Hyperbolic Geometry of Complex Network Data

Konstantin Zuev
http://www.its.caltech.edu/~zuev/

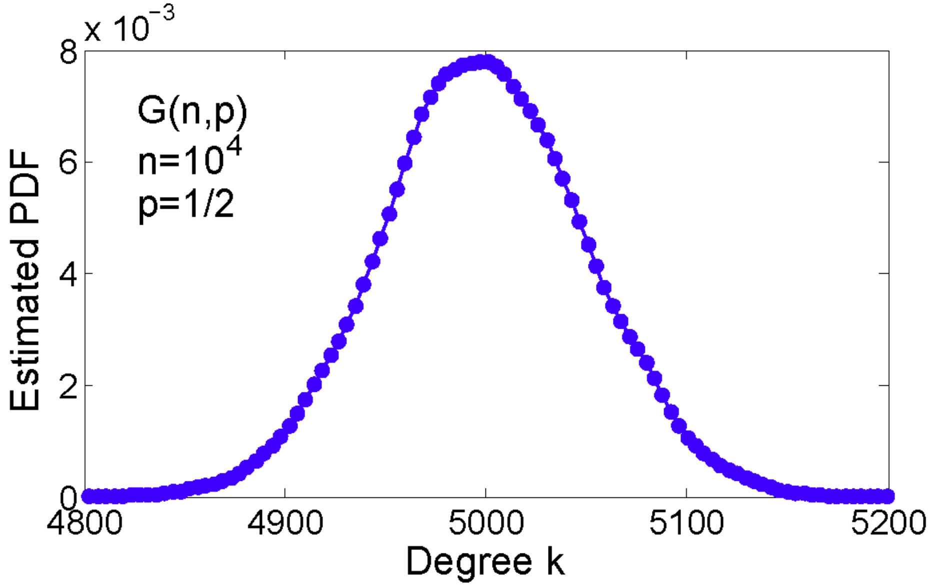
Joint work with D. Krioukov, M. Boguñá, and G. Bianconi

CMX seminar, Caltech
May 24, 2017

How do complex networks grow?

Erdős–Rényi Model G(n,p)

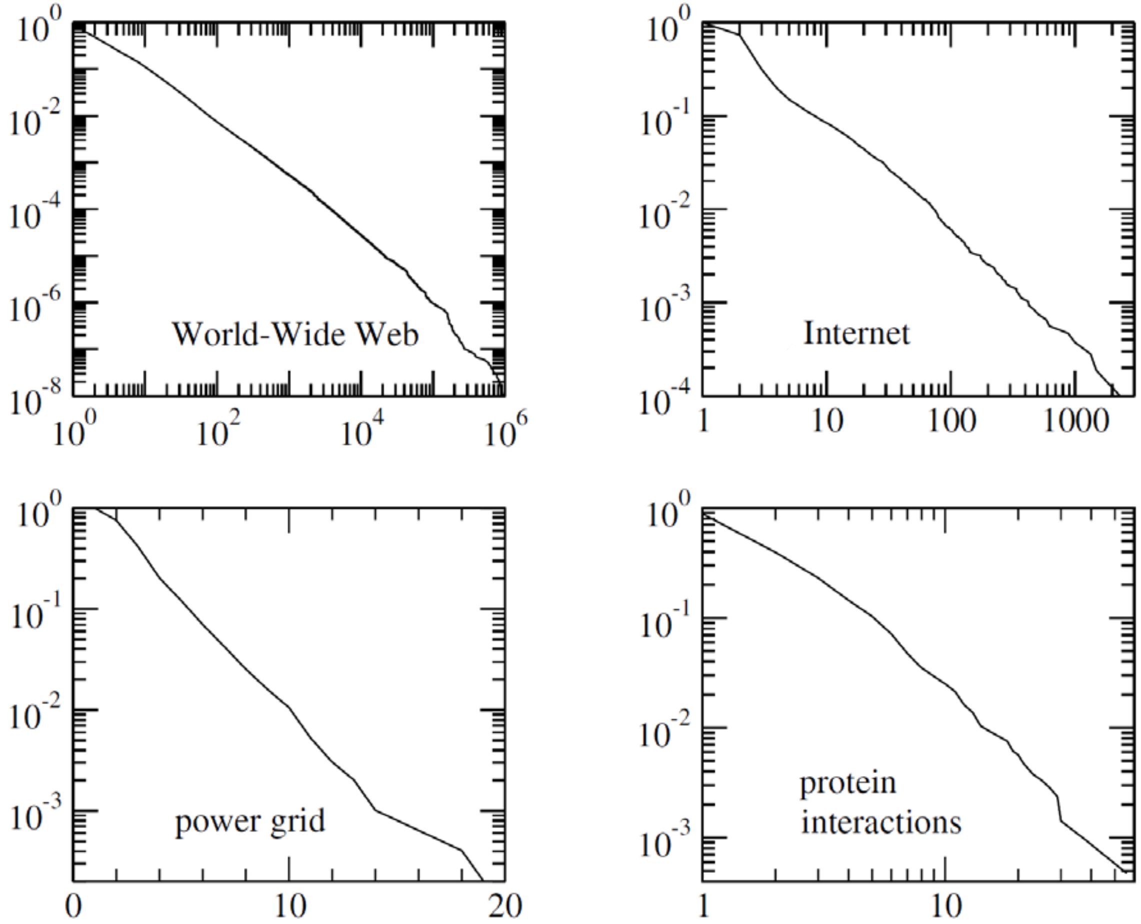
1. Take n nodes
2. Connect every pair of nodes at random with probability p



