tantlab

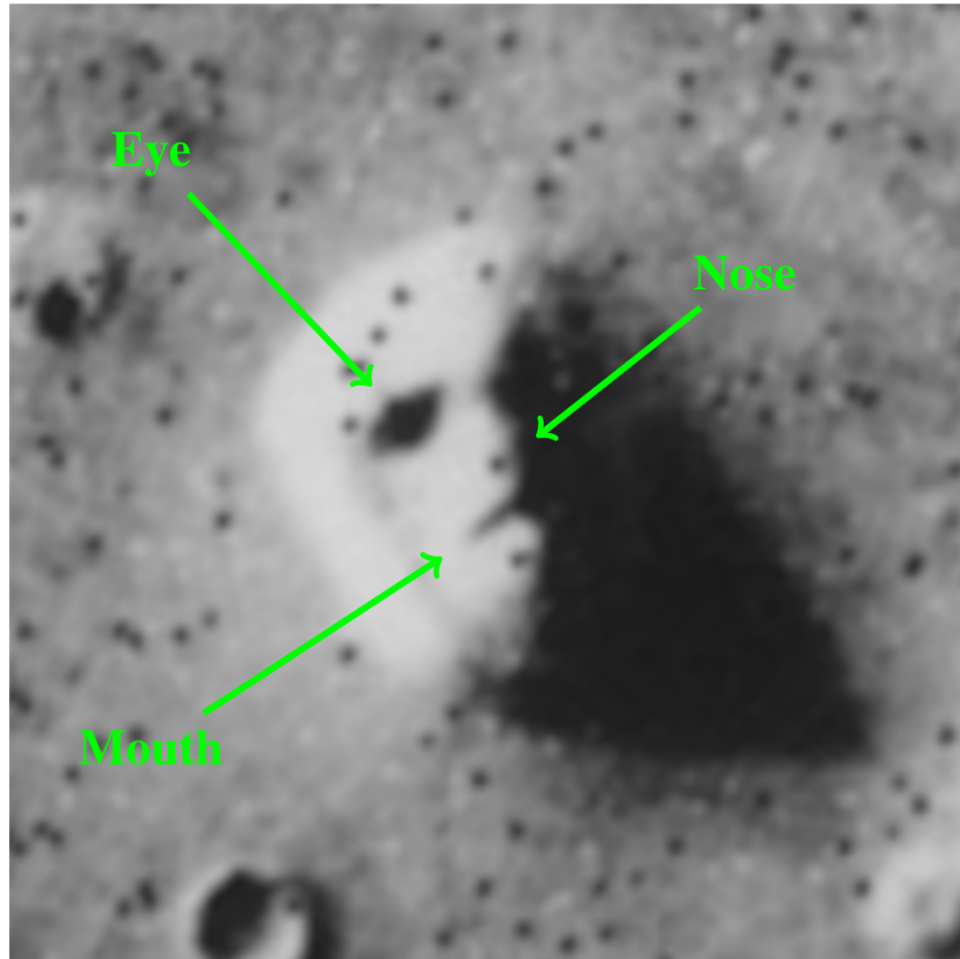# Practices versus modeling
Community detection in network analysis
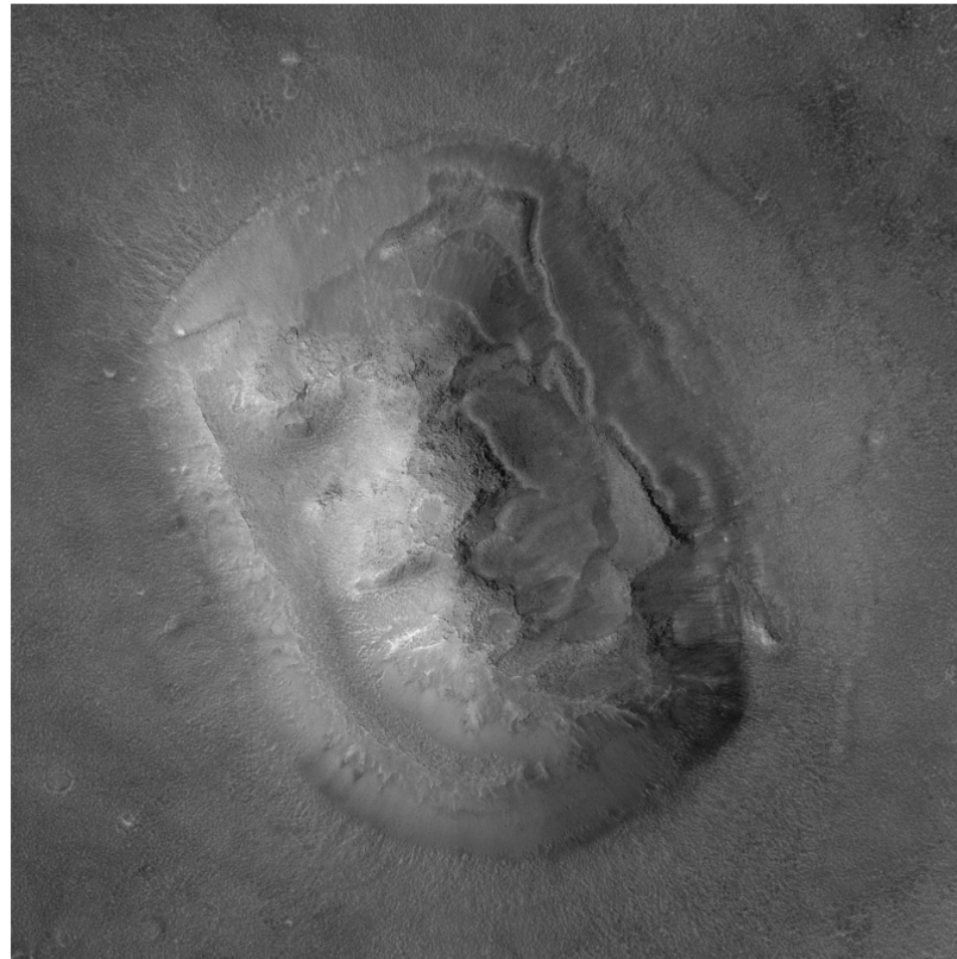
Mathieu Jacomy
Aalborg University Tantlab

