Network Science

#11 Modularity

© 2020 T. Erseghe

Modularity

Newman, Modularity and community structure in networks (2006)
https://www.pnas.org/content/pnas/103/23/8577.full.pdf

Rationale

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

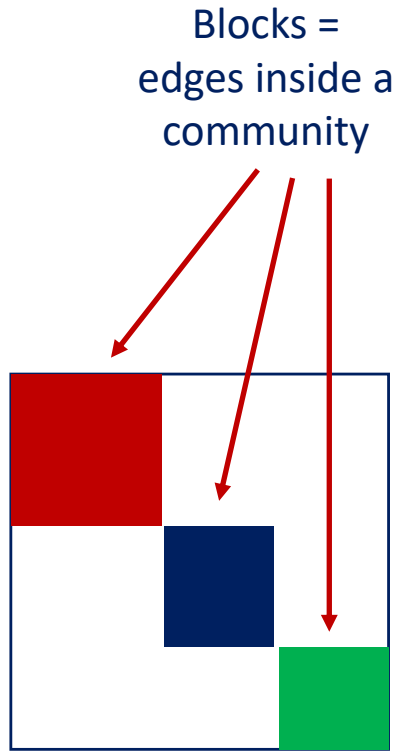
Idea:

- “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”

of edges within groups Q_1

$$Q_1 = \sum_{ij} a_{ij} \cdot \eta(c_i = c_j)$$

- a_{ij} entries of the (binary) adjacency matrix
- η indicating function (=1 if true)
- c_i community (value) of node i



of edges under random rewiring Q_2

The null model !

$$Q_2 = \sum_{ij} p_{ij} \cdot \eta(c_i = c_j)$$

- random rewiring keeps nodes degrees
- wiring probability $p_{ij} = k_i \cdot k_j / 2L$ (see Molloy-Reed)
- $k_i = \sum_j a_{ij}$ = node degree
- $2L = \sum_i k_i$ = # of stubs

probability of a trial

trials

Modularity

Modularity (normalized $-1 \leq Q \leq 1$)

$$Q = (Q_1 - Q_2)/2L$$

$$= 1/2L \cdot \sum_{ij} (a_{ij} - k_i \cdot k_j /2L) \cdot \eta(c_i = c_j)$$

- $Q > 0$ if the edges within groups exceed the (expected) random number
- $Q \in [0.3,0.7]$ for a significant community structure
- Q grows with size of the graph/number of (well-separated) clusters (Good et al, 2009) and cannot use Q to compare graphs which are very different in size
- Can be modified for signed networks

Two communities case

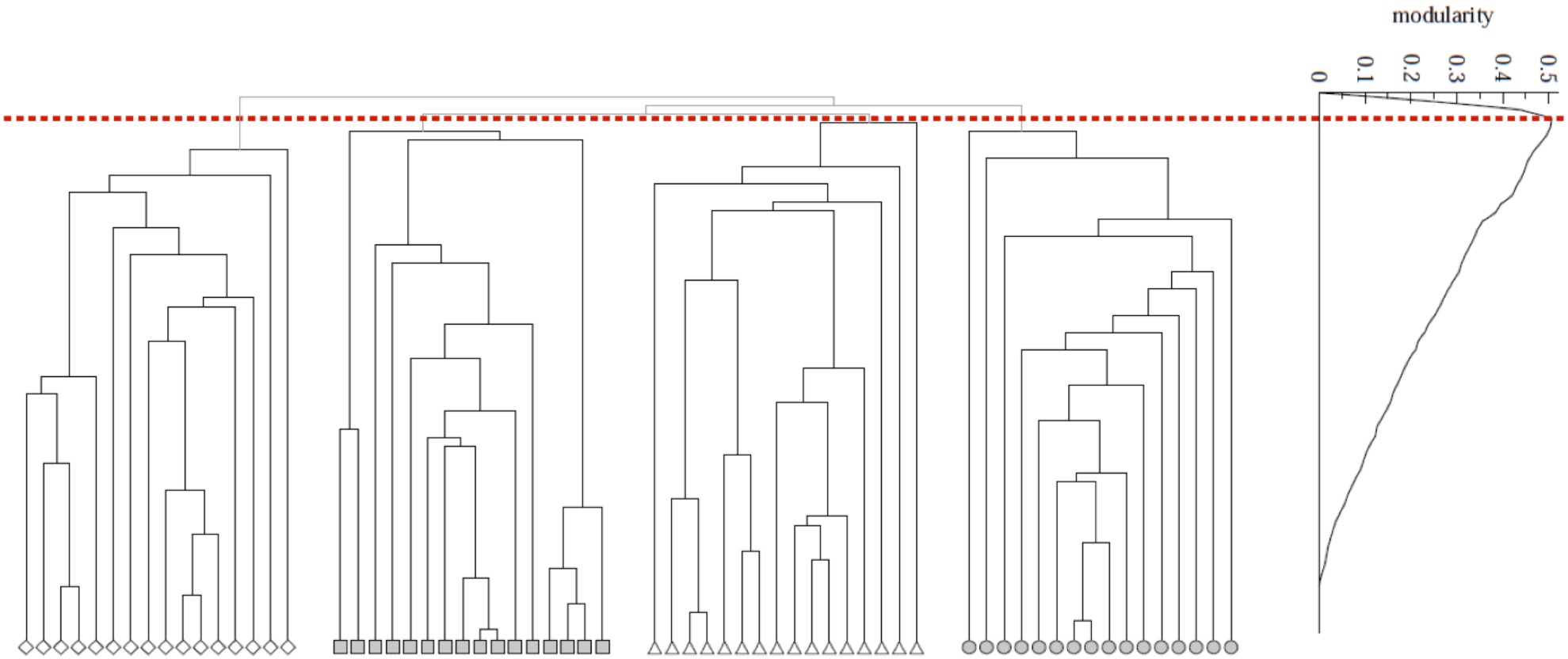
- Community membership signal
  - $s_i = 1$ if $c_i = 1$
  - $s_i = -1$ if $c_i = 2$
- Indicating function $\eta(c_i = c_j) = \frac{1}{2}(1+s_i s_j)$
- Modularity $Q = \sum_{ij} (a_{ij} - k_i \cdot k_j / 2L) \cdot (\cancel{1}+s_i s_j) / 4L$

$$= \sum_{ij} s_i \boxed{(a_{ij} - k_i \cdot k_j / 2L)} s_j / 4L$$

B_{ij} entries build matrix $\boldsymbol{B}$

- Compactly $\boxed{Q = \boldsymbol{s}^T \boldsymbol{B}\, \boldsymbol{s} / 4L}$

Modularity in dendrograms



We can use modularity for selecting the number of clusters

Modularity optimization

- Can use modularity for identifying an optimum community detection (max of modularity function)
- Need a simple algorithm for max Q

Modularity optimization – 2 communities

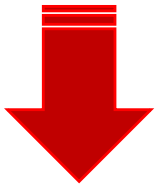
- Find maximum of $\boldsymbol{s}^{\mathsf{T}}\,\boldsymbol{B}\,\boldsymbol{s}$ under the constraint $\boldsymbol{s} \in \{\pm 1\}^{N}$
- Non-trivial NP-complex problem

Spectral approach

The 2-communities case

- $\boldsymbol{B} = \boldsymbol{A} - \boldsymbol{d}\,\boldsymbol{d}^T/D$ real-valued symmetric matrix
- $\boldsymbol{d} = \boldsymbol{A}\,\boldsymbol{1}$ degree vector
- $D = \boldsymbol{1}^T\,\boldsymbol{d}$ number of stubs (weighted version)

- $\boldsymbol{B} = \sum_i \boldsymbol{b}_i\,\boldsymbol{b}_i^T\,\lambda_i$ eigen-decomposition
- $\boldsymbol{b}_i$ normalized eigenvector $|\boldsymbol{b}_i| = 1$
- λ_i eigenvalue



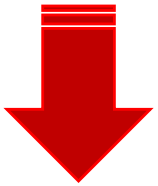
Maximize $\boldsymbol{s}^{\mathsf{T}}\,\boldsymbol{B}\,\boldsymbol{s} = \sum_i (\boldsymbol{b}_i^T\,\boldsymbol{s})^2\,\lambda_i$ under the constraint $s \in \{\pm 1\}^N$

Suboptimum method

Maximize $\boldsymbol{s}^{\mathsf{T}}\,\boldsymbol{B}\,\boldsymbol{s} = \sum_i (\boldsymbol{b}_i{}^T\,\boldsymbol{s})^2\,\lambda_i$ under the constraint $s \in \{\pm 1\}^N$

Keep only the strongest component

Maximize $(\boldsymbol{b}_1{}^T\,\boldsymbol{s})^2\,\lambda_1$ under the constraint $s \in \{\pm 1\}^N$



