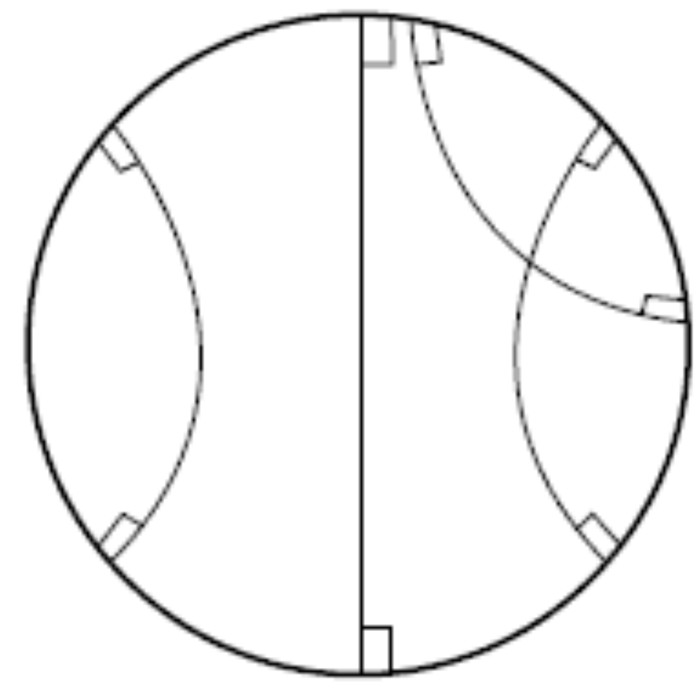
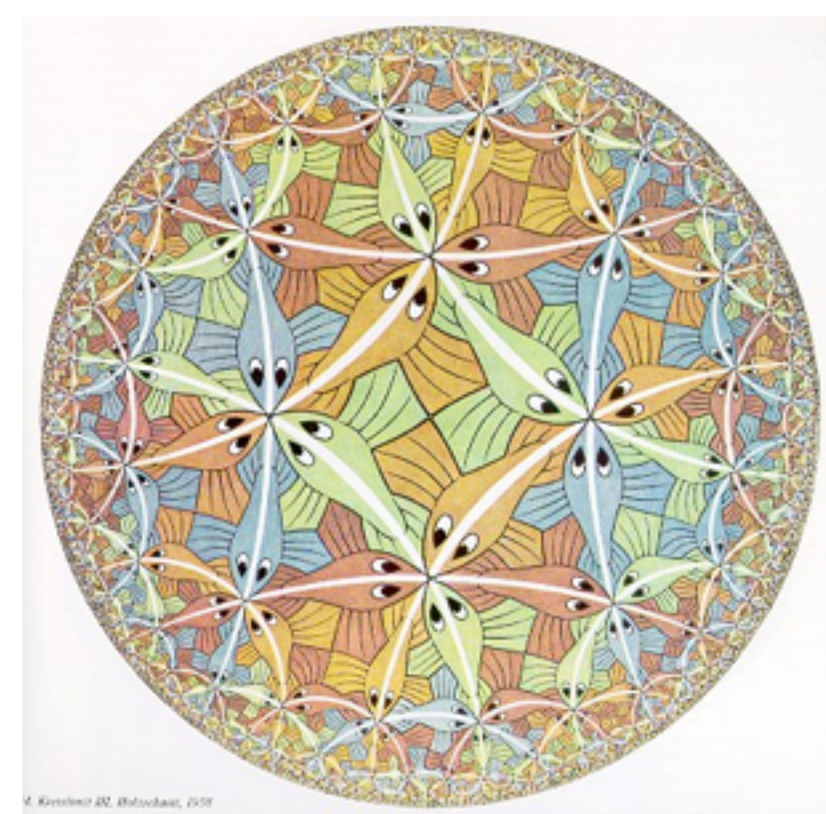
Network Curvature, friendship paradox and dispersion

