graphs according to some generally accepted aesthetic criteria [Eades and Tamassia (1987)]. … Our algorithm does not explicitly strive for these goals, but does well at distributing vertices evenly, making edge lengths uniform, and reflecting symmetry. Our goals for the implementation are speed and simplicity. … We have only two principles for graph drawing: (1) Vertices connected by an edge should be drawn near each other. (2) Vertices should not be drawn too close to each other."

Aesthetic criteria can be dropped.

Früchterman, T.M.J. and Reingold, E.M. (1991) 'Graph drawing by force-directed placement', … d Experience, 21(11), pp. 1129–1164. doi:10.1002/ spe.4380211102.

1997

User benchmark. The tested network has only 16 nodes.

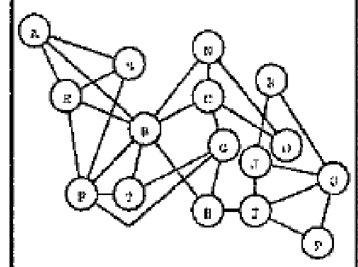
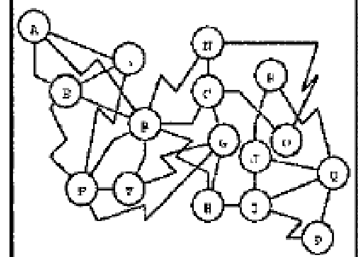
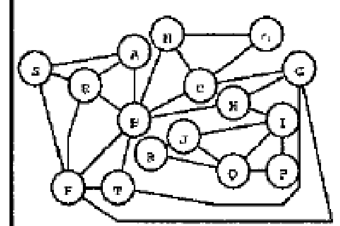
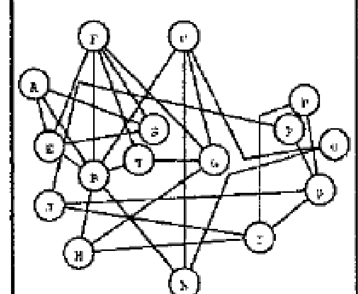
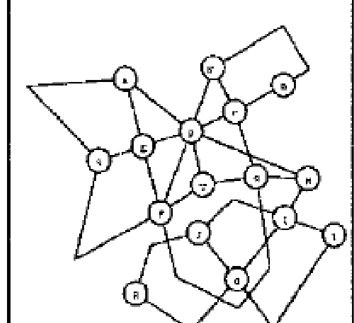
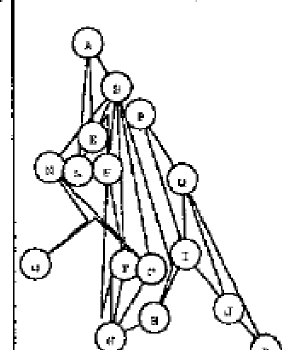
	graph	bend-less	cross-less	minangle	orthog	sym
b+		0.96	0.97	0.38	0.27	0.75
b-		0.47	0.99	0.44	0.28	0.71
c+		0.82	1	0.46	0.33	0.63
c-		0.87	0.88	0.35	0.29	0.84
m+		0.71	0.98	0.62	0.22	0.74
m-		0.82	0.98	0.16	0.26	0.79

Fig. 1. Six of the ten experimental graph drawings, and their aesthetic values.

Purchase, H. (1997) 'Which aesthetic has the greatest effect on human understanding?', in Di Battista, G. (ed.), Graph Drawing. GD 1997. Lecture Notes in Computer Science, vol. 1353, Berlin, Heidelberg: Springer, pp. 248–261. doi:10.1007/3-540-63938-1_67.

2001

GRIP: Graph dRawing with Intelligent Placement*

Pawel Gajer[1] and Stephen G. Kobourov[2]

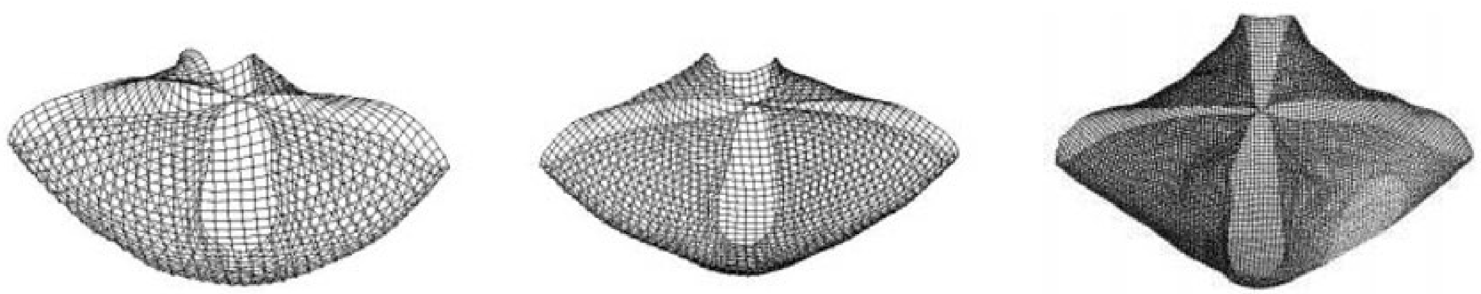


Fig. 5. Knotted rectangular (degree 4) meshes of 1600, 2500, and 10000 vertices.

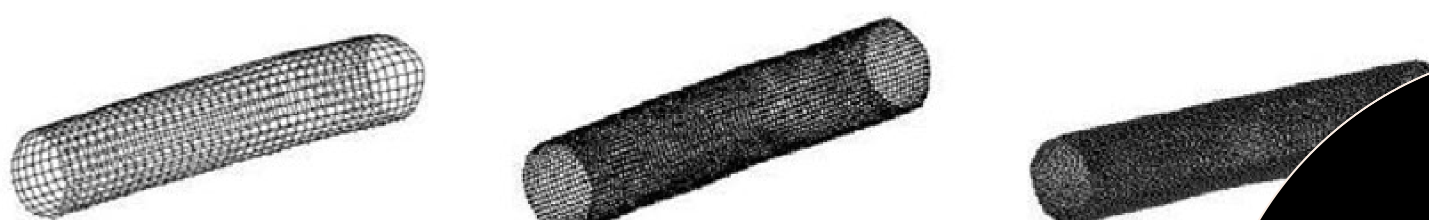


Fig. 6. Cylinders of 1000, 4000, and 10000 vertices.

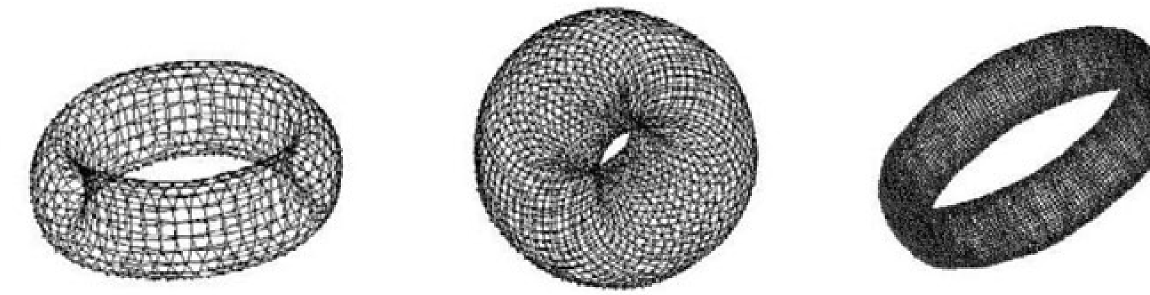


Fig. 7. Tori of various length and thickness: 1000, 2500, and 10000 drawn in four dimensions and projected down to three dimensions.

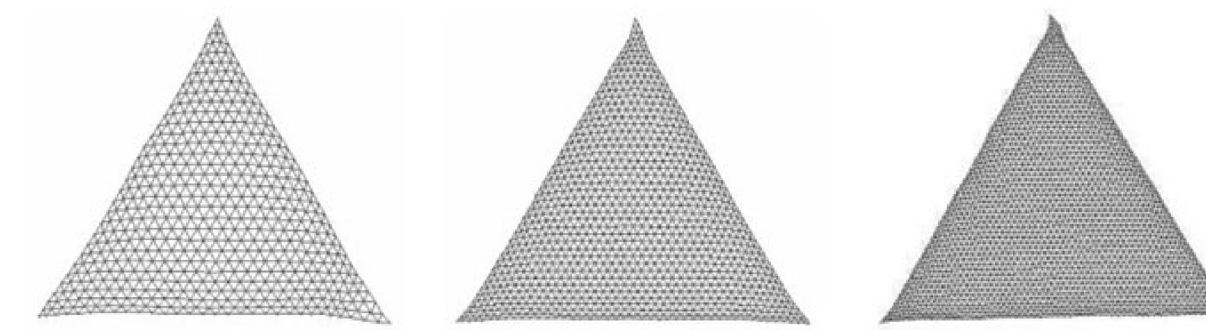


Fig. 8. Triangular (degree 6) meshes of 496, 1035, and 2016 vertices.

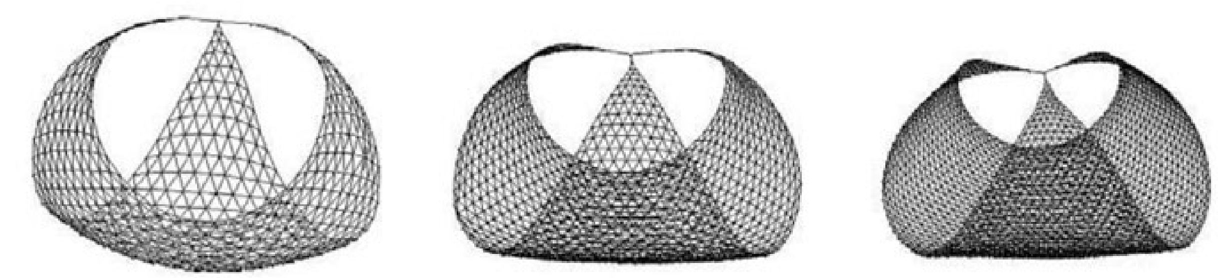
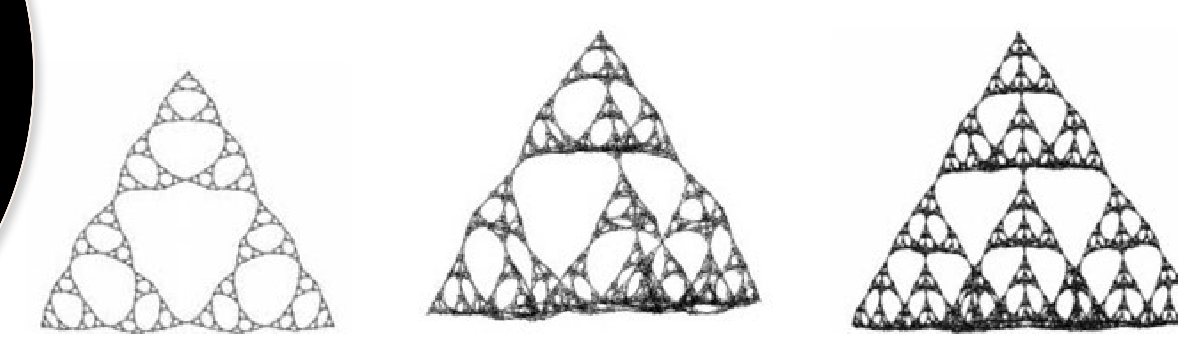


Fig. 9. Knotted triangular (degree 6) meshes of 496, 1035, and 2016 vertices.



Algorithm.
Large networks.
Not empirical.
Not scale-free.

Gajer, P. and Kobourov, S.G. (2001) 'GRIP: Graph drawing with intelligent placement', in Marks, J. (ed.), Graph Drawing. GD 2000. Lecture notes in computer science, vol. 1984. Berlin: Springer Verlag, pp. 222–228.

Large-Graph Layout with the Fast Multipole Multilevel Method

2005

STEFAN HACHUL and MICHAEL JÜNGER
Universität zu Köln, Institut für Informatik

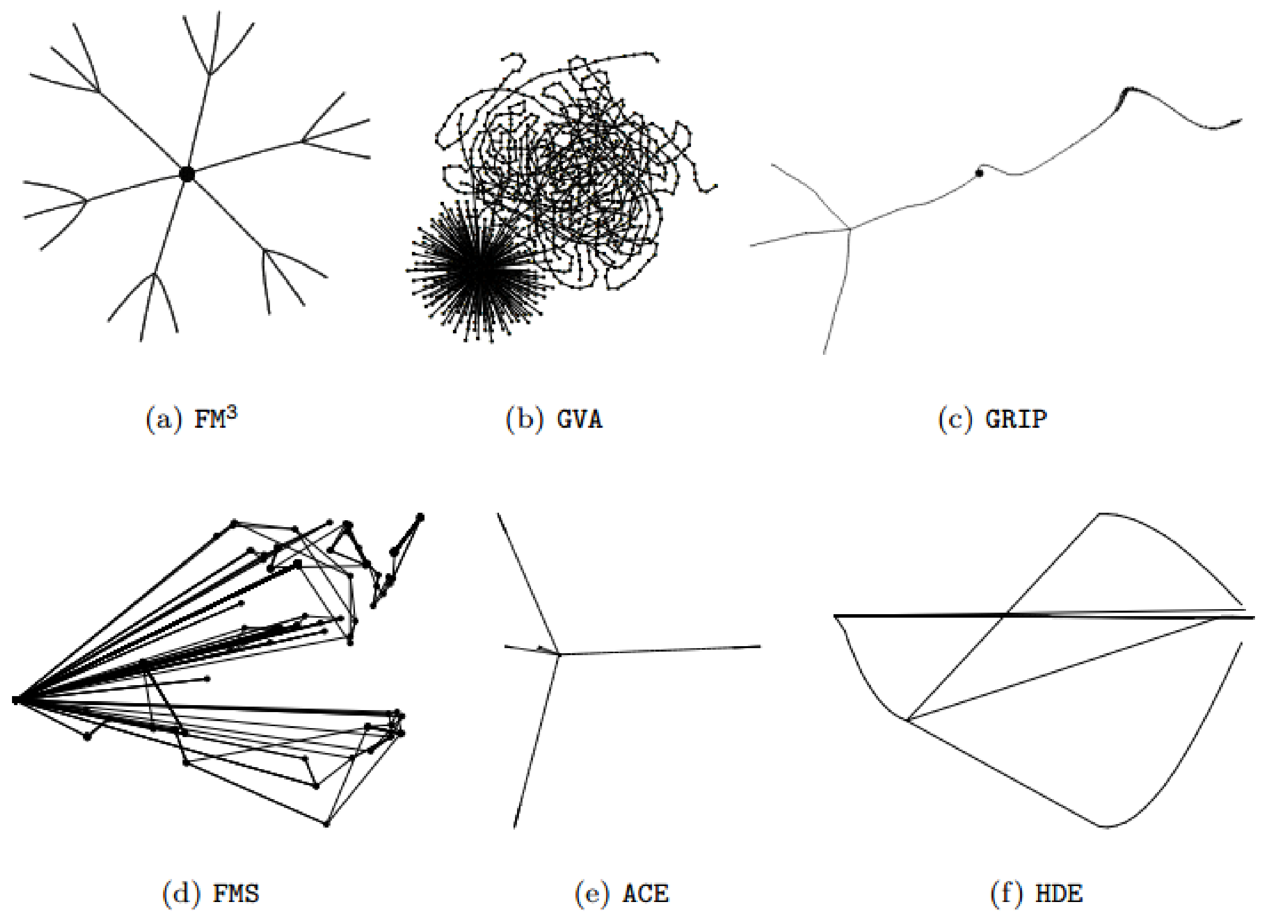
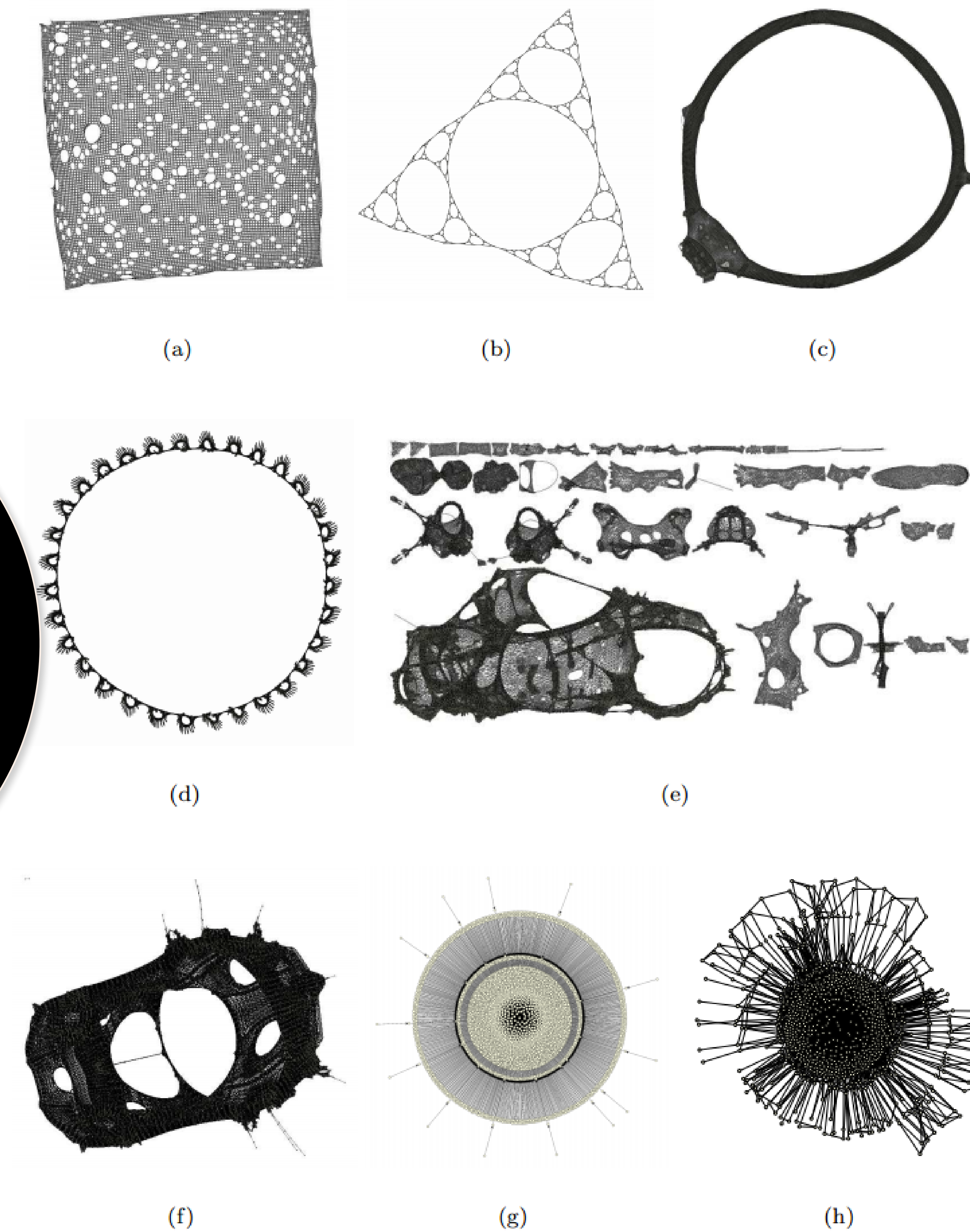


Fig. 12. (a)-(f) Drawings of snowflake_A generated by different algorithms.

