Characterizing the analogy between hyperbolic embedding and community structure of complex networks

Filippo Radicchi

filiradi@indiana.edu filrad.homelinux.org

INFORMATICS, COMPUTING, AND ENGINEERING

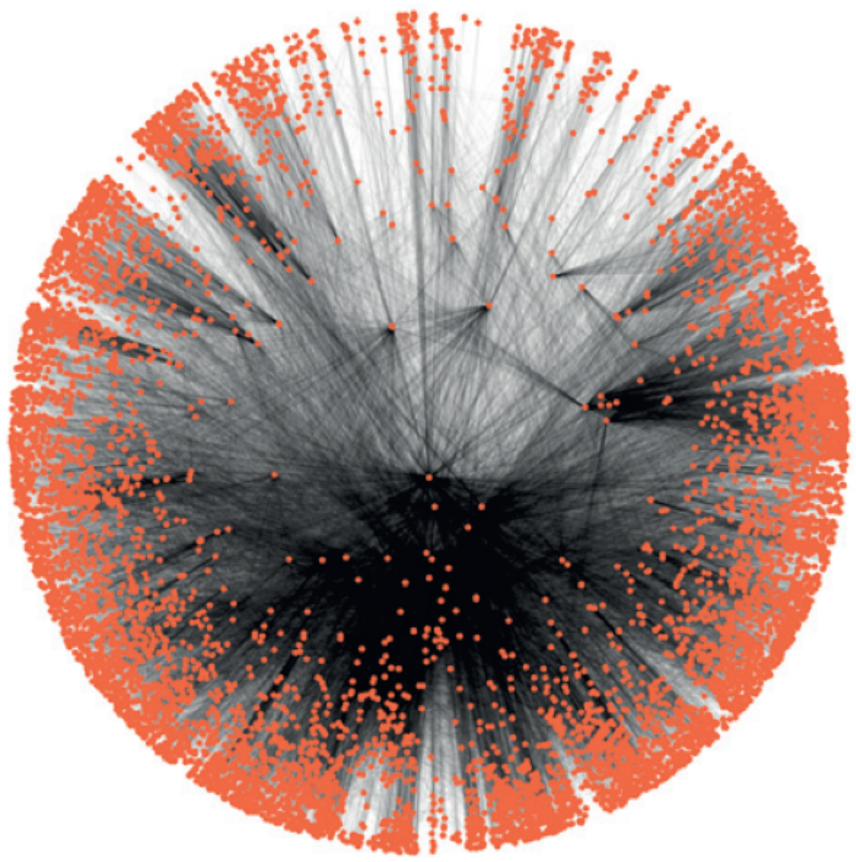
in collaboration with

A. Faqeeh and S. Osat

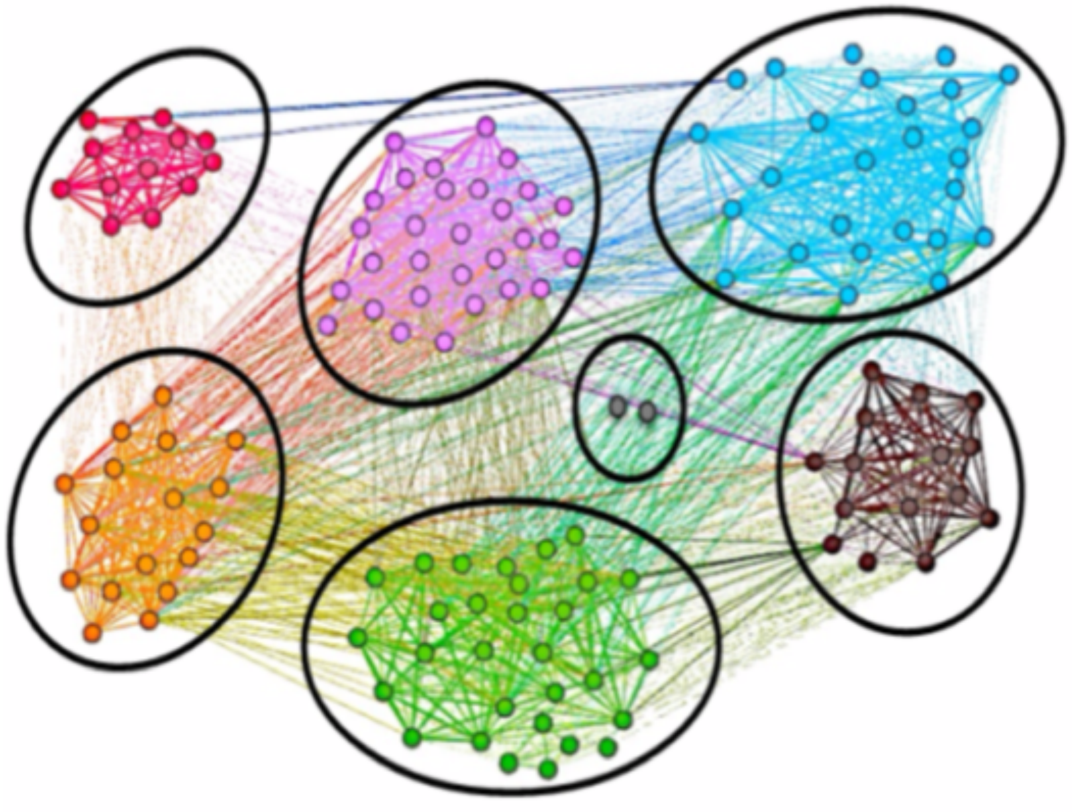
Supported by

Low-dimensional embedding of networks

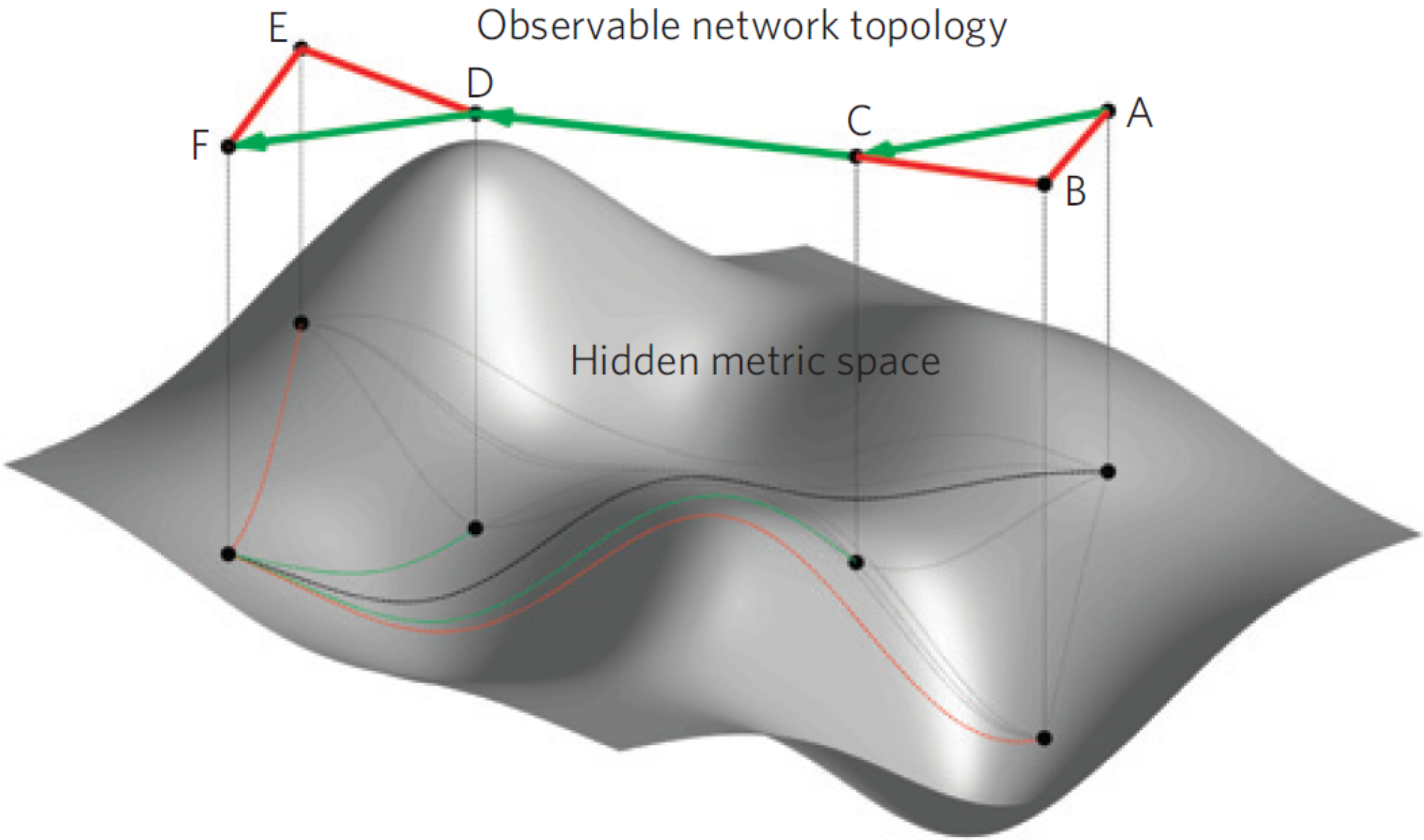
Embedding in the hyperbolic space



Community structure

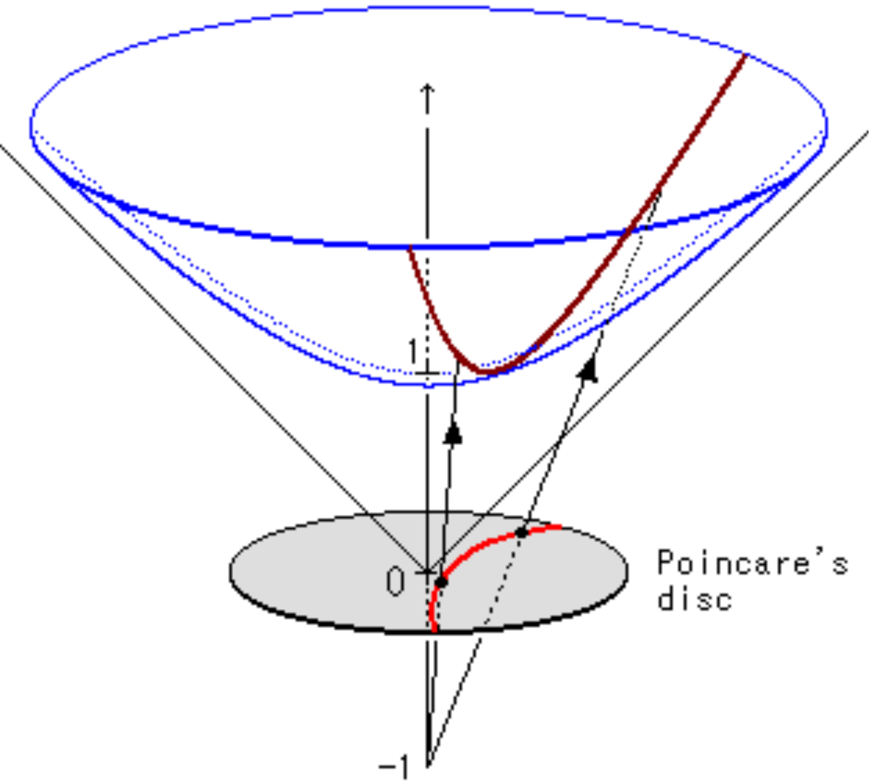


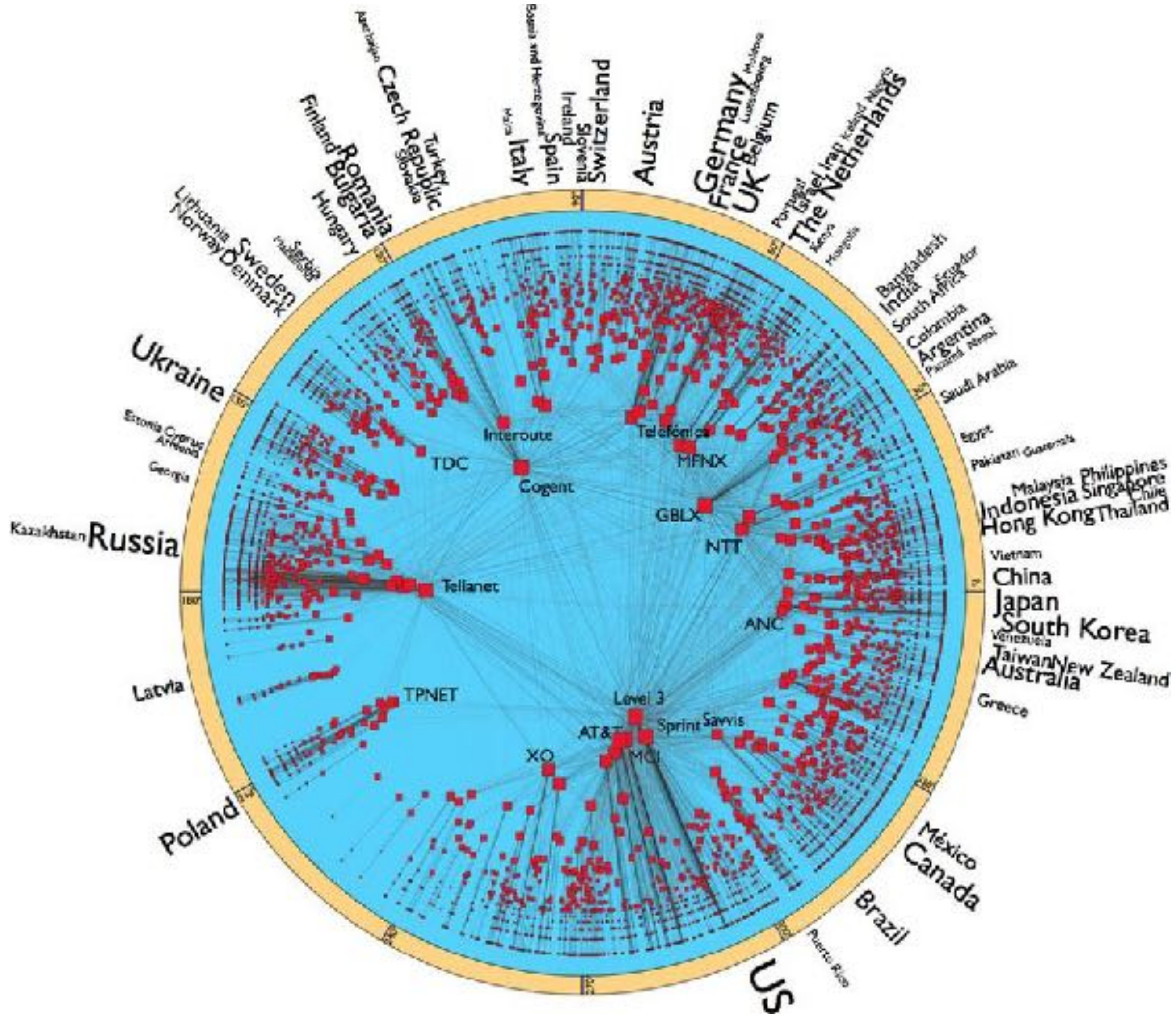
Embedding in metric spaces



M.A. Serrano, D. Krioukov, and M. Boguna, “Self-similarity of complex networks and hidden metric spaces,” Physical Review Letters 100, 078701 (2008).
M. Boguna, D. Krioukov, and K.C. Claffy, “Navigability of complex networks,” Nature Physics 5, 74–80 (2009).

Embedding in the hyperbolic space





D. Krioukov et al. "Curvature and temperature of complex networks," Physical Review E 80, 035101 (2009).

D. Krioukov et al., "Hyperbolic geometry of complex networks," Physical Review E 82, 036106 (2010).

M. Boguna et al., "Sustaining the internet with hyperbolic mapping," Nature Communications 1, 62 (2010).

G. Bianconi and C. Rahmede, "Emergent hyperbolic network geometry," Scientific Reports 7 (2017).

Network models in the hyperbolic space

