Navigability of complex networks

Dmitri Krioukov
CAIDA/UCSD
M. Boguñá, M. Ángeles Serrano, F. Papadopoulos
A. Vahdat, kc claffy

UNT
January, 2010

Complex networks

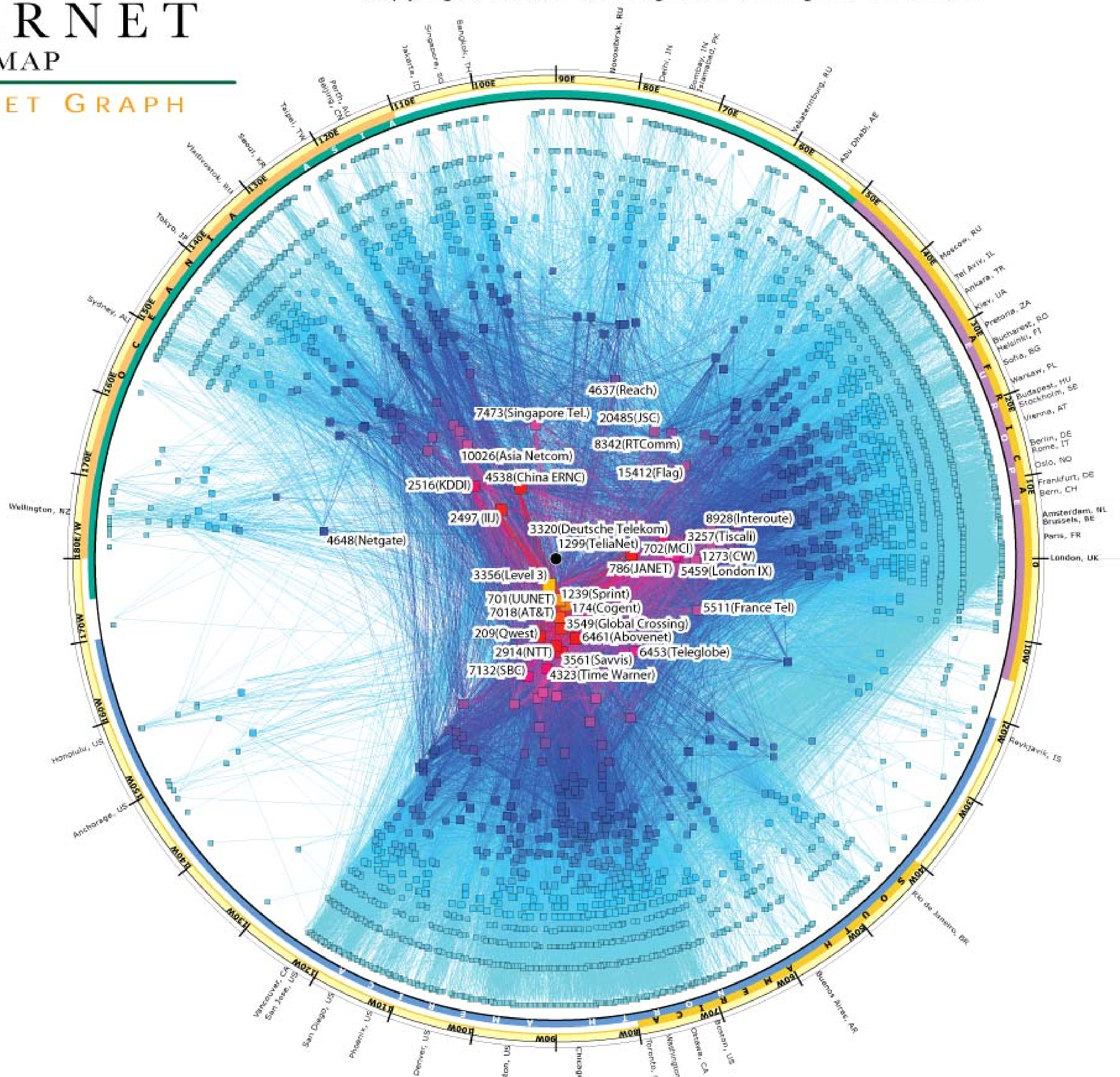
- Technological
  - Internet
  - Transportation
  - Power grid
- Social
  - Collaboration
  - Trust
  - Friendship
- Biological
  - Gene regulation
  - Protein interaction
  - Metabolic
  - Brain

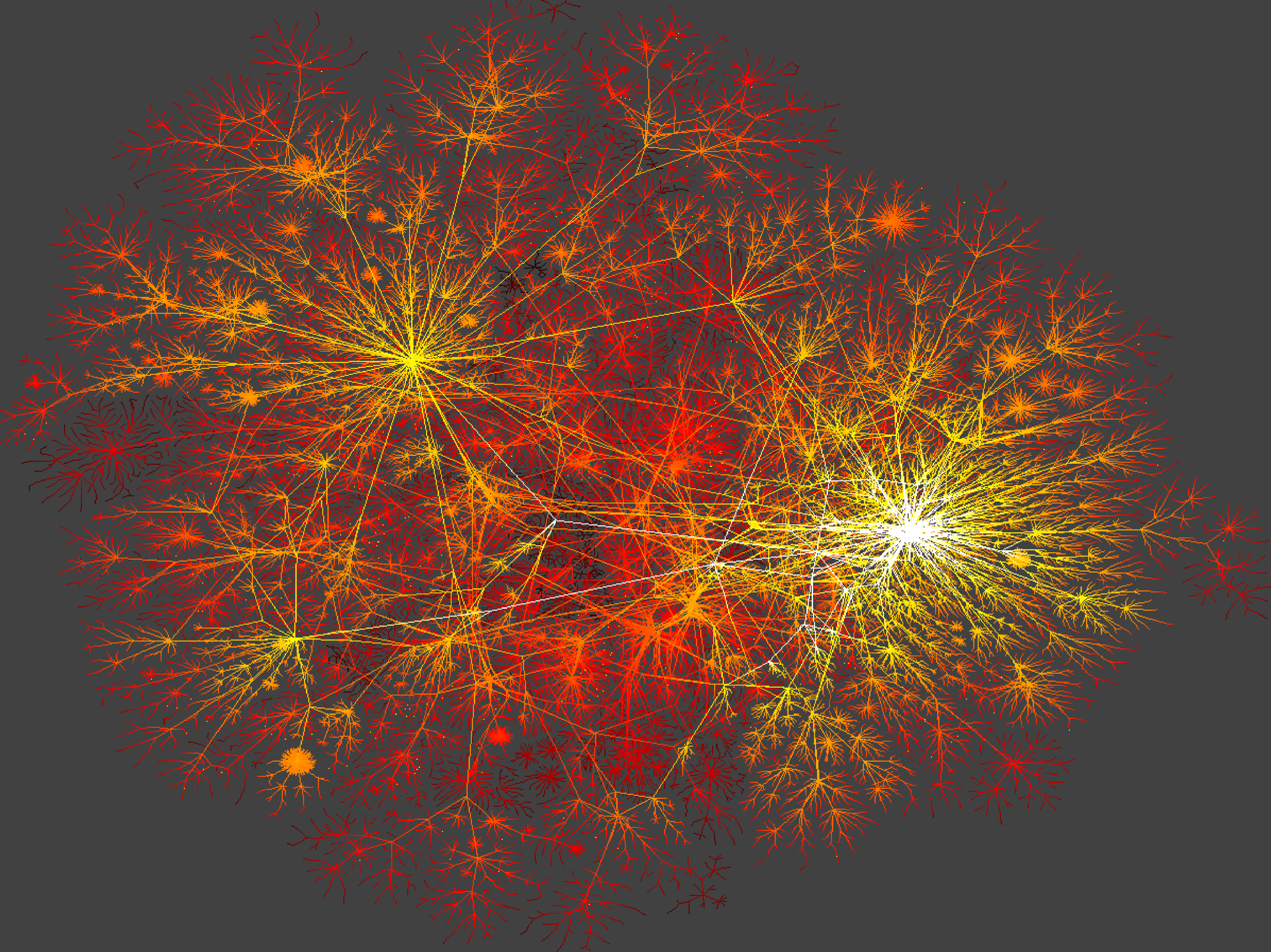
- Can there be anything common to all these networks???

  Naïve answer:
  - Sure, they must be complex
  - And probably quite random
  - But that’s it
- Well, not exactly!