Algorithm.
Large networks.
A few empirical.
A few scale-free.

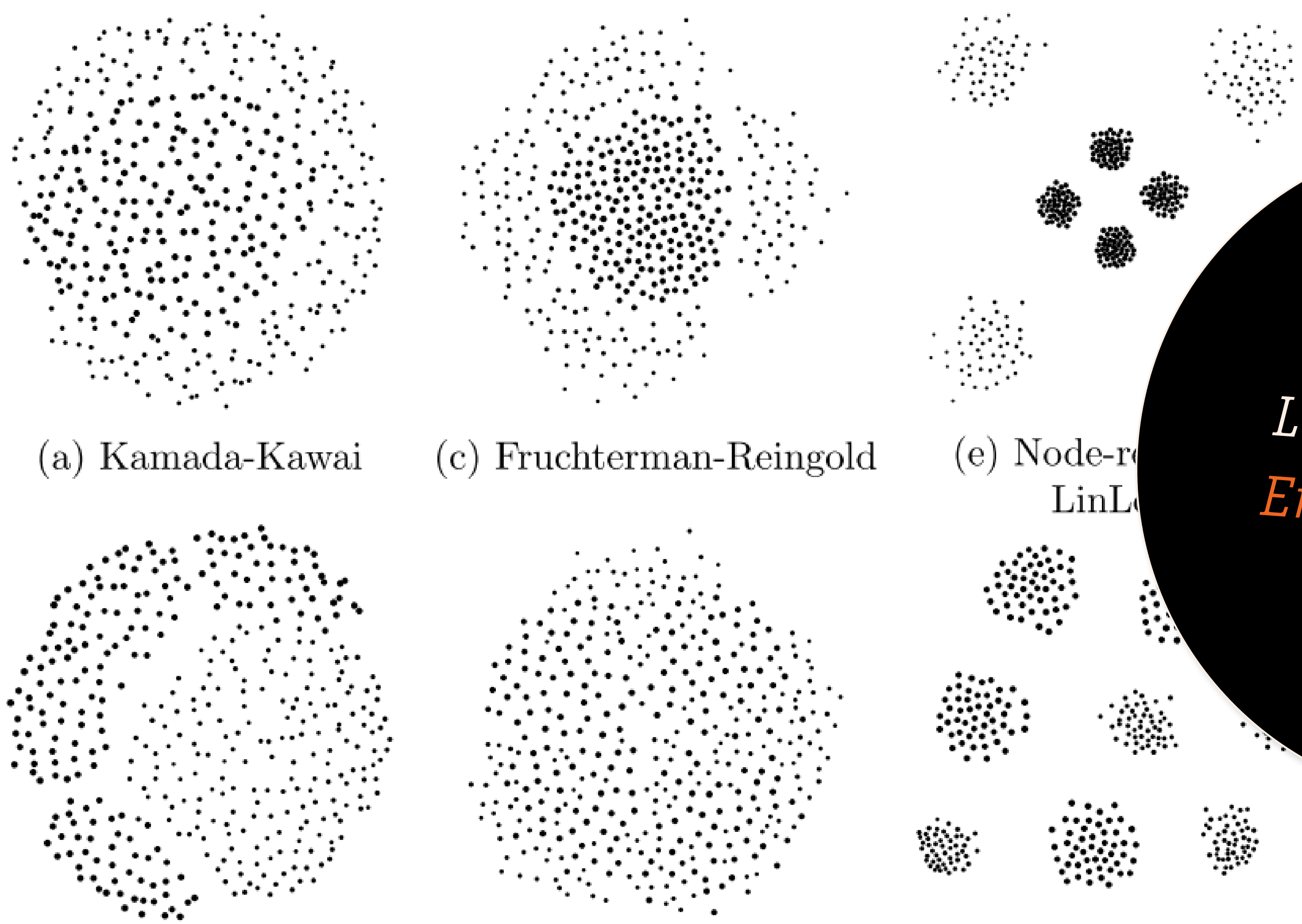


Hachul, S. and Jünger, M. (2005) 'Drawing large graphs with a potential-fieldbased multilevel algorithm', in Pach, J. (ed.), Graph Drawing. GD 2004. Lecture Notes in Computer Science, vol. 3383. Berlin, Heidelberg: Springer, pp. 285–295. doi:10.1007/978-3-540-31843-9_29

Energy Models for Graph Clustering

Andreas Noack
Institute of Computer Science
Brandenburg Technical University, Cottbus, Germany
an@informatik.tu-cottbus.de

2007



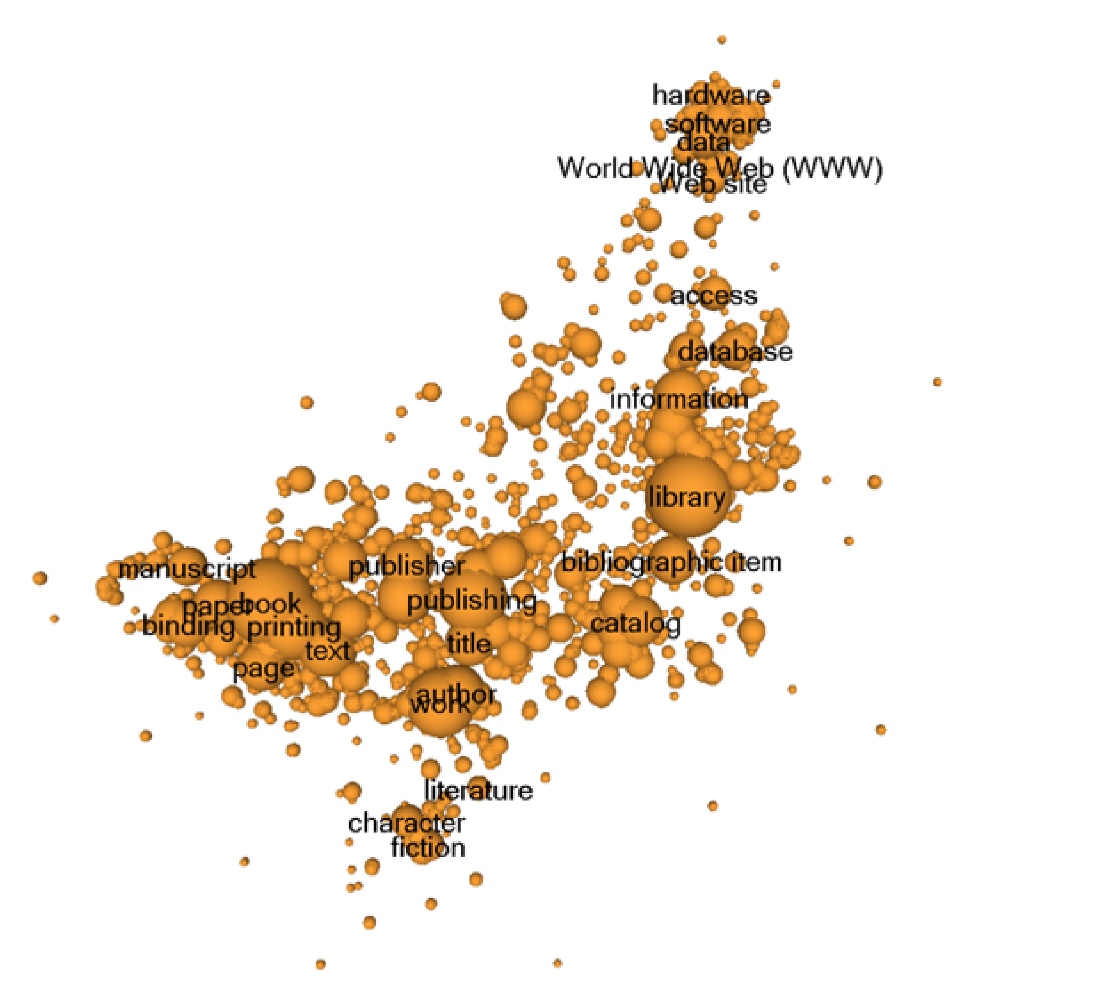
(a) Kamada-Kawai (c) Fruchterman-Reingold (e) Node-r… LinL…

(b) Kamada-Kawai, adapted by Gansner et al. (d) Edge-repulsion Fruchterman-Reingold (f) Edge-repulsion LinLog

Figure 2: Pseudo-random graph

Algorithm.
Large networks.
Empirical again!
Scale-free!

(a) Fruchterman-Reingold model (b) Node-repulsion LinLog model

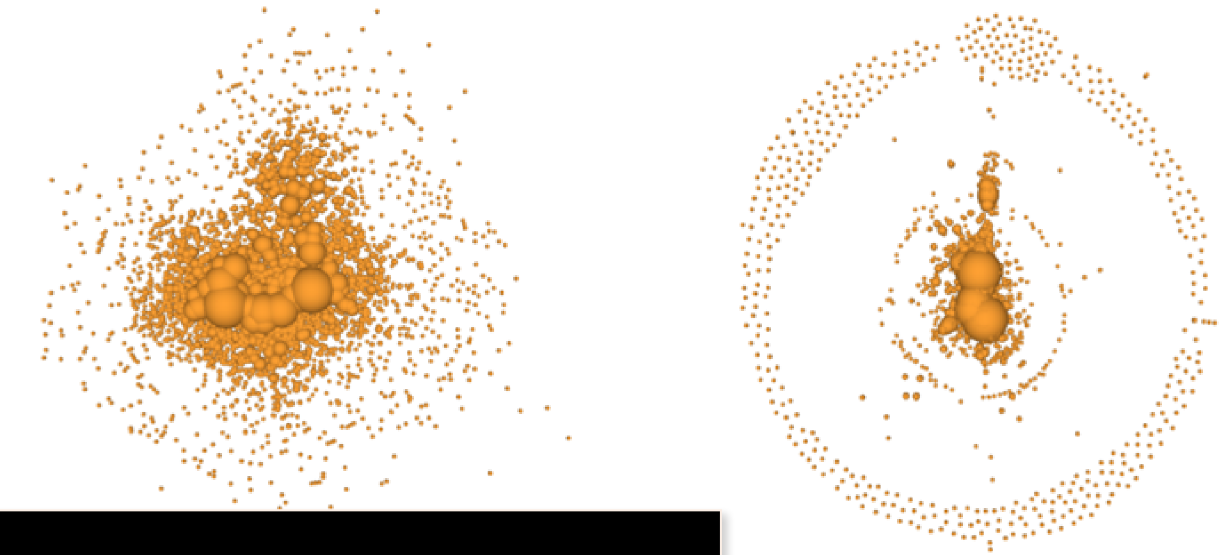


(c) Edge-repulsion LinLog model

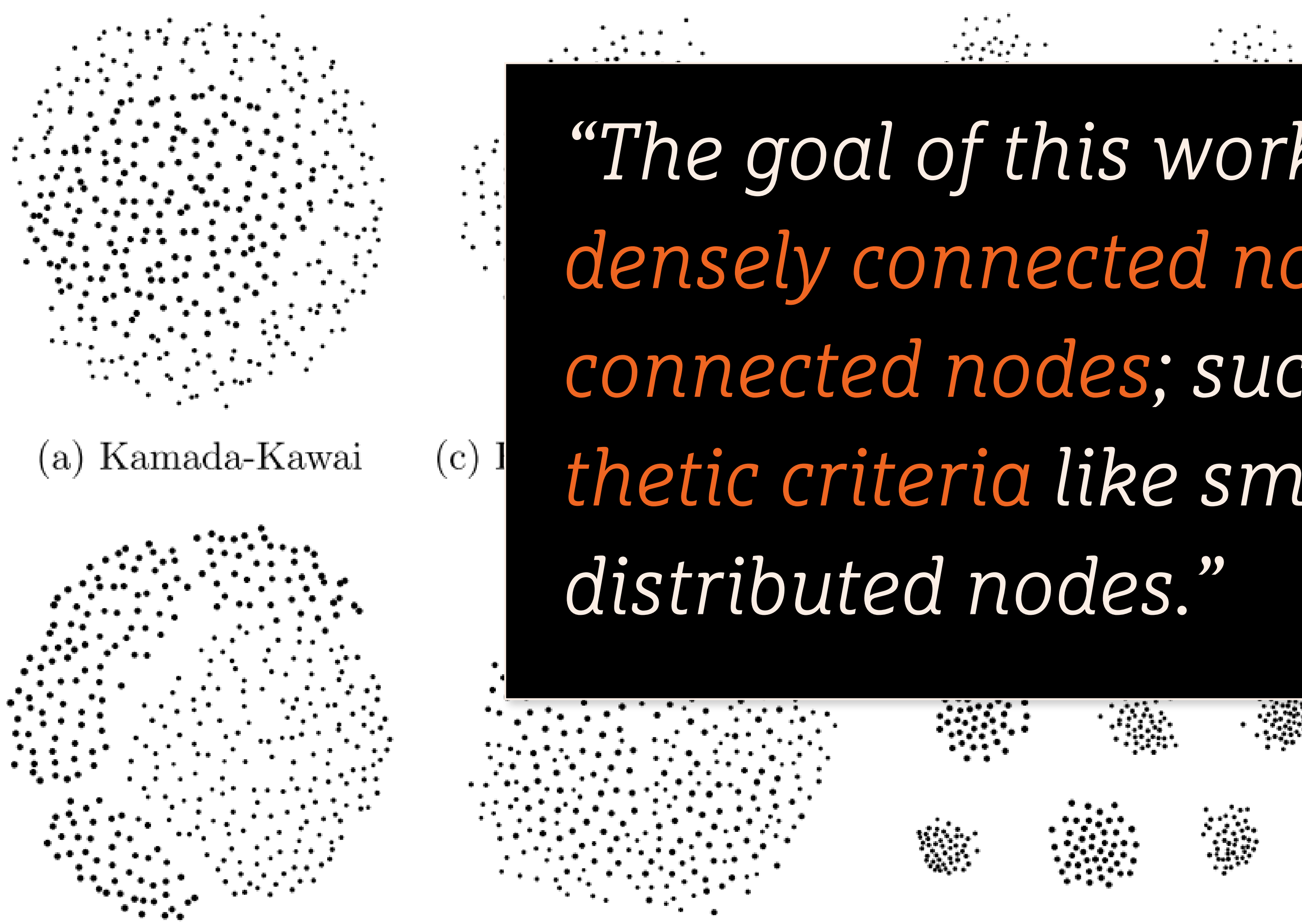
Noack, A. (2007) 'Energy models for graph clustering', Journal of Graph Algorithms and Applications JGAA, 11(112), pp. 453–480.

Energy Models for Graph Clustering

Andreas Noack
Institute of Computer Science
Brandenburg Technical University, Cottbus, Germany
an@informatik.tu-cottbus.de



(b) Node-repulsion LinLog model



(a) Kamada-Kawai

(b) Kamada-Kawai, adapted by Gansner et al.