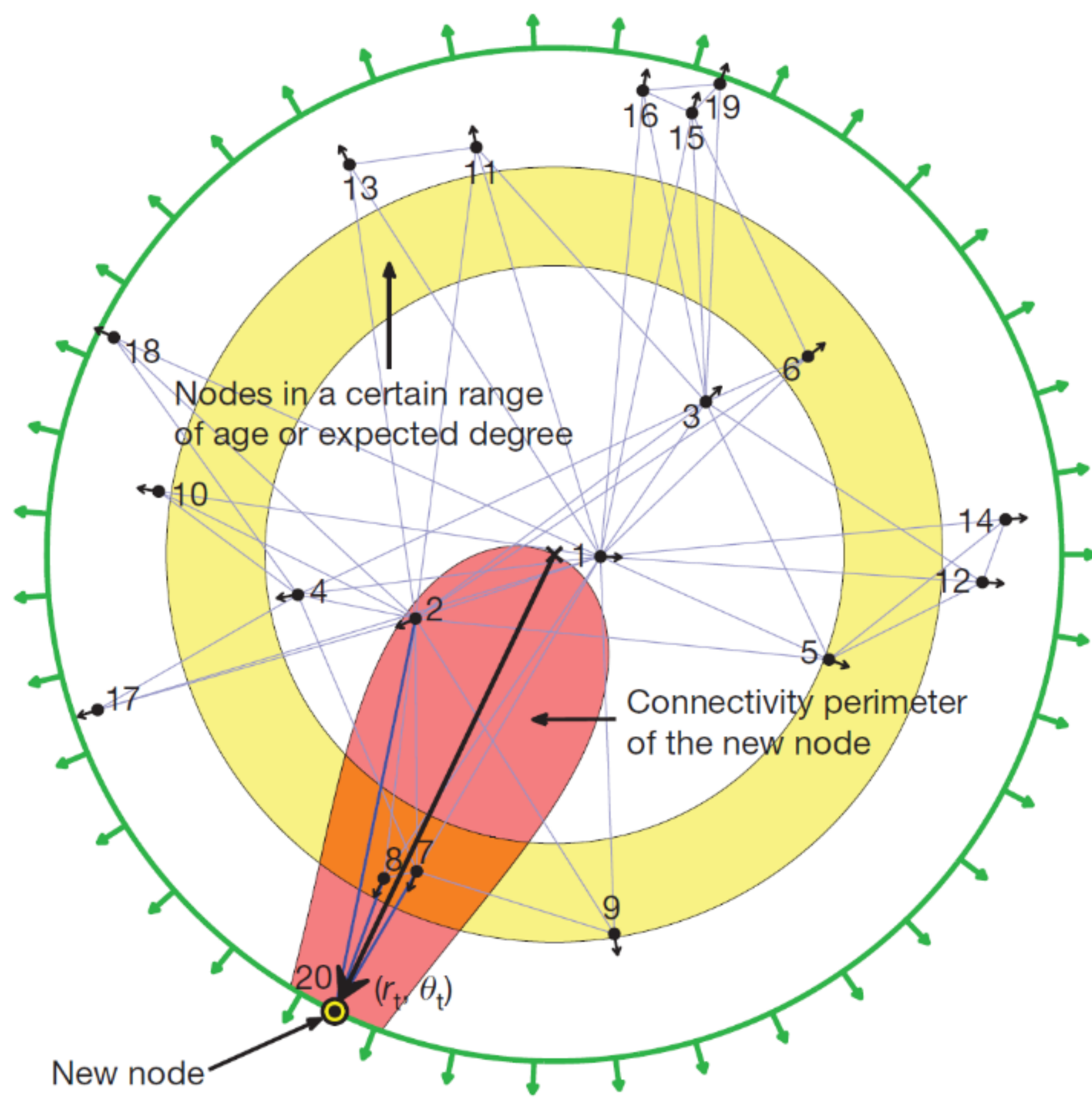
Popularity-similarity optimization model (PSOM)

Every node i is a point in the hyperbolic space (r_i, θ_i)

The probability that nodes i and j are connected is indicated with $p(x_{ij})$

x_{ij} distance of nodes i and j, it includes:

- γ exponent power-law degree distribution
- $\langle k \rangle$ average degree
- T temperature (clustering)



r_i radial coordinate. It accounts for popularity. It is proportional to the degree of the node.

θ_i angular coordinate. Angular difference between nodes coordinates accounts for similarity.

F. Papadopoulos et al., "Popularity versus similarity in growing networks," Nature 489, 537–540 (2012).

Embedding networks in the hyperbolic space

Hypermap

Radial coordinates of nodes (and additional model parameters) are estimated from the observed network

$$r_i \sim k_i$$

Angular coordinates of nodes are inferred from the observed topology by maximizing the likelihood

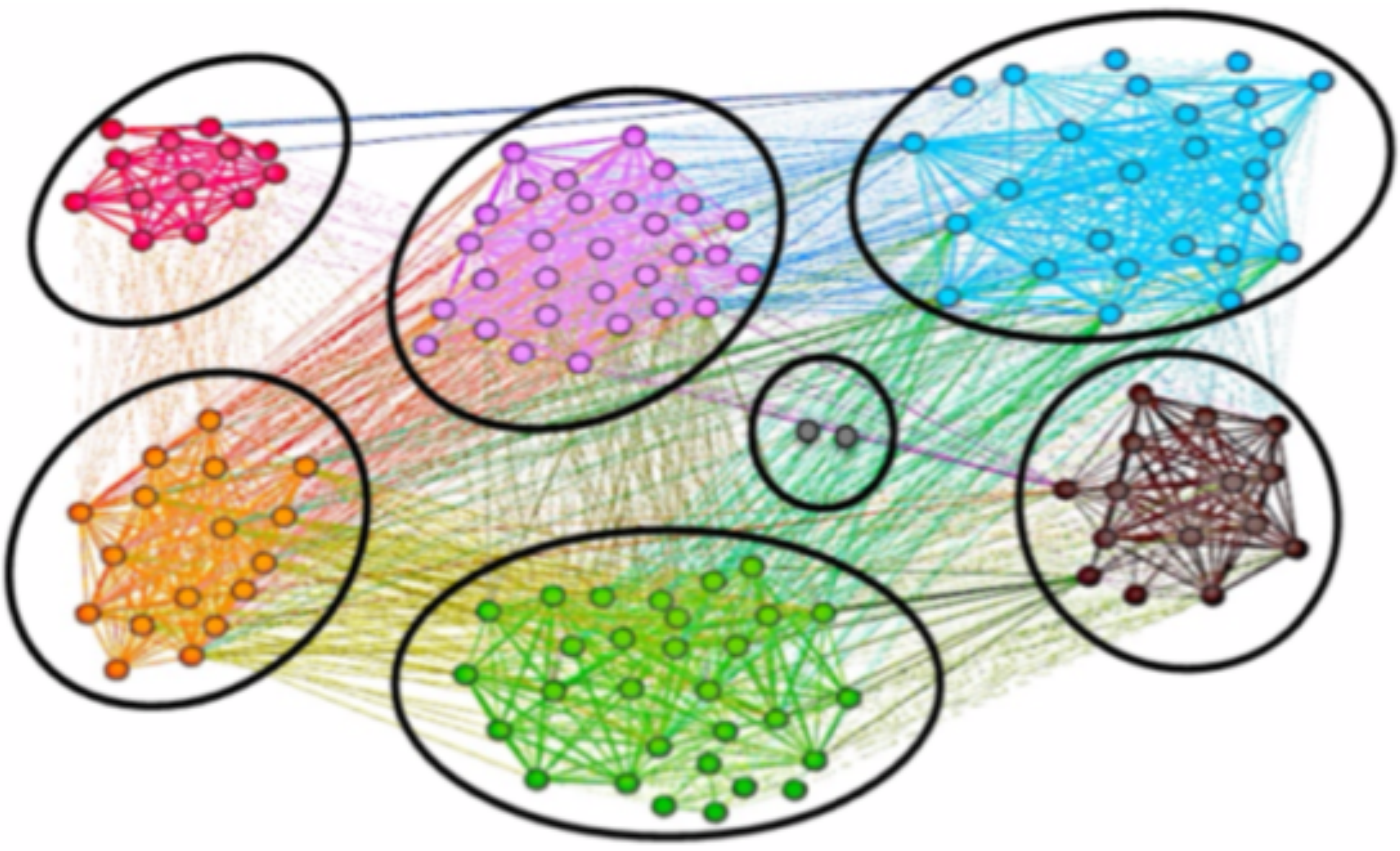
$$L = \prod_{i<j} p(x_{ij})^{A_{ij}} (1 - p(x_{ij}))^{1-A_{ij}}$$

The temperature T is generally treated as a free parameter that can be tuned depending on the application (e.g., most effective routing protocol).

F. Papadopoulos, C. Psomas, and D. Krioukov, "Network mapping by replaying hyperbolic growth," IEEE/ACM Transactions on Networking (TON) 23, 198–211 (2015).

