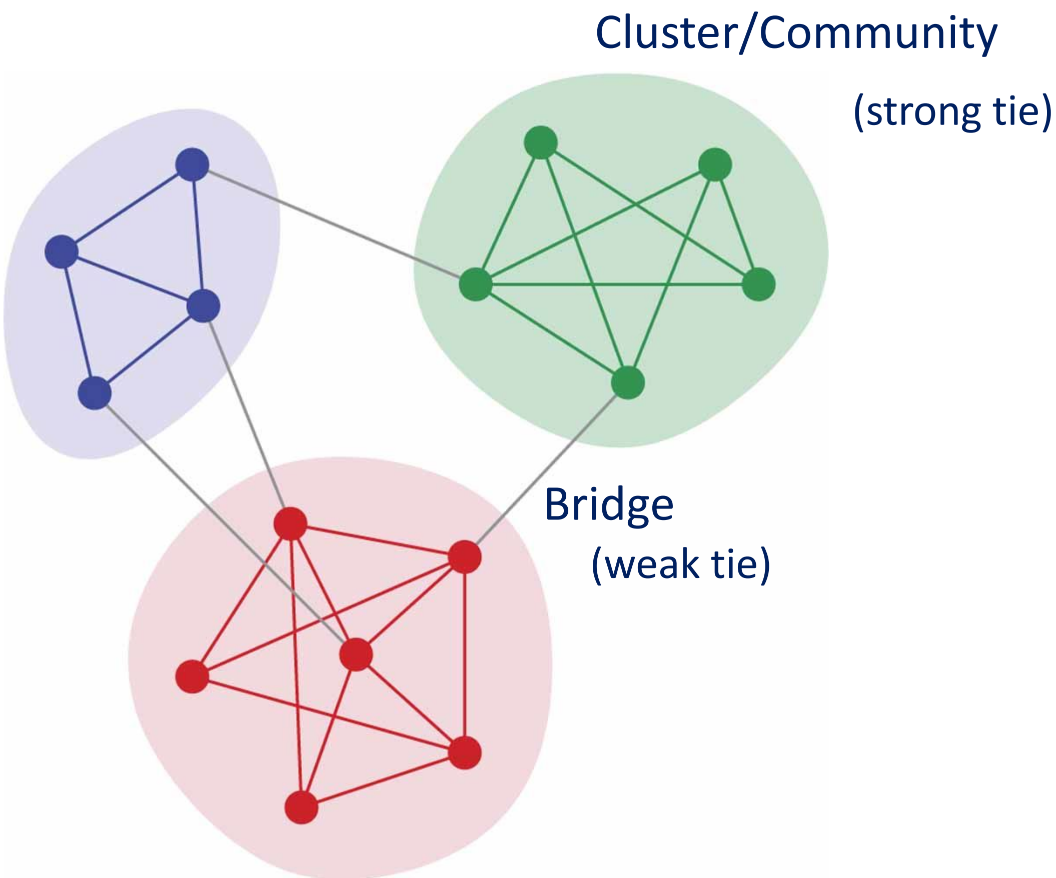
Social Network Analysis

Community detection

Network communities

Conceptual picture of a network



- We often think of networks looking like this
- But, where does this idea come from?

MIME.

Granovetter's explanation

Granovetter, The strength of weak ties [1973]
https://www.jstor.org/stable/pdf/2776392.pdf

Q: How do people discovered their new jobs?

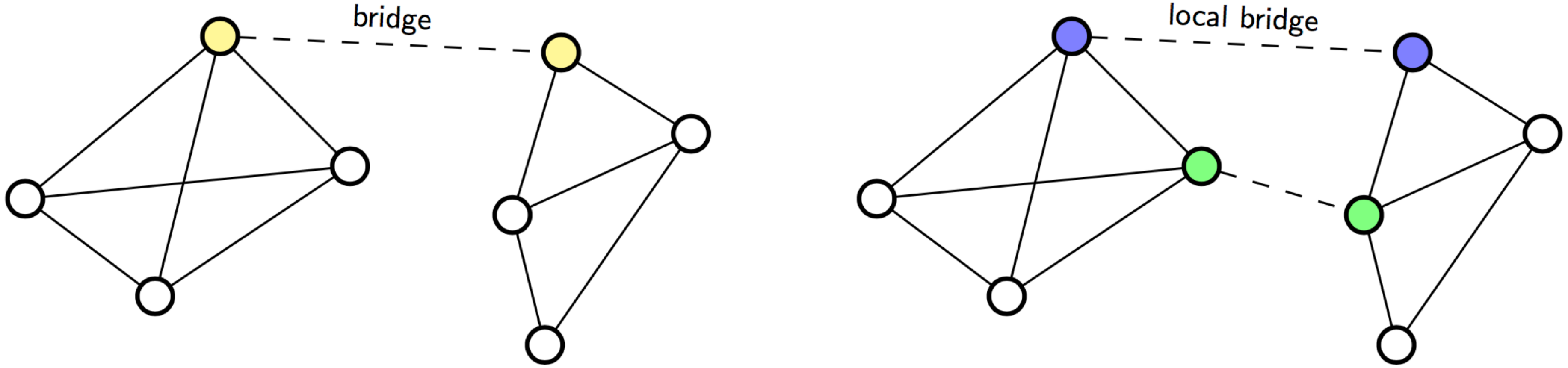
A: Through personal contacts, and mainly through acquaintances rather than through close friends

Remark: Good jobs are a scarce resource

Conclusion:

- Structurally embedded edges are also socially strong, but are heavily redundant in terms of information access
- Long-range edges spanning different parts of the network are socially weak, but allow you to gather information from different parts of the network (and get a job)

Local bridges



- An edge (*i*,*j*) is a **bridge** if deleting it *i* and *j* fall into different components

  this is extremely rare, e.g., because of small world properties

- An edge (*i*,*j*) is a **local bridge** if, by deleting it, *i* and *j* have a span (distance) greater than 2, i.e., if *i* and *j* **do not have friends in common**

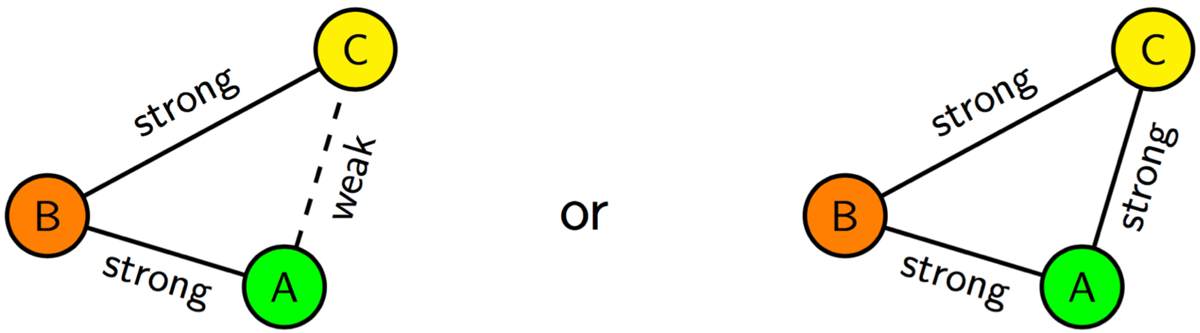
  common friends imply belonging to <u>a triadic closure</u>

Strong triadic closure

Assume two categories of edges:

- Strong ties (close friends)
- Weak ties (acquaintances)

Remark. If node B is strongly tied with A and C, then A and C are very likely to be connected (either weakly or strongly), that is



Strong triadic closure property – If a generic node B is strongly tied with A and C, then A and C are connected (either weakly or strongly)

