Greedy Forwarding in Scale-Free Networks Embedded in Hyperbolic Metric Spaces

Fragkiskos Papadopoulos
CAIDA/UCSD
D. Krioukov, M. Boguñá, A. Vahdat

Outline

- Motivation
- Scale-free Networks and Hyperbolic Metric Spaces
- Greedy Forwarding
- Conclusion

Motivation

- *One of the most serious scaling limitations with the existing Internet architecture: the communication overhead of routing protocols (RFC4984, IAB 2007)*
- Internet routing tables
  - 300,000 prefix entries (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

Motivation (cont.)

- **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 of a network topology (*Krioukov et al. CCR 2007*)
- Complex/scale-free networks, like the Internet, (power-law degree distribution $P(k) \sim 1/k^{\gamma}$, strong clustering) 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?*
- *Many networks in nature (e.g. biological networks) can do this! But how is this possible?*

Milgram's experiments (Sociometry, 1969)

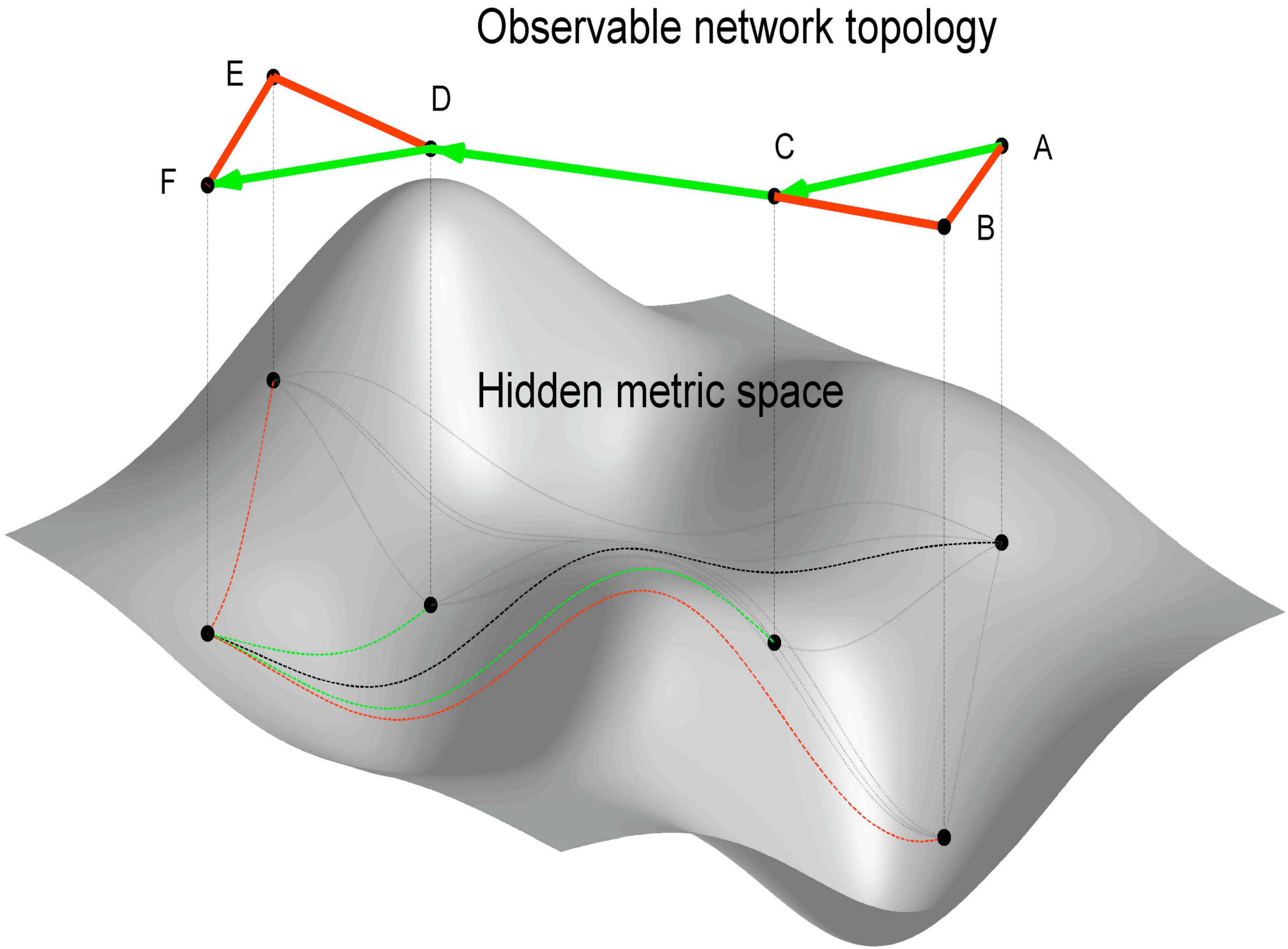
- Settings: random people were asked to forward a letter to a random individual by passing it to their friends who they thought would maximize the probability of letter moving "closer" to the destination
- Results: surprisingly many letters (*30%*) reached the destination by making only *~6* hops on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