(d) Edge-repulsion Fruchterman-Reingold

(f) Edge-repulsion LinLog

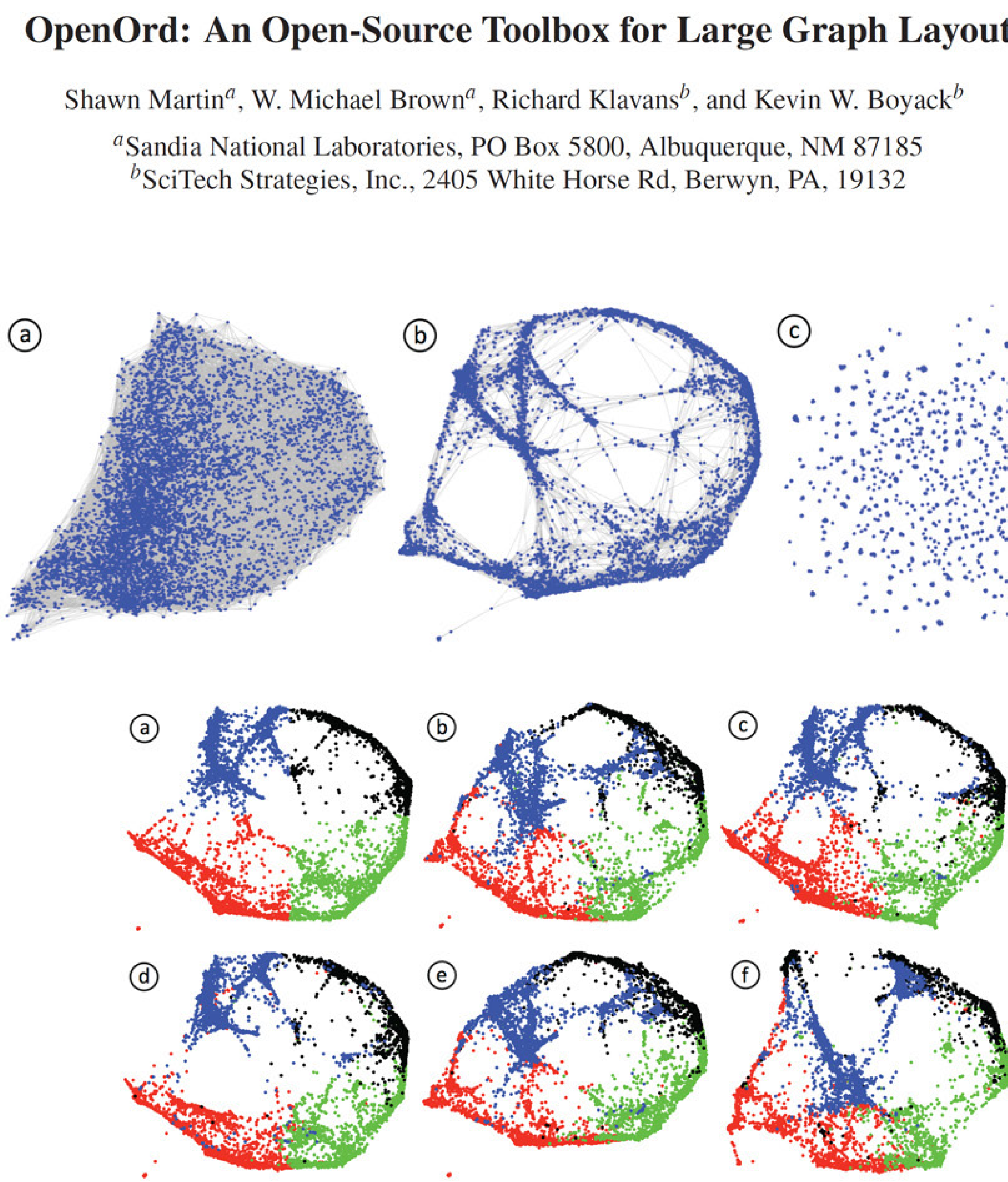
Figure 2: Pseudo-random graph

"The goal of this work are layouts that group densely connected nodes and separate sparsely connected nodes; such layouts often violate aesthetic criteria like small edge lengths or uniformly distributed nodes."

Aesthetic criteria MUST be dropped.

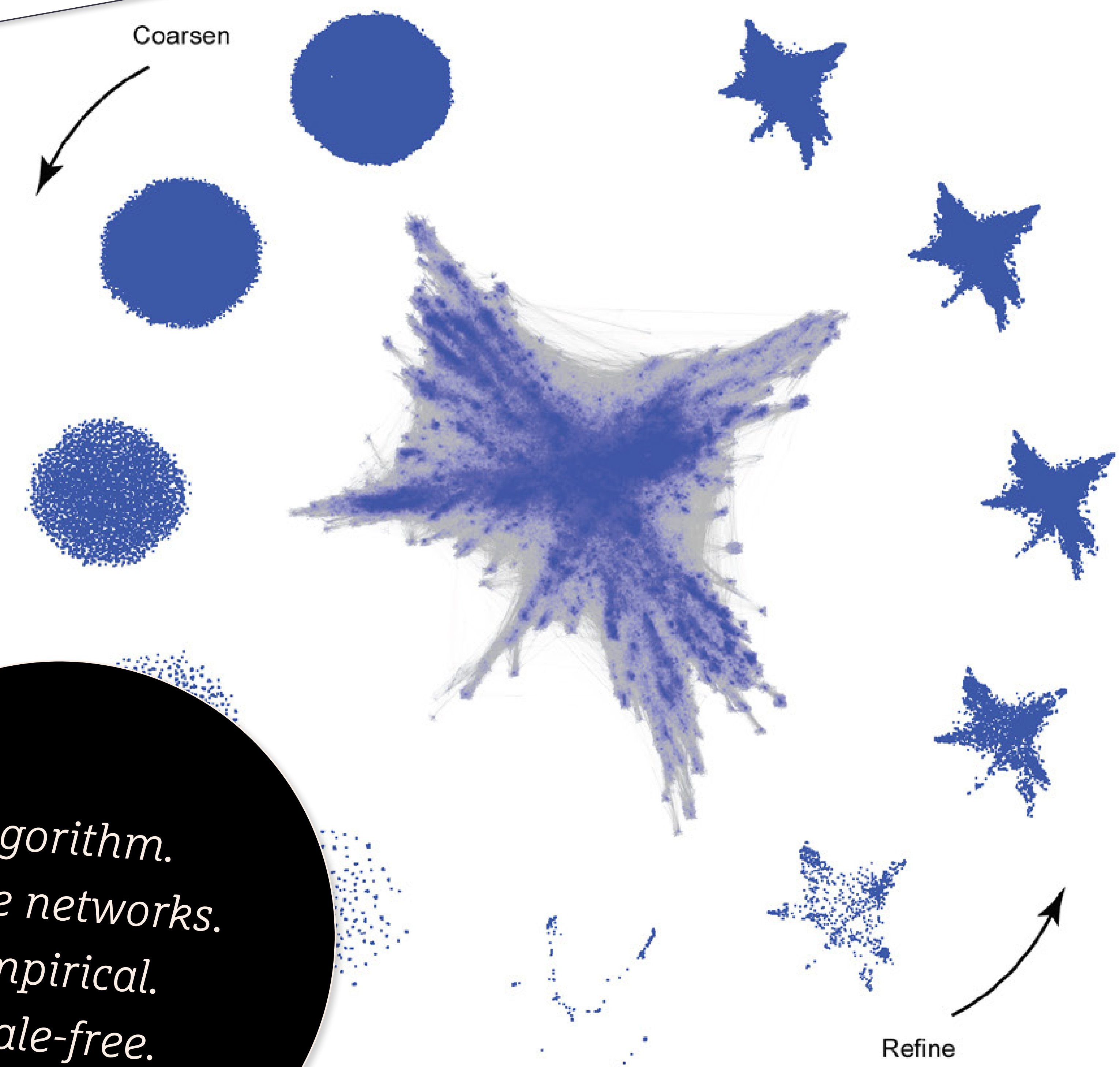
Noack, A. (2007) 'Energy models for graph clustering', Journal of Graph Algorithms and Applications JGAA, 11(112), pp. 453–480.

2011


OpenOrd: An Open-Source Toolbox for Large Graph Layout

Shawn Martin[a], W. Michael Brown[a], Richard Klavans[b], and Kevin W. Boyack[b]

[a]Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185
[b]SciTech Strategies, Inc., 2405 White Horse Rd, Berwyn, PA, 19132



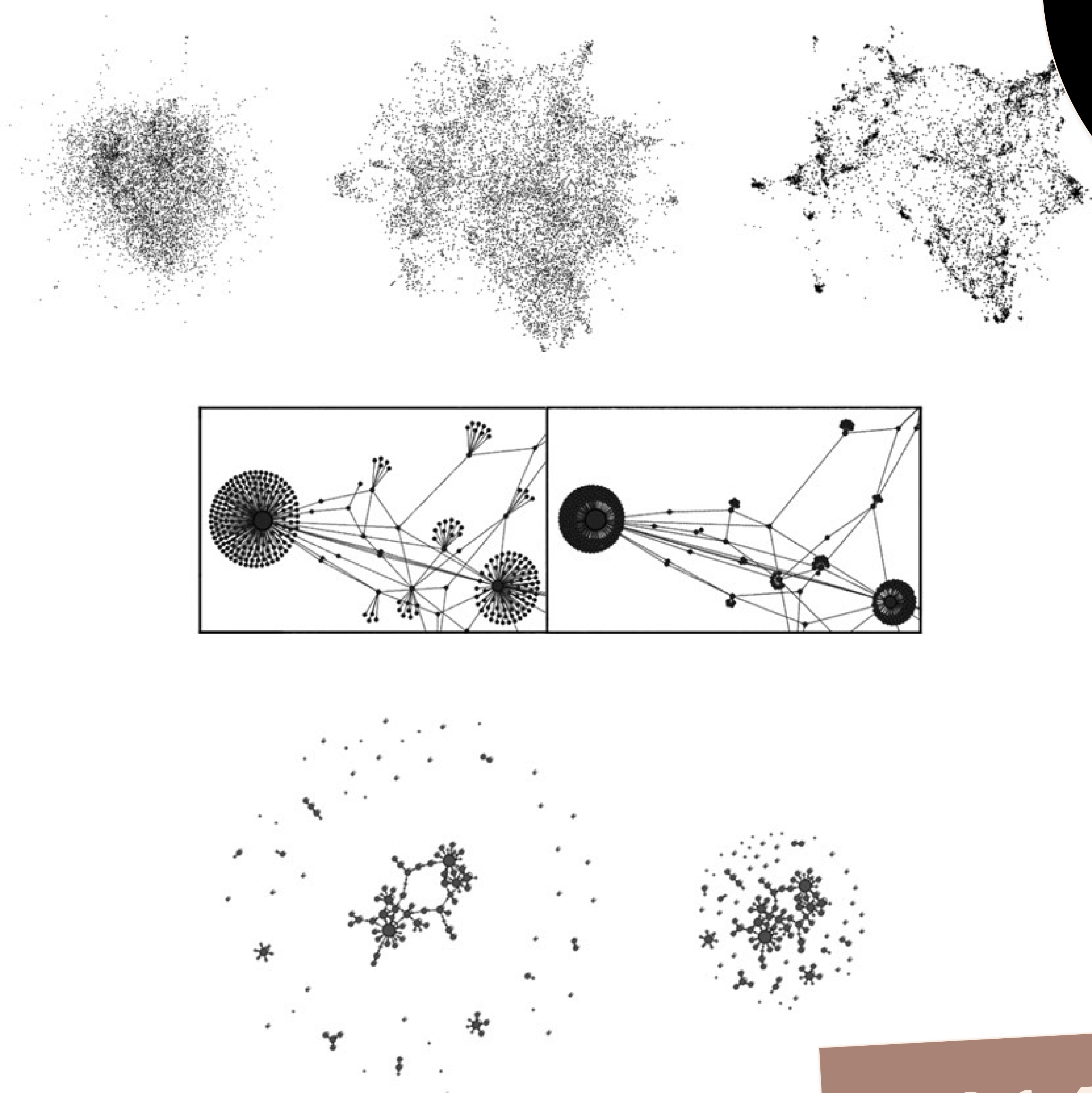
Algorithm.
Large networks.
Empirical.
Scale-free.

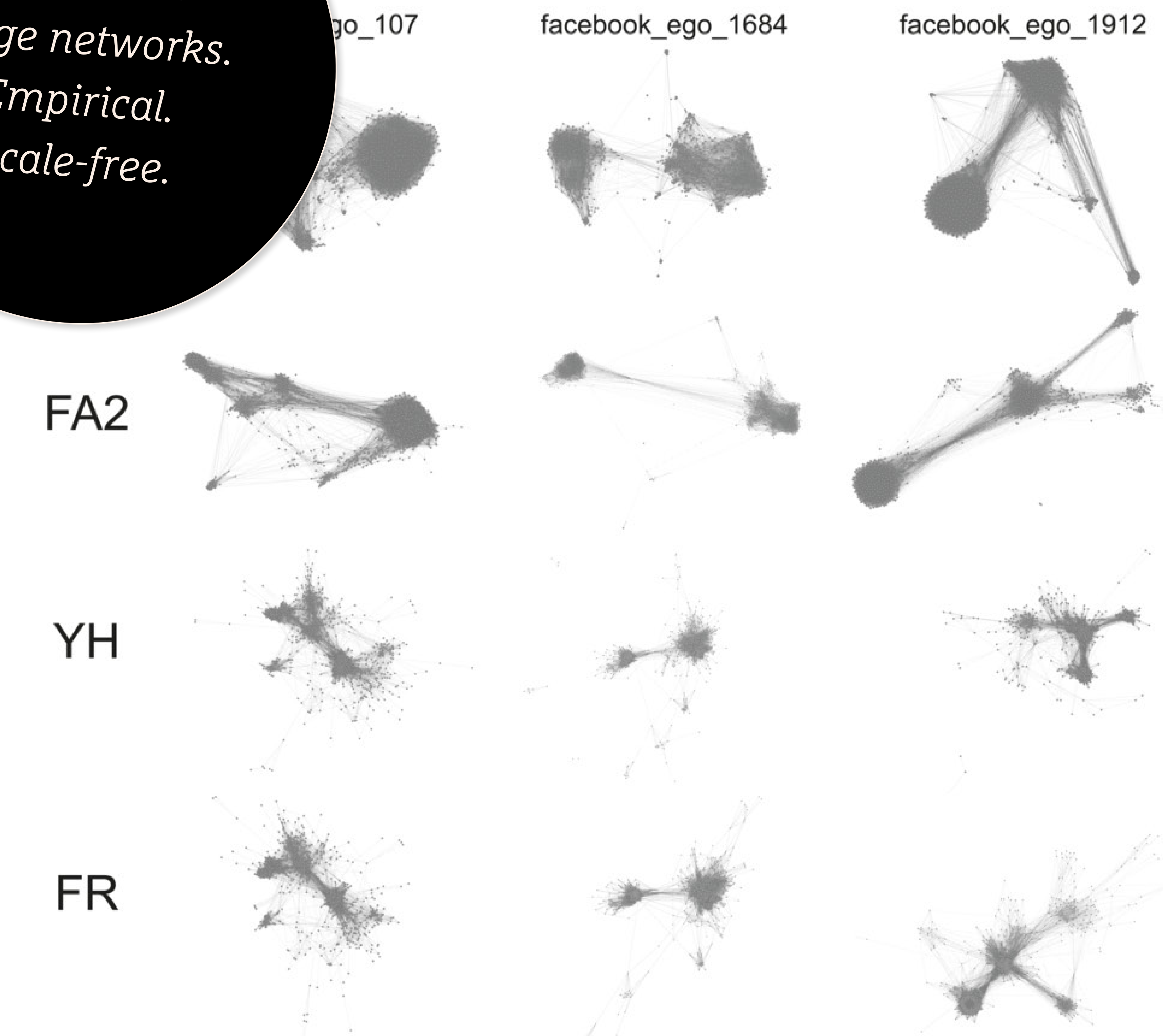
Martin, S., Brown, W.M., Klavans, R. and Boyack, K.W. (2011) 'OpenOrd: An open-source toolbox for large graph layout', Proceedings of SPIE, the International Society for Optical Engineering. Society of Photo-Optical Instrumentation Engineers, San Francisco Airport, California, United States, 24 January 2011, vol. 7868. doi:10.1117/12.871402.

ForceAtlas2, A Continuous Graph Layout Algorithm for Ha
Network Visualization designed for the Gephi software

Mathieu Jacomy*[1,2,3], Tommaso Venturini[1], Sebastien Heymann[3,4] Mathieu Bastian[3]

Algorithm.
Large networks.
Empirical.
Scale-free.

