Social Network Analysis

Other centrality measures

MIME.

Local PageRank

PageRank

Idea:

- the surfer does not necessarily move to one of the links of the page she/he is viewing
- with a certain probability, might jump to a random page

damping factor, typically $c = 0.85$, meaning that 85% of the times the surfer moves to one of the links of the page

the remaining $1 - c$ = 15% of the times the surfer moves to a random page according to a probability vector $\boldsymbol{q}$, e.g., $\boldsymbol{q}=\boldsymbol{1}/N$ for uniform probability

Measuring closeness: Local PageRank

Idea

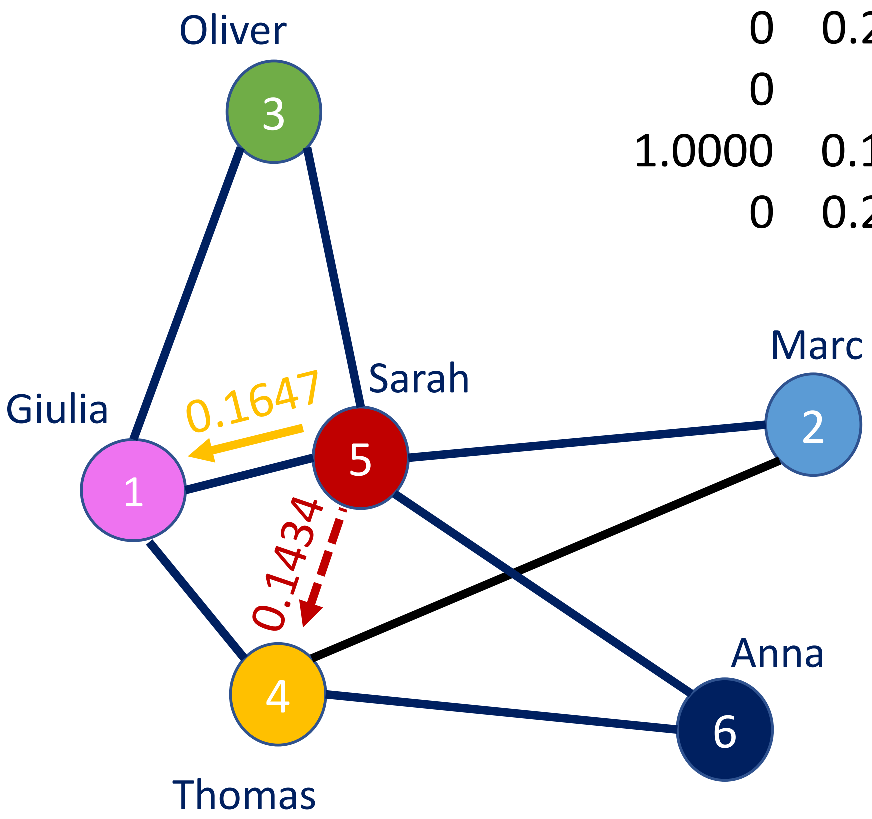
- Measure similarity to node *i* by applying TopicSpecific PageRank with a teleport set with a unique element $S = \{i\}$ and $q = [0 \ldots 0\ 1\ 0 \ldots 0]$

Result

- Measures direct and indirect multiple connections, their quality, degree or weight

Example: who's Sarah's best friend?

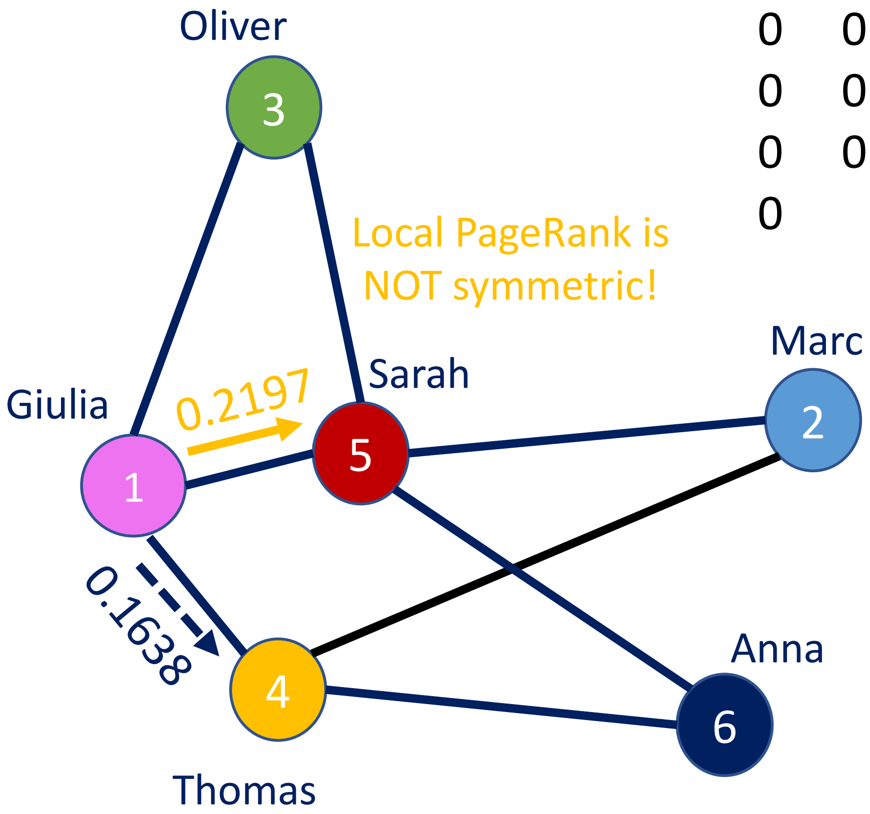
t=1	2	3	4	5
0	0.2125	0.1222	0.2096	0.1290
0	0.2125	0.0319	0.1705	0.0708
0	0.2125	0.0921	0.1369	0.1127
0	0	0.2408	0.0617	0.2043
1.0000	0.1500	0.4811	0.2508	0.4125
0	0.2125	0.0319	0.1705	0.0708



10	20	50	75	100	
0.1743	0.1653	0.1647	0.1647	0.1647	Giulia
0.1238	0.1144	0.1138	0.1138	0.1138	Marc
0.1206	0.1199	0.1199	0.1199	0.1199	Oliver
0.1285	0.1426	0.1434	0.1434	0.1434	Thomas
0.3290	0.3435	0.3444	0.3444	0.3444	Sarah
0.1238	0.1144	0.1138	0.1138	0.1138	Anna

Example: who's Giulia's best friend?

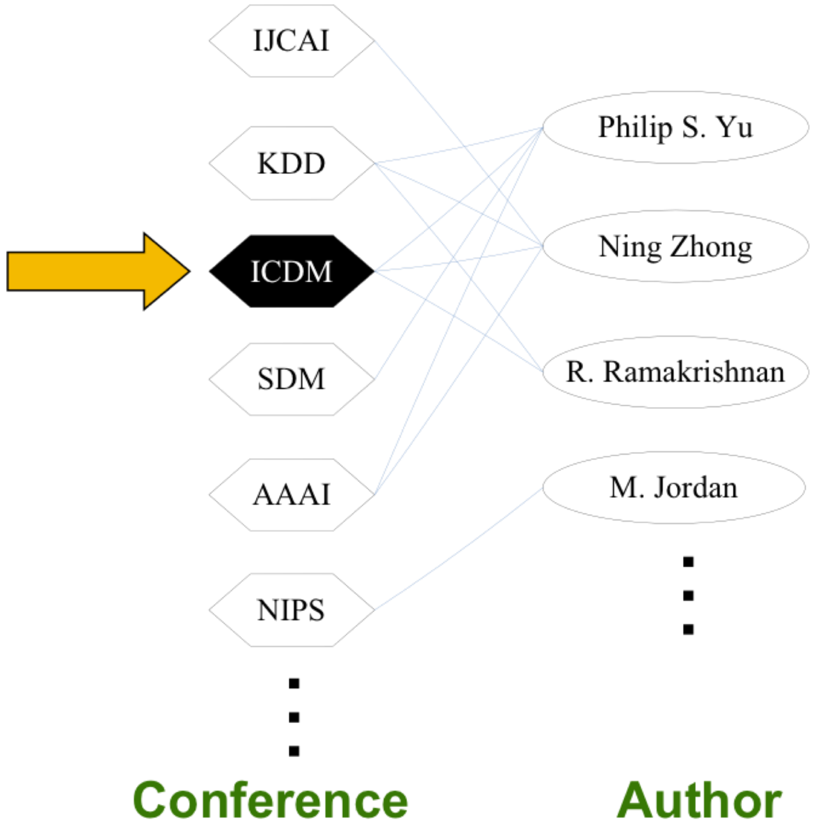
t=1	2	3	4	5
1.0000	0.1500	0.4109	0.2403	0.3404
0	0	0.1405	0.0467	0.1262
0	0.2833	0.1027	0.1510	0.1275
0	0.2833	0.0425	0.2358	0.1078
0	0.2833	0.1629	0.2795	0.1719
0	0	0.1405	0.0467	0.1262



