Popularity versus Similarity in Growing Networks

Fragiskos Papadopoulos
Cyprus University of Technology
M. Kitsak, M. Á. Serrano, M. Boguñá, and
Dmitri Krioukov
CAIDA/UCSD

APS meeting, March 2012

Preferential Attachment (PA)

- *Popularity is attractive*
- If new connections in a growing network prefer popular (high-degree) nodes, then the network has a power-law distribution of node degrees

Issues with PA

- Zero clustering
- PA *per se* is ***impossible*** in real networks
  - It requires global knowledge of the network structure to be implemented
- The popularity preference should be exactly a linear function of the node degree
  - Otherwise, no power laws

No model that would:

- Be simple and universal (like PA)
  - Potentially describing (as a base line) evolution of many different networks
- Yield graphs with observable properties
  - Power laws, strong clustering, to start with
  - But many other properties as well
- Not require any global intelligence
- Be ***validated***

Validation of growth mechanism

- State of the art
  - Here is my new model
  - The graphs that it produces have power laws!
- Almost never the growth mechanism is validated ***directly***
- PA *was* validated directly for many networks, because it is so simple

Paradox with PA validation

- Dilemma
  - PA was validated
  - But PA is impossible
- Possible resolution
  - PA is an emergent phenomenon
  - A consequence of some other underlying processes

Popularity versus Similarity

- Intuition
  - I (new node) connect to you (existing node) not only if you are popular (like Google or Facebook), but also if you are similar to me (like Tartini or free soloing) — homophily
- Mechanism
  - New connections are formed by trade-off optimization between popularity and similarity

Mechanism (growth algorithm)

- Nodes t are introduced one by one
  - t @ 1, 2, 3, …
- Measure of popularity
  - Node's birth time t
- Measure of similarity
  - Upon its birth, node t gets positioned at a random coordinate θ_t in a "similarity" space
  - The similarity space is a circle
  - θ is random variable uniformly distributed on $[0,2\pi]$
  - Measure of similarity between t and s is θ_{st} @ $|\theta_s$ *O* $\theta_t|$

Mechanism (contd.)

- New connections
  - New node t connects to m existing nodes s, s ? t, minimizing $s\theta_{st}$
  - That is, maximizing the product between popularity and similarity

New node t connects to m existing nodes s that minimize

$$s\theta_{st}$$

$$st\frac{\theta_{st}}{2}$$

$$\ln\left(st\frac{\theta_{st}}{2}\right)$$

$$= r_s + r_t + \ln\frac{\theta_{st}}{2}$$

$\approx x_{st}$ — the *hyperbolic* distance between s and t

New nodes connects to m hyperbolically closest nodes

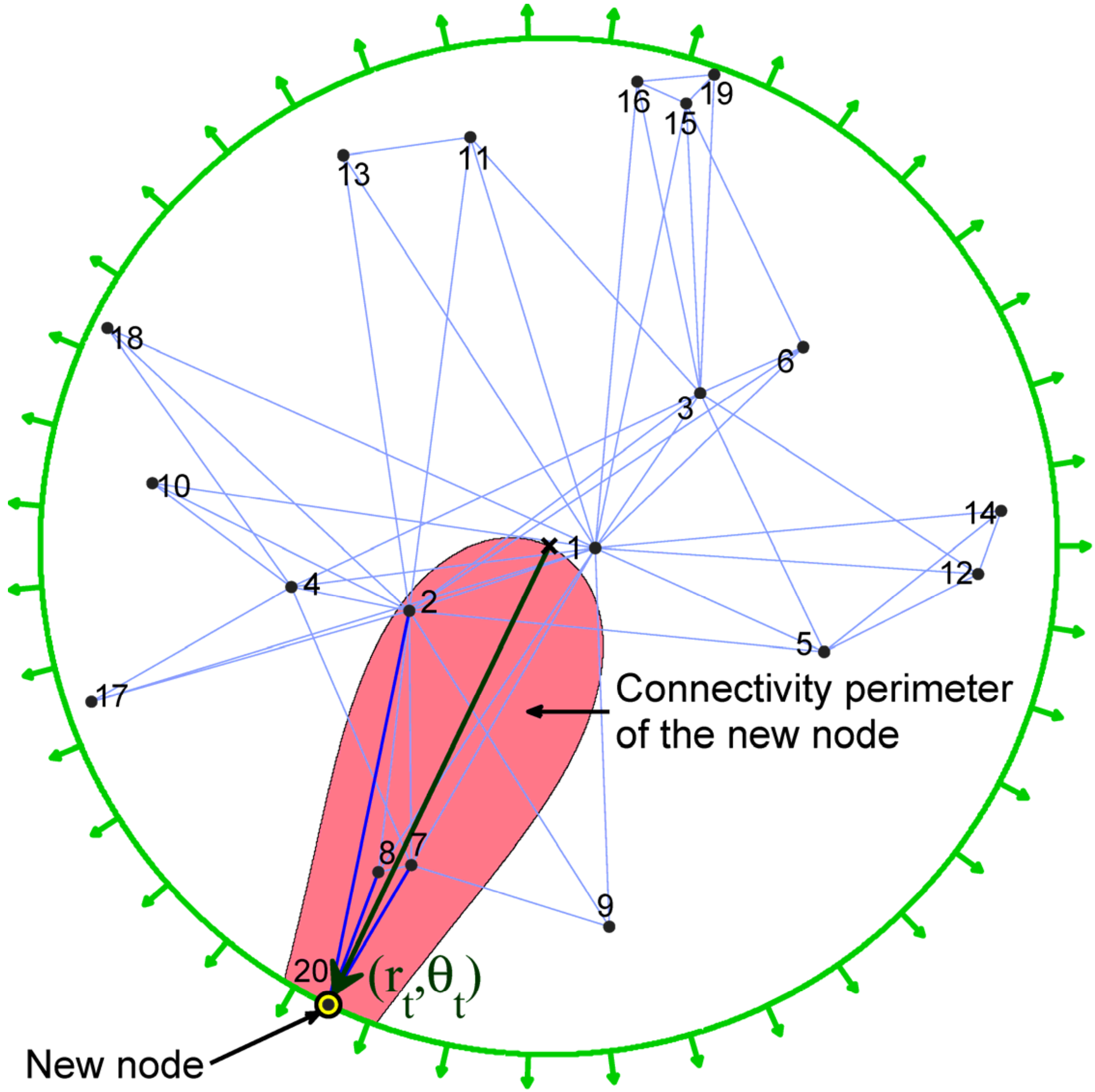
The expected distance to the *m*'th closest node from *t* is