$$\mathbb{P}(k) = e^{-\bar{k}} \frac{\bar{k}^k}{k!}, \quad \bar{k} = Np$$

$\neq$

Real networks are "scale-free"



$$\mathbb{P}(k) \propto k^{-\gamma}, \quad \gamma \in [2,3]$$

Preferential Attachment Mechanism

Barabási–Albert Model

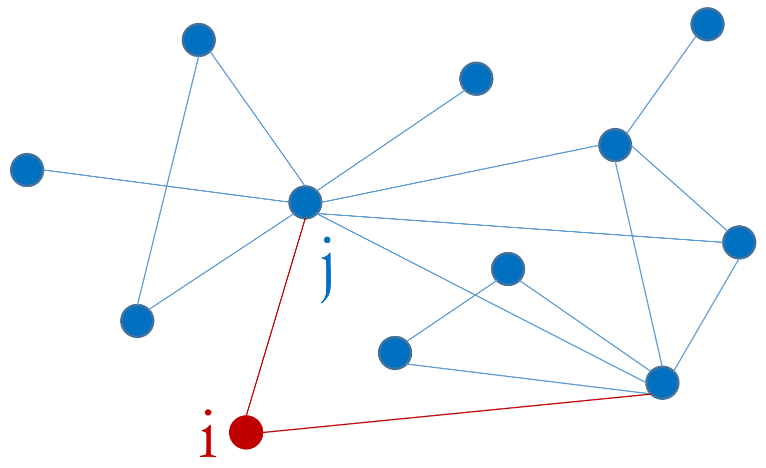
1. Start with n isolated nodes.
   New nodes come one at a time.
2. A new node i connects to m old nodes.
3. The probability that i connects to j

$$\mathbb{P}(j) \propto k_j$$

"rich gets richer"