IPv4 INTERNET
TOPOLOGY MAP

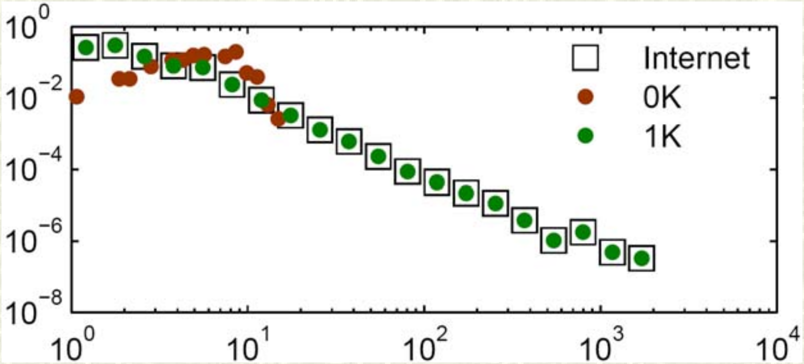
AS-level INTERNET GRAPH

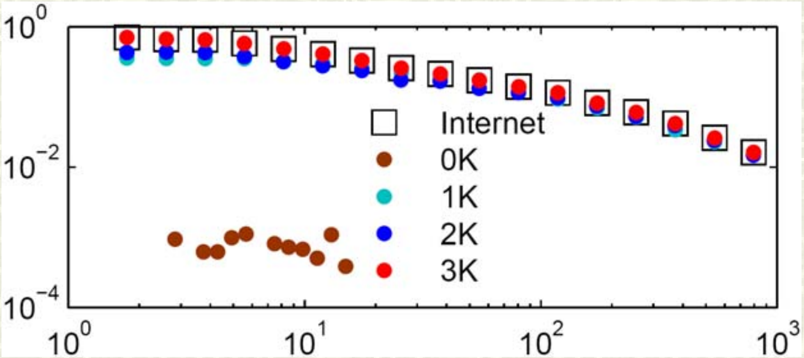




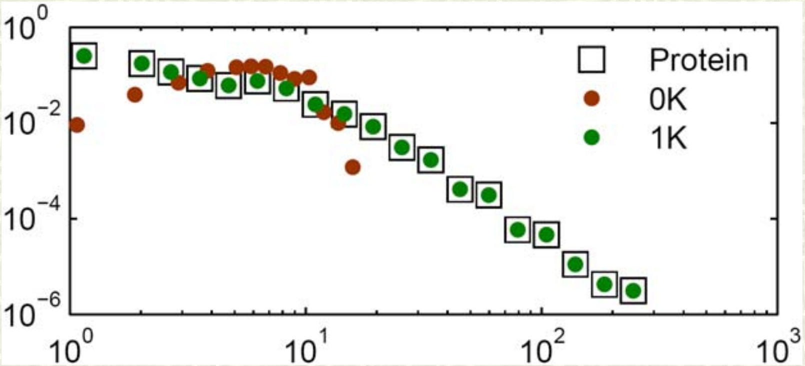
Internet

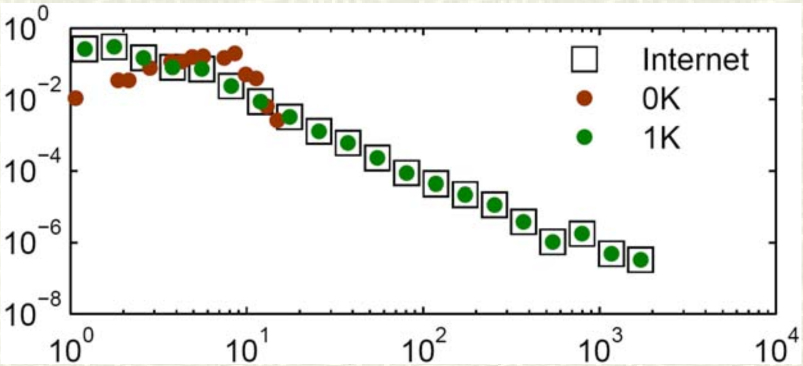
- **Heterogeneity:**
  distribution $P(k)$
  of node degrees k:
  - Real: $P(k) \sim k^{-\gamma}$
  - Random: $P(k) \sim \lambda^k e^{-\lambda}/k!$
- **Clustering:**
  average probability that
  node neighbors are connected:
  - Real: 0.46
  - Random: 6.8×10^{-4}

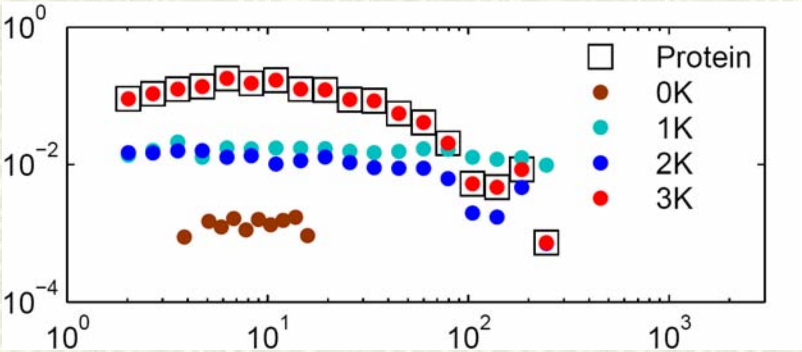


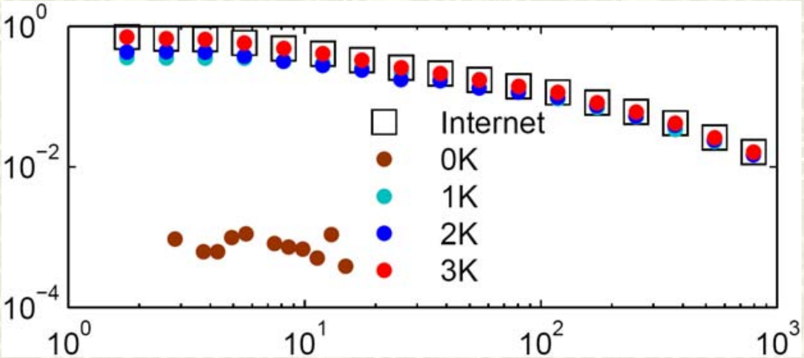


Internet vs. protein interaction









Common **structure** of complex networks: Strong heterogeneity and clustering

Network	**Exponent of the degree distribution**	**Average clustering**
Internet	*2.1*	*0.46*
Air transportation	*2.0*	*0.62*
Actor collaboration	*2.3*	*0.78*
Protein interaction *S. cerevisiae*	*2.4*	*0.09*
Metabolic *E. coli* **and** *S. cerevisiae*	*2.0*	*0.67*
Gene regulation *E. coli* **and** *S. cerevisiae*	*2.1*	*0.09*

Common **function** of complex networks: Transport or signaling phenomena

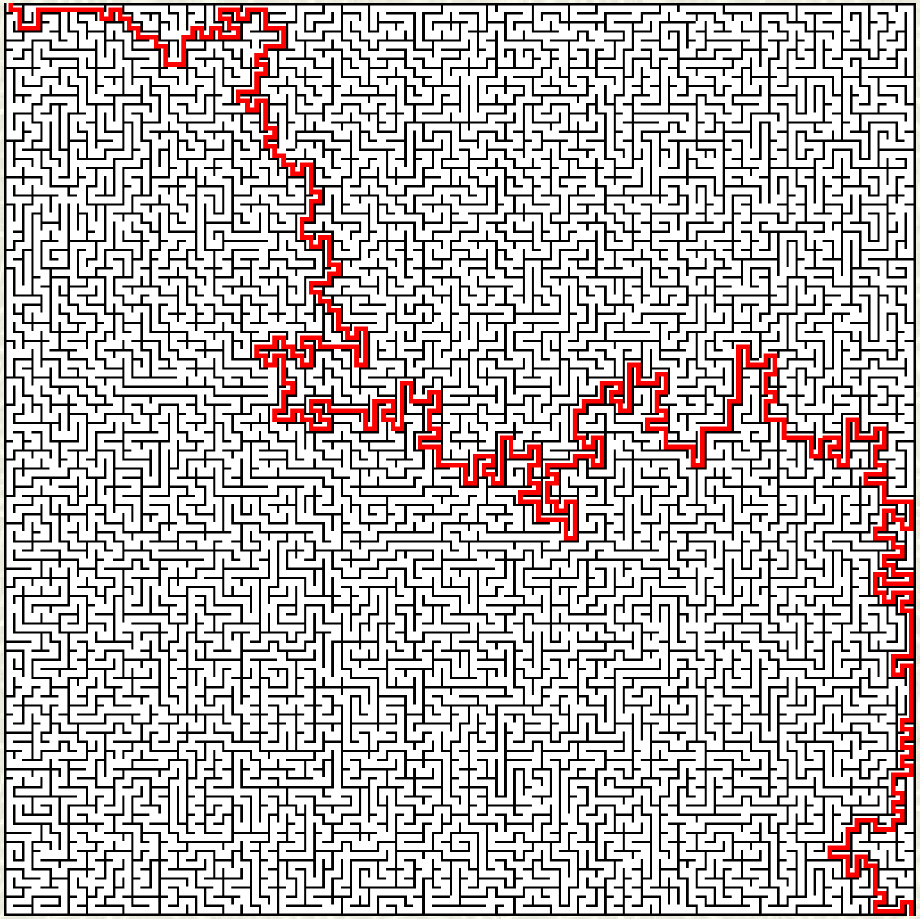
- Examples:
  - Brain
  - Internet
  - Transportation networks
  - Regulatory networks
  - Metabolic networks
  - Food webs
  - Social networks
- But in many networks, nodes do not know the topology of a network, its complex maze

Milgram’s experiments

- Settings: random people were asked to forward a letter to a random individual by passing it to their friends who they thought would maximize the probability of letter moving “closer” to the destination
- Results: surprisingly many letters (*30%*) reached the destination by making only ~*6* hops on average
- Conclusion:
  - People do not know the global topology of the human acquaintance network
  - But they can still find (short) paths through it

Complex networks as complex mazes

- To find a path through a maze is relatively easy if you have its map
- Can you quickly find a path if you are ***in*** the maze and don't have its map?
- Only if you have a compass, which does not lead you to dead ends
- Hidden metric spaces are such compasses