F. Papadopoulos, R. Aldecoa, and D. Krioukov, "Network geometry inference using common neighbors," Physical Review E 92, 022807 (2015).

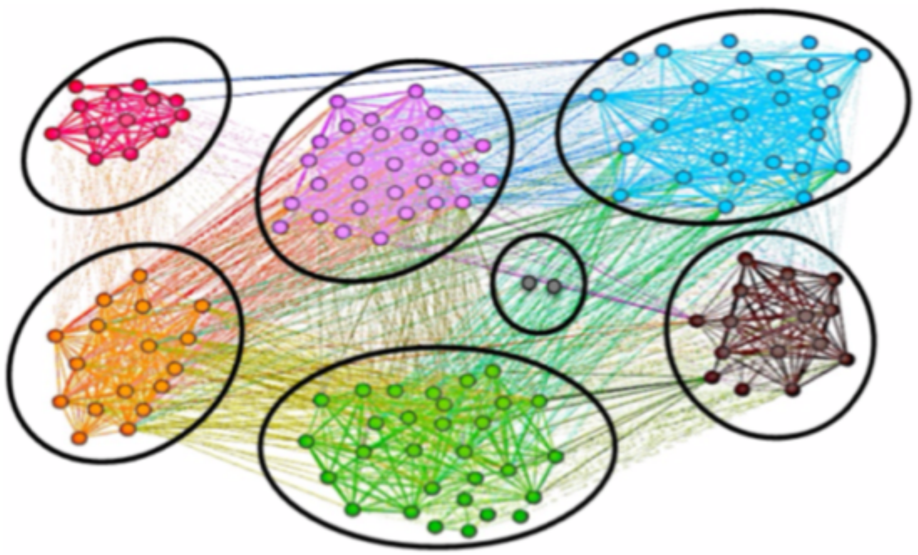
Community structure



S. Fortunato,“Community detection in graphs,”Physics reports 486, 75–174 (2010).

Network models for community structure

Degree-corrected stochastic block model (SBM)

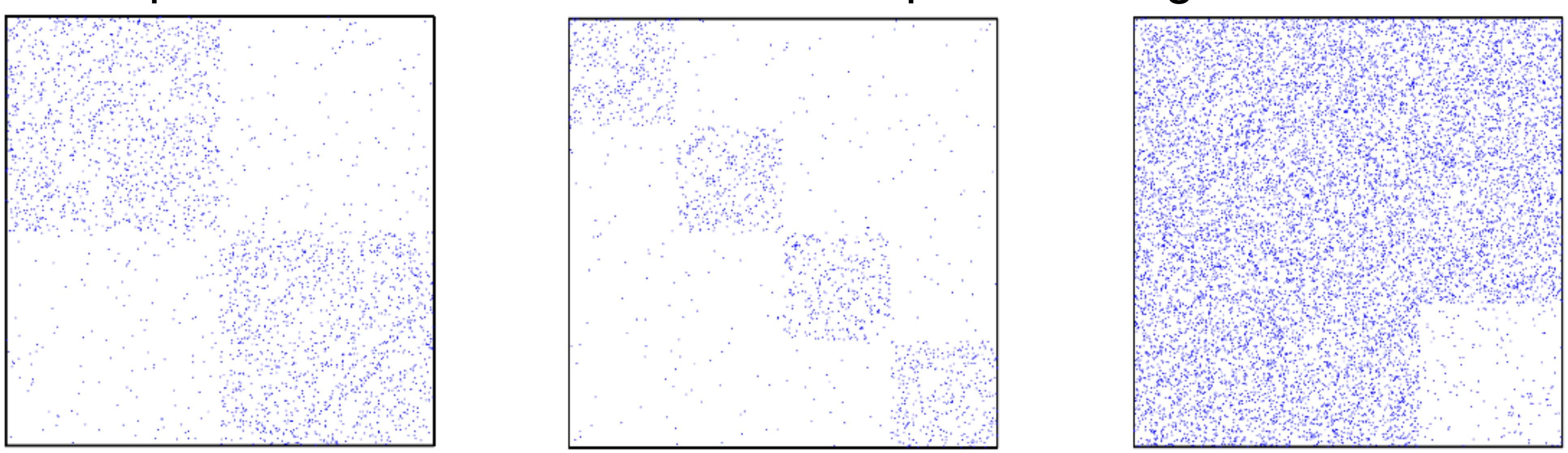


Every node i is represented by the coordinates (k_i, σ_i)

k_i Node degree. It accounts for popularity.

σ_i Node membership. Memberships of pairs of nodes are used to determine their similarity.

Probabilities for pair of nodes to be connected depend on degree and memberships



B. Karrer and M.E.J. Newman, "Stochastic blockmodels and community structure in networks," Physical Review E 83, 016107 (2011).
T.P. Peixoto, "Bayesian stochastic blockmodeling," arXiv preprint arXiv:1705.10225 (2017).

Finding communities in networks

Under the SBM ansatz,
memberships of nodes are inferred
from the observed topology by
maximizing the likelihood

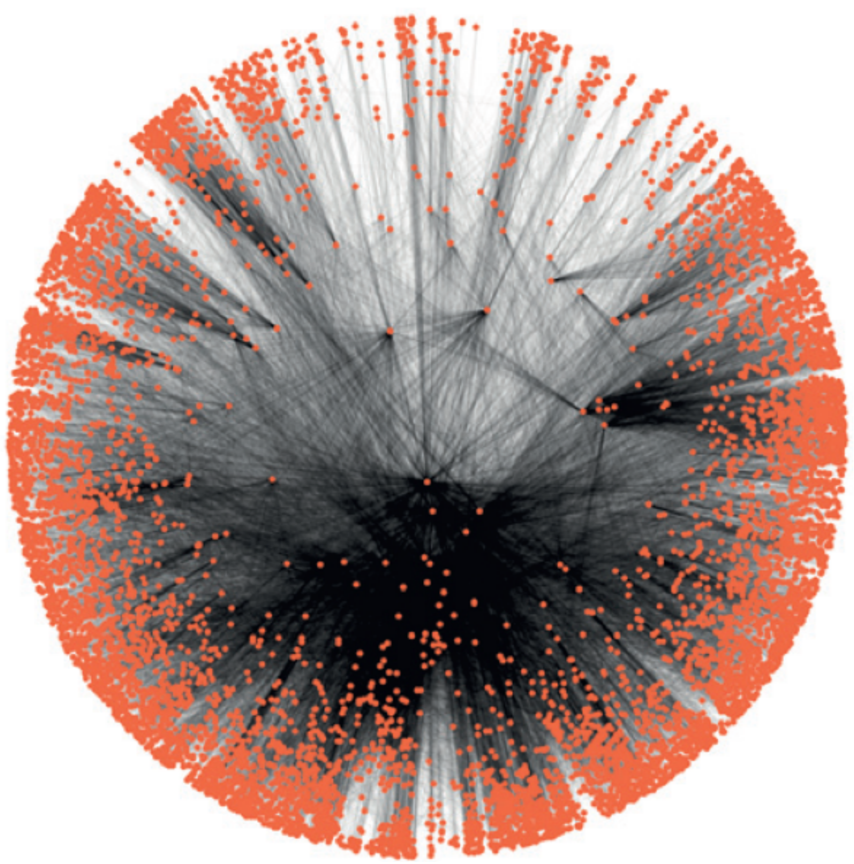
$$L = \prod_{i<j} p(\sigma_i, \sigma_j)^{A_{ij}} (1 - p(\sigma_i, \sigma_j))^{1-A_{ij}}$$

A huge number of methods are available for community detection:
spectral methods, modularity maximization methods, ….

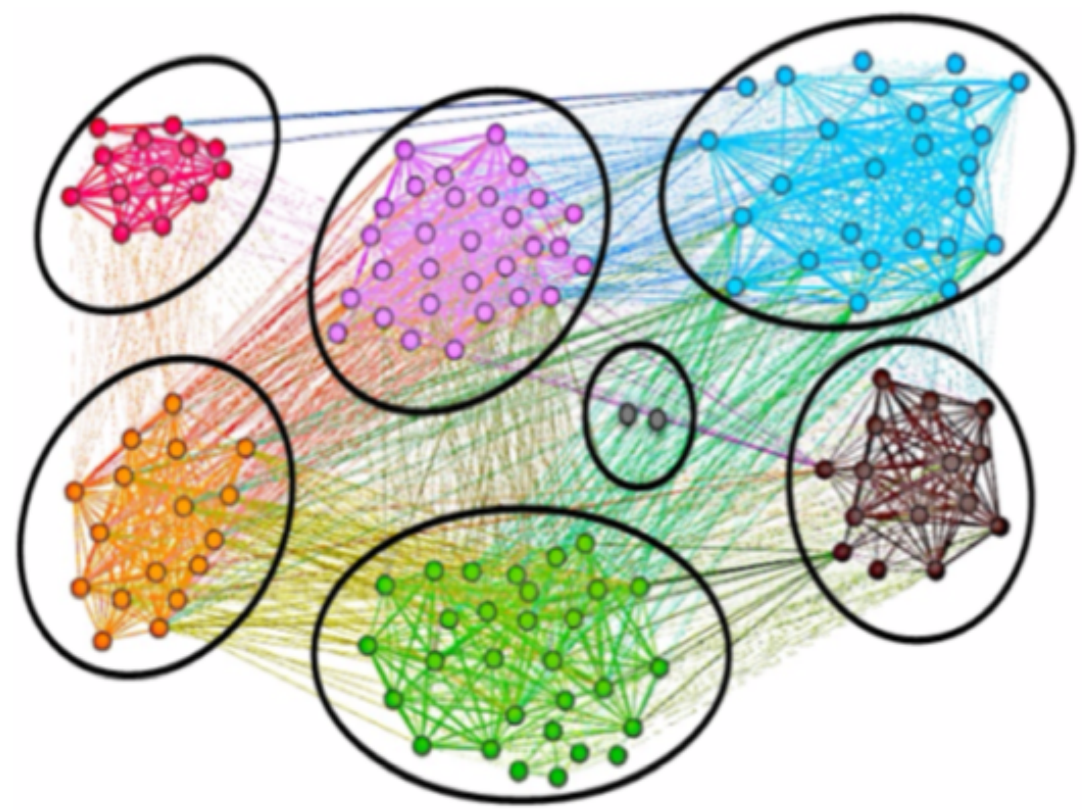
B. Karrer and M.E.J. Newman, "Stochastic blockmodels and community structure in networks," Physical Review E 83, 016107 (2011).
T.P. Peixoto, "Bayesian stochastic blockmodeling," arXiv preprint arXiv:1705.10225 (2017).
S. Fortunato,"Community detection in graphs,"Physics reports 486, 75–174 (2010).

The rationale of the study

Embedding in the hyperbolic space



Community structure



The two representations are different in many respects. However, their basic ingredients are similar. Are the two representations analogous in practical cases? Can we understand the same system properties using either one or the other representation?

Quantifying the analogy

data

network	N
IPv4 Internet	37,542
IPv6 Internet	5,143
C. Elegans, layer 1	248
C. Elegans, layer 2	258
C. Elegans, layer 3	278
D. Melanogaster, layer 1	752
D. Melanogaster, layer 2	633
arXiv, layer 1	1,537
arXiv, layer 2	2,121
arXiv, layer 3	129
arXiv, layer 4	3,669
arXiv, layer 5	608
arXiv, layer 6	336
Physician, layer 1	106
Physician, layer 2	113
Physician, layer 3	110
SacchPomb, layer 1	751
SacchPomb, layer 2	182
SacchPomb, layer 3	2,340
SacchPomb, layer 4	819
Human brain, layer 1	85
Human brain, layer 2	78
Rattus, layer 1	1,866
Rattus, layer 2	529

Rattus, layer 1	1,866
Rattus, layer 2	529
Air/Train, layer 1	69
Air/Train, layer 2	67
ARK200909	24,091
ARK201003	26,307
ARK201012	29,333
Enron emails	33,696
Music chords	2,476
OpenFights Air Transp.	3,397
Human Metabolites	1,436
Human HI-II-14 proteome	4,100
AS Internet	23,748
AS Oregon Interent, $T = 0.58$	6,474
Air Transportation, $T = 0.14$	3,618
P2P, $T = 0.92$	6,299
Euro Roads, $T = 0.28$	1,039
PSOM, $\langle k \rangle = 5, \gamma = 2.1, T = 0.1$	4,114
PSOM, $\langle k \rangle = 5, \gamma = 2.1, T = 0.9$	4,180

39 real networks + 2 instances of the PSOM

Quantifying the analogy

methods

Hyperbolic embedding

Real networks

Publicly available embeddings

Publicly available methods for hyperbolic embedding

PSOM

Generated with publicly available algorithms, embedding given by ground-truth values

KK Kleineberg et al., "Hidden geometric correlations in real multiplex networks," Nature Physics 12, 1076–1081 (2016).
KK Kleineberg et al., "Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks," Physical Review Letters 118, 218301 (2017).
F. Papadopoulos, et al "Popularity versus similarity in growing networks," Nature 489, 537–540 (2012).