$$R_t = r_t - \ln\frac{2r_t}{\pi m}$$ — average degree is fixed to $2m$

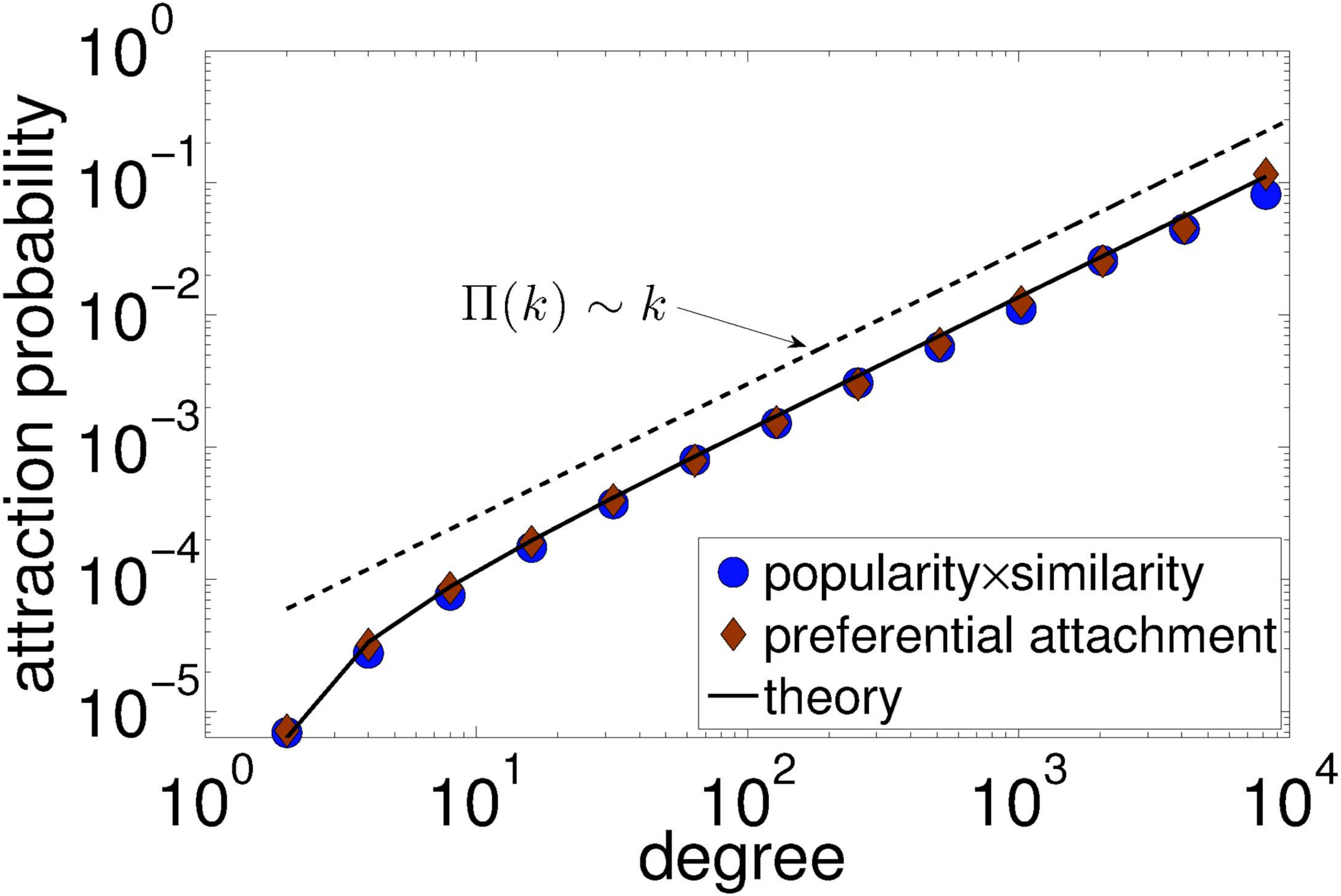
$$R_t = r_t$$ — average degree grows logarithmically with *t*