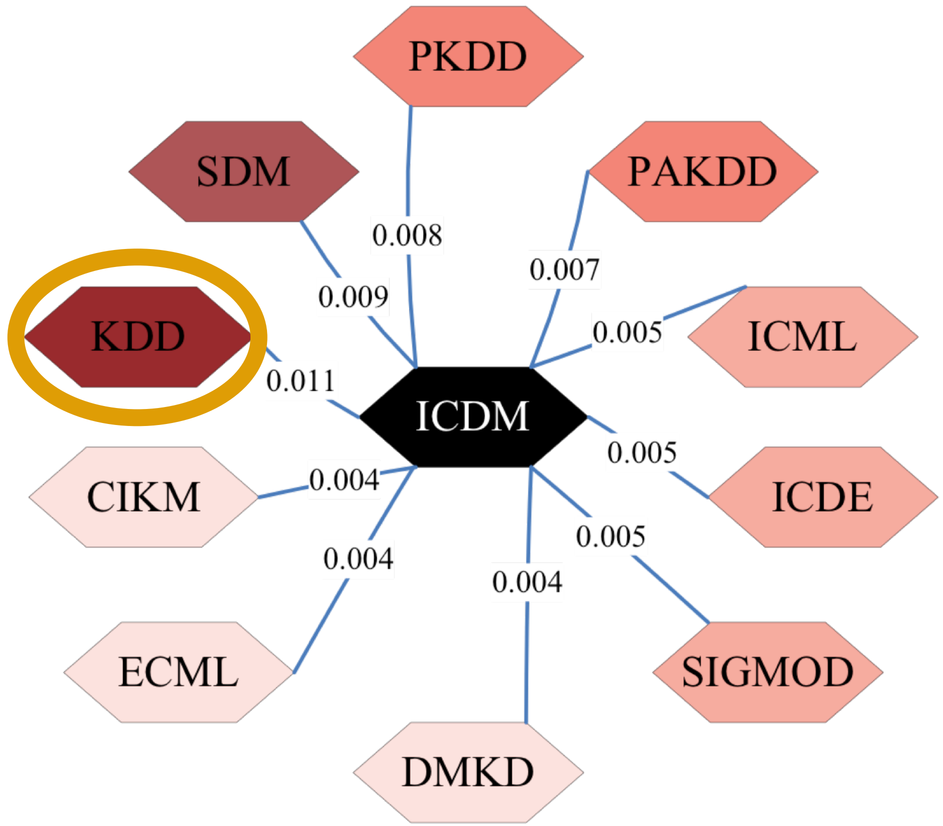
10	20	50	75	100	
0.2909	0.2985	0.2989	0.2989	0.2989	Giulia
0.0848	0.0926	0.0931	0.0931	0.0931	Marc
0.1309	0.1313	0.1314	0.1314	0.1314	Oliver
0.1763	0.1645	0.1638	0.1638	0.1638	Thomas
0.2324	0.2204	0.2197	0.2197	0.2197	Sarah
0.0848	0.0926	0.0931	0.0931	0.0931	Anna

Example

What is the most related conference to ICDM?



Top 10 ranking results

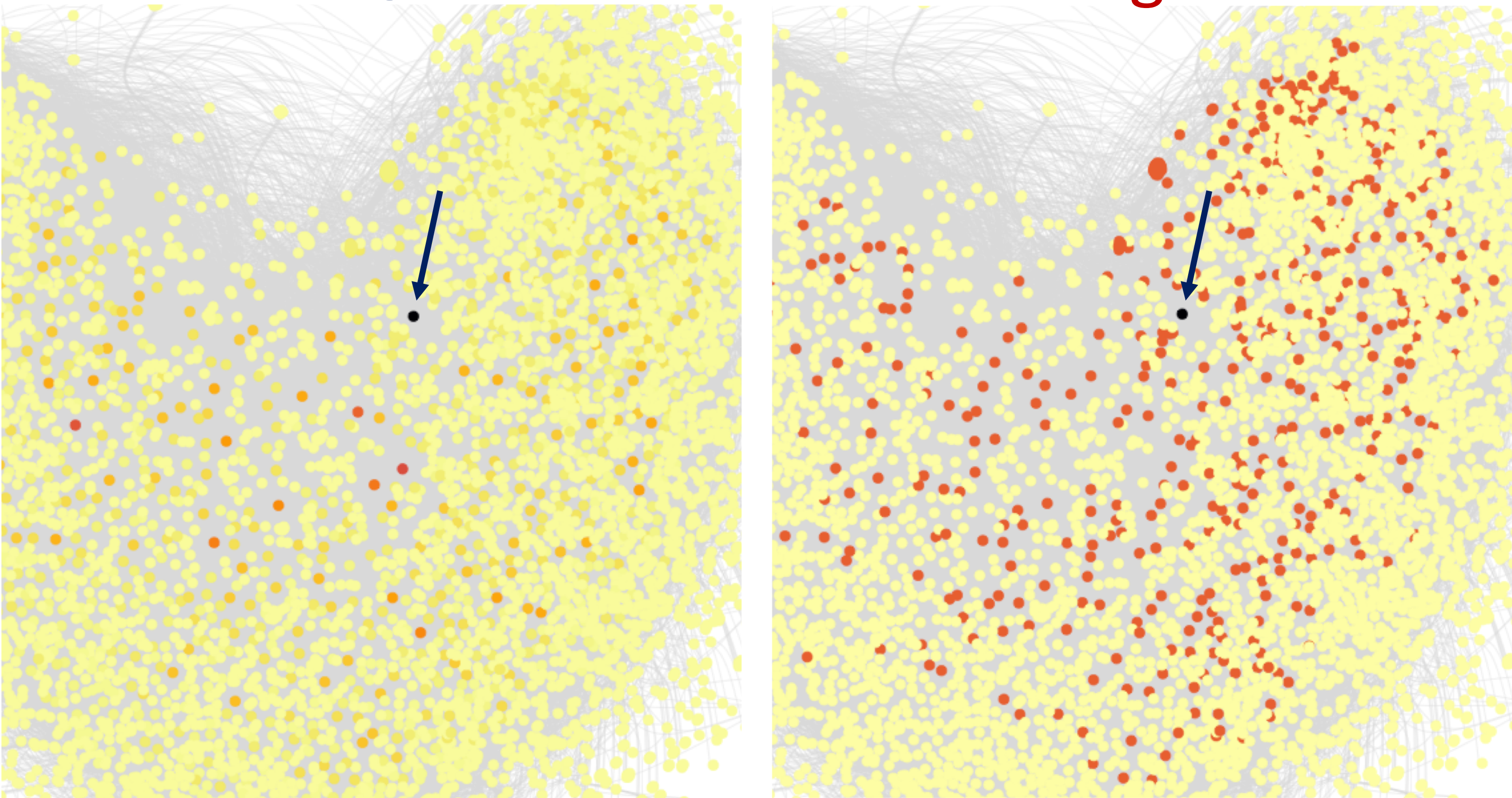


ICDM = international conf. on data mining
KDD = knowledge discovery and data mining

Local PageRank (authorities , *A*)