Community structure

Publicly available implementations of

Louvain

Infomap

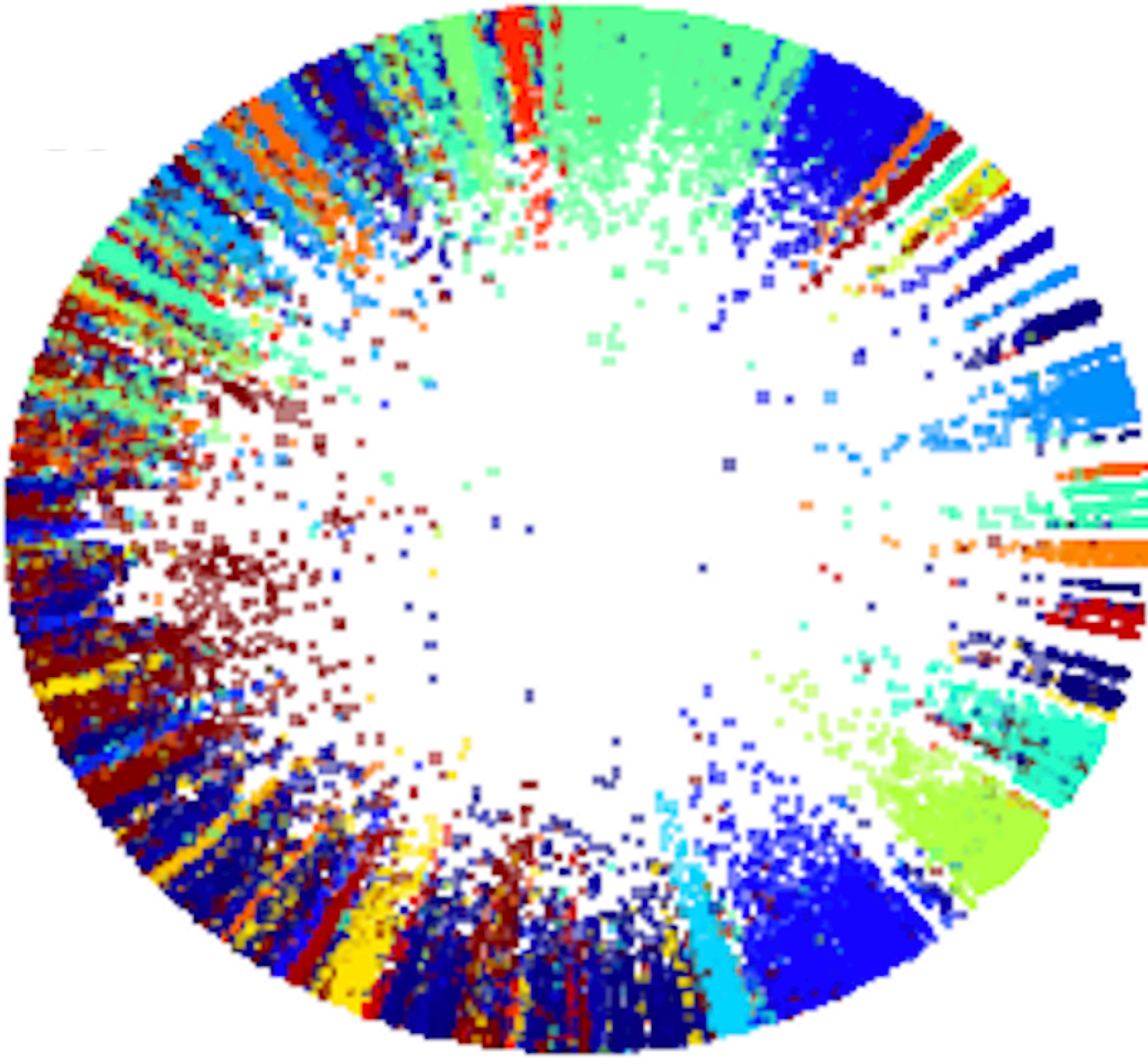
algorithm by Ronhovde and Nussinov

LFR benchmark graphs

VD Blondel et al., "Fast unfolding of communities in large networks," Journal of statistical mechanics: theory and experiment 2008, P10008 (2008).
M. Rosvall and C.T. Bergstrom, "Maps of random walks on complex networks reveal community structure," PNAS 105, 1118–1123 (2008).
P. Ronhovde and Z. Nussinov, "Local resolution-limit- free potts model for community detection," Phys. Rev. E 81, 046114 (2010).
A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," Physical review E 78, 046110 (2008).

Quantifying the analogy

IPv4 Internet



Positions of points are determined by the hyperbolic embedding of the network

Colors identify the community membership of the nodes according to Louvain (C = 31 communities)

Z. Wang, et al., "Hyperbolic mapping of complex networks based on community information," Physica A: Statistical Mechanics and its Applications 455, 104–119 (2016).

Quantifying the analogy

Systematic analysis

Angular coherence of a community

$$\xi_g \, e^{i\phi_g} = \frac{1}{n_g} \sum_{j=1}^{N} \delta_{\sigma_j, g} \, e^{i\theta_j}$$

Angular coherence of a community partition

$$\bar{\xi} = \frac{1}{N} \sum_{g=1}^{C} n_g \xi_g$$

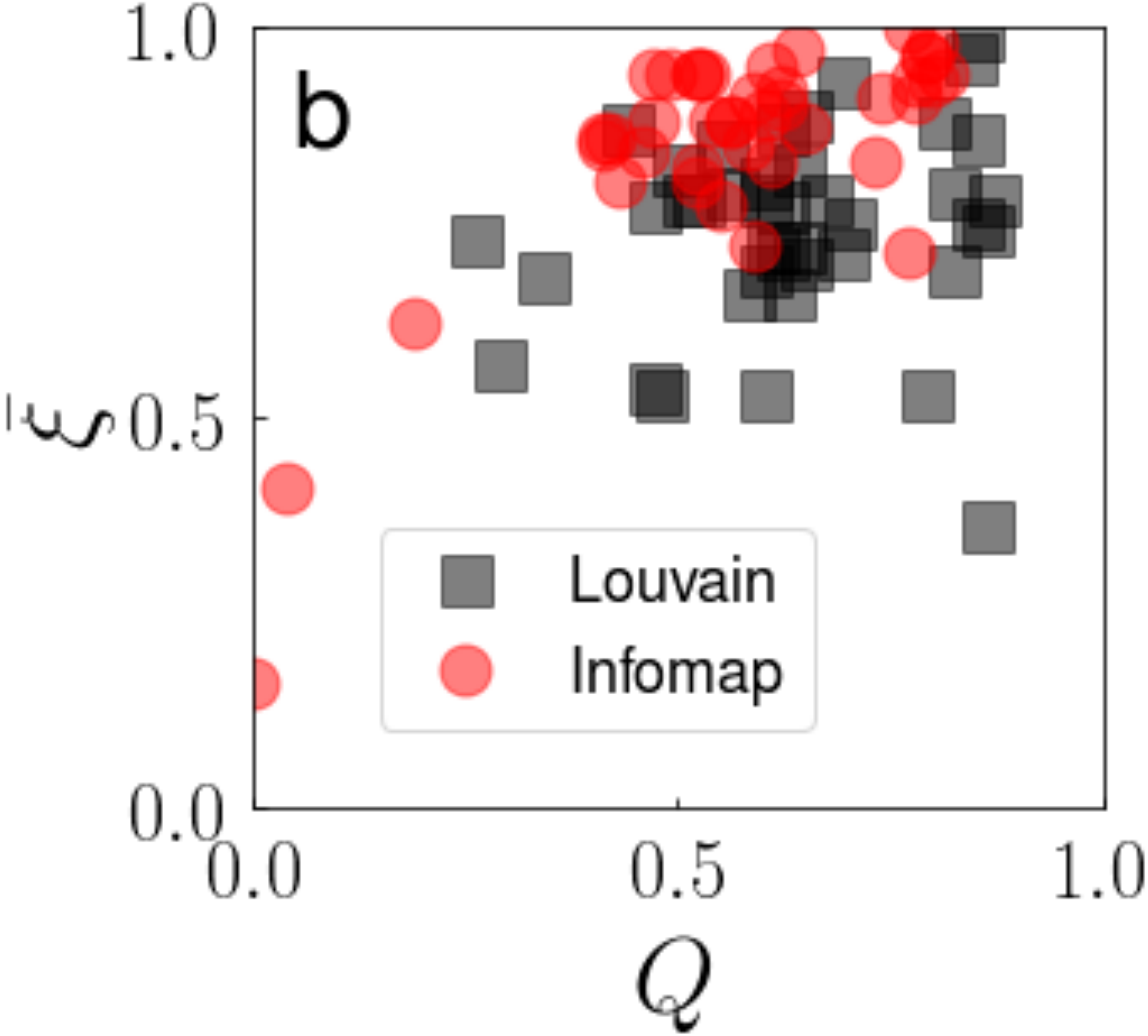
Strength of the community partition is measured with the modularity function Q

		Louvain			Infomap		
network	N	C	Q	$\bar{\xi}$	C	Q	$\bar{\xi}$
IPv4 Internet	37,542	31	0.61	0.72	1,625	0.47	0.94
IPv6 Internet	5,143	19	0.48	0.53	418	0.41	0.86
C. Elegans, layer 1	248	9	0.65	0.70	29	0.61	0.83
C. Elegans, layer 2	258	9	0.50	0.82	23	0.46	0.84
C. Elegans, layer 3	278	7	0.44	0.87	11	0.42	0.86
D. Melanogaster, layer 1	752	17	0.64	0.82	70	0.59	0.91
D. Melanogaster, layer 2	633	17	0.64	0.72	68	0.60	0.89
arXiv, layer 1	1,537	32	0.87	0.78	130	0.81	0.94
arXiv, layer 2	2,121	35	0.86	0.74	190	0.79	0.96
arXiv, layer 3	129	10	0.81	0.88	17	0.78	0.93
arXiv, layer 4	3,669	46	0.82	0.69	290	0.74	0.91
arXiv, layer 5	608	23	0.85	0.86	61	0.79	0.96
arXiv, layer 6	336	17	0.84	0.96	38	0.80	0.98
Physician, layer 1	106	8	0.51	0.78	13	0.52	0.80
Physician, layer 2	113	10	0.56	0.79	14	0.55	0.77
Physician, layer 3	110	9	0.60	0.53	18	0.59	0.72