if ј ¬ 2

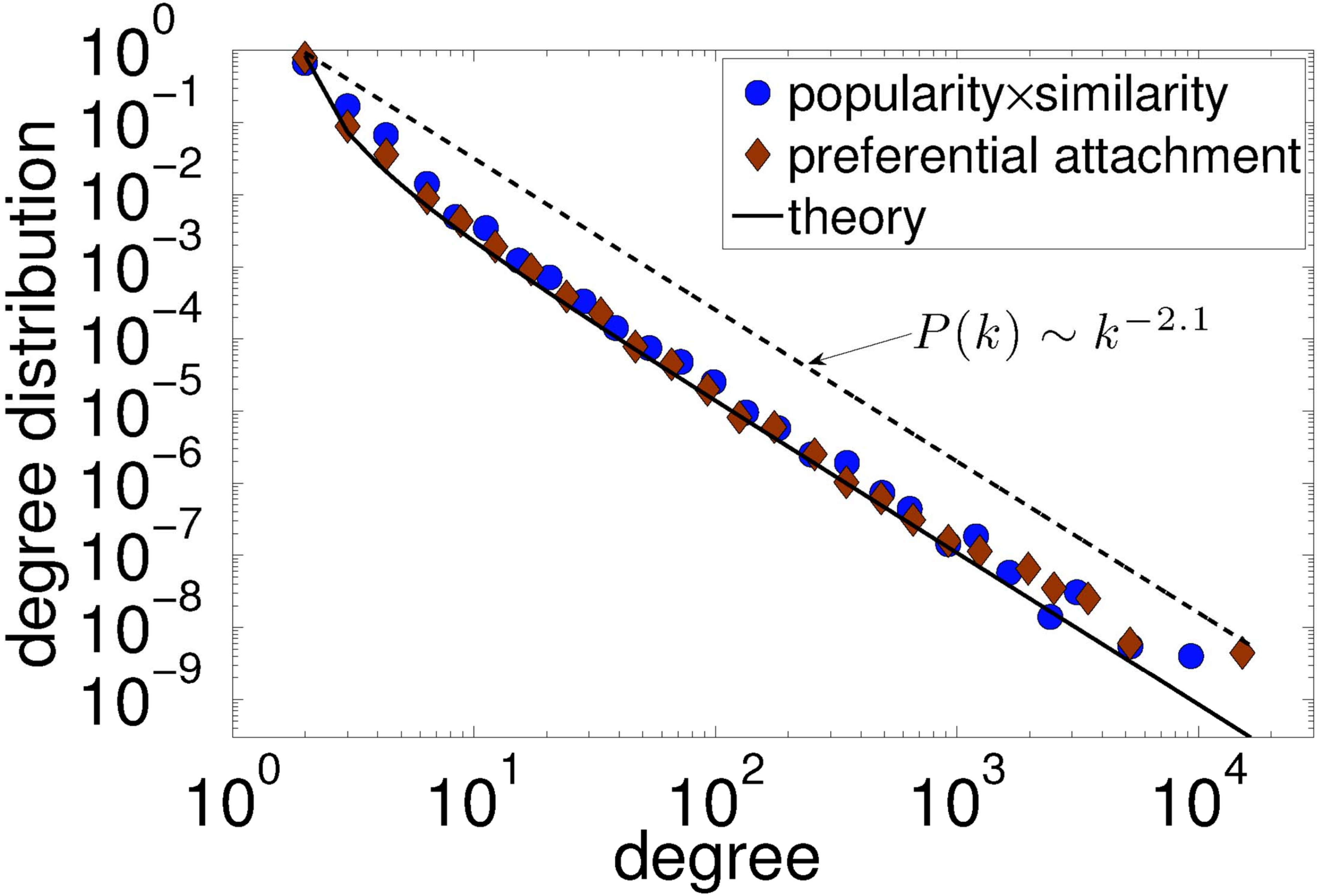
New node *t* is located at radial coordinate $r_t \sim \ln t$,
and connects to all nodes within distance $R_t \sim r_t$