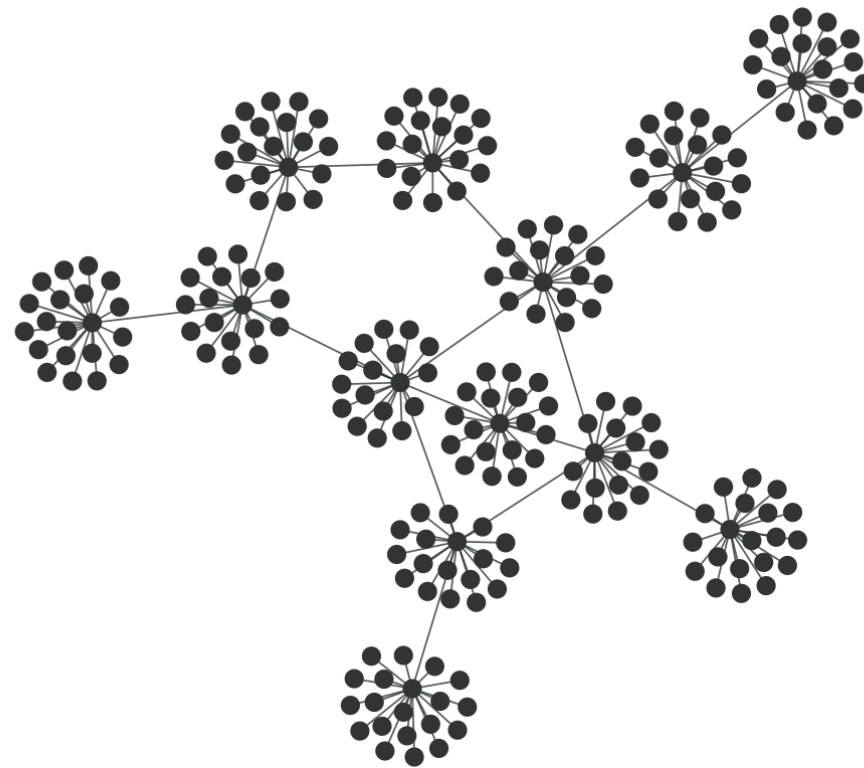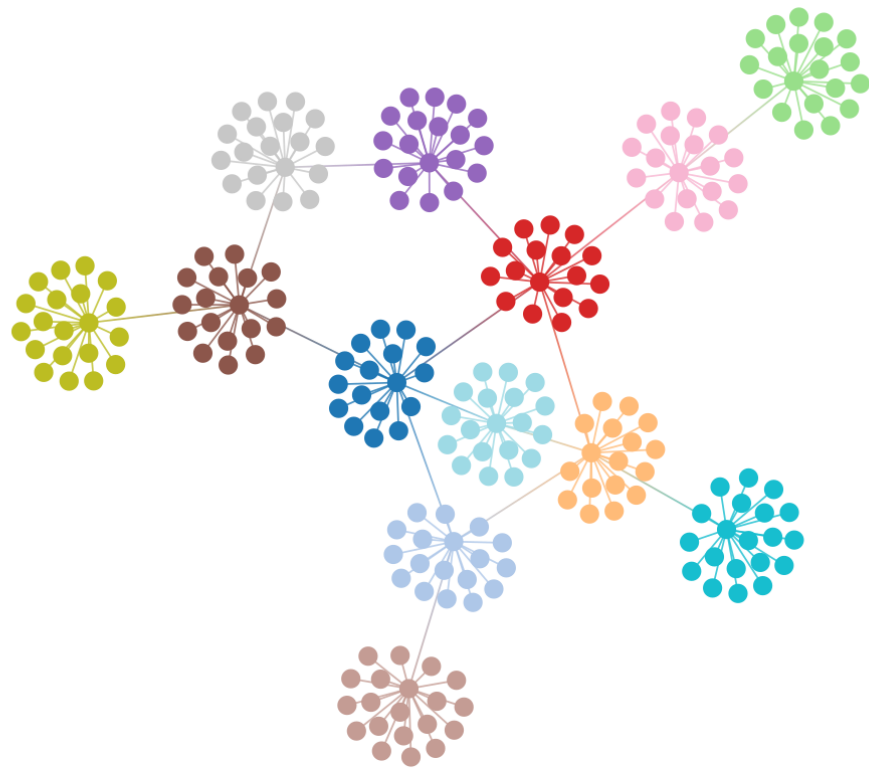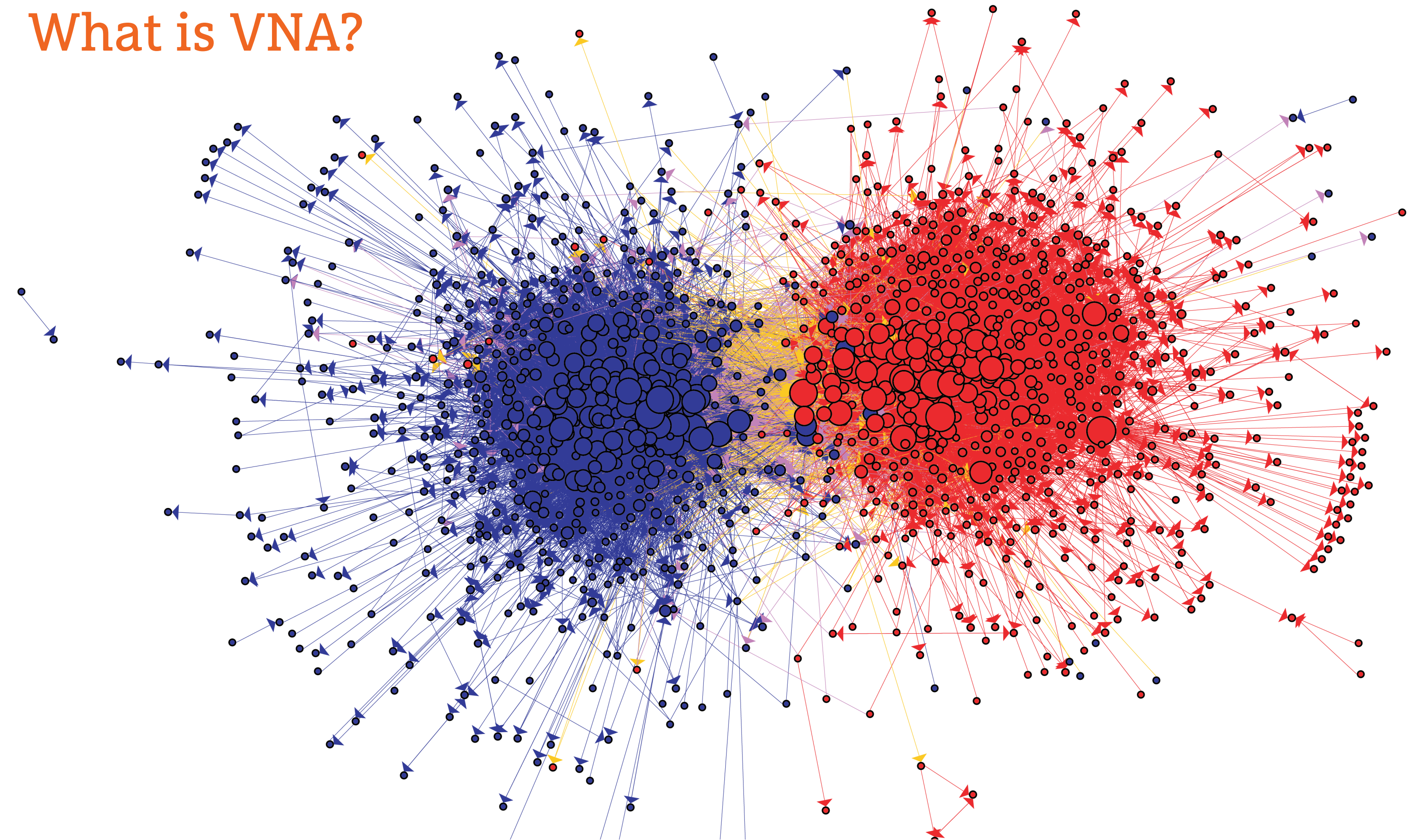| Description	Explanation
A face	A mountain

It starts with a disagreement.

Are there communities?

