Quantmetry
Data Science Consulting

Graph Theory

… looking for communities …

Aurélia Nègre
Data Scientist
anegre@quantmetry.com

Alberto Guggiola
Data Scientist
aguggiola@quantmetry.com

Berlin Buzzwords
13th June 2017

Who are we?

Quantmetry
Data Science Consulting

- 40 Data Scientists
- From proofs of concept to production
- Fraud detection, predictive maintenance, customer insights ...

Aurélia Nègre & Alberto Guggiola

© Quantmetry 2017 | Diffusion interdite sans accord

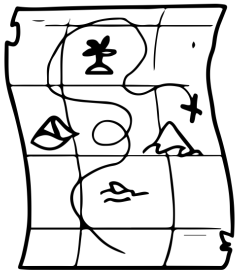
A graph: a structure made up of nodes and links

Social network

Transportation network

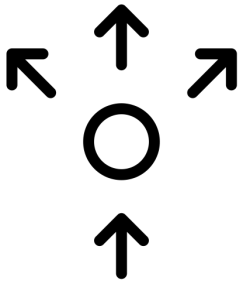
https://fr.wikipedia.org/wiki/Analyse_des_r%C3%A9seaux_sociaux

Some use cases of graph theory



Finding the best route

- Transportation systems
- Social networks

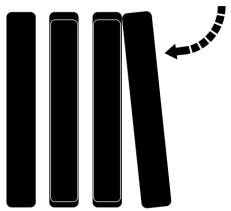


Spreading

- Viral marketing
- Vaccination campaigns

Structural importance

- Leaders?
- Google PageRank



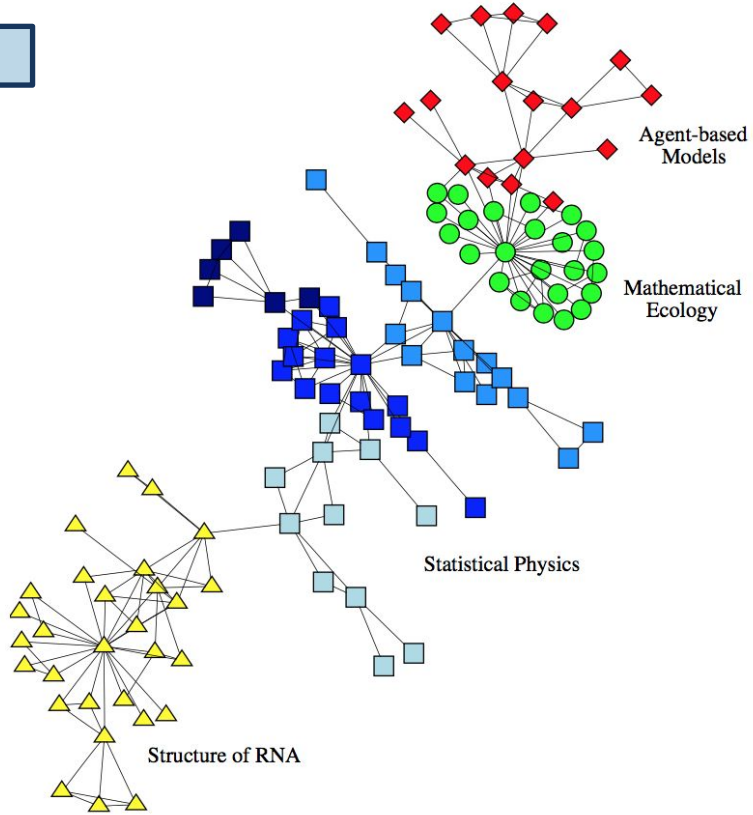
Domino effects

- Security systems
- Economics
- Infrastructures

© Quantmetry 2017 | Diffusion interdite sans accord

Community detection: looking for a structure

Co-authorship network