$$\Downarrow$$

Scale-free networks with $\gamma = 3$

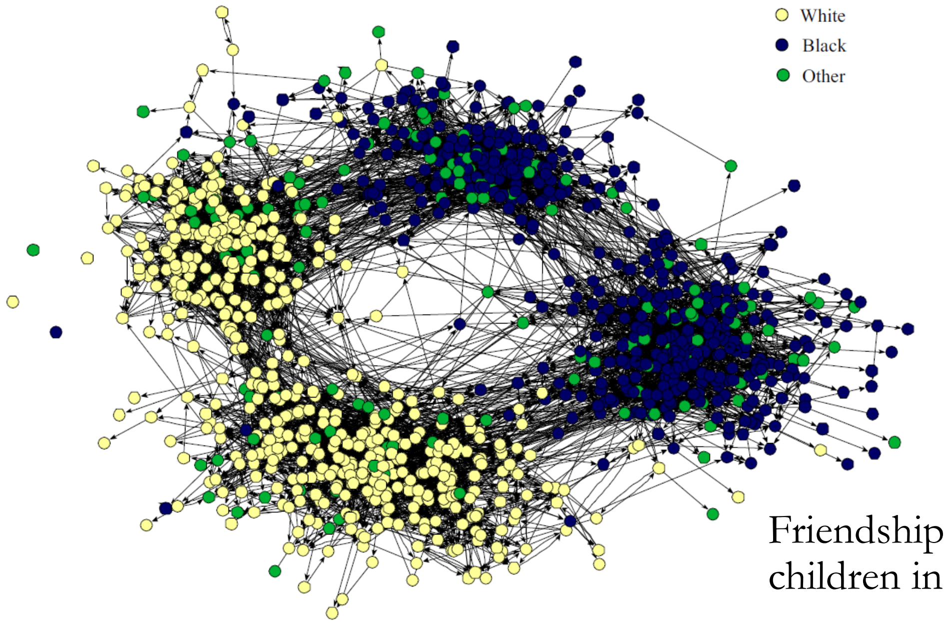


Issues with PA

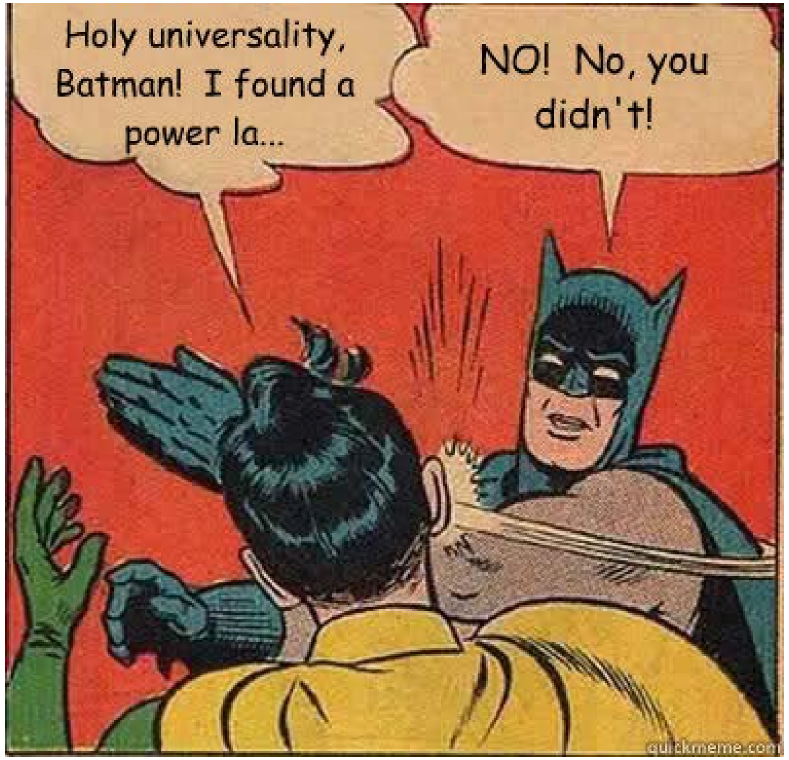
- Zero clustering
- No communities

Universal Properties of Complex Networks

- Heavy-tail degree distribution ("scale-free" networks) } PA
- Strong clustering ("many triangles")
- Community structure