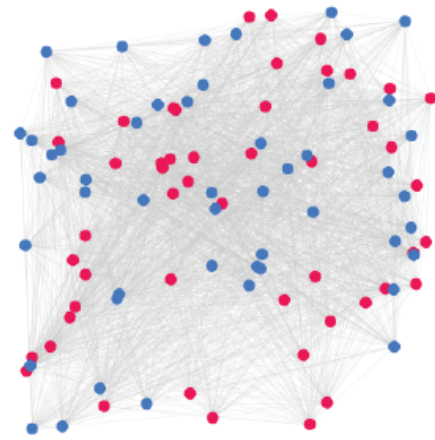*Tiago Peixoto. https://skewed.de/tiago/blog/descriptive-inferential#fig-generation*

What is VNA?

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.*
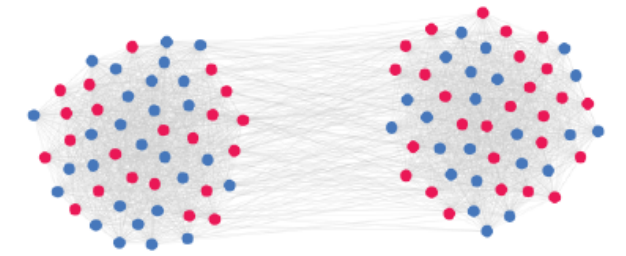
Color =
ATTRIBUTE X

No color

Is there an
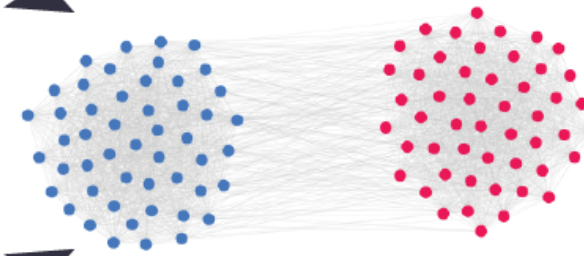attribute-topology
CORRELATION?

YES   or   NO

No layout

Layout =
TOPOLOGY
(edges)

Nodes of same
attribute X
are more often
connected

Attribute X
and edges
are unrelated
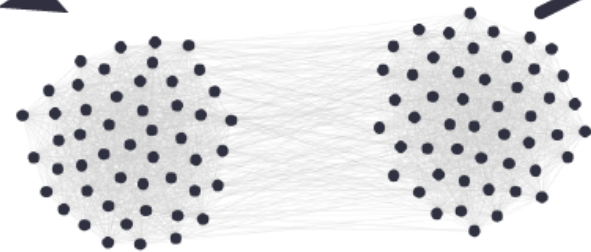
What some people do:

Obtaining groups from community detection.

Most of the time, the network
has no obvious divide.
Where are the groups?

C. Elegans network, LinLog energy model. Network from Watts, D. and Strogatz, S. (1998) 'Collective dynamics of "smallworld" networks', Nature, 393, pp. 440–442.

Most of the time, the network has no obvious divide.
Where are the groups?
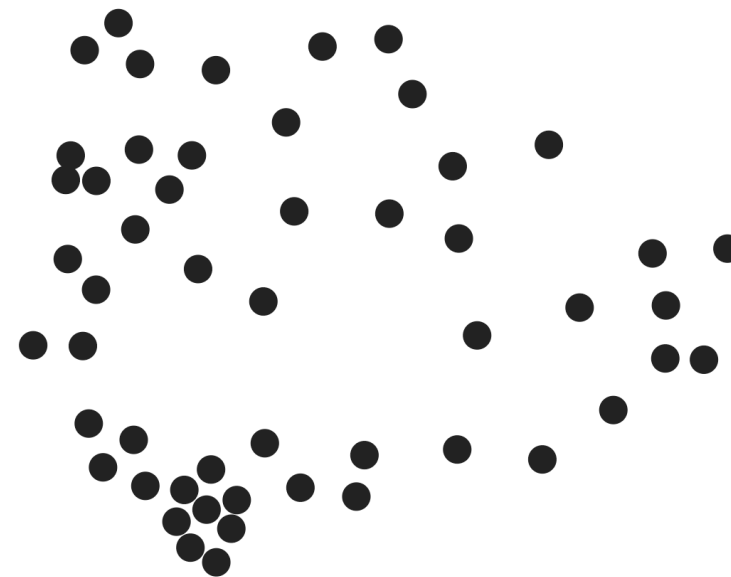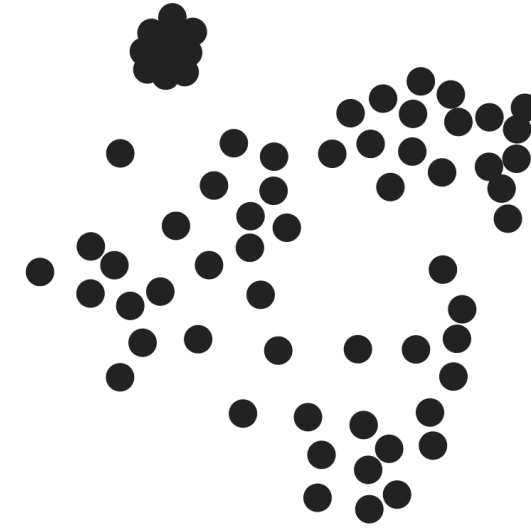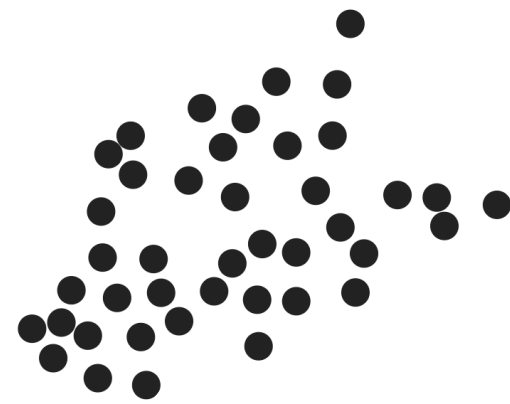
If only it were like this!

Most of the time, the network
has no obvious divide.
Where are the groups?

But it's often
like that :(

What do
we do?

