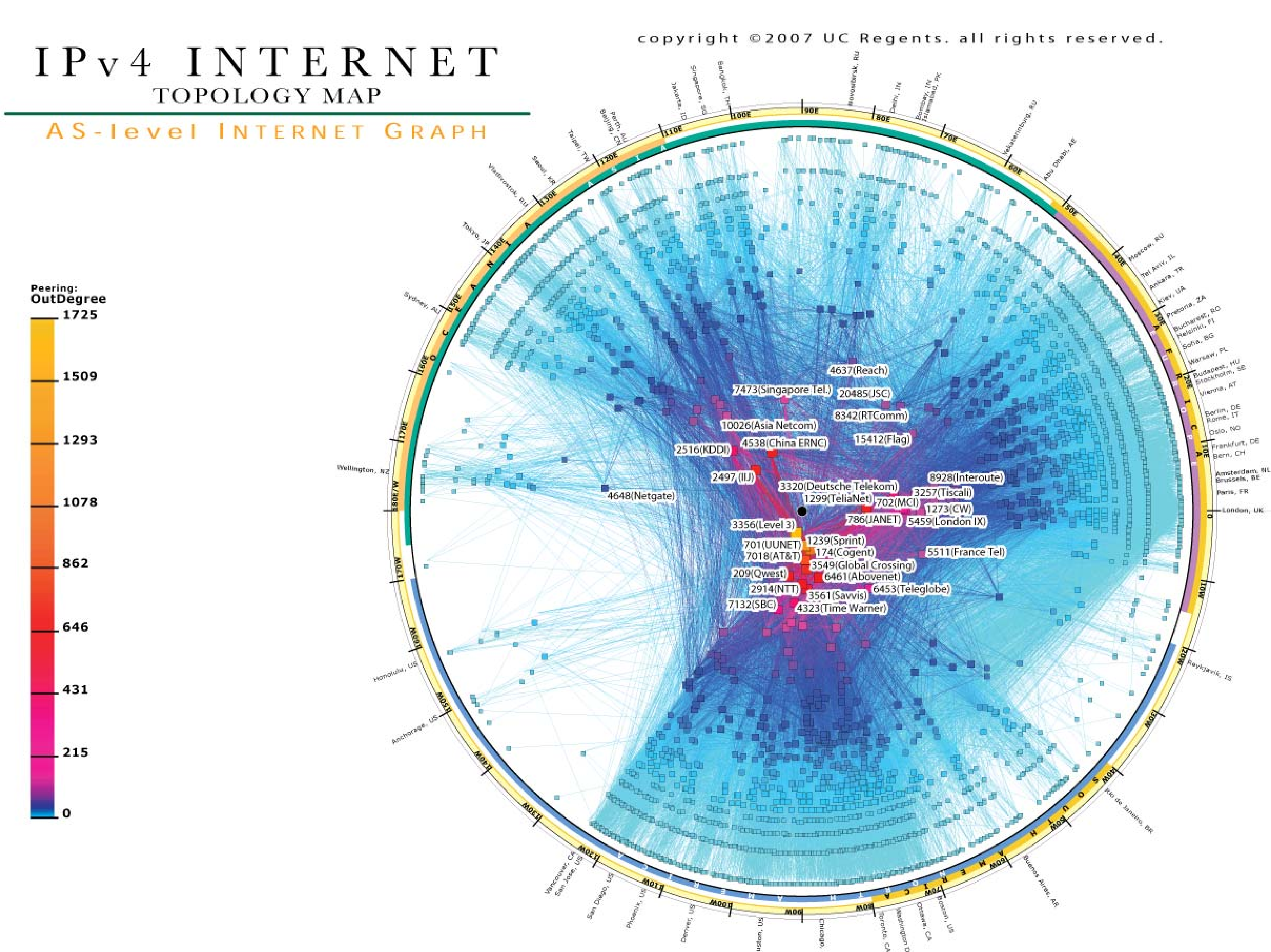
# Hyperbolic geometry and scale-free topology of complex networks

**Dmitri Krioukov**
CAIDA/UCSD
**F. Papadopoulos, M. Boguñá, A. Vahdat**

# Internet

- Microscopic view ("design")
  - IP/TCP, routing protocols
  - Routers
  - Per-ISP router-level topologies
- Macroscopic view ("non-design")
  - Global AS-level topology is a cumulative result of local, decentralized, and rather complex interactions between AS pairs
  - Surprisingly, in 1999, it was found to look completely differently than engineers had thought
    - It is not a grid, tree, or classical random graph
    - It shares all the main features of topologies of other complex networks
      - scale-free (power-law) node degree distributions ($P(k) \sim k^{-\gamma}$, $\gamma \in [2,3]$)
      - strong clustering (large numbers of 3-cycles)
- The big problem is that "design" has now to deal with "non-design"
  - Routing protocols have to find and promptly update paths to all destinations in the Internet



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)
7473(Singapore Tel.)
20485(JSC)
8342(RTComm)
10026(Asia Netcom)
4538(China ERNC)
15412(Flag)
2516(KDDI)
2497(IIJ)
3320(Deutsche Telekom)
8928(Interoute)
3257(Tiscali)
1299(TeliaNet)
702(MCI)
1273(CW)
4648(Netgate)
786(JANET)
5459(London IX)
3356(Level 3)
1239(Sprint)
701(UUNET)
174(Cogent)
7018(AT&T)
5511(France Tel)
3549(Global Crossing)
209(Qwest)
6461(Abovenet)
2914(NTT)
6453(Teleglobe)
3561(Savvis)
7132(SBC)
4323(Time Warner)
ASIA
OCEANIA
EUROPE
AFRICA
SOUTH AMERICA
NORTH AMERICA
London, UK
Paris, FR
Amsterdam, NL
Brussels, BE
Frankfurt, DE
Bern, CH
Oslo, NO
Berlin, DE
Rome, IT
Vienna, AT
Budapest, HU
Stockholm, SE
Warsaw, PL
Sofia, BG
Bucharest, RO
Helsinki, FI
Pretoria, ZA
Kiev, UA
Ankara, TR
Tel Aviv, IL
Moscow, RU
Abu Dhabi, AE
Yekaterinburg, RU
Bombay, IN
Islamabad, PK
Delhi, IN
Novosibirsk, RU
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
Washington DC
Ottawa, CA
Boston, US
Buenos Aires, AR
Rio de Janeiro, BR
Reykjavik, IS
0
10E
20E
30E
40E
50E
60E
70E
80E
90E
100E
110E
120E
130E
140E
150E
160E
170E
180E/W
170W
160W
150W
140W
130W
120W
110W
100W
90W
80W
70W
60W
50W
40W
30W
20W
10W

# Routing practice

- Global (DFZ) routing tables
  - 300,000 prefix entries (and growing)
  - 30,000 ASs (and growing)
- Routing overhead/convergence
  - BGP updates
    - 2 per second on average
    - 7000 per second peak rate
  - Convergence after a single event can take up to tens of minutes
- Problems with design?
  - Yes and no

# Routing theory

- There can be no routing algorithm with the number of messages per topology change scaling better than linearly with the network size in the worst case
- Small-world networks are this worst case

- *Is there any workaround?*
- *If topology updates/convergence is so expensive, then may be we can route without them, i.e., without global knowledge of the network topology?*
- *Let us look at the existing systems*

# Navigability of complex networks

- In many (if not all) existing complex networks, nodes communicate without any global knowledge of network topologies; examples:
  - Social networks
  - Neural networks
  - Cell regulatory networks
- How is this possible???

