Map Equation Centrality: A Map Equation-based Community-Aware Centrality Score

NetSciX 2022

arXiv:2201.12590

Christopher Blöcker[1], Juan Carlos Nieves[2], Martin Rosvall[1]

christopher.blocker@umu.se

[1]Integrated Science Lab
Department of Physics
Umeå University

[2]Department of Computing Science
Umeå University

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

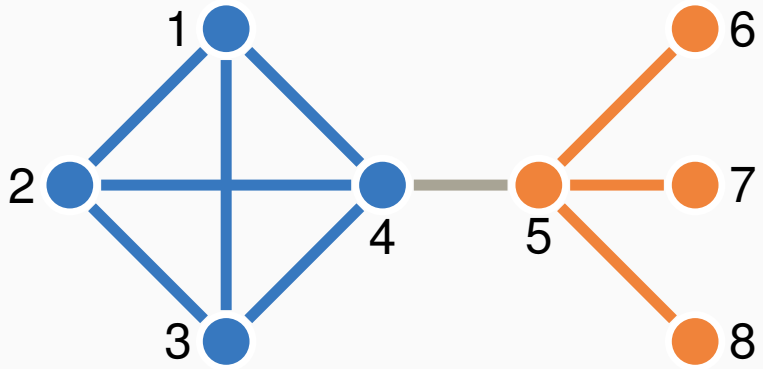
UMEÅ UNIVERSITY

icelab INTEGRATED SCIENCE LAB

Motivation

Question

How influential are the nodes? Does 4 or 5 have more influence?