Rik Sarkar

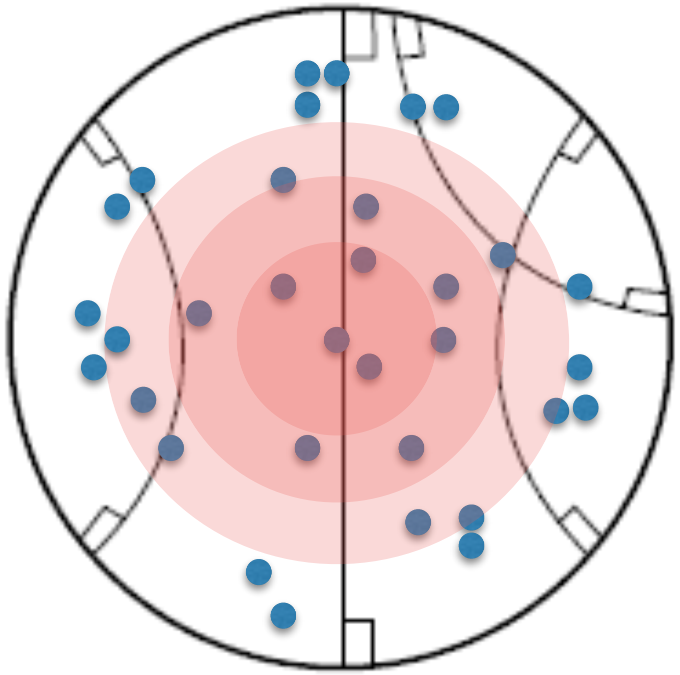
Recap: Hyperbolic distances





- Points in a disk
  - Shortest paths along circular curves bent toward the center
  - Similar to internet paths being bent toward the core
  - Distances look cramped close to the boundaries

Internet emulates hyperbolic metrics



- Shavitt, Tankel. ACM ToN 2008.