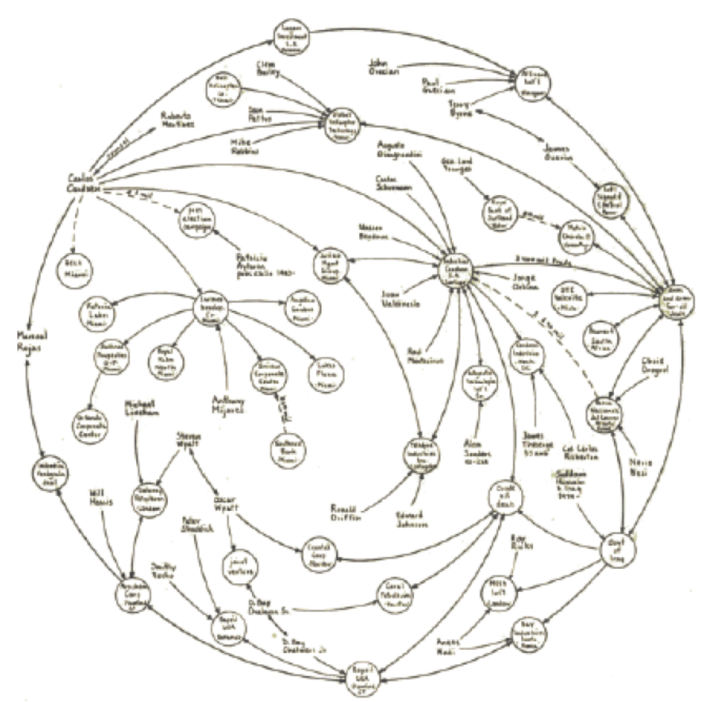
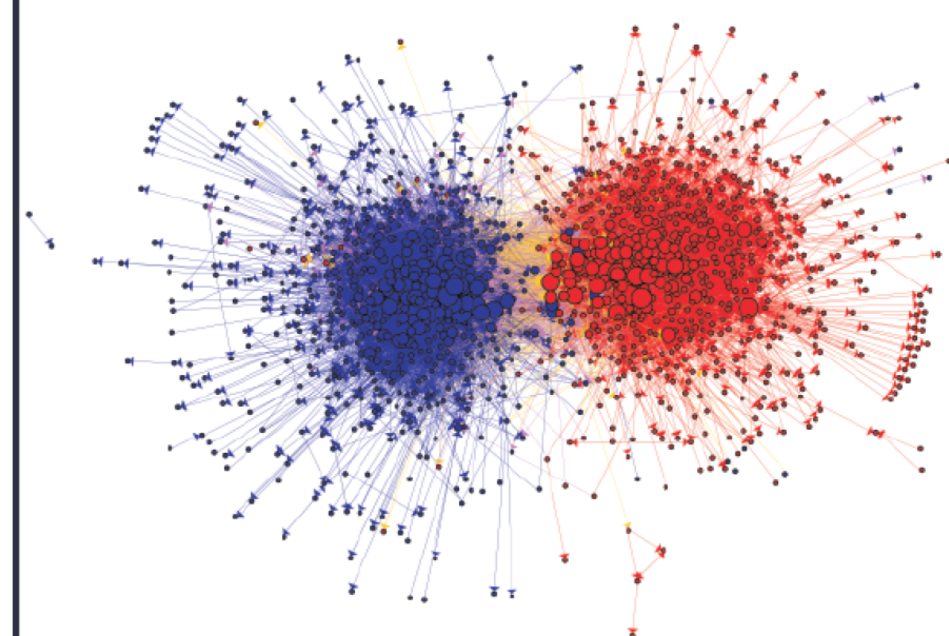




2014

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.

Interpretation regime:

	Diagrammatic	*Topological*
Typical data visualized	Diagram, 10-100 nodes, homogeneous degree distribution 	Complex network, 100-10M nodes, hyperconnected hubs and long tail 
Interpretative tasks	Identify nodes, follow paths.	Describe density distribution in different areas of the image.

TAKEAWAYS

Theory has ALWAYS followed practices.

With a considerable delay.

The field consists of recipes all the way down.
I mean: the field is full of heuristics.

In short: we don't know why it works.

3. Algorithm designers say they know why it works

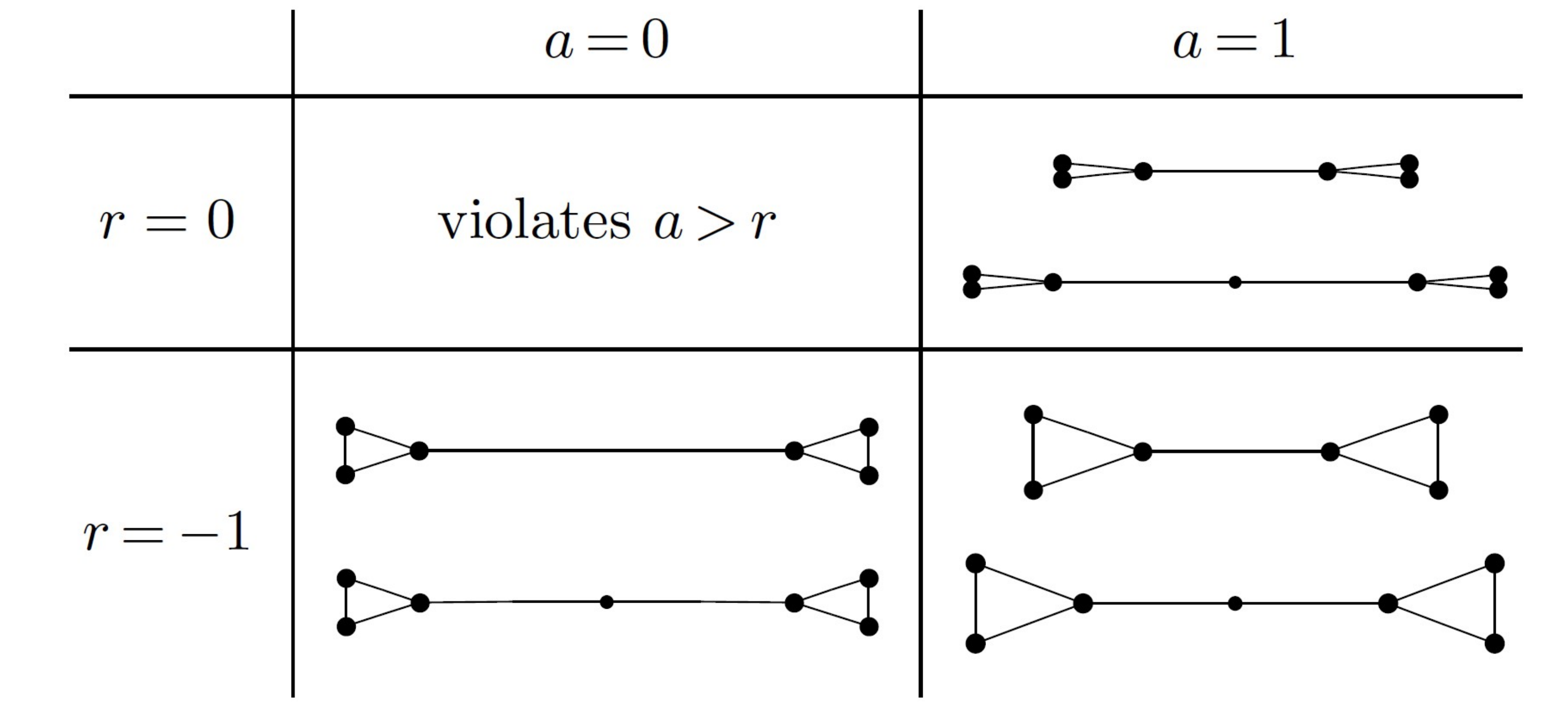


FIG. 2: Layouts with optimal (a, r)-energy for different values of a and r. All vertices and edges have weight 1, except for the small vertex between the triangles which has weight 0.

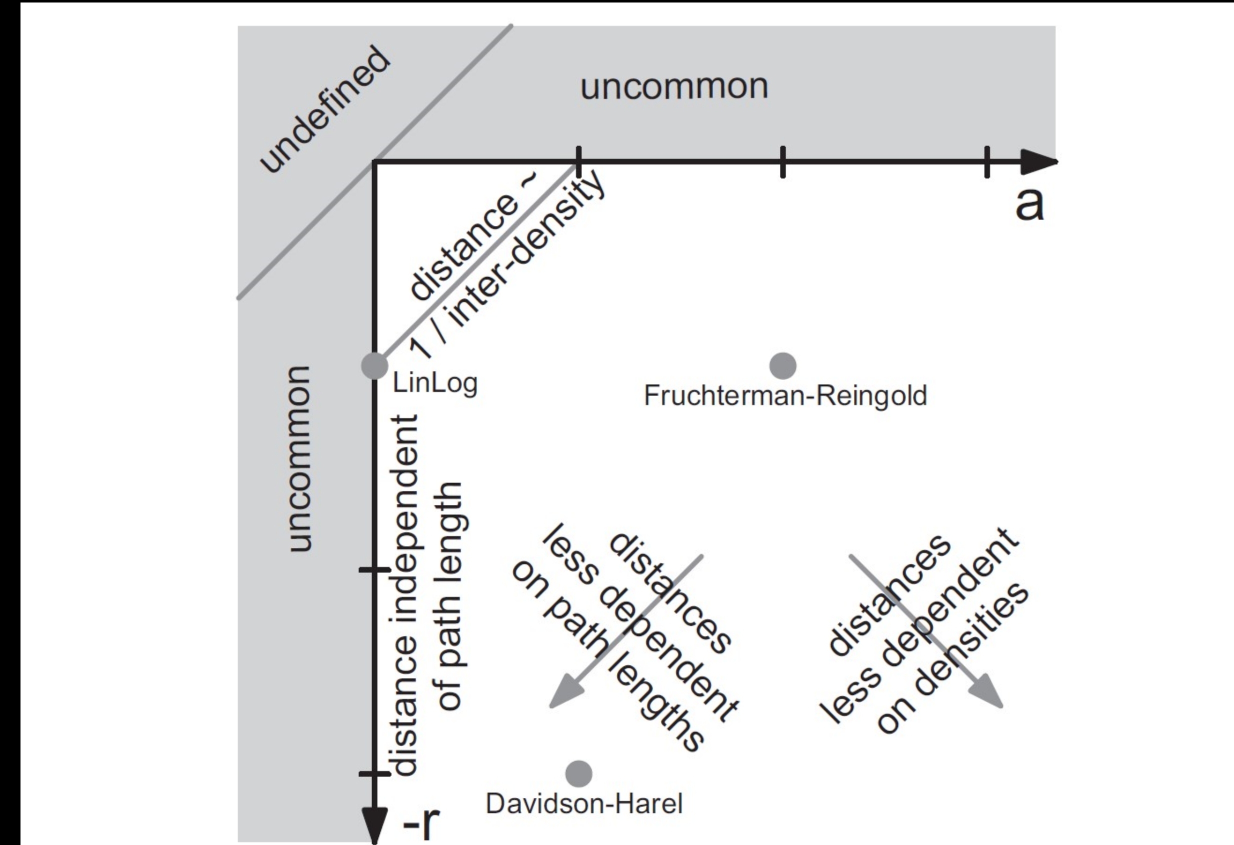


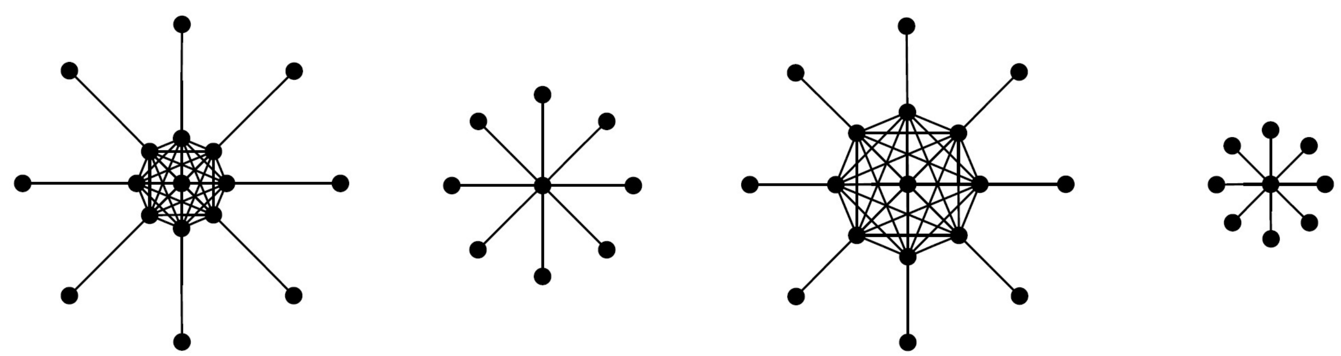
FIG. 3: Impact of the parameters a and r on the optimal layouts of the (a, r)-energy model.

Noack says:

- The goal is cluster separability (I concur)
- It is decided by the attraction and repulsion forces (yes, BUT...)
- The optimal forces are linear and logarithmic. Hence "LinLog". (I concur)

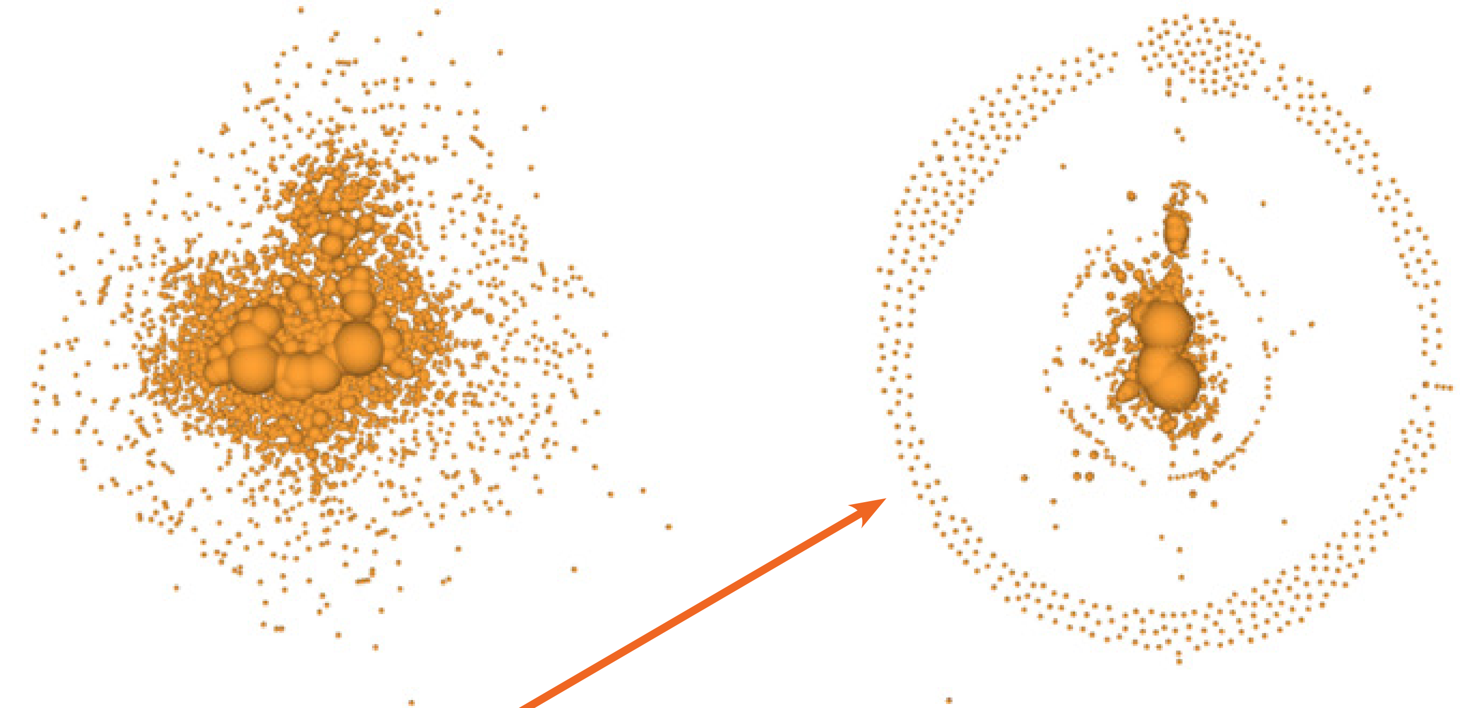
Noack, A. (2007) 'Energy models for graph clustering', Journal of Graph Algorithms and Applications JGAA, 11(112), pp. 453–480.