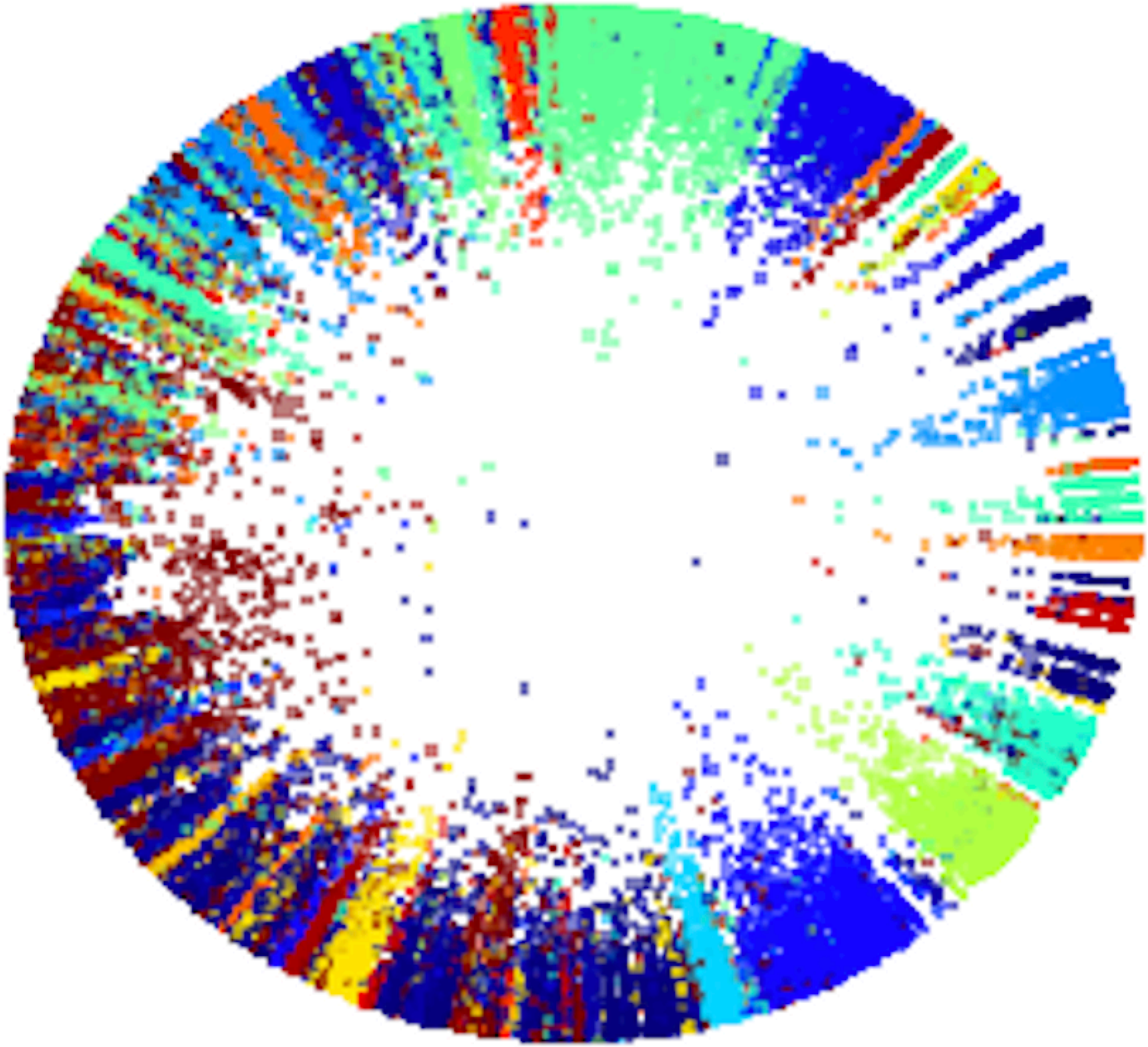
Y. Kuramoto, Chemical oscillations, waves, and turbulence (Dover Publications, New York, 1984).
M.E.J Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E 69, 026113 (2004).

Quantifying the analogy

Systematic analysis



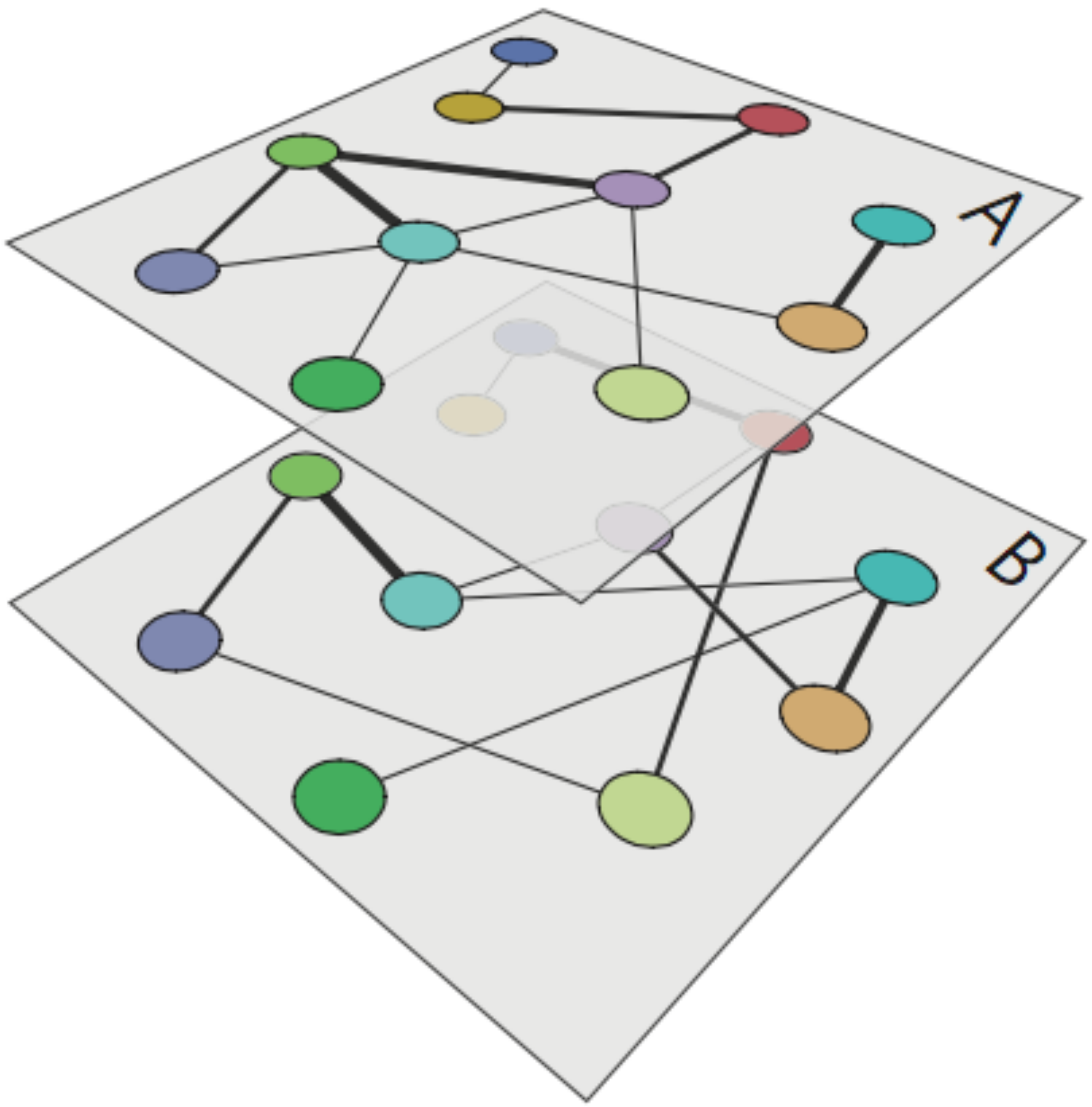
Consequences of the analogy

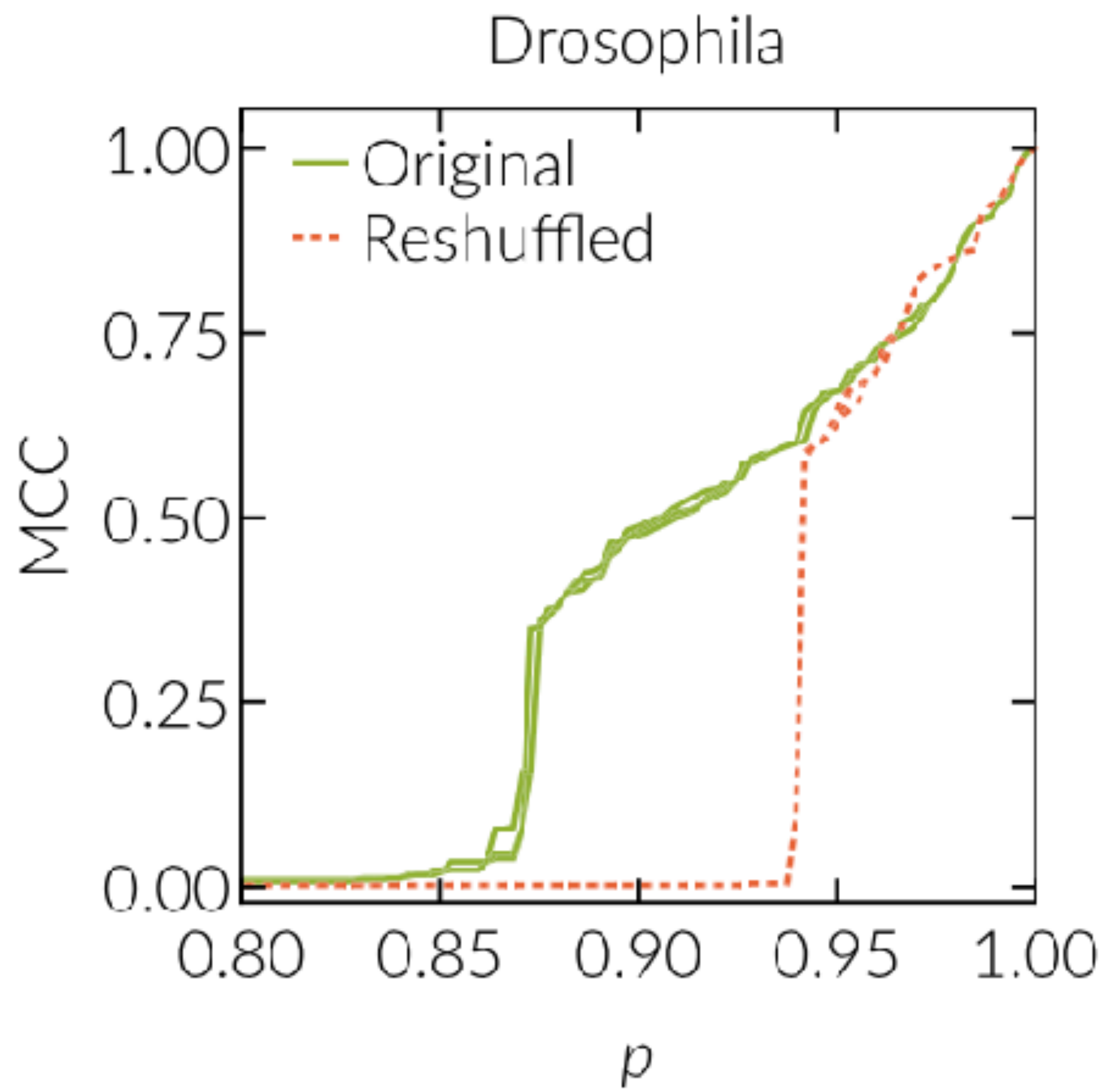


Can we interpret physical properties of networks deduced from their hyperbolic embedding using community structure only?

Robustness of multiplex networks

under targeted attack

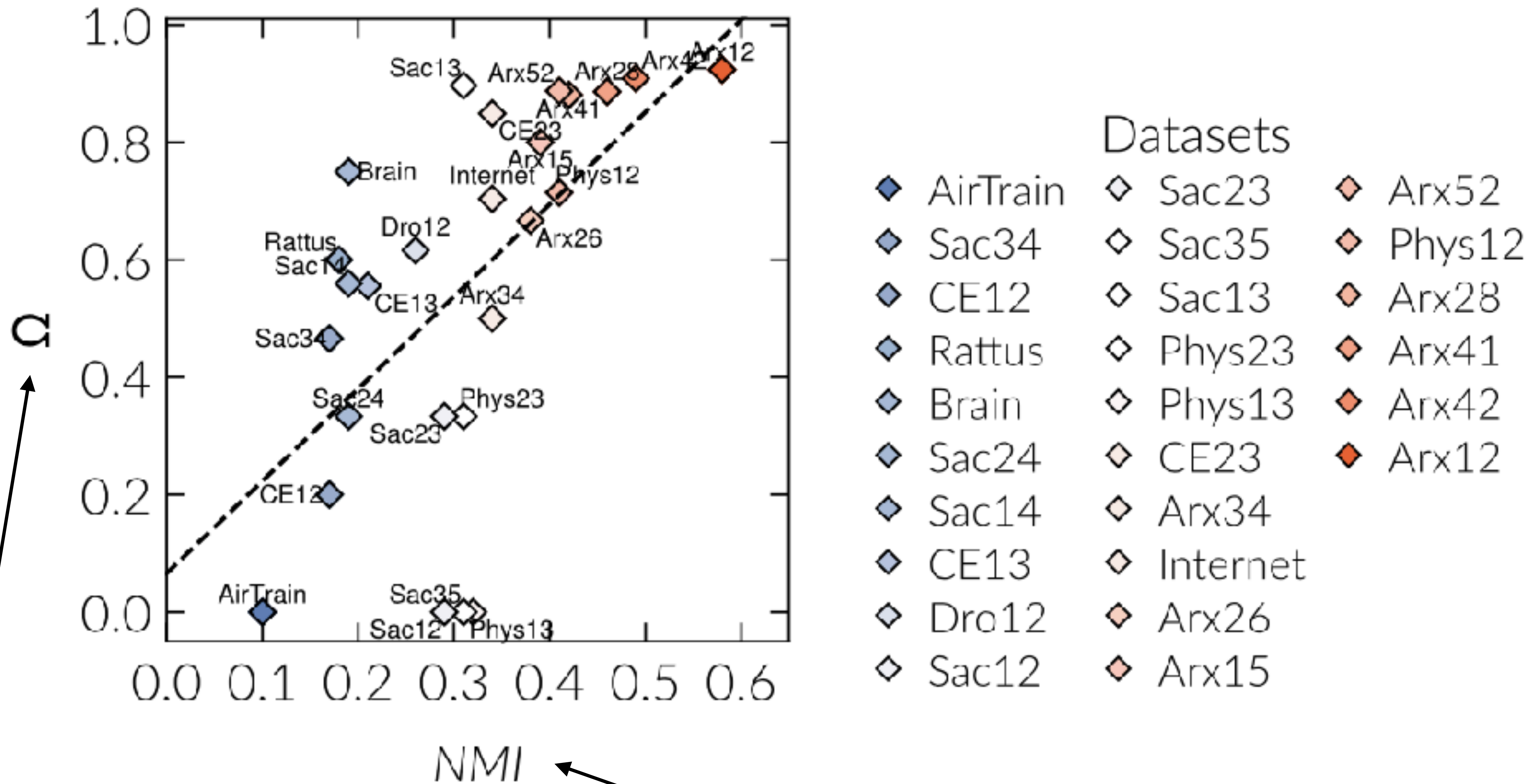




KK Kleineberg et al., "Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks," Physical Review Letters 118, 218301 (2017).