$\boldsymbol{b}_1$ can be extracted by a power iteration

$\boldsymbol{s} = \mathrm{sign}(\boldsymbol{b}_1)$

Modularity optimization

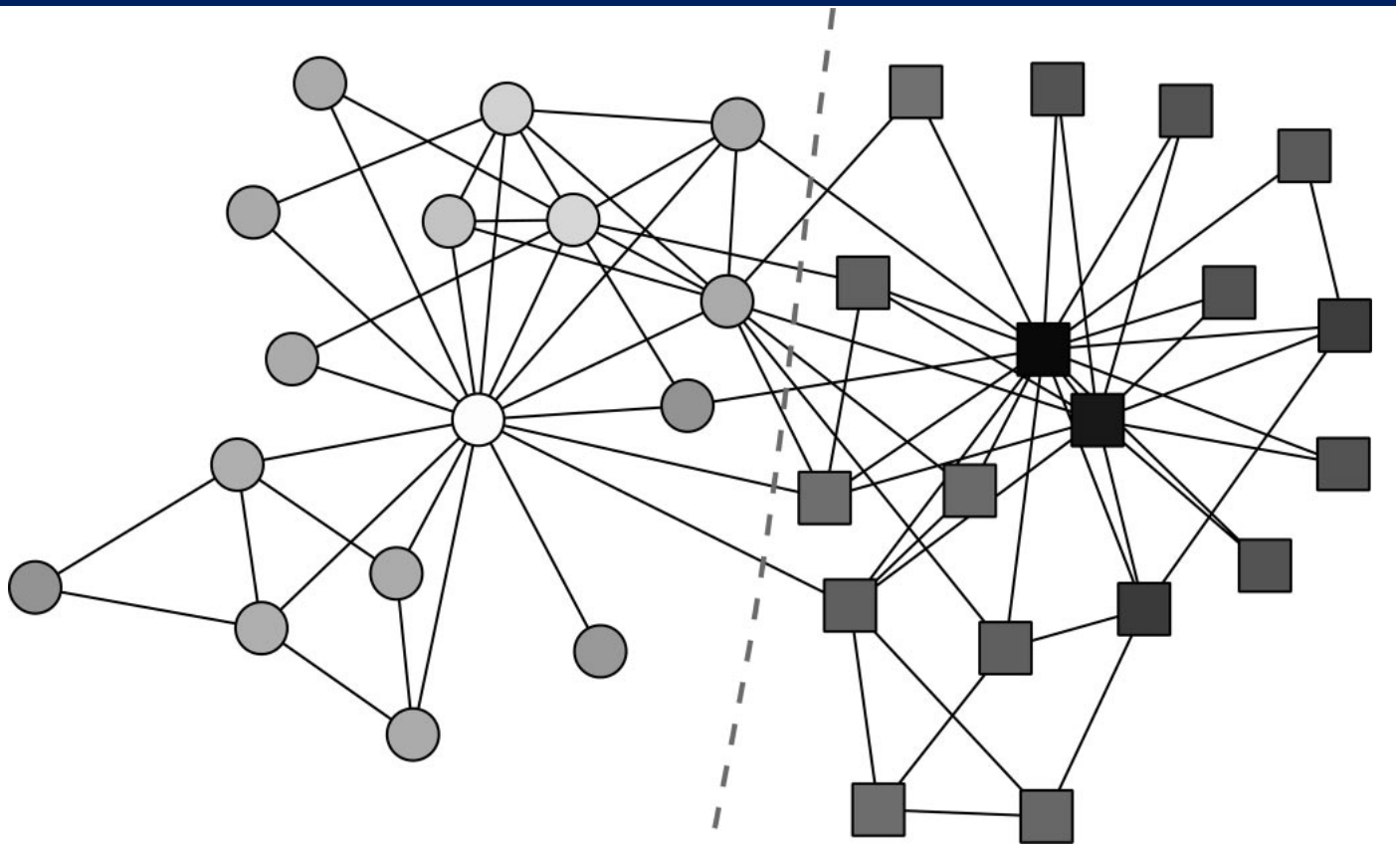


Fig. 2. Application of the eigenvector-based method to the karate club network of ref. 23. Shapes of vertices indicate the membership of the corresponding individuals in the two known factions of the network, and the dotted line indicates the split found by the algorithm, which matches the factions exactly. The shades of the vertices indicate the strength of their membership, as measured by the value of the corresponding elements of the eigenvector.

Modularity optimization

Fast modularity optimization

- Start from a single community and $Q = 0$
- Hierarchically bisection (partition in two) each community
  - Find the leading eigenvector $\boldsymbol{b}_1$ of $\boldsymbol{B}$
  - Identify the partition through the signs of $\boldsymbol{b}_1$
  - Eventually refine the partition by (iteratively) moving one vertex at-a-time to the other group , and by confirming the move if Q increases (Kernigan & Lin, 1970)
- If a proposed split does not cause modularity to increase, declare community indivisible

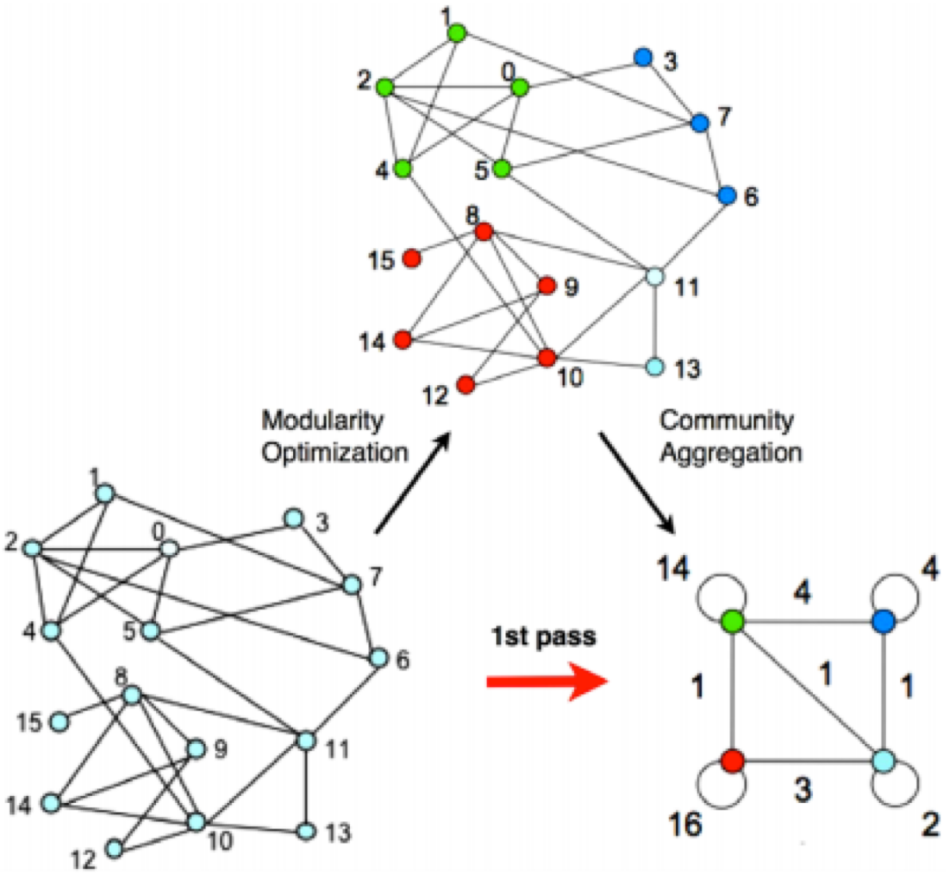
Louvain algorithm

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

A scalable approach

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

Hierarchical approach

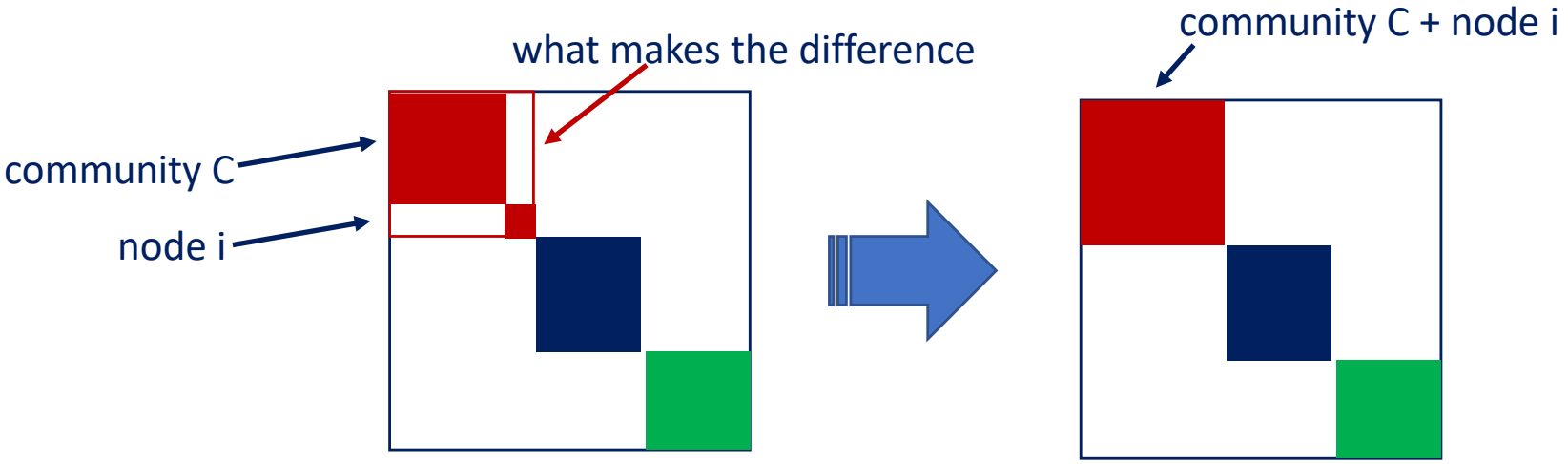


- Each node is a community @ start
- Phase 1: modularity is optimized by allowing only local changes of communities
- Phase 2: the communities found are aggregated (sum of links) in order to build a new network of communities