Hyperbolic model for networks

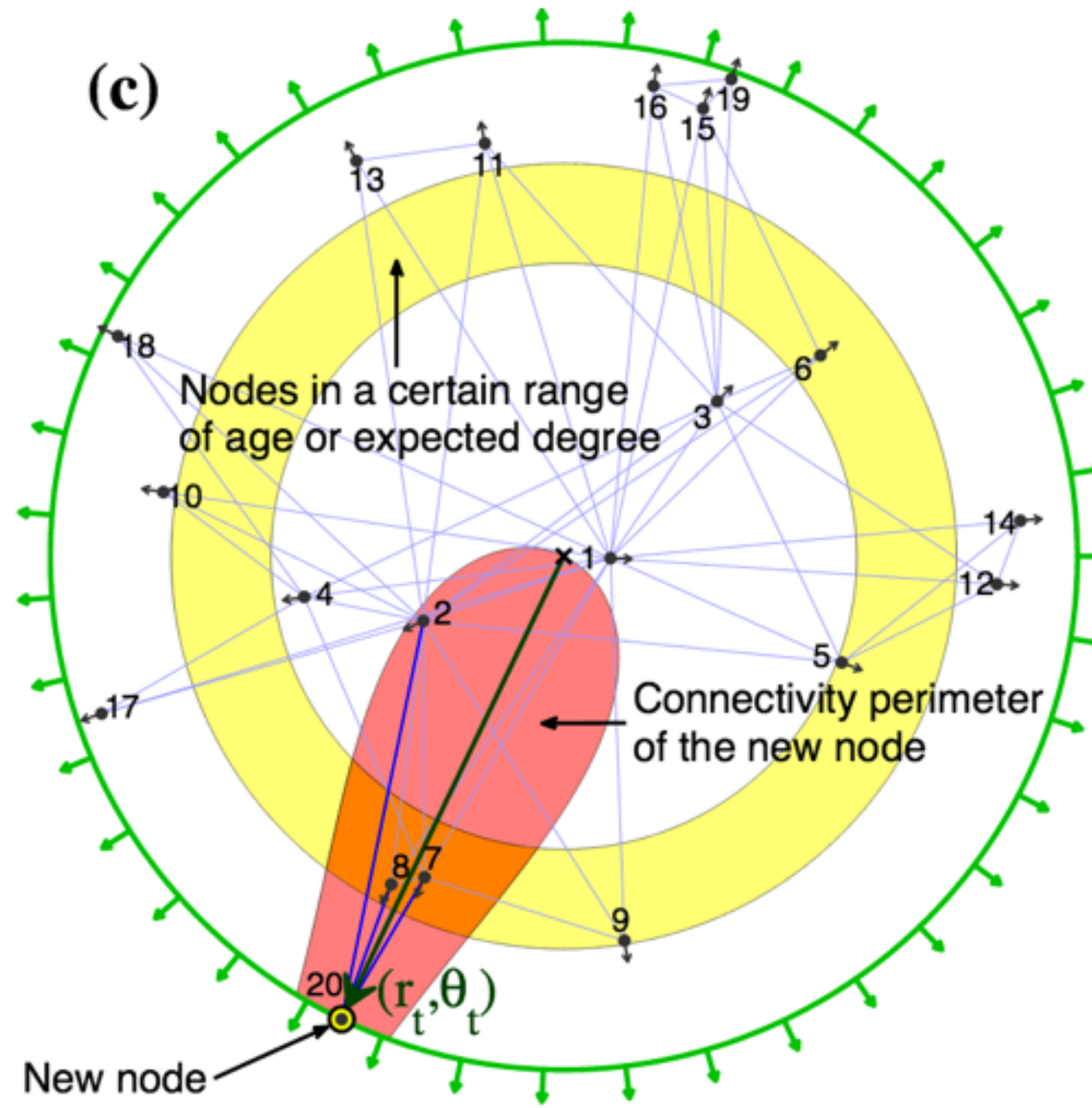
- People connect to popular “central” nodes
  - Preferential attachment. Hubs. Cause small diameters.
- People connect to other “similar” nodes
  - Similar in location, or interests, or communities
  - Similar means small distance in some measure
- Preferential attachment does not model this well
  - Cannot model the clustering properties

Popularity/similarity model

- Put all nodes on the plane at polar coord: (r, θ)
- Popularity: Distance from the center
  - Like preferential attachment, earlier nodes are popular
  - If a node appears at time t, its distance from center is $r = \ln t$
- Interests/features for similarity: Represented by angle
  - θ
  - Two nodes a,b are similar if $|\theta_a - \theta_b|$ is small.