Hidden metric space explanation
(J. Kleinberg, Nature 2000)

- All nodes exist in a Euclidean (hidden) 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

Greedy forwarding (J. Kleinberg)

- To reach a destination, each node forwards the message/packet to the neighbor that is closest to the destination in the hidden space.


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

Greedy forwarding (cont.)

- Works well (high probability of successfully finding the destination, shortest paths) only if the network topology is ***congruent*** with the hidden space
- Kleinberg's model based on Euclidean underlying space produces only *k-regular* graphs
- But most real networks, including the Internet, are scale-free (power-law degree distribution, strong clustering)

Questions raised:

I. What are the metric spaces underlying real networks?

II. What is the efficiency of greedy forwarding in these networks?

Outline

- Motivation
- Scale-free Networks and Hyperbolic Metric Spaces
- Greedy Forwarding
- Conclusion

Hidden metric spaces underlying real networks are hyperbolic

- Network nodes can often be classified hierarchically, at least approximately (*Watts et al. Science 2002*)
- 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

Main hyperbolic property (the exponential expansion of space)

- Circle lengths and disc areas grow with their radius R as

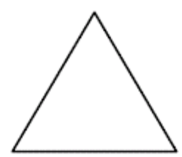
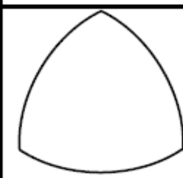
$$\sim e^R$$

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

$$\sim b^R$$

  where b is the tree branching factor
- The metric structures of hyperbolic spaces and trees are essentially the same

Hyperbolic geometry properties

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

Synthetic networks embedded in hyperbolic spaces

- What network topologies emerge in the simplest possible settings involving hidden hyperbolic metric spaces?

The model

1. Let N be the number of nodes in the network

   Fix the hyperbolic disc radius to R according to $N = \kappa e^{R/2}$

   Constant κ controls the average degree

2. Uniformly distribute nodes in the disc, by assigning to each node two coordinates:

   (i) Angular θ uniformly distributed in $[0, 2\pi]$

   (ii) Radial r with density $\rho(r) = \dfrac{\sinh r}{\cosh R - 1} \approx e^{(r-R)} \sim e^{r}$

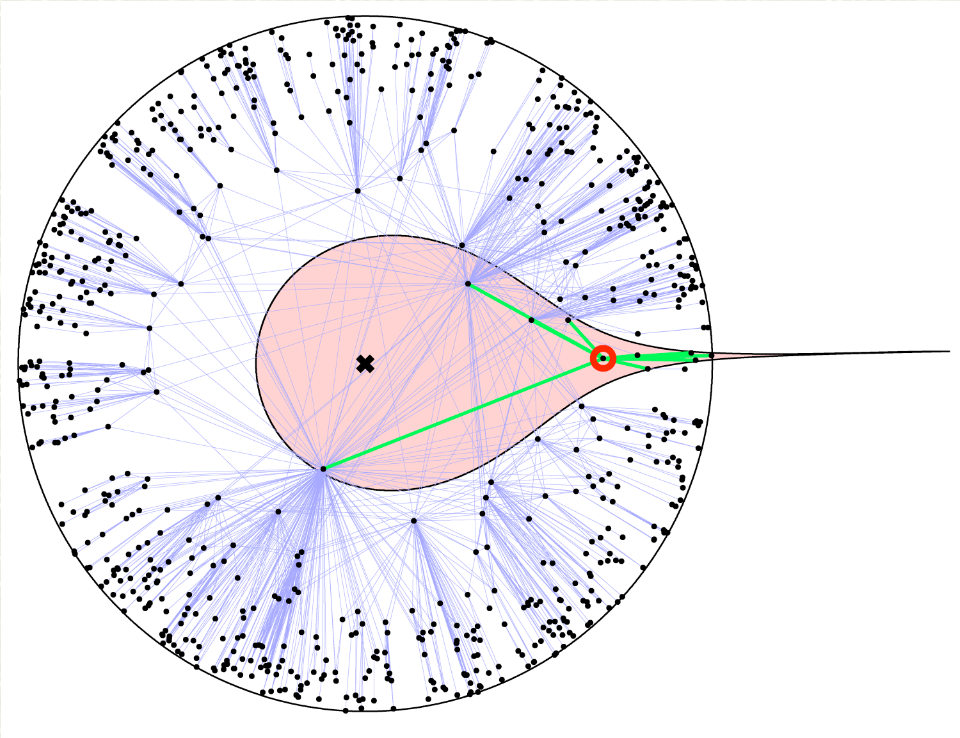
   (Recall: circle length is $2\pi \sinh r$)

