Hyperbolic geometry of complex networks

Dmitri Krioukov
CAIDA/UCSD
dima@caida.org
F. Papadopoulos, M. Boguñá, A. Vahdat, and kc claffy

Complex networks

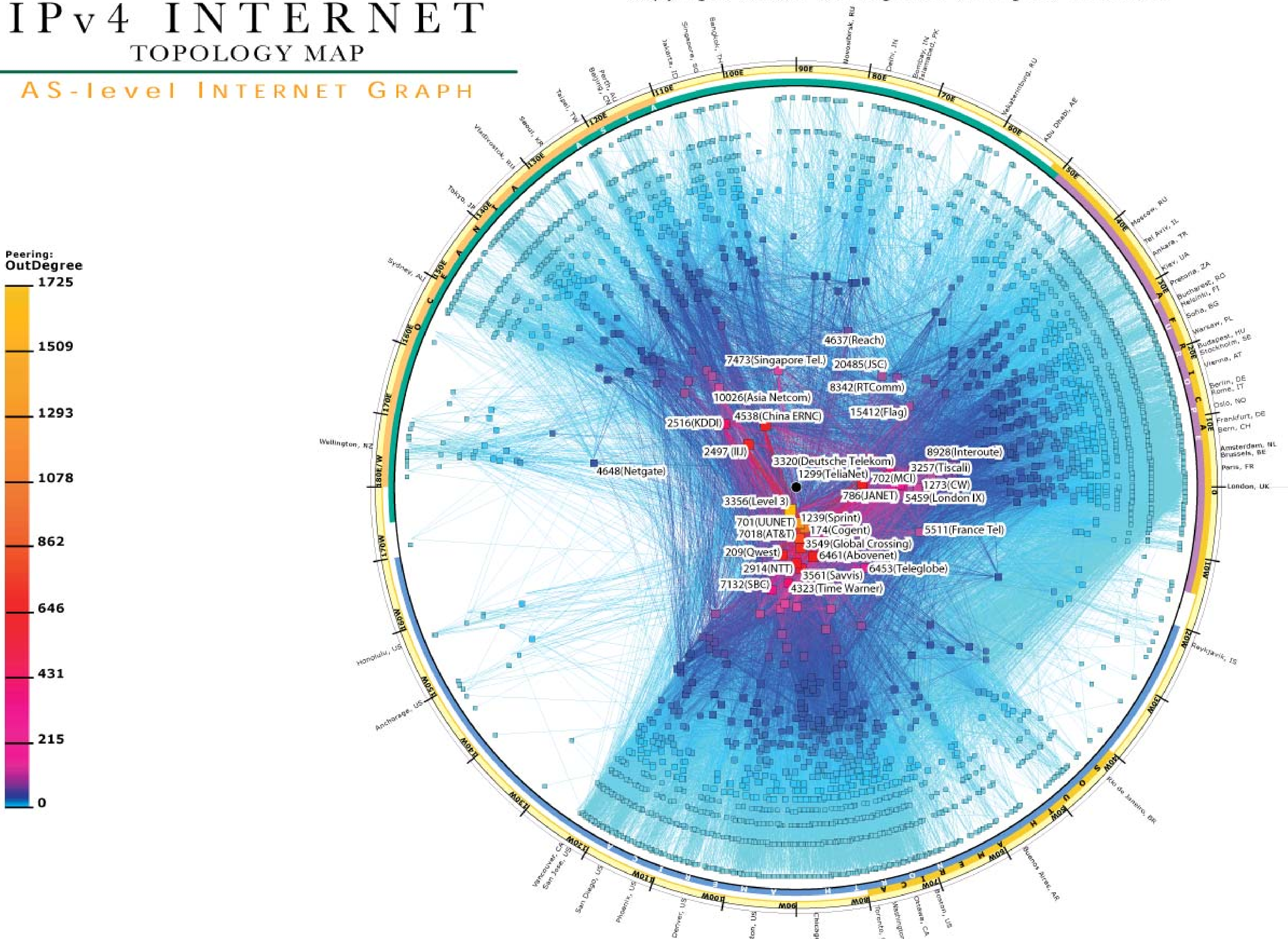
- Technological
  - Internet
  - Transportation
  - Power grid
- Social
  - Collaboration
  - Trust
  - Friendship
- Biological
  - Gene regulation
  - Protein interaction
  - Metabolic
  - Brain

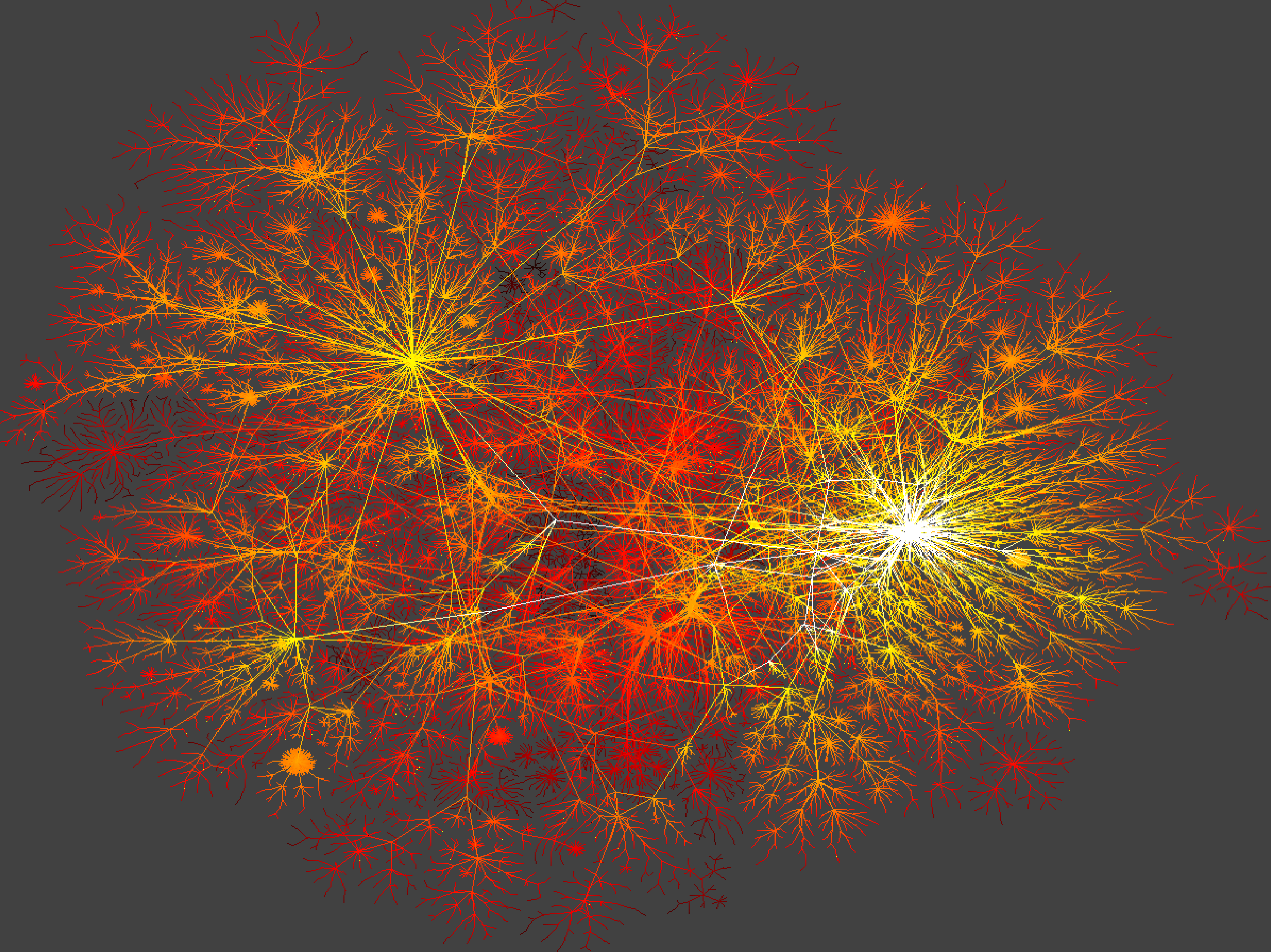
- Can there be anything common to all these networks???