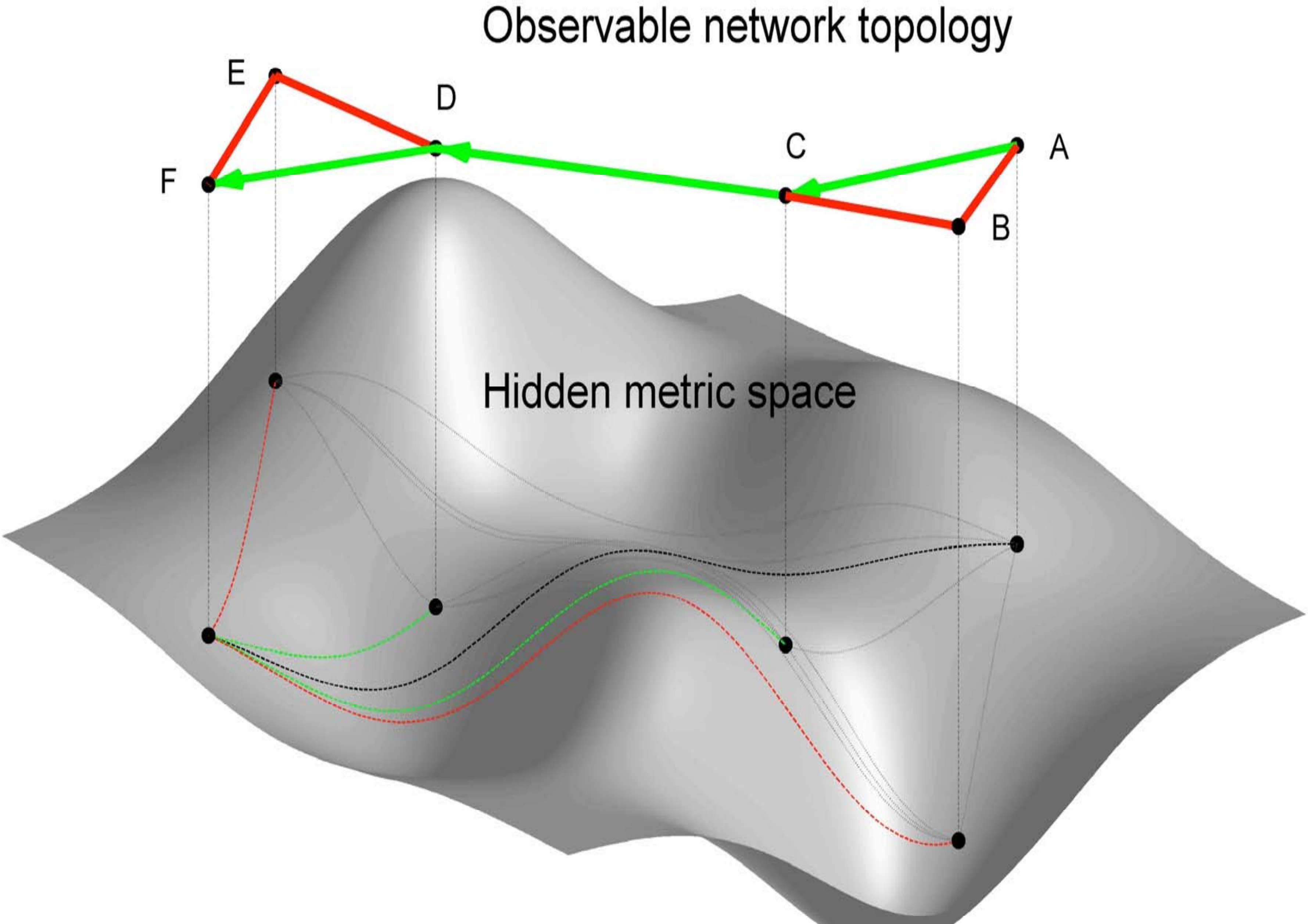


Take home message

- *Hidden metric spaces* explain common **structure <u>and</u> function** of complex networks

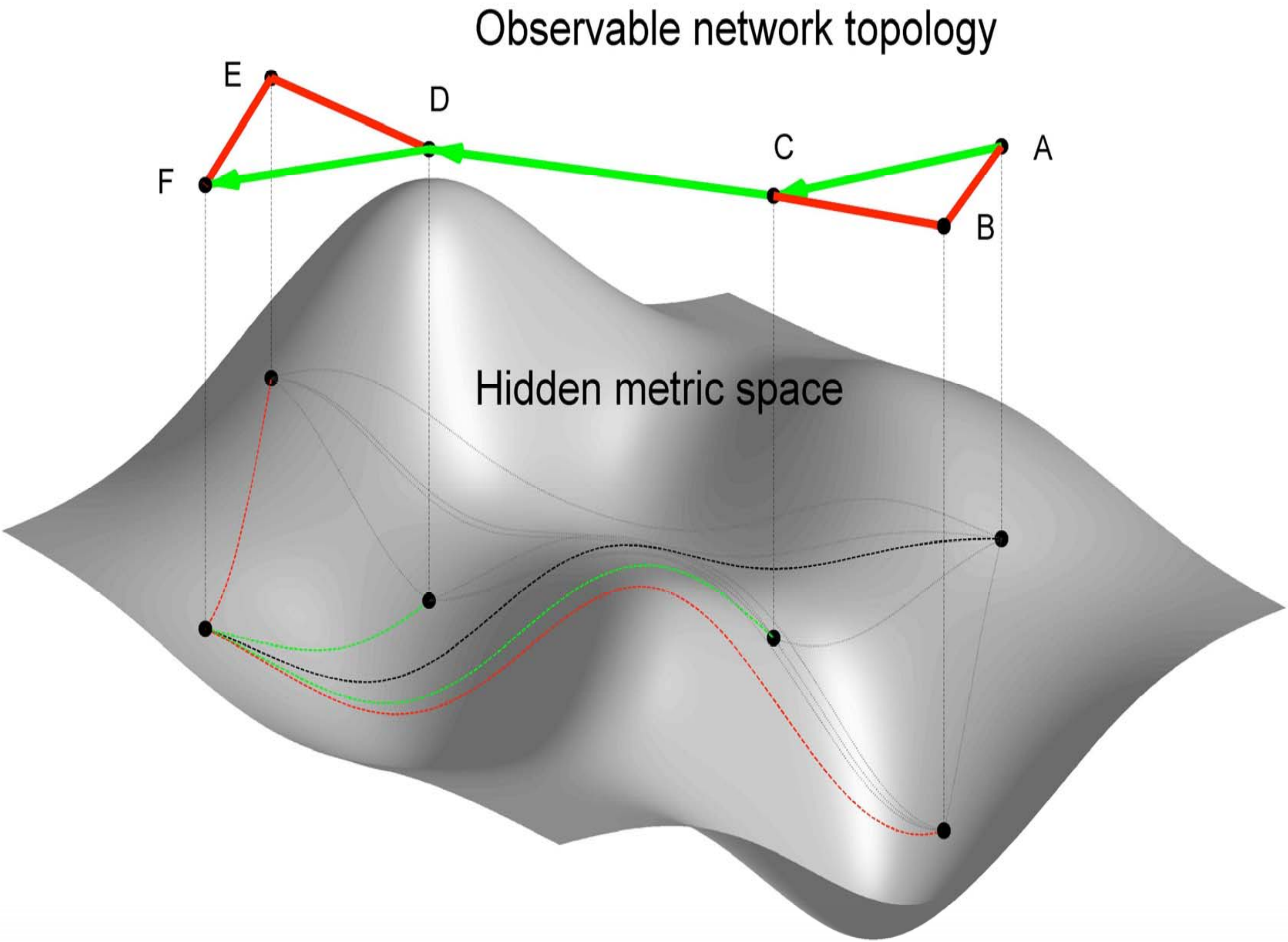
Hidden metric spaces

- All nodes in a network exist in a metric space
- Distances in this space abstract node similarities
  - More similar nodes are closer in the space
- Network consists of links that exist with probability that decreases with the hidden distance
  - More similar/close nodes are more likely to be connected


Observable network topology
E
D
C
A
F
B
Hidden metric space

Navigation by greedy routing

- To reach a destination, each node forwards information to the one of its neighbors that is closest to the destination in the hidden space


Observable network topology
E
D
C
A
F
B
Hidden metric space

Navigability metrics

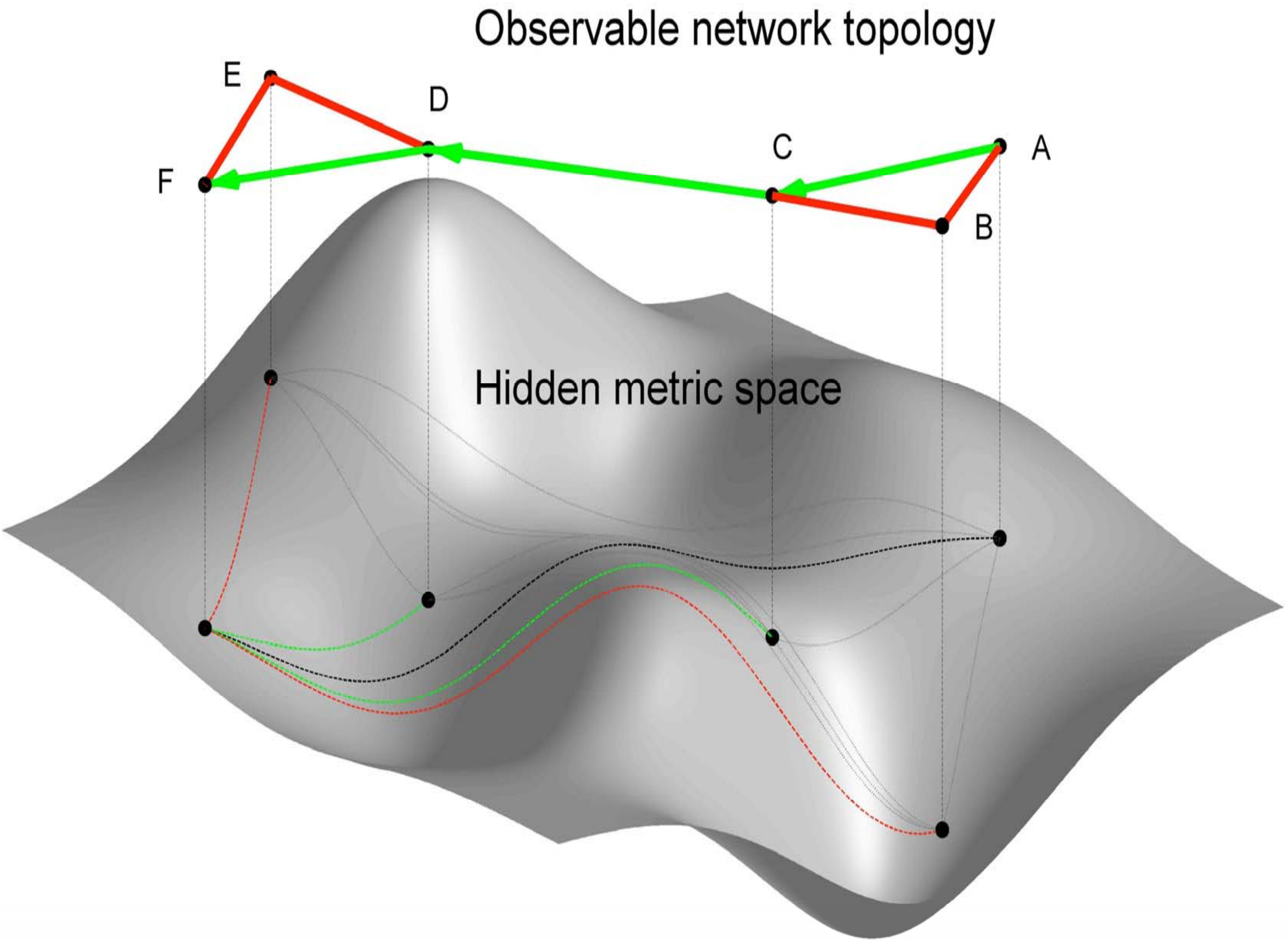
- Stretch
  - how much longer greedy paths are with respect to shortest paths in the network
- Success ratio
  - what percentage of greedy paths reach their destination without getting stuck at local minima, i.e., nodes that do not have any neighbors closer to the destination than themselves