Robustness of multiplex networks

under targeted attack



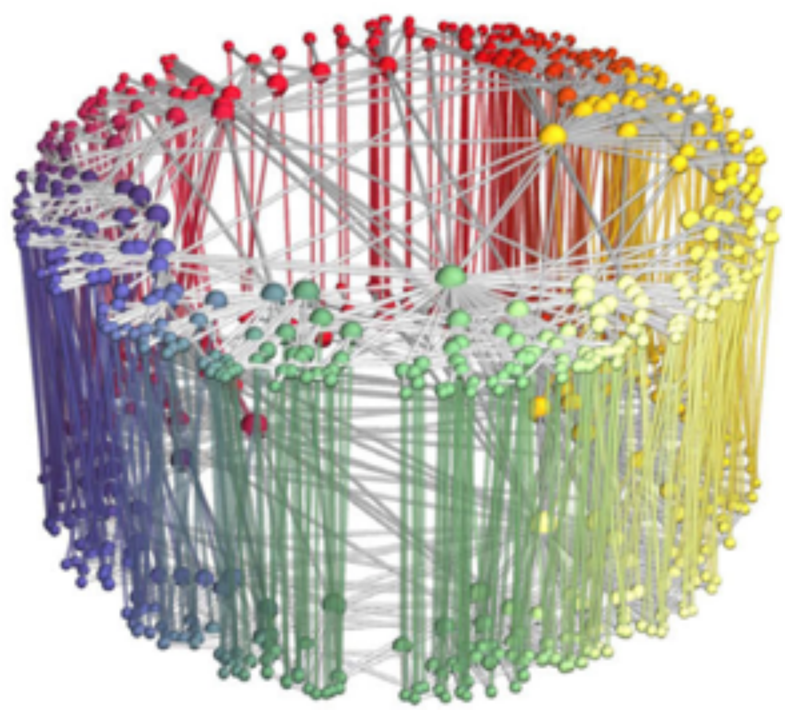
robustness

geometric correlation

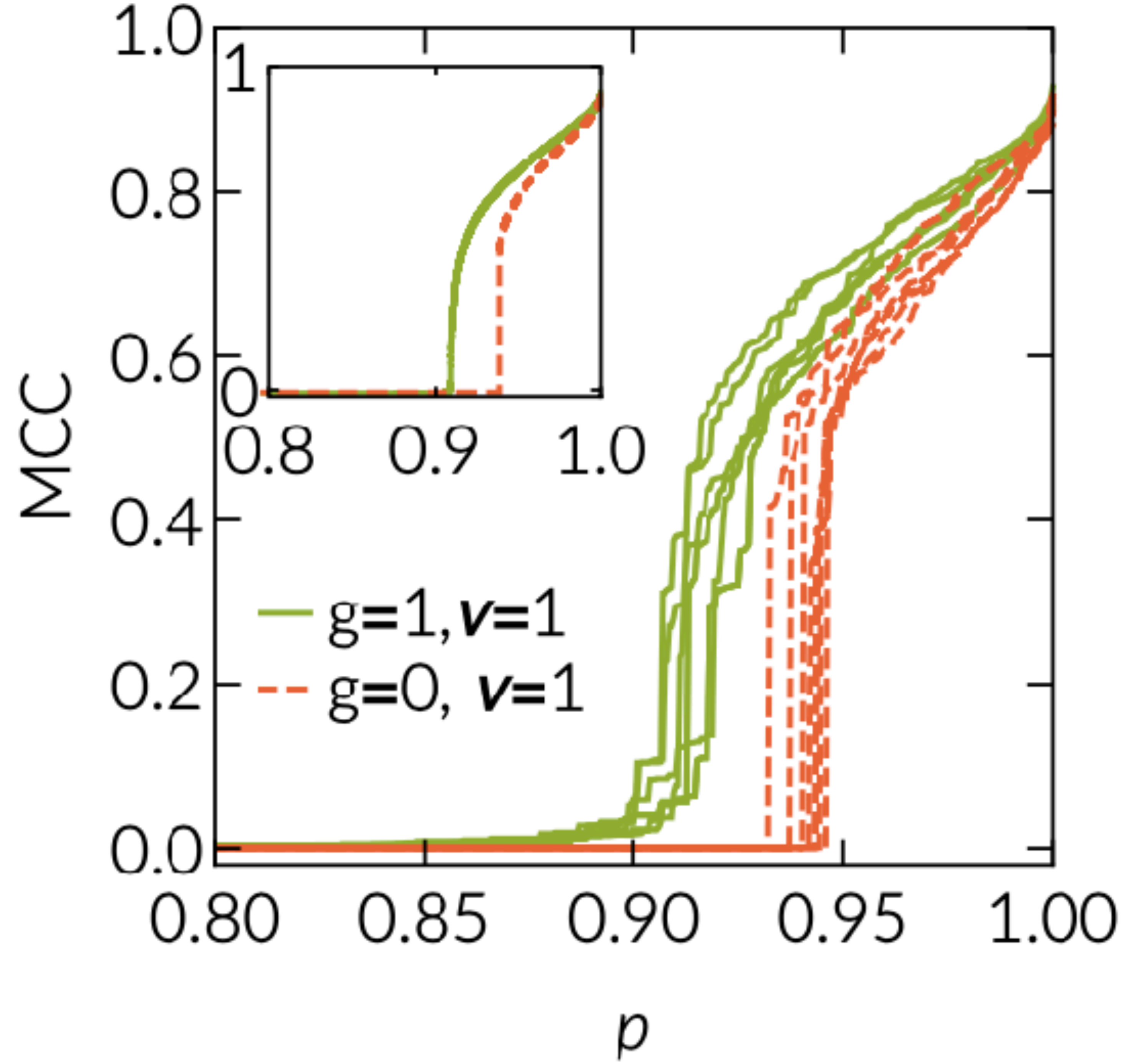
KK Kleineberg et al., “Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks,” Physical Review Letters 118, 218301 (2017).

Robustness of multiplex networks

under targeted attack



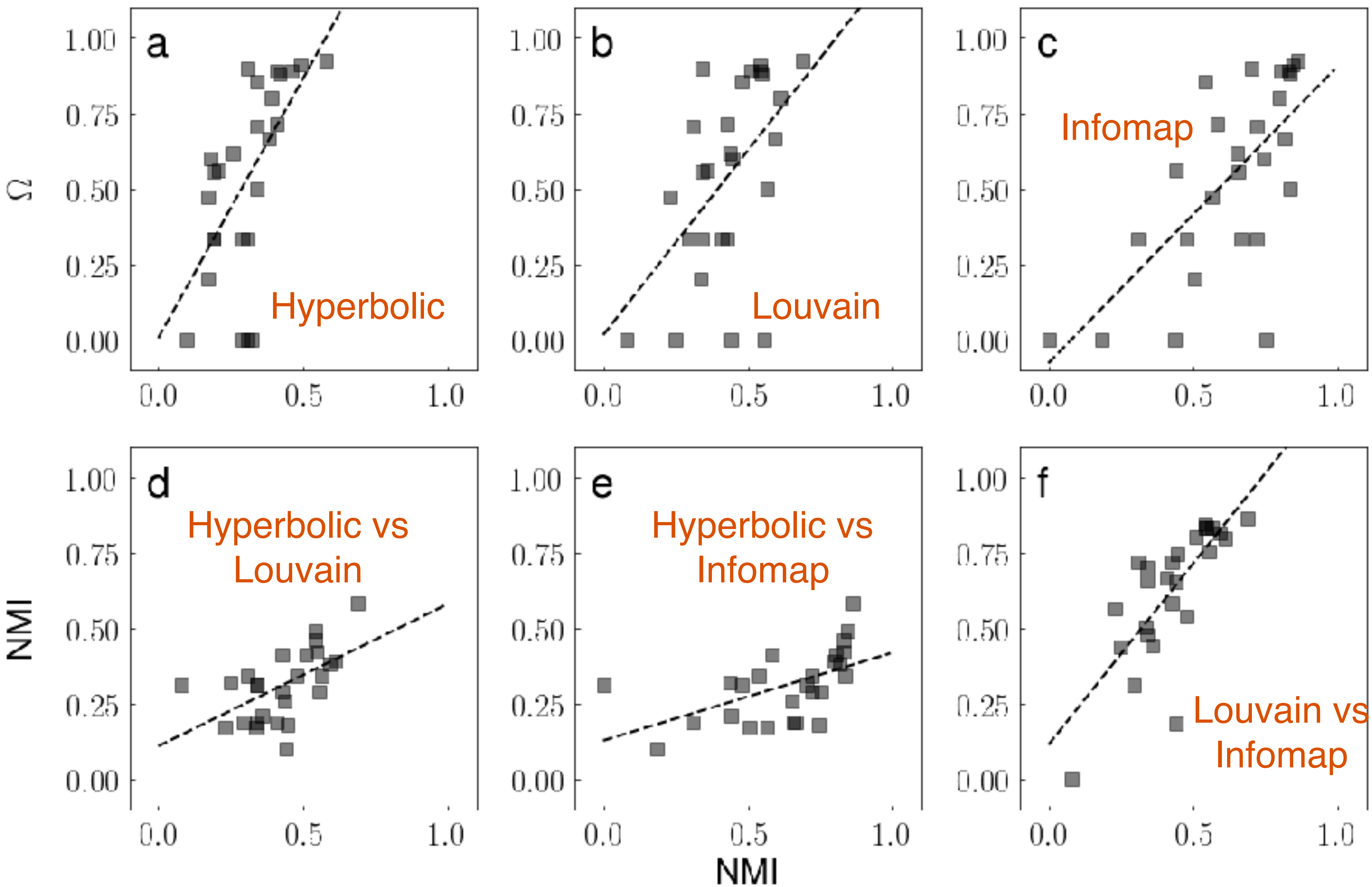
Synthetic network model with tunable correlation among radial and angular coordinates



KK Kleineberg et al., “Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks,” Physical Review Letters 118, 218301 (2017).