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

IPv4 INTERNET
TOPOLOGY MAP

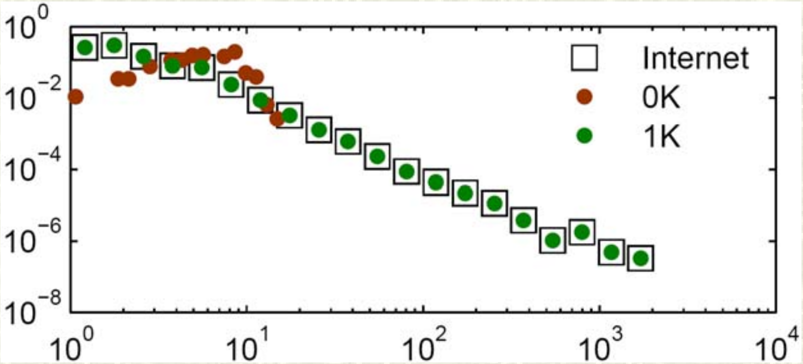
AS-level INTERNET GRAPH

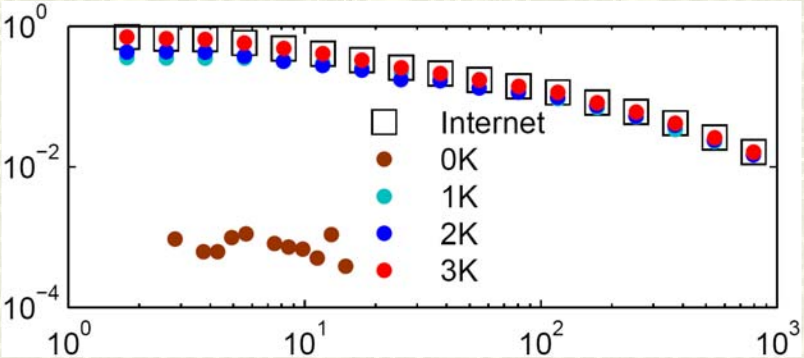




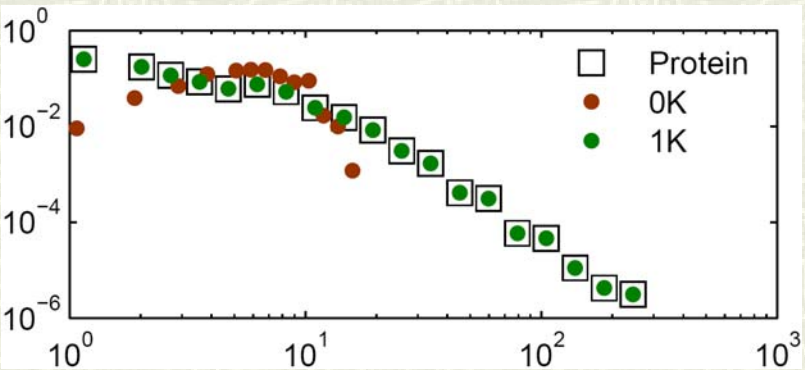
Internet

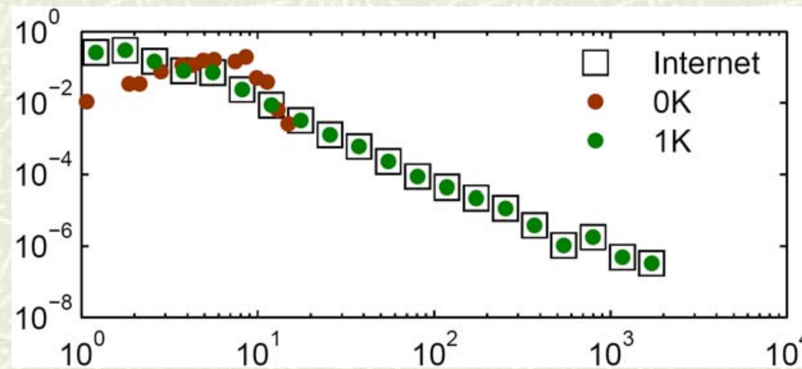
- **Heterogeneity:**
  distribution $P(k)$
  of node degrees k:
  - Real: $P(k) \sim k^{-\gamma}$
  - Random: $P(k) \sim \lambda^k e^{-\lambda}/k!$
- **Clustering:**
  average probability that
  node neighbors are connected:
  - Real: 0.46
  - Random: 6.8×10^{-4}

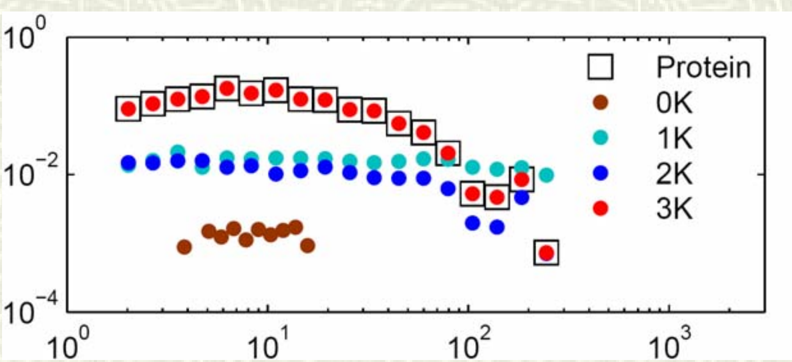


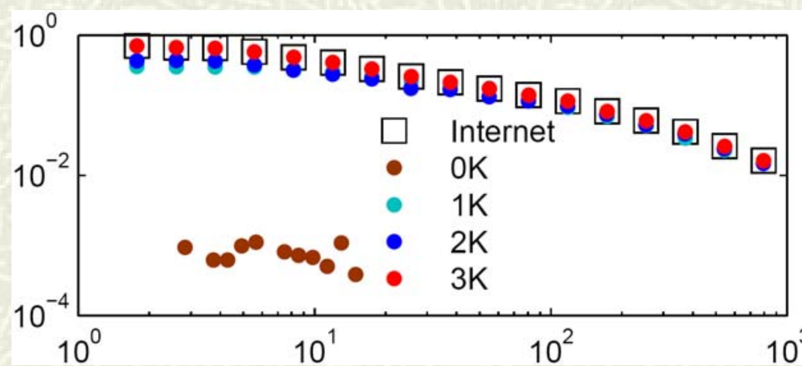


Internet vs. protein interaction









Strong heterogeneity and clustering as common features of complex networks

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	*2.4*	*0.09*
Metabolic	*2.0*	*0.67*
Gene regulation	*2.1*	*0.09*

Any other common features?

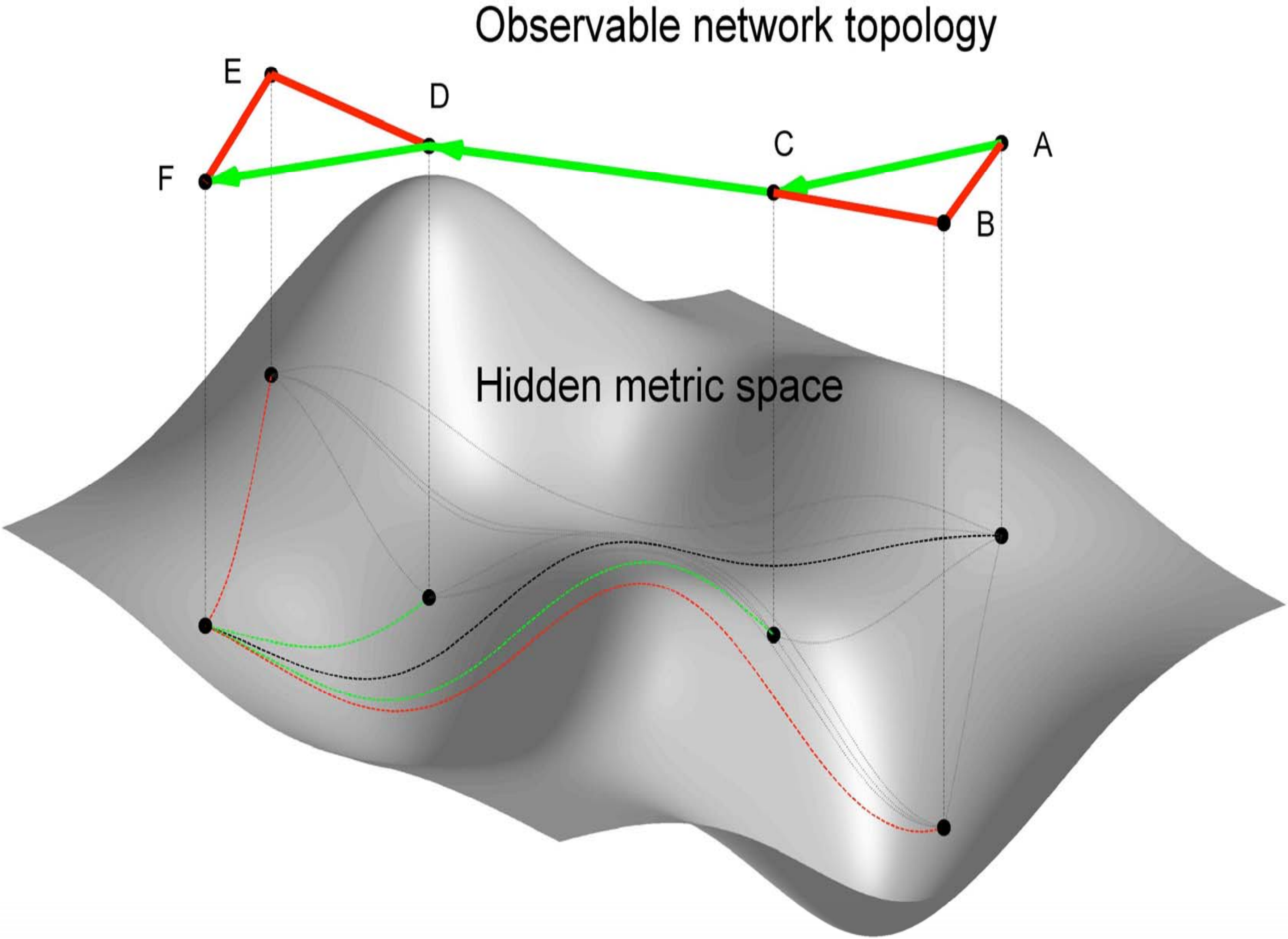
- Heterogeneity, clustering, some randomness, and their consequences:
  - Small-world effect (prevalence of short paths)
  - High path diversity (abundance of different paths between the same pair of nodes)
  - Robustness to random breakdowns
  - Fragility to targeted attacks
  - Modular/hierarchical organization
- pretty much exhaust all the commonalities—the networks are quite different and unique in all other respects
- Can we explain these two fundamental common features, heterogeneity and clustering?

Hidden metric space explanation

- All nodes 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

Mathematical perspective: Graphs embedded in manifolds

- All nodes exist in “two places at once”:
  - graph
  - hidden metric space, e.g., a Riemannian manifold
- There are two metric distances between each pair of nodes: observable and hidden:
  - hop length of the shortest path in the graph
  - distance in the hidden space



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

Hidden metric spaces explain the complex network **structure**

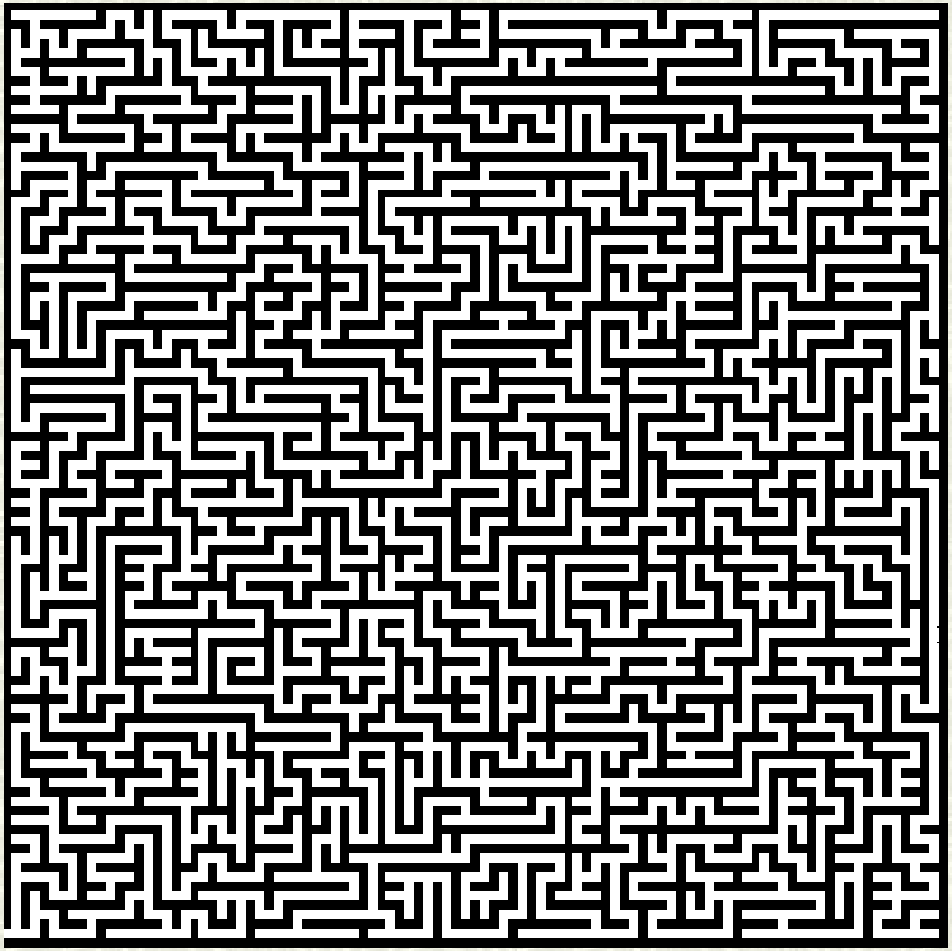
- **Clustering** is a consequence of the metric property of hidden spaces
- **Heterogeneity** is a consequence of their negative curvature (hyperbolic geometry)