Properties to focus on

- Structure
  - clustering
  - heterogeneity
- Function
  - stretch
  - success ratio

Clustering

- Clustering is a direct consequence of the triangle inequality in hidden metric spaces


Observable network topology
E
D
C
A
F
B
Hidden metric space

First empirical evidence: self-similarity of clustering

- Hidden metric spaces appear as the only reasonable explanation of one fine property of real networks – clustering self-similarity

Phys Rev Lett, v.100, 078701, 2008

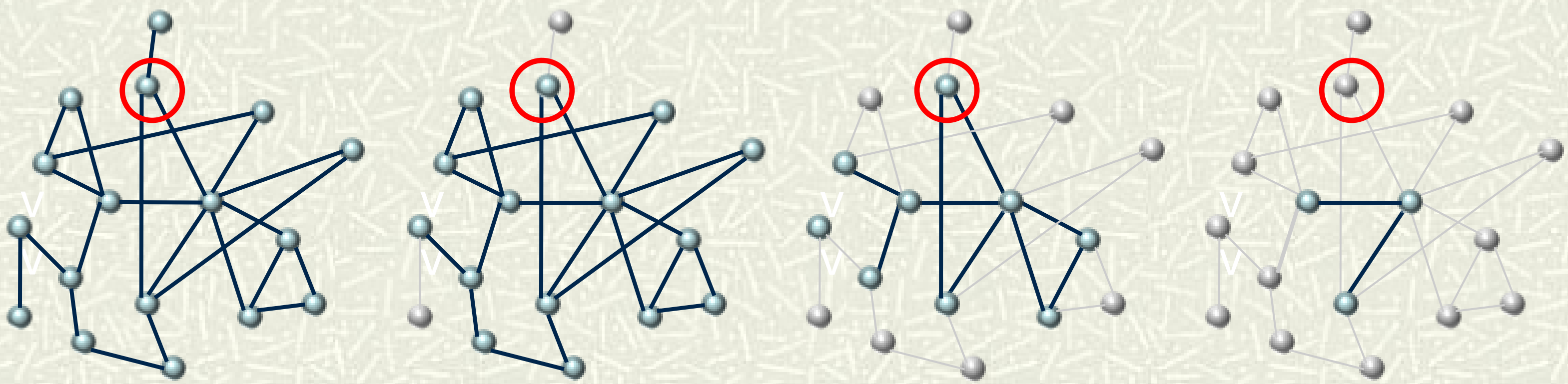
Clustering self-similarity

- Consider four networks
  - a real one, whose metric space we do not know
  - a synthetic one, with a modeled metric space underneath
  - randomized versions of both networks
- Degree-renormalize all four networks
- Compare clustering before and after renormalization:
  - original networks (real and synthetic): clustering is the same
  - randomized networks (real and synthetic): clustering is **not** the same
- Suggesting that as the synthetic network,
  the real network also has some metric structure underneath,
  which gets destroyed by randomization

Phys Rev Lett, v.100, 078701, 2008

Degree renormalization

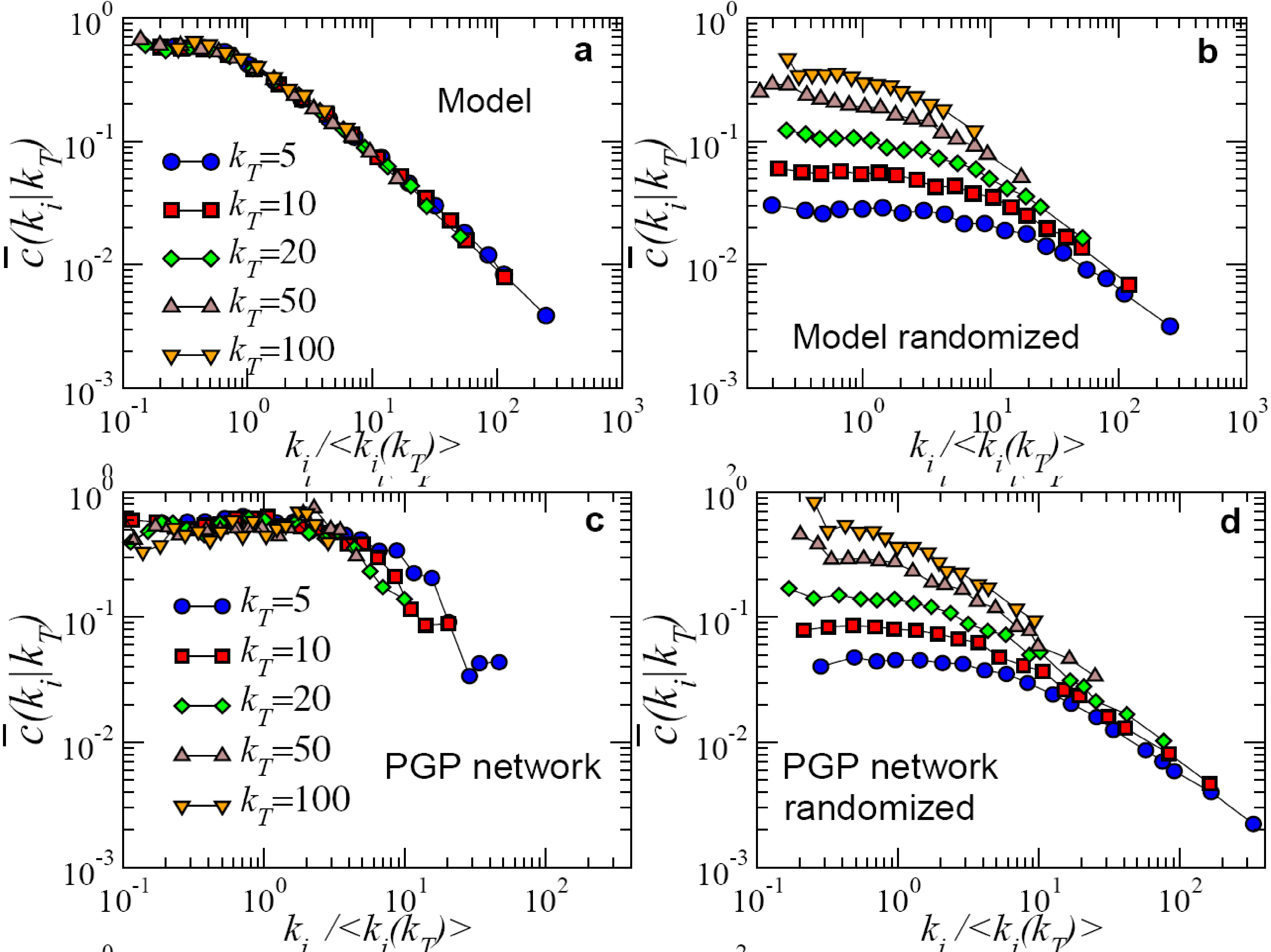
$k > k_T$



$k_T = 0$ $k_T = 1$ $k_T = 2$ $k_T = 3$

$$k_i / < k_i(k_T) >$$

Phys Rev Lett, v.100, 078701, 2008

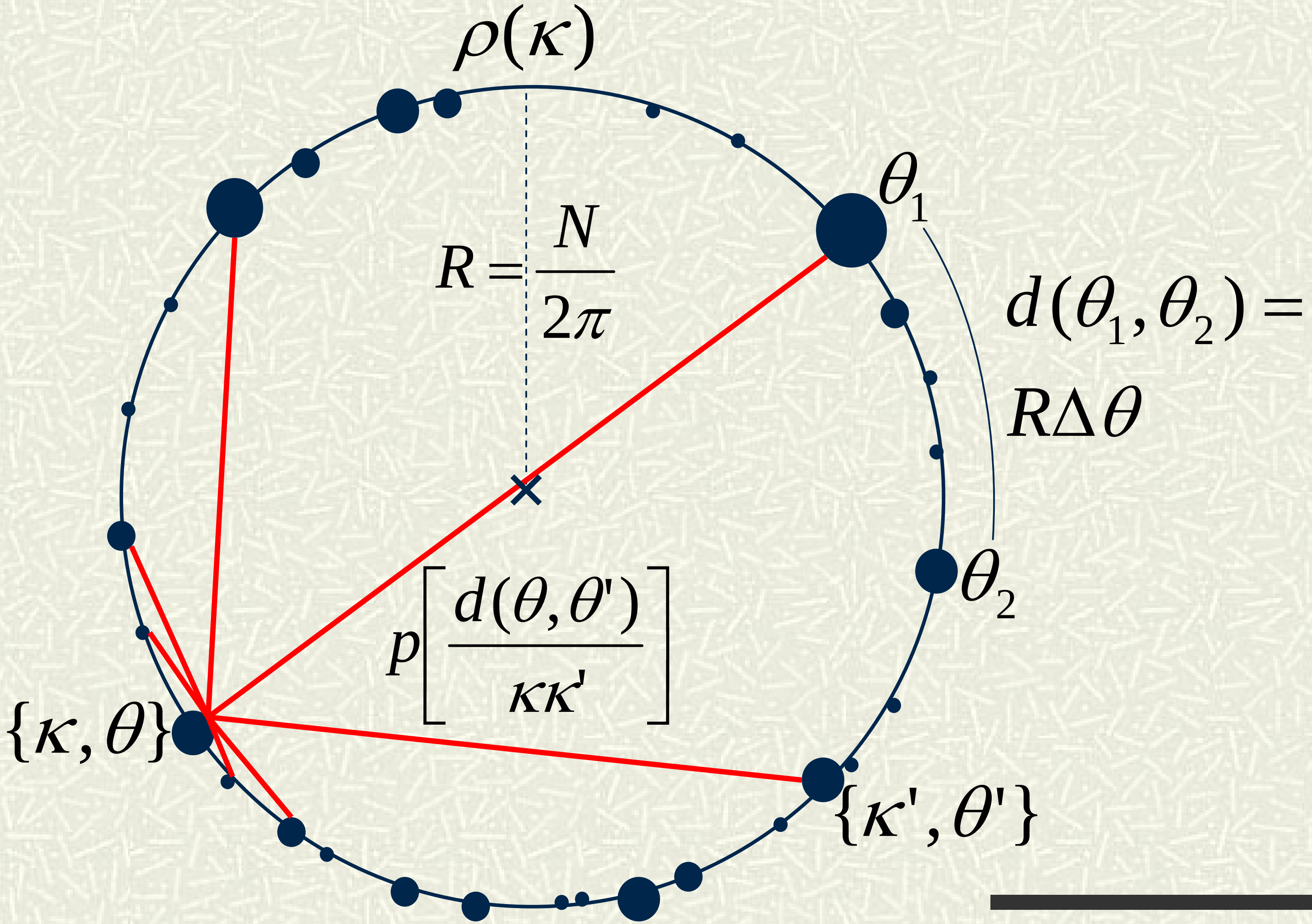

a
Model
k_T=5
k_T=10
k_T=20
k_T=50
k_T=100
$\bar{c}(k_i|k_T)$
$k_i/<k_i(k_T)>$
b
Model randomized
c
PGP network
d
PGP network
randomized