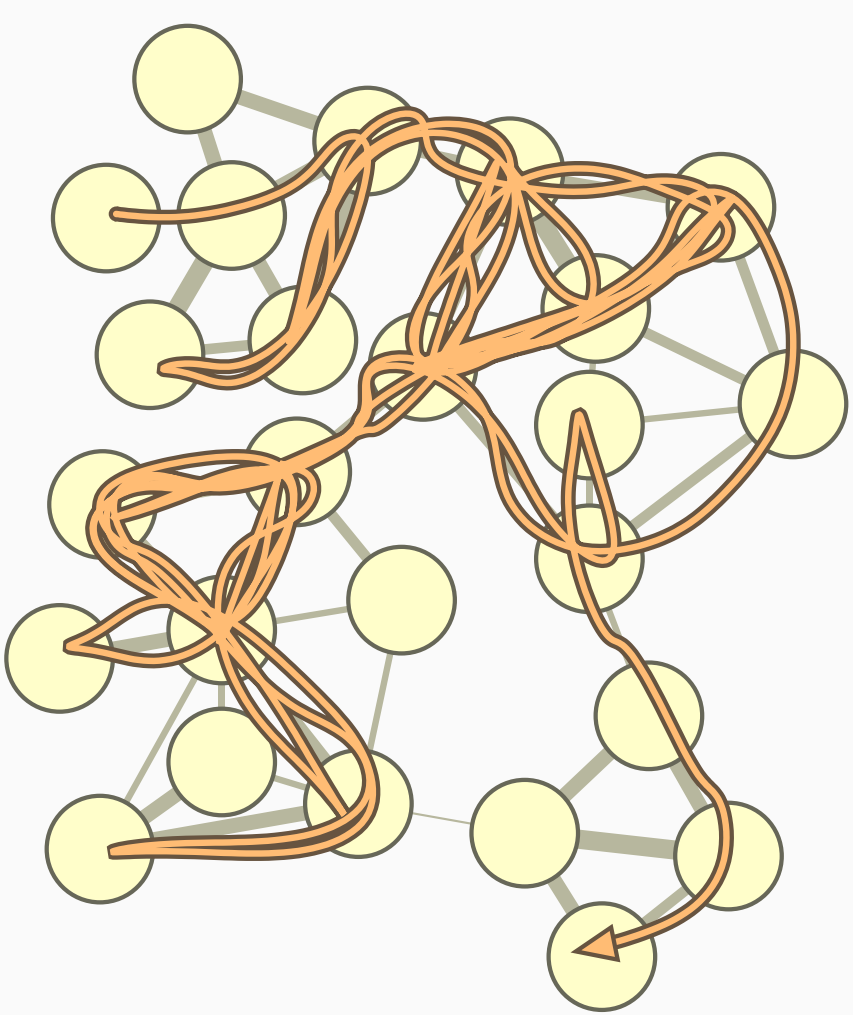


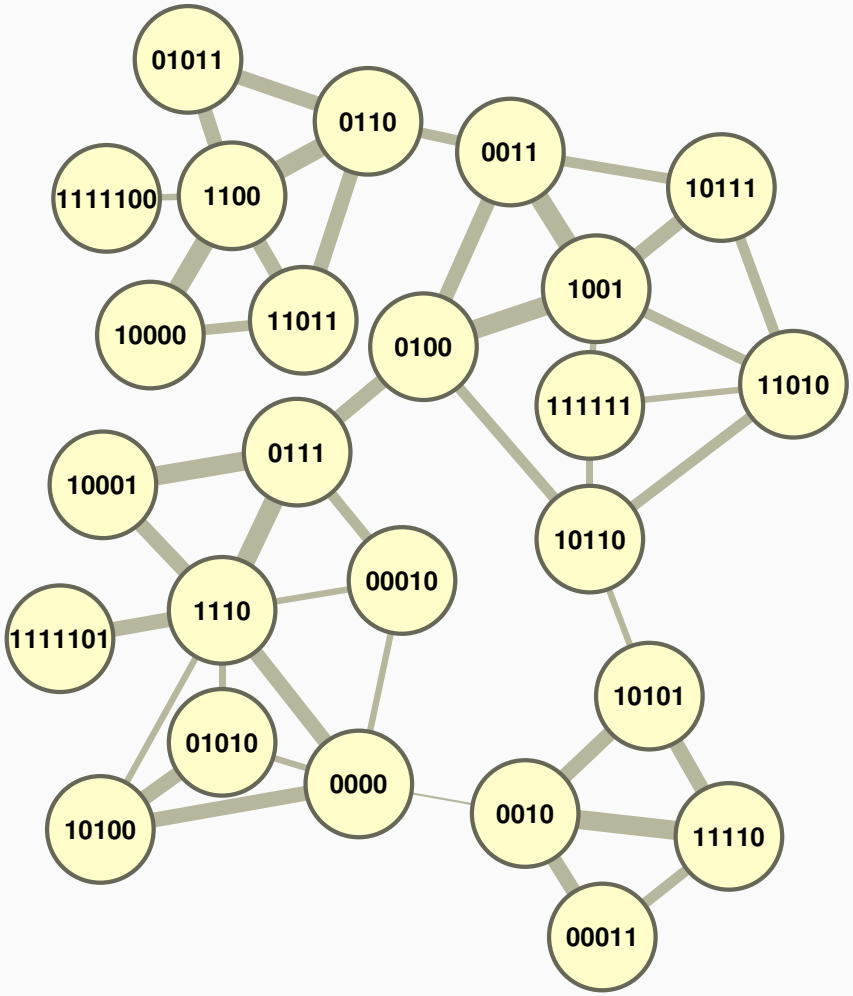
Idea

Use communities to determine influence.