Hidden metric spaces explain the complex network **function**

- **Transport** or **signaling** to specific destinations is a common function of many complex networks:
  - Transportation
  - Internet
  - Brain
  - Regulatory networks
- But in many networks, nodes do not know the topology of a network, its complex maze

Complex networks as complex mazes

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

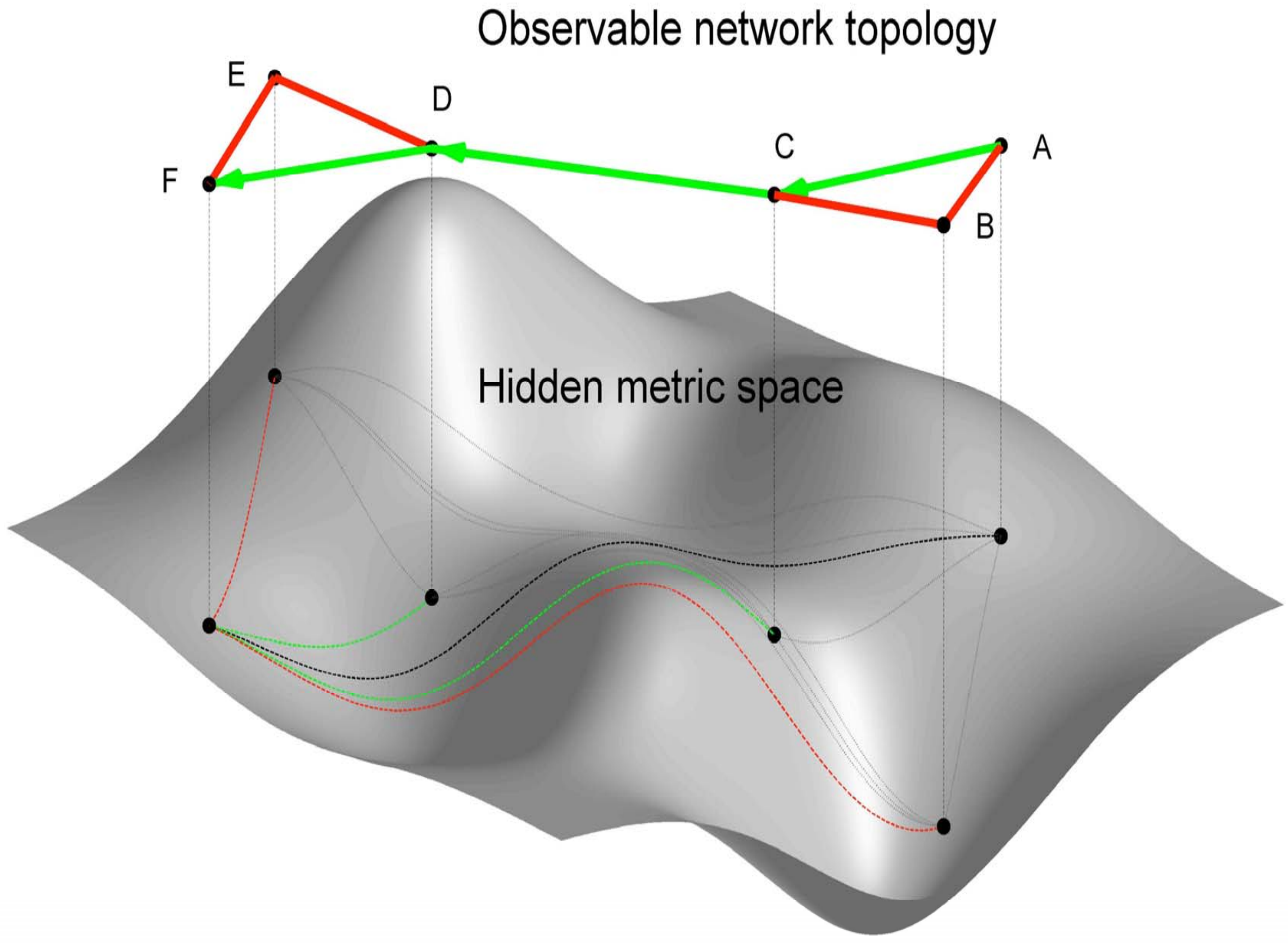


Milgram's experiments

- 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

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

Result #1: Hidden metric spaces do exist

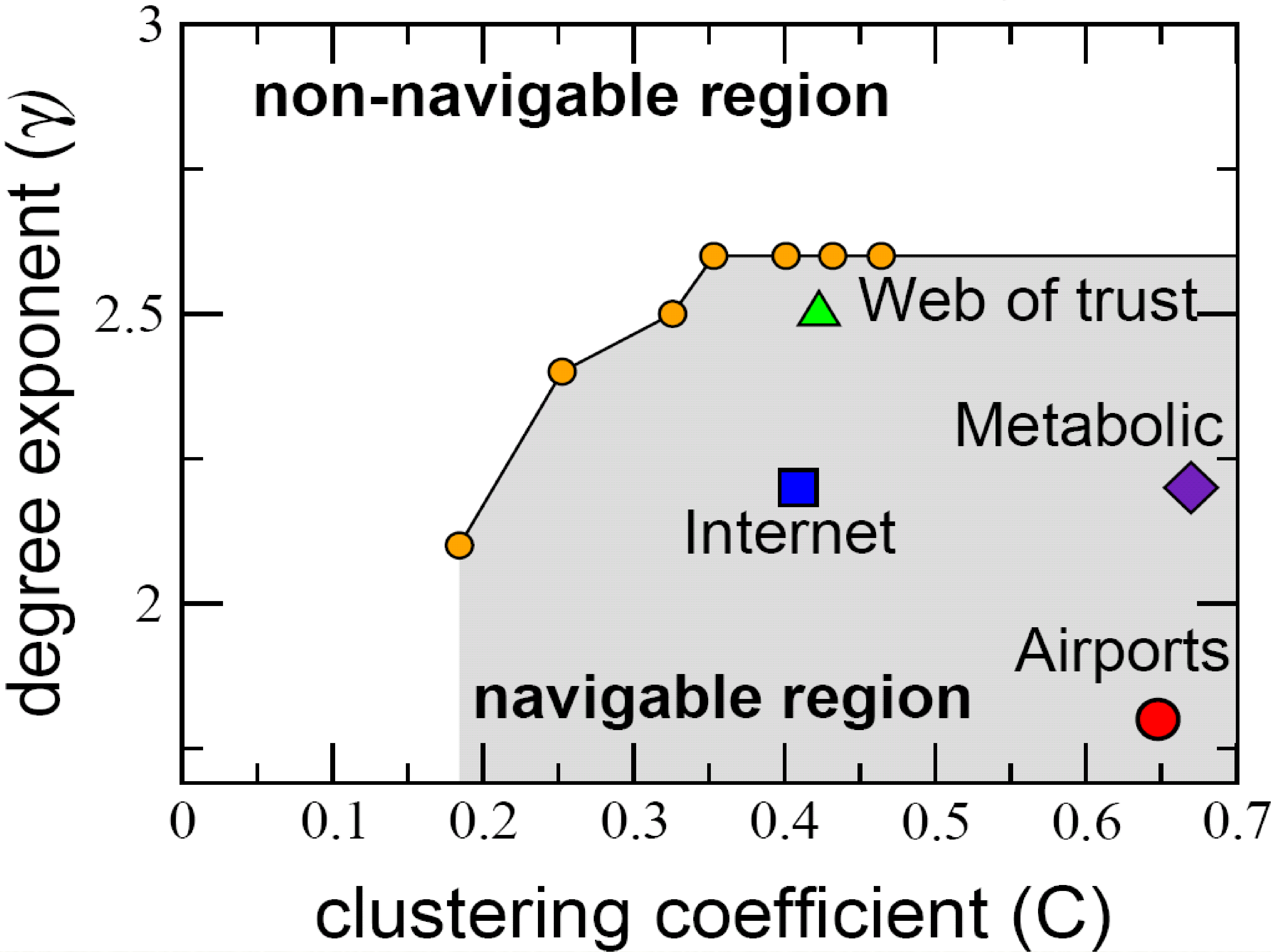
- Their existence appears as the only reasonable explanation of one peculiar property of the topology of real complex networks – self-similarity of clustering

Phys Rev Lett, v.100, 078701, 2008

Result #2:
Complex network topologies are navigable

- Specific values of degree distribution and clustering observed in real complex networks correspond to the highest efficiency of greedy routing
- Which implicitly suggests that complex networks do evolve to become navigable
- Because if they did not, they would not be able to function

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



non-navigable region
navigable region
Web of trust
Metabolic
Internet
Airports
degree exponent (γ)
clustering coefficient (C)
3
2.5
2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Result #3:
Successful greedy paths are shortest

- Regardless the structure of the hidden space, complex network topologies are such, that all successful greedy paths are asymptotically shortest
- But: how many greedy paths are successful does depend on the hidden space geometry

Phys Rev Lett, v.102, 058701, 2009

Result #4:

In hyperbolic geometry, all paths are successful

- Greedy routing in complex networks, including the real AS Internet, embedded in hyperbolic spaces, is always successful and always follows shortest paths
- Even if some links are removed, emulating topology dynamics, greedy routing finds remaining paths if they exist, without recomputation of node coordinates
- The reason is the exceptional congruency between complex network topology and hyperbolic geometry