Modeling hidden metric spaces the simplest way – by a circle

- N nodes are randomly placed on a circle of radius
  $$N/(2\pi)$$
  - so that the node density is uniform ($=1$) on the circle
- All N nodes are assigned a random variable κ, the node expected degree, drawn from
  $$\rho(\kappa) = (\gamma-1)\kappa^{-\gamma}$$
- Each pair of nodes is connected with probability p, which must be an integrable function of
  $$\chi \sim \Delta\theta / (\kappa\kappa')$$
  - where $\Delta\theta$ is the angular distance between nodes, and κ, κ' are their expected degrees

Phys Rev Lett, v.100, 078701, 2008

The $\mathbb{S}^1$ model



Properties of the $\mathbb{S}^1$ model

- The model generates networks with
  - any (heterogeneous) degree distribution
    - by choosing different $\rho(\kappa)$
  - any clustering
    - by choosing different $p(\chi)$
  - a simple metric space underneath (obviously)
    - so that we can study network navigability
- Therefore, using the model, and having the metric space fixed (to $\mathbb{S}^1$), we may ask the question:
  - What combinations of degree distribution and clustering lead to maximum navigability?

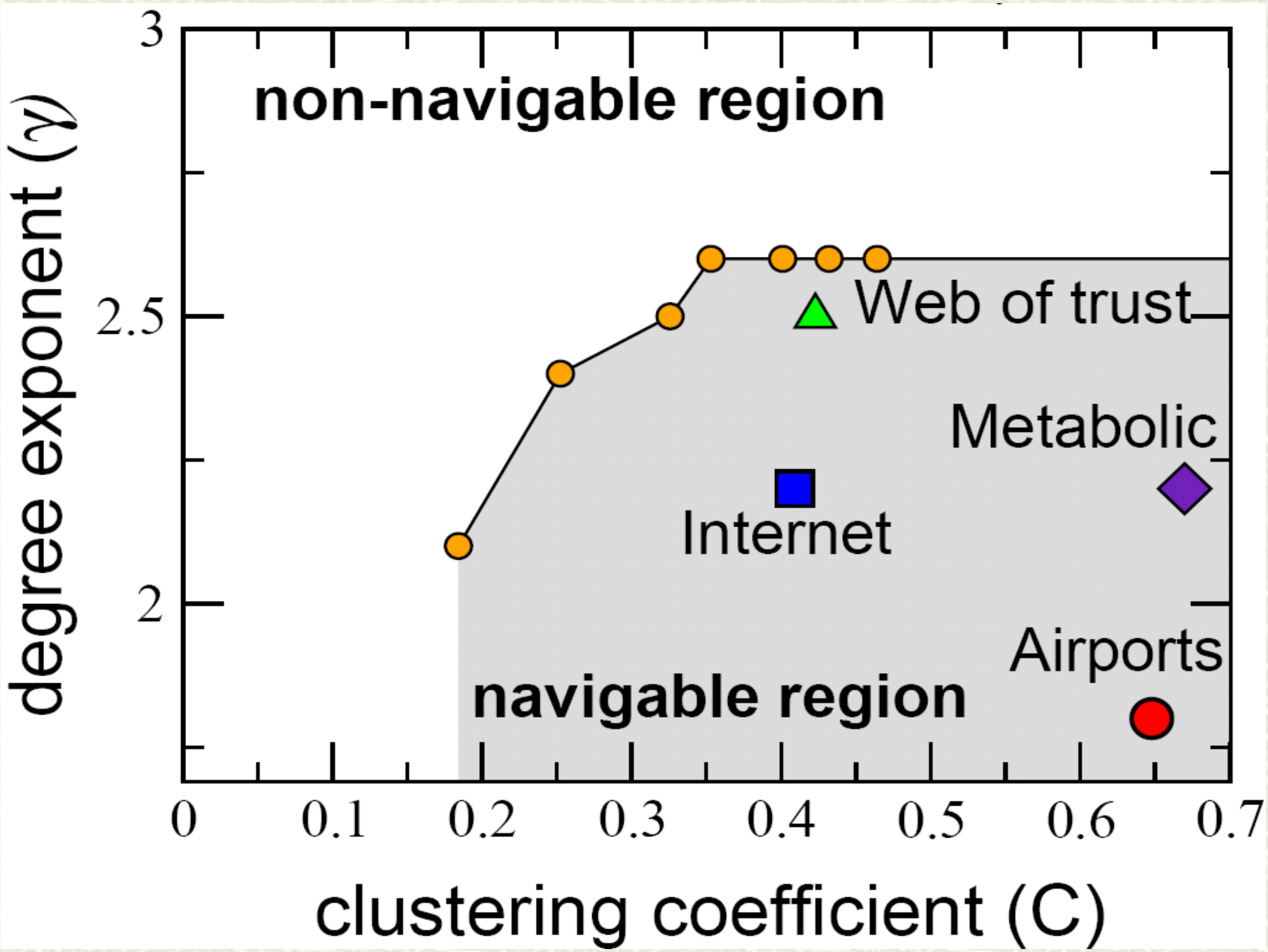
Phys Rev Lett, v.100, 078701, 2008

Ultrasmall **stretch** of ultrasmall worlds

- All successful greedy paths are asymptotically shortest (stretch = 1) in heterogeneous topologies with strong clustering
  - In fact, this statement holds for any uniform and isotropic hidden space geometry
  - But the success ratio does depend on this geometry

Phys Rev Lett, v.102, 058701, 2009

Success ratio in the $\mathbb{S}^1$ model



Nature Physics, v.5, p.74-80, 2009

Complex networks are navigable

- Specific values of degree distribution and clustering observed in real complex networks correspond to the highest efficiency of greedy routing in the $\mathbb{S}^1$ model
- Which implicitly suggests that complex networks evolve to navigable configurations
- If they did not, they would not be able to function

Nature Physics, v.5, p.74-80, 2009

One caveat and one question

- The maximum value of the success ratio observed in the $\mathbb{S}^1$ model is 65%
- Is there a space, other than $\mathbb{S}^1$, that brings the maximum success ratio close to 100%?

Nature Physics, v.5, p.74-80, 2009

One answer

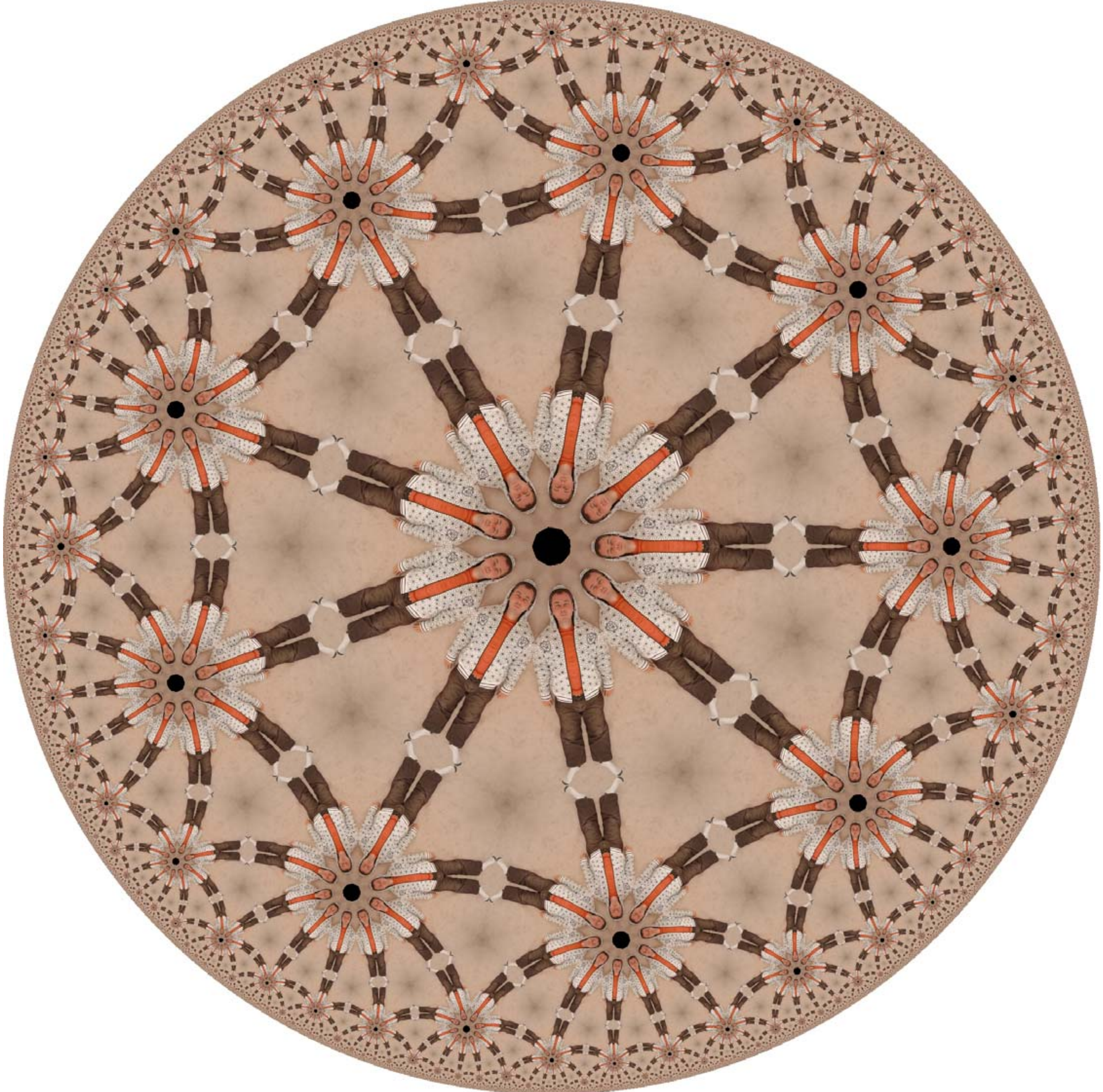
- The answer is yes!
- The maximum success ratio reaches 100% if the hidden space is hyperbolic ($\mathbb{H}^2$)