We can delegate finding groups
to an algorithmic technique:
community detection
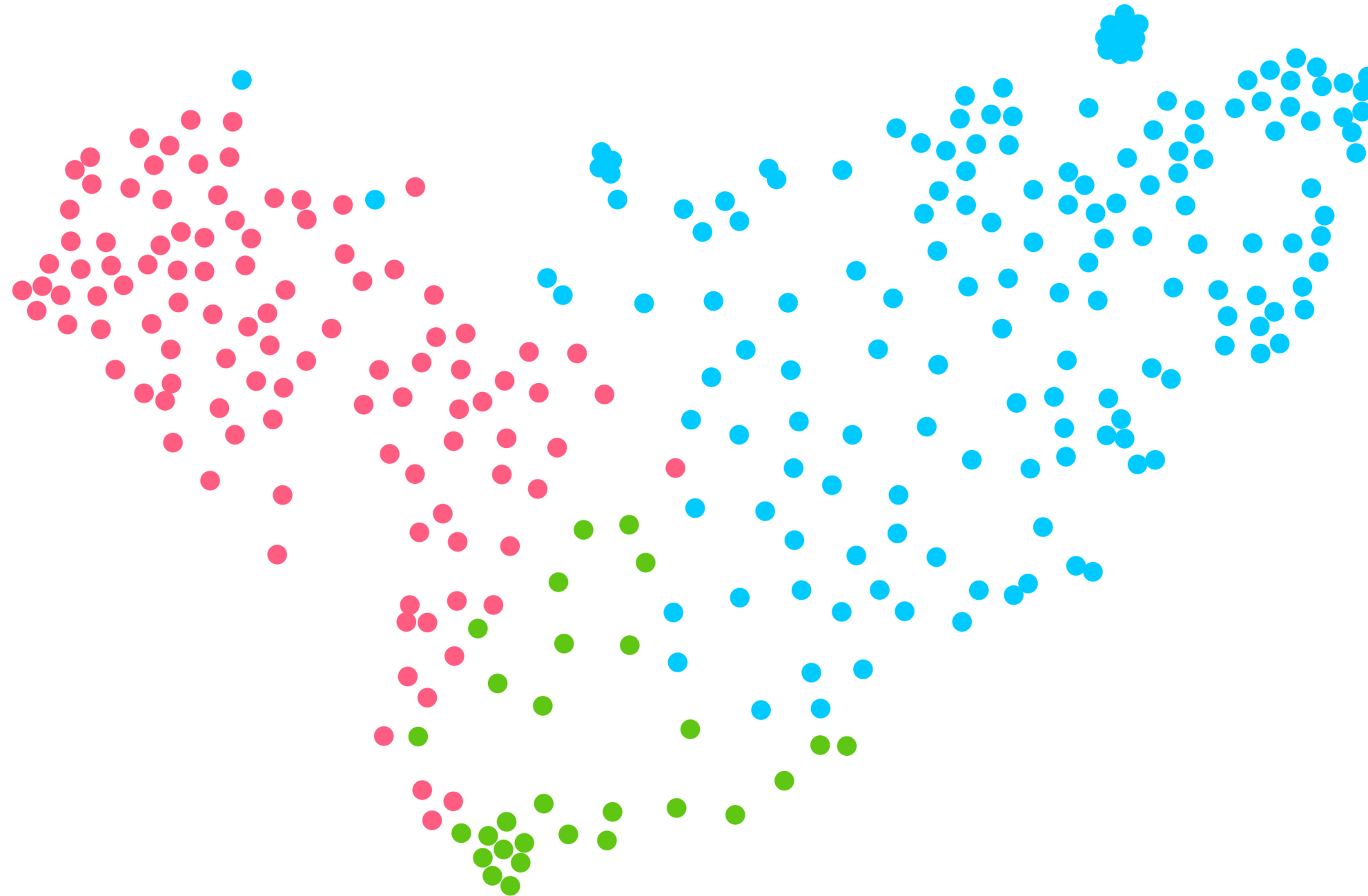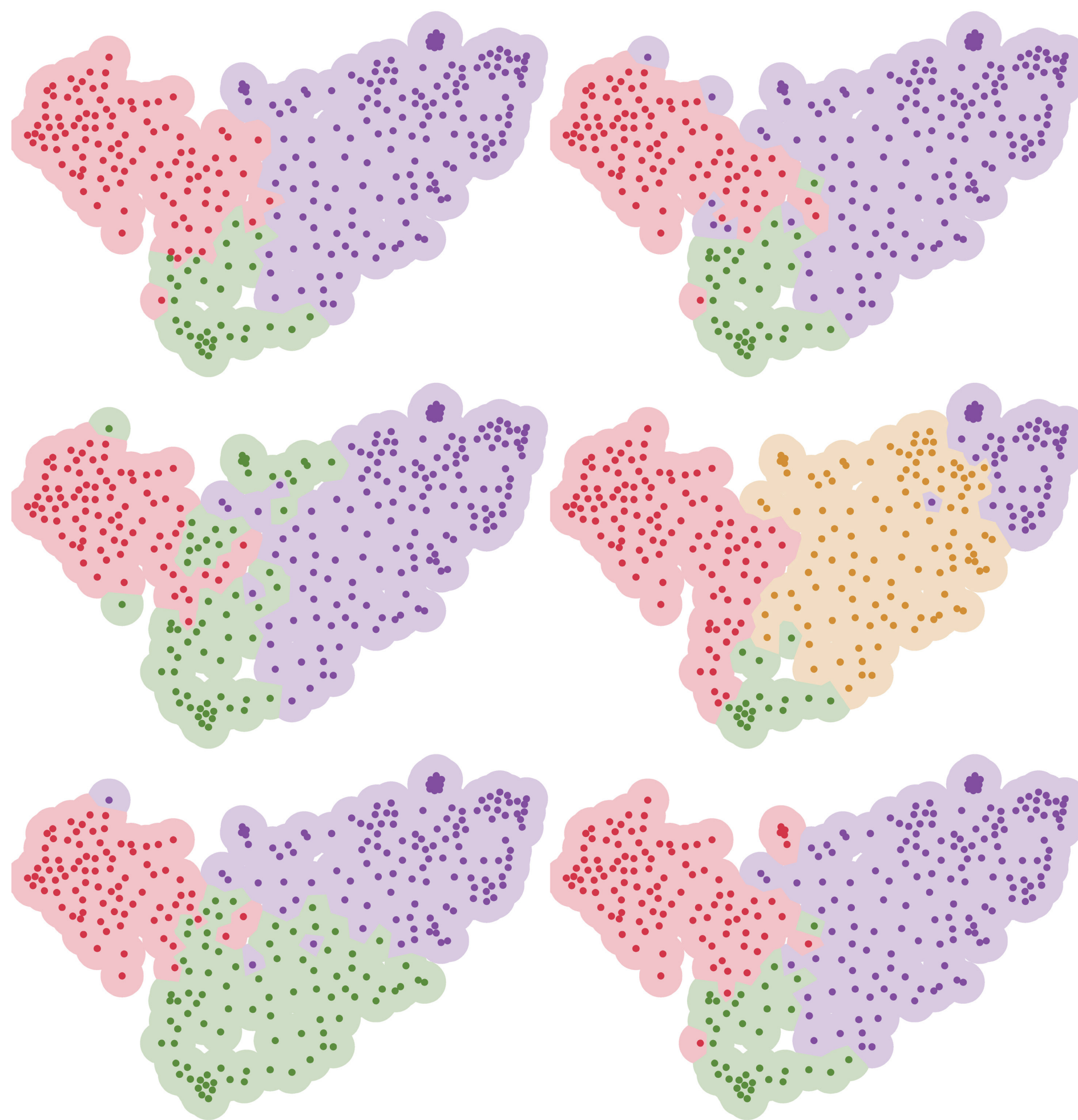
C. Elegans network, LinLog energy model. Network from Watts, D. and Strogatz, S. (1998) 'Collective dynamics of "smallworld" networks', Nature, 393, pp. 440–442.

What even counts as a group?

Peixoto, T.P. (2020) 'Revealing consensus and dissensus between network partitions', arXiv preprint arXiv:2005.13977.