Friendship network of children in a U.S. school

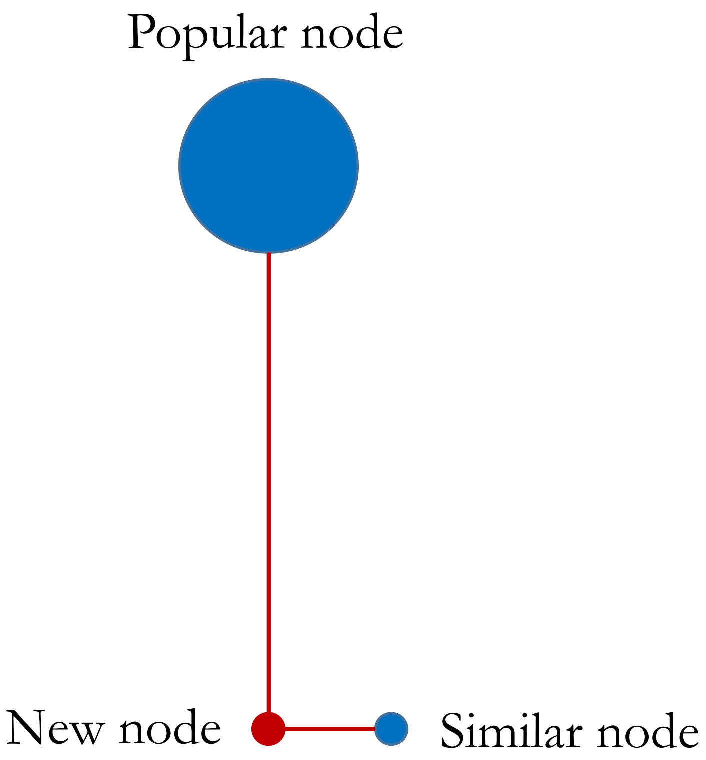


Popularity versus Similarity

Intuition

How does a new node make connections?

- It connects to popular nodes
  - Preferential Attachment
- It connects to similar nodes
  - "Birds of feather flock together"
  - Homophily



Key idea: new connections are formed by trade-off between popularity and similarity

Popularity-Similarity Model

In a growing network:

- The popularity of node $t = 1, 2, \ldots$ is modeled by its birth time
- The similarity of t is modeled by θ_t distributed over a "similarity space" $\mathbb{S}^1$
    - The angular distance θ_{st} quantifies the similarity between s and t.

$$\theta_{st} = \pi - |\pi - |\theta_s - \theta_t||$$

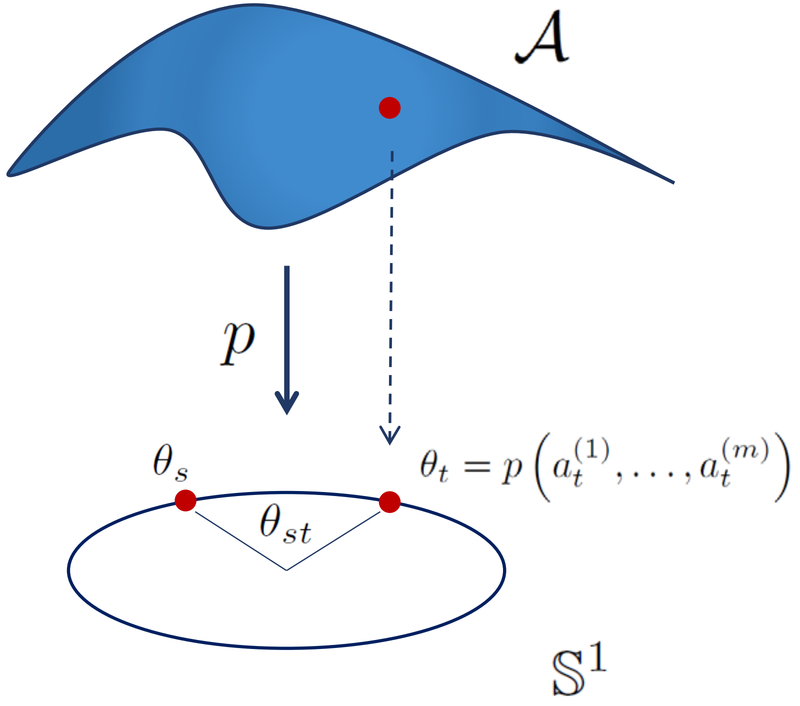
Mechanism:

a new node t connects to an existing node $s < t$

if s is both popular and similar to t, that is if:

$$\boxed{s^{\beta}\theta_{st} \text{ is small}}$$

β controls the relative contributions of popularity and similarity



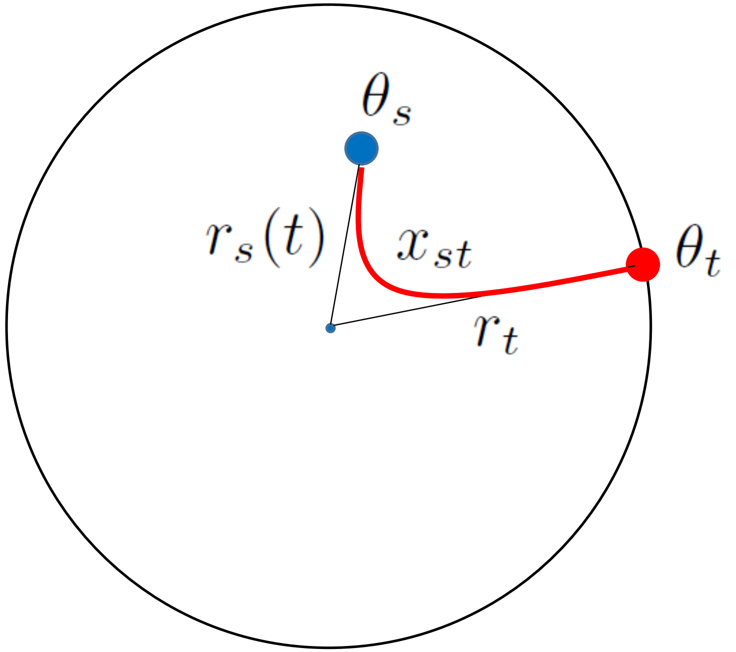
Complex Network in a Hyperbolic Plane

- The angular coordinate of node s is its similarity θ_s
- The radial coordinate of node s is $r_s = \ln s$
- Let it grow with time: $r_s(t) = \beta r_s + (1-\beta) r_t$
- Let x_{st} be the hyperbolic distance between s and t

$$x_{st} = r_s(t) + r_t + \ln \frac{\theta_{st}}{2} = \ln \left(\frac{t^{2-\beta} s^{\beta} \theta_{st}}{2} \right)$$

- Minimization of $s^{\beta}\theta_{st} \Leftrightarrow$ minimization of x_{st}
- New nodes connect to