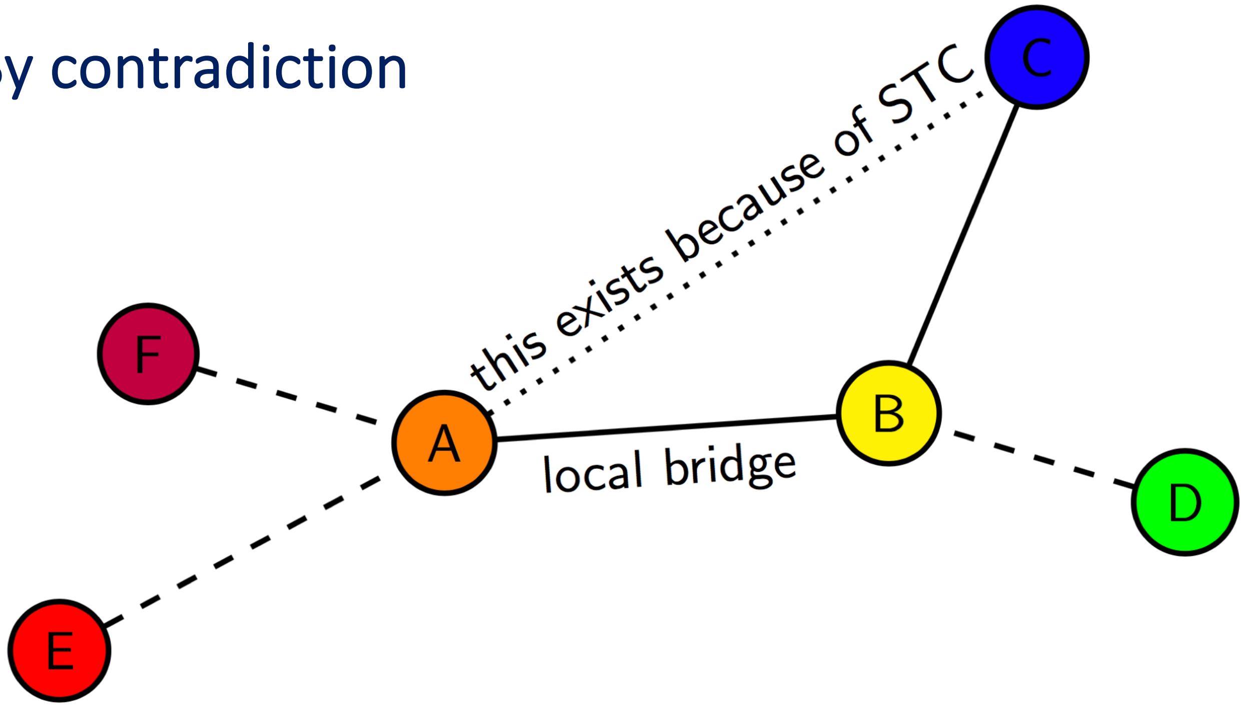
Granovetter's claim

Claim:

- Under the strong triadic closure property, local bridges are weak ties (if at least one of their nodes belongs to at least two strong ties)

Proof:

- By contradiction



Community detection

- Granovetter's theory suggests that networks are composed of tightly connected sets of nodes (i.e., communities), loosely connected between them
- We want to be able to automatically find such densely connected group of nodes
- Applications in

  Social networks

  Functional brain networks in neuroscience

  Scientific interactions

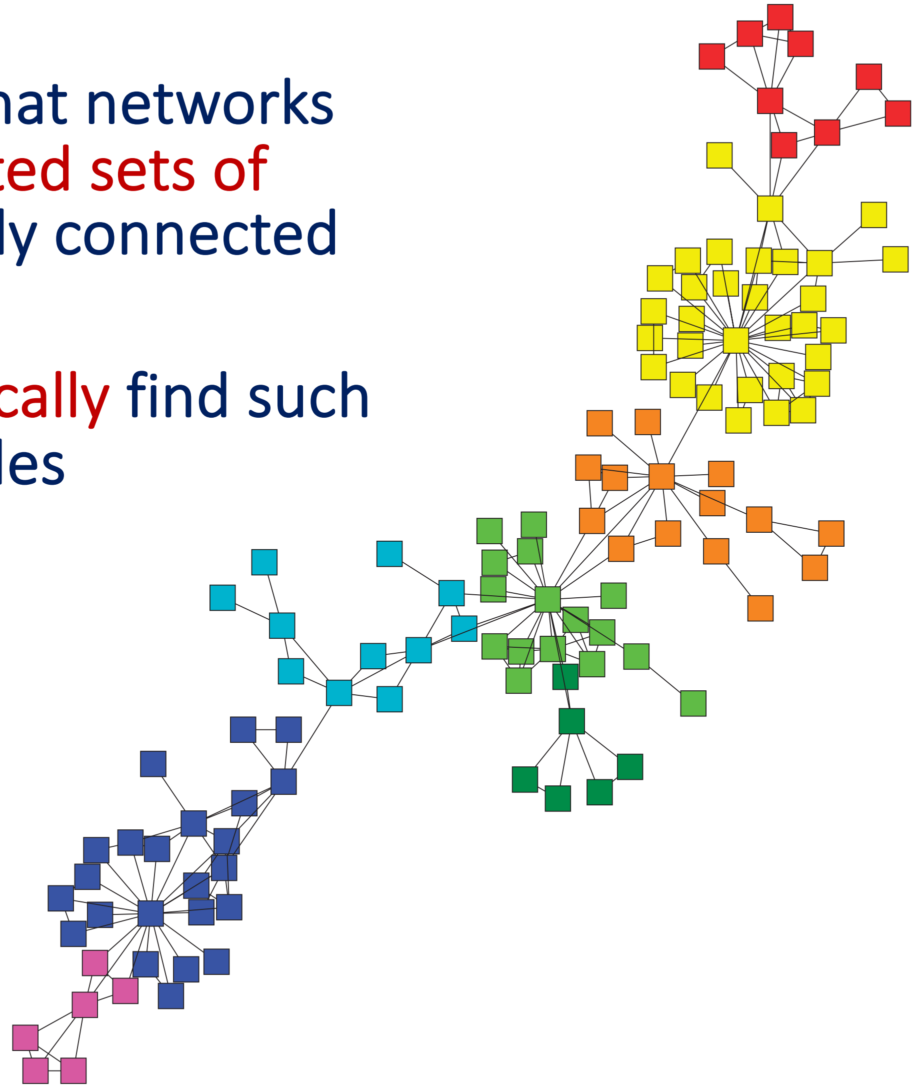
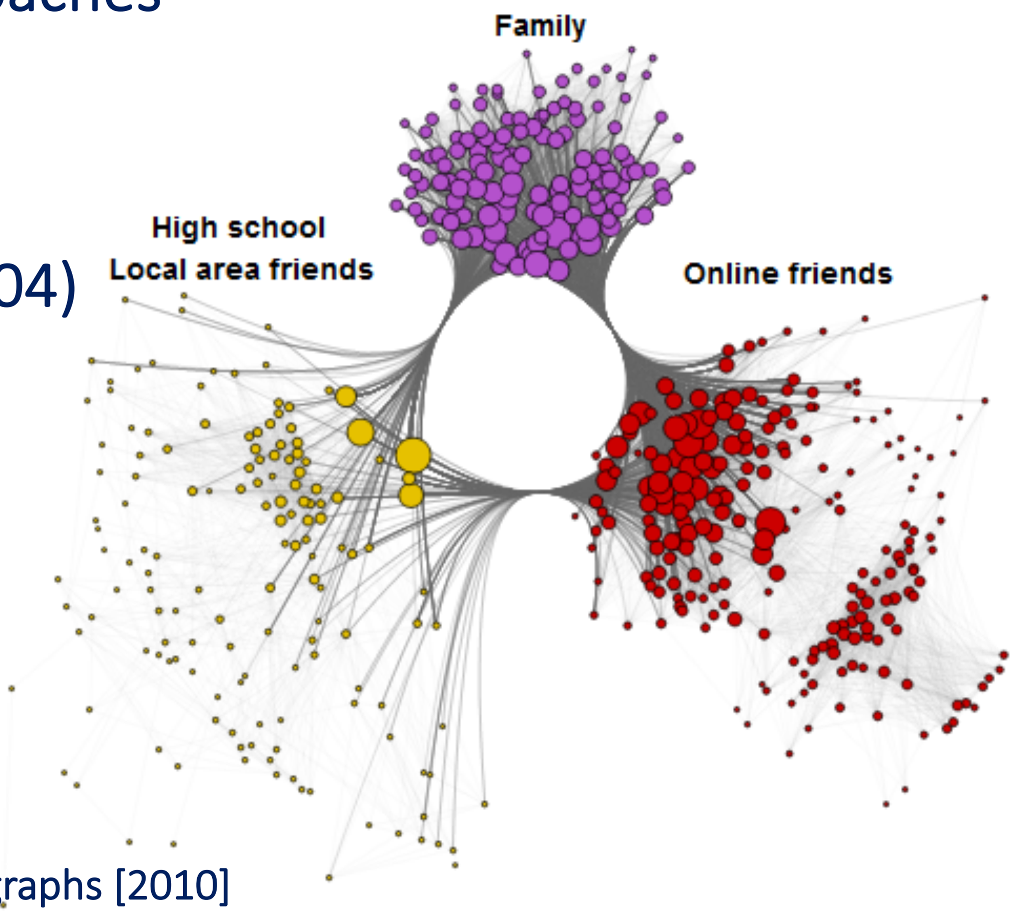


Figure 2 | A network of collaborations among scientists at a research

Community detection

Some relevant algorithms/approaches

- ❑ Dendrograms
- ❑ Girvan-Newman (2001)
- ❑ Modularity optimization (2004)
- ❑ Spectral clustering (2002)



Find a complete list in:

Fortunato, Community detection in graphs [2010]

https://www.sciencedirect.com/science/article/pii/S0370157309002841

Overlapping communities

Lescovec, Lang, Dasgupta, Mahoney, 2008

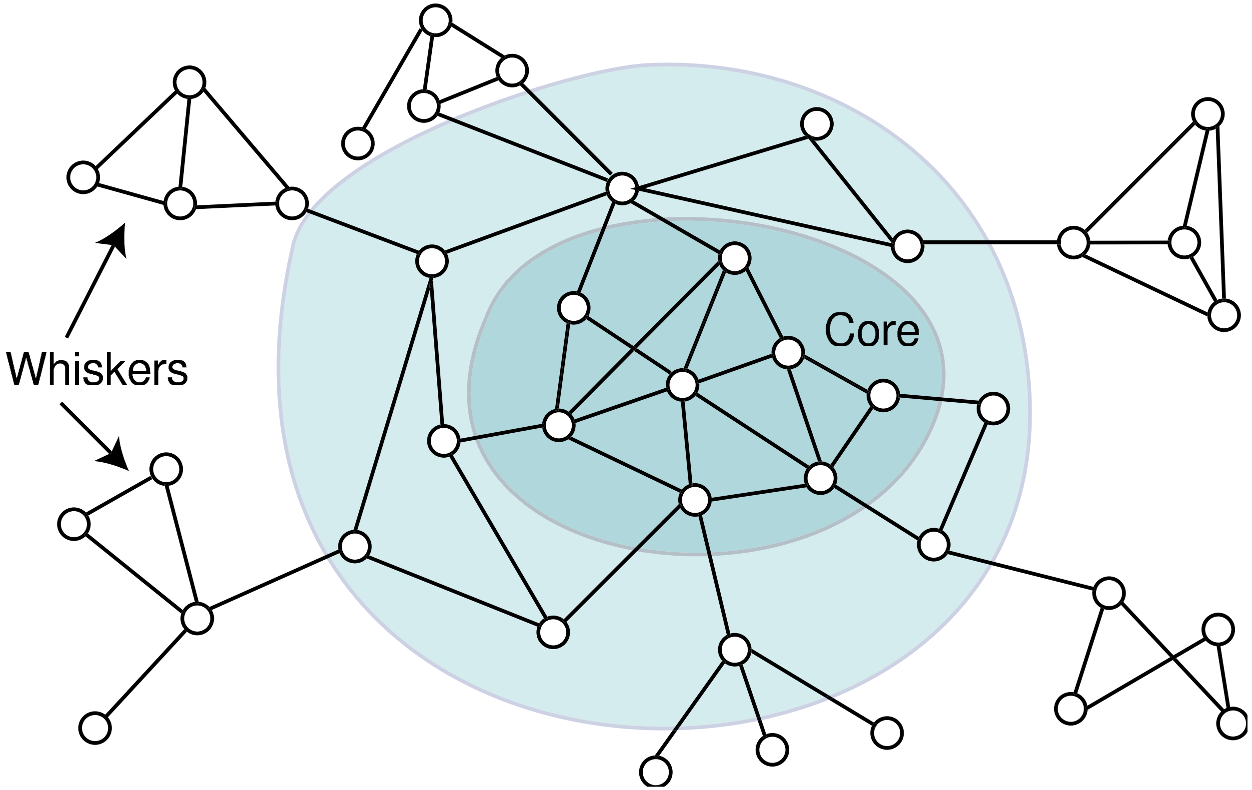
Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters

https://arxiv.org/abs/0810.1355

MIME.

The core-periphery model

Small, peripheral clusters



Caricature of network structure

Can we find a justification for this?

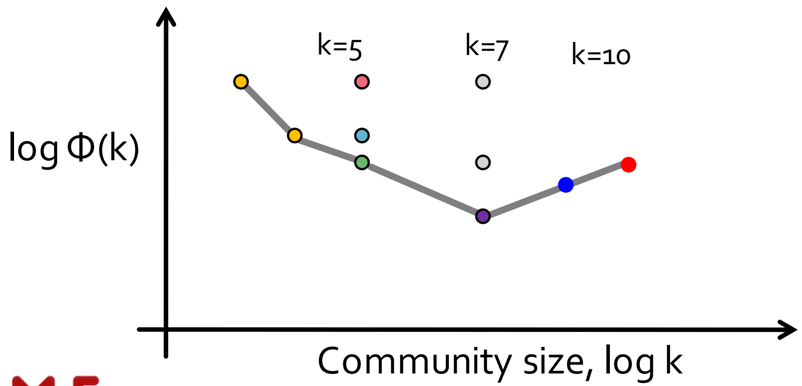
Network community profile

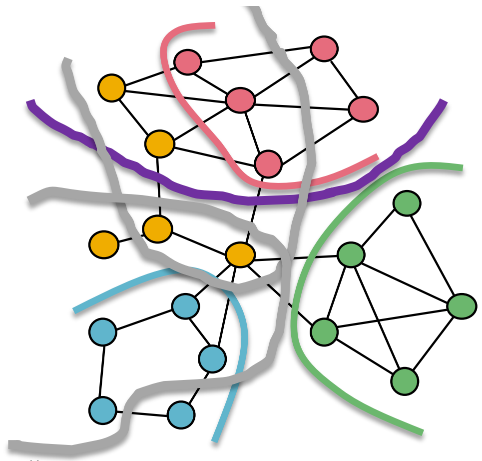
Conductance $\phi(S)$ – a metric for clusters

- S is a good cluster if it has many edges internally and few pointing outside

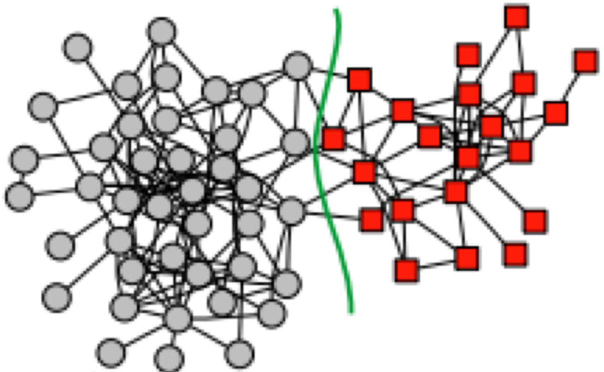
Network community profile – a metric for networks

- $\boldsymbol{\Phi}(k) = \min_{|S|=k} \phi(S)$
- Shows the best score for communities of order k

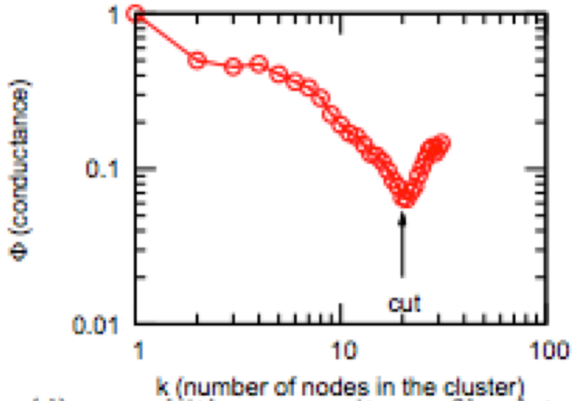




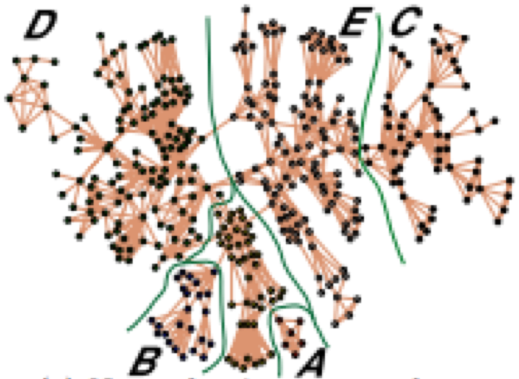
Examples



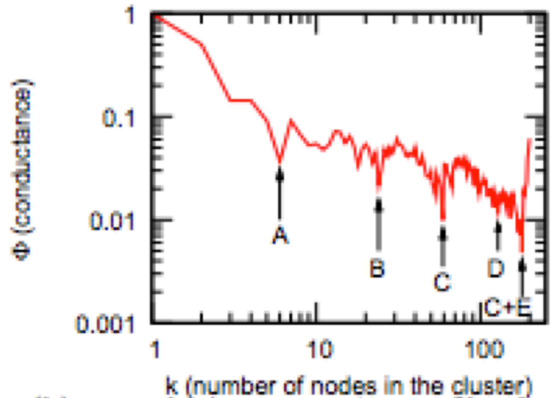
(c) Dolphins social network . . .