Empirical data
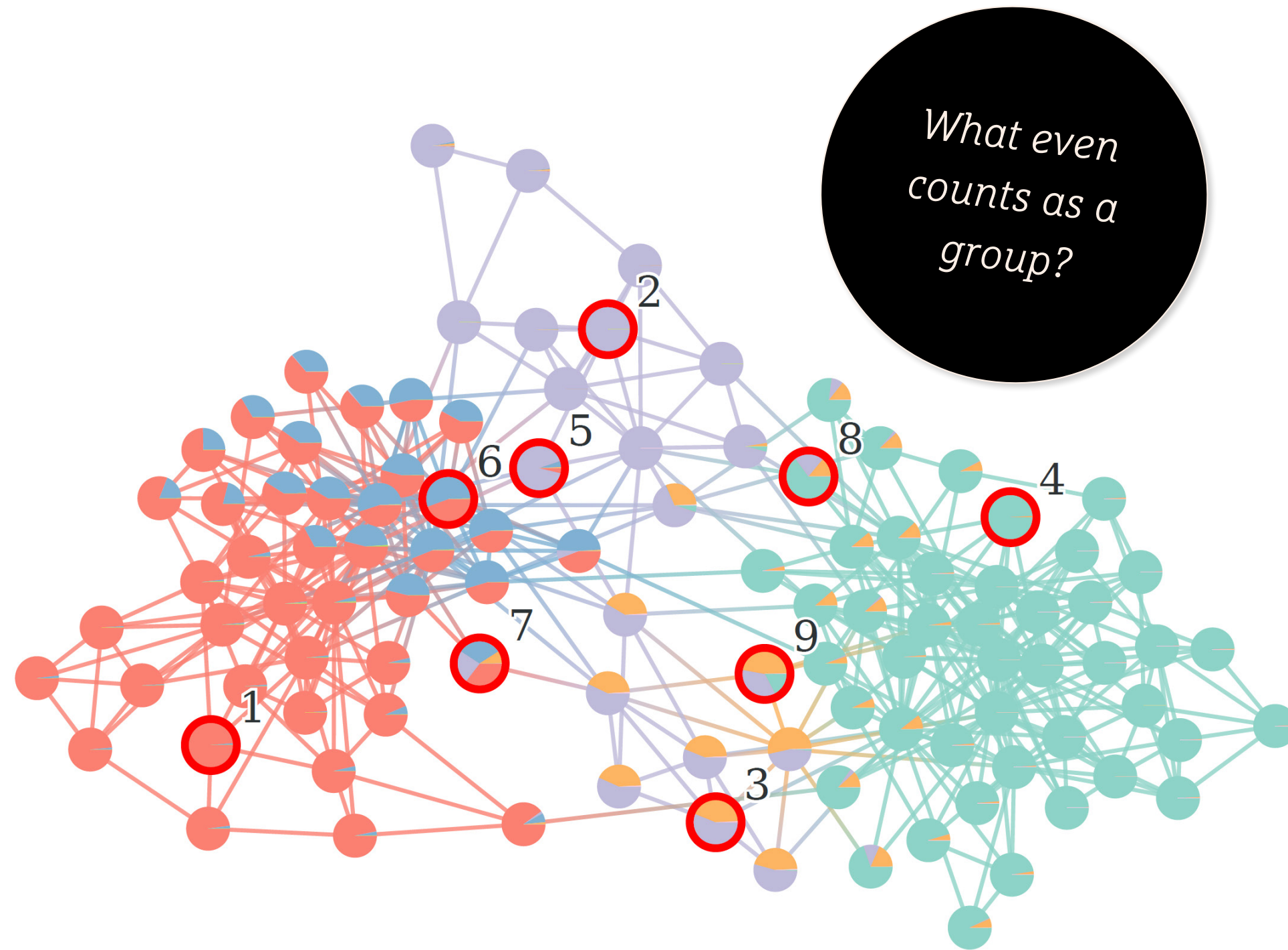
COMMUNITY
DETECTION

Set of partitions
in partial agreement

Empirical data

FORCE-
DRIVEN
LAYOUT

Visual heterogeneity

Tiago Peixoto's argument:

1. There are descriptive and inferential methods (of community detection).

2. Descriptive methods "do not articulate precisely what constitutes community structure" while inferential methods "start with an explicit definition of what constitutes community structure".
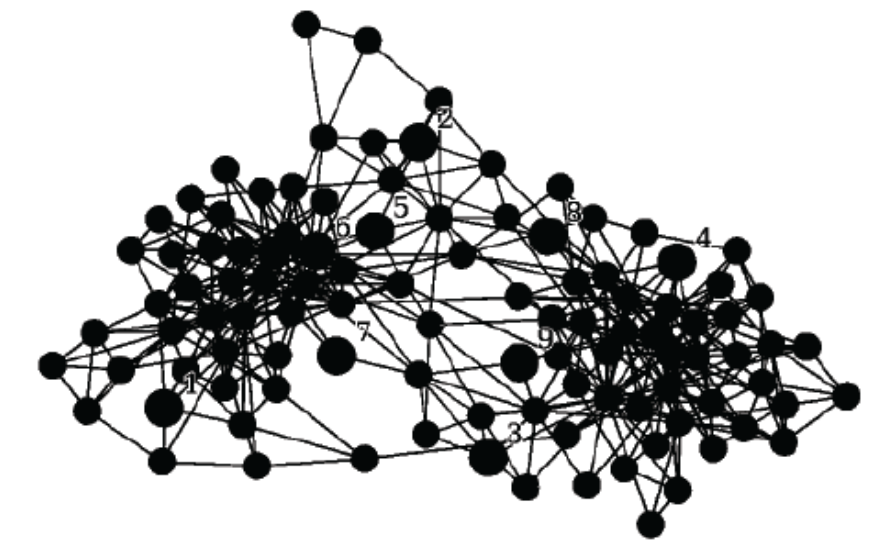
3. Even though "descriptive clustering approaches arise naturally" in a number of practical situations, they "carry no explanatory power", contrary to inferential methods.

4. The communities obtained from descriptive methods "can be seen and described, but they cannot explain". *In his argument, the power to "explain" boils down to being able to predict when nodes are connected by looking at which communities they belong to, for a given (explicit) generative process.*

5. The most-used benchmarks in the literature are in fact not suitable to descriptive methods. Inferential methods are easier to benchmark because they state the model they aim to fit.
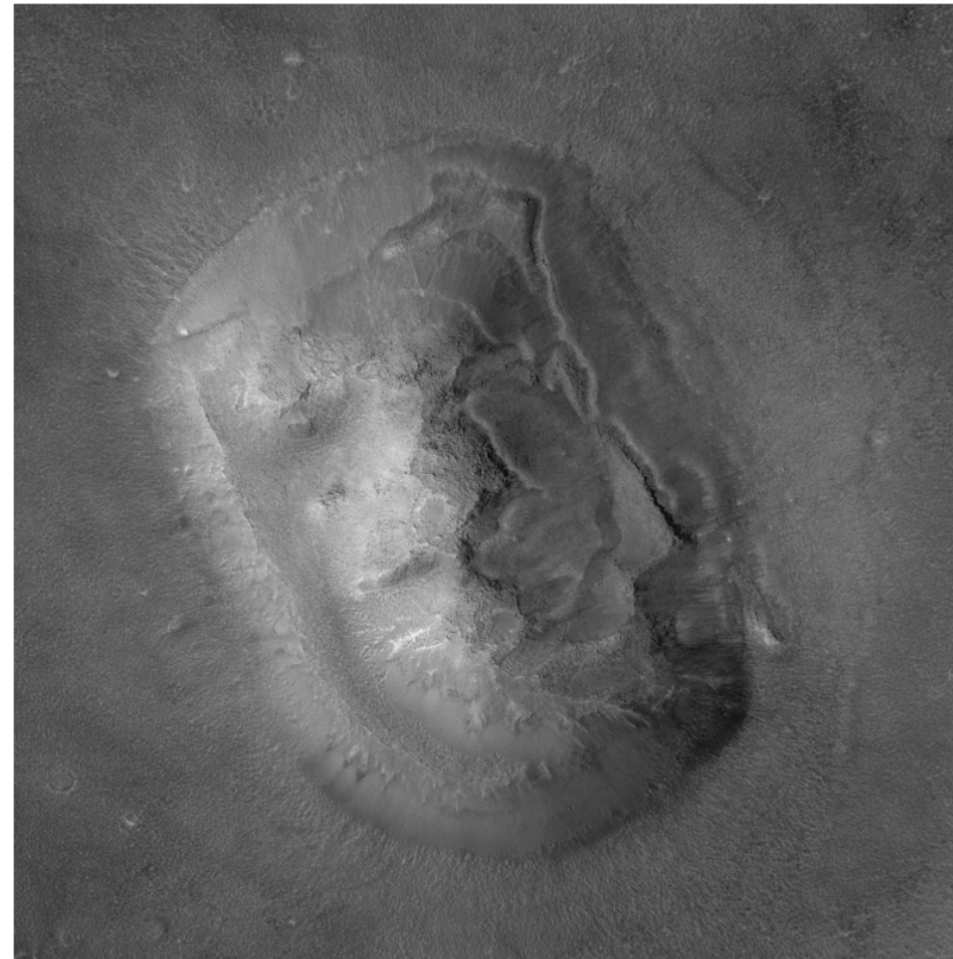
6. "every descriptive method can be mapped to an inferential one, according to some implicit model." Descriptive methods are inferential methods that do not state their model, which makes them inherently worse. "There is no such thing as a 'model-free' community detection method."