  hyperbolically closets existing nodes



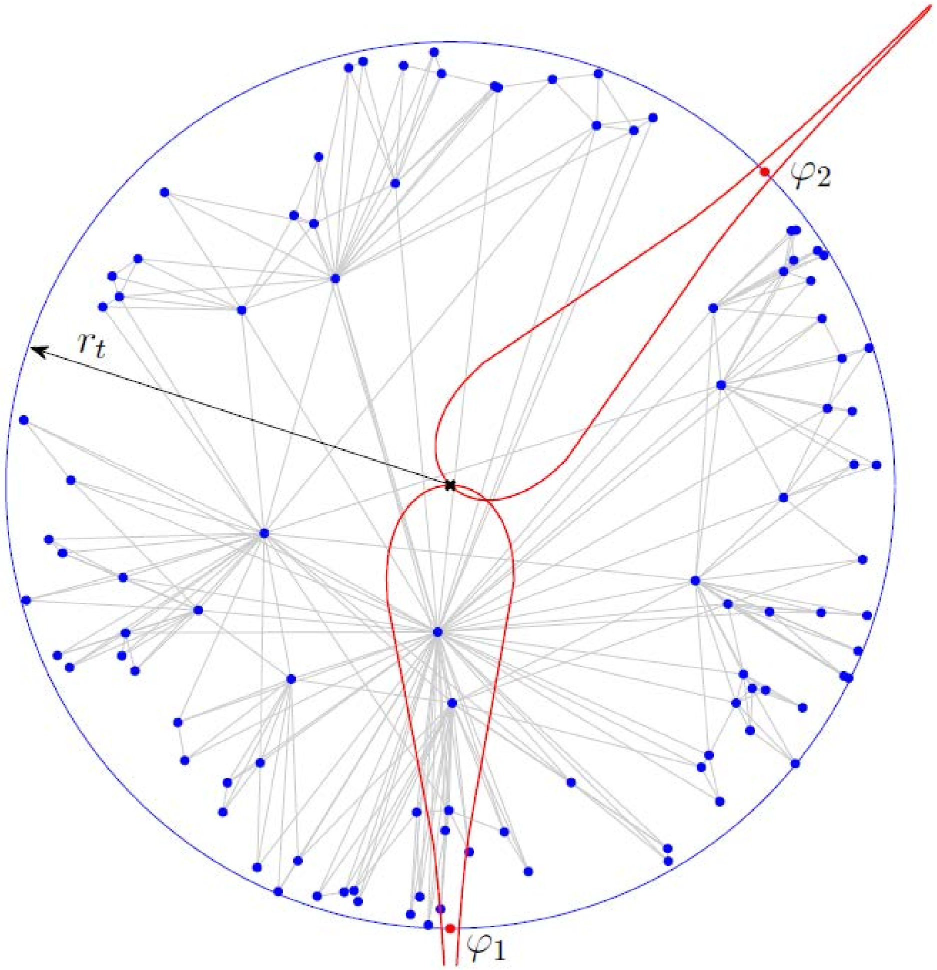
Hyperbolic Disk

We call this mechanism: Geometric Preferential Attachment

Geometric Preferential Attachment

How does a new node t find its position θ_t in the similarity space $\mathbb{S}^1$?

- Fashion: $\mathbb{S}^1$ contains "hot" regions.
- Attractiveness of $\varphi \in \mathbb{S}^1$ for a new node t is the number of existing nodes in $D_\varphi(r_t)$
- The higher the attractiveness of φ, the higher the probability that $\theta_t = \varphi$



GPA Model of Growing Networks

1. Initially the network is empty. New nodes $t = 1, \ldots$ appear one at a time.
2. The angular (similarity) coordinate θ_t of t is determined as follows:
   a. Sample $\varphi_i \sim U[0, 2\pi]$, $i = 1, \ldots, t$ uniformly at random (candidate positions)
   b. Compute the attractiveness $A_t(\varphi_i)$ for all candidates
   c. Set $\theta_t = \varphi_i$ with probability

$$\mathbb{P}(\theta_t = \varphi_i) \propto A_t(\varphi_i) + \Lambda$$

$\Lambda \geqslant 0$ is initial attractiveness

3. The radial (popularity) coordinate of node t is set to $r_t = \ln t$
   The radial coordinates of existing nodes $s < t$ are updated to

$$r_s(t) = \beta r_s + (1 - \beta) r_t$$

$\beta \in [0, 1]$ models popularity fading

4. Node t connects to m hyperbolically closet existing nodes.

GPA as a Model for Real Networks

- GPA is the first model that generates networks with
  - Power law degree distribution
  - Strong clustering
  - Community structure

$\Big\}$ "Universal" properties of complex networks

- Can we estimate the model parameters from the network data?

- The model has three parameters:
  - m the number of links established by every new node $\bar{k} = 2m$
  - β the speed of popularity fading $\gamma = 1 + 1/\beta$
  - Λ the initial attractiveness MLE