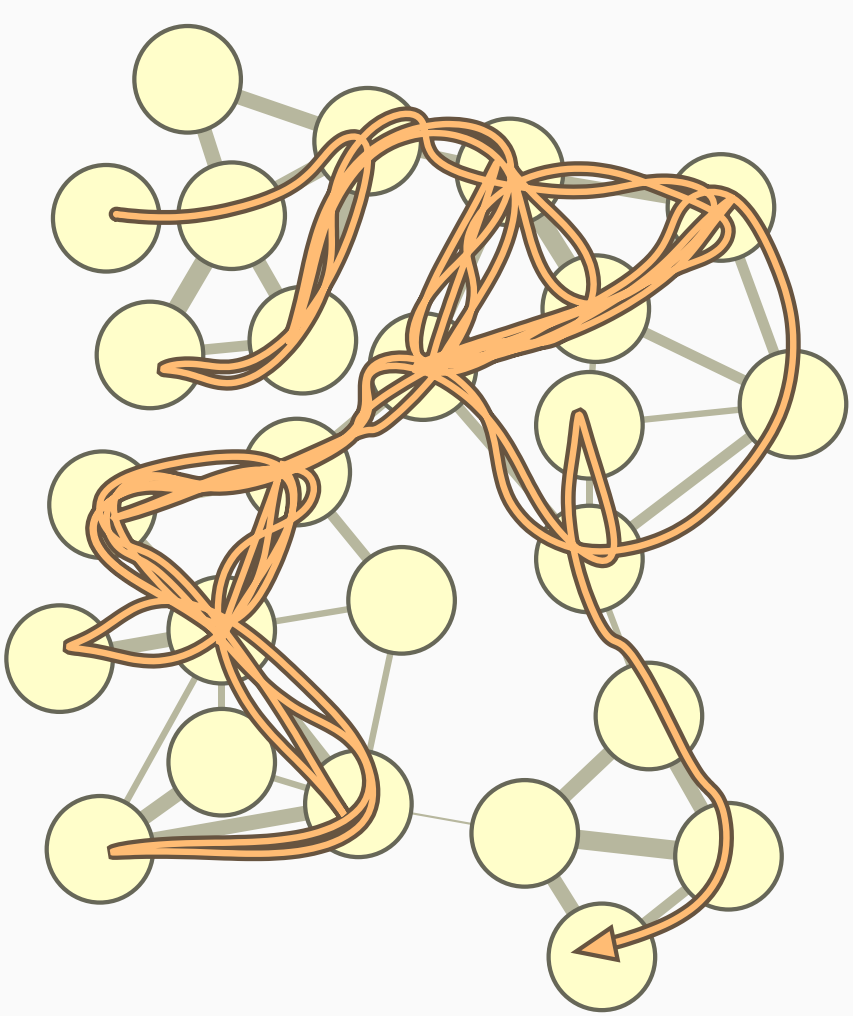
Background:
The Map Equation Framework

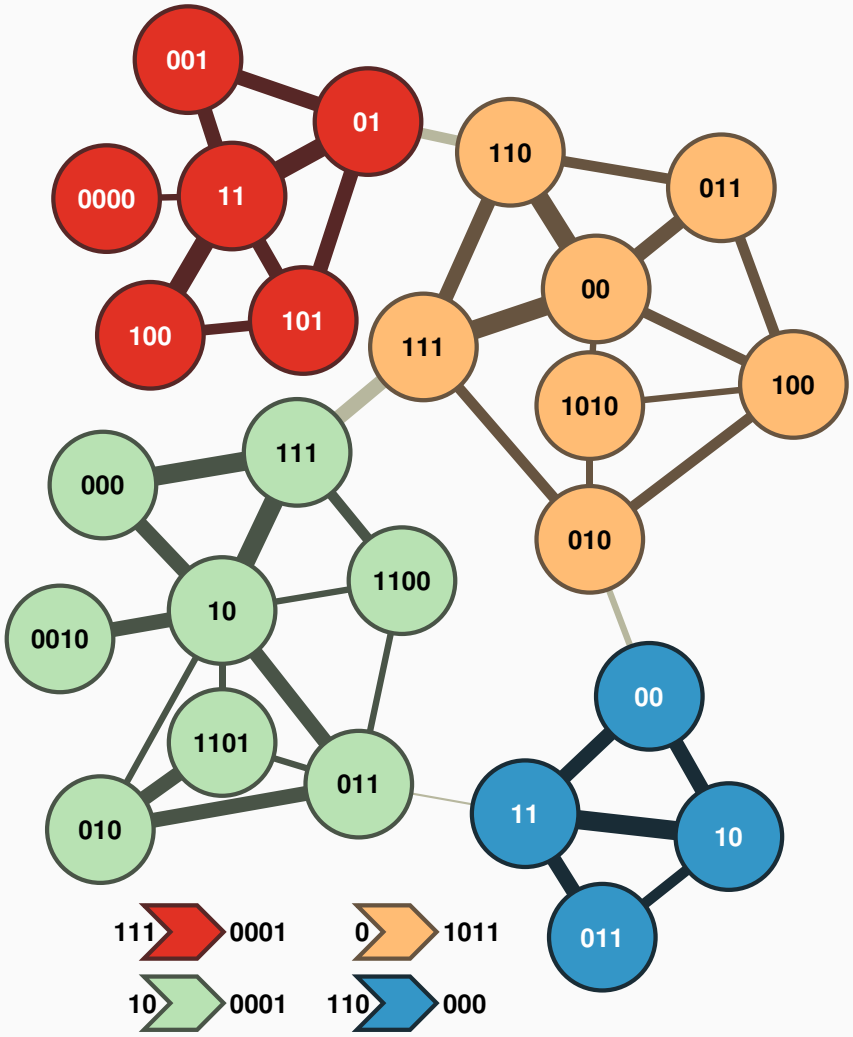
The Map Equation Framework





The Map Equation Framework





$$L(M) = qH(\mathcal{Q}) + \sum_{m \in M} p_m H(\mathcal{P}_m)$$

Network Vitality

Network Vitality

Network Vitality

Given a graph G and a function f that operates on graphs, the vitality μ with respect to some node u is defined as (Koschützki 2005)

$$\mu(G, u) = f(G) - f(G - \{u\}),$$

where $G - \{u\}$ denotes the graph G with node u removed.

This was used to define *modularity vitality* (Magelinski 2021)