Description — A face

Eye
Nose
Mouth
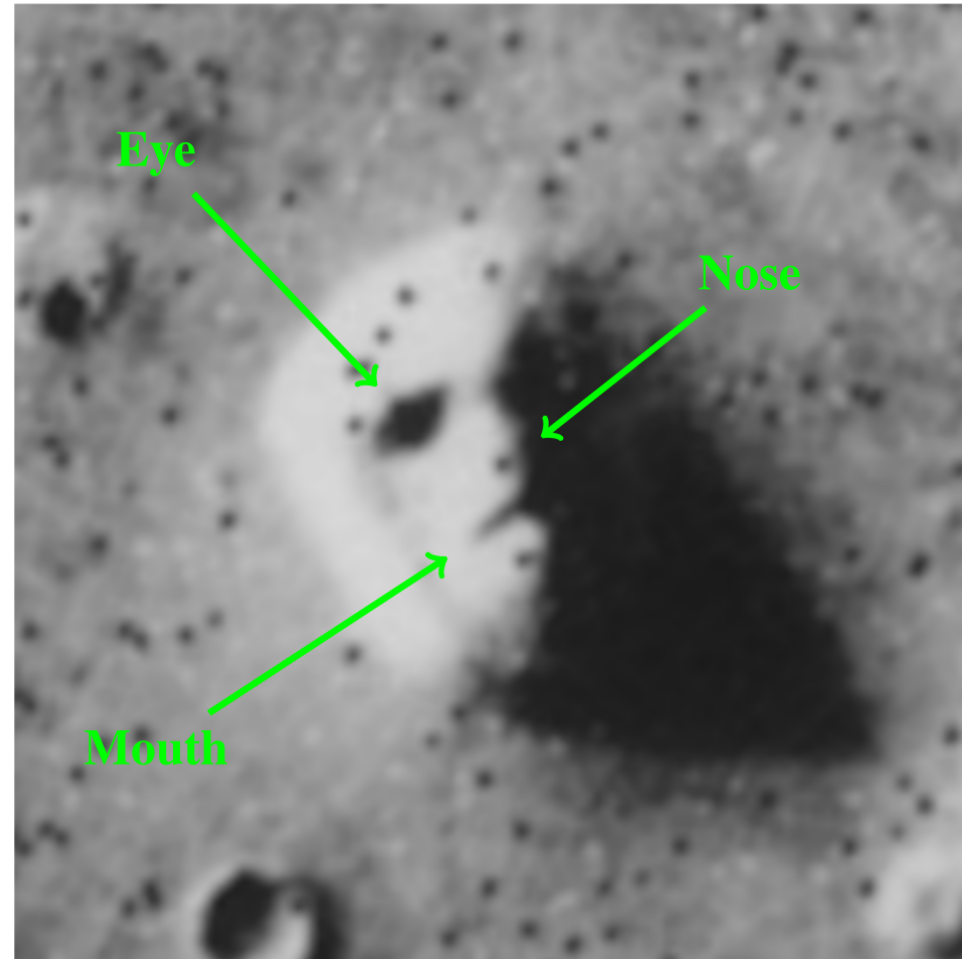
Explanation — A mountain

Bayesian inference applied to community detection:

Imagine a process that generates links given nodes and a set of rules involving groups of nodes: the probability that two nodes get linked depends on their groups. If you know the rules, you can call it a model: they generate different networks of the same model.

Bayesian inference is like a guessing game: if you have the nodes, the links, and the model (the rules), can you guess which node was in which group? Inference measures the likelihood of the guess.
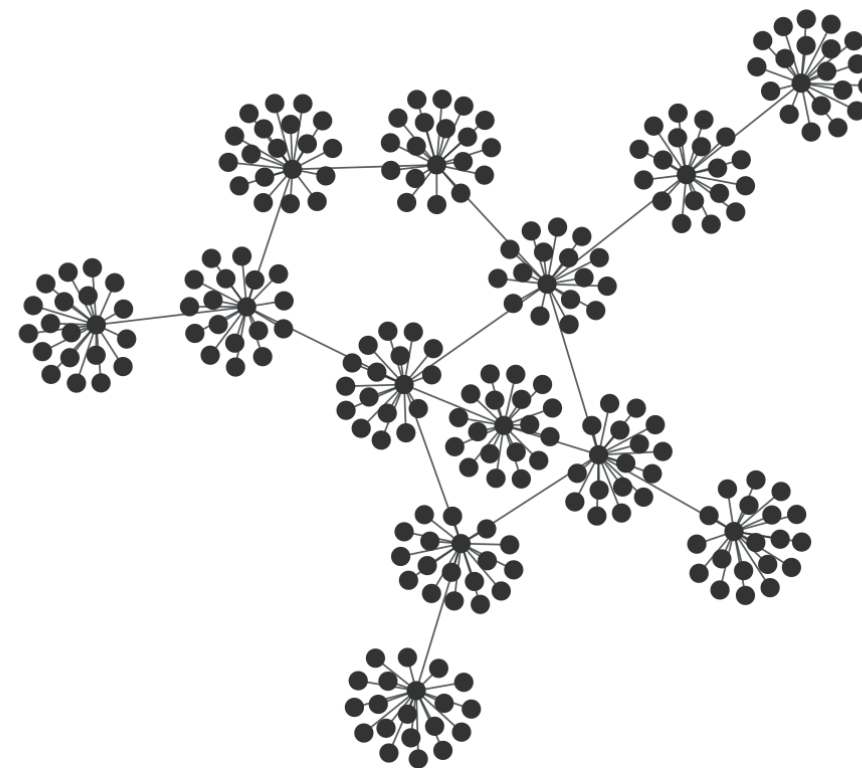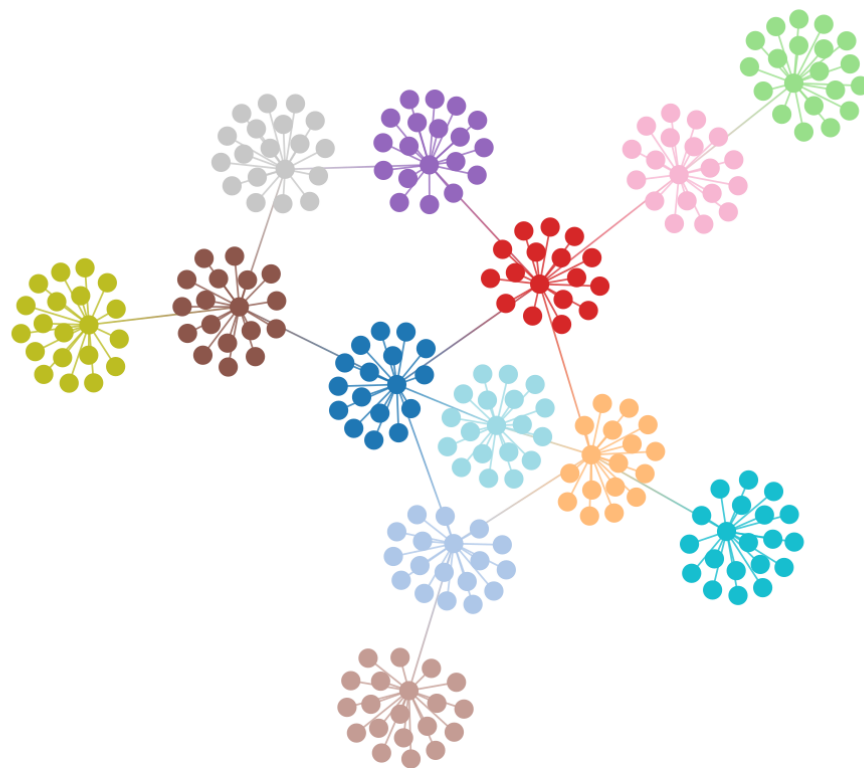
Is the inference approach superior?
- Yes, if the model is appropriate
- Yes, because you know the model
- No, if the model cannot be assumed
- No, if you do not use the model to explain
- No, if the result is unusable for your own goal.