Local PageRank

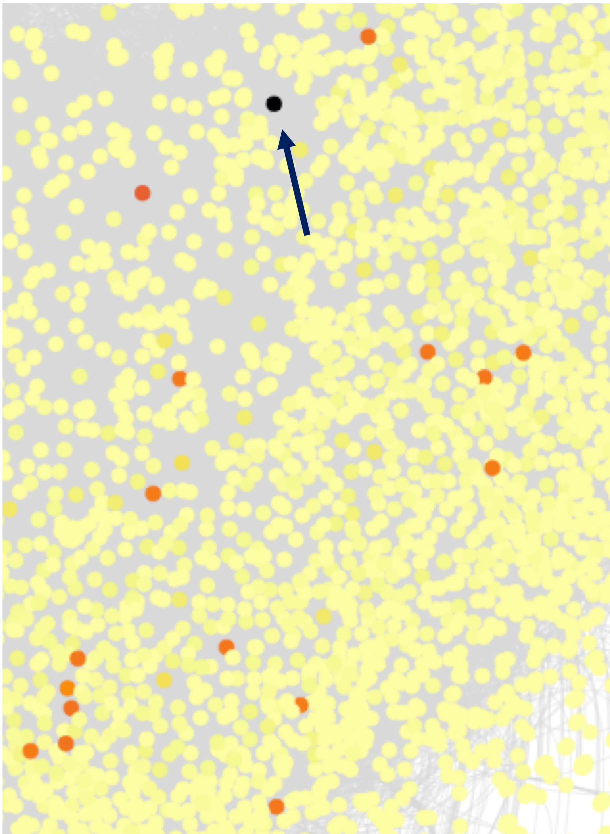
1-hop
out-neighbours



neighbours authority score =
local node → neighbours

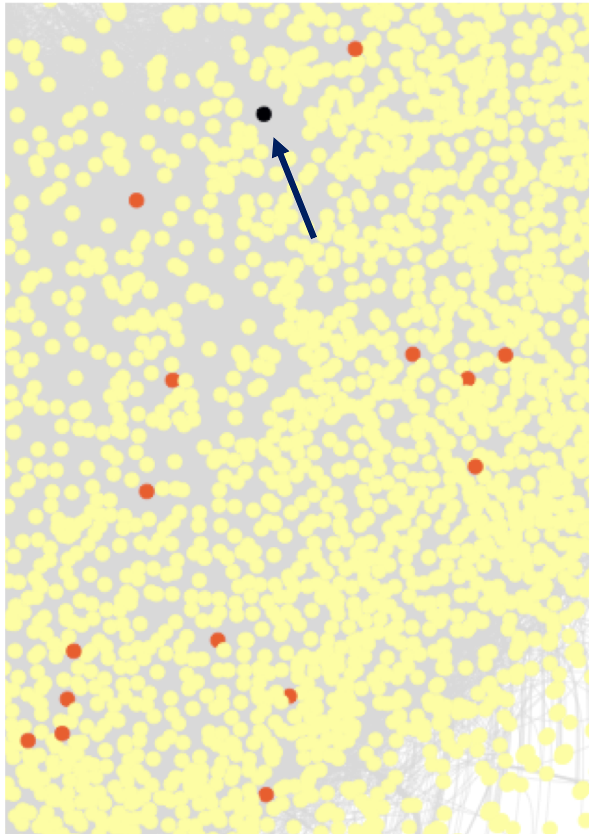
Local PageRank (hubs, A^T)

Local PageRank



neighbours hub score =
neighbours → local node

1-hop
in-neighbours



How can we use PageRank?

Want to measure proximity/similarity to a node? Local PageRank

Want to know about a specific topic? TopicSpecific PageRank

… appropriately select your teleport vector ***q*** !

MIME.

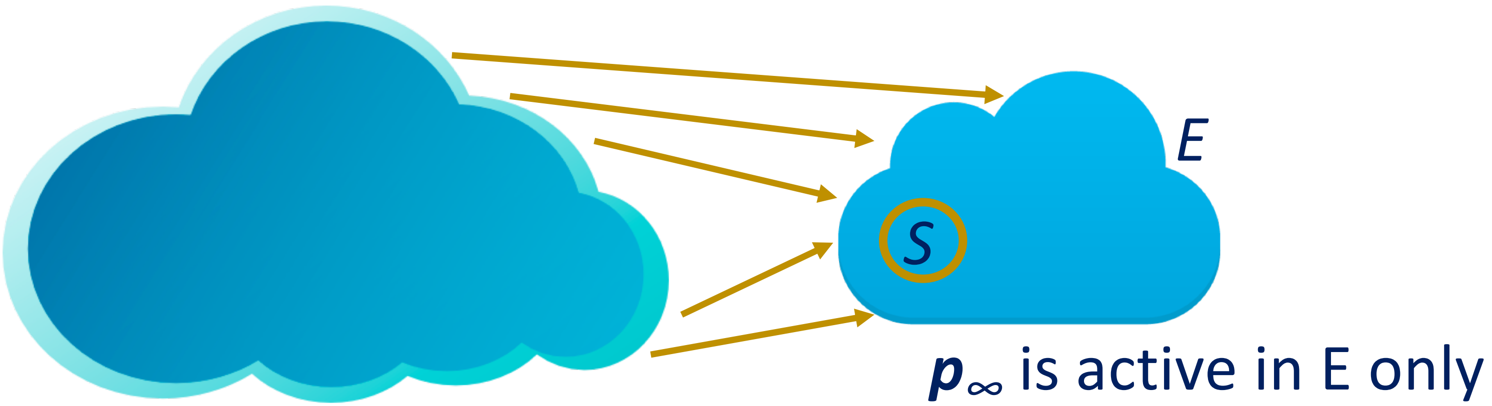
Topic specific PageRank

Idea

- Bias the random walk towards a topic specific teleport set *S* of nodes, i.e., make sure that $\boldsymbol{q}$ is active in *S* only
- *S* should contain only pages that are relevant to the topic

Result

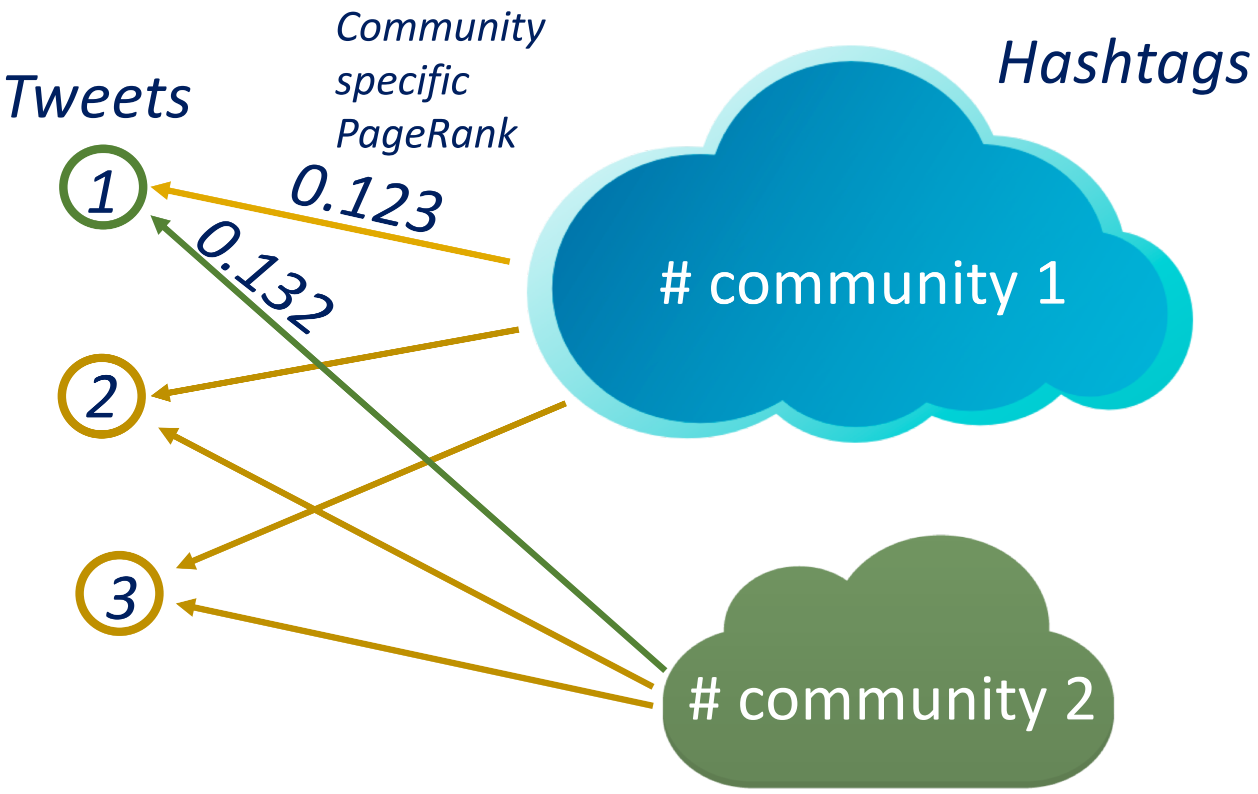
- The random walk deterministically ends in a small set *E*, containing *S*, and being in some sense close to it