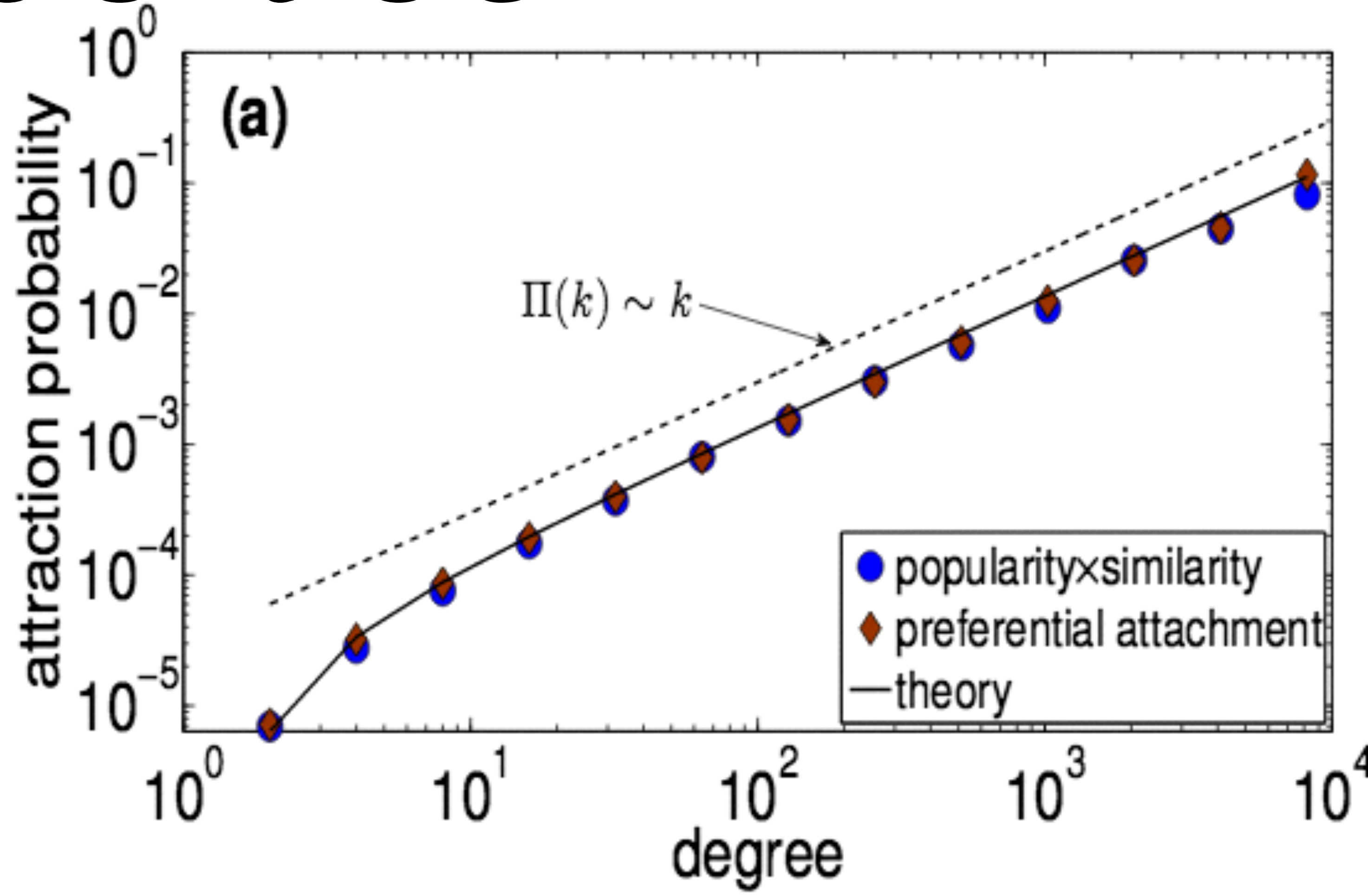
Edge attachments

- A new node appears at time t
  - Sets r = ln t
  - Sets θ = random
  - It connects to the k nearest nodes in hyperbolic distance
- Central nodes are older and higher degree