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

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)



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

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

Consensus clustering

Consensus clustering

Lancichinetti & Fortunato, Consensus clustering in complex networks, 2012
https://www.nature.com/articles/srep00336

1. Apply Louvain to *A* to yield *P* community detections (partitions)

2. Compute the consensus matrix *D*
   - D_{ij} is the fraction of partitions in which vertices *i* and *j* are assigned to the same cluster
   - entries below a chosen threshold are set to zero

3. Apply Louvain to *D* to yield *P* partitions
   - if the partitions are all equal, stop
   - otherwise go back to 2.

Consensus clustering example

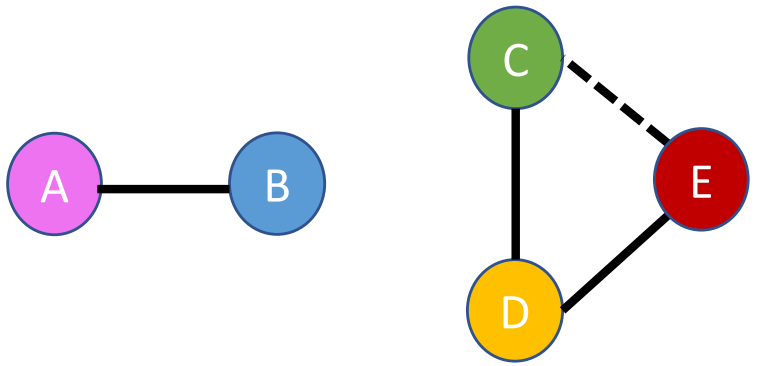
P=4 community assignments

1	1	3	2
1	2	3	2
2	3	2	1
2	4	1	1
3	5	1	1

5x5 consensus matrix (unnormalized)

4	3	0	0	0
3	4	0	0	0
0	0	4	2	1
0	0	2	4	2
0	0	1	2	4

Resulting "average" network



Resolution limit

The resolution limit

- prevents the algorithms in detecting small communities
- arises because the null model assumes that each node has an equal probability of connecting to every other node

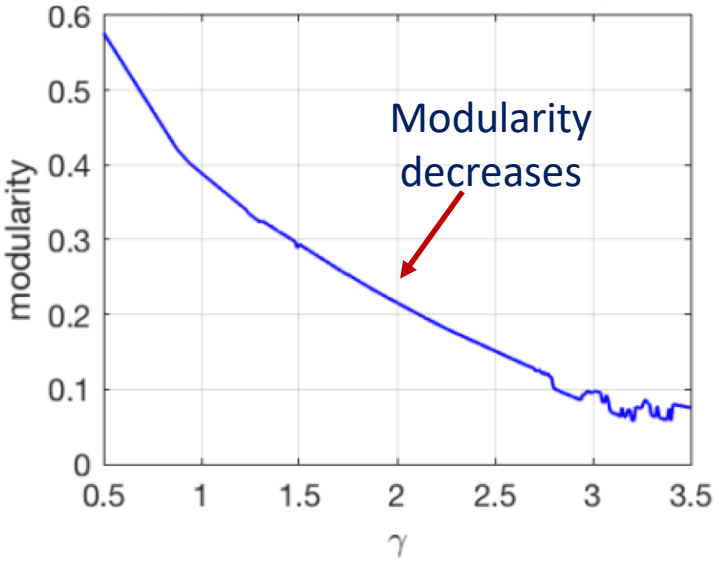
Generalized modularity

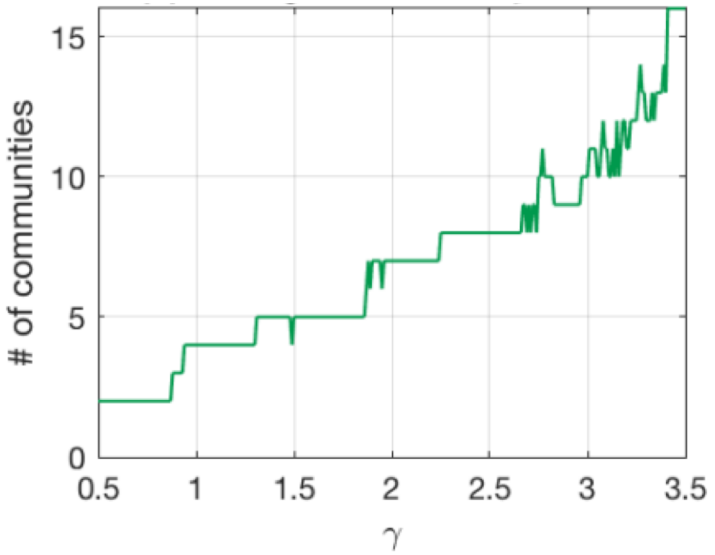
$$Q = 1/2L \cdot \sum (a_{ij} - \gamma\, k_i \cdot k_j /2L) \cdot \eta(c_i = c_j)$$

tunable value

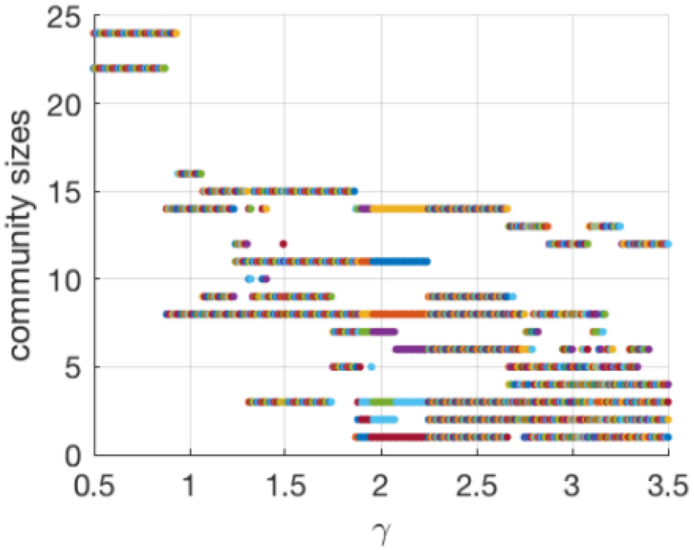
Performance wrt γ

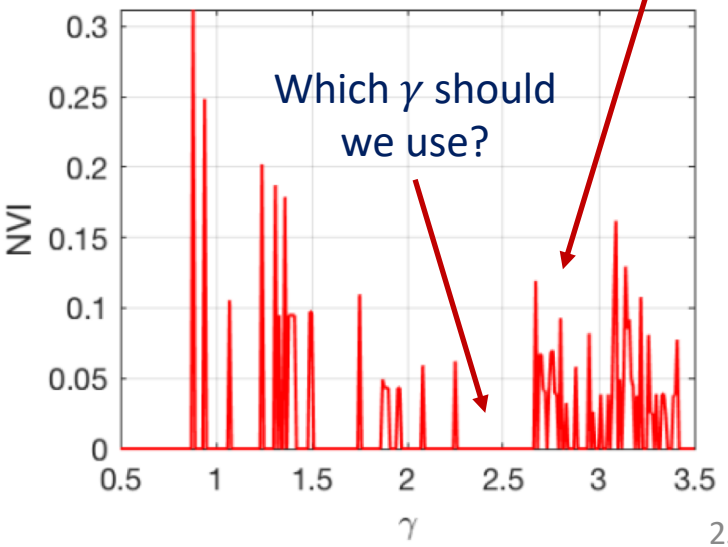
Nastaran Amini, Community and Hub Detection in Human Functional Brain Networks, 2020





Clean thanks to consensus clustering



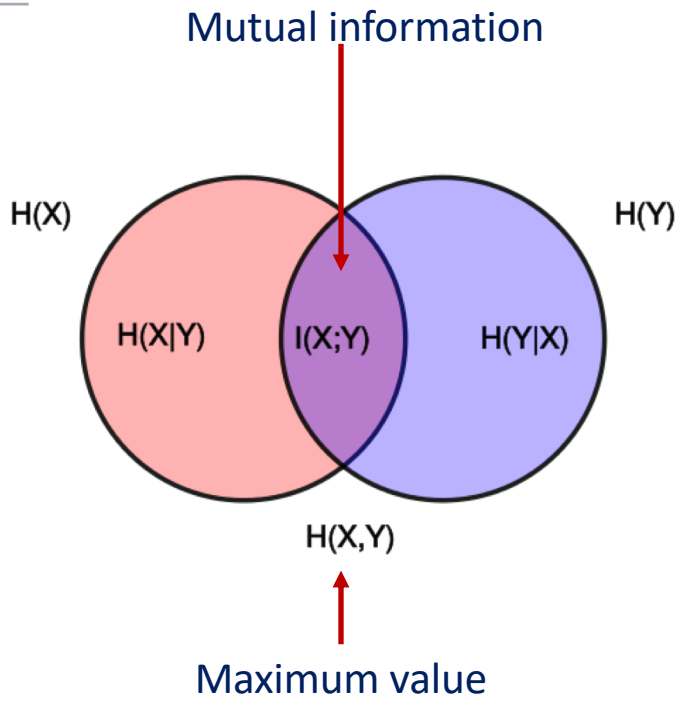


MI = mutual information

Mutual information

From Wikipedia, the free encyclopedia
(Redirected from Mutual Information)