# Hidden metric space explanation

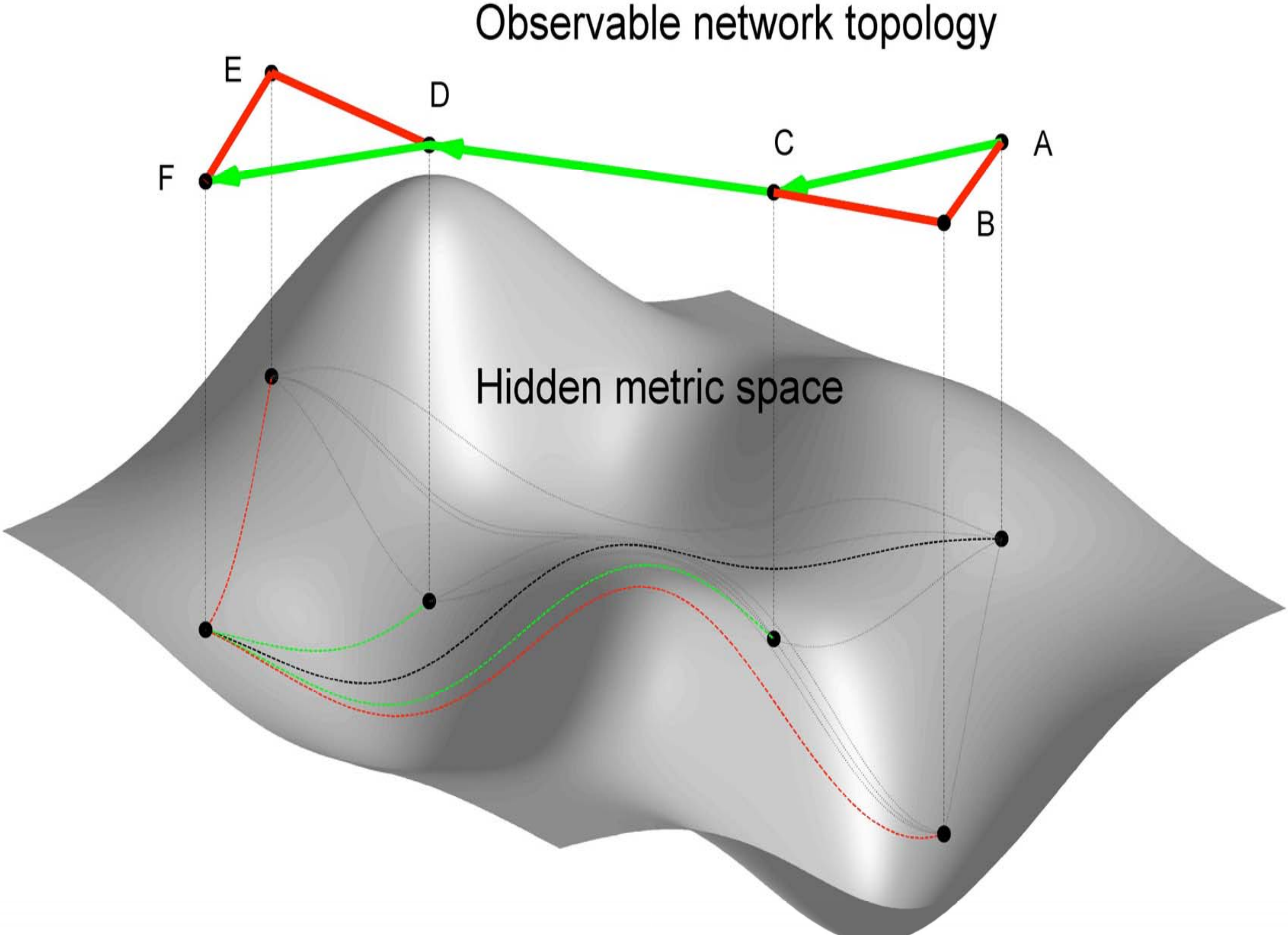
- All nodes exist in a metric space
- Distances in this space abstract node similarities
  - More similar nodes are closer in the space
- Similarities are defined by network-specific node attributes and are often based on/related to the community structure
  - The more communities in common, the more similar the two nodes
- Network consists of links that exist with probability that decreases with the hidden distance
  - More similar/close nodes are more likely to be connected

# Hidden schizophrenia

- All nodes exist in “two places at once”:
  - network
  - hidden metric space
- There are two metric distances between each pair of nodes: observable and hidden:
  - hop length of the shortest path in the network
  - distance in the hidden space

# Greedy routing (Kleinberg)

- 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

# Questions raised by the approach

- What is the hidden space?
- What are the node positions in it?
- What is the connection probability?
- How efficient is the greedy routing process?
  - How often greedy-routing paths get stuck at nodes that do not have any neighbors closer to the destination than themselves
  - How closely greedy-routing paths follow the shortest paths in the network
- What topologies are navigable, i.e., congruent w.r.t. greedy routing, i.e., make it efficient?

# Hidden spaces are metric spaces

- Using the simplest metric space (a circle), we show that
  - the triangle inequality in hidden spaces
    - transitivity of being similar/close

  explains

  - strong clustering in real networks
    - transitivity of being connected
- It also explains their self-similarity

# Navigability mechanisms

- More navigable networks are networks with
  - more heterogeneous node degree distributions
    - more hubs
  - stronger clustering
    - stronger influence of hidden distances on links
    - stronger congruency between hidden geometries and observed topologies
    - stronger congruency between greedy and shortest paths
- Greedy routing paths follow navigable path pattern

# Hidden geometries

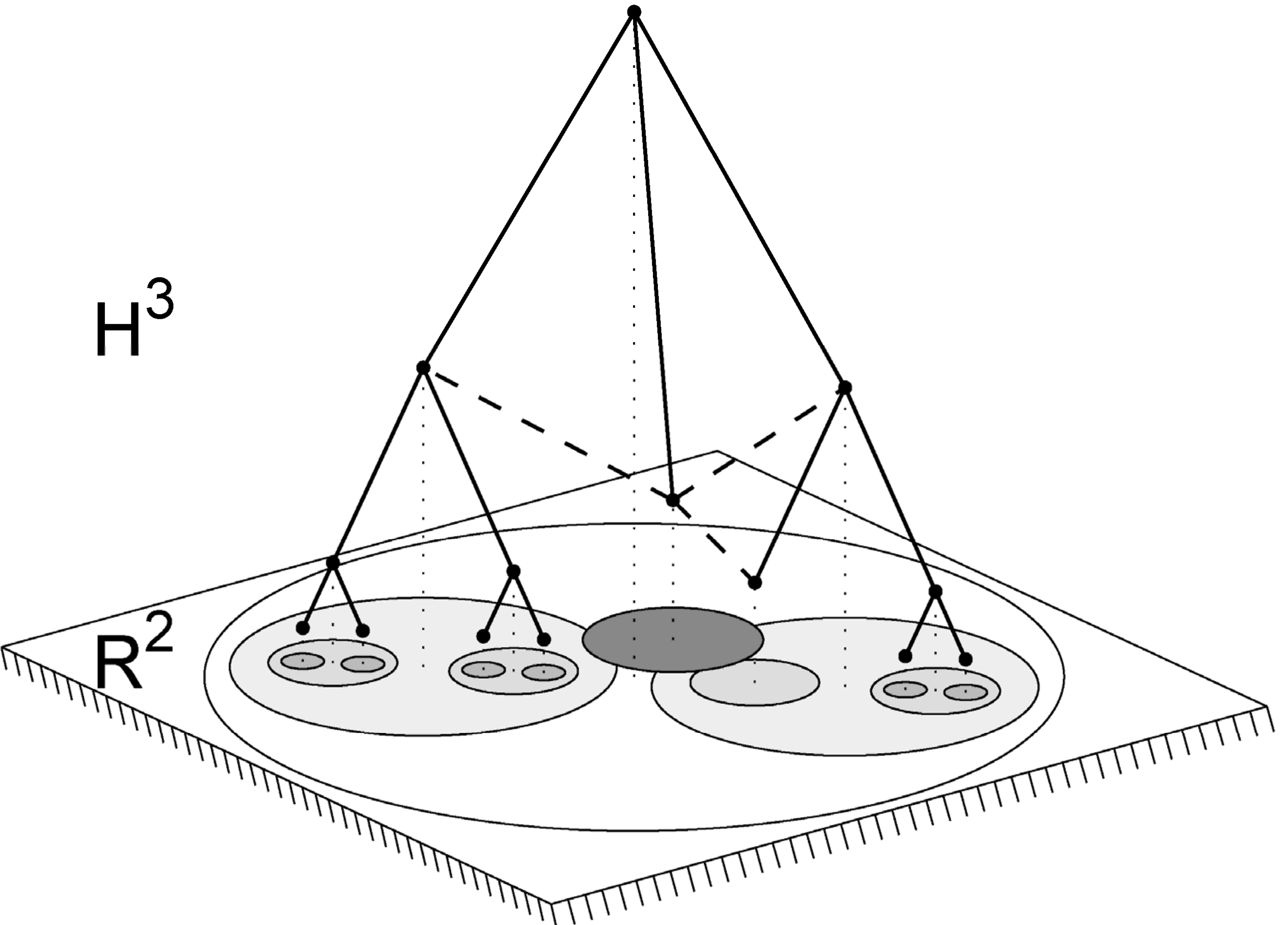
- What hidden geometries are maximally congruent with the navigability mechanisms of the observed complex network topologies?