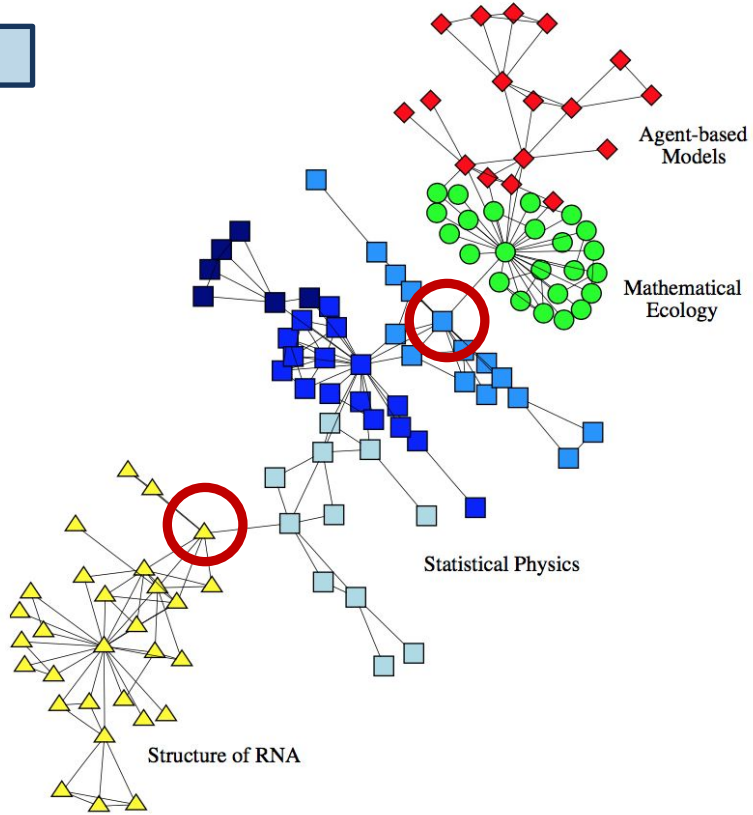


Girvan & Newman, 2002

© Quantmetry 2017 | Diffusion interdite sans accord

Community detection: looking for a structure

Co-authorship network



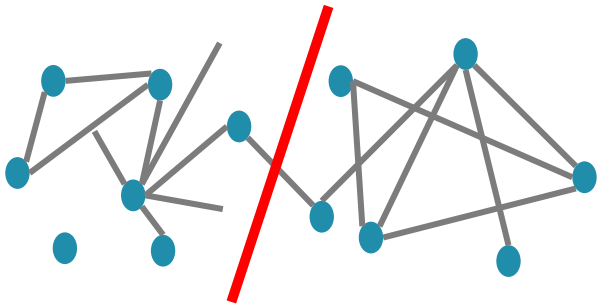
Girvan & Newman, 2002

Where are the « bridges »?

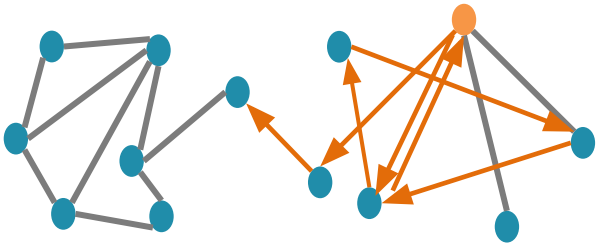
© Quantmetry 2017 | Diffusion interdite sans accord

Two approaches for finding clusters

Cutting the bridges



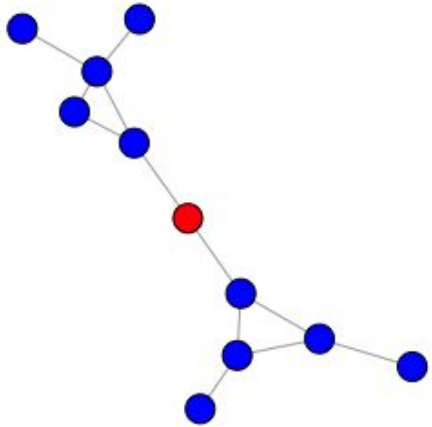
Gathering the most connected elements



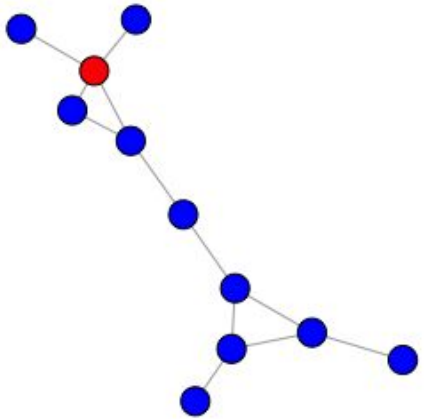
© Quantmetry 2017 | Diffusion interdite sans accord

Different ways of measuring nodes importance

A global importance:
the betweenness centrality



A local importance:
the degree



Several alternatives, including Google PageRank

Tools

Non distributed analytical libraries

NetworkX

NetworkX Home | Documentation | Download | Developer (Github)

High-productivity software for complex networks

graph-tool | Efficient network analysis

igraph

Distributed analytical libraries

GraphX

Databases

neo4j

© Quantmetry 2017 | Diffusion interdite sans accord

Demo Time using LinkedIn data

© Quantmetry 2017 | Diffusion interdite sans accord

Q&A