In probability theory and information theory, the **mutual information** (**MI**) of two random variables is a measure of the mutual dependence between the two variables. More specifically, it quantifies the "amount of information" (in units such as shannons, commonly called bits) obtained about one random variable through observing the other random variable. The concept of mutual information is intimately linked to that of entropy of a random variable, a fundamental notion in information theory that quantifies the expected "amount of information" held in a random variable.



e.g., similarity between two different community assignments

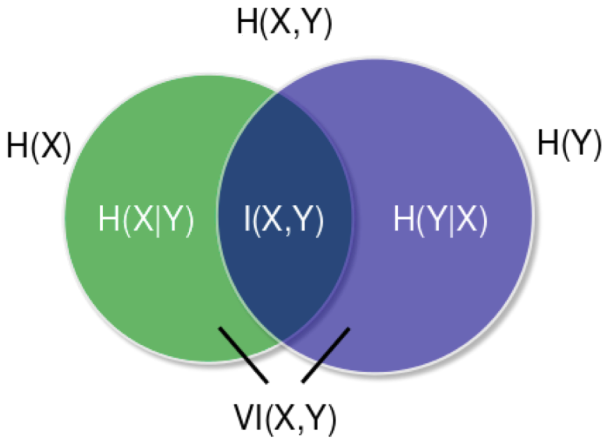
VI = variation of information

Variation of information

From Wikipedia, the free encyclopedia

In probability theory and information theory, the **variation of information** or **shared information distance** is a measure of the distance between two clusterings (partitions of elements). It is closely related to mutual information; indeed, it is a simple linear expression involving the mutual information. Unlike the mutual information, however, the variation of information is a true metric, in that it obeys the triangle inequality.[1] [2] [3]

e.g., dissimilarity between two different community assignments



VI = variation of information

Fraction of nodes belonging to the ith community in the 1st assignment

$$\mathrm{VI}(X;Y) = -\sum_{i,j} r_{ij} \left[\log(r_{ij}/p_i) + \log(r_{ij}/q_j)\right]$$

Fraction of nodes belonging to the ith community in the 1st assignment and to the jth community in the 2nd assignment

Fraction of nodes belonging to the jth community in the 2nd assignment

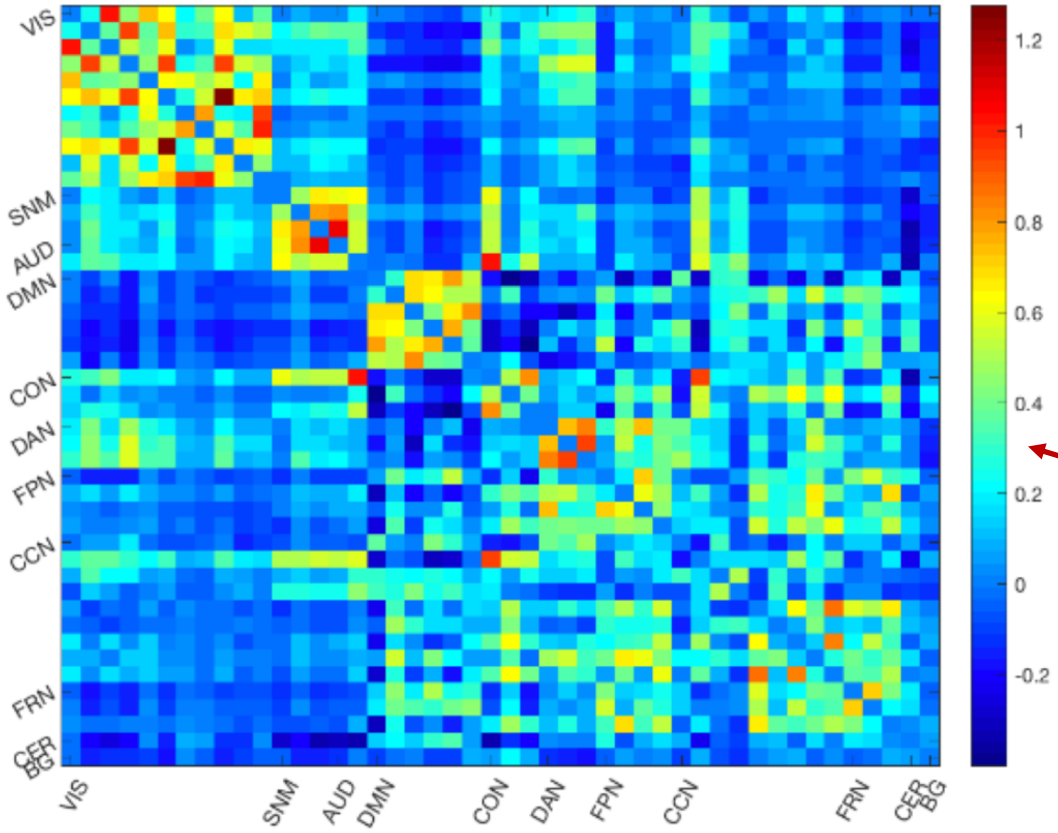
An application example

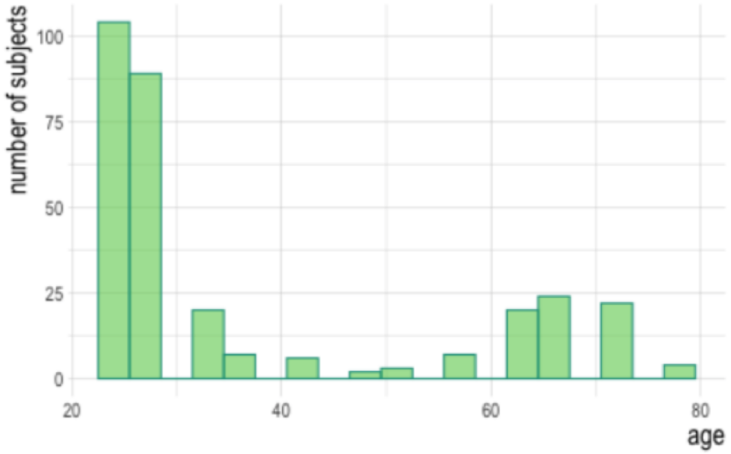
Community detection in functional brain networks (fMRI data)

The challenge

How do brain regions interconnect?