# Hidden metric spaces are hyperbolic

- Network nodes can often be classified hierarchically, at least approximately
- Hierarchies are tree-like structures
- Hyperbolic geometry is the geometry of tree-like structures
  - Formally: trees embed almost isometrically in hyperbolic spaces, not in Euclidean ones

# Hierarchies in real networks

- Community structure in Wikipedia:
  - editors are nodes; articles are communities
  - the more articles you edited, the closer to the top you are
  - similarity between two editors is defined by how many articles they both edited
- Hierarchies of overlapping sets
  - map nodes to sets of communities they are members of
  - define the similarity between two nodes as a measure of the overlap of their community sets


$H^3$
$R^2$

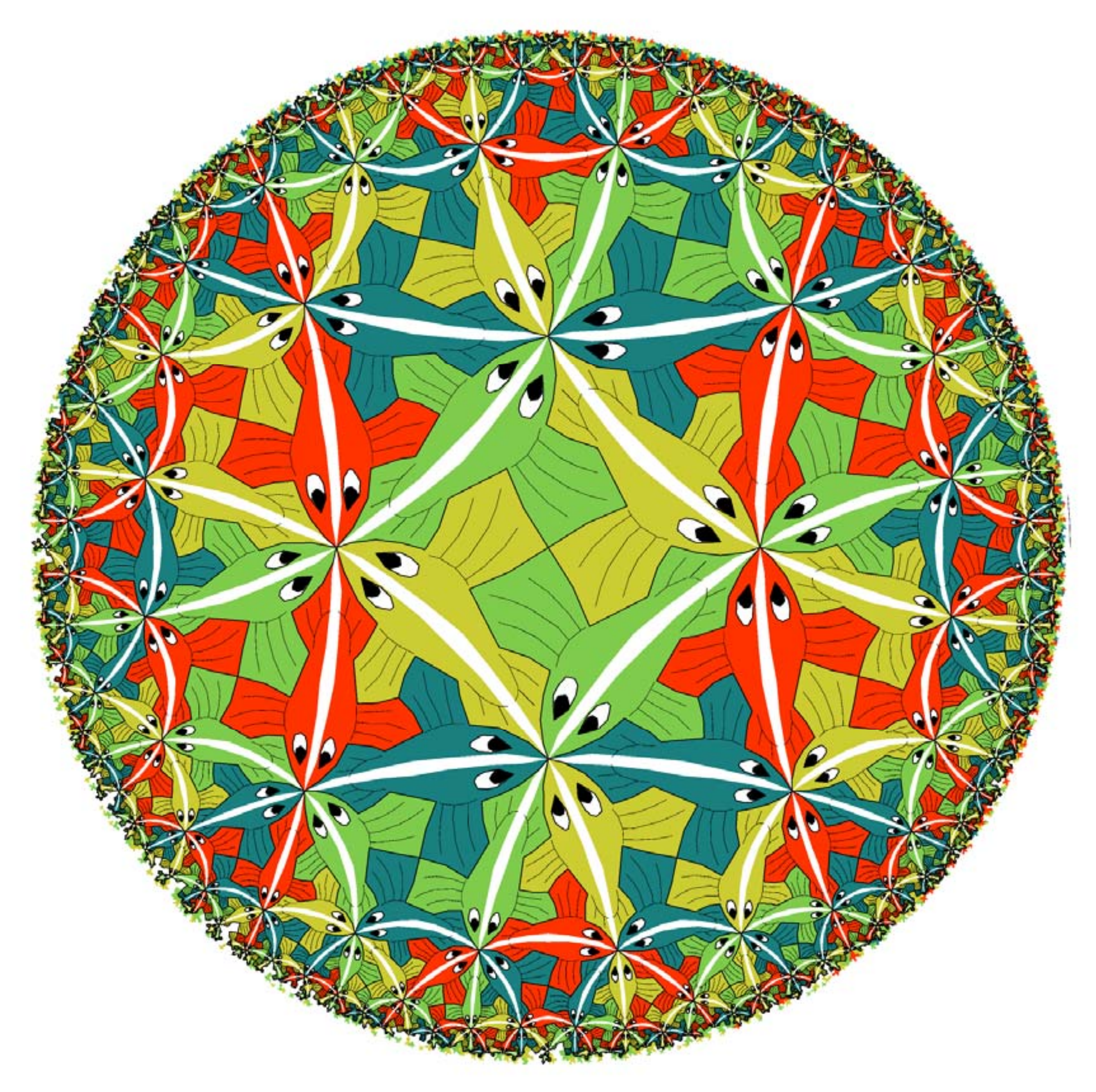
# Hyperbolic geometry rising

- The mapping between balls in $R^d$ $B(x,r)$ and points $\alpha = (x,r)$ in $H^{d+1}$ satisfies
  - 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. $|\alpha-\alpha'| \leq C$

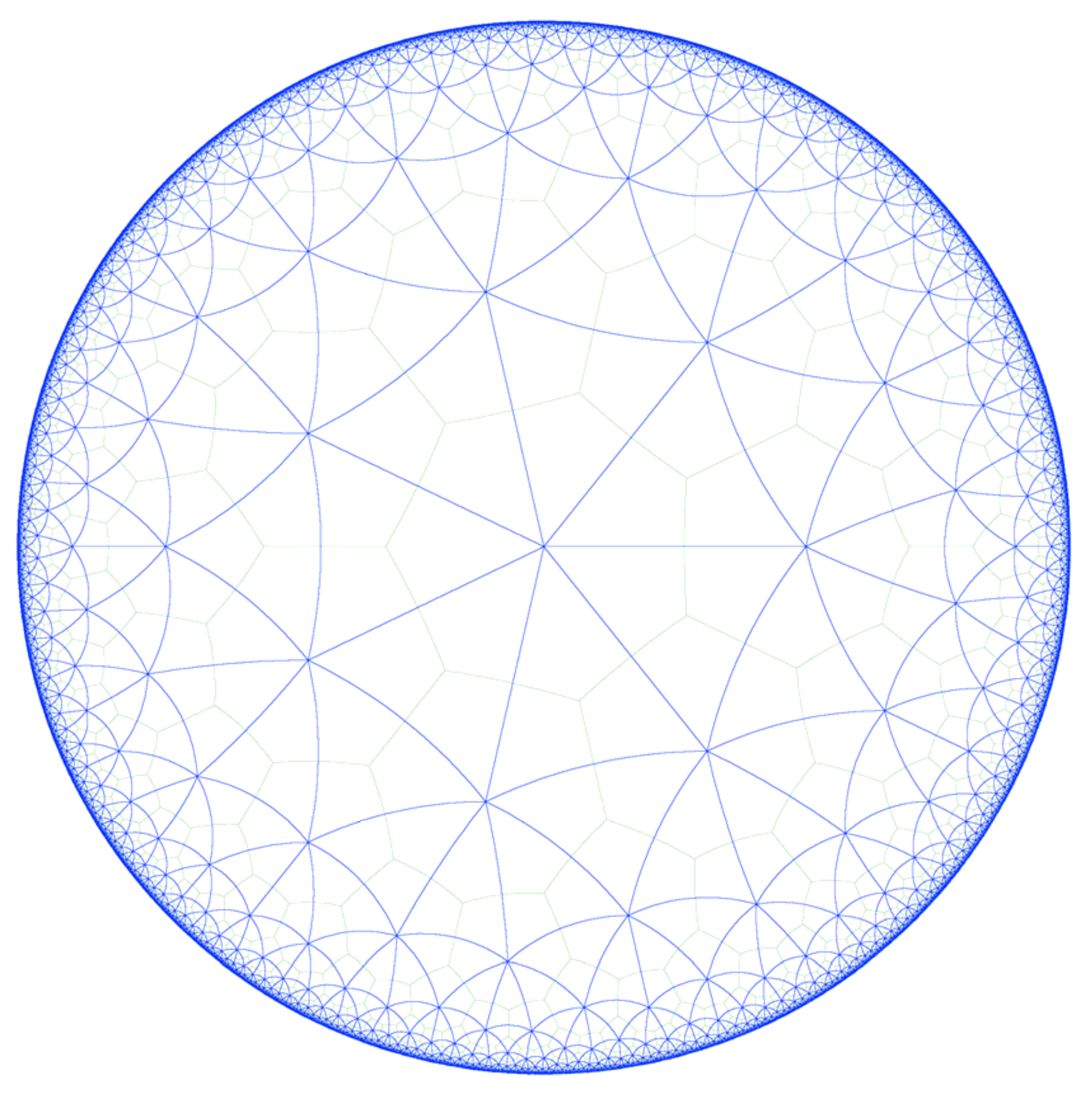
# Hyperbolic geometry defined

- Geometry in which through a point not belonging to a line passes not one but infinitely many lines parallel to the given line

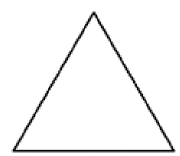
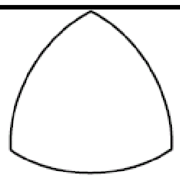
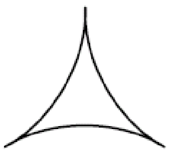
# Hyperbolic art



# Hyperbolic tessellation



# Geometry properties

| Property | Euclid. | Spherical | Hyperbolic |
|---|---|---|---|
| Curvature | $0$ | $1$ | $-1$ |
| Parallel lines | $1$ | $0$ | $\infty$ |
| Triangles are | normal | thick | thin |
| Shape of triangles |  |  |  |
| Sum of angles | $\pi$ | $> \pi$ | $< \pi$ |
| Circle length | $2\pi R$ | $2\pi \sin R$ | $2\pi \sinh R$ |
| Disc area | $2\pi R^2/2$ | $2\pi(1 - \cos R)$ | $2\pi(\cosh R - 1)$ |