Result #5:
Emergence of topology from geometry

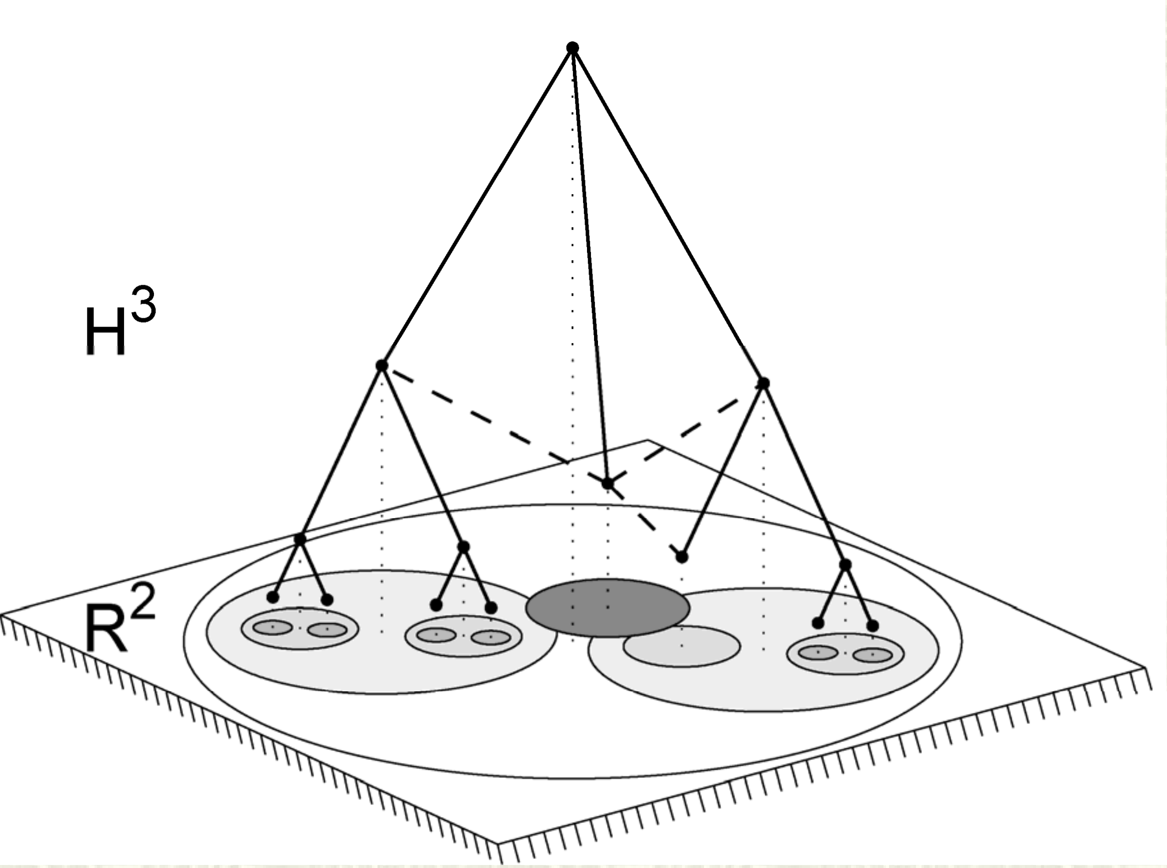
- The two main properties of complex network topology are direct consequences of the two main properties of hyperbolic geometry:
  - Scale-free degree distributions are a consequence of the exponential expansion of space in hyperbolic geometry
  - Strong clustering is a consequence of the fact that hyperbolic spaces are metric spaces

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

Motivation for hyperbolic spaces under complex networks

- Nodes in complex networks can often be hierarchically classified
  - Community structure (social and biological networks)
  - Customer-provider hierarchies (Internet)
  - Hierarchies of overlapping balls/sets (all networks)
- Hierarchies are (approximately) trees
- Trees embed almost isometrically in hyperbolic spaces

Mapping between balls $B(x,r)$ in $\mathbb{R}^d$ and points $\alpha = (x,r)$ in $\mathbb{H}^{d+1}$



- If $|\alpha - \alpha'| \leq C$, then there exist $k(C)$ s.t. $k^{-1} \leq r/r' \leq k$ and $|x - x'| \leq k\, r$
- If $|x - x'| \leq k\, r$ and $k^{-1} \leq r/r' \leq k$, then there exist $C(k)$ s.t. $|\alpha - \alpha'| \leq C$

Metric structure of hyperbolic spaces

- The volume of balls and surface of spheres grow with their radius r as

$$e^{\alpha r}$$

where $\alpha = (-K)^{1/2}(d-1)$, K is the curvature and d is the dimension of the hyperbolic space

- The numbers of nodes in a tree within or at r hops from the root grow as

$$b^{r}$$

where b is the tree branching factor

- The metric structures of hyperbolic spaces and trees are essentially the same ($\alpha = \ln b$)

Hidden space in our model: hyperbolic disc

- Hyperbolic disc of radius R, where $N = c\, e^{R/2}$, N is the number of nodes in the network and c controls its average degree
- Curvature $K = -1$

Node distribution in the disc: uniform

- Uniform angular density

$$\rho_\theta(\theta) = 1/(2\pi)$$

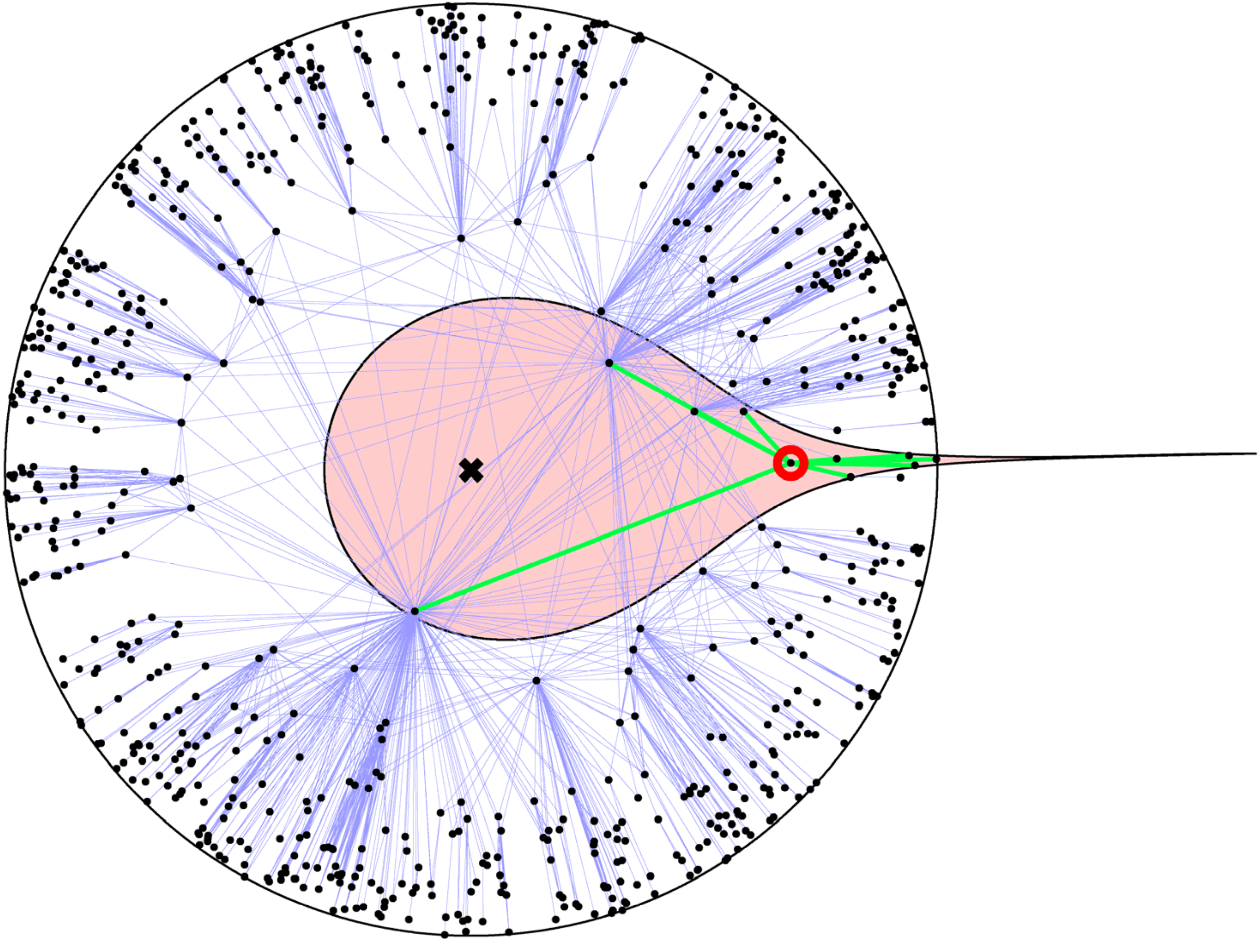
- Exponential radial density

$$\rho(r) = \sinh r \,/\, (\cosh R - 1) \approx e^{r-R}$$

Connection probability: step function

- Connected each pair nodes located at (r,θ) and (r',θ'), if the hyperbolic distance x between them is less than or equal to R, where

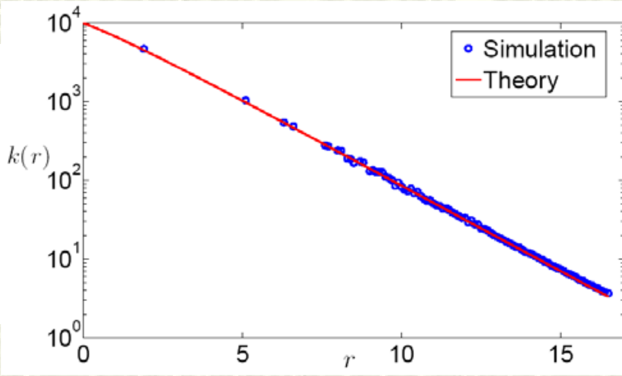
$$\cosh x = \cosh r \cosh r' - \sinh r \sinh r' \cos \Delta\theta$$



Average node degree at distance r from the disc center

- Terse but exact expression

$$k(r) = \delta \left\{ (\cosh R - 2)\pi + 2 \left(\cosh\chi \arccos \frac{\cosh r \cosh\chi - \cosh R}{\sinh r \sinh \chi} \right. \right.$$
$$\left. + \cosh R \arctan \frac{\cosh\chi - \cosh R \cosh r}{\sqrt{(\cosh r - \cosh(R-\chi))(\cosh(R+\chi) - \cosh r)}} \right) \qquad (42)$$
$$\left. - \arctan \frac{8(\cosh r - \cosh R \cosh\chi)\sqrt{(\cosh r - \cosh(R-\chi))(\cosh(R+\chi) - \cosh r)}}{16(\cosh r - \cosh R\cosh\chi)\cosh r - 8\cosh^2 r + 4(\cosh^2 R + \cosh^2 R \cosh^2\chi + \cosh^2\chi) - 1} \right\}.$$