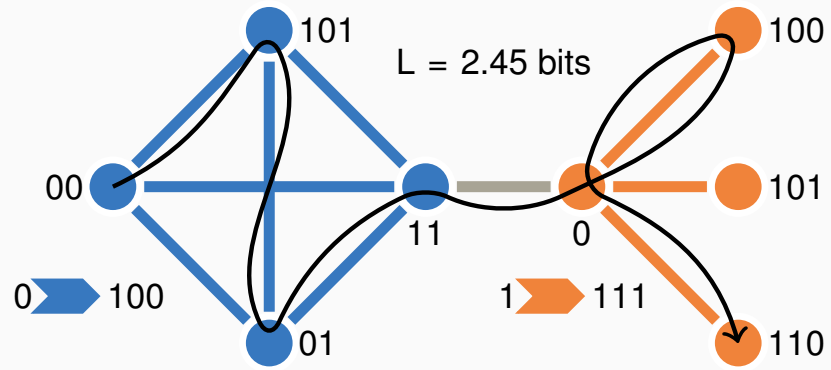
We take a slightly different approach with the map equation

- to keep the visit rates and modules unchanged, we do **not** remove nodes
- instead we "silence" them, that is we ignore them when describing the random walk
  $\rightarrow$ the per-step description length changes
- Idea: silencing a more important nodes has a larger effect

Silencing Nodes

Silencing Nodes

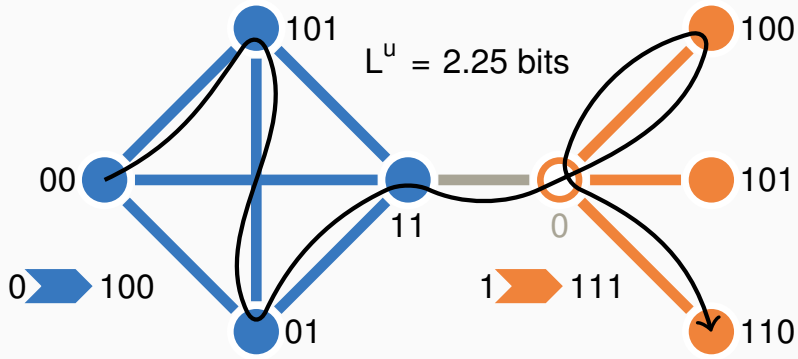
This is the starting point, nothing happened yet.



sender: 0 00 101 01 11 100 1 0 100 0 110 (22 bits)