   Non-uniform node distribution if $\rho(r) = \dfrac{\alpha \sinh \alpha r}{\cosh \alpha R - 1} \approx \alpha e^{\alpha(r-R)} \sim e^{\alpha r}$

The model (cont.)

3. Connect each pair of nodes if their hyperbolic distance $d \leq R$

$$d = \cosh^{-1}\left(\cosh(r)\cosh(r') - \sinh(r)\sinh(r')\cos(\Delta\theta)\right)$$



Average node degree at distance r from the disc center

$$\bar{k}(r) \approx \kappa e^{\frac{1}{2}R} \left\{ \frac{2\alpha}{\pi(\alpha - \frac{1}{2})} e^{-\frac{1}{2}r} + \left(1 - \frac{2\alpha}{\pi(\alpha - \frac{1}{2})} \right) e^{-\alpha r} \right\}$$

Important part:

$\bar{k}(r) \sim e^{-\beta r}$ where $\beta = \alpha$ if $\alpha \leq \frac{1}{2}$, $\beta = \frac{1}{2}$ otherwise

Node degree distribution

It is given by the combination of exponentials to yield a power law

Since $\bar{k}(r) \sim e^{-\beta r}$, $r(k) \sim -\frac{1}{\beta}\ln k$, and since $\rho(r) \sim e^{\alpha r}$:

$$P(k) \approx \rho[r(k)]\left|r'(k)\right| \sim k^{-\gamma},$$

where $\gamma = 1 + \frac{\alpha}{\beta} =$

2 if $\alpha \leq 1/2$; or

$2\alpha+1$ otherwise

Thus, we can create networks with any γ in [2, 3] like the vast majority of complex networks (AS Internet: $\gamma \approx 2.1$)

Clustering (number of 3-cycles)

Our modeled networks also have strong clustering (triangle inequality)

But previous connection probability does not allow tuning it

It yields higher clustering than in the Internet

We solve this problem by introducing the following connection probability:

$p(d) = \dfrac{1}{1 + e^{\frac{1}{2T}(d-R)}}$, where T a parameter that tunes clustering

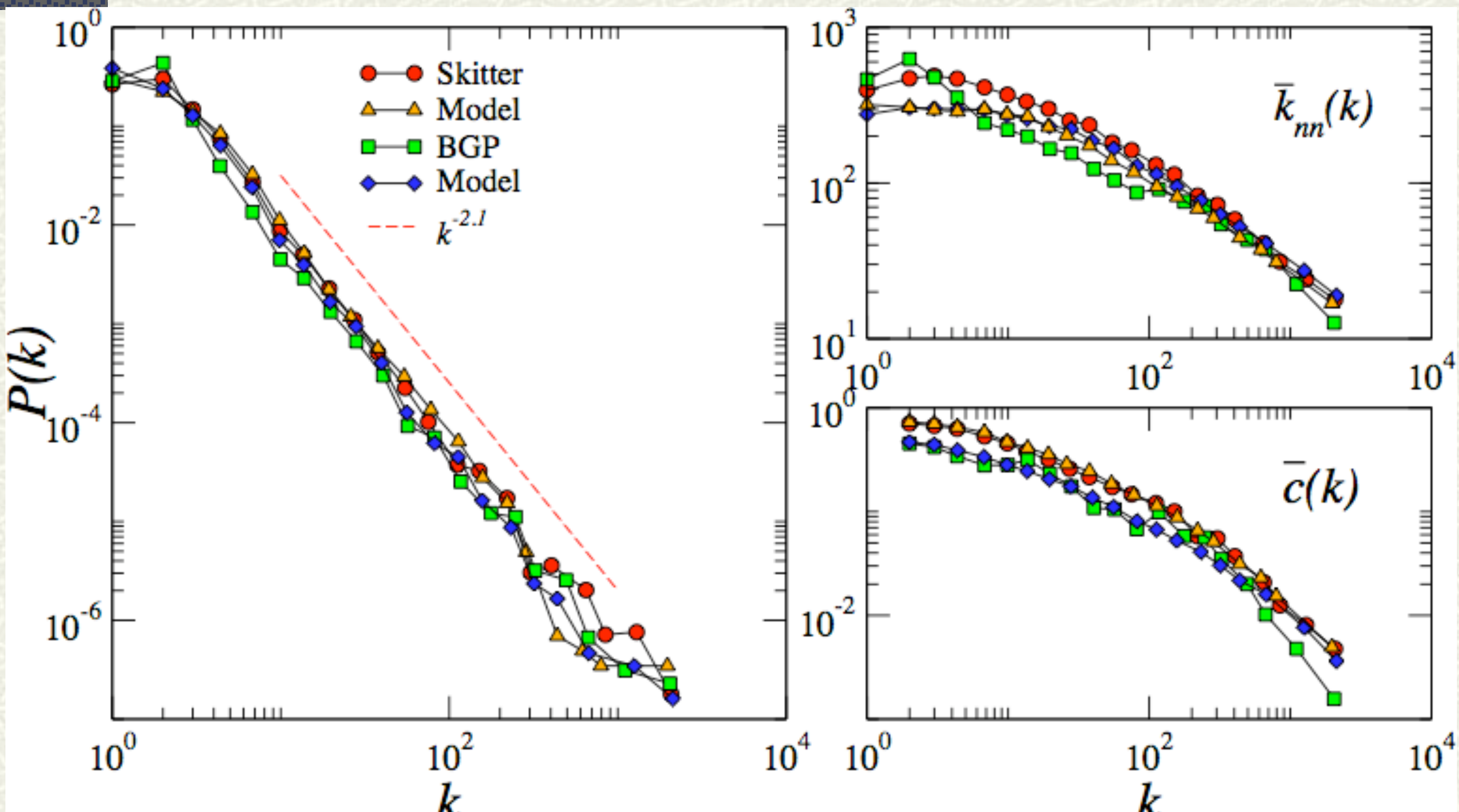
$T \to 0$ $p(d)$ becomes the step fucntion (clustering is maximized)