Practice: coding the network as empirical material. One uses the coding as a means to build a discourse. This comes with constraints: not too many communities, etc.
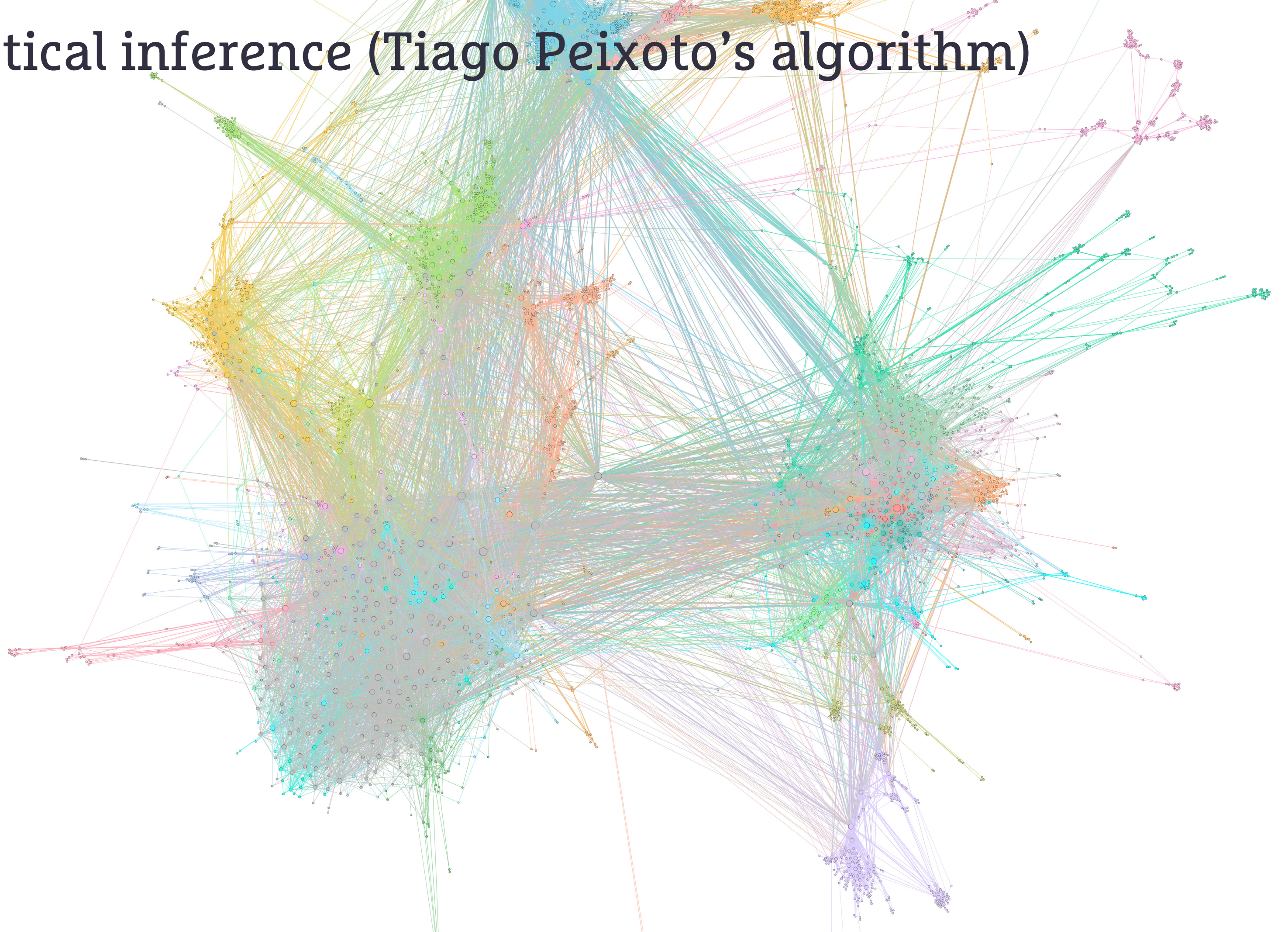
Countries (ground truth)
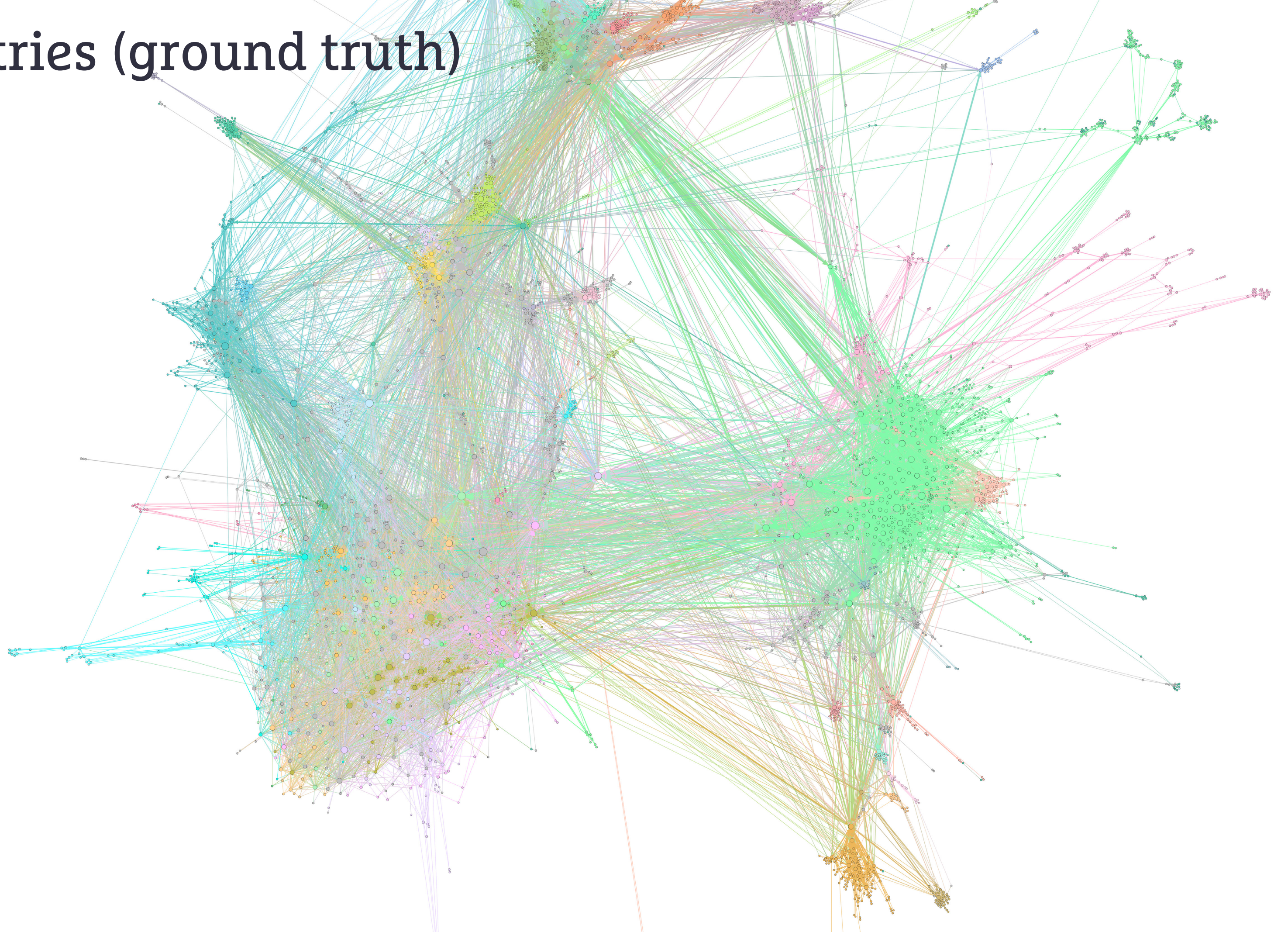
Modularity clustering ("Louvain")
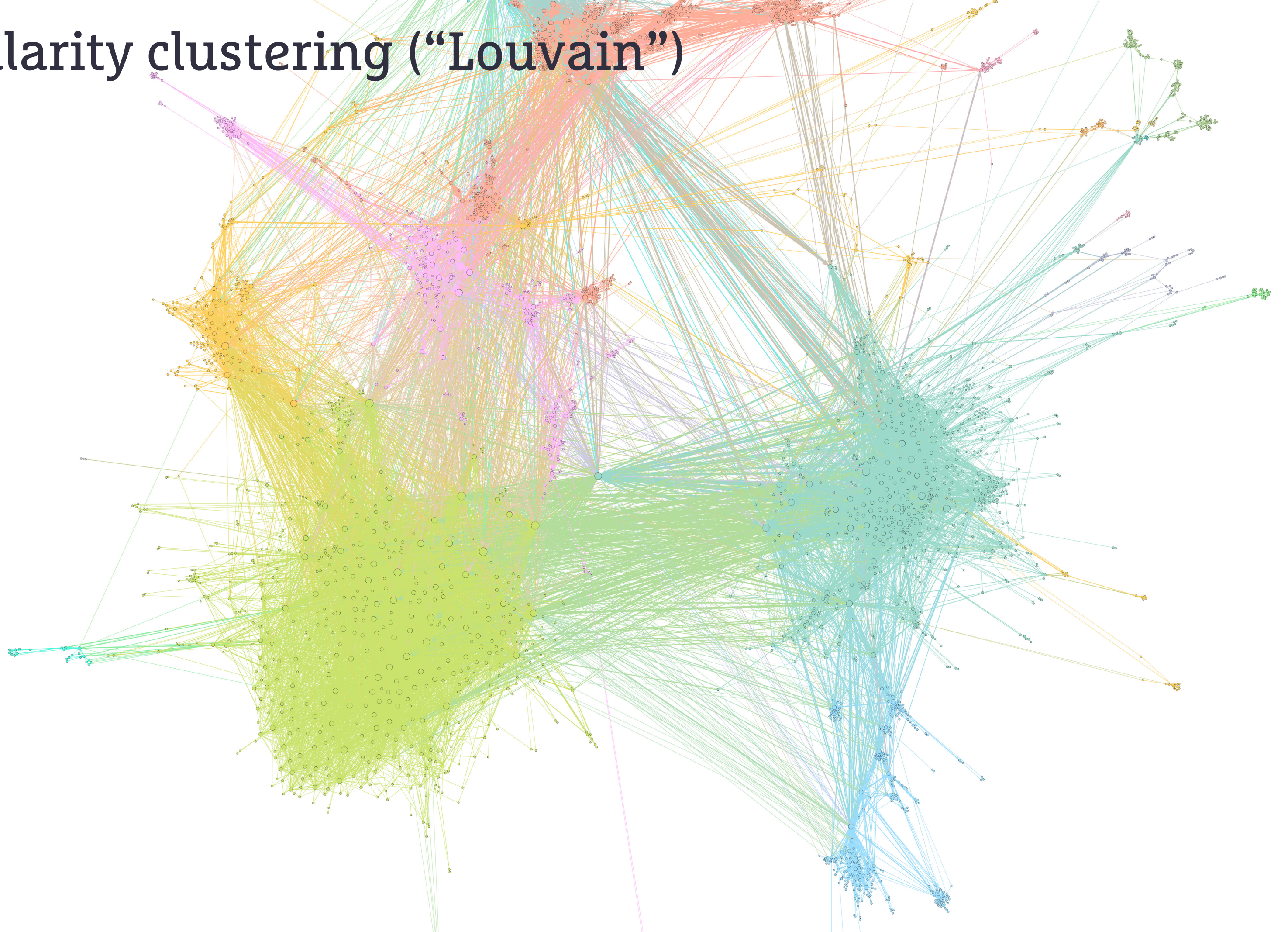
Statistical inference (Tiago Peixoto's algorithm)

Modularity clustering gives more homogeneous clusters. Bias or feature?
- For what we want to do, it is a feature
- For explanatory models, it is a bias
- Either way, it is a methodological commitment that we must know about!

Those who make algorithms do not generally know what other people will do with them. Epistemic cultures repurpose tools and methods. And those who use the algorithms do not necessarily understand them well.

Thank you for your attention.

@jacomyma
http://reticular.hypotheses.org

Computational order

GAPS

GAPS

NO GAPS

✓ AGREEMENT   ✗ DISAGREEMENT   ✓ AGREEMENT

GAPS

NO GAPS

NO GAPS

Visual order