4. Yet algo designers don’t actually know



(a) (Node-repulsion) Fruchterman-Reingold

(b) Edge-repulsion Fruchterman-Reingold



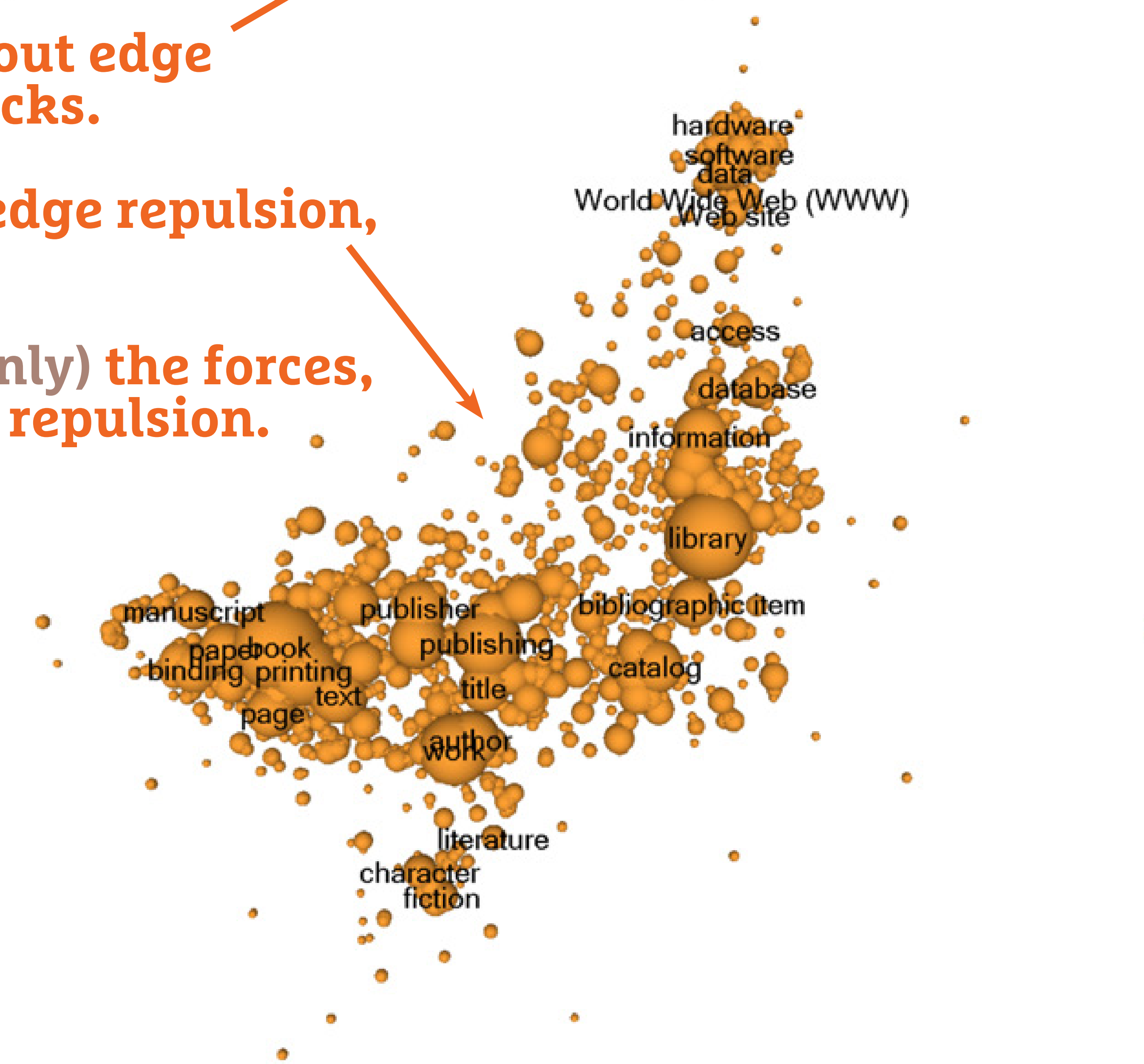
(a) Fruchterman-Reingold model

(b) Node-repulsion LinLog model

LinLog without edge repulsion sucks.

While with edge repulsion, it’s great.

It’s not (mainly) the forces, it’s the edge repulsion.
‘=_=



(c) Edge-repulsion LinLog model

Noack, A. (2007) ‘Energy models for graph clustering’, Journal of Graph Algorithms and Applications JGAA, 11(112), pp. 453–480.

5. Who cares?

Users care that:

- It works in practice
- They can justify the method

An algo designer does not care because:

- Their algo works,
- it has a rationale…
- …that passed peer review.

Job done, right?

As long as we agree to not look under the hood, this is good science.

There is a risk to reopening a solved problem:

- Reopening a “cold case” may bring back existing conflicts/tensions
- It is inherently controversial (it goes against the consensus)
- There might be no benefits to the time and effort spent on it

Yet benefits may still come out:

- Innovation or breakthrough (who knows)
- Better methodological grounding
- Better explanation (teaching)

Interlude
What am I up against?

A tool maker’s perspective


What do the algorithm users want?
(and what do the tool makers want?)
GECKO DAVINCI
Évangélistes (17%)
Gauche/LFI (33%)

Interlude: what am I up against?

The user of a tool or algorithm
may have different goals
than those expected by the author
of that tool or algorithm.

Is this controversial?

Interlude: what am I up against?

Here is a point of view you find in the digital humanities

BLACK BOX

The tool is
EASY TO USE
HIDING COMPLEXITY

The user is
LAZY THINKING
UNREFLECTIVE

Interlude: what am I up against?

Here is a point of view you find in the digital humanities

$\overbrace{\hspace{8em}}^{\text{CRITICAL THINKING}}$

The tool is
DIFFICULT TO USE
EXPOSING COMPLEXITY

The user is
HARD THINKING
REFLECTIVE

Interlude: what am I up against?

Here is a point of view you find in the digital humanities

BLACK BOX

CRITICAL THINKING

Ease-of-use axis

The tool is
EASY TO USE
HIDING COMPLEXITY

The tool is
DIFFICULT TO USE
EXPOSING COMPLEXITY

The user is
LAZY THINKING
UNREFLECTIVE

The user is
HARD THINKING
REFLECTIVE

I oppose this view!

Interlude: what am I up against?

The problem with ignoring the user's own needs

Interlude: what am I up against?

The problem with ignoring the user's own needs

a. ...but Mjölnir
cannot be lifted!

b. So the user finds another way:
They hit Mjölnir with the nail.

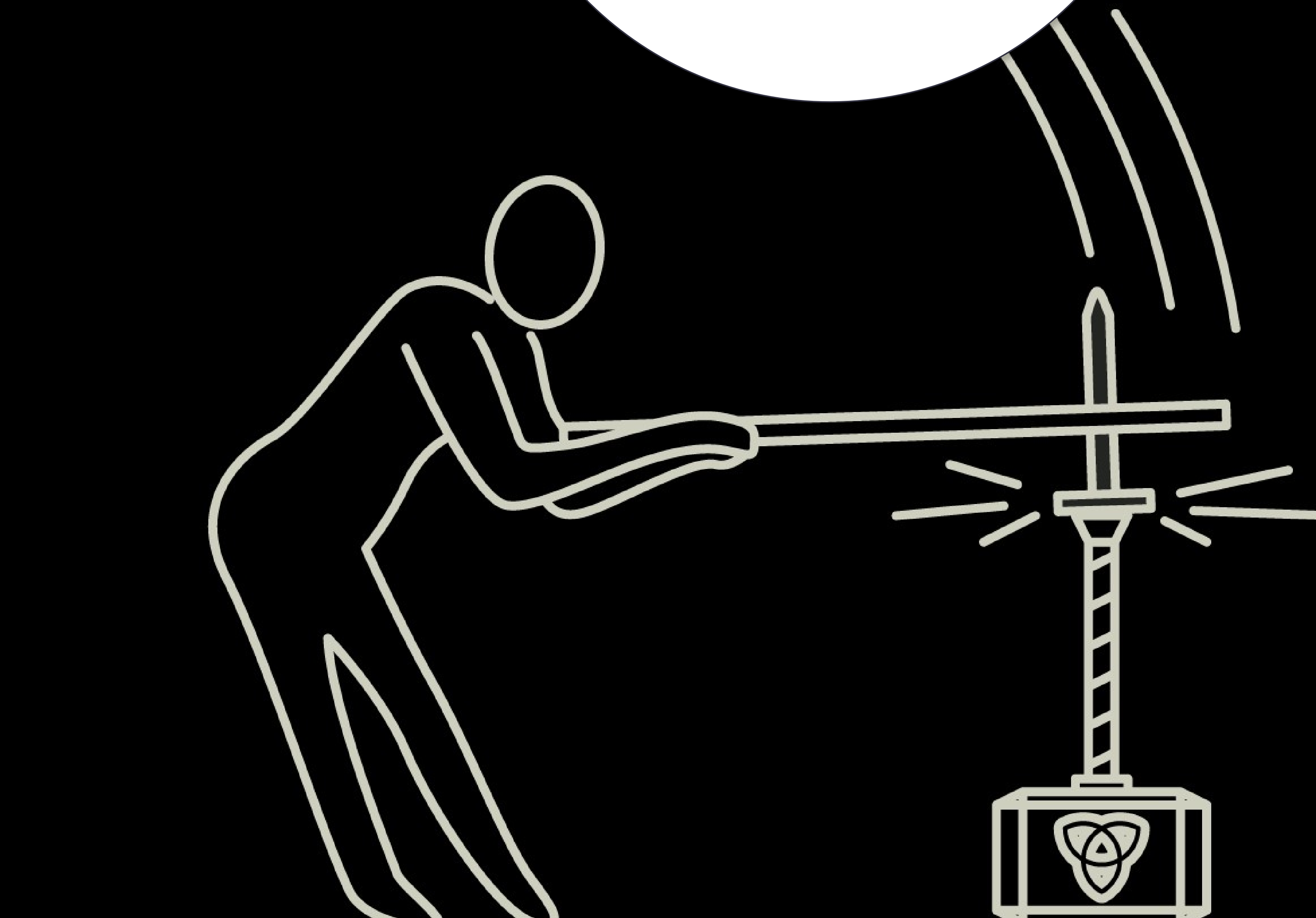
Interlude: what am I up against?

The problem with ignoring the user's ov

Time constrained situations;
Learning by doing;
Testing;
...

The user needs to hammer the nail...

a. *...but Mjölnir cannot be lifted!*



b. *So the user finds another way: They hit Mjölnir with the nail.*

Interlude: what am I up against?

In my abstract for this conference I wrote:

“Observing network analysis practices shows that users have their own epistemic culture.”

(You’ll soon meet my Reviewer #2)

What are epistemic cultures:

“amalgams of arrangements … which, in a given field, make up how we know what we know. *… cultures that create and warrant knowledge.”*

Knorr-Cetina, K. (1999). Epistemic cultures: How the sciences make knowledge. Harvard university press.

Meet my reviewer #2

"If this contribution is accepted, it is to highlight how caricatural science has become in some areas where criticizing it is considered a problem, the user is always right even when wrong (because epistemic culture etc.), existing practices are perfect, and not understanding is great."

Allégorie du seum
Gustave Doré, Le Louvre (definitely not)

Why would “*users have their own goals*”
imply that “*their practices are unproblematic*”?

→ Because you think that
practices should be bound
by academic authority.

Yet academic authority follows from the practices.

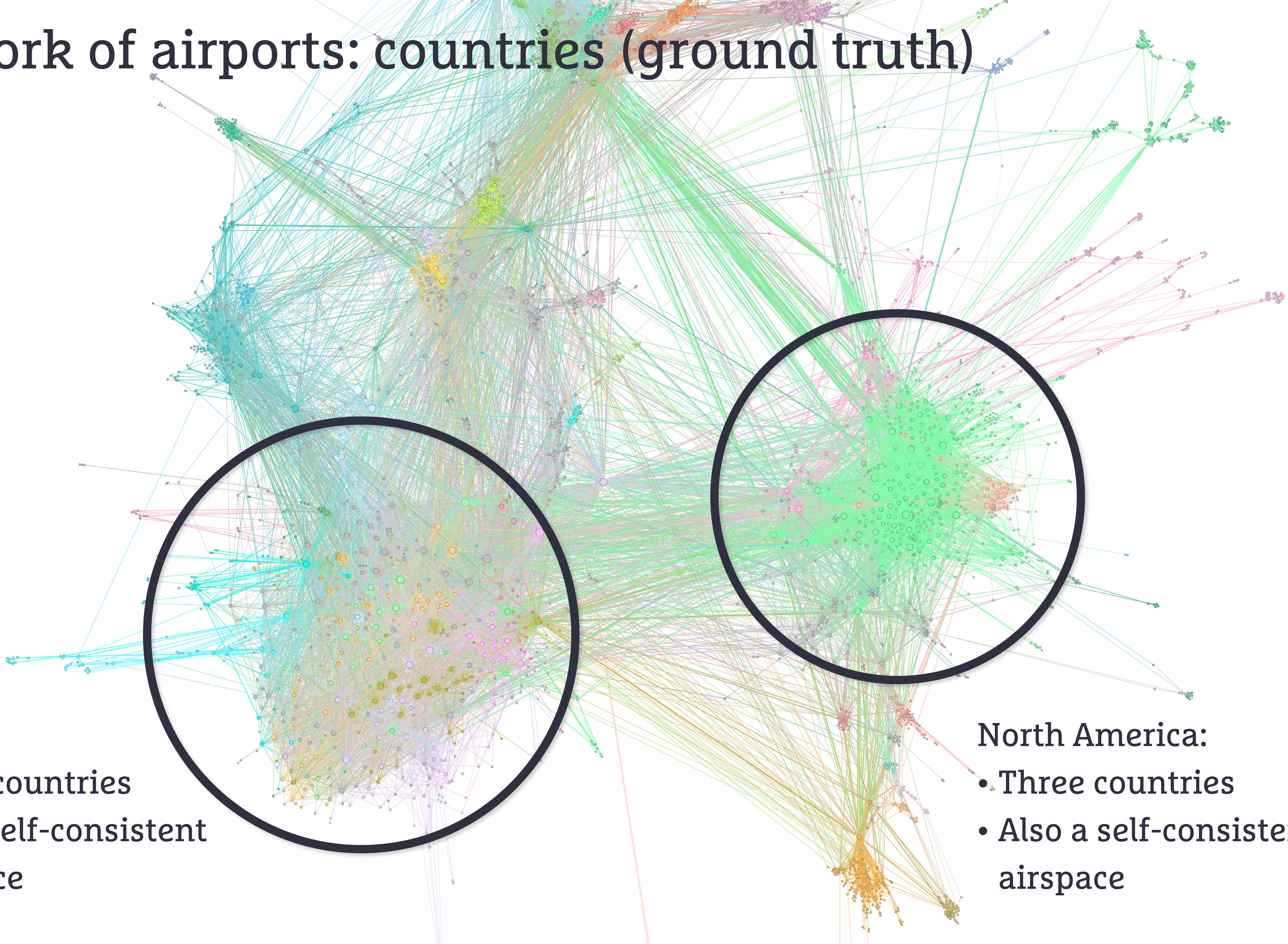
Case #2: Community detection (in networks)

This case in short:

1. People like the Louvain method
2. The Leiden and Bayesian Inference methods are claimed to be superior by their designers
3. Some users still prefer Louvain
4. That's because they do something else than what algorithm designer consider should be done with these algos
5. Those designers still contend that these users are wrong
6. This boils down to my reviewer #2: Users cannot have their own goals

Network of airports: countries (ground truth)

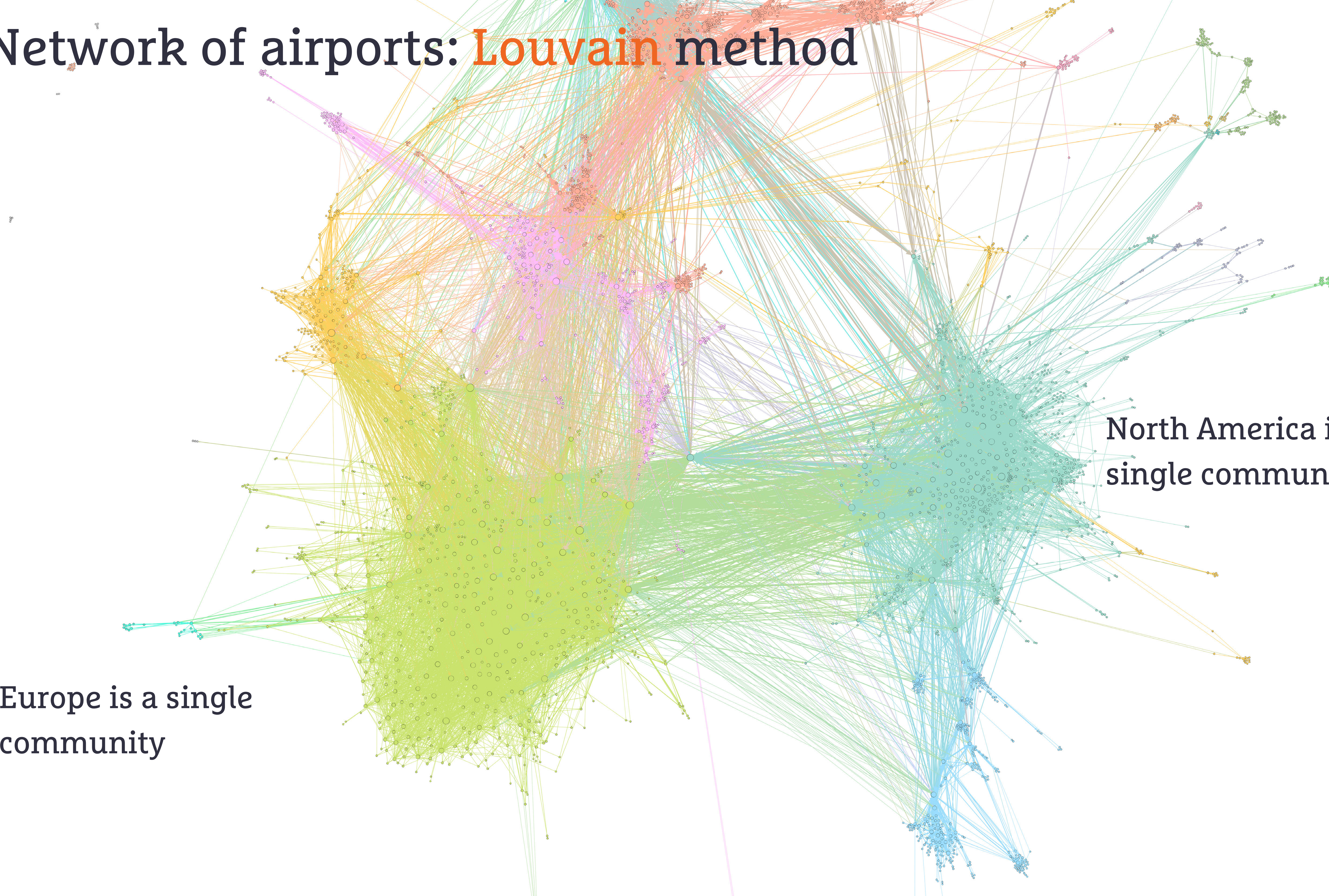
Network of airports: countries (ground truth)