- Simple approximation:

$$k(r) \approx (4c/\pi)\, e^{(R-r)/2}$$

Degree distribution

- Since $\rho(r) \sim e^{r}$ and $k(r) \sim e^{-r/2}$,

$$P(k) = \rho[r(k)]\,|r'(k)| \sim k^{-3}$$

- Power-law degree distribution naturally emerges as a simple consequence of the exponential expansion of hyperbolic space

Generalizing the model

- Curvature

$$K = -\zeta^2$$

- and node density:

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

- lead to the average degree at distance r

$k(r) \sim e^{-\zeta r/2}$ if $\alpha/\zeta \geq 1/2$; or

$k(r) \sim e^{-\alpha r/2}$ otherwise

Generalized degree distribution

- Degree distribution

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

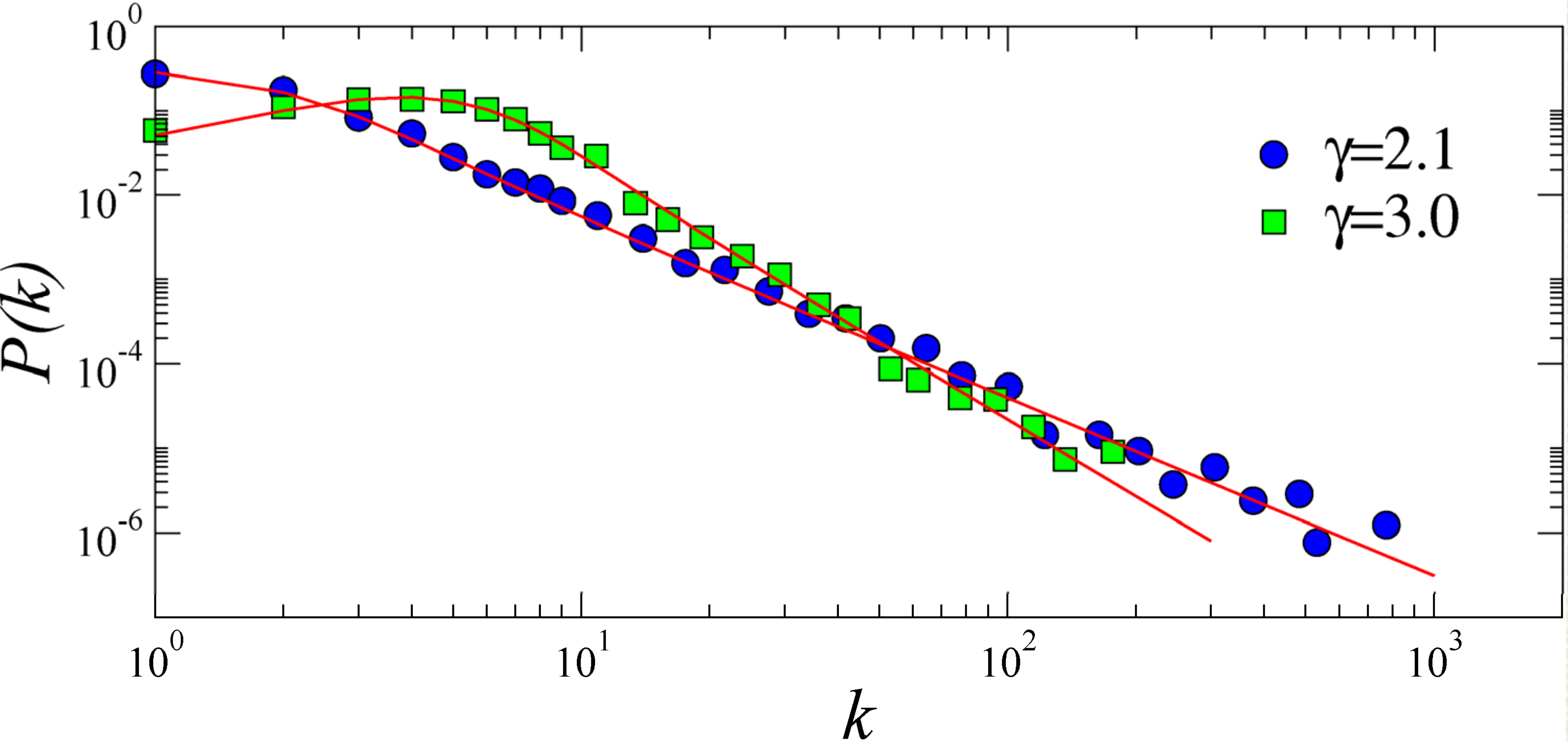
- where

$$\gamma = 2\,\alpha/\zeta + 1 \qquad \text{if } \alpha/\zeta \geq 1/2$$

$$\gamma = 2 \qquad \text{otherwise}$$

- Uniform node density ($\alpha = \zeta$) yields $\gamma = 3$ as in the standard preferential attachment

Node degree distribution: theory vs. simulations



The other way around

- We have shown that scale-free topology naturally emerges from underlying hyperbolic geometry
- Now we will show that hyperbolic geometry naturally emerges from scale-free topology

The $\mathbb{S}^1$ model

- The hidden metric space is a circle of radius
$$N/(2\pi)$$
- The node density is uniform ($=1$) on the circle
- Nodes are assigned an additional hidden variable κ, the node expected degree, drawn from
$$\rho_\kappa(\kappa) = (\gamma - 1)\kappa^{-\gamma}$$
- To guarantee that $k(\kappa) = \kappa$, the connection probability must be an integrable function of
$$\chi \sim N\Delta\theta / (\kappa\kappa')$$
- where $\Delta\theta$ is the angle between nodes, and κ, κ' are their expected degrees

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

- Formal change of variables
  $\kappa = e^{\zeta(R-r)/2}$ (cf. $k(r) \sim e^{-\zeta r/2}$ in $\mathbb{H}^2$)
- where
  $\zeta/2 = \alpha/(\gamma - 1)$ (cf. $\gamma = 2\,\alpha/\zeta + 1$ in $\mathbb{H}^2$)
- yields density
  $\rho(r) = \alpha\, e^{\alpha(r-R)}$ (as in $\mathbb{H}^2$)
- and the argument of the connection probability

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

- where

$$x = r + r' + (2/\zeta)\ln(\Delta\theta/2)$$

  is approximately the hyperbolic distance between nodes on the disc

Fermi connection probability

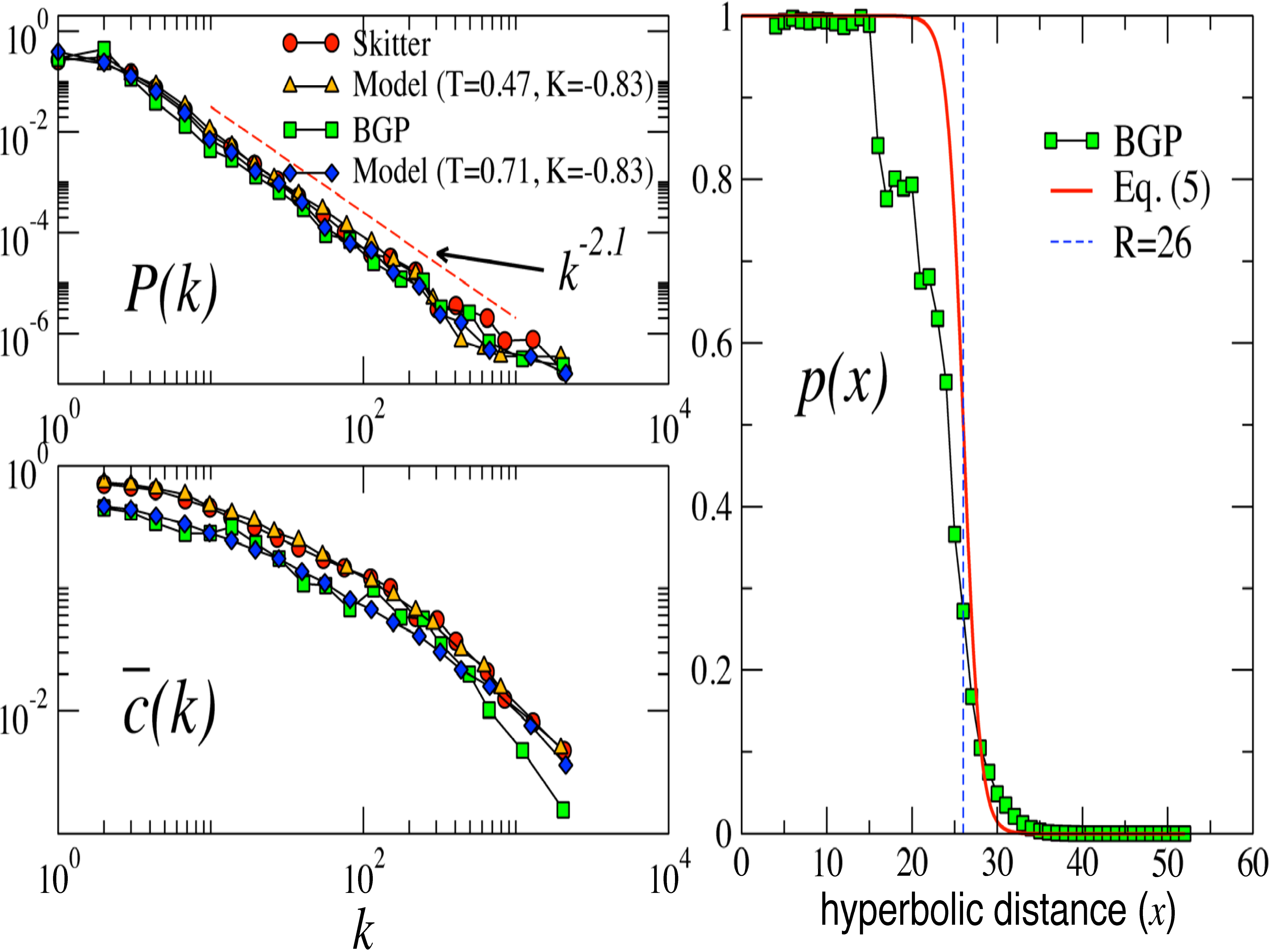
- Connection probability can be any function of χ
- Selecting it to be $1 / (1 + \chi^{1/T})$, $T \geq 0$, i.e.,
  $$p(x) = 1 / (1 + e^{\zeta(x-R)/(2T)})$$
- allows to fully control clustering between its maximum at $T = 0$ and zero at $T = 1$
- At $T = 0$, $p(x) = \Theta(R-x)$, i.e., the step function
- At $T = 1$ the system undergoes a phase transition, and clustering remains zero for all $T \geq 1$
- At $T = \infty$ the model produces classical random graphs, as nodes are connected with the same probability independent of hidden distances

Physical interpretation

- Hyperbolic distances x are energies of corresponding links-fermions
- Hyperbolic disc radius R is the chemical potential
- Clustering parameter T is the system temperature
- Two times the inverse square root of curvature $2/\zeta$ is the Boltzmann constant