Example: tweets & hastag communities

Idea

- Assign a tweet to the # community it is closer to



Closeness centrality

What is Closeness?

Closeness centrality

From Wikipedia, the free encyclopedia

In a connected graph, **closeness centrality** (or **closeness**) of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the *closer* it is to all other nodes.

Closeness was defined by Bavelas (1950) as the reciprocal of the **farness**,[1][2] that is:

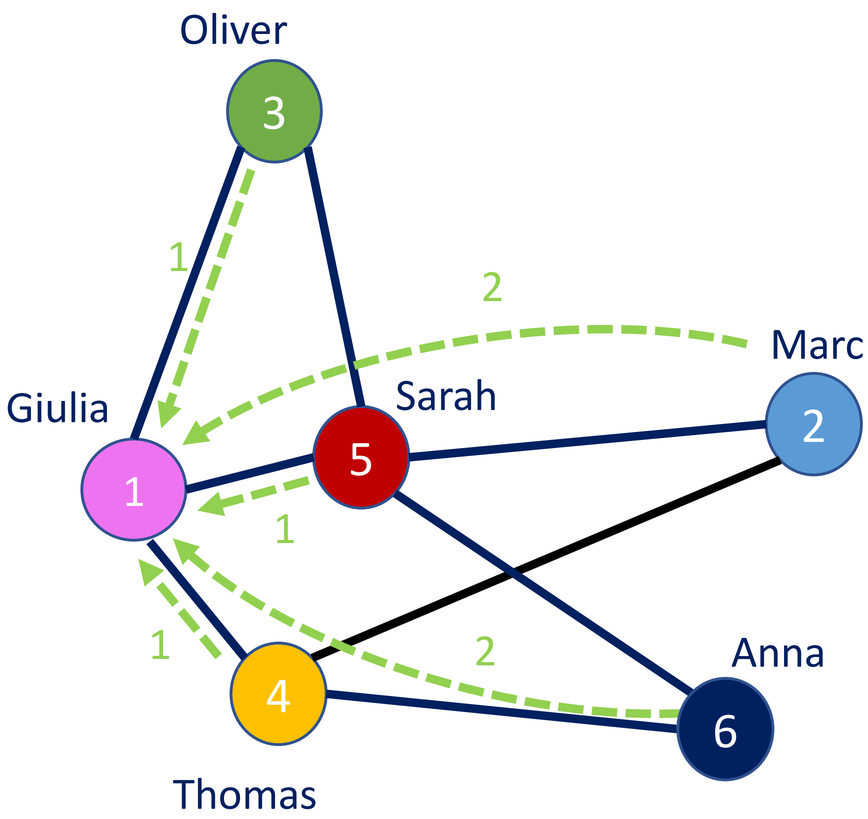
$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

where $d(y,x)$ is the distance between vertices x and y.

Rationale: the node which is the easiest to reach, the one which is the best for spreading information

Example

count the lengths of the shortest paths
leading to Giulia
1 + 2 + 1 + 2 + 1 = 7



Closeness

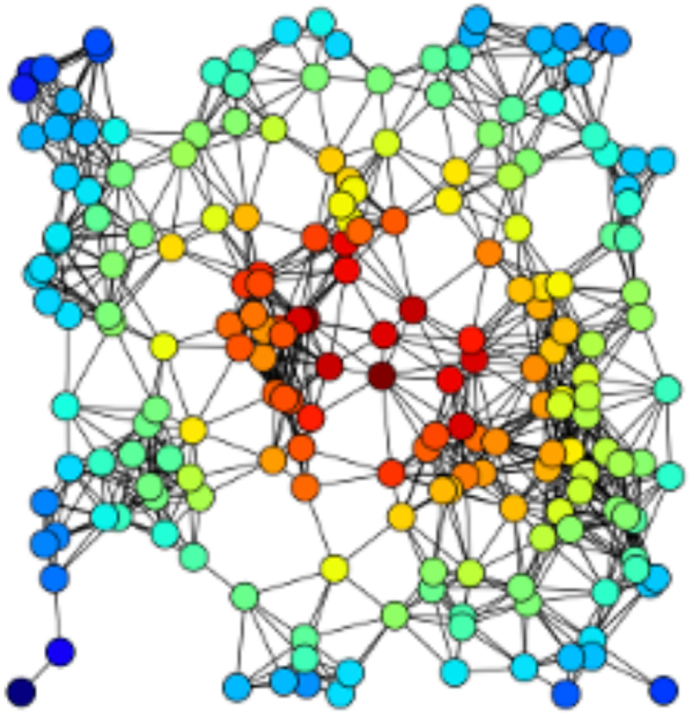
0.1429	Giulia
0.1250	Marc
0.1250	Oliver
0.1429	Thomas
0.1667	Sarah
0.1250	Anna

Sarah is the preferred node for spreading information

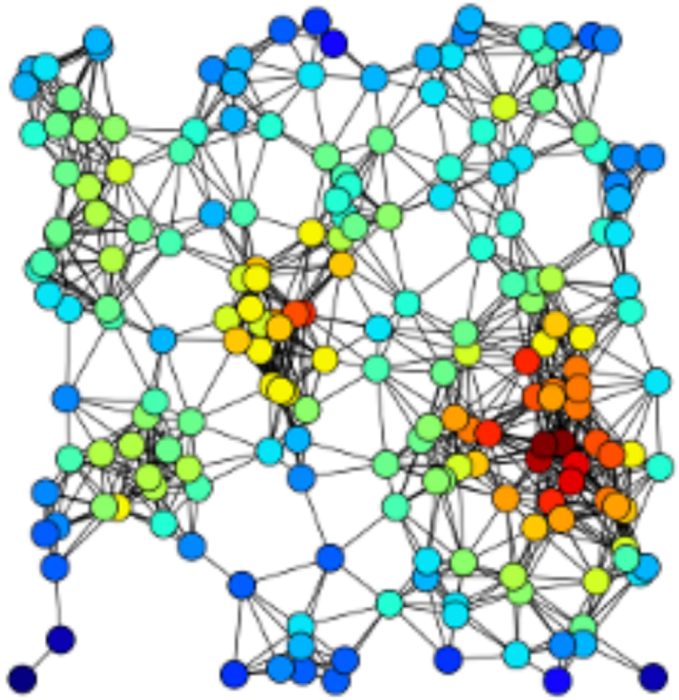
C(Giulia) = 1/7
= 0.1429

Closeness versus Degree centrality

Closeness



Degree



Harmonic centrality

In disconnected graphs [edit]