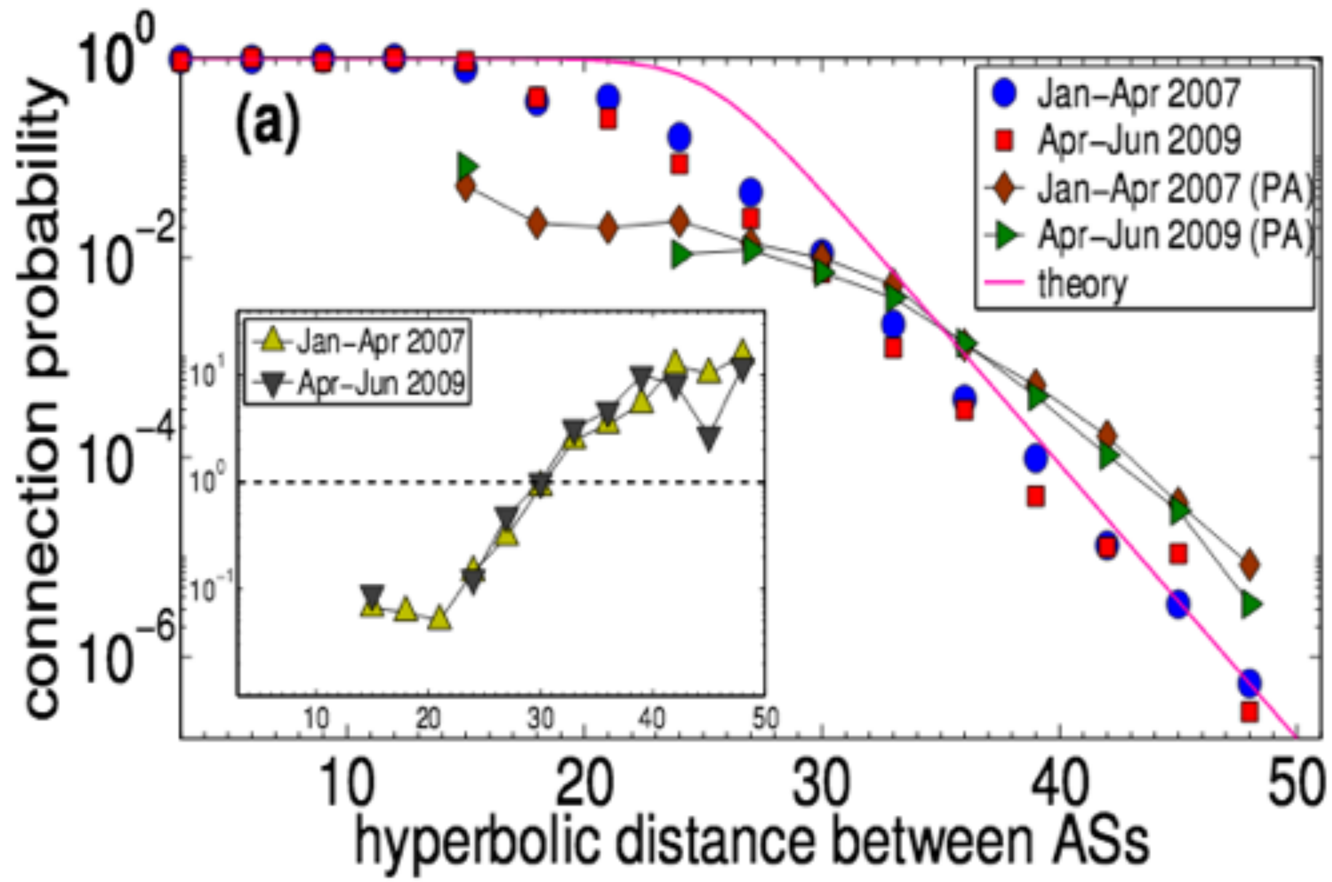
Properties

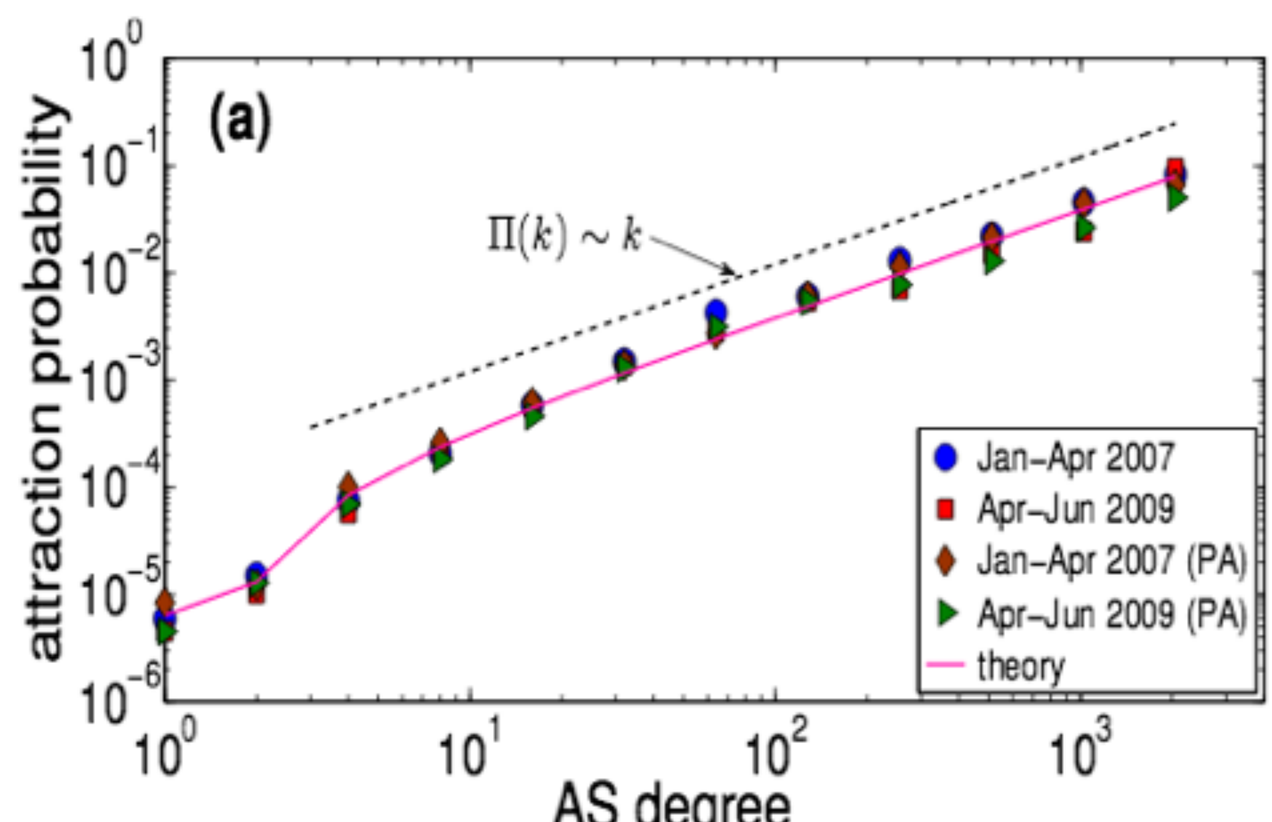
- Creates power law distribution
- Creates strong clustering
  - Different from pref. attachment
  - More realistic in real networks