Hyperbolic embedding of real complex networks

- Measure the average degree, degree distribution exponent, and clustering in a real network
- Map those to the three parameters in the model (c, α/ζ, T)
- Use maximum-likelihood techniques (e.g., the Metropolis-Hastings algorithm) to find the hyperbolic node coordinates



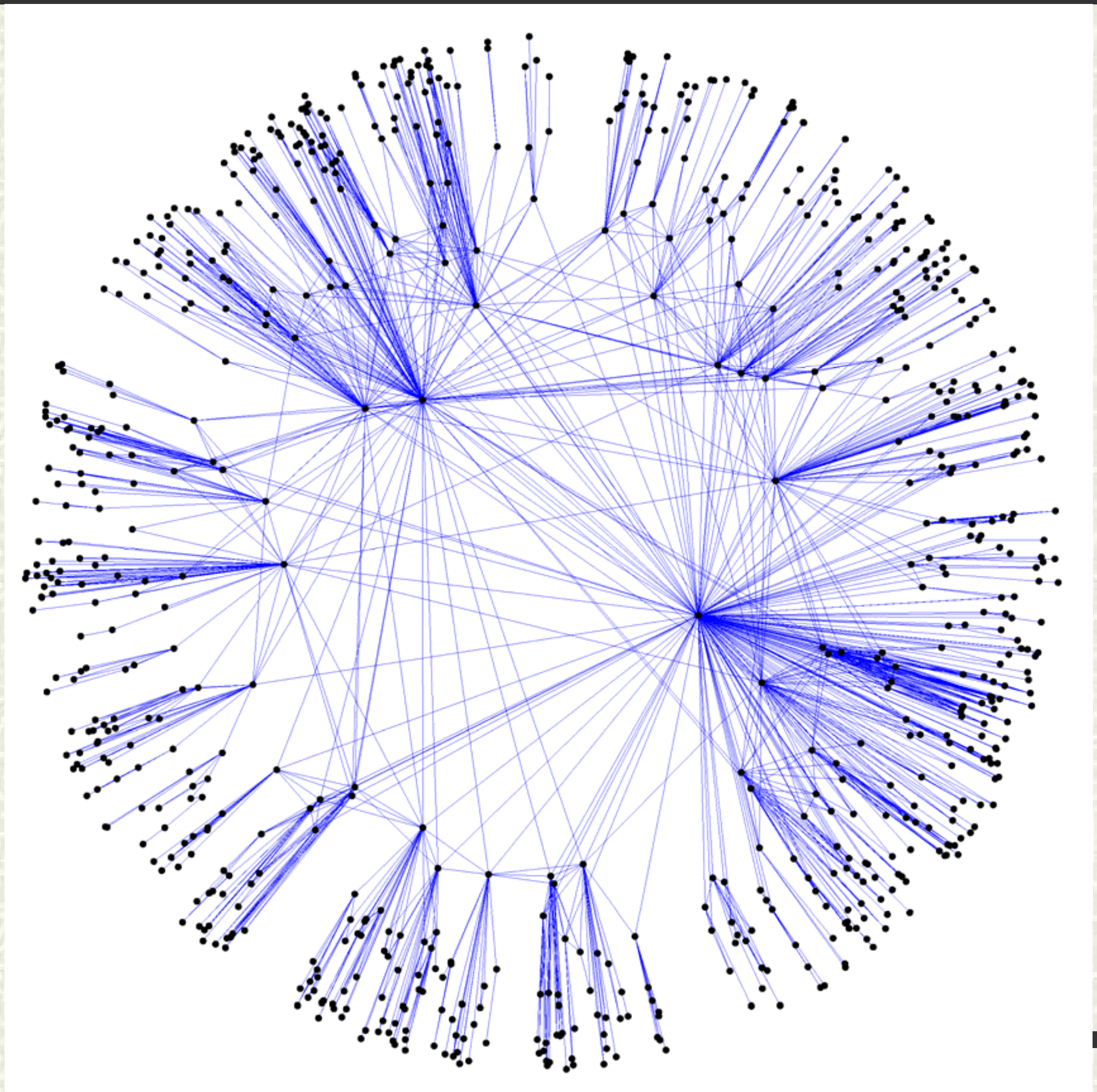
Skitter
Model (T=0.47, K=-0.83)
BGP
Model (T=0.71, K=-0.83)
k^-2.1
P(k)
c̄(k)
k
p(x)
BGP
Eq. (5)
R=26
hyperbolic distance (x)

Navigation in $\mathbb{S}^1$ and $\mathbb{H}^2$

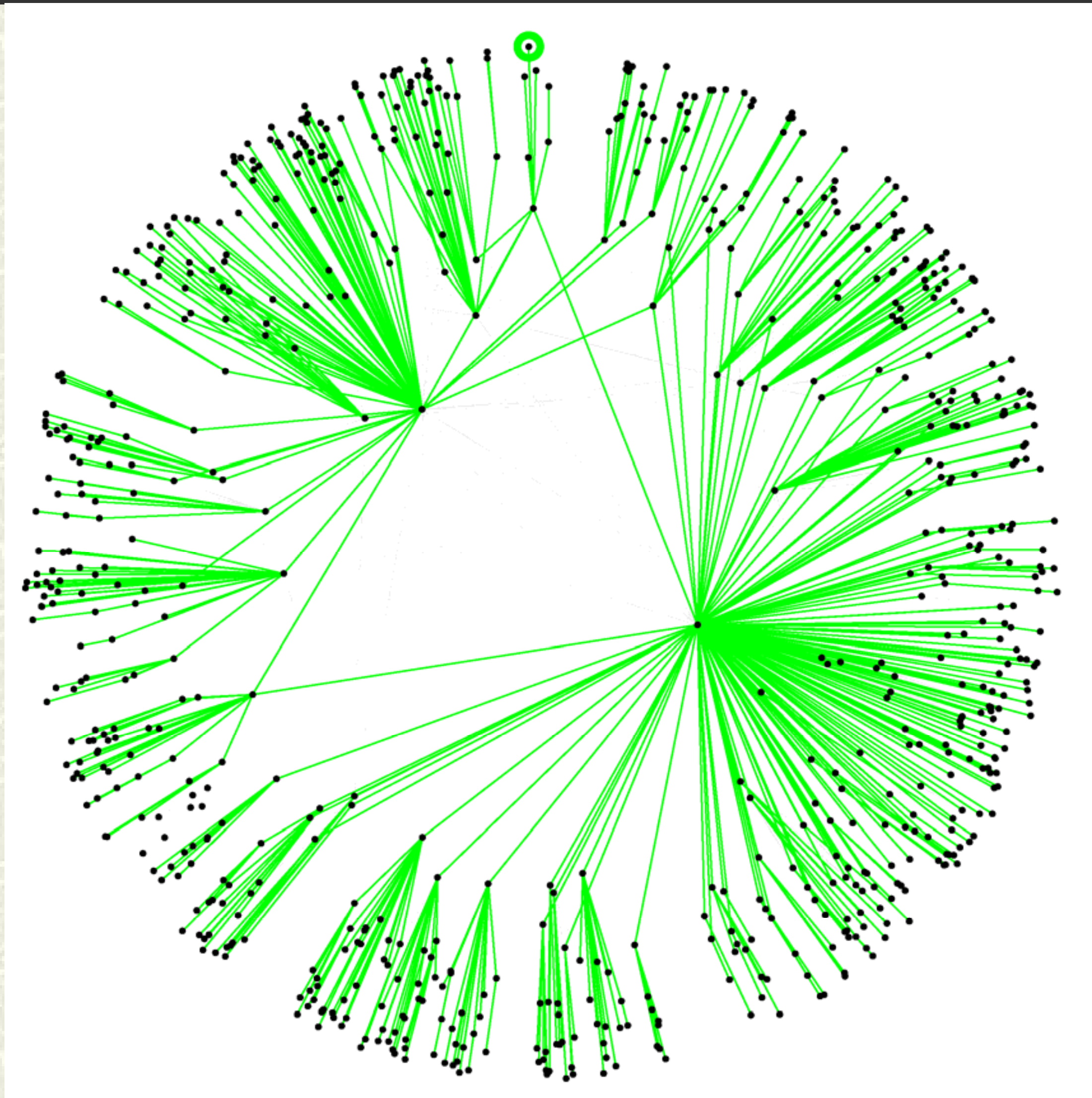
- The $\mathbb{S}^1$ and $\mathbb{H}^2$ models are essentially equivalent in terms of produced network topologies
- But what distances, $\mathbb{S}^1$ or $\mathbb{H}^2$, should we use to navigate the network?
- Successful greedy paths are asymptotically shortest
- But what about success ratio?

	Embedded Internet	Synthetic networks
$\mathbb{S}^1$	76%	$\leq$ 70%
$\mathbb{H}^2$	95%	$\leq$ 100%