(d) . . . and it's community profile plot

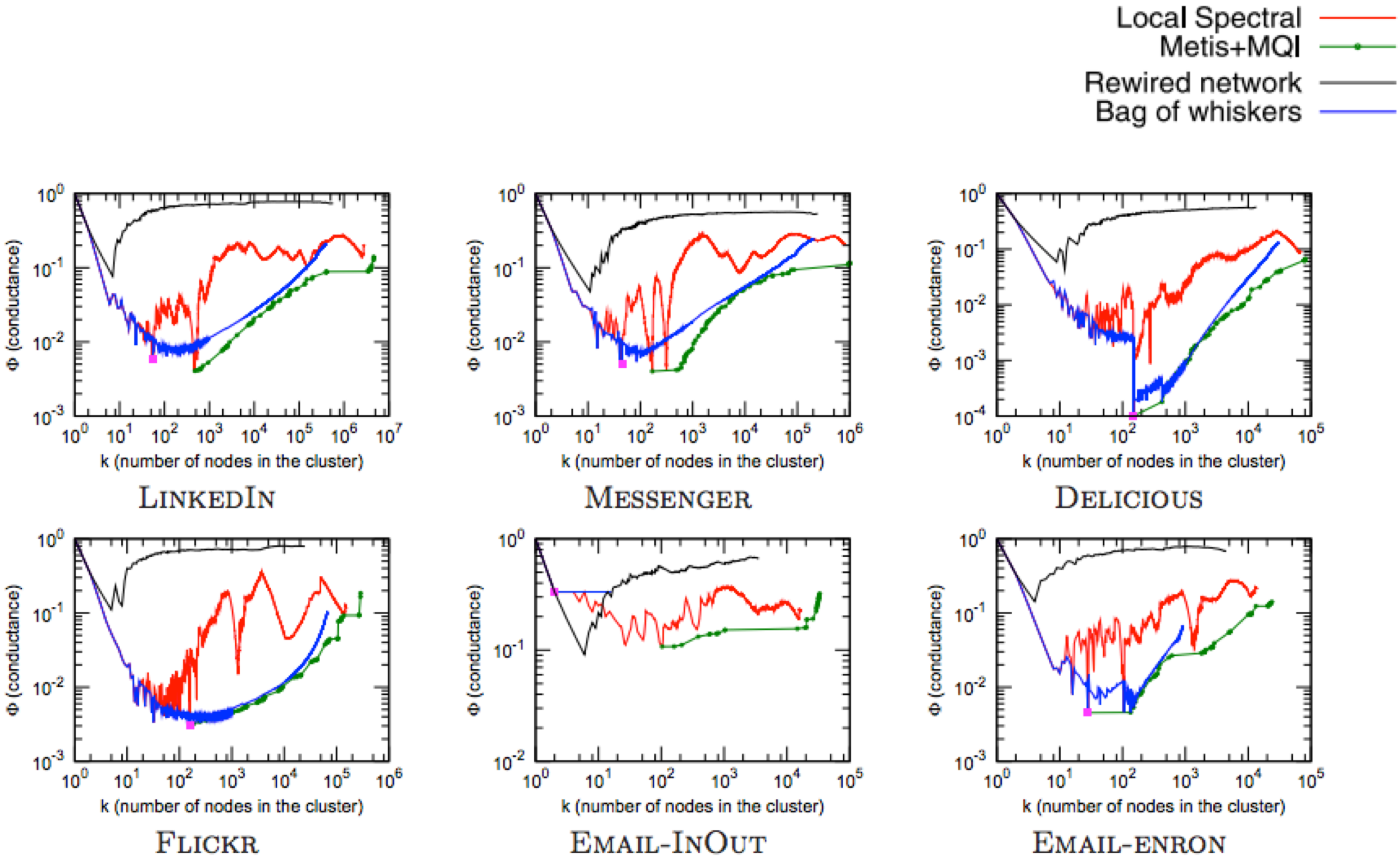


(g) Network science network . . .

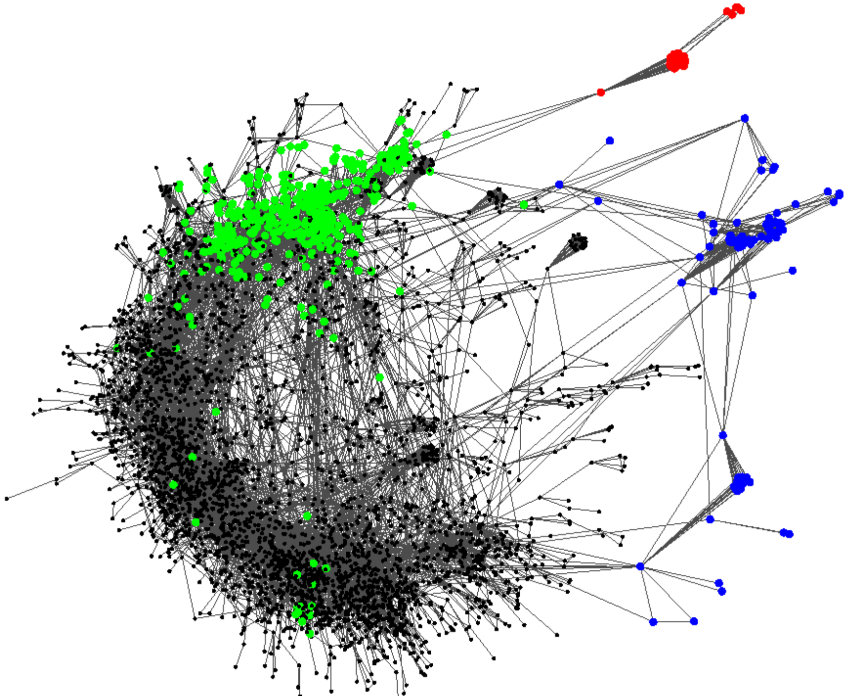


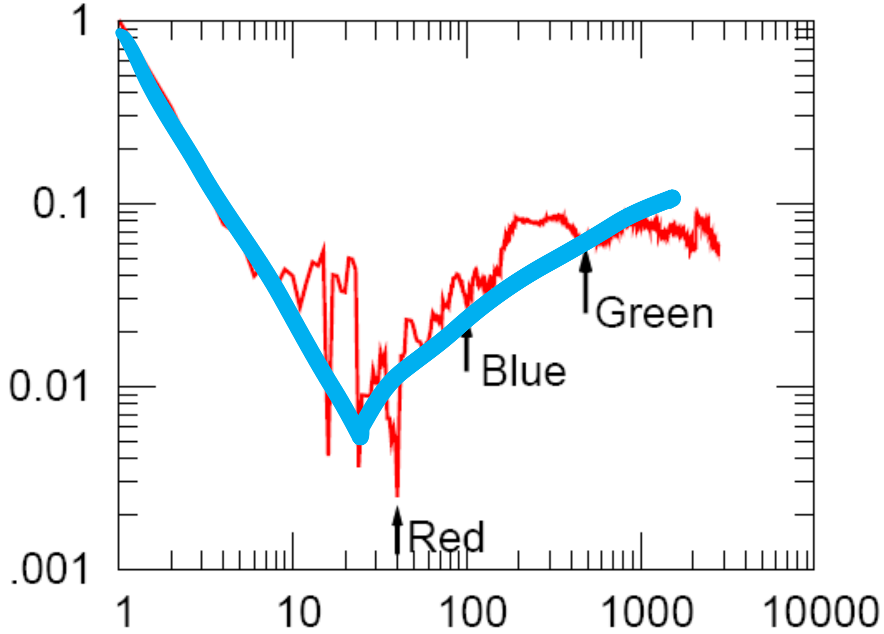
(h) . . . and it's community profile plot