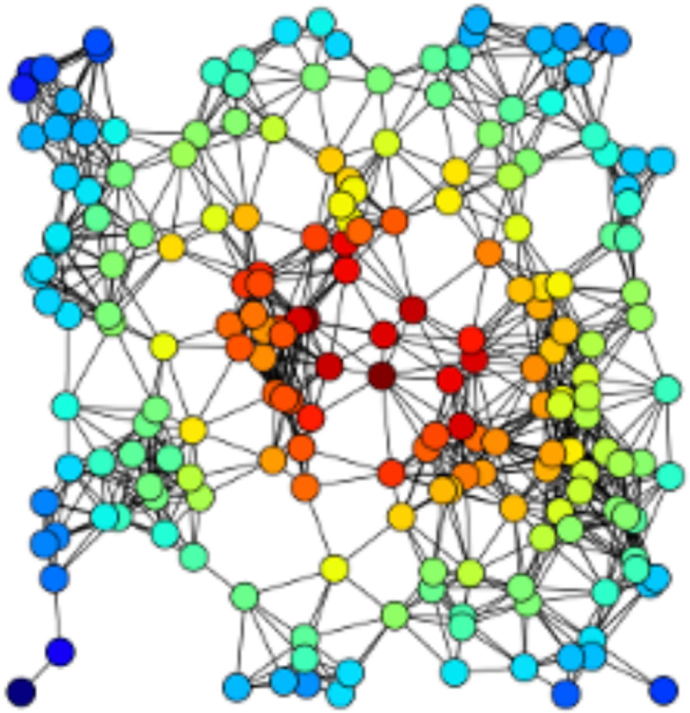
When a graph is not strongly connected, a widespread idea is that of using the sum reciprocal of distances, instead of the reciprocal of the sum of distances, with the convention $1/\infty = 0$:

$$H(x) = \sum_{y \neq x} \frac{1}{d(y, x)}.$$

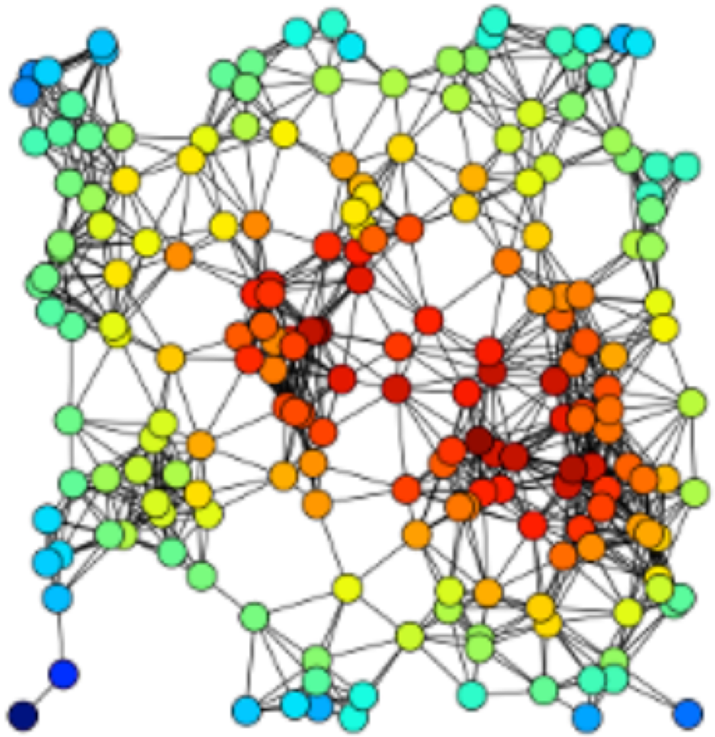
The most natural modification of Bavelas's definition of closeness is following the general principle proposed by Marchiori and Latora (2000)[3] that in graphs with infinite distances the harmonic mean behaves better than the arithmetic mean. Indeed, Bavelas's closeness can be described as the denormalized reciprocal of the arithmetic mean of distances, whereas harmonic centrality is the denormalized reciprocal of the harmonic mean of distances.

Closeness versus Harmonic centrality

Closeness



Harmonic



Betweenness centrality

What is Betweenness?

Betweenness centrality

From Wikipedia, the free encyclopedia

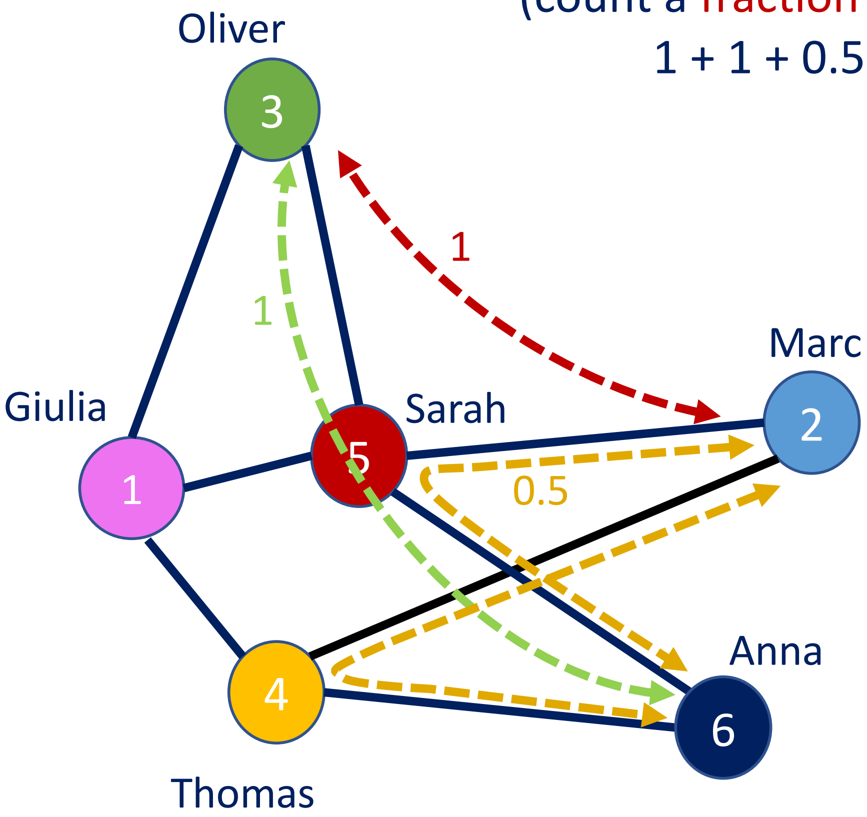
In graph theory, **betweenness centrality** is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.

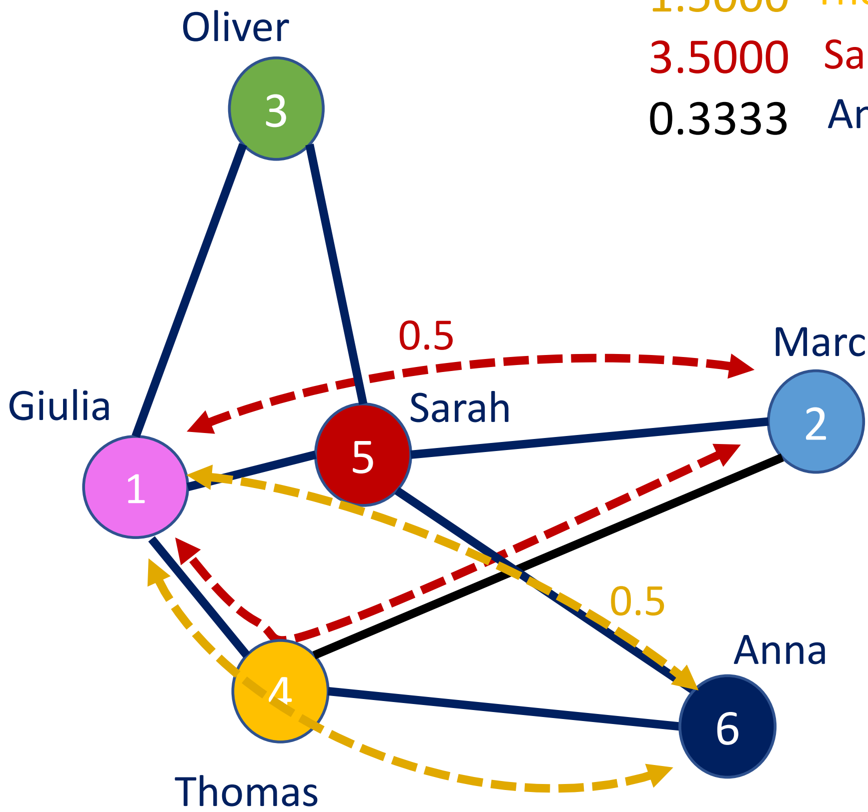
Betweenness centrality was devised as a general measure of centrality:[1] it applies to a wide range of problems in network theory, including problems related to social networks, biology, transport and scientific cooperation. Although earlier authors have intuitively described centrality as based on betweenness, Freeman (1977) gave the first formal definition of betweenness centrality.