$T \to 1$ clustering goes to 0

Model vs. AS Internet

- We will compare the *dK*-properties (Mahadevan et al. ACM SIGCOMM 2006)
  - Degree distribution
  - Clustering
  - Degree correlations
- Reproducing these also reproduces a number of other topological properties

Model vs. AS Internet



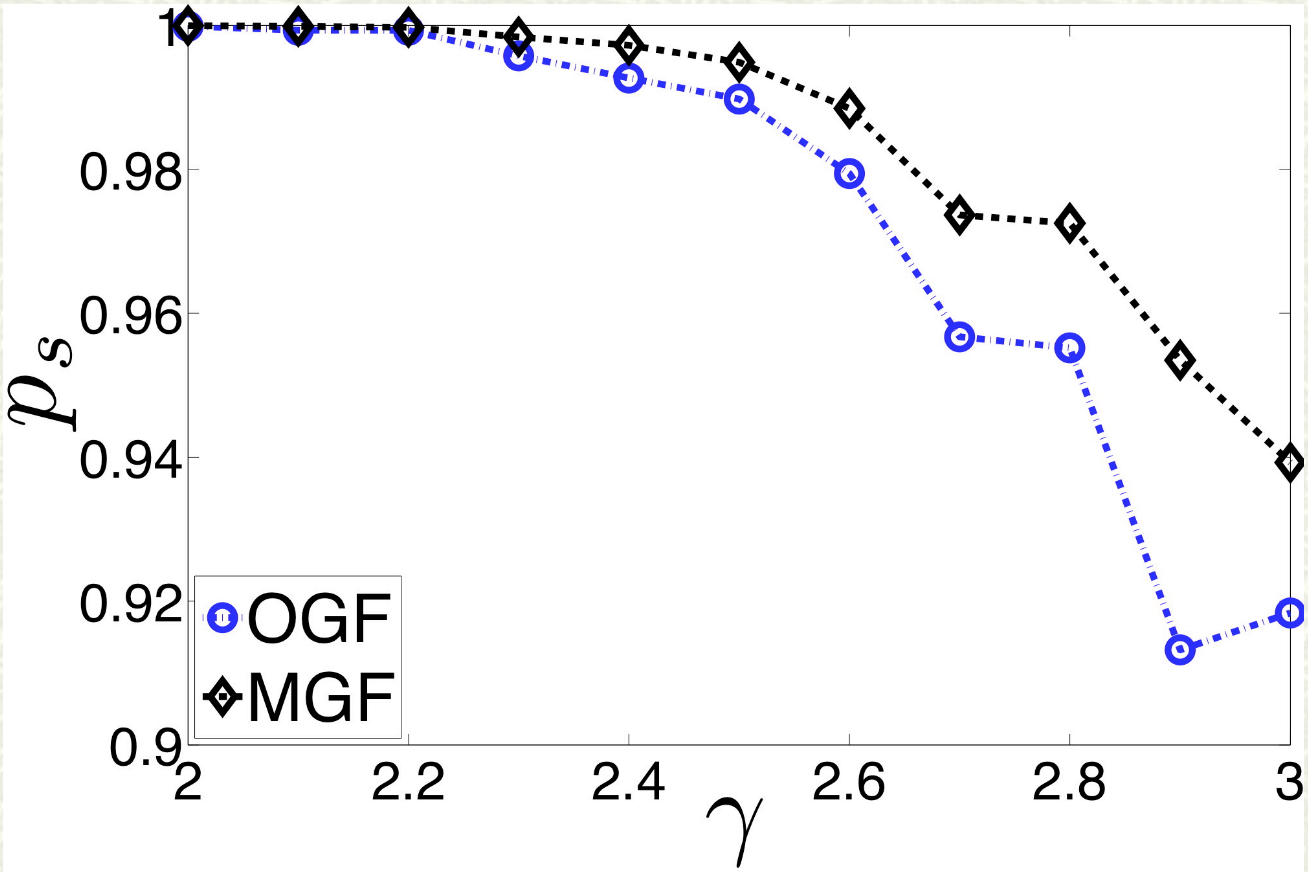
Outline

- Motivation
- Scale-free Networks and Hyperbolic Metric Spaces
- Greedy Forwarding
- Conclusion