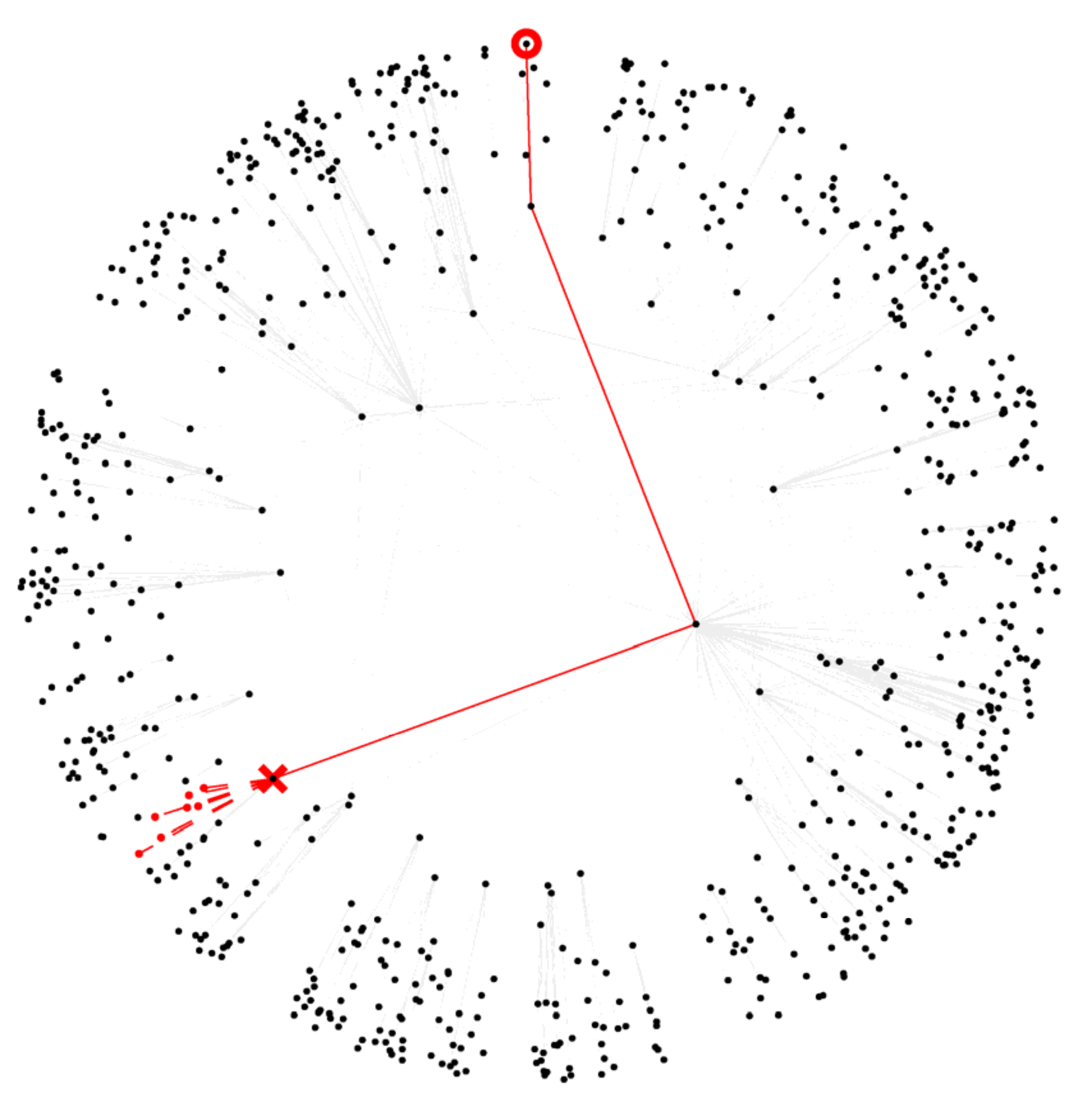
Visualization of a modeled network



Successful greedy paths



Unsuccessful greedy paths

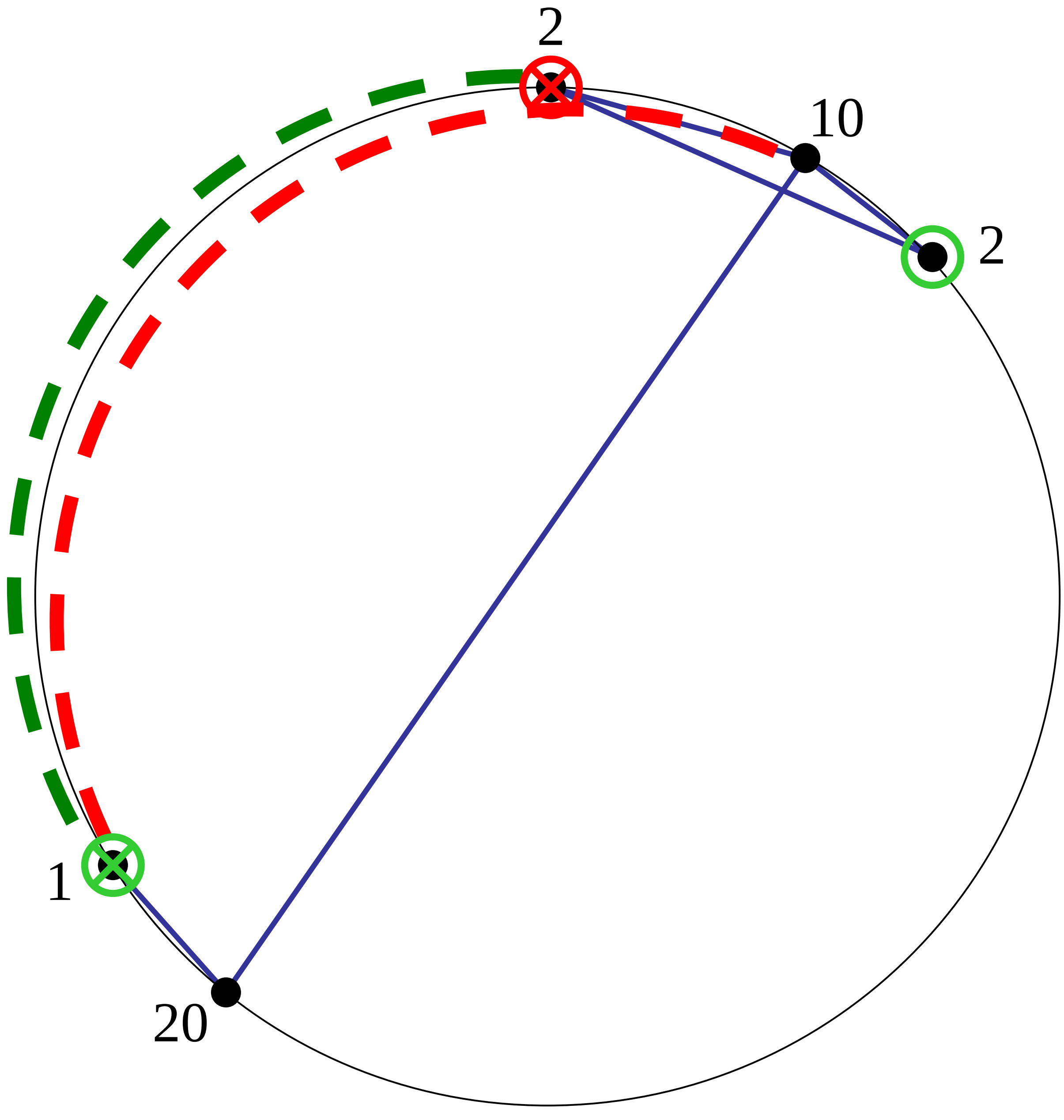


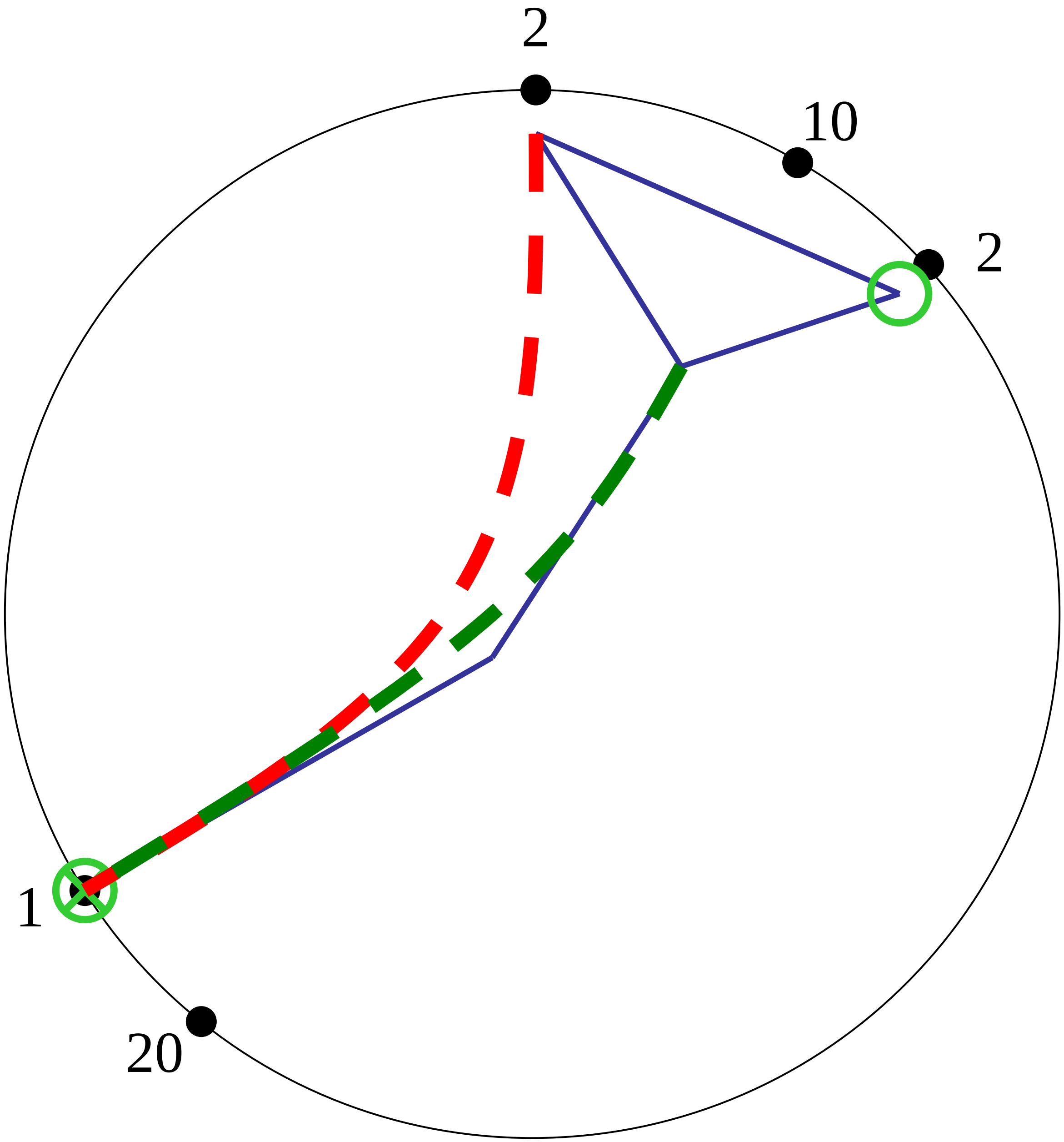
Robustness of greedy routing in $\mathbb{H}^2$ w.r.t. topology perturbations

- As network topology changes, the greedy routing efficiency deteriorates very slowly
- For example, for synthetic networks with $\gamma \leq 2.5$, removal of up to *10%* of the links from the topology degrades the percentage of successful path by less than *1%*