Phys Rev E, v.80, 035101(R), 2009

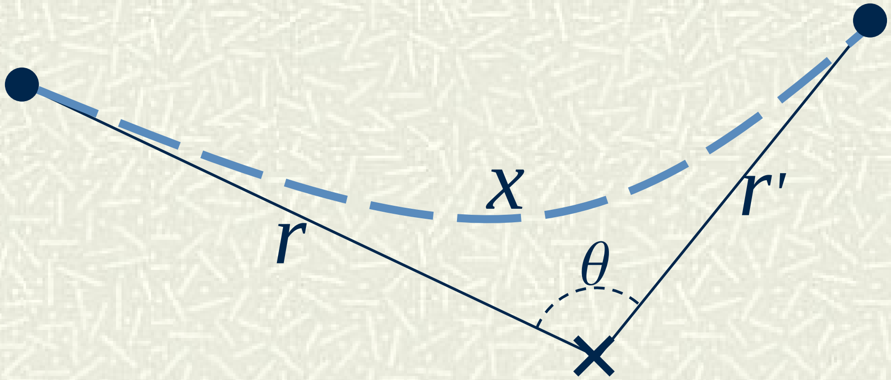
Two facts on hyperbolic geometry

- Exponential expansion of space
- Distance calculations

Phys Rev E, v.80, 035101(R), 2009



Hyperbolic distance



- $\cosh x = \cosh r \cosh r' - \sinh r \sinh r' \cos \theta$
- $x \approx r + r' + 2 \ln \sin(\theta/2)$
- $x \approx r + r' + 2 \ln(\theta/2)$

Phys Rev E, v.80, 035101(R), 2009

The $\mathbb{S}^1$-to-$\mathbb{H}^2$ transformation

- Change of variables from κ (expected degree) to r (radial coordinate)

$$\kappa = e^{(R-r)/2}$$

  - where $R = 2 \ln(N/c)$

- yields the radial node density

$$\rho(r) = \alpha\, e^{\alpha(r-R)}$$

  - where
    $\alpha = (\gamma - 1) / 2$

- and the argument of the connection probability

$$\chi = e^{(x-R)/2}$$

  - where x is the hyperbolic distance between nodes

Phys Rev E, v.80, 035101(R), 2009

The other way around: The native $\mathbb{H}^2$ model

- The hidden space is the simplest hyperbolic space – a disc (of radius $R = 2 \ln(N/c)$)
- Distribute nodes (quasi-)uniformly on it:
  - the angular node density is uniform
  - the radial node density is exponential (because the space is hyperbolic!)

$$\rho(r) = \alpha\, e^{\alpha(r-R)}$$

- Connect each pair of nodes with probability

$$p[\chi] = p[e^{(x-R)/2}]$$

- The resulting node degree distribution is a power law

$$P(k) \sim k^{-\gamma}$$

  - where $\gamma = 2\,\alpha + 1$

Phys Rev E, v.80, 035101(R), 2009

Two properties of the $\mathbb{H}^2$ model

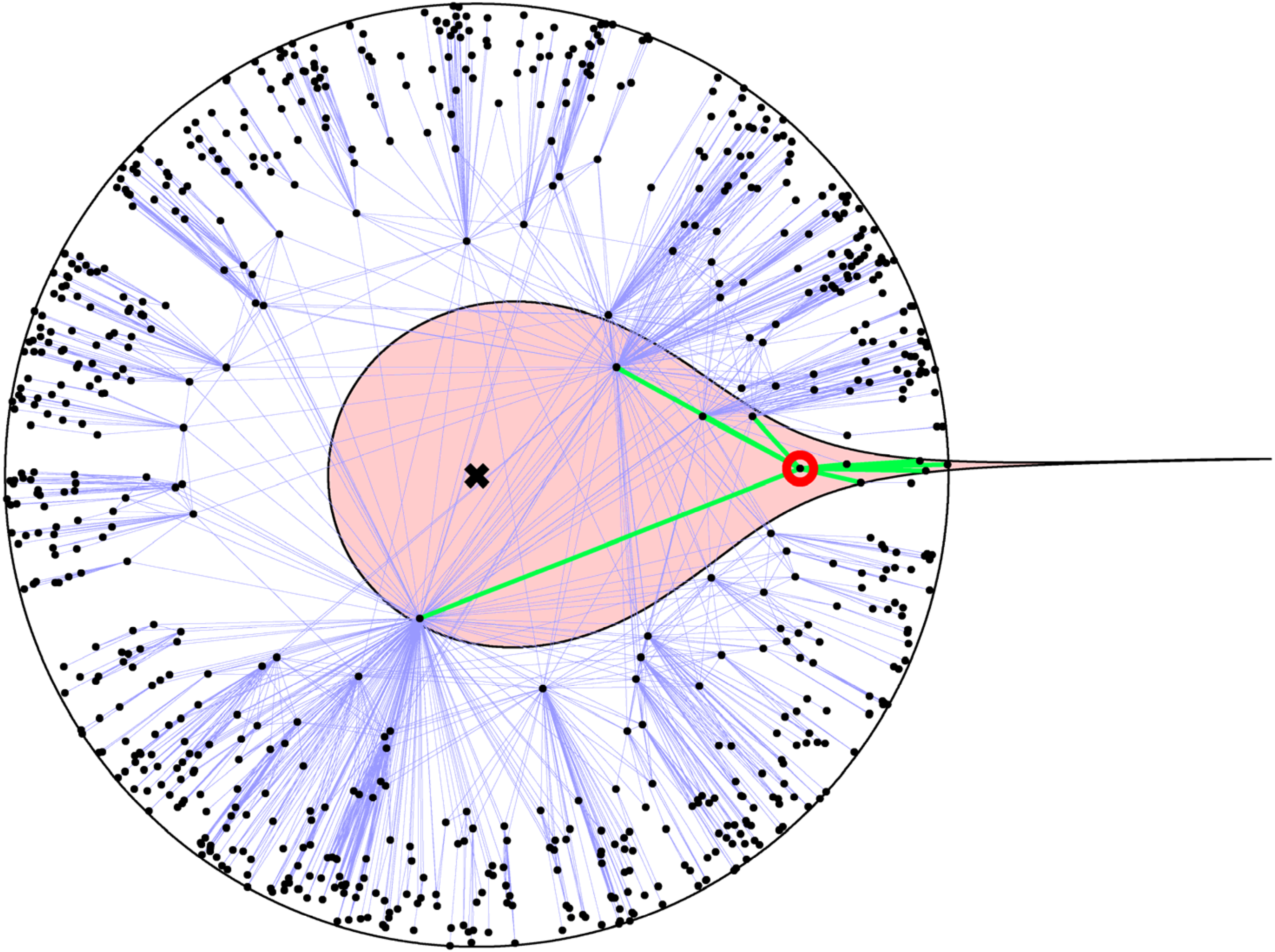
- Network heterogeneity emerges naturally as a simple consequence of the exponential expansion of space in hyperbolic geometry
- The choice of the Fermi-Dirac connection probability

$$p[x] = \frac{1}{e^{(x-R)/(2T)}+1} \xrightarrow[T\to 0]{} \Theta(R-x)$$

yields the following physical interpretation:

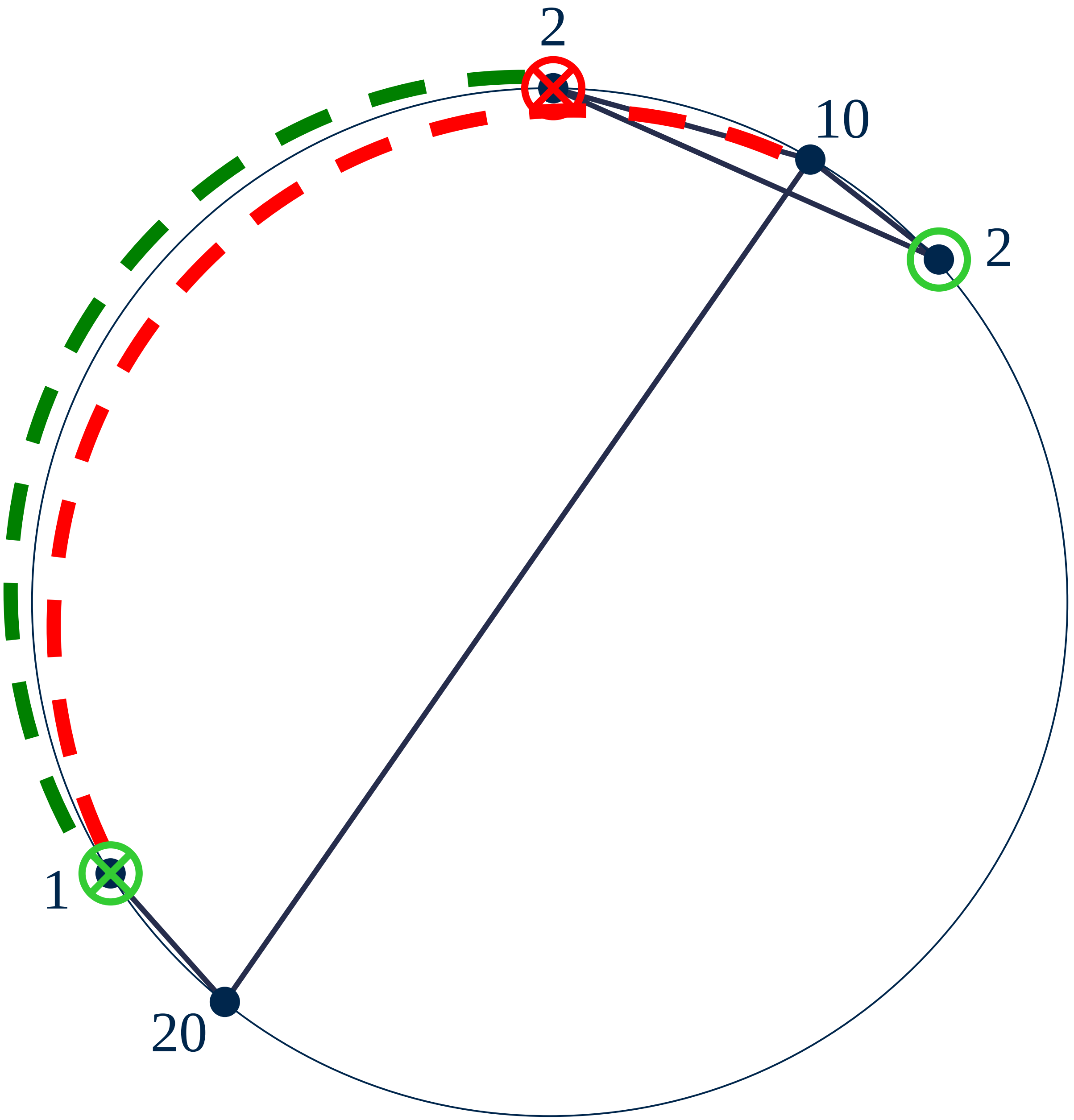
  - Hyperbolic distances x are energies of the corresponding links-fermions
  - Hyperbolic disc radius R is the chemical potential
  - Clustering-controlling parameter T is the system temperature