fMRI correlation matrix



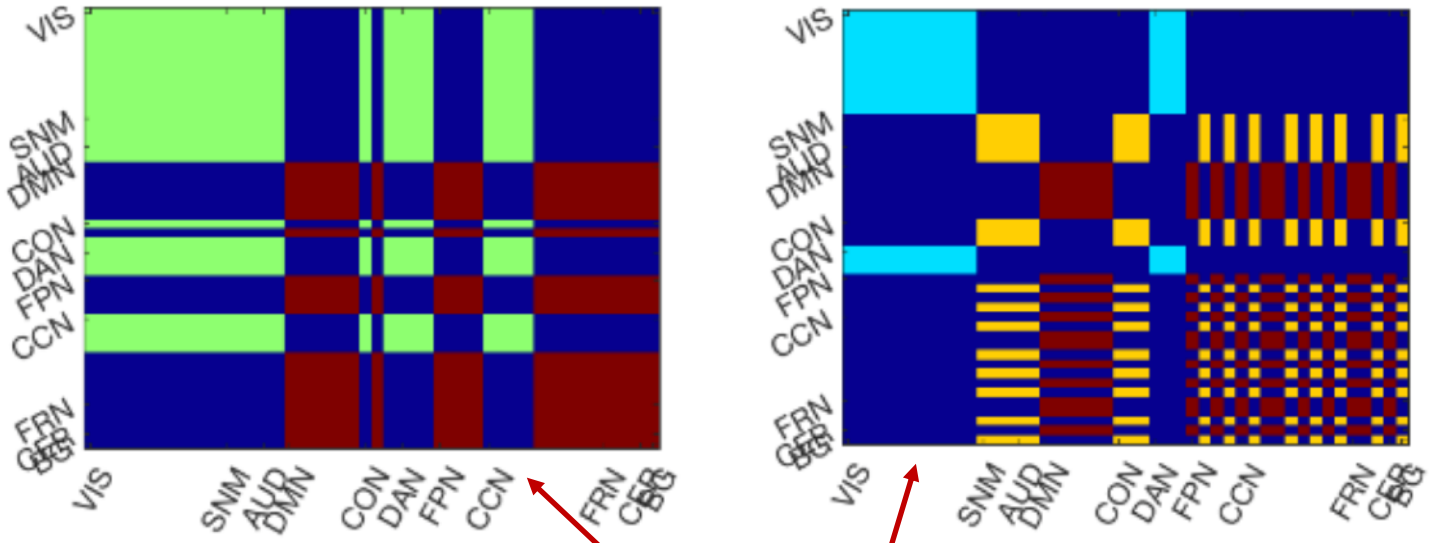


Age histogram of the 308 subjects under study

Can run Louvain on it to identify meaningful patterns for each subject

The patterns

3 patterns per subject

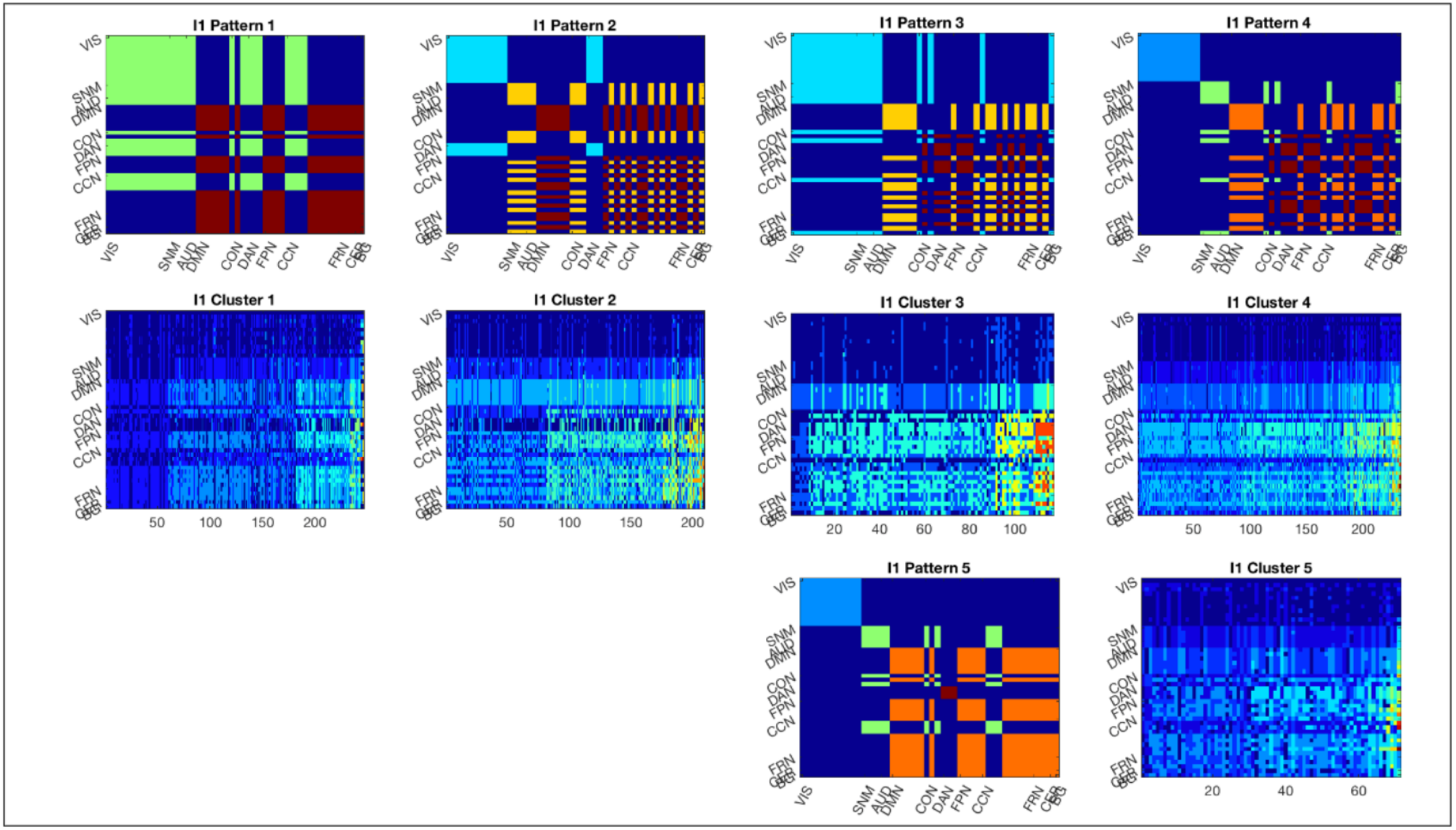


Can we cluster them?
Yes, using **Louvain** on a **similarity** matrix

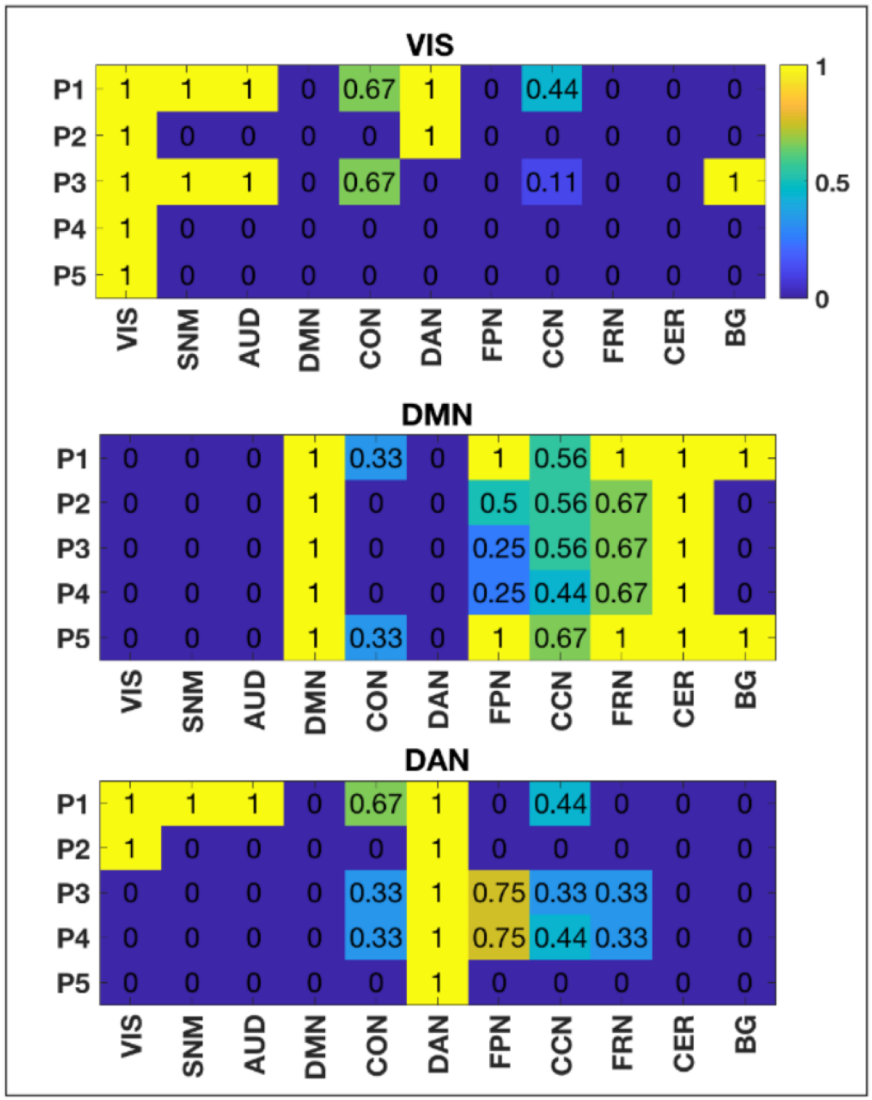
How to measure their **similarity**?