Silencing Nodes

(a) use the same code but don't encode visits to the silenced node

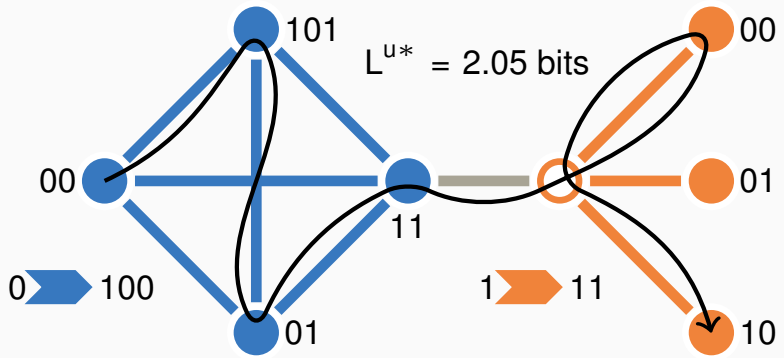


sender: 0 00 101 01 11 100 1 100 110 (20 bits)

$$L^u(G,\mathrm{M}) = \overbrace{qH(Q)}^{\text{index level}} + \overbrace{\sum_{m\in\mathrm{M},m\neq m_u} p_m H(P_m)}^{\text{modules without } u}$$

$$- \overbrace{\sum_{p\in P_{m_u}\setminus\{p_u\}} p\log_2\frac{p}{p_{m_u}}}^{\text{module with } u}$$

(b) design a new code without a codeword for the silenced node



sender: 0 00 101 01 11 100 1 00 10 (18 bits)

$$L^{u*}(G,\mathrm{M}) = \overbrace{qH(Q)}^{\text{index level}} + \overbrace{\sum_{\mathrm{m}\in\mathrm{M},\mathrm{m}\neq m_u} p_{\mathrm{m}} H(P_{\mathrm{m}})}^{\text{modules without } u}$$

$$- \overbrace{\sum_{p\in P_{m_u}\setminus\{p_u\}} p\log_2\frac{p}{p_{m_u}-p_u}}^{\text{module with } u}$$

Define: $\lambda(G,\mathrm{M},u) = L^u(G,\mathrm{M}) - L^{u*}(G,\mathrm{M}) = -(p_{m_u}-p_u)\log_2\frac{p_{m_u}-p_u}{p_{m_u}}$

Map Equation Centrality

Map Equation Centrality

$$\lambda(G, \mathsf{M}, u) = -(p_{\mathsf{m}_u} - p_u) \log_2 \frac{p_{\mathsf{m}_u} - p_u}{p_{\mathsf{m}_u}}$$

Interpretation: map equation centrality captures the "harm" that a node causes to others by its existence. Without it, the other nodes could have shorter codewords.

This only affects nodes **within the same module**!