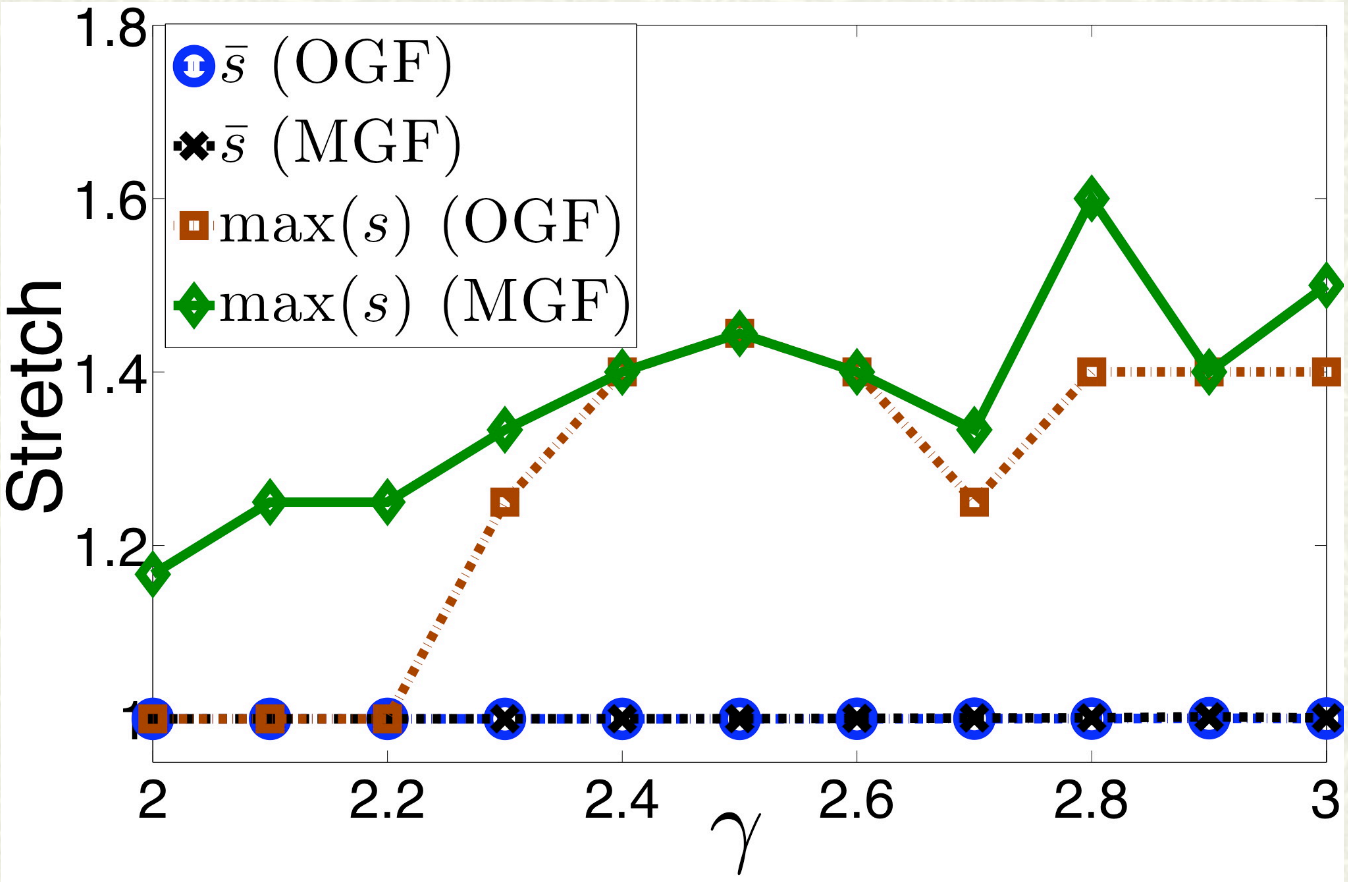
Greedy forwarding efficiency

- Two simple algorithms
  - Original Greedy Forwarding (OGF): select closest neighbor to destination, drop the packet if no one closer than current hop
  - Modified Greedy Forwarding (MGF): select closest neighbor to destination, drop the packet if a node sees it twice