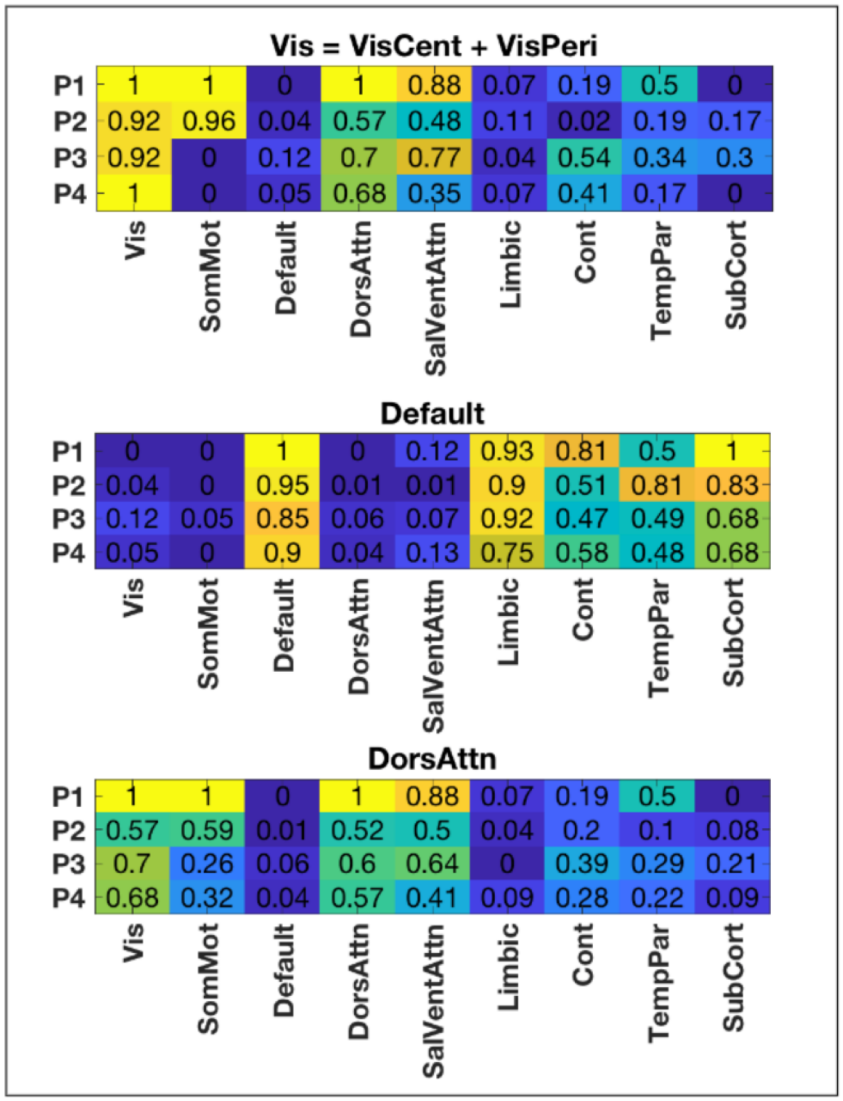
- NVI ☹
- **Dice correlation** ☺

The clusters



The result


VIS
P1 1 1 1 0 0.67 1 0 0.44 0 0 0
P2 1 0 0 0 0 1 0 0 0 0 0
P3 1 1 1 0 0.67 0 0 0.11 0 0 1
P4 1 0 0 0 0 0 0 0 0 0 0
P5 1 0 0 0 0 0 0 0 0 0 0
VIS SNM AUD DMN CON DAN FPN CCN FRN CER BG
1
0.5
0
DMN
P1 0 0 0 1 0.33 0 1 0.56 1 1 1
P2 0 0 0 1 0 0 0.5 0.56 0.67 1 0
P3 0 0 0 1 0 0 0.25 0.56 0.67 1 0
P4 0 0 0 1 0 0 0.25 0.44 0.67 1 0
P5 0 0 0 1 0.33 0 1 0.67 1 1 1
VIS SNM AUD DMN CON DAN FPN CCN FRN CER BG
DAN
P1 1 1 1 0 0.67 1 0 0.44 0 0 0
P2 1 0 0 0 0 1 0 0 0 0 0
P3 0 0 0 0 0.33 1 0.75 0.33 0.33 0 0
P4 0 0 0 0 0.33 1 0.75 0.44 0.33 0 0
P5 0 0 0 0 0 1 0 0 0 0 0
VIS SNM AUD DMN CON DAN FPN CCN FRN CER BG


Vis = VisCent + VisPeri
P1 1 1 0 1 0.88 0.07 0.19 0.5 0
P2 0.92 0.96 0.04 0.57 0.48 0.11 0.02 0.19 0.17
P3 0.92 0 0.12 0.7 0.77 0.04 0.54 0.34 0.3
P4 1 0 0.05 0.68 0.35 0.07 0.41 0.17 0
Vis SomMot Default DorsAttn SalVentAttn Limbic Cont TempPar SubCort
Default
P1 0 0 1 0 0.12 0.93 0.81 0.5 1
P2 0.04 0 0.95 0.01 0.01 0.9 0.51 0.81 0.83
P3 0.12 0.05 0.85 0.06 0.07 0.92 0.47 0.49 0.68
P4 0.05 0 0.9 0.04 0.13 0.75 0.58 0.48 0.68
Vis SomMot Default DorsAttn SalVentAttn Limbic Cont TempPar SubCort
DorsAttn
P1 1 1 0 1 0.88 0.07 0.19 0.5 0
P2 0.57 0.59 0.01 0.52 0.5 0.04 0.2 0.1 0.08
P3 0.7 0.26 0.06 0.6 0.64 0 0.39 0.29 0.21
P4 0.68 0.32 0.04 0.57 0.41 0.09 0.28 0.22 0.09
Vis SomMot Default DorsAttn SalVentAttn Limbic Cont TempPar SubCort

Another application example

The rise of #climateaction in the time of the FridaysForFuture movement: A semantic network analysis, Social Networks, 2022
https://www.sciencedirect.com/science/article/pii/S037887332200057

#climatechange before/after Greta

What to search in Twitter?

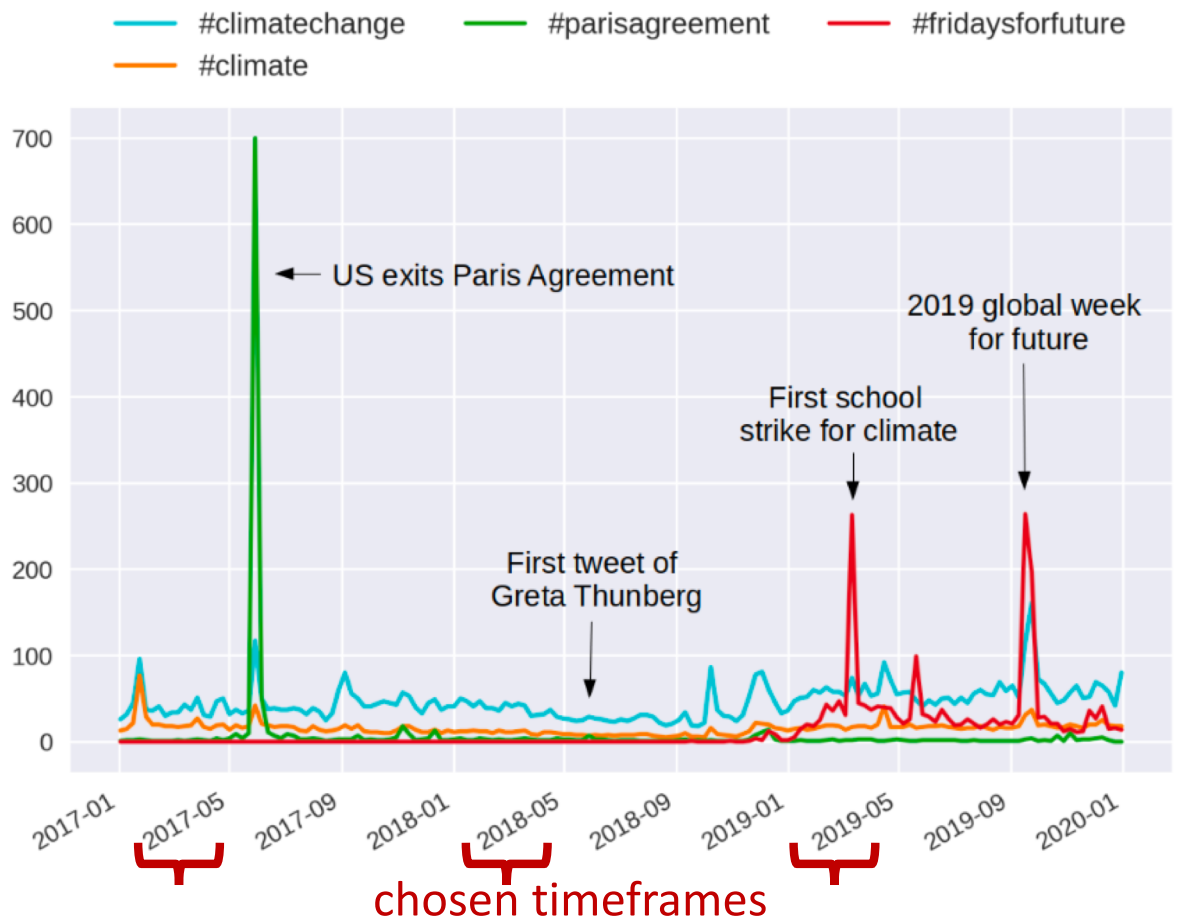


Figure 1: Historical twitter trends for some hashtags related to climate action where values represent 1/10,000th of 1% of tweets.