Rationale: the node which takes you elsewhere (bridge, broker)

Example

count the # of shortest paths
passing through Sarah
(count a fraction if more than one path)
1 + 1 + 0.5 + 0.5 + 0.5 = 3.5



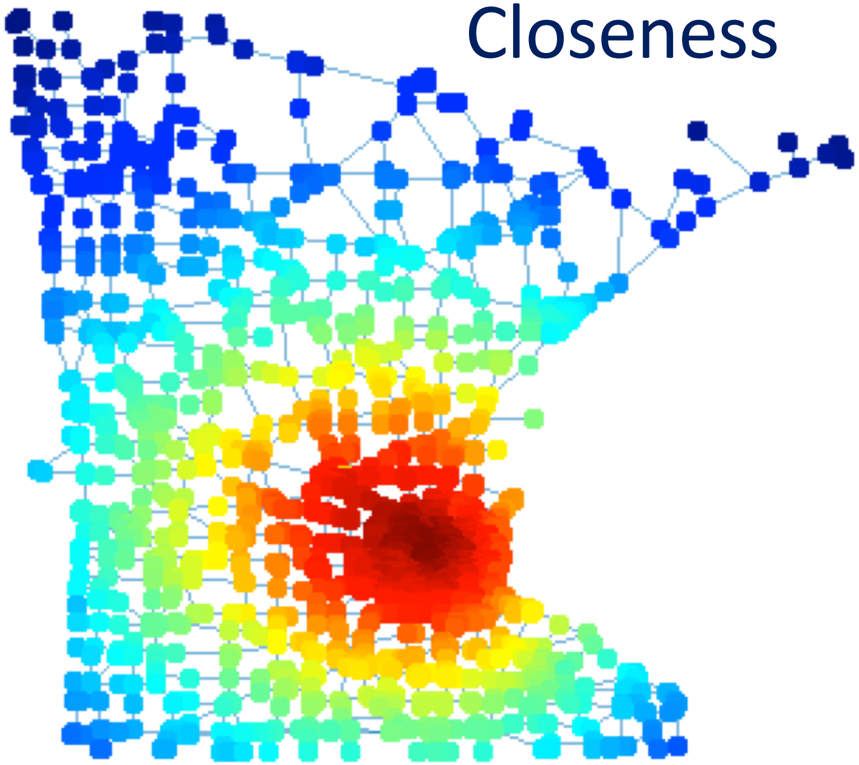


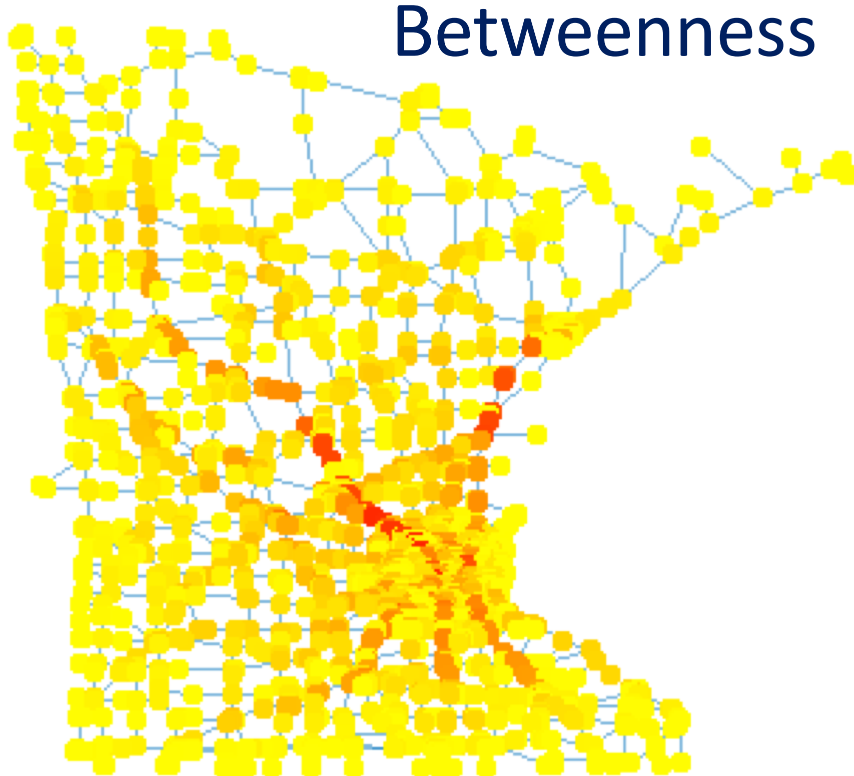
Betweenness

1.3333	Giulia
0.3333	Marc
0	Oliver
1.5000	Thomas
3.5000	Sarah
0.3333	Anna

Closeness versus Betweenness centrality

Minnesota road network



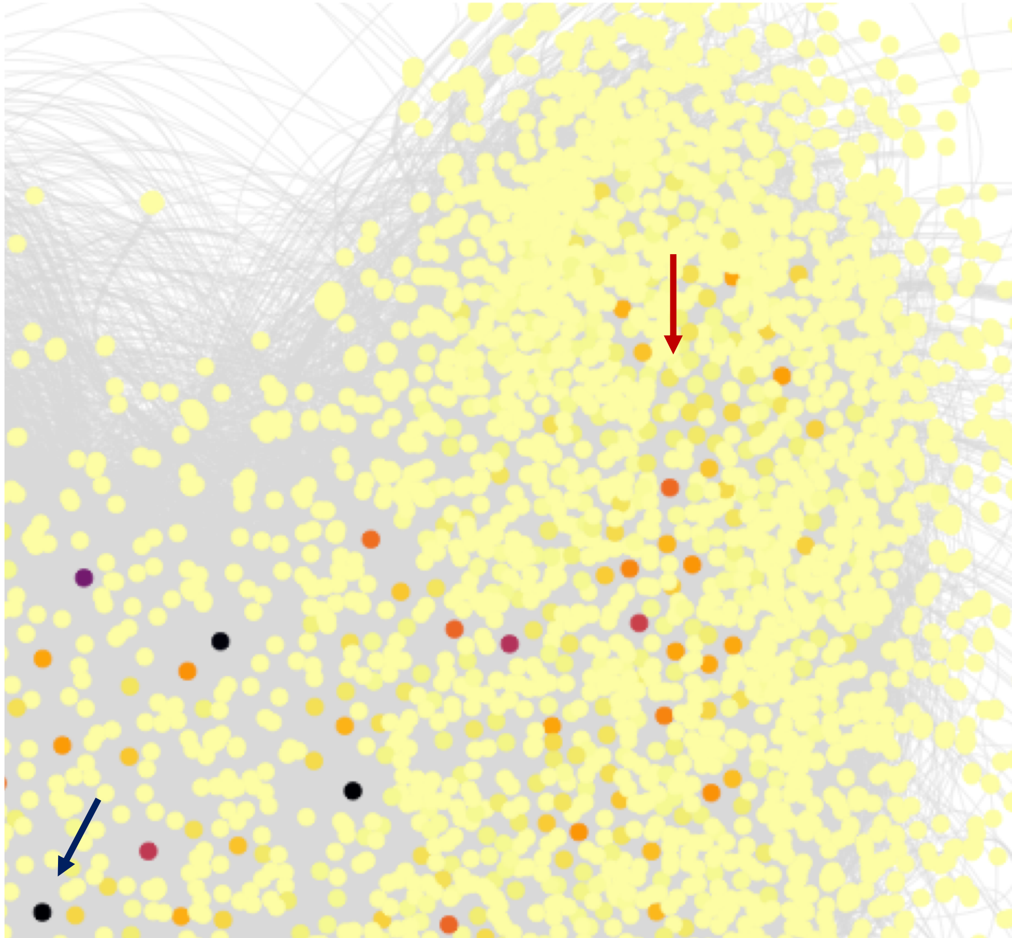


Closeness is a measure of center of gravity (best node from which to spread info)