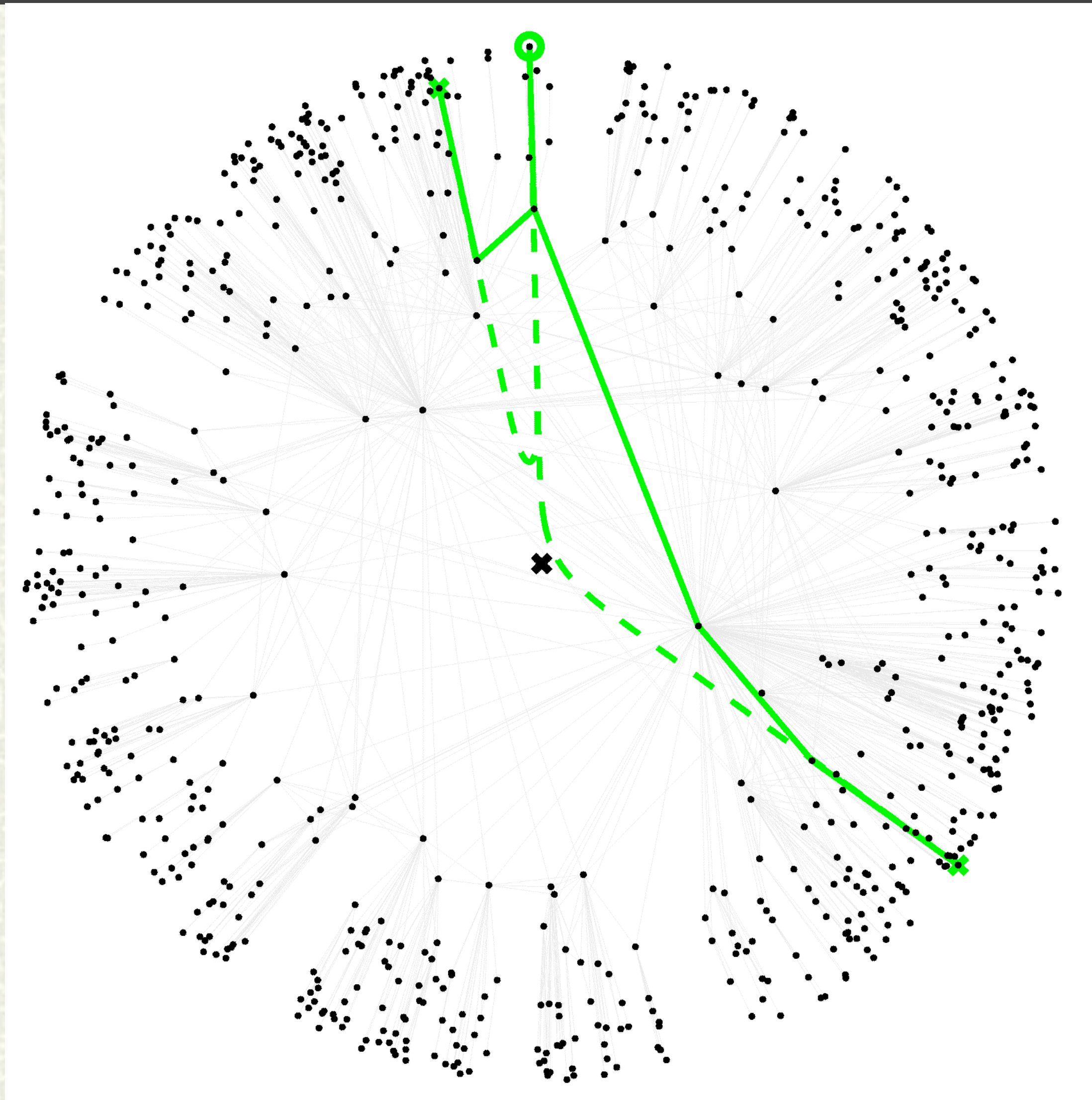
Percentage of successful paths



Average and maximum stretch