Social network examples



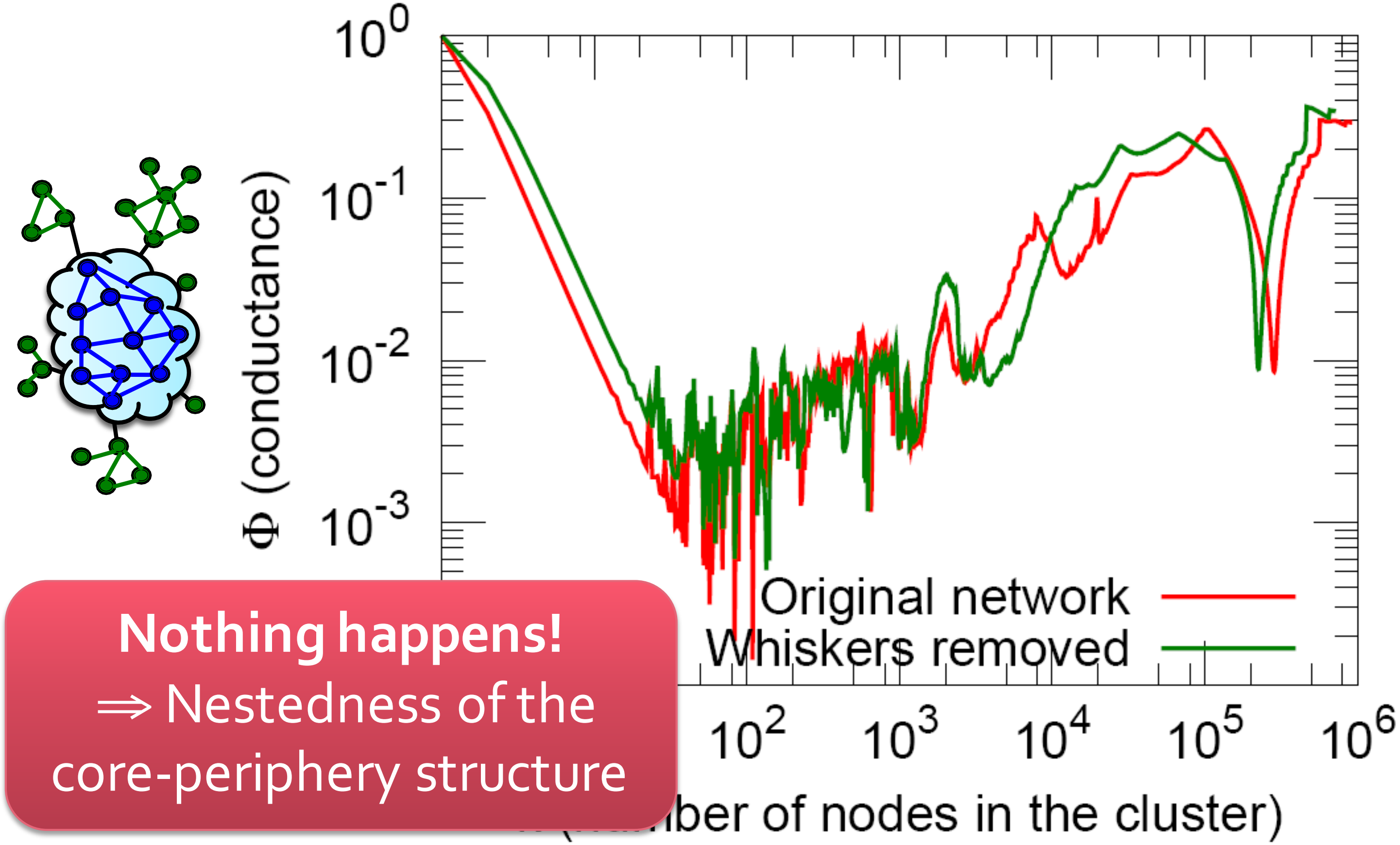
V shape of NCP





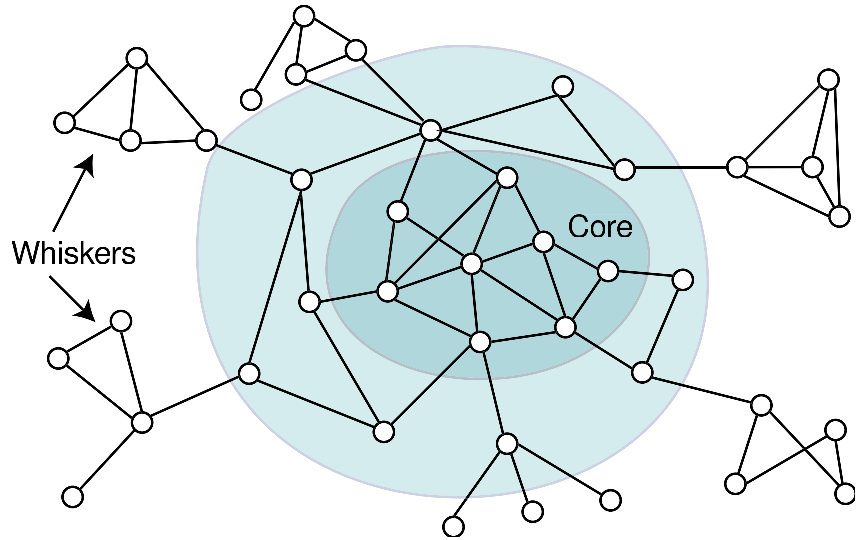
- Dips in the graph correspond to the good clusters
- Slope corresponds to the dimensionality of the network
- The V shape is common in large (social) networks
- Best clusters have about 100 nodes
- Large clusters get worse and worse performance

What if we remove good clusters?



Nothing happens!
⇒ Nestedness of the core-periphery structure

Overlapping communities model

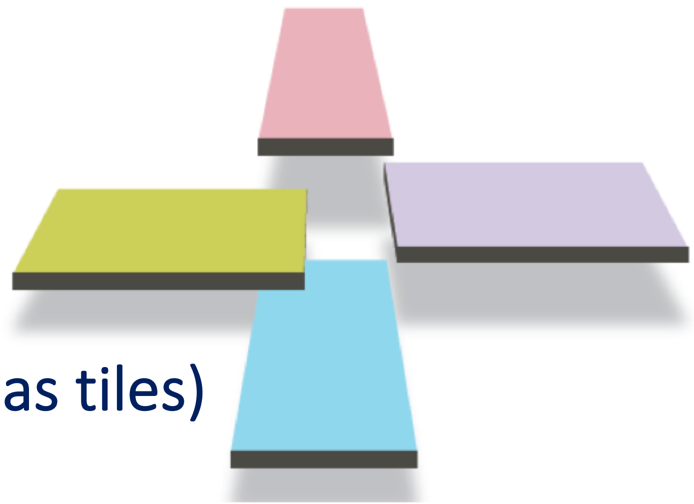


Wiskers

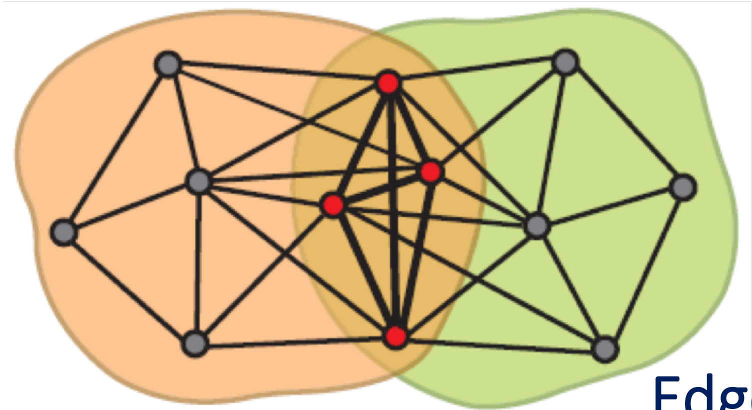
- ❑ are typically of size 100
- ❑ are responsible of good communities

Core

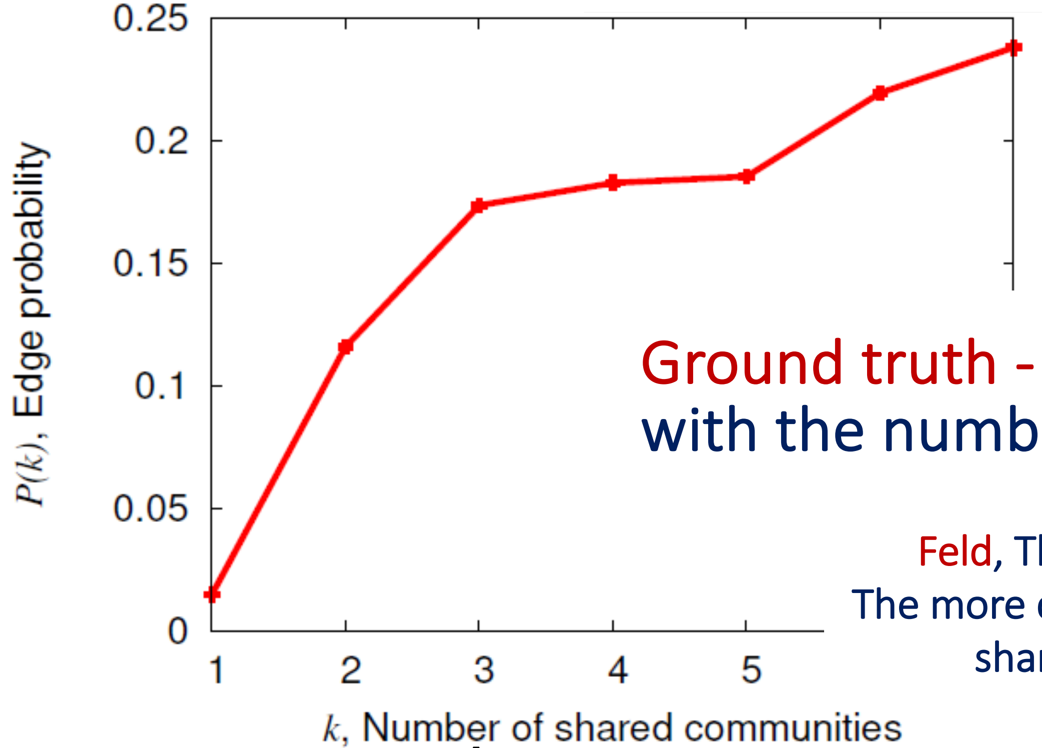
- ❑ denser and denser region
- ❑ contains 60% nodes and 80% edges
- ❑ a region where communities overlap (as tiles)