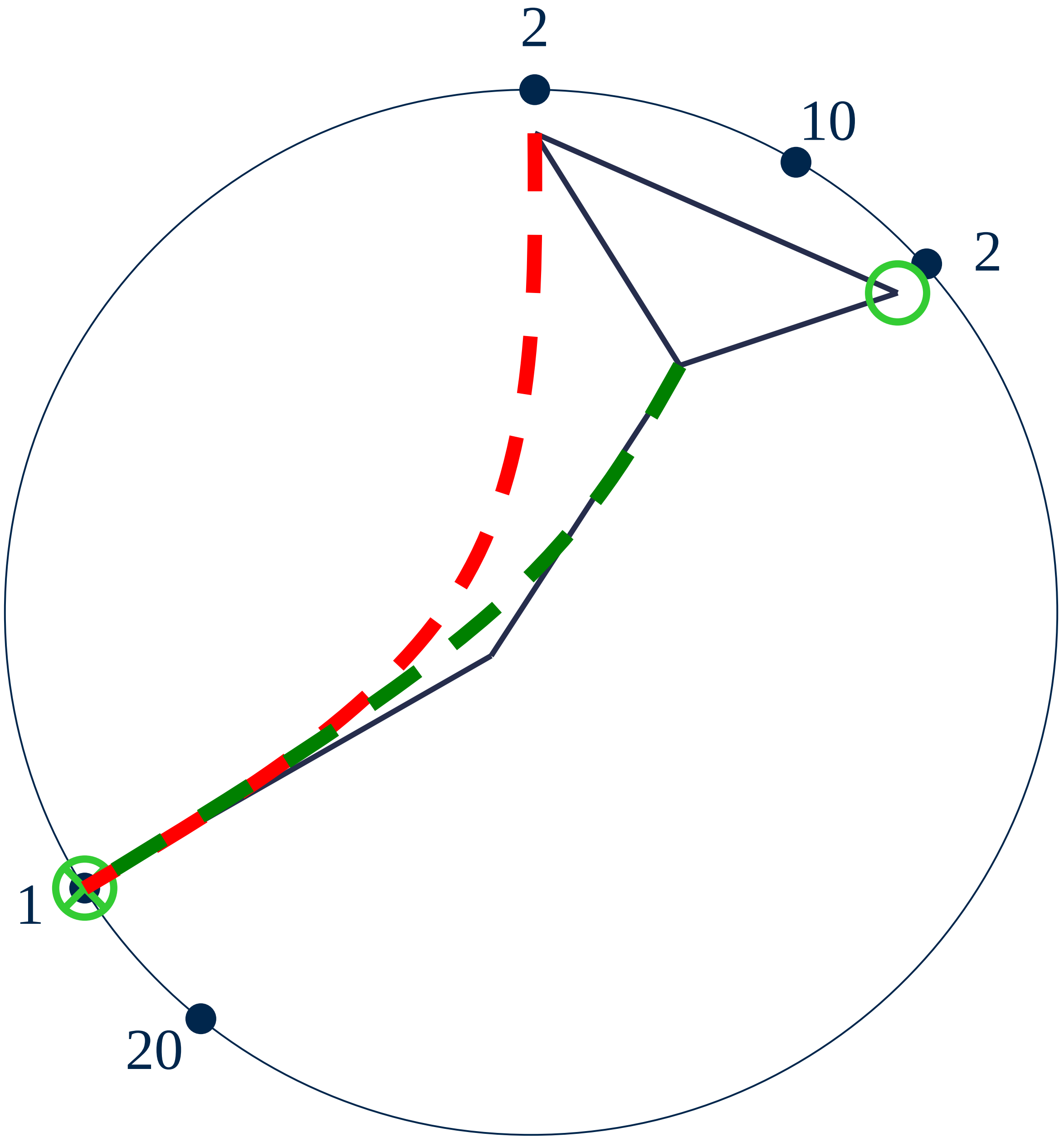
Phys Rev E, v.80, 035101(R), 2009

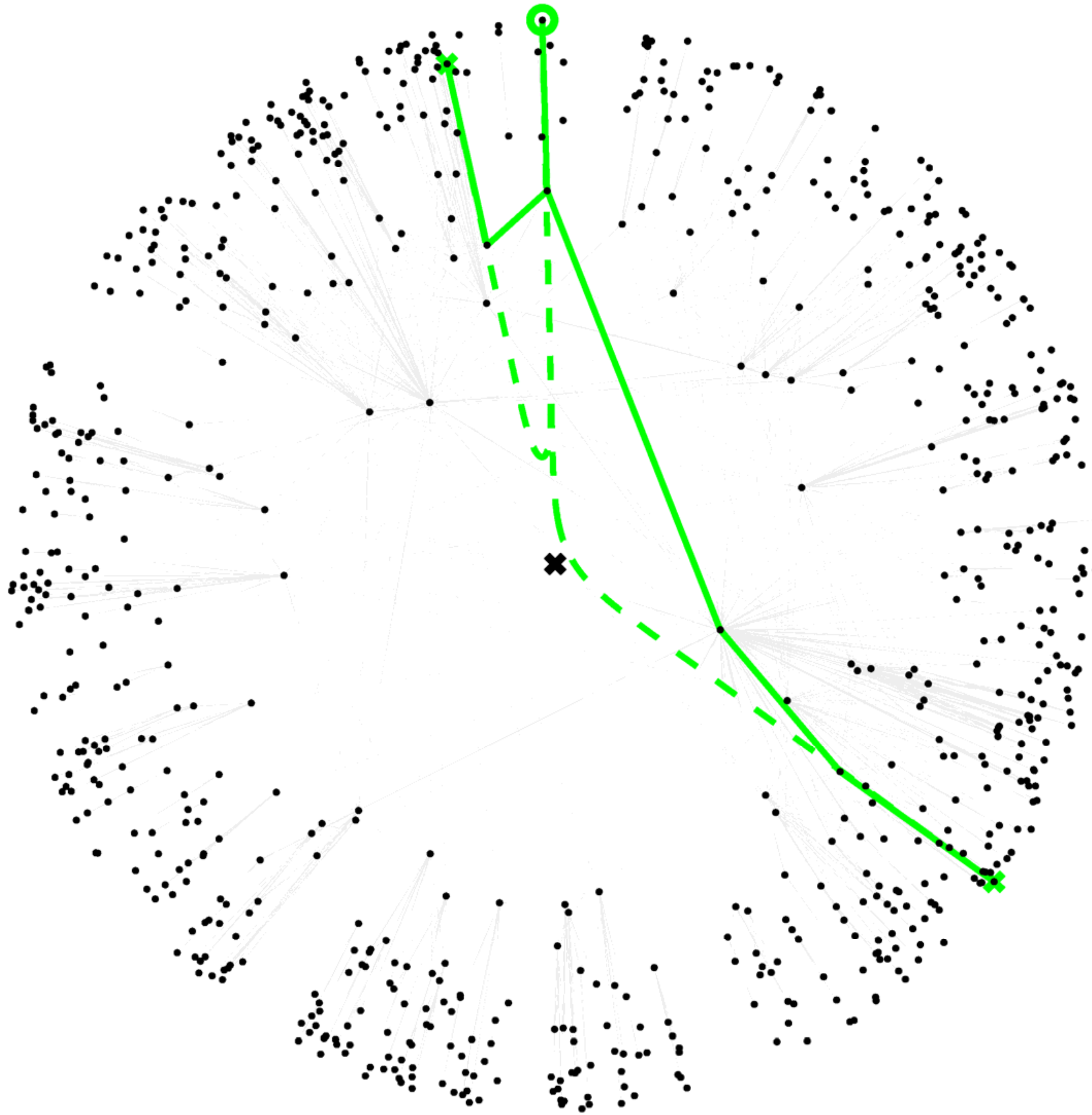


Why navigation in $\mathbb{H}^2$ is more efficient than in $\mathbb{S}^1$

- Because nodes in the $\mathbb{S}^1$ model are not connected with probability which depends solely on the $\mathbb{S}^1$ distances $\sim\Delta\theta$
- Those distances are rescaled by node degrees to $\chi \sim \Delta\theta / (\kappa\kappa')$ (to guarantee that $k(\kappa) = \kappa$)
- These rescaled distances are hyperbolic (after the κ-to-r change of variables)
- Intuitively, navigation is more efficient if it uses more congruent distances, i.e., those with which the network is built