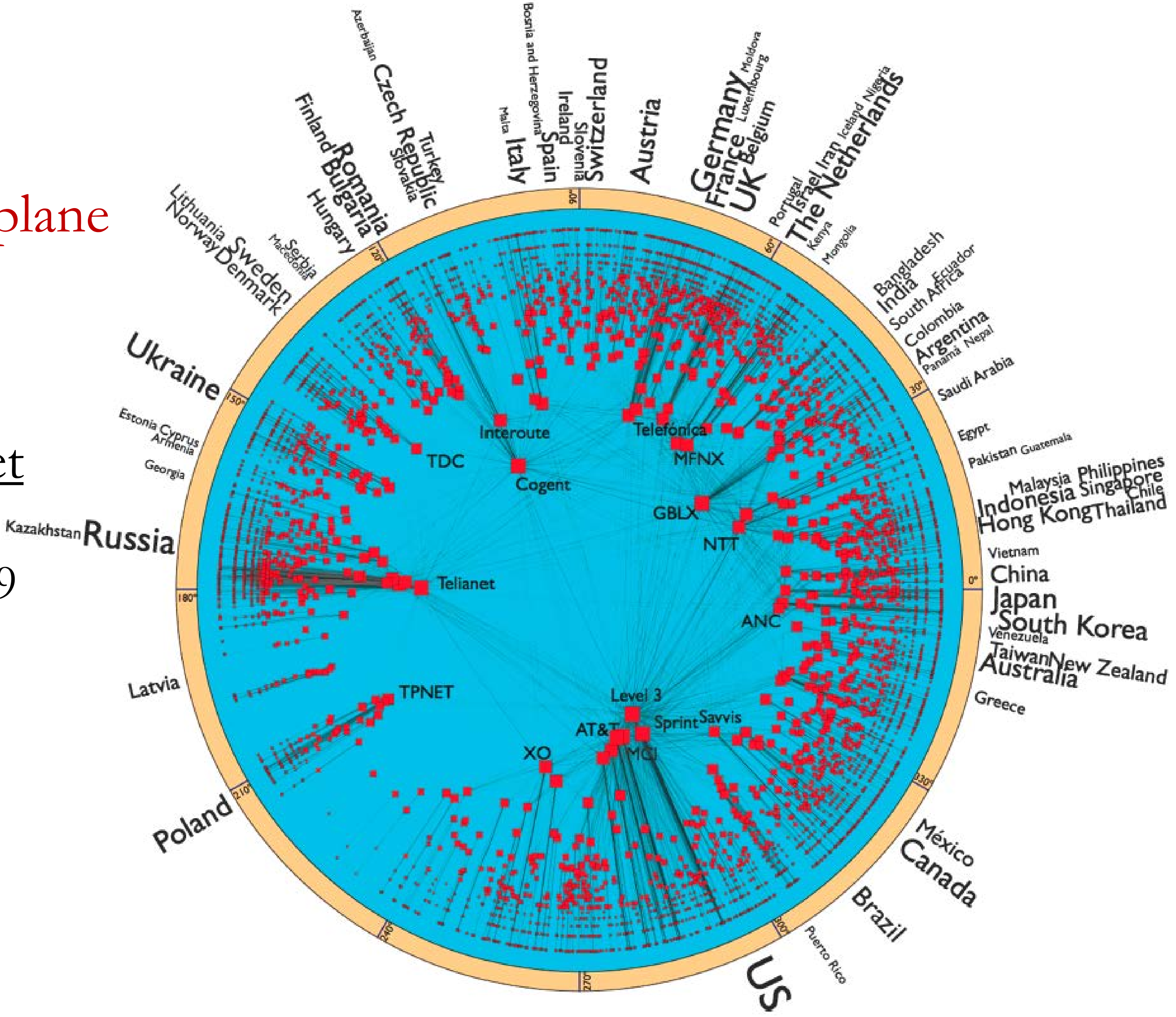
Hyper Map

Embed G into the hyperbolic plane

$$G \longmapsto \{(r_i, \theta_i)\}$$

<u>Hyperbolic Atlas of the Internet</u>

- Data: CAIDA
- Internet topology as of Dec 2009
- Nodes are autonomous systems
- Two ASs are connected if they exchange traffic
- Node size $\propto \log k$
- Font size $\propto \log \#\text{ASs}$



References

- General text on Complex Networks
  - M. Newman *Networks: An Introduction* 2009 aka **Big Black Book**
- Hyper Map
  - F. Papadopoulos et al "Network Mapping by Replaying Hyperbolic Growth" *IEEE/ACM Transactions on Networking*, 2015 (first arXiv version 2012)
- Geometric Preferential Attachment
  - K. Zuev et al "Emergence of Soft Communities from Geometric Preferential Attachment" *Nature Scientific Reports 2015.*