Overlapping communities model



Edge density is bigger in the overlap



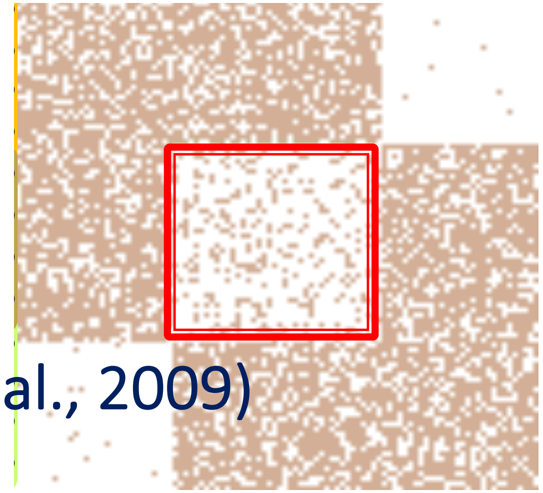
Amazon

Ground truth - Edge probability increases with the number of shared communities

Feld, The focused organization of social ties, [1981]
The more different foci (communities) that two individuals share, the more likely is that they will be tied

Overlapping communities model

most assume a wrong overlapping model !

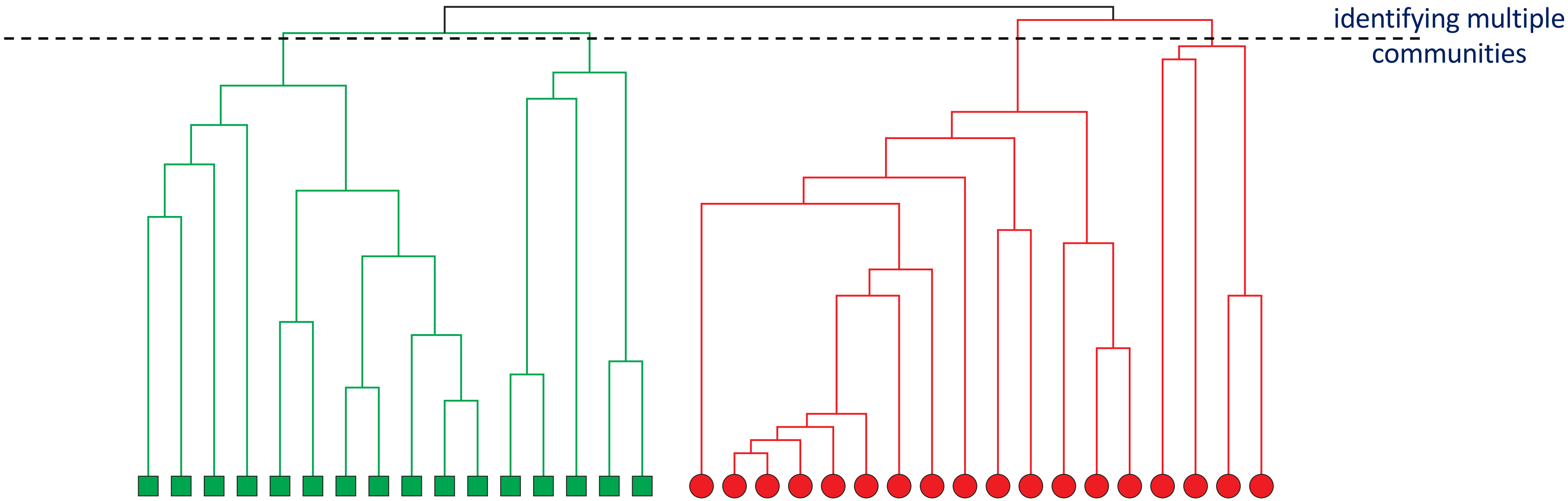


Available algorithms

- Clique percolation (Palla et al., 2005)
- Link clustering (Ahn et al., 2010) (Evans et al., 2009)
- Clique expansion (Lee et al., 2010)
- Mixed membership stochastic model (Airoldi et al., 2008)
- Bayesian matrix factorization (Psorakis et al., 2011)
- …
- BigCLAM (Yang and Lescovec, 2013)
- …

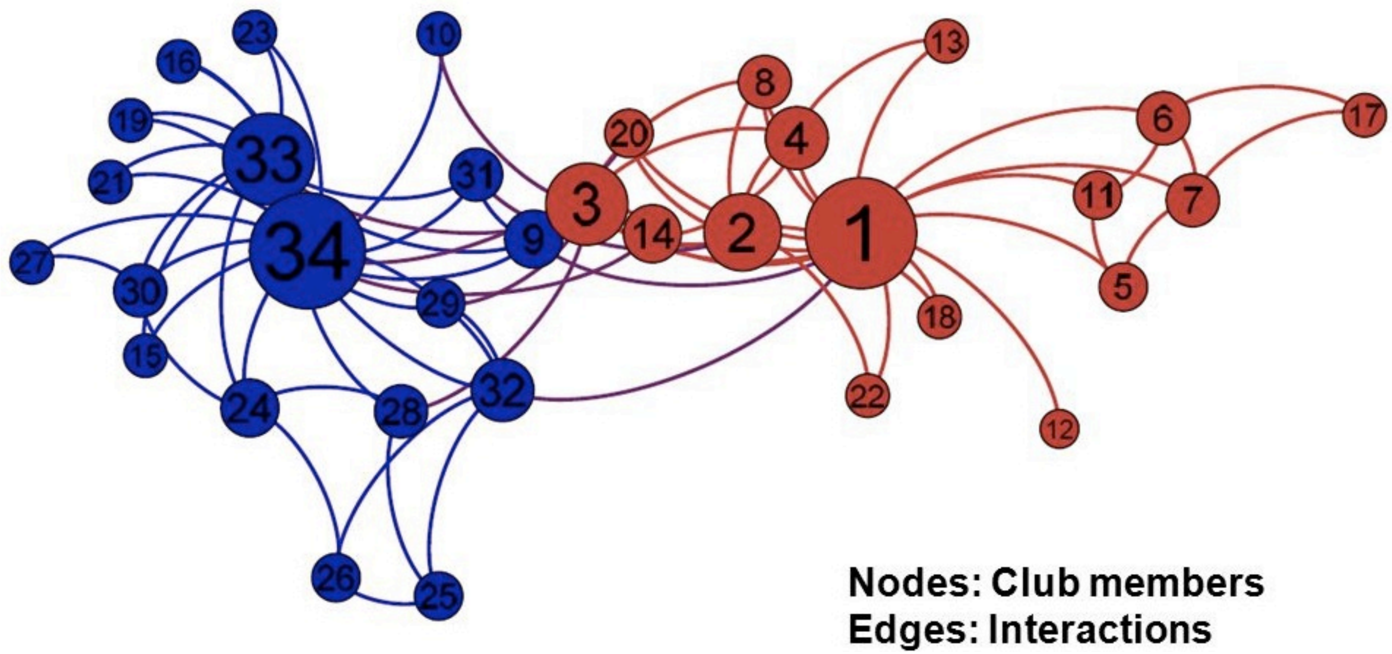
Dendrograms

Dendrograms



- A (agglomerative) hierarchical clustering algorithm
- Progressively add edges, from the strongest and ending with the weakest ones
- Example for Zachary's Karate club network

Zachary's Karate club (social) network



- Ground truth
- Observe social ties and rivalries in a university club
- During observation conflict led the group to split
- Split could be explained by a minimum cut

Pros and cons of dendrograms

Pros and cons

- Performance strongly depends on the chosen weight (local weight definitions typically provide weak solutions)
- Can be agglomerative or divisive, but adding strongest weights is in general weaker that deleting weaker ones
- May provide poor results
- Useful method, far from perfect

MIME.