2
10
2
1
20

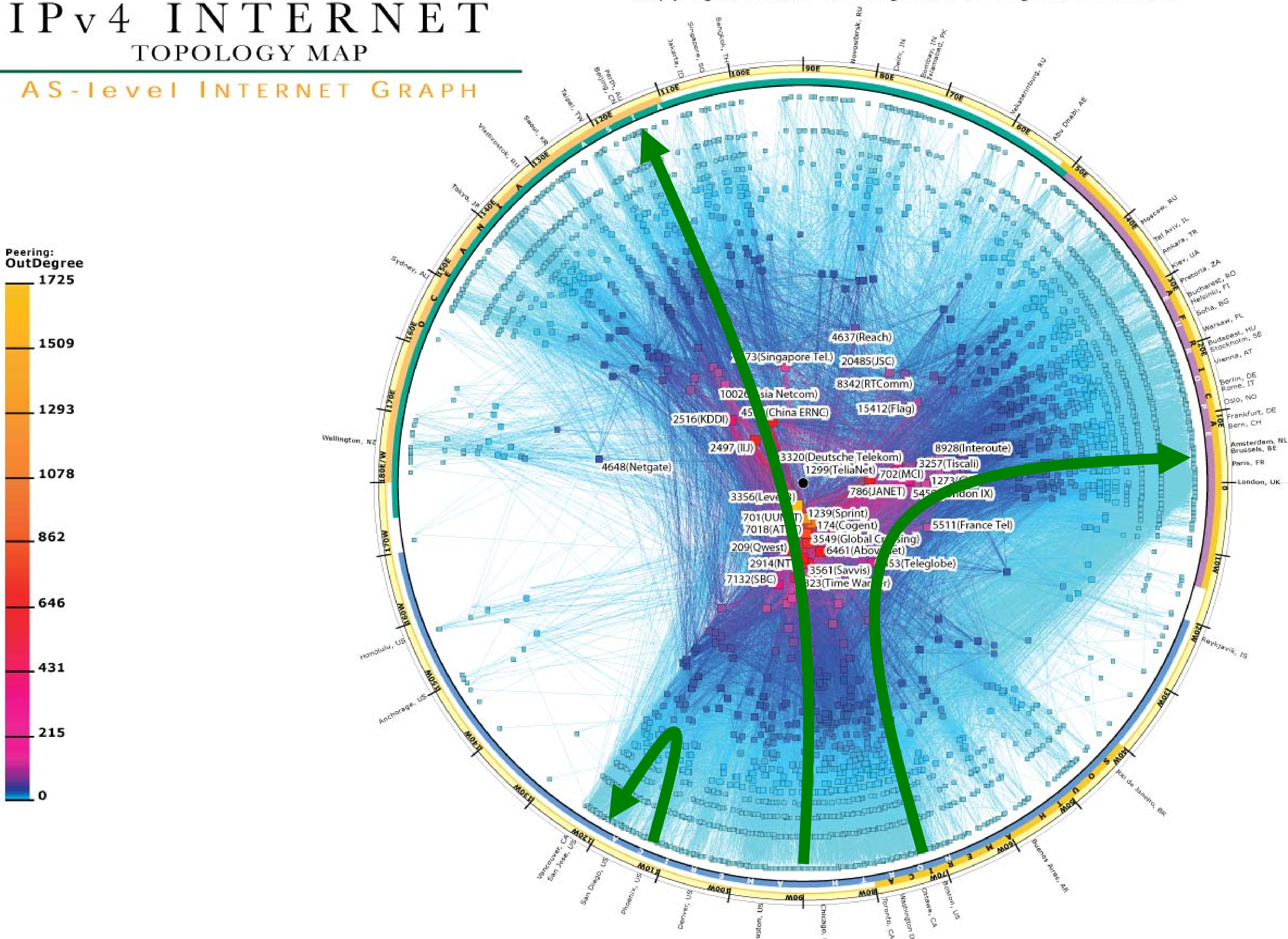

2
10
2
1
20



IPv4 INTERNET

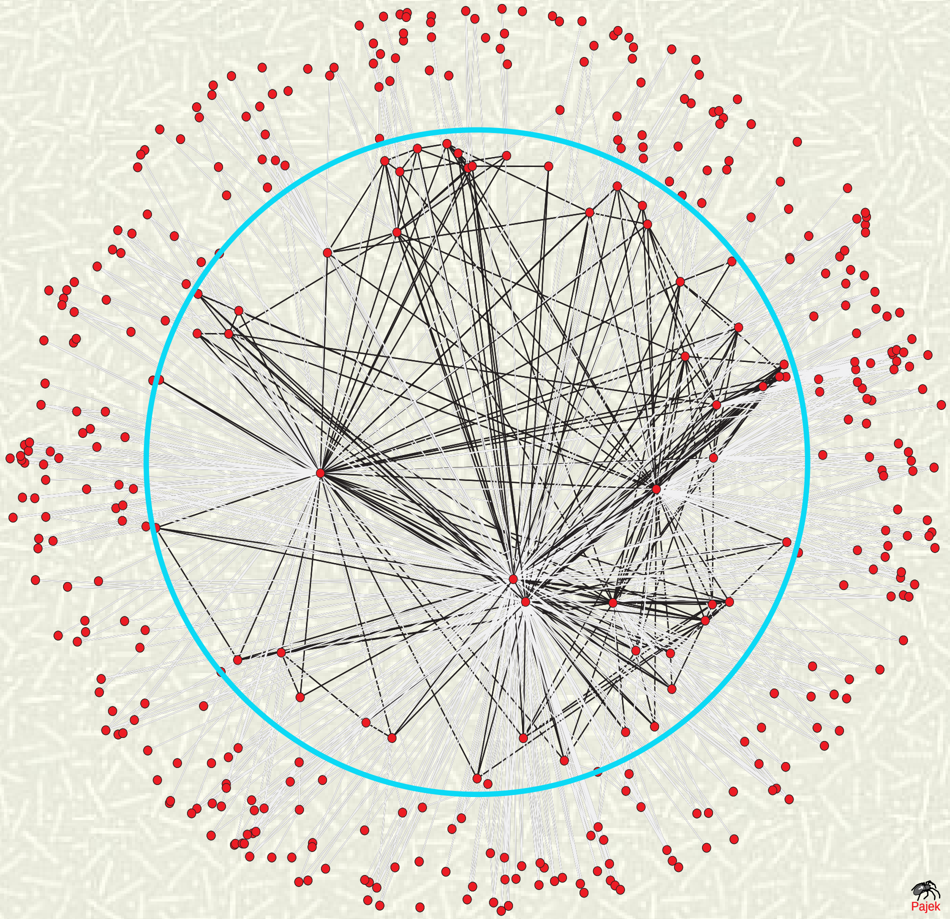
TOPOLOGY MAP

AS-level INTERNET GRAPH



…and back to self-similarity

- Degree-thresholding renormalization is a homothety along the radial coordinate
- Such homotheties are symmetry transformations in hyperbolic geometry
- Self-similarity of complex networks proves not only that hidden spaces exist, but also that they are hyperbolic



Why hyperbolic spaces

- Nodes in complex networks can often be hierarchically classified
  - Community structure (social and biological networks)
  - Customer-provider hierarchies (Internet)
  - Hierarchies of overlapping attribute sets (all networks)
- Hierarchies are (approximately) trees
- Hyperbolic spaces are tree-like
  (trees embed almost isometrically in them)

Phys Rev E, v.80, 035101(R), 2009

Take home message

- Hidden hyperbolic metric spaces explain the common structure and function of complex networks:
  - Structure:
    - Strong clustering is a consequence of the fact that hyperbolic spaces are metric
    - Heterogeneity is a consequence of their negative curvature
  - Function:
    - Stretch is 1, i.e., all greedy paths are shortest
    - Success ratio is 100%, i.e., all greedy paths are successful

Current work

- We have a formal proof that stretch is 1, but we do not yet have a formal proof that success ratio is 100%
- Mapping real networks to their metric spaces. Two paths:
  - **Brute force**: use statistical inference techniques (e.g., MLE) to map a network to a model space
    - Requires involved manual intervention
    - Algorithm running times are prohibitive for large networks
  - **Constrictive**: construct a map based on intrinsic node similarities
    - What node attributes to choose to compute similarities
    - Many similarity metrics exist. Which one to choose?

One immediate application

- Succeeded in brute-force mapping the Internet to $\mathbb{H}^2$
  - Stretch is almost 1
  - Success ratio is almost 100%
- Thus resolving long-standing scalability problems with existing Internet routing
  - Existing Internet routing is based on global knowledge of the topology
  - If topology changes, the information about the change must be diffused to all the routers
  - Which involves enormous and ever-growing communication overhead
    - over-whelmed routers fail,
      endangering the performance and stability of the global Internet
    - black holes have started appearing in the Internet already
  - Greedy routing does not require any global knowledge,
    thus resolving the scaling limitations of routing in the Internet

Future potential applications

- Upon successful mapping a network to its metric space, we can:
  - provide a new perspective on community detection
    - instead of splitting nodes into discrete communities, we have a continuous measure of similarity for each two nodes – their hidden distance
    - “communities” are then zones of higher node density in the hidden space
  - improve recommender systems
    - based on user similarity, predict what movies (Netflix) or goods (Amazon) a user will like
  - predict what conditions lead to the appearance of undesirable local minima – examples of such ***lethal dead ends*** include:
    - cancer
    - protein mis-folding
    - brain mal-function
    - etc

Further details

- M. Ángeles Serrano, D. Krioukov, and M. Boguñá,
  Self-Similarity of Complex Networks and Hidden Metric Spaces,
  Physical Review Letters, v.100, 078701, 2008
- M. Boguñá, D. Krioukov, and kc claffy,
  Navigability of Complex Networks,
  Nature Physics, v.5, p.74-80, 2009
- M. Boguñá and D. Krioukov,
  Navigating Ultrasmall Worlds in Ultrashort Time,
  Physical Review Letters, v.102, 058701, 2009
- D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá,
  Curvature and Temperature of Complex Networks,
  Physical Review E, v.80, 035101(R), 2009