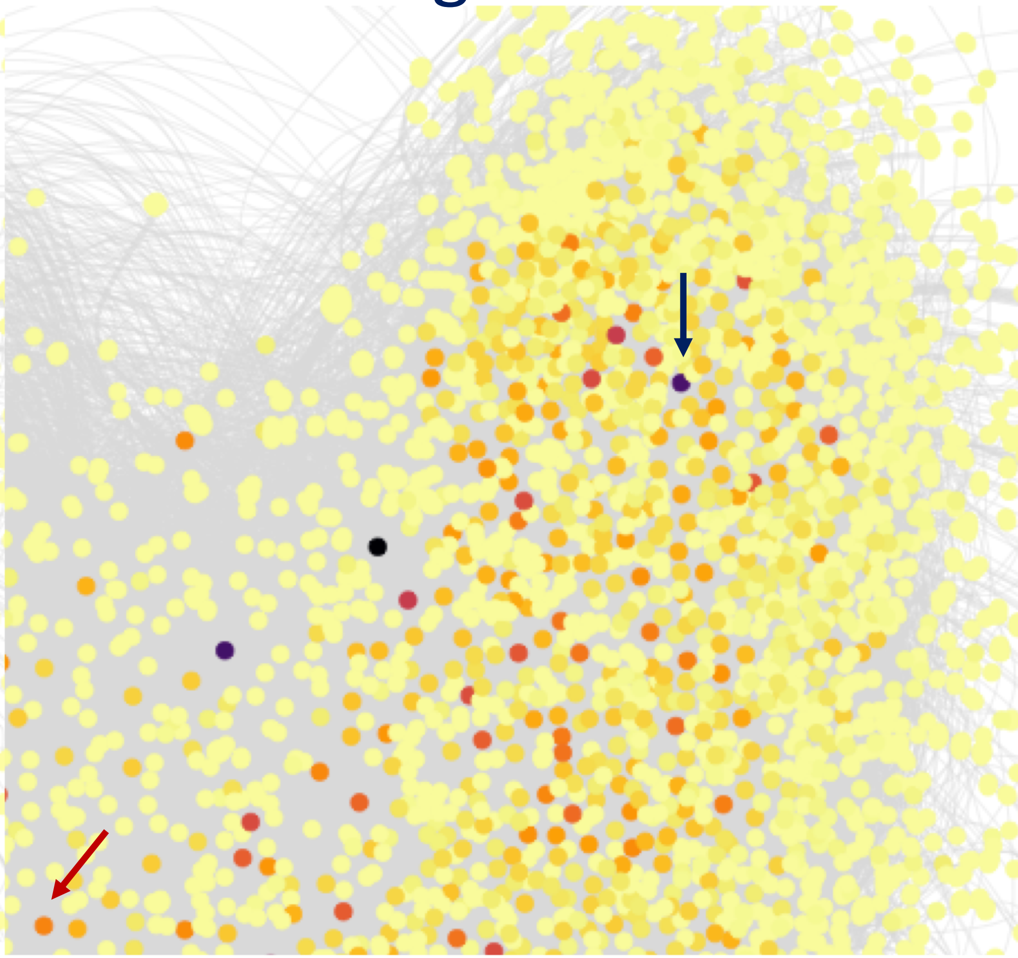
Betweenness is a measure of brokerage (i.e., being a bridge)

Betweenness versus PageRank centrality

Betweenness



PageRank



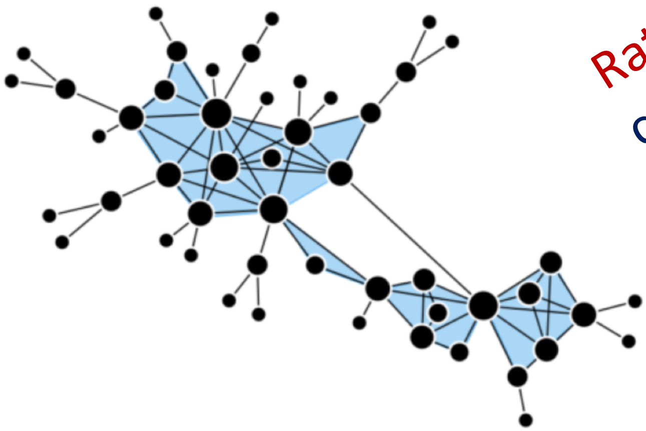
MIME

Clustering coefficient

What is the Clustering coefficient?

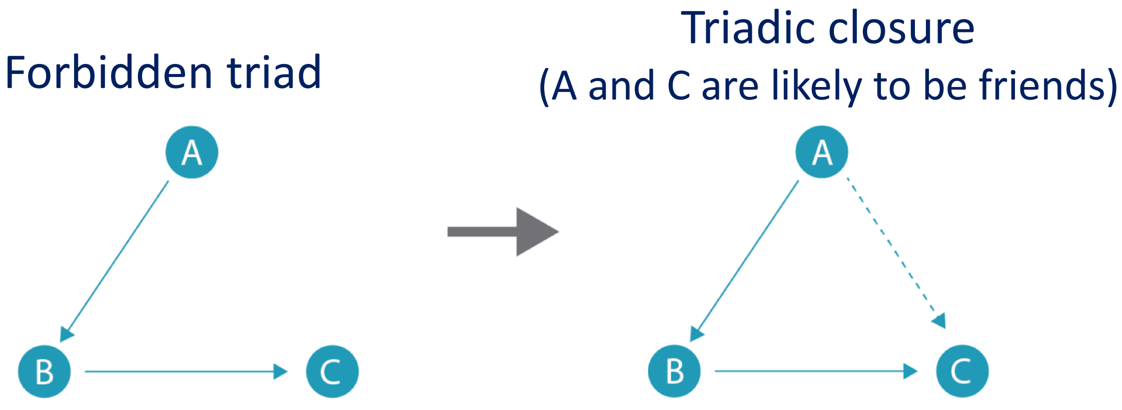
Local clustering coefficient [edit]

The **local clustering coefficient** of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph). Duncan J. Watts and Steven Strogatz introduced the measure in 1998 to determine whether a graph is a small-world network.



Rationale: how strongly connected is the network locally / general indication of the graph's tendency to be organized into clusters

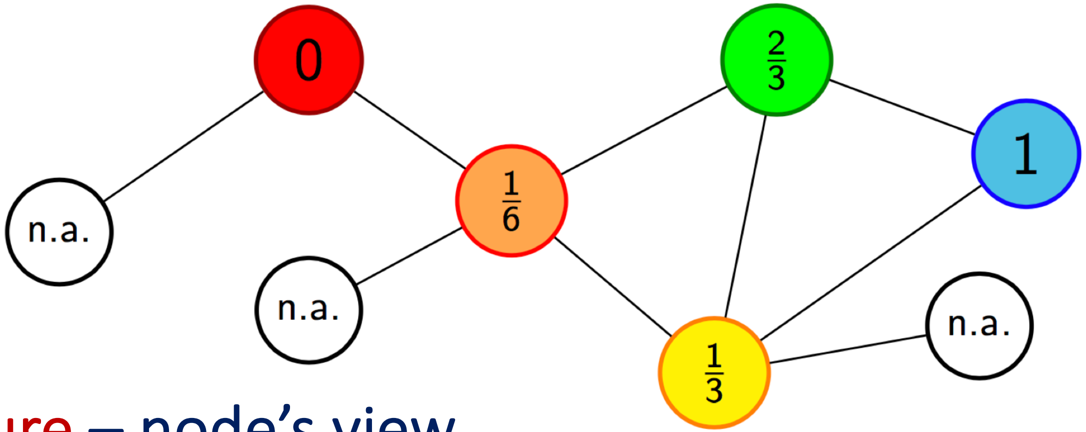
Triadic closure



Triadic closure

- A and C are likely to have the opportunity to meet because they have a common friend B
- The fact that A and C is friends with B gives them the basis of trusting each other
- B may have the incentive to bring A and C together, as it may be hard for B to maintain disjoint relationships