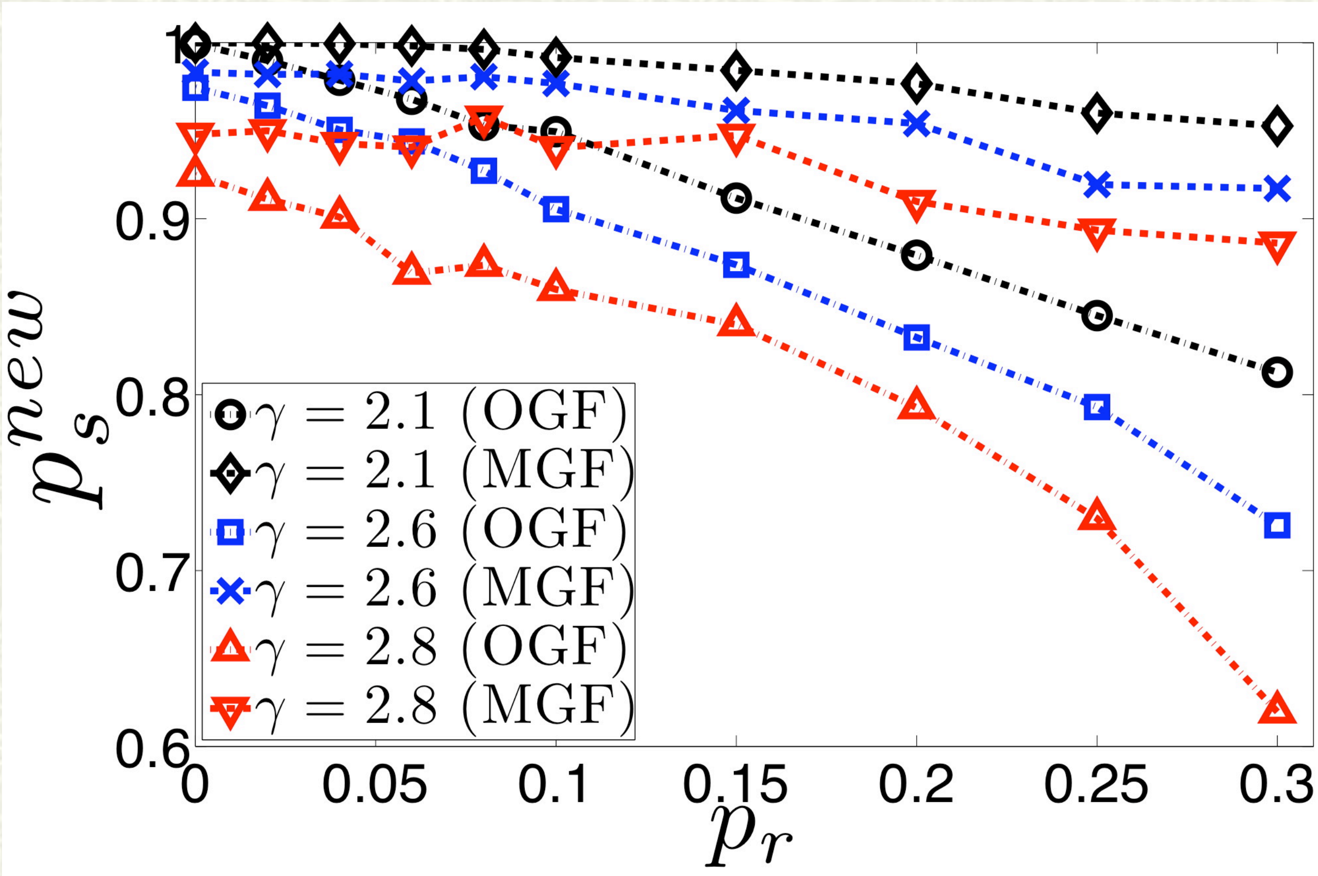
Hyperbolic geometry vs. scale-free topology