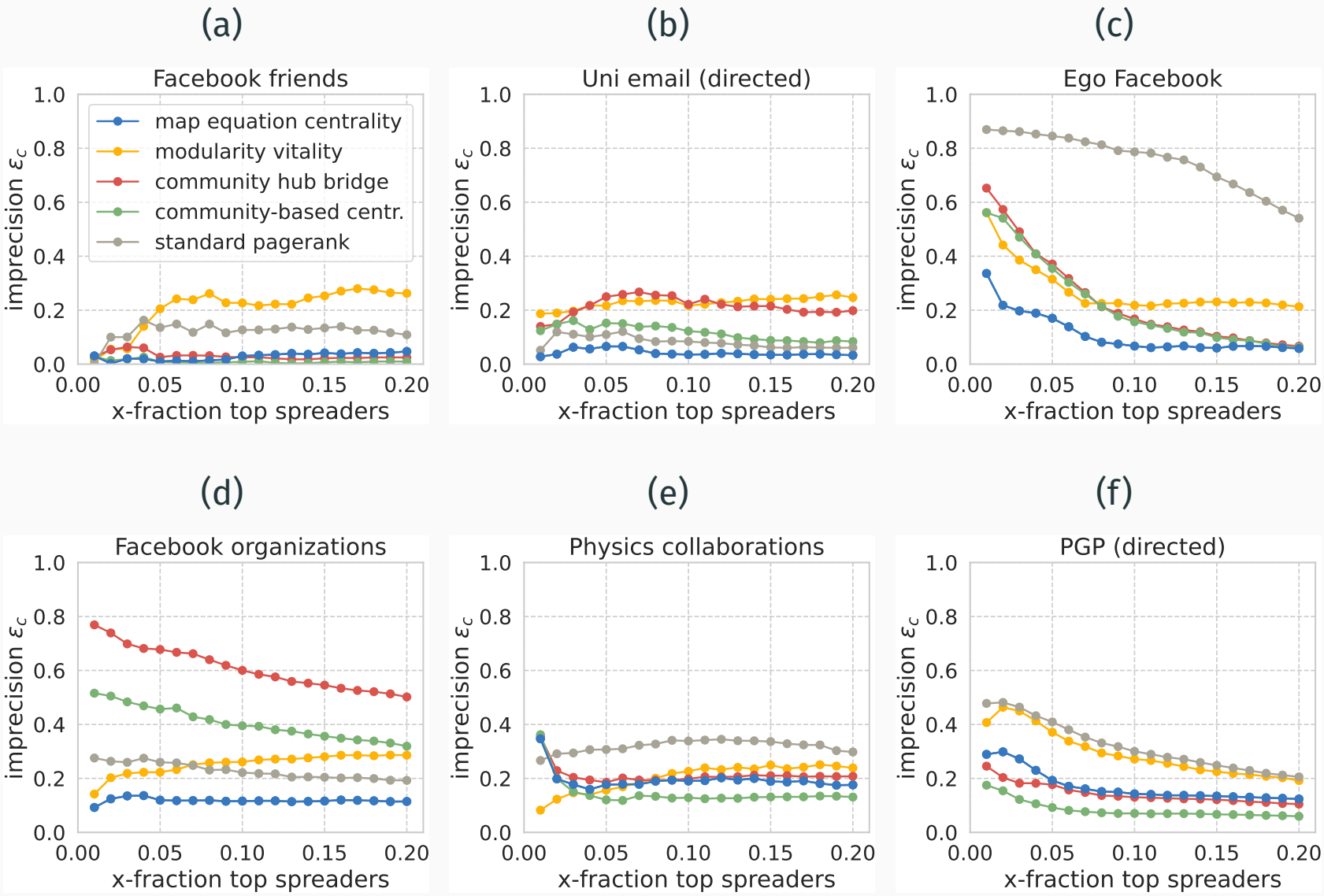
Advantage: map equation centrality is true to the map equation.

Evaluation

Evaluation

- Determine influence of nodes in real-world networks
- We follow the approach of (Rajeh 2021)
  - Node u's influence $\leftarrow$ SIR spreading power
    expected number of nodes that get infected when u is the single starting point of an SIR disease
  - Select the top spreaders according to centrality measure c and SIR
  - Measure performance using imprecision $\epsilon_c(x) = 1 - \frac{M_c(x)}{M_{\text{SIR}}(x)}$
    - x: fraction of spreaders to identify
    - M_c: average spreading power for top x-fraction according to c
    - M_{SIR}: average spreading power for top x-fraction according to SIR
- Compare with 3 community-aware baselines and standard PageRank

Evaluation


(a)
Facebook friends
(b)
Uni email (directed)
(c)
Ego Facebook
(d)
Facebook organizations
(e)
Physics collaborations
(f)
PGP (directed)
map equation centrality
modularity vitality
community hub bridge
community-based centr.
standard pagerank
imprecision ϵ_c
x-fraction top spreaders
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20

Conclusion

Conclusion

Problem

We consider node centrality from a community-based perspective.

Our Contribution

We propose map equation centrality, an information-theoretic centrality measure based on the map equation.

Results

Map equation centrality performs well in determining influential nodes in real-world networks.

Acknowledgements

Thank you for your attention!