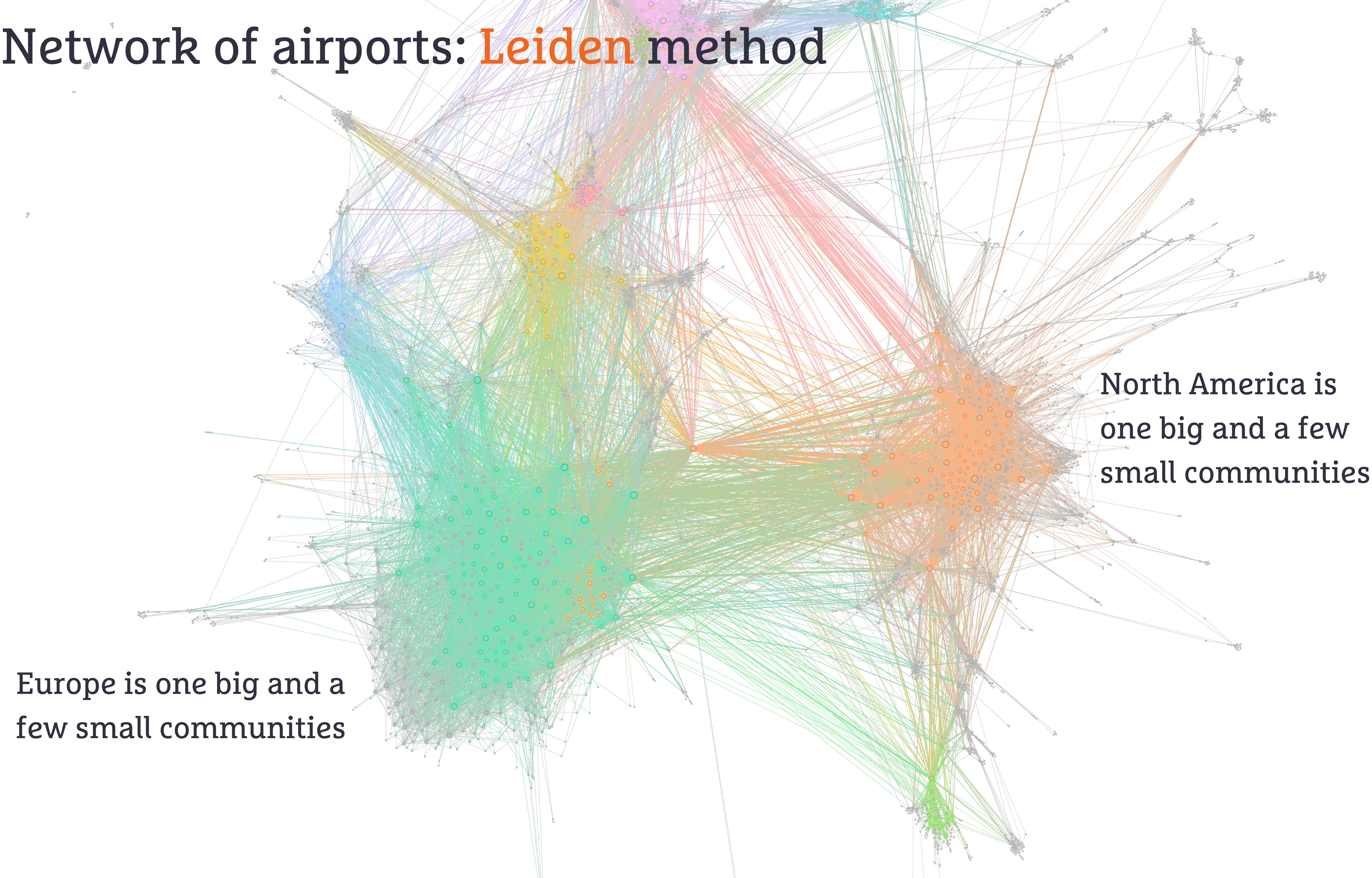
Europe:
- Many countries
- But a self-consistent airspace

North America:
- Three countries
- Also a self-consistent airspace

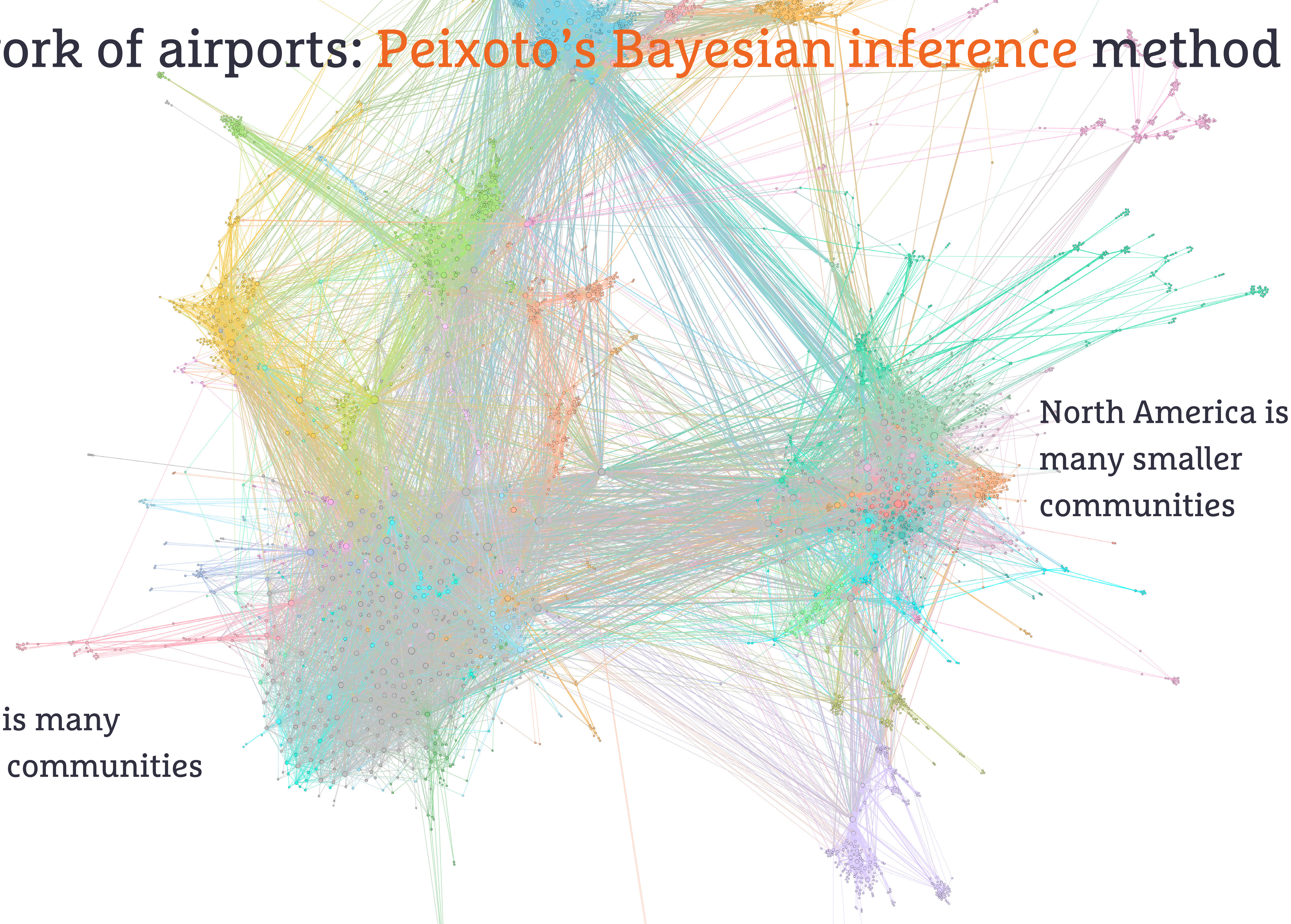
Network of airports: Louvain method



Network of airports: Leiden method



Network of airports: Peixoto’s Bayesian inference method



North America is many smaller communities

Europe is many smaller communities

What makes Leiden and Bayesian Inference better

LOUVAIN

With Louvain you can set “resolution”:
How big you want the communities.

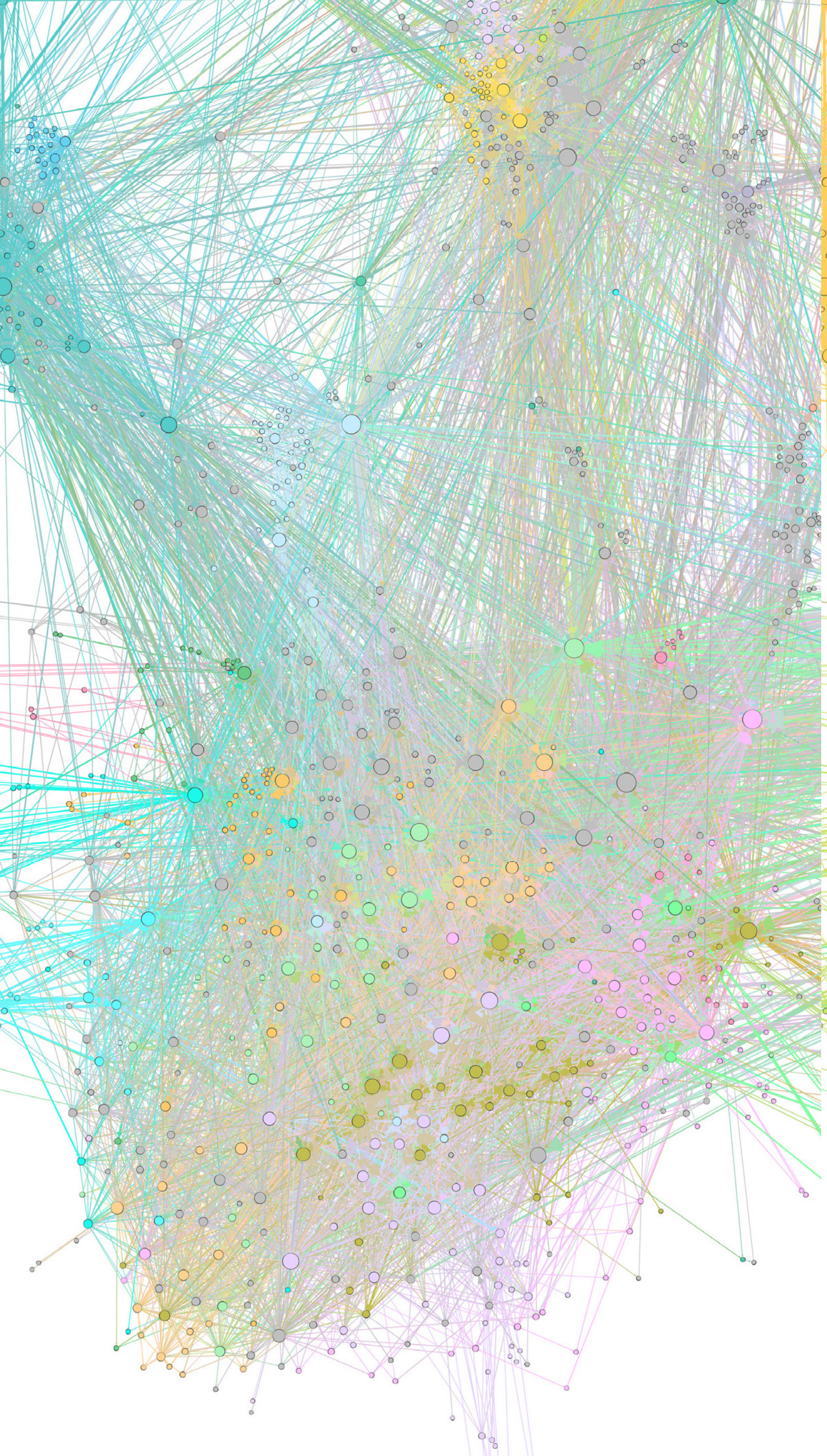
The Louvain method has a known bias:
It finds same-size communities.

LEIDEN

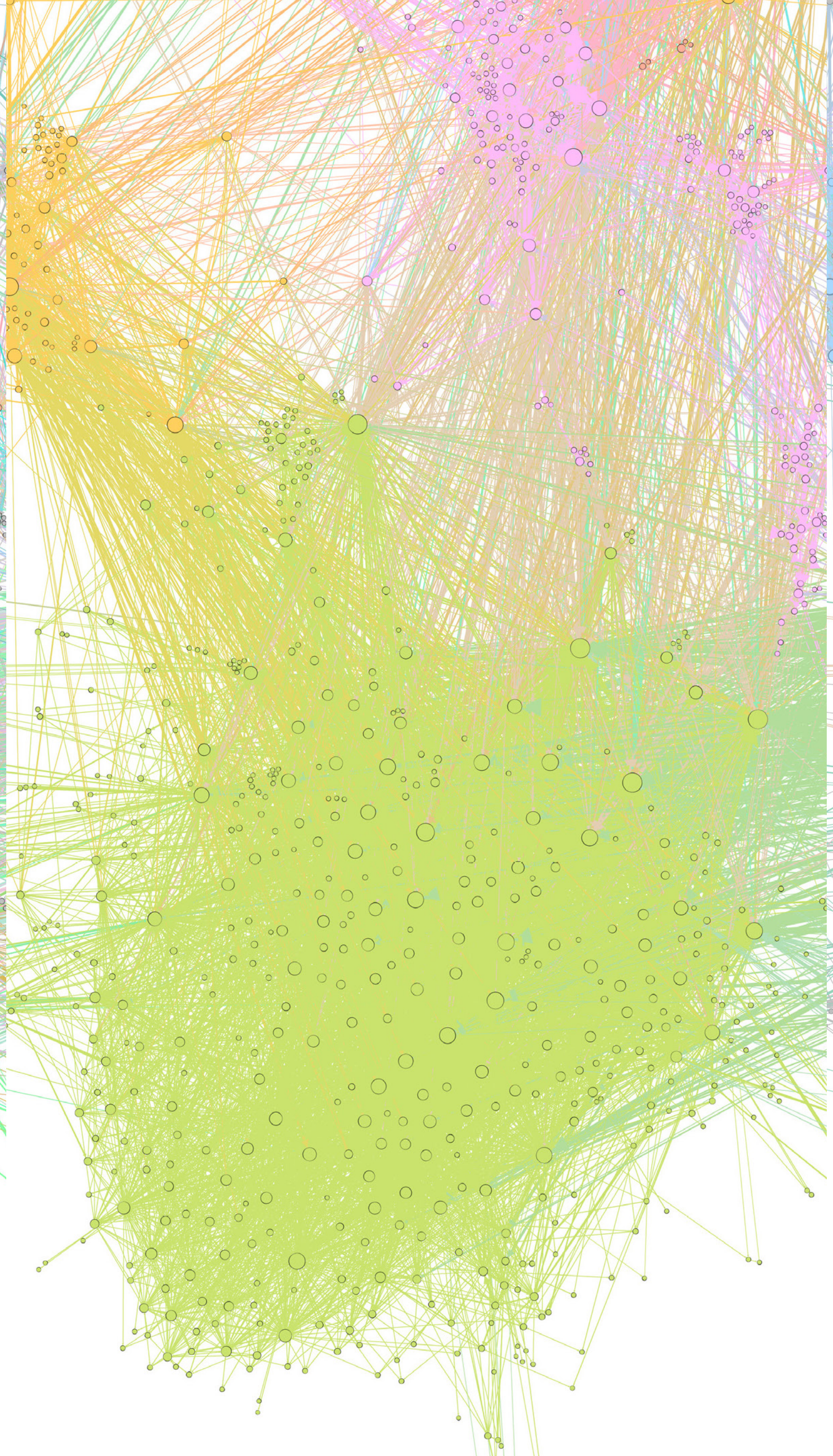
The Leiden method fixes that bias.
Still has the resolution setting.
Better mathematical justification.