Modeling the internet

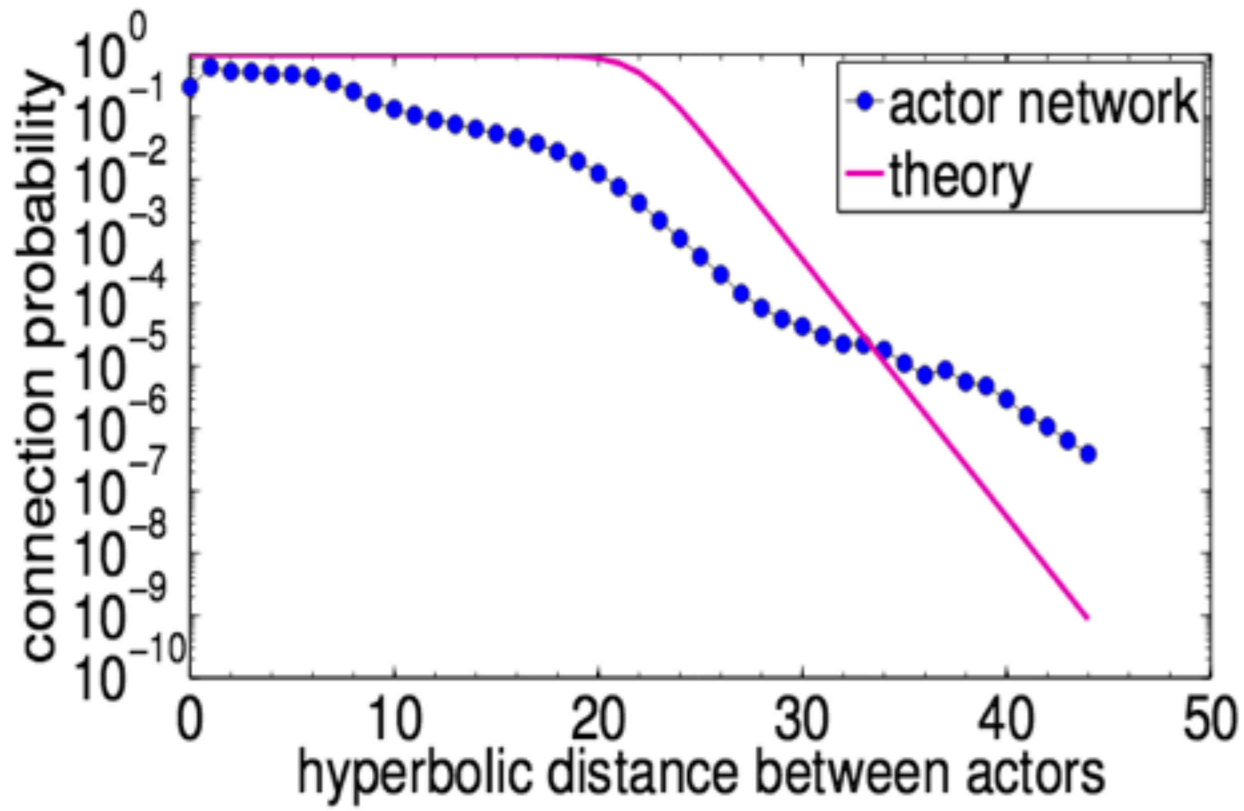
- A suitable hyperbolic embedding gives very good model of connection probabilities
- Similar results in other power law networks





Actor networks

- Does *not* work equally well



- Popularity vs Similarity in Growing Networks
- Papdopoulos et al. Nature 2012.

Hyperbolic geometry

- Useful in modeling metrics with exponential growth (number of nodes within distance x)
  - E.g. balanced binary tree
  - Many parameters may have such properties
    - Position in a hierarchy
    - Topological types of paths in a domain
    - Subsets of items

Few other things

Friendship paradox

- Your friends have more friends than you do!
- Are you less social than others?

Friendship paradox

- The paradox:
  - If you ask everyone to report their degrees, you get the average degree
  - If you ask everyone to report the average degrees of their friends and take the averages of all,
    - you get more than the overall average degree!
- Most of us have some popular friends (hence they are popular)
  - If you pick a random friend of a random person, (random edge)
    - This friend is relatively likely to be popular, since popular nodes have more edges

Friendship paradox

- Average degree of nodes:
  - A node with degree d(v) contributes d(v) once
- Average degree of a friend:
  - Each person picks a friend and counts degree
  - A node with degree d(v) contributes d(v) times, with total contribution $d(v)^2$
  - A few nodes with relatively high d(v) can skew the count
  - https://en.wikipedia.org/wiki/Friendship_paradox
  - S. L. Feld, Why your friends have more friends than you do, American journal of sociology, 1991

Identify spouses or romantic partners

- Suppose you have the facebook graph
- Only the graph and nothing else
- Can you identify which edges correspond to spouses or romantic partners?