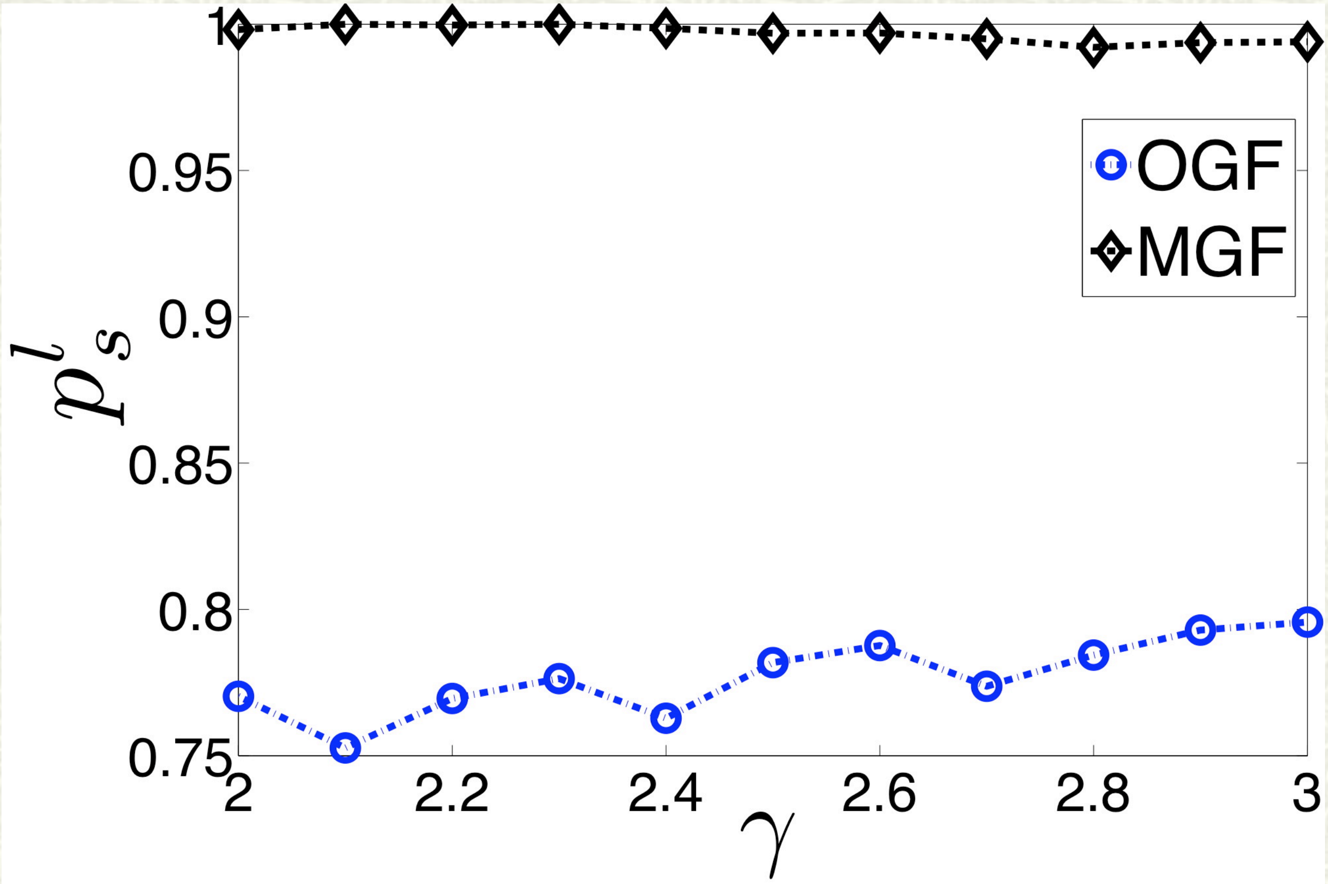
Robustness of greedy forwarding w.r.t. network dynamics

- Two scenarios
  - Scenario 1: Randomly remove a percentage of links and compute the new success ratio
  - Scenario 2: Remove a link and compute the precentage of paths that were going through it and are still successful (have found a by-pass)

Percentage of successful paths (dynamic networks, scenario 1)



Percentage of successful paths (dynamic networks, scenario 2)



In summary

- Scale-free networks are congruent w.r.t. hidden hyperbolic geometries
  - Greedy paths follow shortest paths
- This congruency is robust w.r.t. network dynamics
  - There are many shortest paths between the same source and destination
  - If some of them go away, others remain available, and greedy forwarding 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

- Given a scale-free network, how do we find its coordinates in the space *using only local information???*
- Applications to network overlays, e.g. P2P