Louvain algorithm

Blondel, Guillaume, Lambiotte, Lefebvre (2008)
Fast unfolding of communities in large networks
https://arxiv.org/abs/0803.0476

Modularity

Want to:

- measure of how well a network is partitioned into communities (i.e., sets of tightly connected nodes)
- solve the problem of selecting the number of partitions

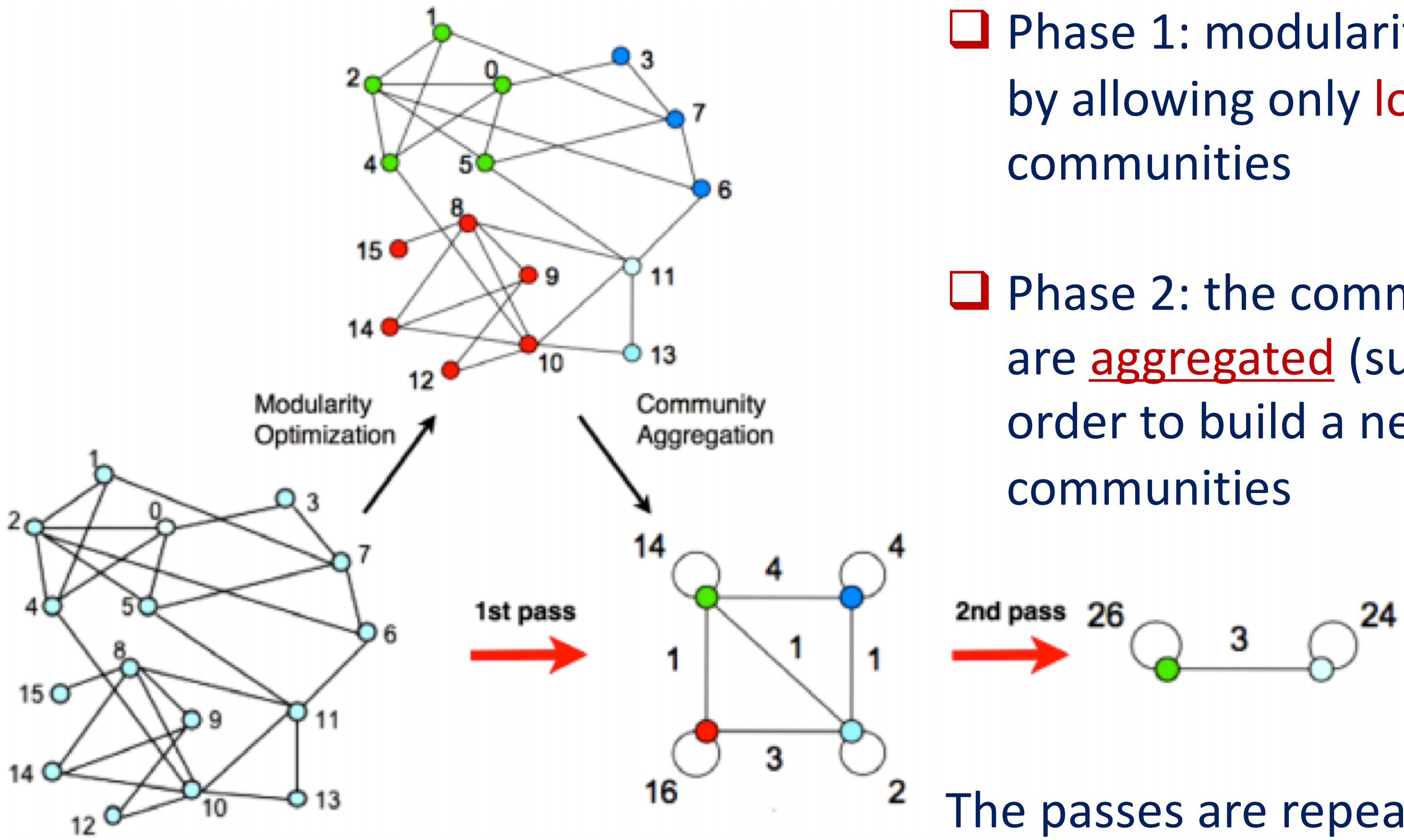
Idea:

- “If the number of edges between two groups is only what one would expect on the basis of random chance, then few thoughtful observers would claim this constitutes evidence of meaningful community structure”
- Modularity is “the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random”

A scalable approach

- Spectral approach robust but complex
- Need a scalable approach → Louvain
- A greedy technique
- Reference implementation in Python, R, MatLab

Hierarchical approach



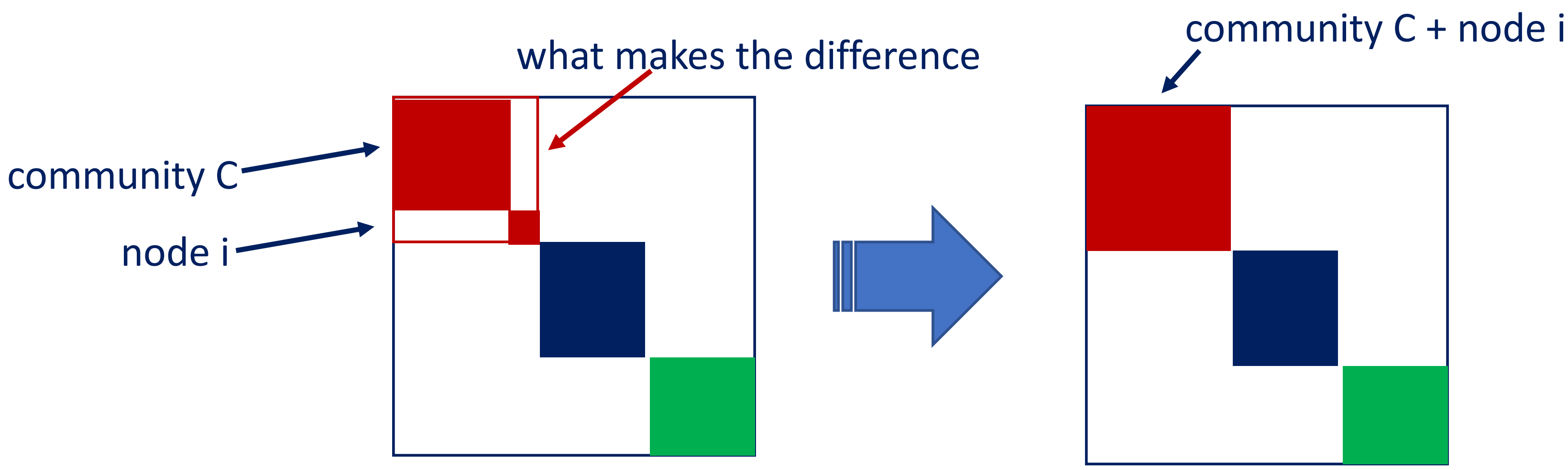
- ❑ Each node is a community @ start
- ❑ Phase 1: modularity is optimized by allowing only local changes of communities
- ❑ Phase 2: the communities found are aggregated (sum of links) in order to build a new network of communities

The passes are repeated iteratively until no increase of modularity is possible

MIME.

Local changes – easy to calculate

- for each node *i* consider the neighbours *j* of *i*
- evaluate the gain of modularity that would take place by removing *i* from its community and by placing it in the community of *j*
- node *i* is then placed in the community for which this gain is maximum (and positive)



$$\Delta Q = 2 \sum_{j \in C \cap N_i} (a_{ij} - d_i d_j / D)/D$$

Louvain: characteristics

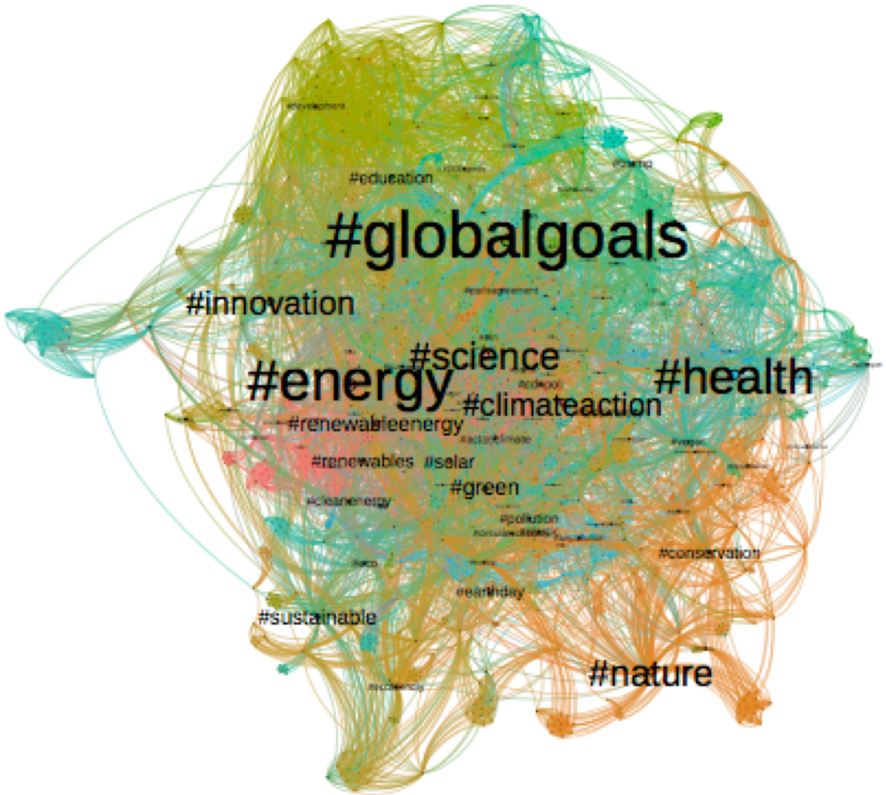
- Implements modularity optimization
- Scalable (low complexity)
- Effective
- Available as the reference implementation in any programming language
- A greedy technique (in the order the nodes are searched)