Robustness of multiplex networks

interpreted with communities

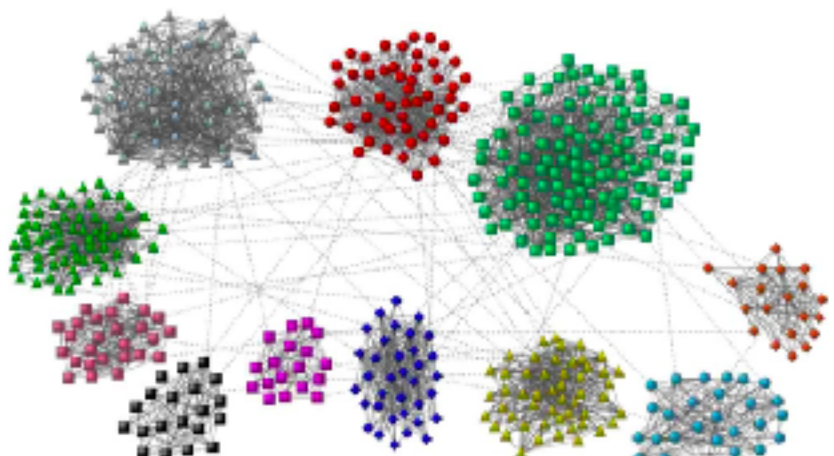


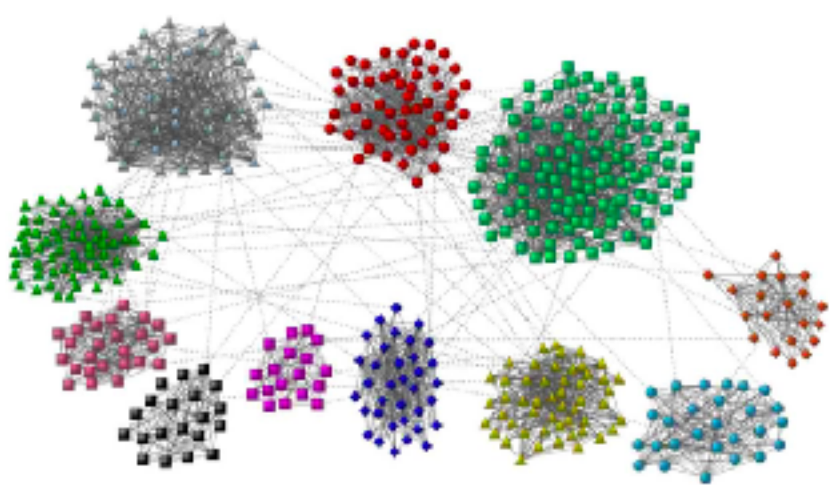
NMI is defined as in L. Danon et al., “Comparing community structure identification,” Journal of Statistical Mechanics: Theory and Experiment 2005, P09008 (2005).

Robustness of multiplex networks

interpreted with communities

1) Create two identical network instances of the LFR model, strength of community structure can be tuned by varying the mixing parameter value





2) Shuffle labels of nodes in one of the layers to destroy degree-degree correlations and edge overlap

A) Shuffling is allowed only among pairs of nodes within the same community

NMI = 1

B) Shuffling is allowed among all pairs of nodes

NMI = 0

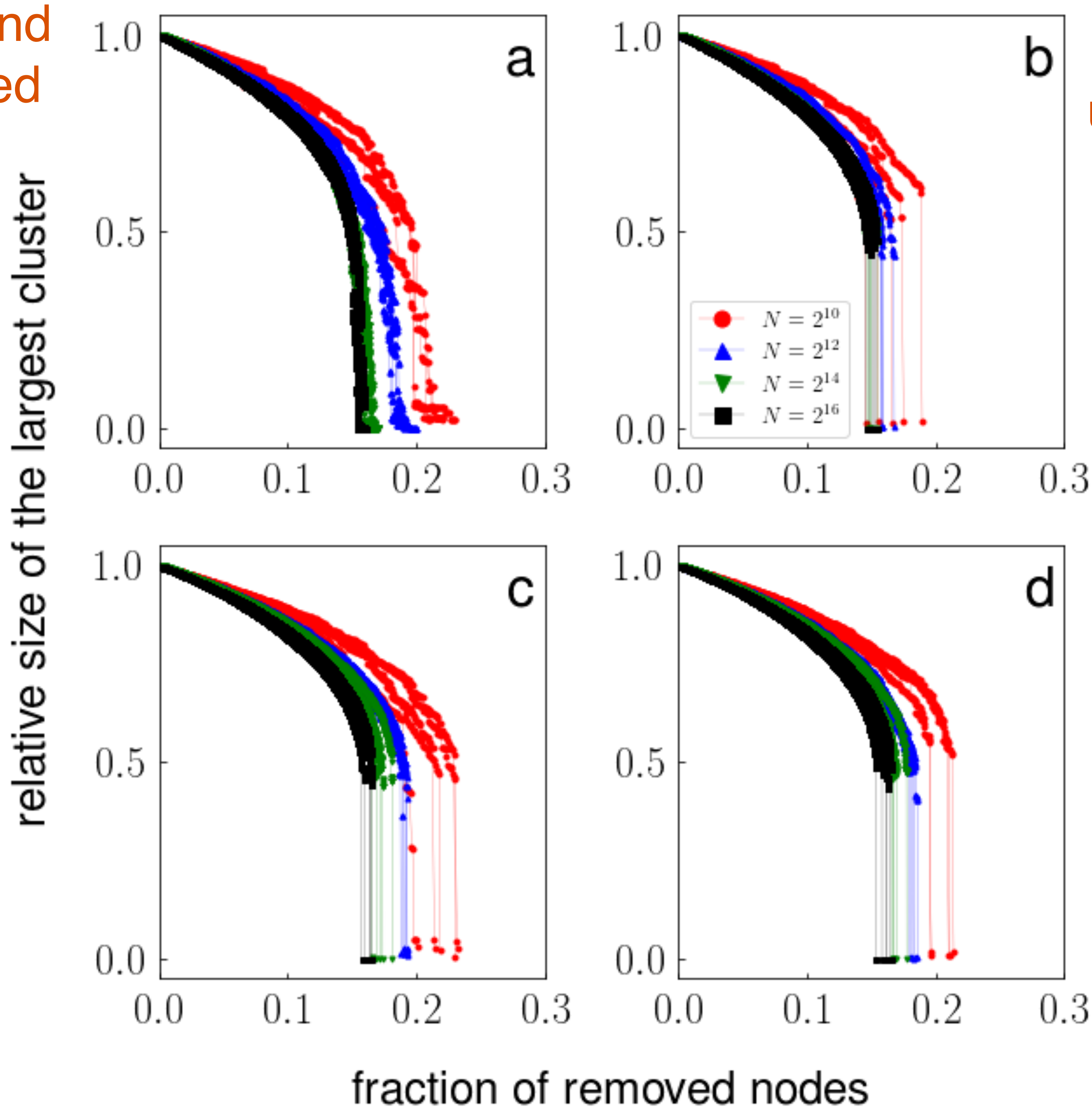
A. Lancichinetti, S. Fortunato, and F. Radicchi, “Benchmark graphs for testing community detection algorithms,” Physical review E 78, 046110 (2008).

Robustness of multiplex networks

LFR model $C = \sqrt{N}, S = \sqrt{N}$

Strong and correlated

Strong and uncorrelated



Weak and correlated

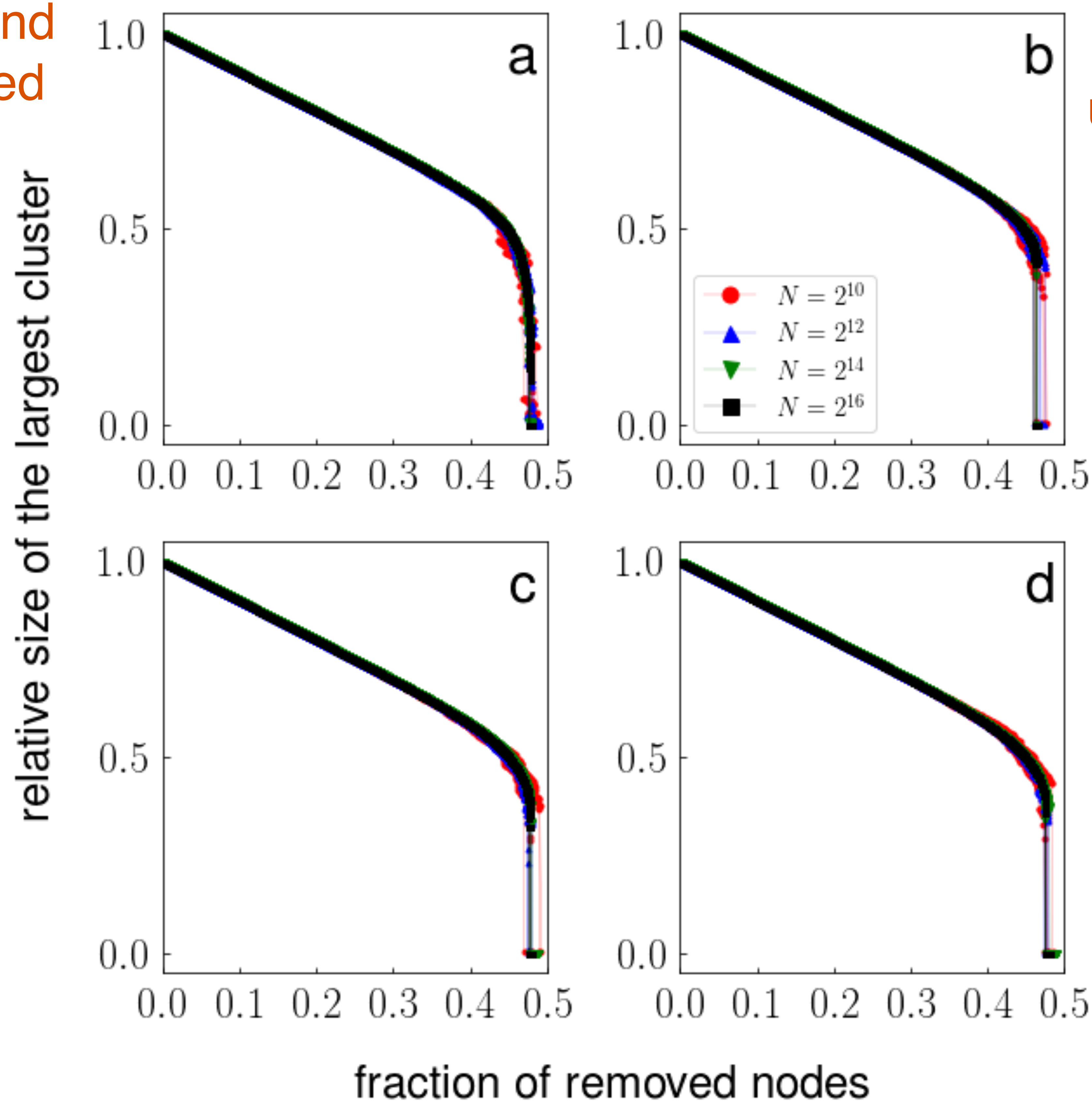
Weak and uncorrelated

Robustness of multiplex networks

LFR model $C = N/64, S = 64$

Strong and correlated

Strong and uncorrelated



Weak and correlated

Weak and uncorrelated

Navigability of networks

greedy routing

algorithm

Strategy for delivering a packet from a source node s to a target node t

At every stage of the algorithm, the packet seating on node i chooses the next move according to the rule

$$j^{(i)}_{(best)} = \arg \min_{j \in N_i} d(j,t)$$

If a packet reaches the target, it is considered delivered

If a packet visits a second time the same node, it is considered lost

metrics of performance

For random pairs of nodes s and t

z , success rate

<R> , average length of successful paths

Z <1/R> , efficiency

M. Boguna et al., "Sustaining the internet with hyperbolic mapping," Nature Communications 1, 62 (2010).

M. Boguna, D. Krioukov, and K.C. Claffy, "Navigability of complex networks," Nature Physics 5, 74–80 (2009).

Navigability of networks

greedy routing and community structure

$$j^{(i)}_{(best)} = \arg \min_{j \in N_i} d(j,t)$$

We define a measure of “distance” among pairs of nodes in the stochastic block model

$$d(j,t) = \beta D_{\sigma_j,\sigma_t} - (1-\beta) \ln k_j$$

D_{σ_j,σ_t} distance between modules in the stochastic block model, calculated using the log of the observed density of connections between communities