Identify spouses or romantic partners

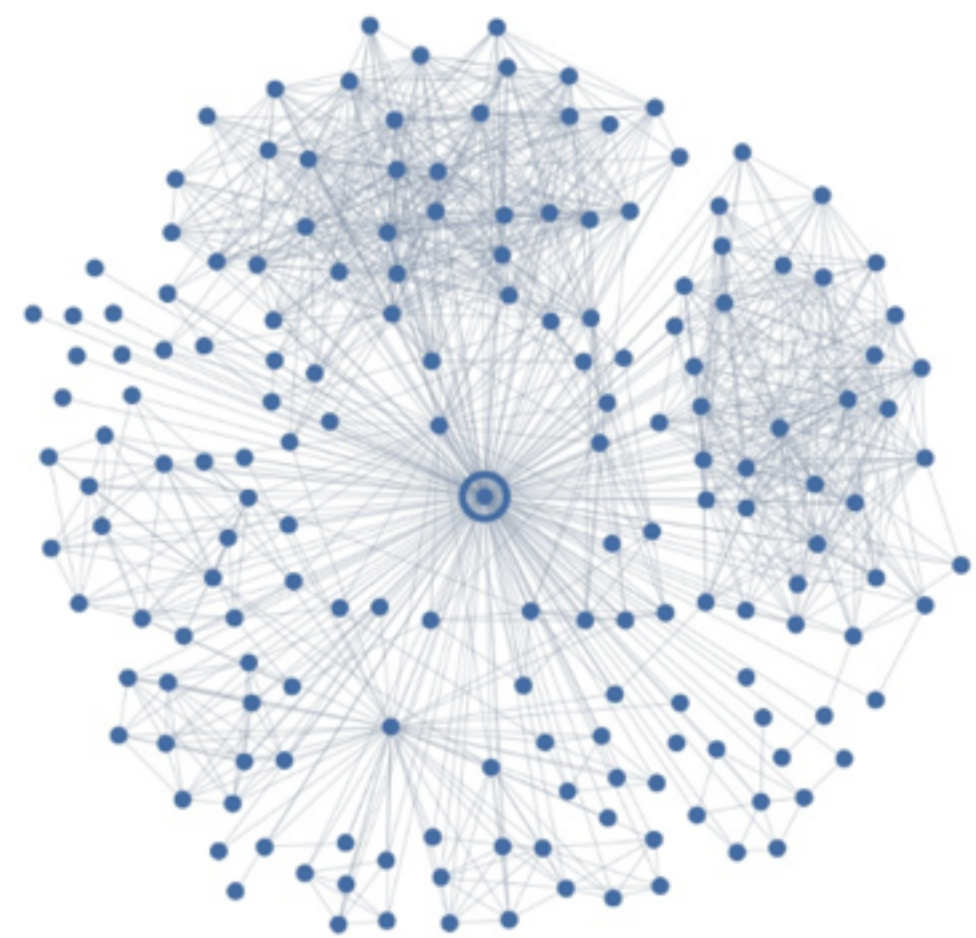
Identify spouses or romantic partners

- Tie strengths are important
- Romantic ties tend to be of high strength, more likely to transmit information
- Do you expect romantic links to have high embeddedness (number/fraction of common friends)?

- People have clusters of friend circles

- Work, school, college, hobbies

- Edges in these have high embeddedness, even if they are not strong friends



- Spouses usually know some friends in each-others different circles
  - The edge does not have high embeddedness
    - Compared to links in groups such as school/college
  - But, it has a dispersed structure:
    - There are several mutual friends, but the mutual friends are not well connected among themselves

Dispersion

- dispersion between u,v
- Notations:
    - C(u,v): Common friends of u, v
    - G_u : Subgraph induced by u and all neighbors of u
    - d_{uv} : distance measured in G_u-{u,v}: Without using u or v

$$disp(u, v) = \sum_{s,t \in C(u,v)} d_{uv}(s, t)$$

Dispersion

$$disp(u,v) = \sum_{s,t \in C(u,v)} d_{uv}(s,t)$$

- Increases with more mutual friends
- Increases when these friends are far in the graph
- It is possible to use other distance measures
- Good results with d = 1 if no direct edge, 0 otherwise

Normalized dispersion

- Use norm(u,v) = disp(u,v)/embed(u,v)
  - 48% accuracy
- Apply recursively, to weigh higher nodes with high dispersion
  - Gives 50.5% accuracy
  - 60% accuracy for married couples
- High accuracy considering hundreds of friends
- Works better than usual machine learning based on posts, visits, photos etc features
  - Best results with combination of features

- Backstrom and Kleinberg. Romantic partnerships and dispersion of social ties, ACM CSCW 2014