can be solved by consensus clustering

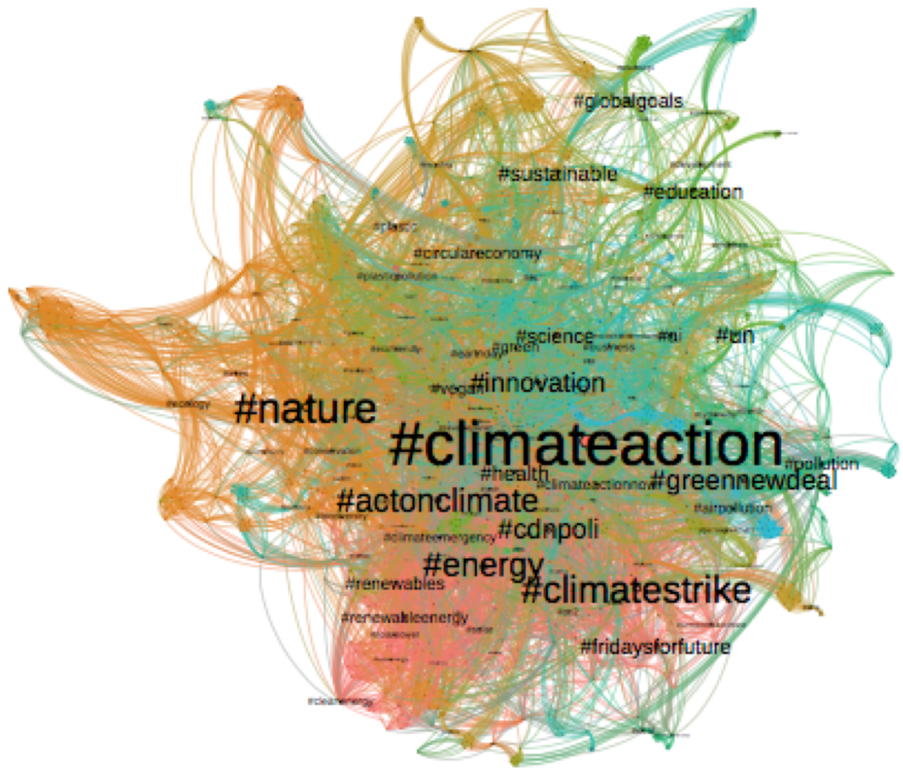
Example

#climateaction

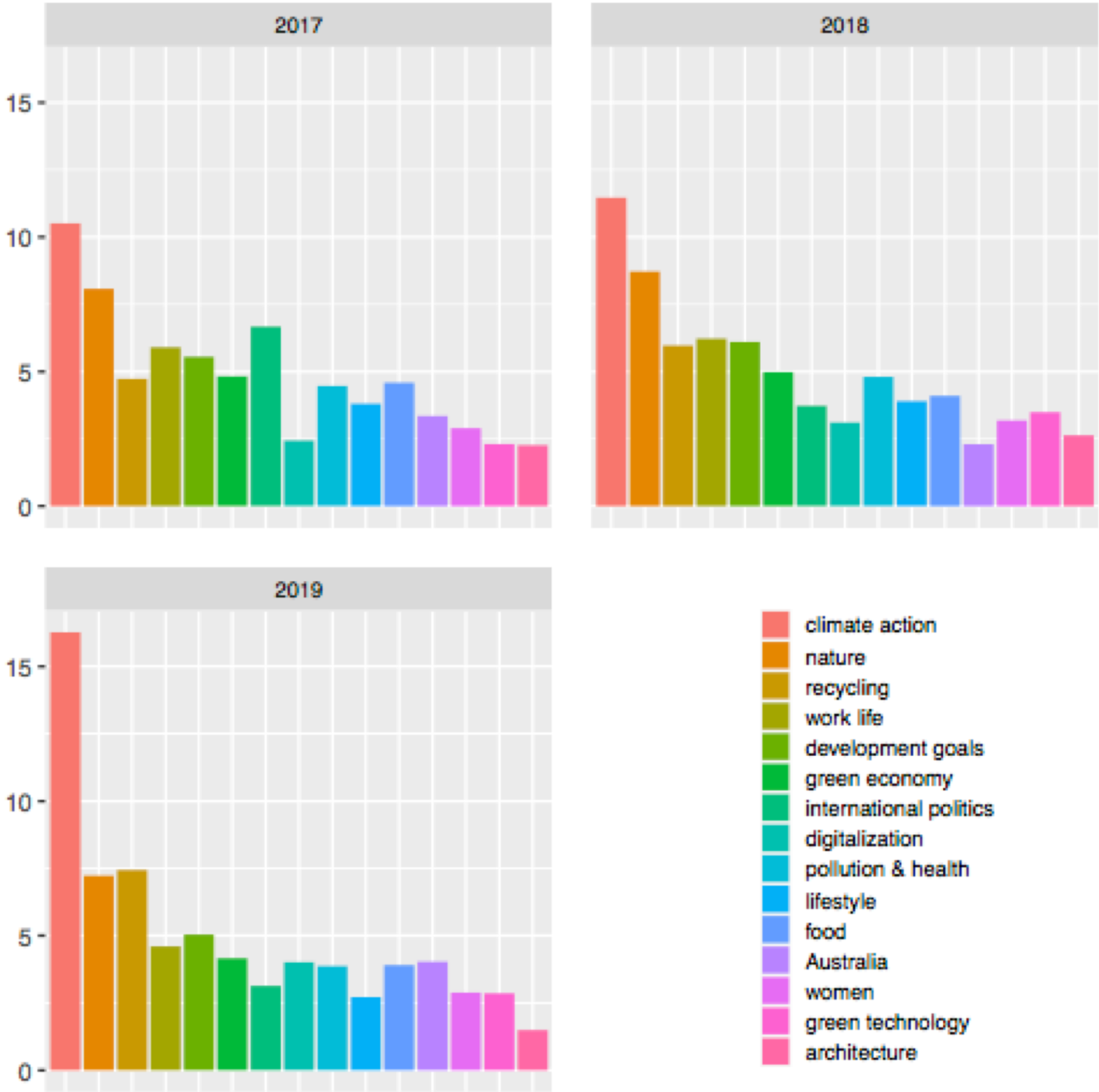
(b) Hashtag network in 2018



(c) Hashtag network in 2019



#climateaction



MIME.

#climateaction

#	Community name	Descriptive hashtags	Brief description
1	climate action	#climateaction, #actonclimate, #energy, #science, #cdnpoli, #renewableenergy, #renewables, #greennewdeal, #climatestrike	calls to action related to climate change
2	nature	#nature, #earthday, #conservation, #biodiversity, #oceans, #ecology, #trees, #forests, #wildlife	photos ad videos about naturalistic environments and animals
3	recycling	#innovation, #circulareconomy, #plastic, #sustainabledevelopment, #recycling, #ecofriendly, #recycle	business solutions for the circular economy, and recycling techniques
4	work life	#leadership, #employment, #creativity, #partnerships, #decentwork, #career	professional-life and working environment aspects
5	developments goals	#globalgoals, #education, #parisagreement, #un, #2030agenda, #community, #migration, #teachsdgs	2030 Global Goals for Sustainable Development
6	green economy	#green, #eco, #sugarcane, #ecofashion, #sustainablefashion, #vegetarian	promoting green and eco-friendly products
7	international politics	#trump, #epa, #resist, #coal, #p2, #environmentaljustice, #tcot, #usa, #2a, #oil, #theresistance, #eu	political topics
8	digitalization	#ai, #iot, #dataviz, #data, #bigdata, #digital, #smartcity, #digitaltransformation, #smarthome	methods and procedures for the digital transformation and innovations
9	pollution and health	#health, #pollution, #airpollution, #cities, #healthforall, #publichealth, #wellbeing, #airquality, #worldhealthday	topics of air pollution and public health
10	lifestyle	#weather, #travel, #coffee, #worldmetday, #europe, #spring, #thursdaythoughts, #london, #sxsw, #snow, #summer, #noaa, #greenland	big variety of free-time-related topics
11	food	#agriculture, #food, #zerohunger, #foodsecurity, #regenerativeagriculture, #insect, #urbanfarming, #learn, #foodtech	food issues and food technologies
12	Australia	#auspol, #extinctionrebellion, #climatecrisis, #greatbarrierreef, #stopadani, #australia, #extinction, #factsmatter, #ausvotes, #actnowforfuture, #brisbane	climate collective actions in Australia
13	women	#genderequality, #women, #womensday, #gender, #internationalwomensday, #iwd2018, #sdg5, #unea4, #localgov, #solvedifferent, #women4climate	gender-related topics

MIME.

Questions ?