k_j degree of node j

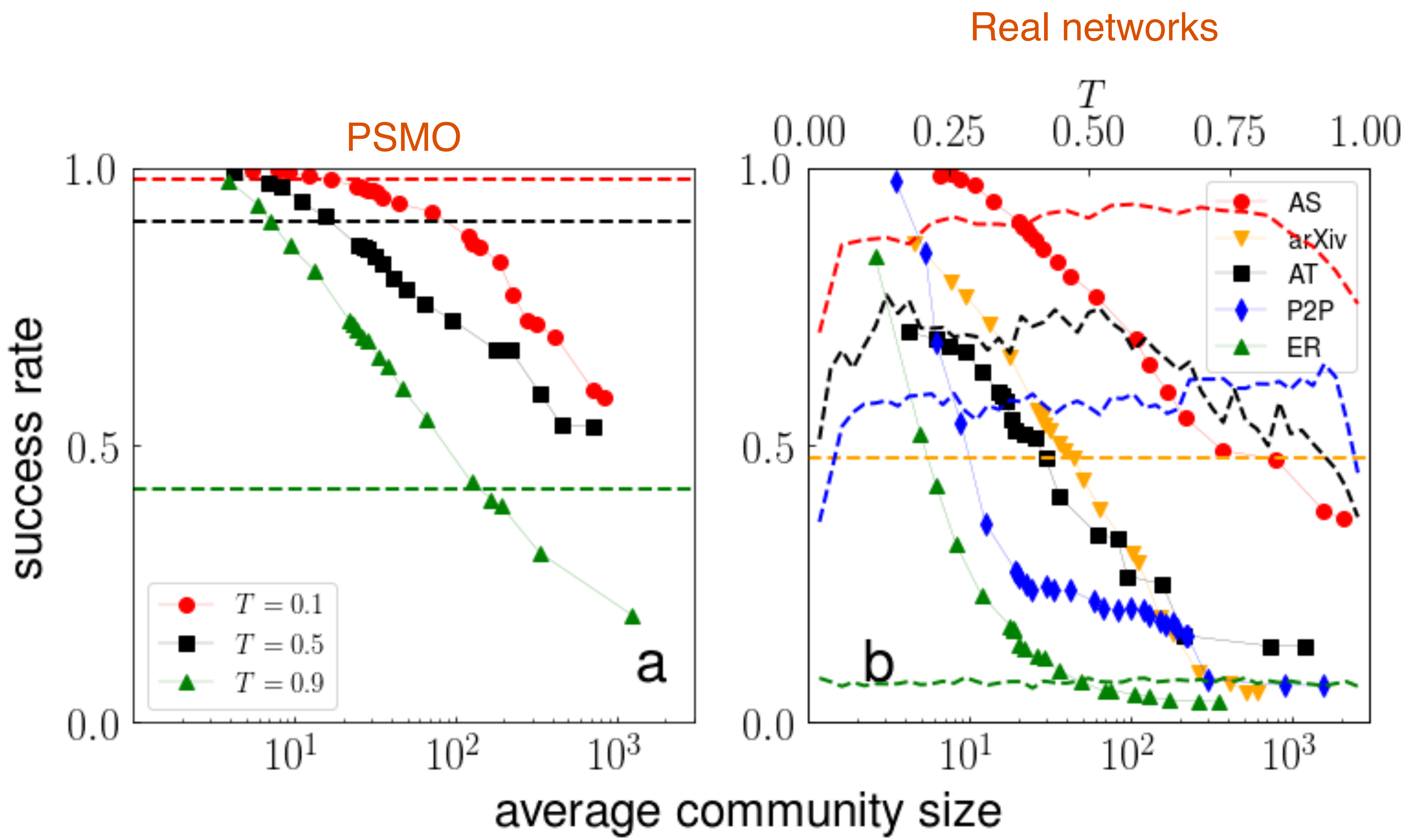
σ_j module of node j

β weighting parameter (we chose the value that maximizes performance)

We vary the size S and the number C of the communities by changing the resolution parameter of the the algorithm by Ronhovde and Nussinov

Navigability of networks

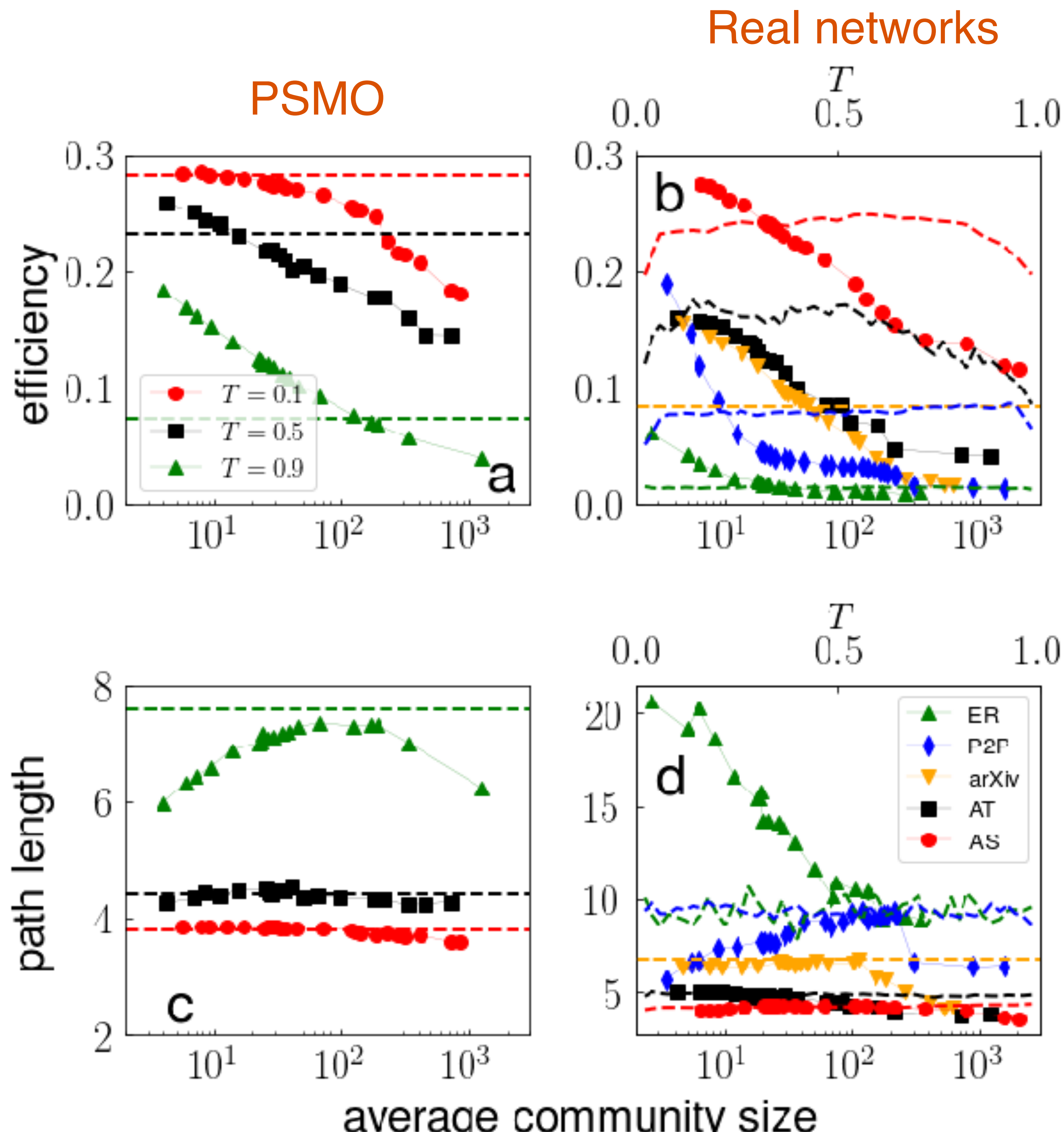
Success rate



F. Papadopoulos et al. “Network mapping by replaying hyperbolic growth,” IEEE/ACM Transactions on Networking (TON) 23, 198–211 (2015).

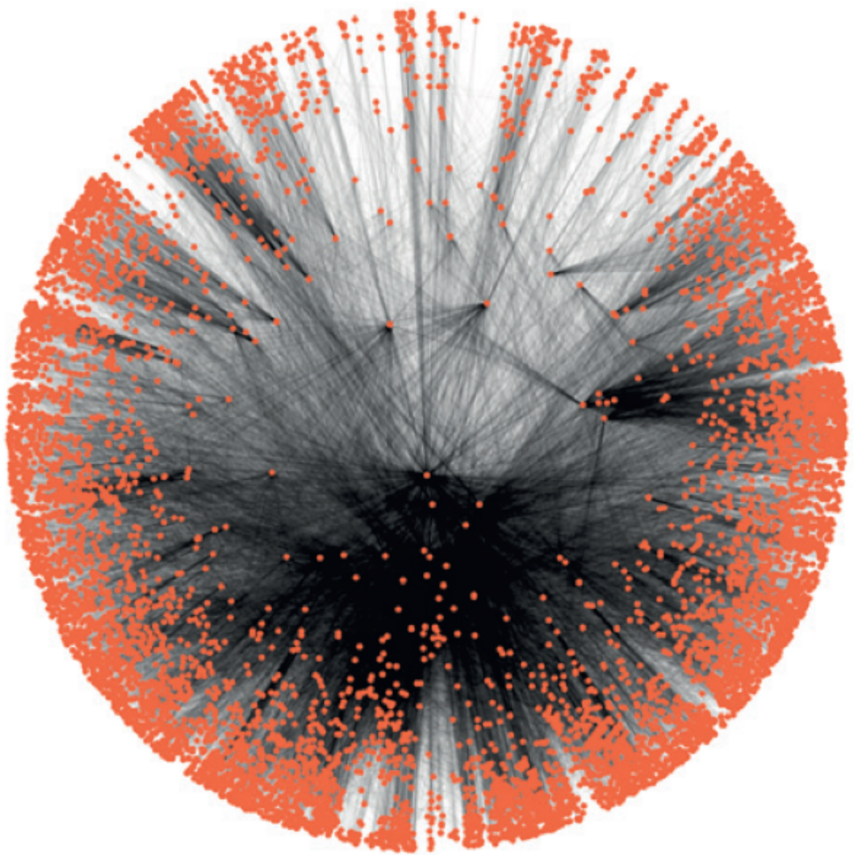
Navigability of networks

Other metrics of performance

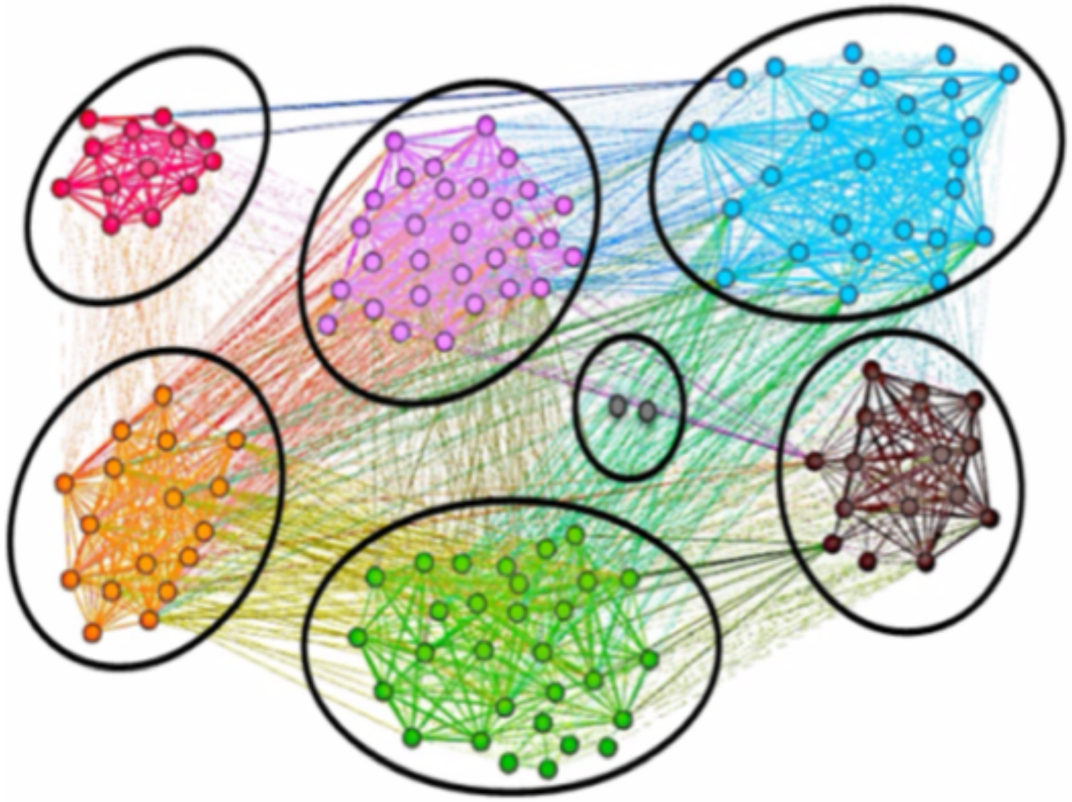


The analogy

Embedding in the hyperbolic space



Community structure



The analogy holds for real and artificial networks

Physical properties of networks can be (equally well) explained using either framework

Implications of the analogy

- Inter-community structure in networks may have geometric organization, meaning that at the global level, geometry dominates, while at the local scale, community memberships prevail

- Real networks may be modeled by a graphon consisting of a mixture of latent-spatial and block-like structures.

Characterizing the analogy between hyperbolic embedding and community structure of complex networks

A. Faqeeh, S. Osat and F. Radicchi

Phys. Rev. Lett. **121,** 098301 (2018)