PEIXOTO’s Bayesian Inference

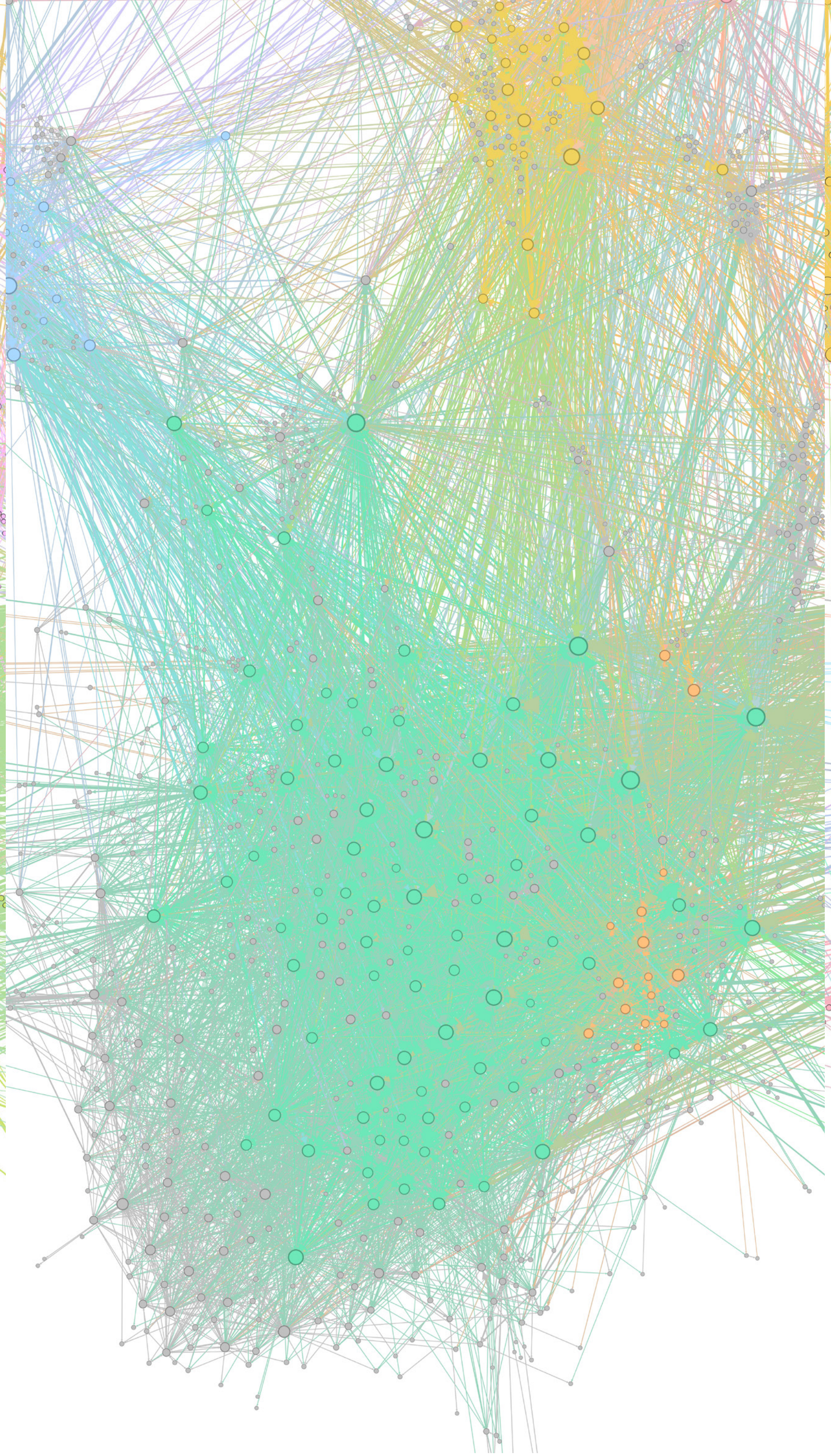
Also fixes the bias.
Even better mathematical justification.
(explicitates model assumptions)
Has no resolution. (non-parametric)



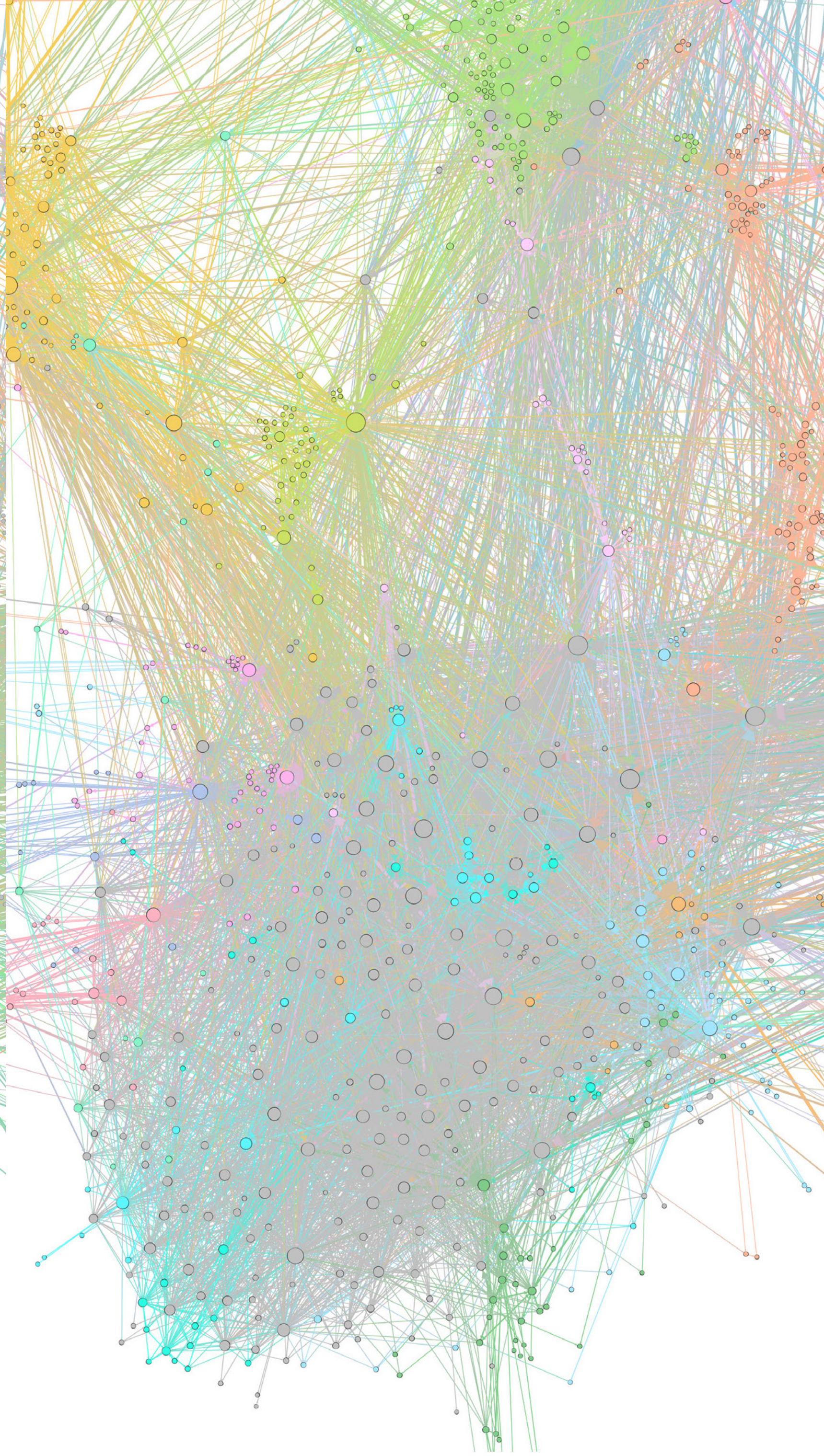
Countries
(ground truth?)
(or maybe not?)



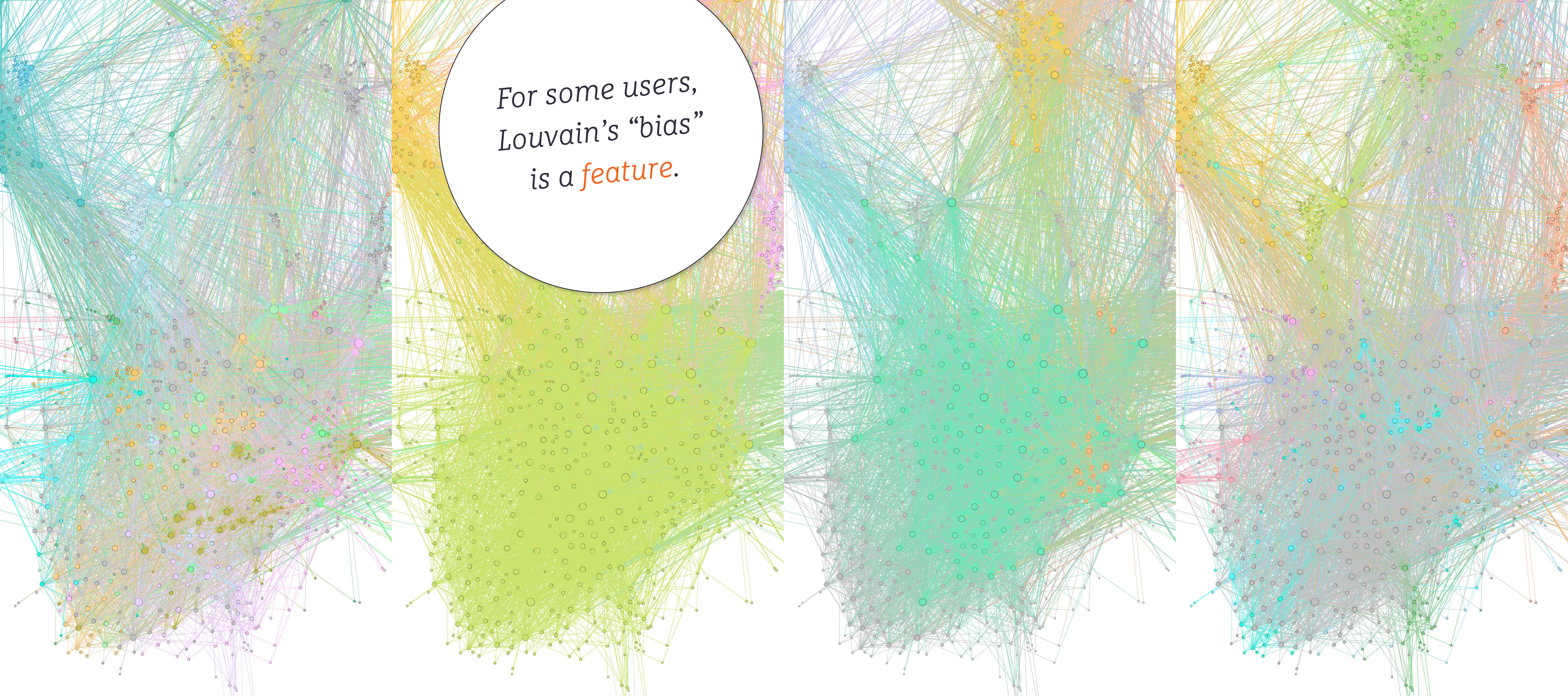
LOUVAIN
Few big chunks.
Useful to summarize the structure of the network.



LEIDEN
Mixed chunks.
Useful to retrieve macro & micro structures.



BAYESIAN INFERENCE
Can predict communities that follow the model's assumptions.



Countries
(ground truth?)
(or maybe not?)

LOUVAIN
Few big chunks.
Useful to summarize the structure of the network.

LEIDEN
Mixed chunks.
Useful to retrieve macro & micro structures.

BAYESIAN INFERENCE
Can predict communities that follow the model's assumptions.

Tiago Peixoto's argument

Descriptive methods like Louvain "do not articulate precisely what constitutes community structure" contrary to inferential methods.

Therefore they "carry no explanatory power." The communities obtained from descriptive methods "can be seen and described, but they cannot explain."

"Every descriptive method can be mapped to an inferential one, according to some implicit model."

Descriptive methods are inferential methods that do not state their model, which makes them inherently worse.

"There is no such thing as a 'model-free' community detection method."

→ For Tiago Peixoto, you would be wrong to prefer Louvain over Bayesian inference.

→ But he assumes that you always aim to predict, since there always is a model.

Tiago Peixoto. https://skewed.de/tiago/blog/descriptive-inferential

Sympathy

Dear algorithm designer,

Users will repurpose your creation, and if you want to criticize it, you must put the effort to understand why.

I am sorry for your loss,
Mathieu

PS: I’ve been there. Feel free to give me a call.

Conclusion:
How to keep the gate of the algorithm

“Scientific and technical work is made invisible by its own success.”
— Bruno Latour

When technology works, the science of understanding why is often undone.

Yet it could help us:

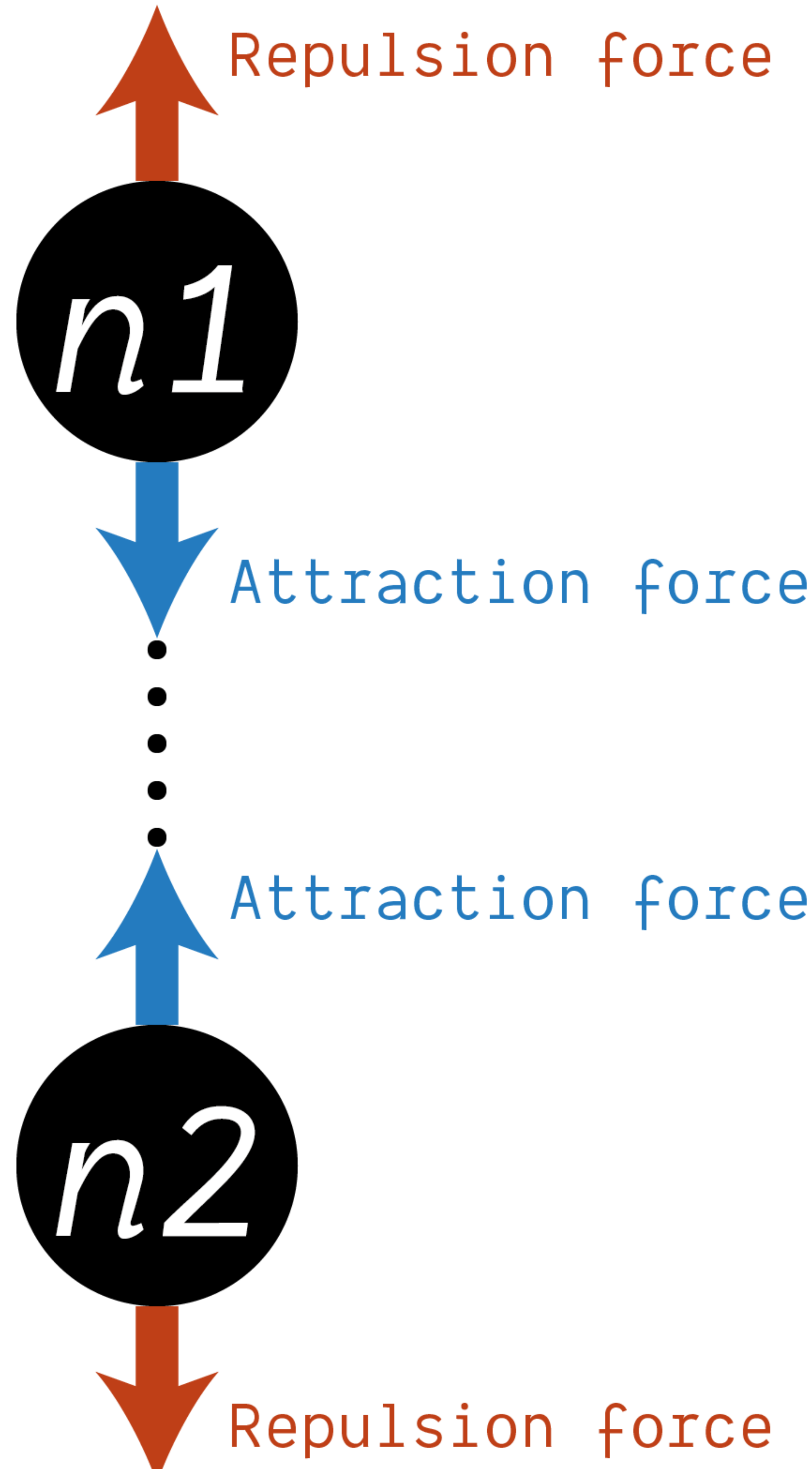
- Find new purposes to existing algorithms (invented by users)
- Find out when users are actually wrong, and help them improve
- Do science that supports existing practices