Refining hashtags

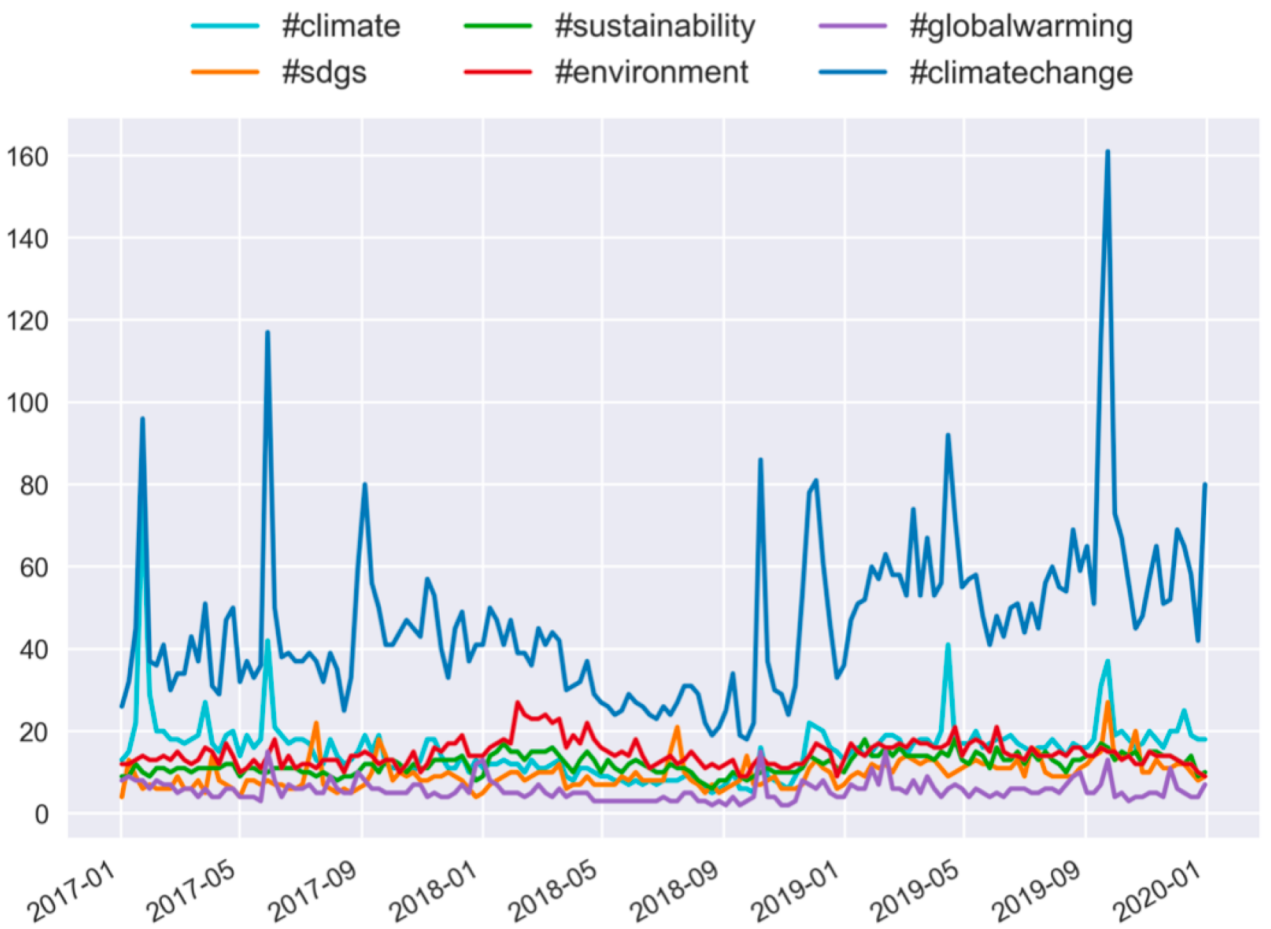
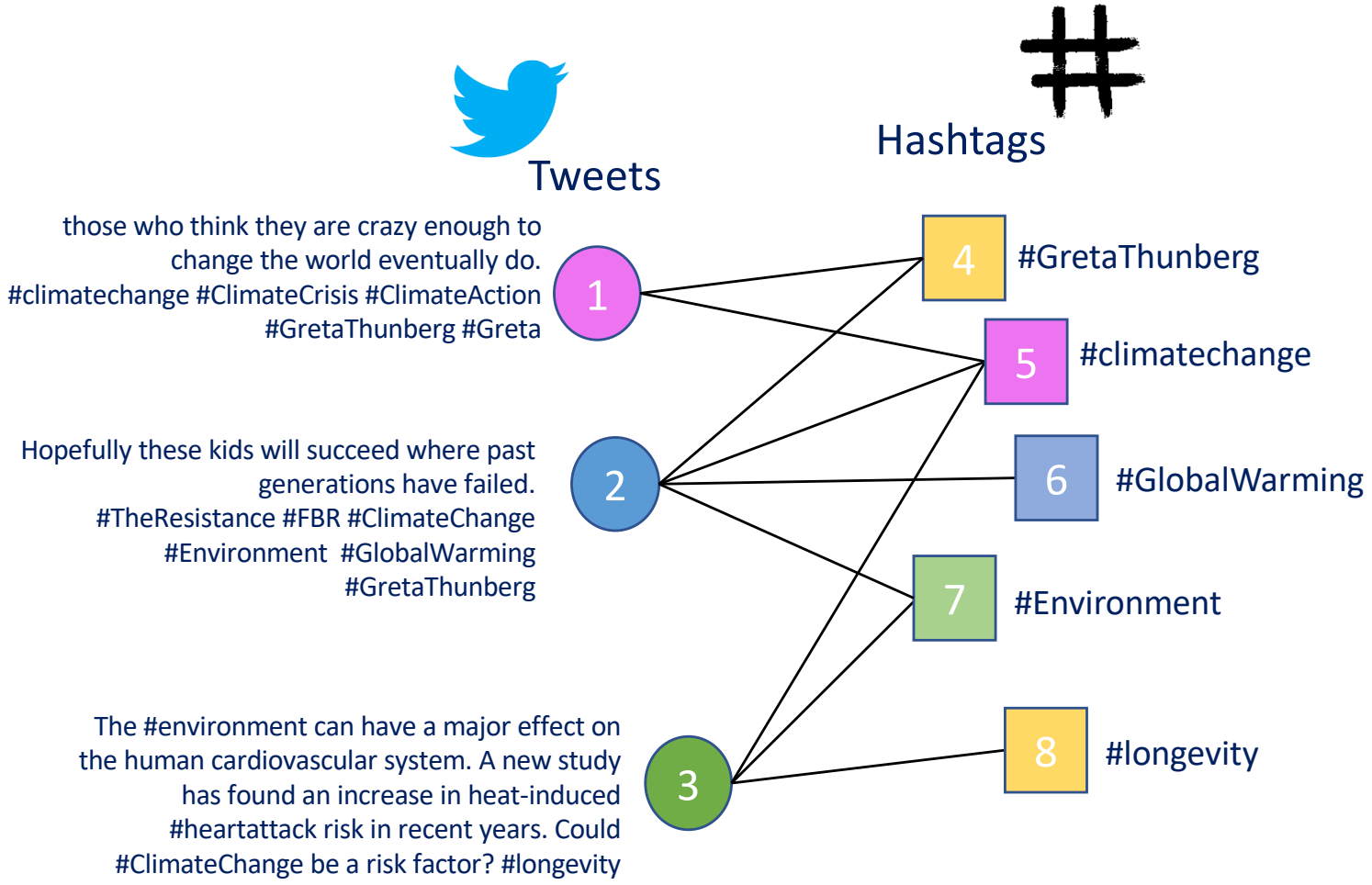
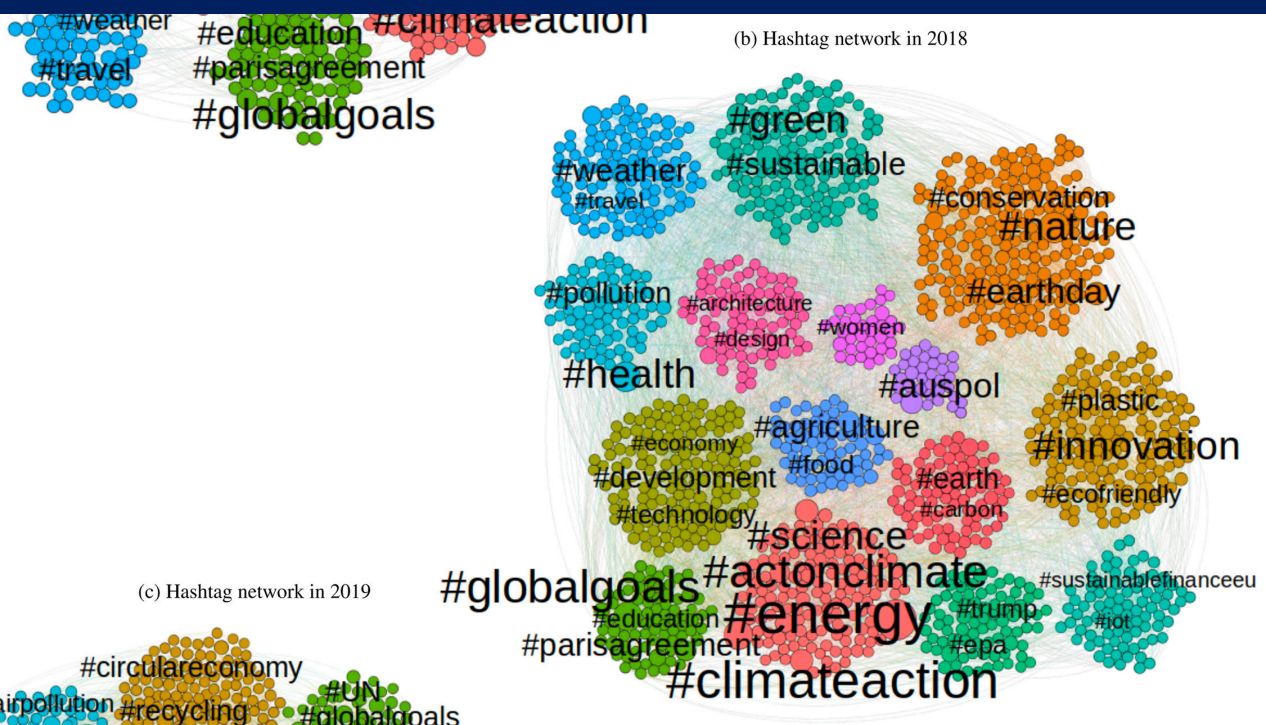


Figure 2: Historical twitter trends for the selected neutral hashtags.