# Main hyperbolic property

- 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 of radius R, where $N = \kappa\, e^{R/2}$, $N$ is the number of nodes in the network and $\kappa$ controls its average degree
  - Average degree is fixed (by $\kappa$) to the same value ($\sim 6$, like in many real networks) for all modeled networks

# Node distribution

- Number of nodes $n(r)$ located at distance $r$ from the disc center is

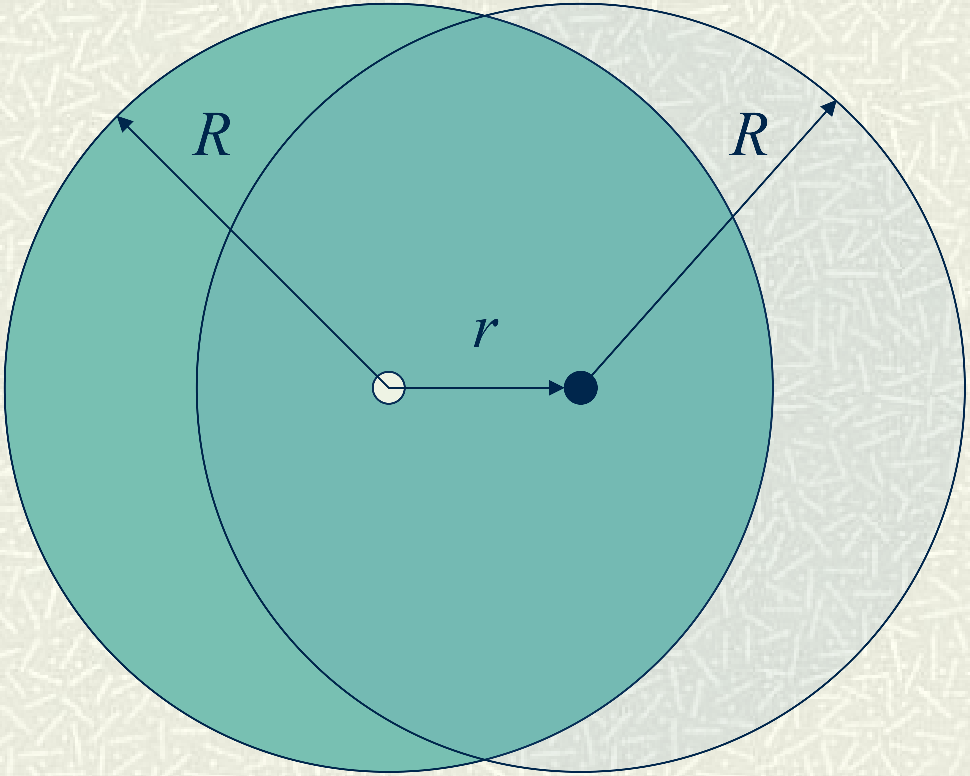
$$n(r) \sim e^{\alpha r}$$

where $\alpha = 1$ corresponds to the uniform node distribution in the hyperbolic plane of curvature $-1$

# Connection probability

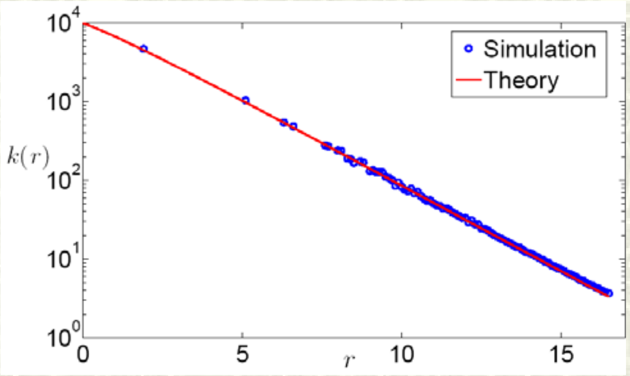
- Connect each two nodes if the distance between them is less than or equal to *R*

# Average node degree at distance $r$ from the disc center



# Average node degree at distance *r* from the disc center

- For $\alpha = 1$, we obtain a terse but exact expression




- For other $\alpha$:

$$k(r) \sim e^{-\beta r}$$

where

$\beta = \alpha$ if $\alpha \leq ½$

$\beta = ½$ otherwise

# Node degree distribution

- Is given by the combination of exponentials to yield a power law

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

  where

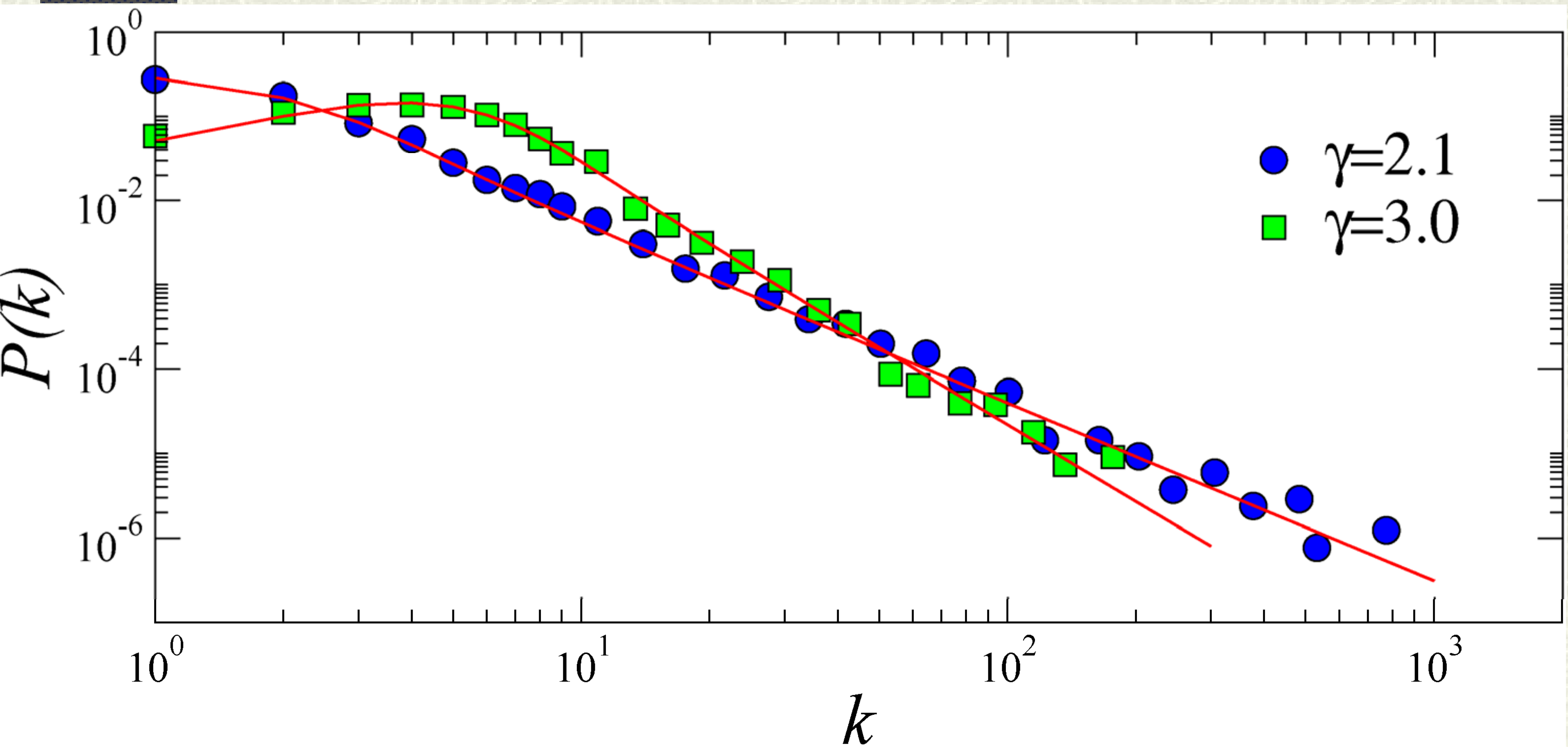
  $\gamma = 1 + \alpha/\beta =$

  $2$ if $\alpha \leq ½$; or

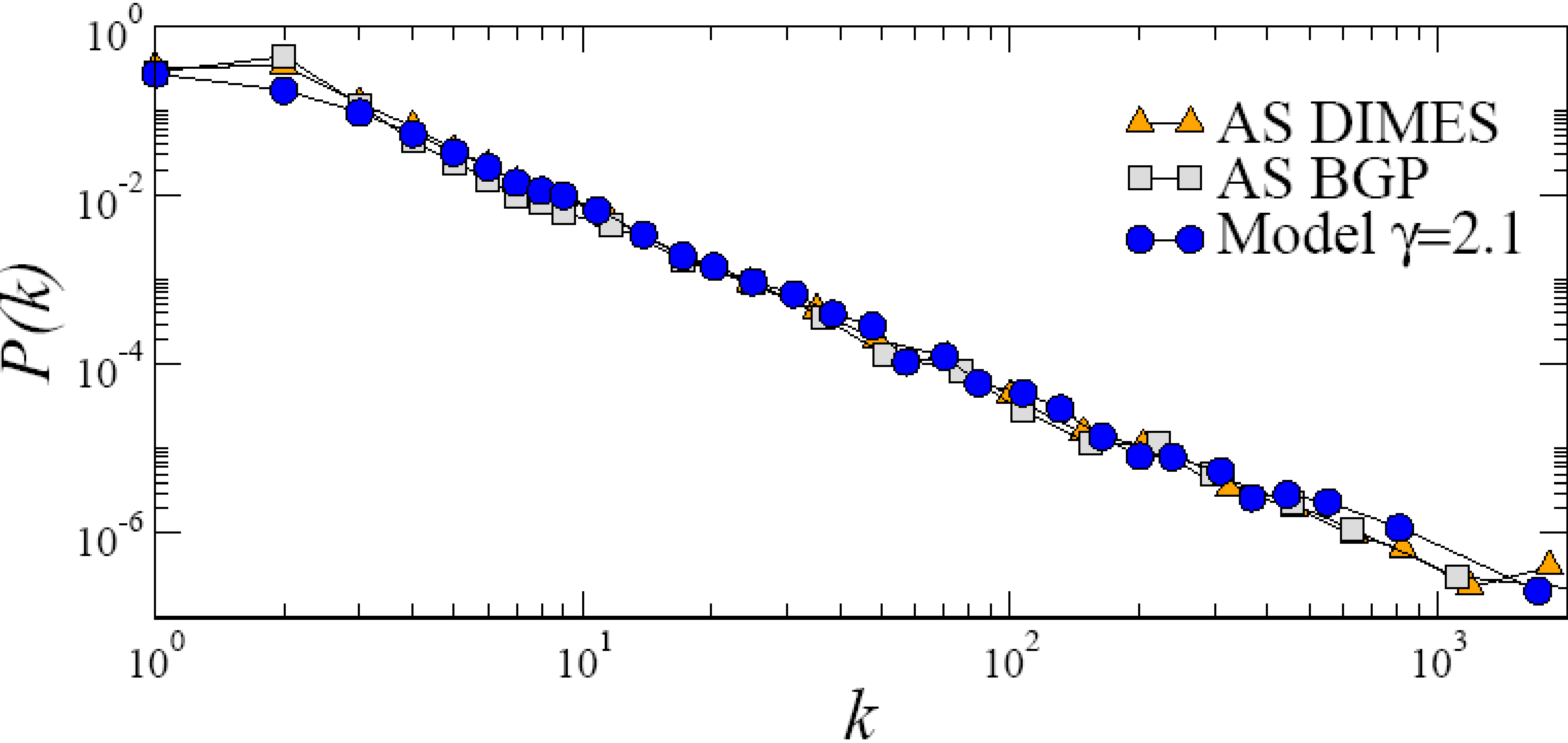
  $2\alpha + 1$ otherwise

- The uniform node distribution in the plane $(\alpha = 1)$ yields $\gamma = 3$

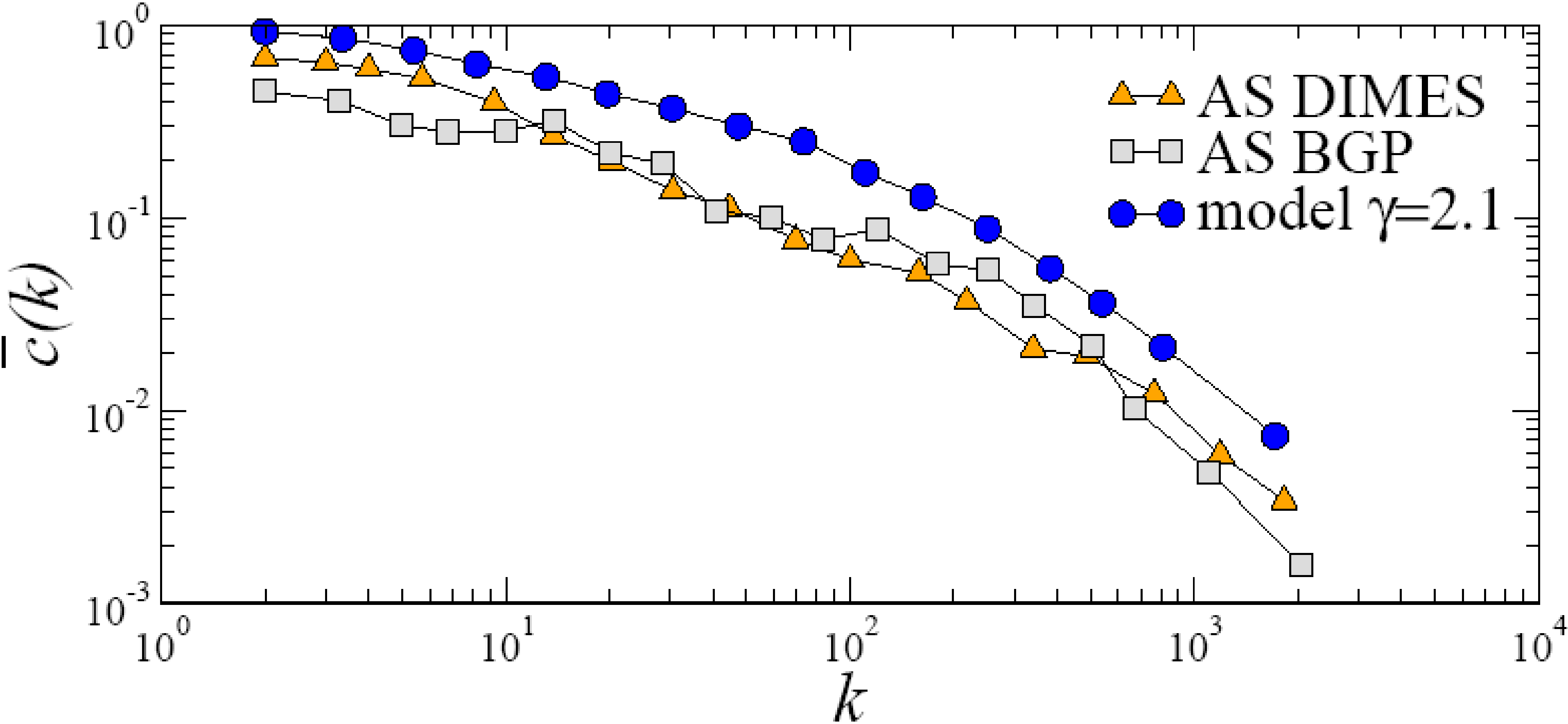
# Node degree distribution: theory vs. simulations



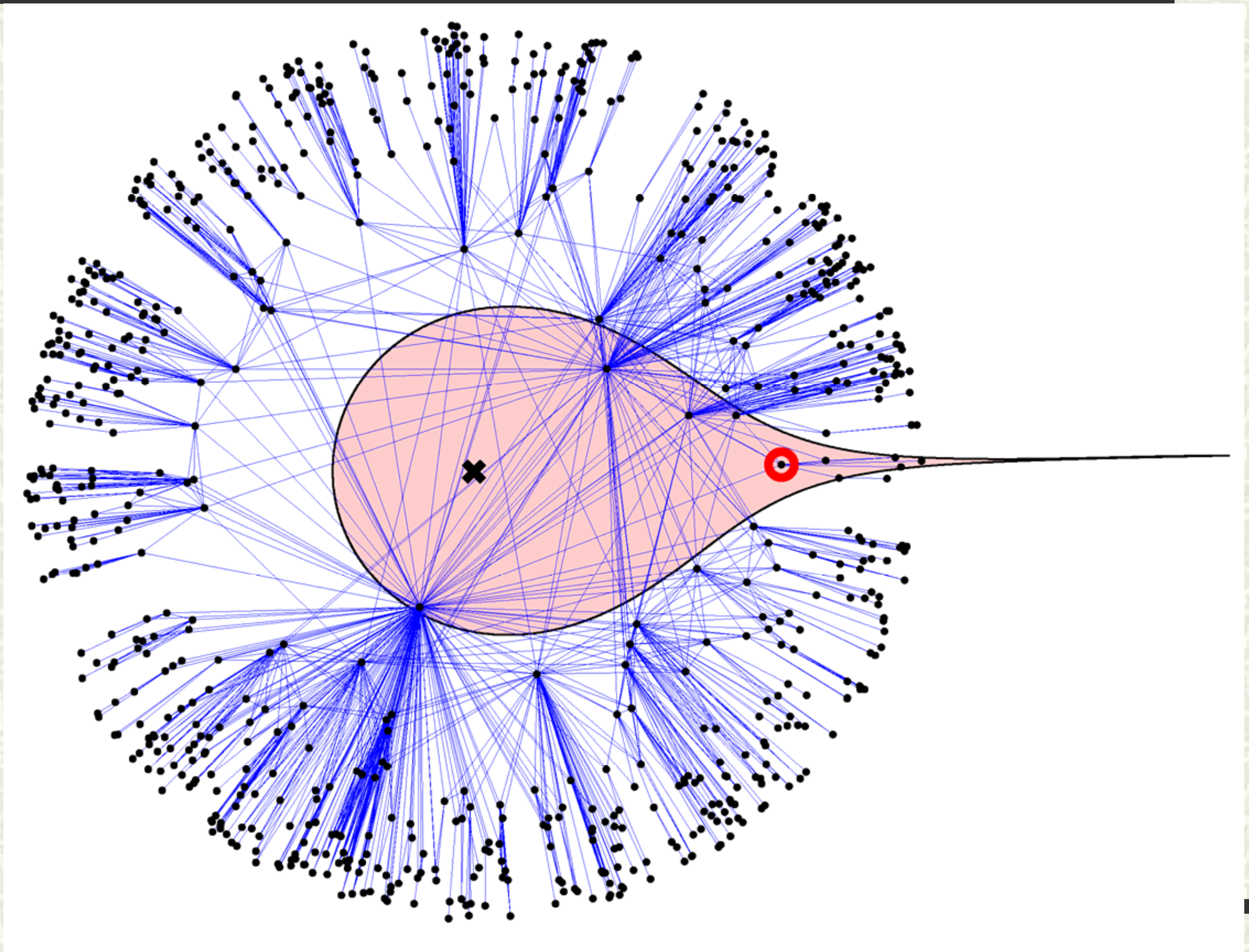
# Node degree distribution: model vs. Internet