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

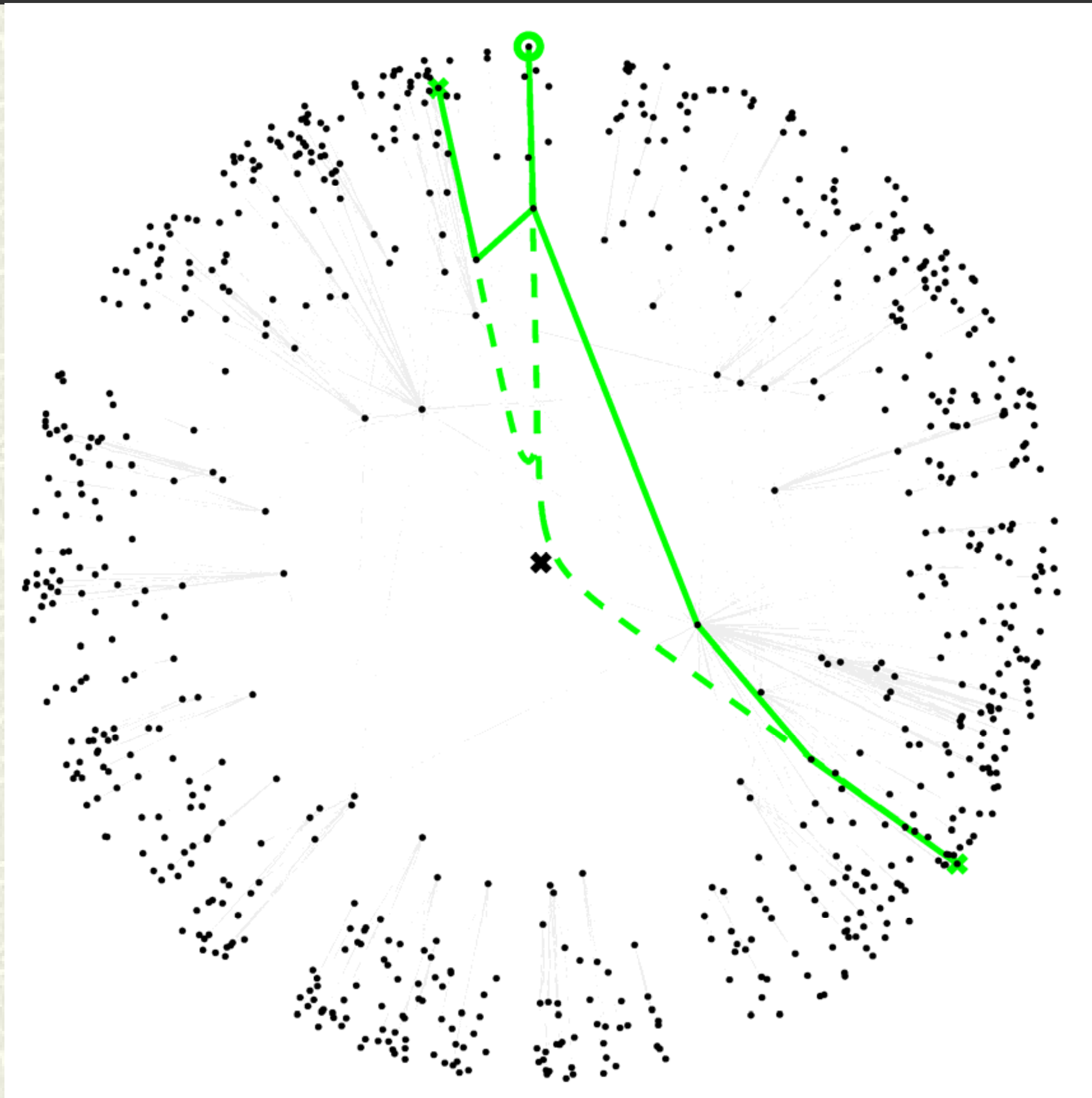
- Because nodes in the $\mathbb{S}^1$ model are not connected with probability which depends solely on the $\mathbb{S}^1$ distances $N\Delta\theta$
- Those distances are rescaled by node degrees to $\chi \sim N\Delta\theta/(\kappa\kappa')$, and we have shown that these rescaled distances are essentially hyperbolic if node degrees are power-law distributed
- Intuitively, navigation is better if it uses more congruent distances, i.e., those with which the network is built

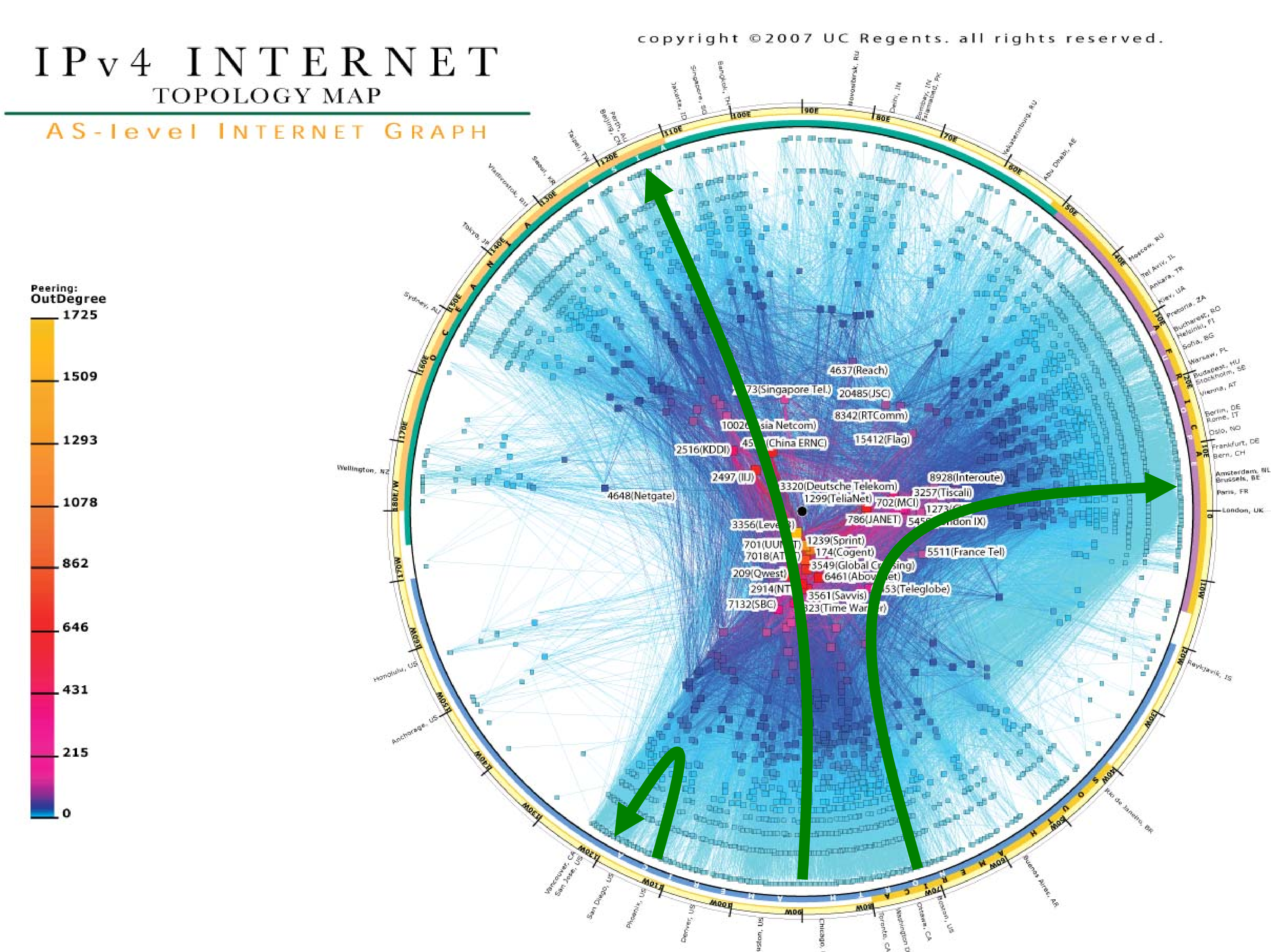

2
10
2
1
20



2
10
2
1
20

Shortest paths in scale-free graphs and hyperbolic spaces




IPv4 INTERNET
TOPOLOGY MAP
AS-level INTERNET GRAPH
copyright ©2007 UC Regents. all rights reserved.
Peering:
OutDegree
1725
1509
1293
1078
862
646
431
215
0
4637(Reach)
20485(JSC)
8342(RTComm)
15412(Flag)
2516(KDDI)
2497 (IIJ)
3320(Deutsche Telekom)
1299(TeliaNet)
702(MCI)
8928(Interoute)
3257(Tiscali)
786(JANET)
4648(Netgate)
1239(Sprint)
174(Cogent)
5511(France Tel)
209(Qwest)
3561(Savvis)
7132(SBC)
London, UK
Amsterdam, NL
Brussels, BE
Paris, FR
Frankfurt, DE
Bern, CH
Oslo, NO
Berlin, DE
Rome, IT
Vienna, AT
Moscow, RU
Tel Aviv, IL
Ankara, TR
Kiev, UA
Pretoria, ZA
Sofia, BG
Warsaw, PL
Abu Dhabi, AE
Yekaterinburg, RU
Novosibirsk, RU
Delhi, IN
Bombay, IN
Bangkok, TH
Singapore, SG
Jakarta, ID
Perth, AU
Beijing, CN
Taipei, TW
Seoul, KR
Vladivostok, RU
Tokyo, JP
Sydney, AU
Wellington, NZ
Honolulu, US
Anchorage, US
Vancouver, CA
San Jose, US
San Diego, US
Phoenix, US
Denver, US
Chicago, US
Toronto, CA
Ottawa, CA
Boston, US
Buenos Aires, AR
Rio de Janeiro, BR
Reykjavik, IS
ASIA
OCEANIA
EUROPE
NORTH AMERICA
SOUTH AMERICA

In summary

- Hidden hyperbolic metric spaces explain, simultaneously, the two main topological characteristics of complex networks
  - scale-free degree distributions (by negative curvature)
  - strong clustering (by metric property)
- Complex network topologies are congruent with hidden hyperbolic geometries
  - Greedy paths follow shortest paths that approximately follow shortest hidden paths, i.e., geodesics in the hyperbolic space
    - Both topology and geometry are *tree-like*
- This congruency is robust w.r.t. topology dynamics
  - There are many link/node-disjoint shortest paths between the same source and destination that satisfy the above property
    - *Strong clustering* (many by-passes) boosts up the path diversity
  - If some of shortest paths are damaged by link failures, many others remain available, and greedy routing still finds them

Conclusion

- To efficiently route without topology knowledge, the topology should be both hierarchical (tree-like) and have high path diversity (not like a tree)
- Complex networks do borrow the best out of these two seemingly mutually-exclusive worlds
- Hidden hyperbolic geometry naturally explains how this balance is achieved

Applications

- Greedy routing mechanism in these settings may offer virtually infinitely scalable information dissemination (routing) strategies for future communication networks
  - Zero communication costs (no routing updates!)
  - Constant routing table sizes (coordinates in the space)
  - No stretch (all paths are shortest, stretch=*1*)
- Interdisciplinary applications
  - systems biology: brain and regulatory networks, cancer research, phylogenetic trees, protein folding, etc.
  - data mining and recommender systems
  - cognitive science