Bipartite graph



Communities



Communities (1/2)

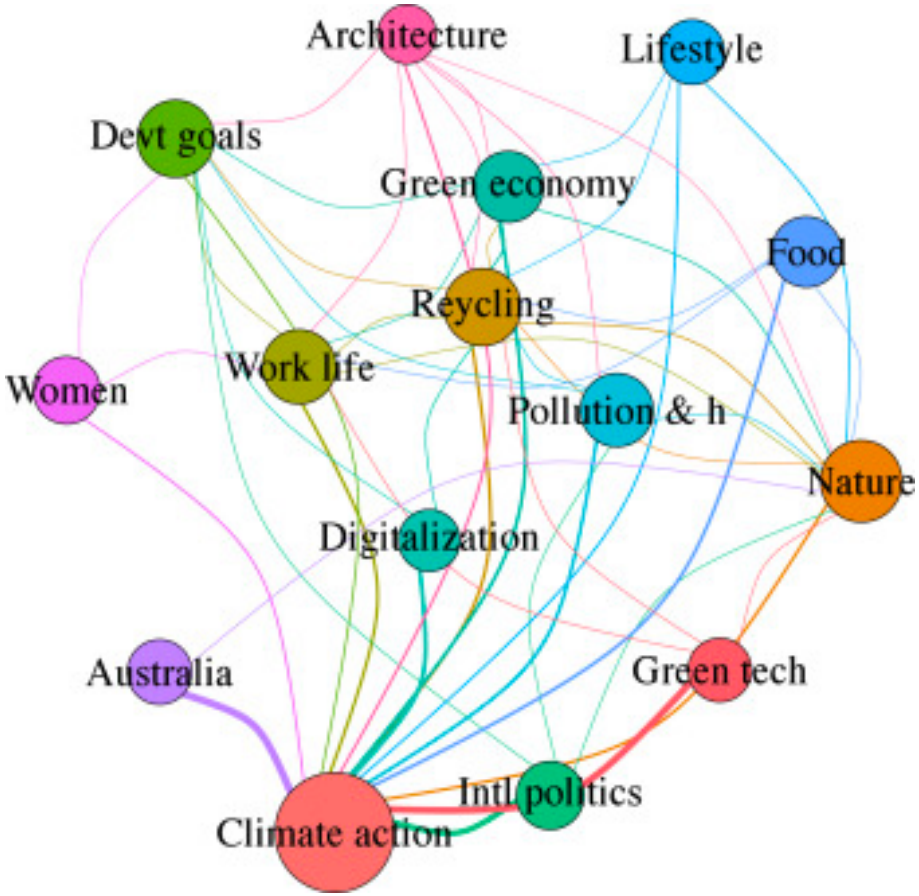
Assigned by community-specific PageRank

#	Community name	Brief description	Descriptive hashtags	Descriptive tweet
1	climate action	calls to action related to climate change	#climateaction, #actonclimate, #energy, #science, #cdnpoli, #renewableenergy, #renewables, #greennewdeal, #climatestrike	We too are taking part to the #GlobalClimateStrike in #Torino. If you want to make the world a better place, take a look at yourself and then make the change #FridaysForFuture #climatestrike #climatechange #globalstrikeforfuture #globalwarming #StrikeForClimate https://t.co/gGFJE1wOkk
2	nature	photos ad videos about naturalistic environments and animals	#nature, #earthday, #conservation, #biodiversity, #oceans, #ecology, #trees, #forests, #wildlife	Traces in rock may be the oldest evidence of life on Earth e... https://t.co/aJ9zI5U2Bp #nature #Science #climate https://t.co/kLx6nkXAsR
3	recycling	business solutions for the circular economy, and recycling techniques	#innovation, #circulareconomy, #plastic, #sustainabledevelopment, #recycling, #ecofriendly, #recycle	Buy less; buy smarter - why you should buy less and how https://t.co/dXIm6tWmVT #Sustainability #turnthetideonplastic #make2019count #circulareconomy #circularthinking
4	work life	professional life and working environment aspects	#leadership, #employment, #creativity, #partnerships, #decentwork, #career	CREATIVE WORK: Respect the Dignity of All Types of Work https://t.co/8trFtZRHNf #creativity #millennials #boomers #YoungAdults #selfies #students #employment #workers #money #unemployment #satisfaction #technology #Innovation #sustainability #compensation #income #poverty https://t.co/hM5p7YzDPn
5	developments goals	2030 Global Goals for Sustainable Development	#globalgoals, #education, #parisagreement, #un, #2030agenda, #community, #migration, #teachsdgs	Which of the #GlobalGoals are you supporting as an African Youth? #SDGs #TheAfrikanLegacy #TheAfricaWeWant #YouthForAfrica #YouthForChange https://t.co/B4ZMNzvPcS
6	green economy	promoting green and eco-friendly products	#green, #eco, #sugarcane, #ecofashion, #sustainablefashion, #vegetarian	Our 4th panellist for #TheTrueCost film night is Fashion Design student Olivia Jane Riley! She works to promote sustainable and #ecofashion, and believes that #sustainability is the next big shift that will change the industry. https://t.co/omGn1YKOYH #FashionExperts #fastfashion https://t.co/7llCaDW8LH
7	international politics	political topics	#trump, #epa, #resist, #coal, #p2, #environmentaljustice, #tcot, #usa, #2a, #oil, #theresistance, #eu	#China backing #TeslaMotors electric car. Trump backing #coal . Making America Return to 1950. #blacklung @realDonaldTrump ≠ #environment
8	digitalization	methods and procedures for the digital transformation and innovations	#ai, #iot, #dataviz, #data, #bigdata, #digital, #smartcity, #digitaltransformation, #smarthome	#smartcity in a box - #INFOGRAPHICS #Security #sustainability #efficiency #people #smartcities #smartgrid #smarttraffic #IoT #bigdata #blockchain #CloudComputing #startups #digitaltransformation #HealthTech #FinTechs https://t.co/HLJWOjZBBx

Communities (2/2)

9	pollution and health	topics of air pollution and public health	#health, #pollution, #airpollution, #cities, #healthforall, #publichealth, #wellbeing, #airquality, #worldhealthday	Toxic chemicals come out of polystyrene products when heated ,and spill into food or drink that is inside.Polystyrene chemicals leak into drinks by a small amount. photo credit: 5gyres #sustyvibes #staredownonpollution #pollution #SOTSonduty #environment #sustyfacts https://t.co/SsVw1HEGre
10	lifestyle	big variety of free-time-related topics	#weather, #travel, #coffee, #worldmetday, #europe, #spring, #thursdaythoughts, #london, #sxsw, #snow, #summer, #noaa, #greenland	Top Beach Destinations reviewed! https://t.co/qbleTwY1S3 #travel #travellight #travelpreneur #sustainability #laptoplifestyle https://t.co/gQMe9ynNf7
11	food	food issues and food technologies	#agriculture, #food, #zerohunger, #foodsecurity, #regenerativeagriculture, #insect, #urbanfarming, #learn, #foodtech	Extreme temperature changes are damaging food production throughout the world. Agriculture needs to become climate-smart so we can achieve #ZeroHunger https://t.co/WuuCSGsHLJ #climatechange #WorldMetDay https://t.co/LaHfsRlChL
12	Australia	climate collective actions in Australia	#auspol, #extinctionrebellion, #climatecrisis, #greatbarrierreef, #stopadani, #australia, #extinction, #factsmatter, #ausvotes, #actnowforfuture, #brisbane	@giveadam_ @ozrobotgirl She's a lone voice for #WorldHeritage! When will the REST of @AustralianLabor catch up! #environment #extinction @bmucinc @bmcsnsw @CrMarkGreenhill @brenthoare @RomolaHollywood @MehreenFaruqi @billshortenmp @Tony_Burke @Mark_Butler_MP #auspol #2019Elections #AusVotes2019 https://t.co/UDAgmFk1TF
13	women	gender-related topics	#genderequality, #women, #womensday, #gender, #internationalwomensday, #iwd2018, #sdg5, #unea4, #localgov, #solvedifferent, #women4climate	Hands up for #GenderEquality! #YouthCSW61 we are all about inclusion @UNWOMEN4Youth #abilityvsdisability #CRPD #CelebratingDiversity #SDGs https://t.co/XqYWKDSiSk
14	green technology	technological and sustainable innovations	#earth, #carbon, #jobs, #blockchain, #emissions, #cleantech, #engineering, #startups, #ghg, #electric, #natural, #paris, #life, #mining	This #Norwegians Easter research break is brilliant. Hopefully some progress in understanding more of #integratedreporting and #sustainability. Tomorrow most probably connecting it to #blockchain #businessmodels I talked of last week.
15	architecture	architecture topics	#architecture, #fashion, #design, #construction, #greenbuilding, #building, #webinar, #steamdrills, #5star, #innovative, #free, #interiordesign	.@WorldGBC's @CDomGreen on how #greenbuildings can help achieve many of the #globalgoals #SDGs https://t.co/16Un9htc6T

Relations among communities



Questions ?