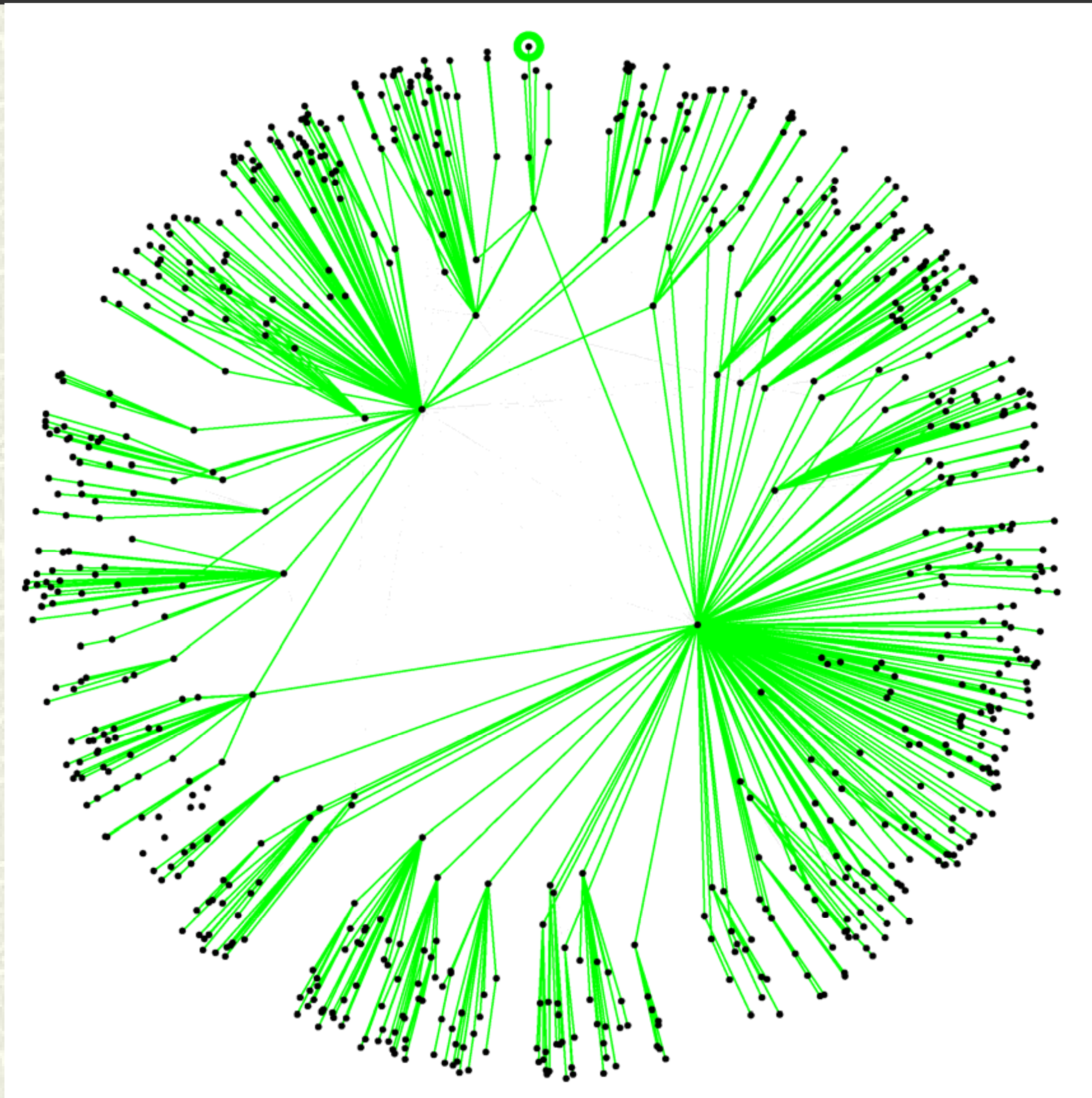
# Clustering: model vs. Internet



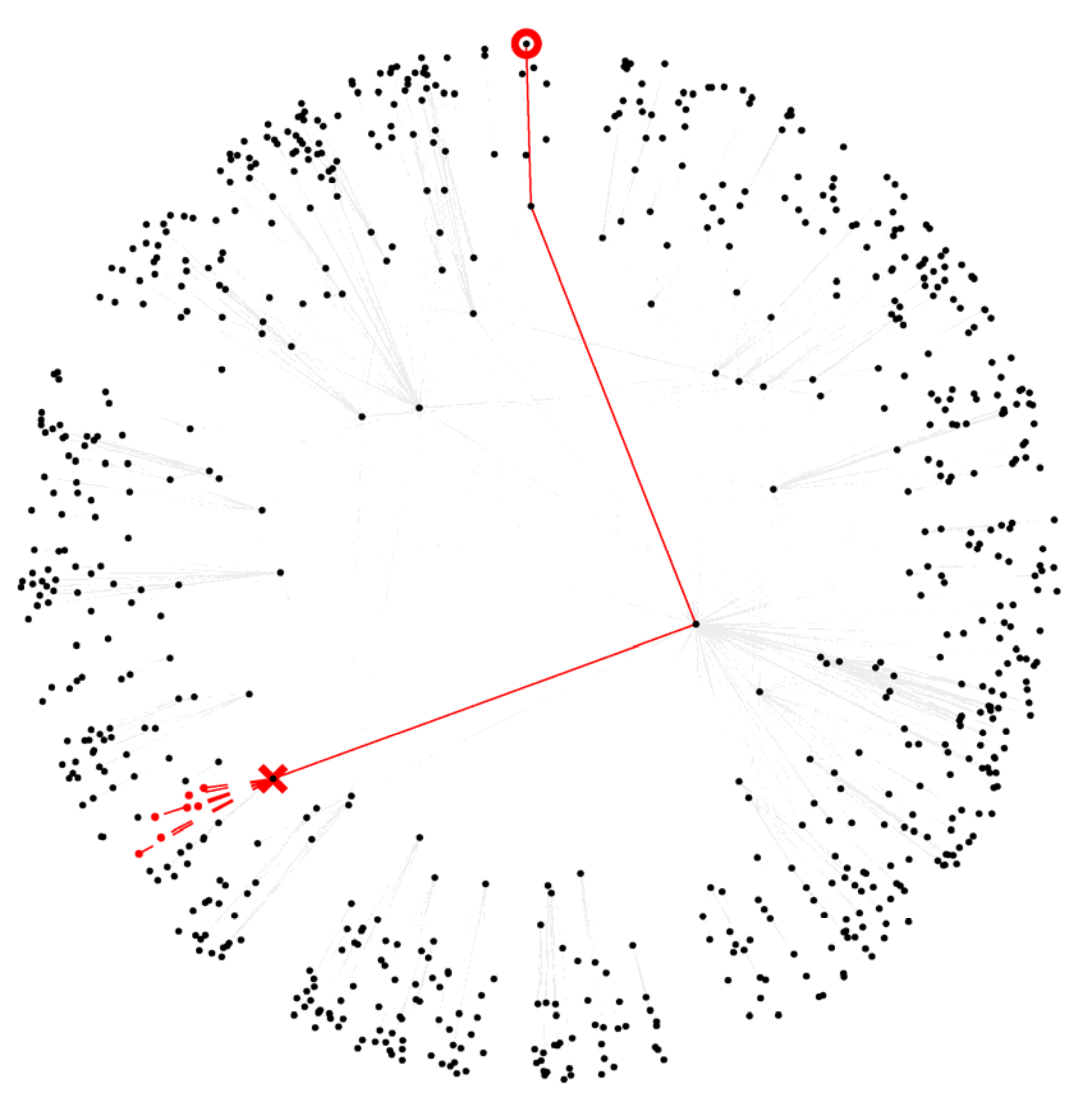
# Visualization of a modeled network



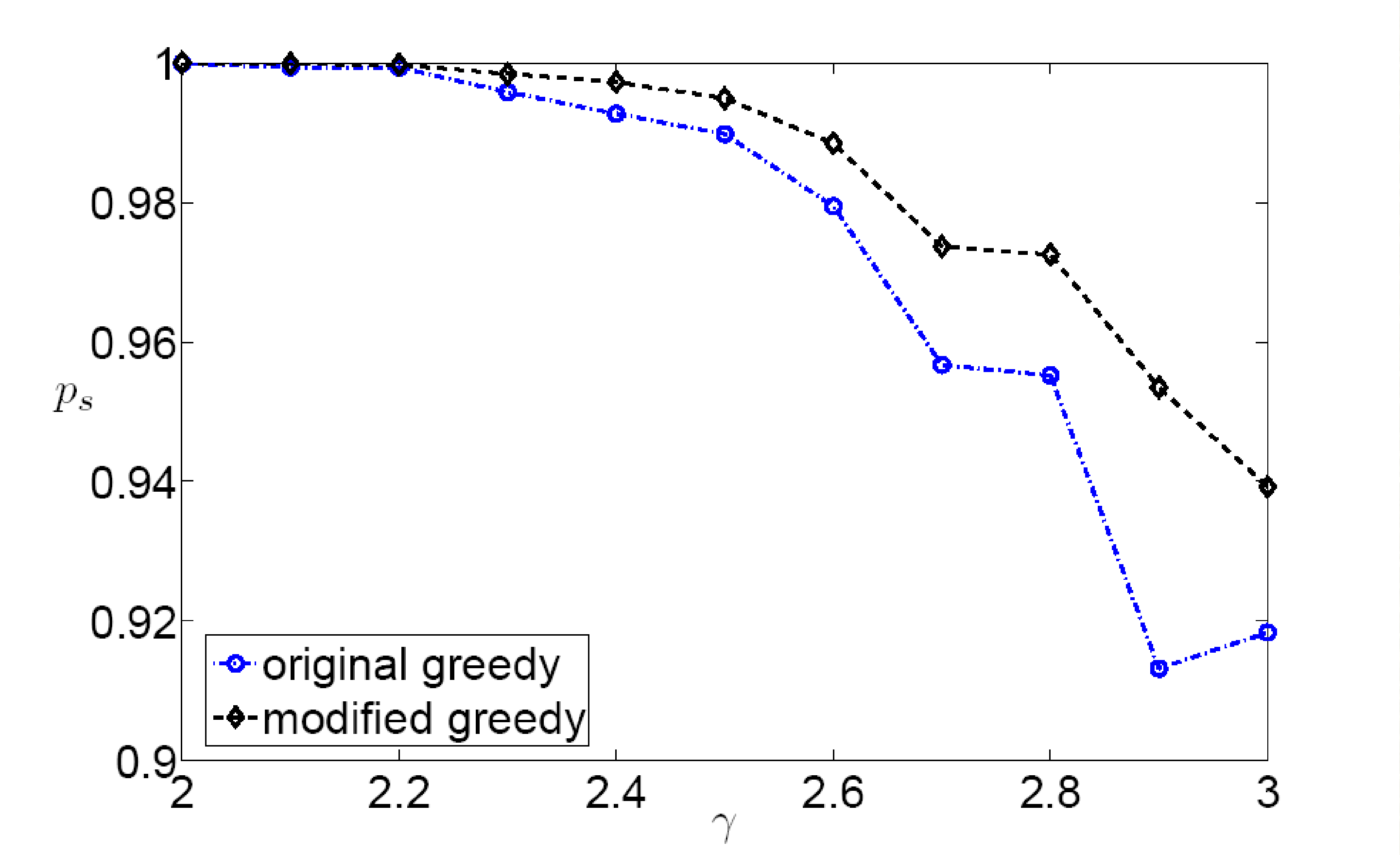
# Successful greedy paths



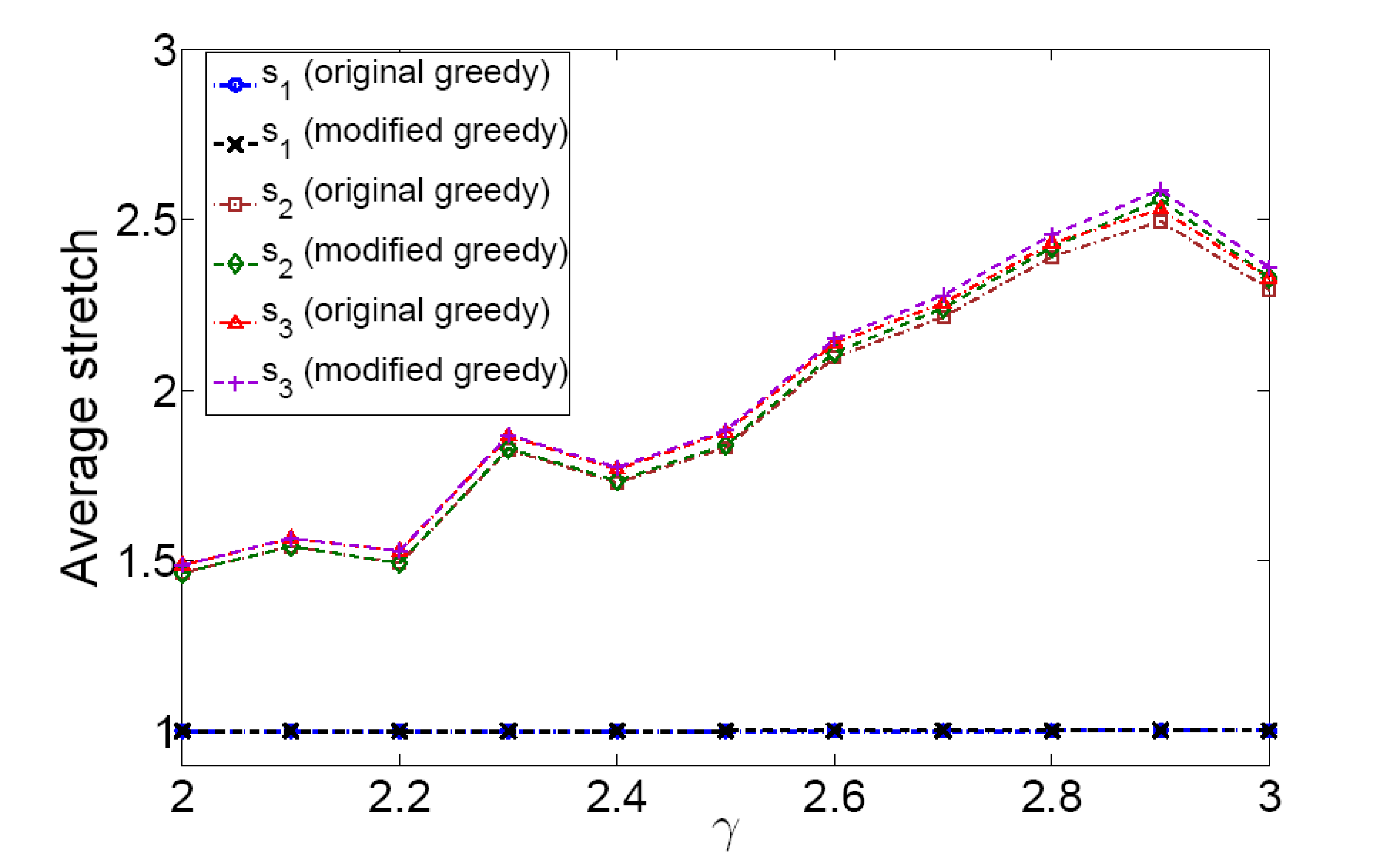
# Unsuccessful greedy paths



# Percentage of successful paths



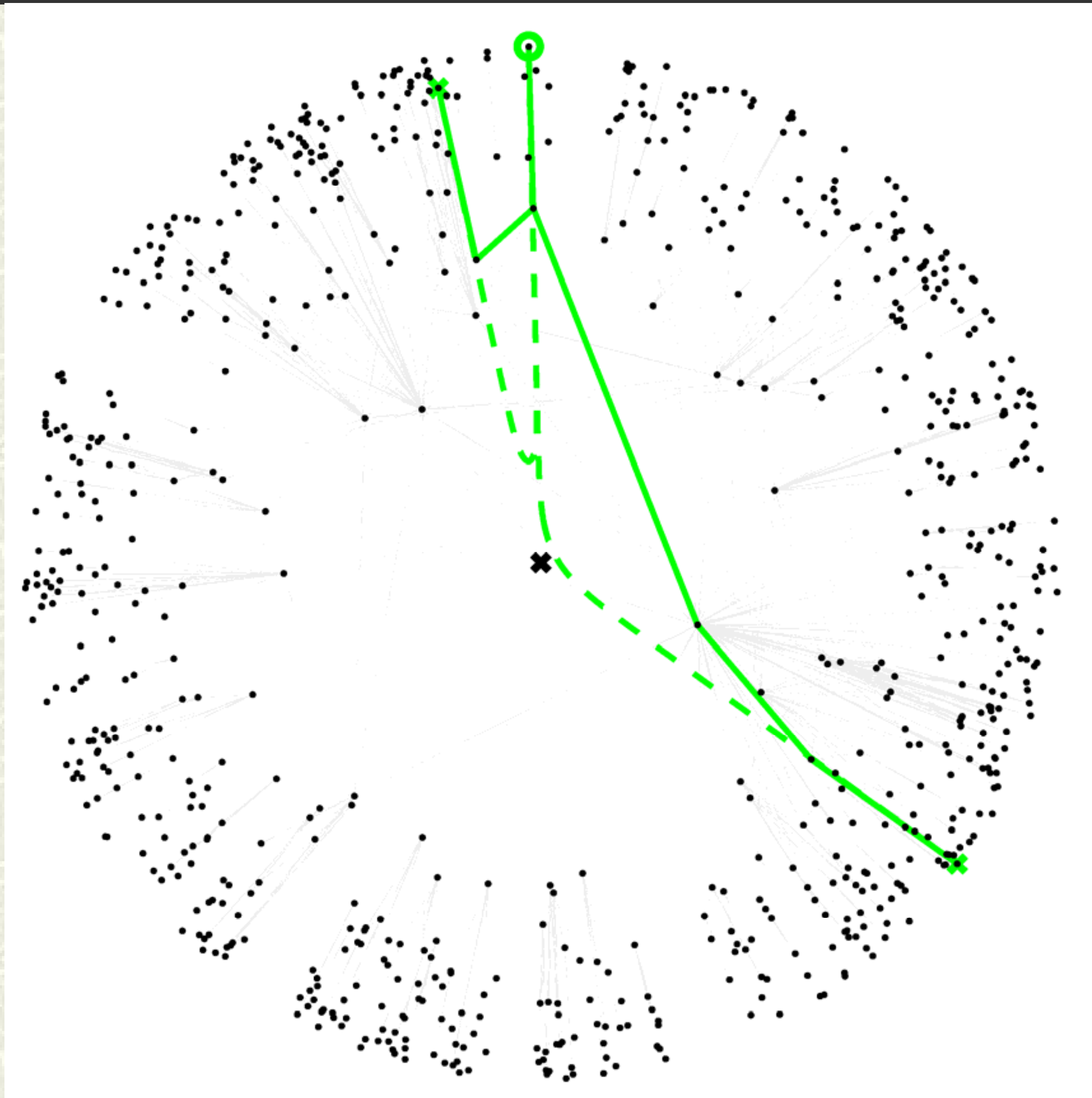
# Multiplicative average stretch



# Robustness of greedy routing w.r.t. network dynamics

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

# Hyperbolic geometry vs. scale-free topology

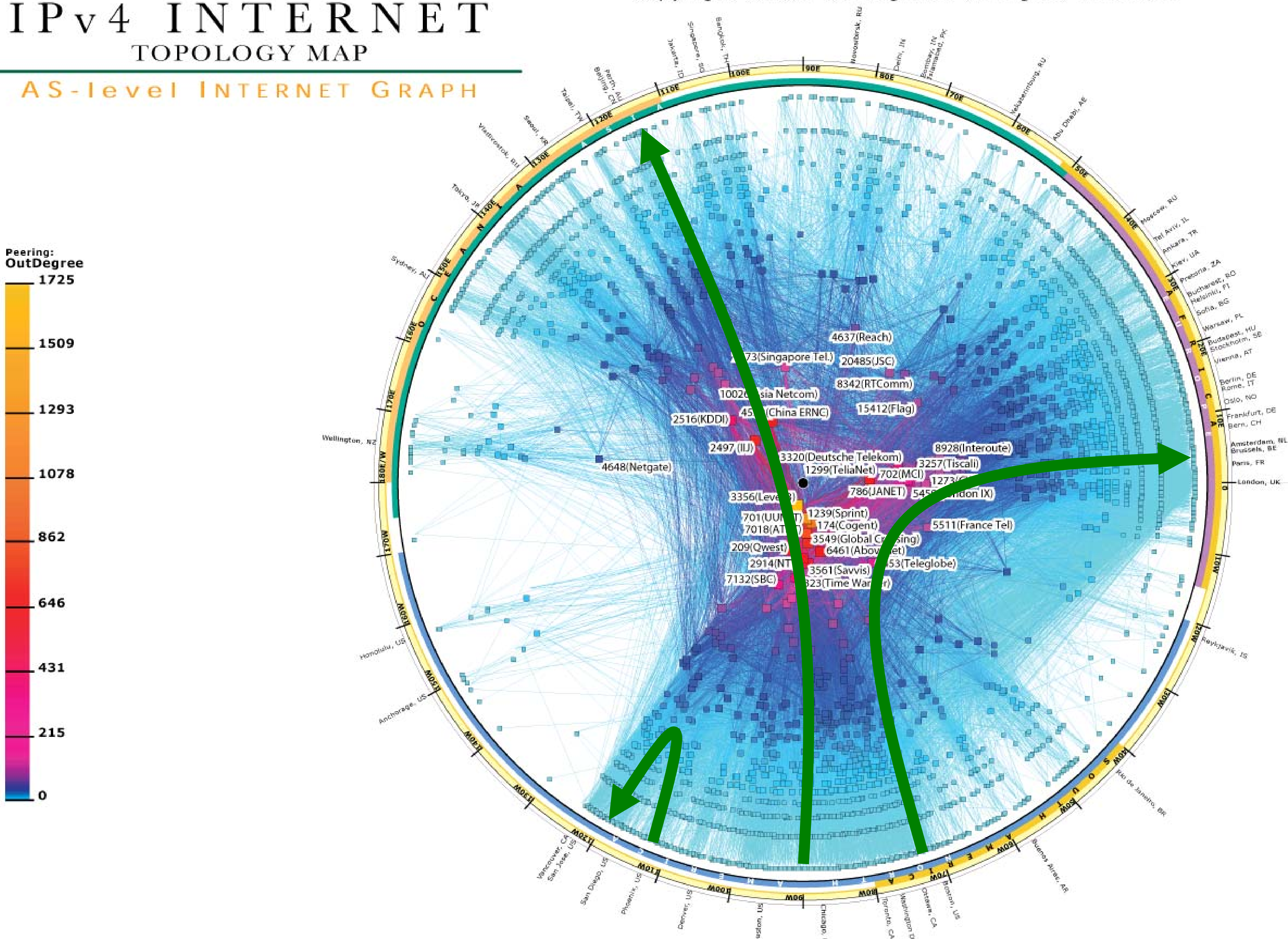


# IPv4 INTERNET

## TOPOLOGY MAP

## AS-level INTERNET GRAPH





# In summary

- Scale-free networks are congruent w.r.t. hidden hyperbolic geometries
  - Greedy paths follow shortest paths that approximately follow shortest hidden paths, i.e., geodesics in the hyperbolic space
- This congruency is robust w.r.t. network dynamics/evolution
  - There are many shortest paths between the same source and destination that satisfy the above properties
  - If some of them go away, others remain available, and greedy routing still finds them

# Conclusion

- Hidden hyperbolic metric spaces explain, simultaneously, the two main topological characteristics of complex networks
  - scale-free degree distributions
  - strong clustering
- Greedy routing mechanism in these settings may offer virtually infinitely scalable routing algorithms 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*)

# Directions for future research

- Find the structure of hidden metric spaces underlying real networks
  - measure similarities based on *intrinsic, network-specific* node attributes (instead of similarities reconstructed from the network topology, e.g., by community detection algorithms)
  - obtain a finite metric space
  - find the most appropriate (least distorted) embeddings for it into continuous model spaces
- Find the coordinates of nodes in hidden spaces
  - given a model space, can nodes “know” (compute) their positions in it using only local information???
- Applications in routing, search, recommender systems, systems biology, cognitive science, protein folding, etc.