Clustering coefficient and triadic closure



A measure for triadic closure – node's view

- Clustering coefficient C_i
- Counts the fraction of pairs of neighbours which form a triadic closure with node *i*

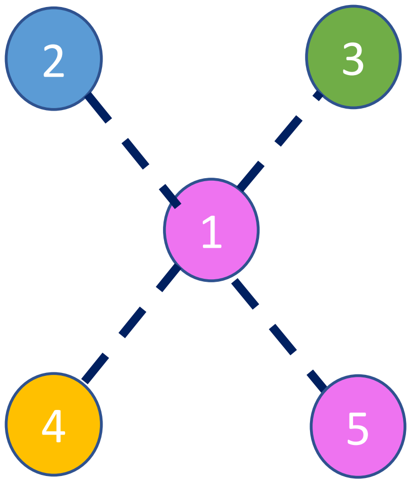
$$C_i = \frac{1}{|\mathcal{N}_i|(|\mathcal{N}_i|-1)} \sum_{\substack{(j,k)\in\mathcal{N}_i^2 \\ j\neq k}} \mathrm{tc}_{i,j,k}$$

where $tc_{ijk} = 1$ if the triplet (i,j,k) forms a triadic closure, and zero otherwise

Examples

not connected neighbourhood

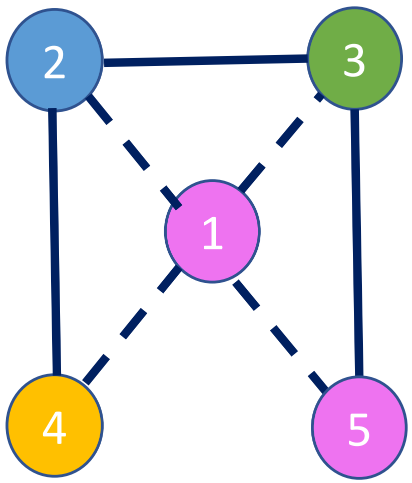
$C_1 = 0$



$\langle C \rangle = 0$

weakly connected neighbourhood

$C_1 = ½ = 3 / 6$

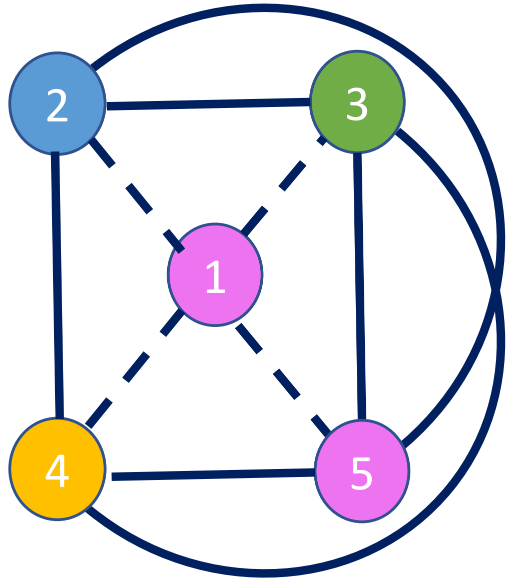


$C_2 = C_3 = ⅔, C_4 = C_5 = 1$

$\langle C \rangle = 0.766$

strongly connected neighbourhood

$C_1 = 1 = 6 / 6$

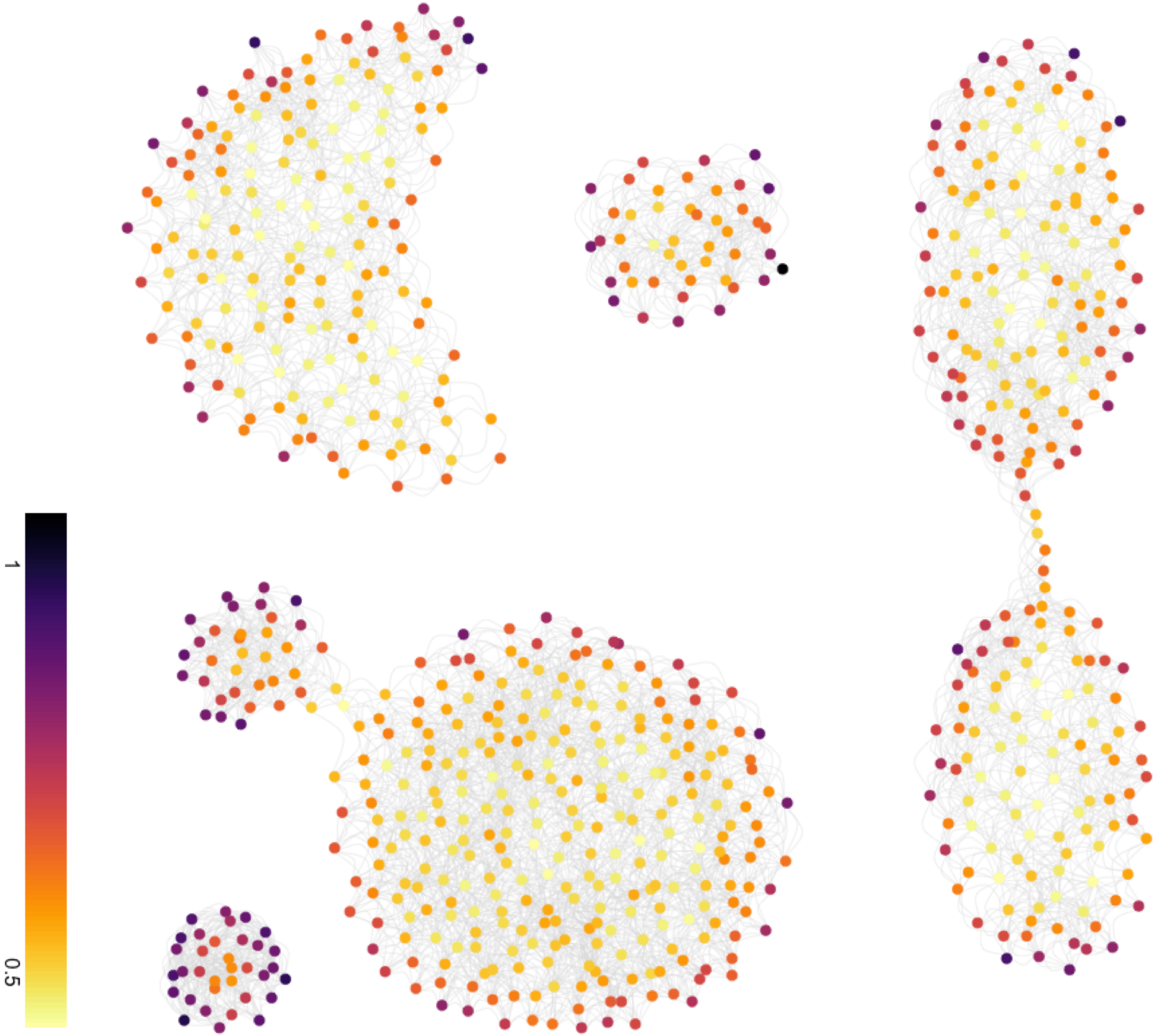


$\langle C \rangle = 1$

Warning

But clustering coefficient is generally hard to see and visual interpretation is considered unreliable

Visual example



Wrap up

Centrality measures wrap-up

Centrality measure	Technical property	Meaning
Degree (in/out)	Measures number (and quality) of connections	Cohesion, Self-monitoring, Entrepreneurship, Extraversion, Popularity
PageRank (authorithies/hubs)	Measures number (and quality) of direct and indirect connections	Cohesion, Self-monitoring, Entrepreneurship, Extraversion, Authority, Closeness/Similarity/Friendship (with a direction), Dependence
Closeness	Measures length of min paths	Visual centrality, Influence, Significant spreading points, Outliers
Betweenness	Measures number of min paths	Brokerage, Structural holes, Ostracism

Take-aways

https://reticular.hypotheses.org/1745