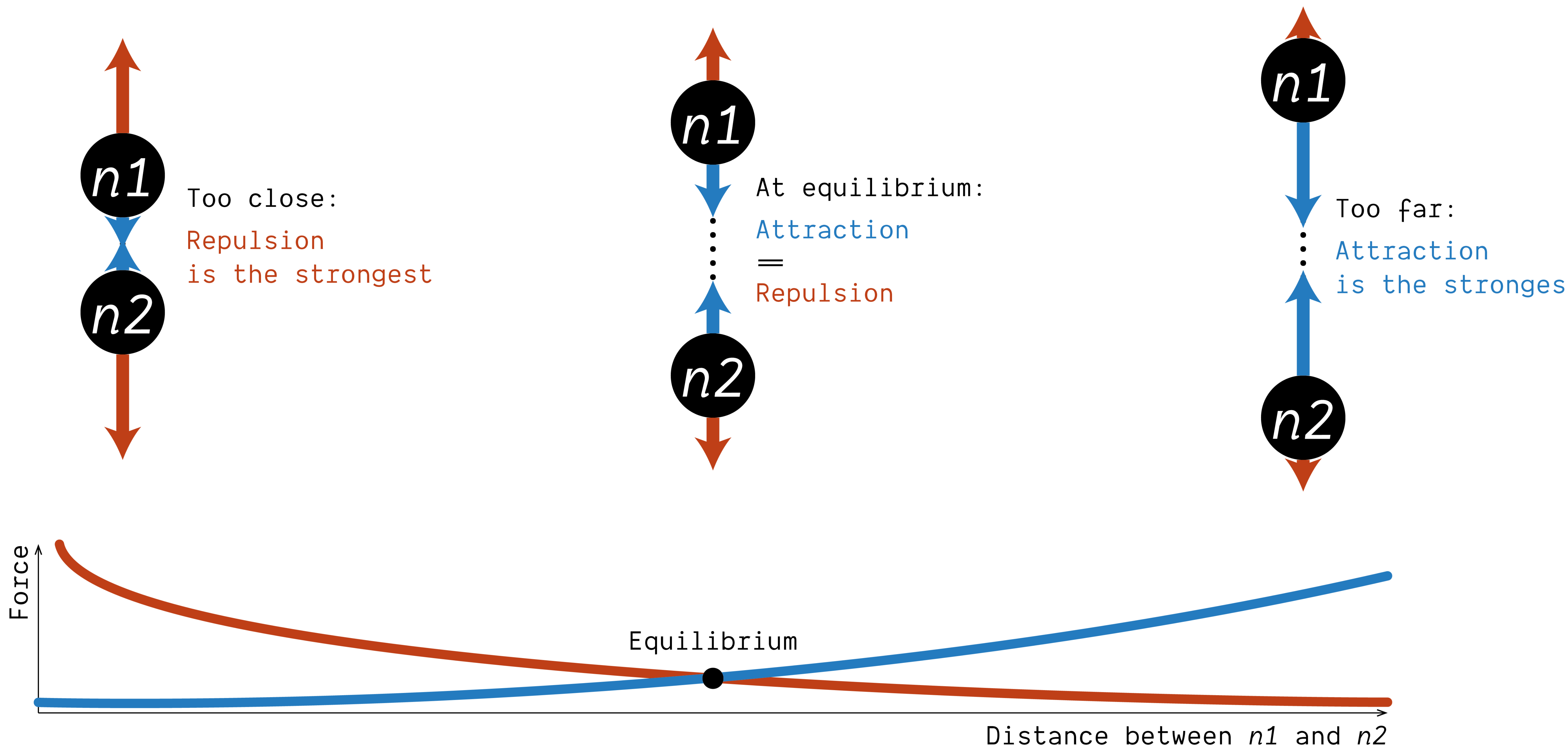
My wishful thinking for doing that undone:

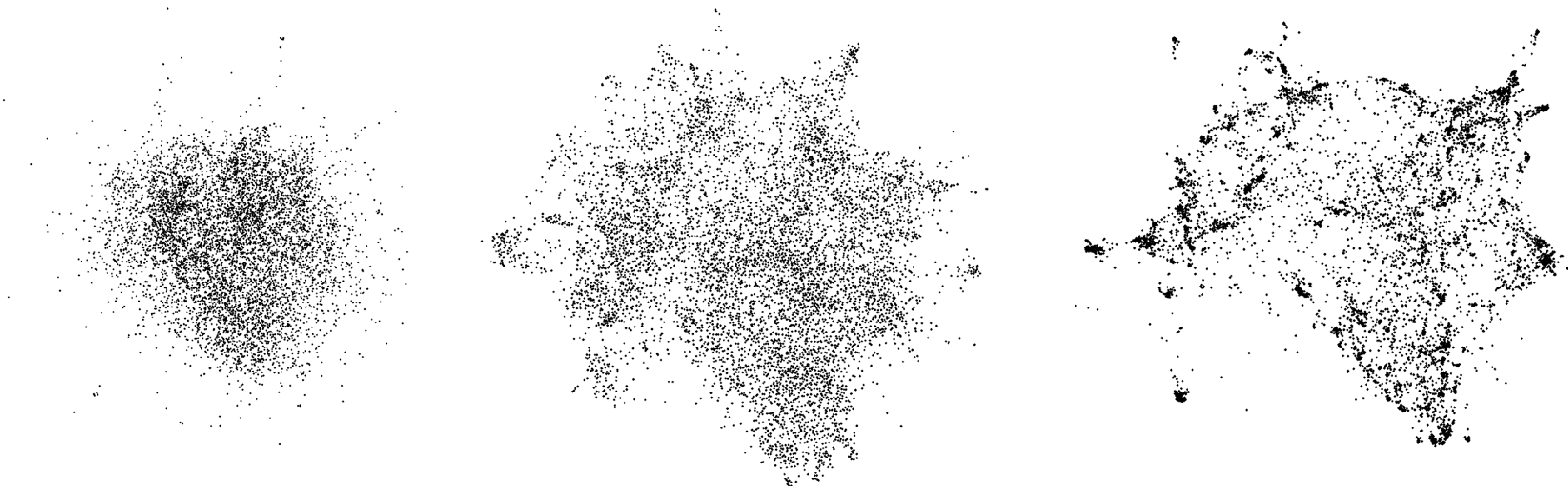
1. Peer review should allow an algorithm author to not know why it works. (they may still show it does!)
2. Algo designers should face user practices. Understanding before gatekeeping.
3. Explaining should have academic currency. Not just novelty or efficiency.

Thank you for your attention.

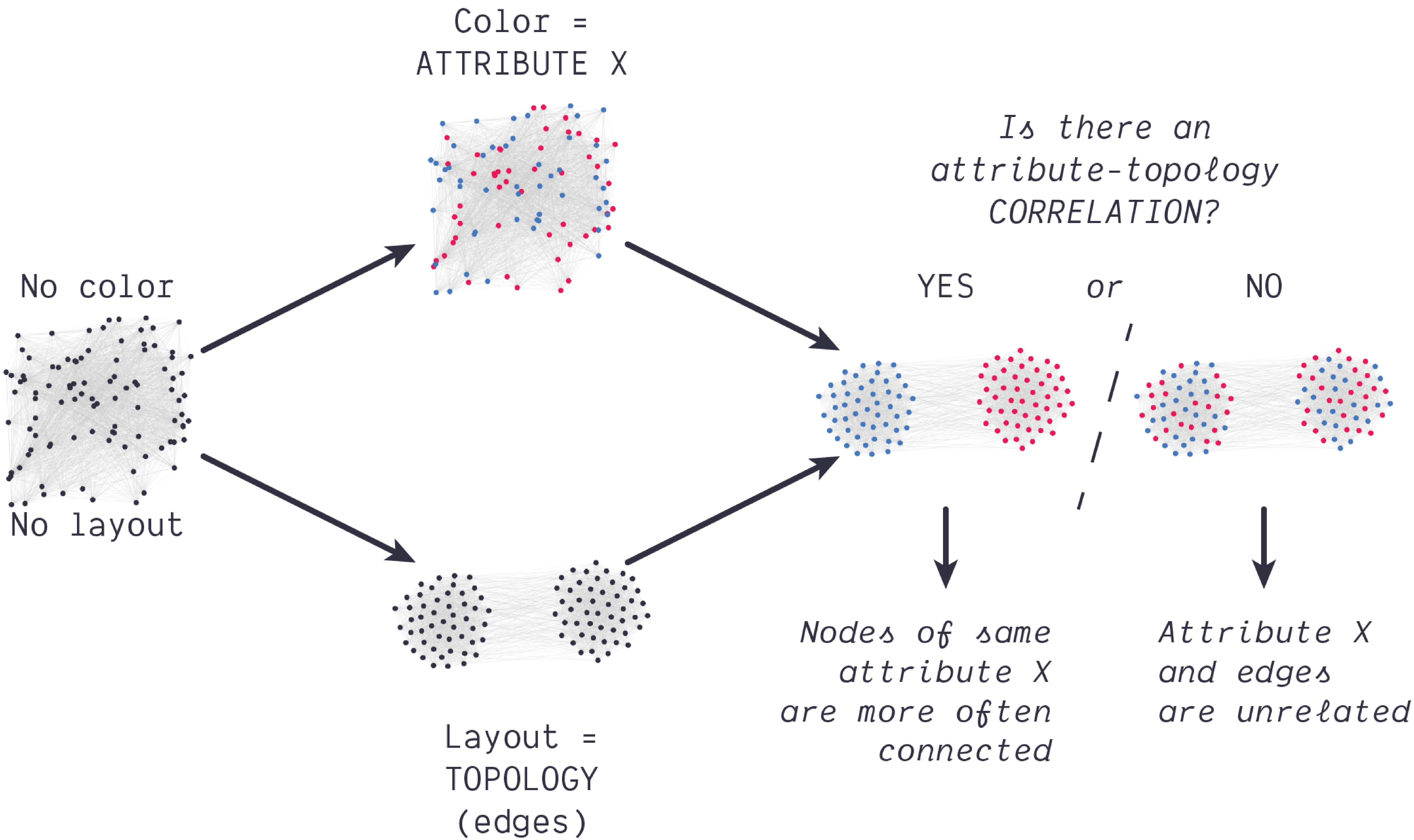
@jacomyma@mas.to
reticular.hypotheses.org


Repulsion force
n1
Attraction force
Attraction force
n2
Repulsion force


n1
n2
Too close:
Repulsion
is the strongest
n1
n2
At equilibrium:
Attraction
=
Repulsion
n1
n2
Too far:
Attraction
is the strongest
Force
Equilibrium
Distance between n1 and n2



Different algorithms produce different results.


Color =
ATTRIBUTE X
No color
No layout
Layout =
TOPOLOGY
(edges)
Is there an
attribute-topology
CORRELATION?
YES
or
NO
Nodes of same
attribute X
are more often
connected
Attribute X
and edges
are unrelated

The blue dots gather on the left, the red dots gather on the right (image layer).

...which means that...

The layout placed the nodes of the same kind together (layout layer).

...which means that...

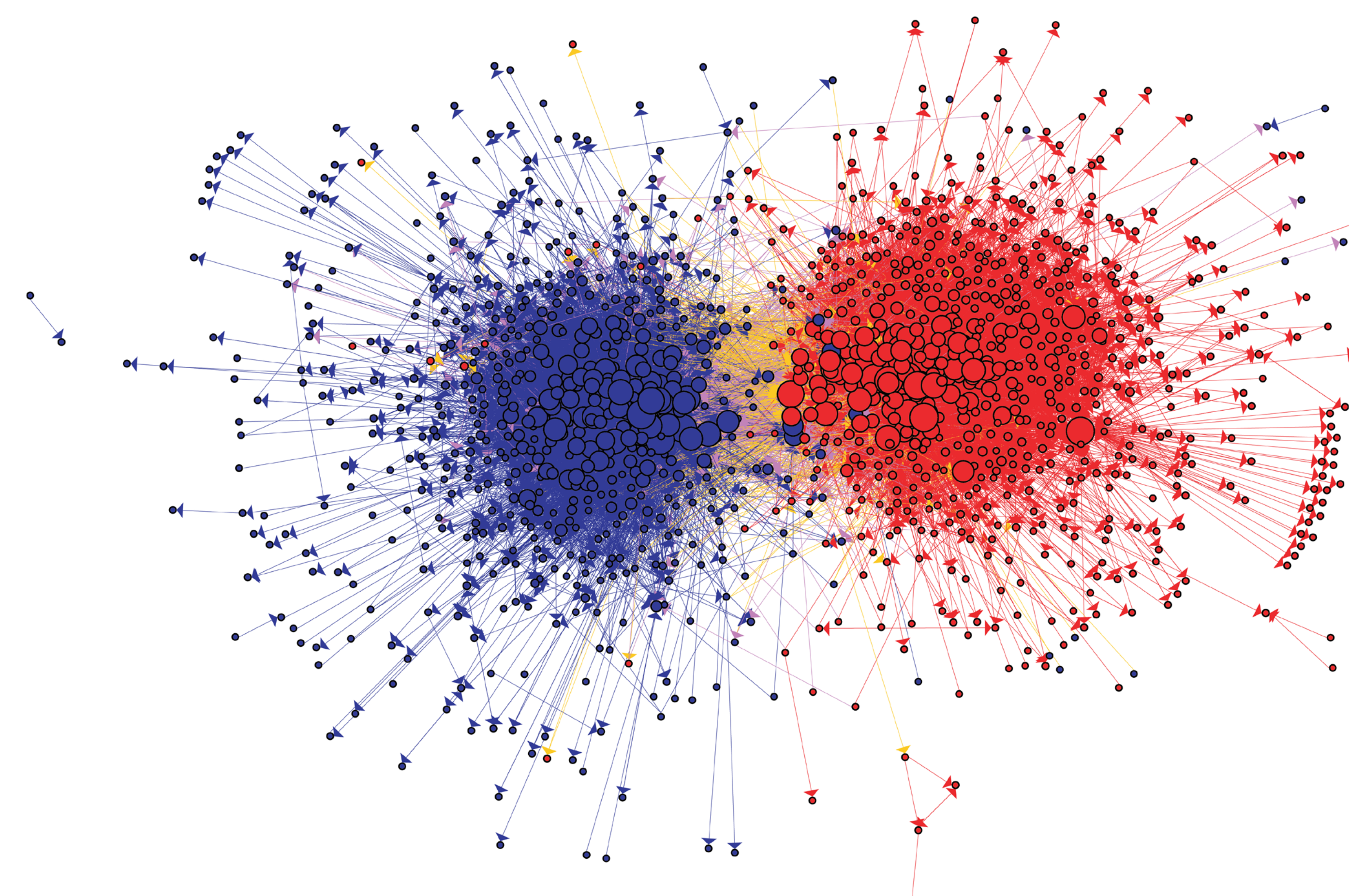
The blogs tend to connect more with blogs of the same political affiliation (network layer).

...which means that...

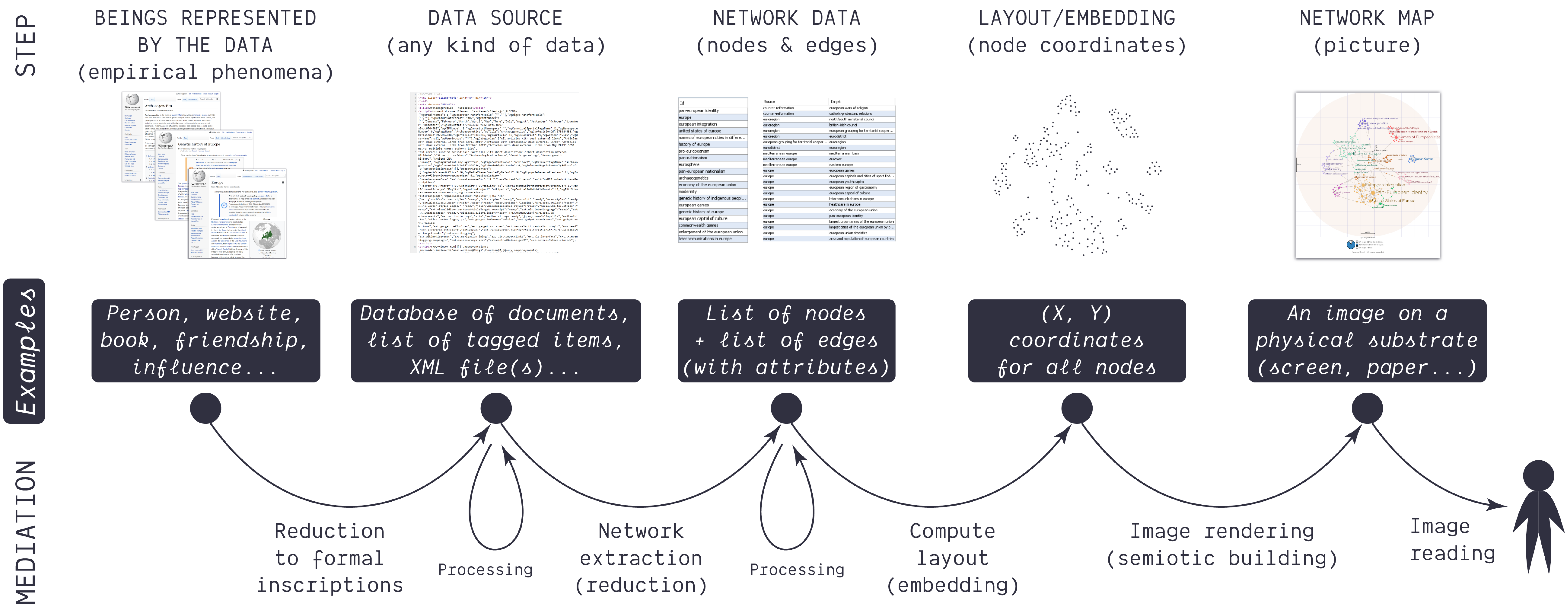
When bloggers add a blog to their blogroll, it generally has a similar political content (phenomenon layer).

...interpretation:

The behavior of political bloggers features homophily (tendency to link to the same) which results in the polarization of the political web.



Layers of mediation



Situating visual network analysis, Jacomy, 2021.