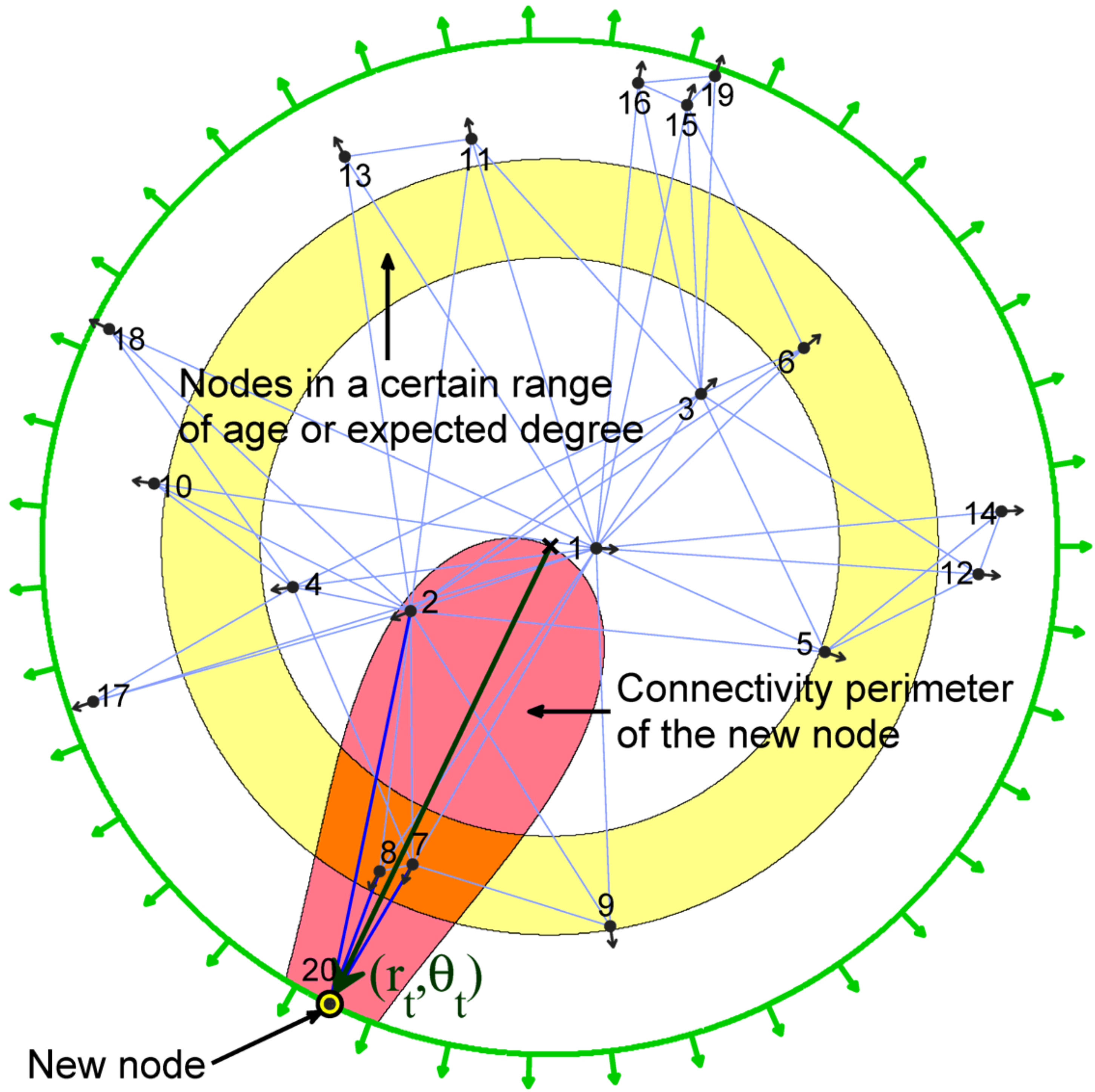
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
(r_t,θ_t)
Connectivity perimeter
of the new node
New node



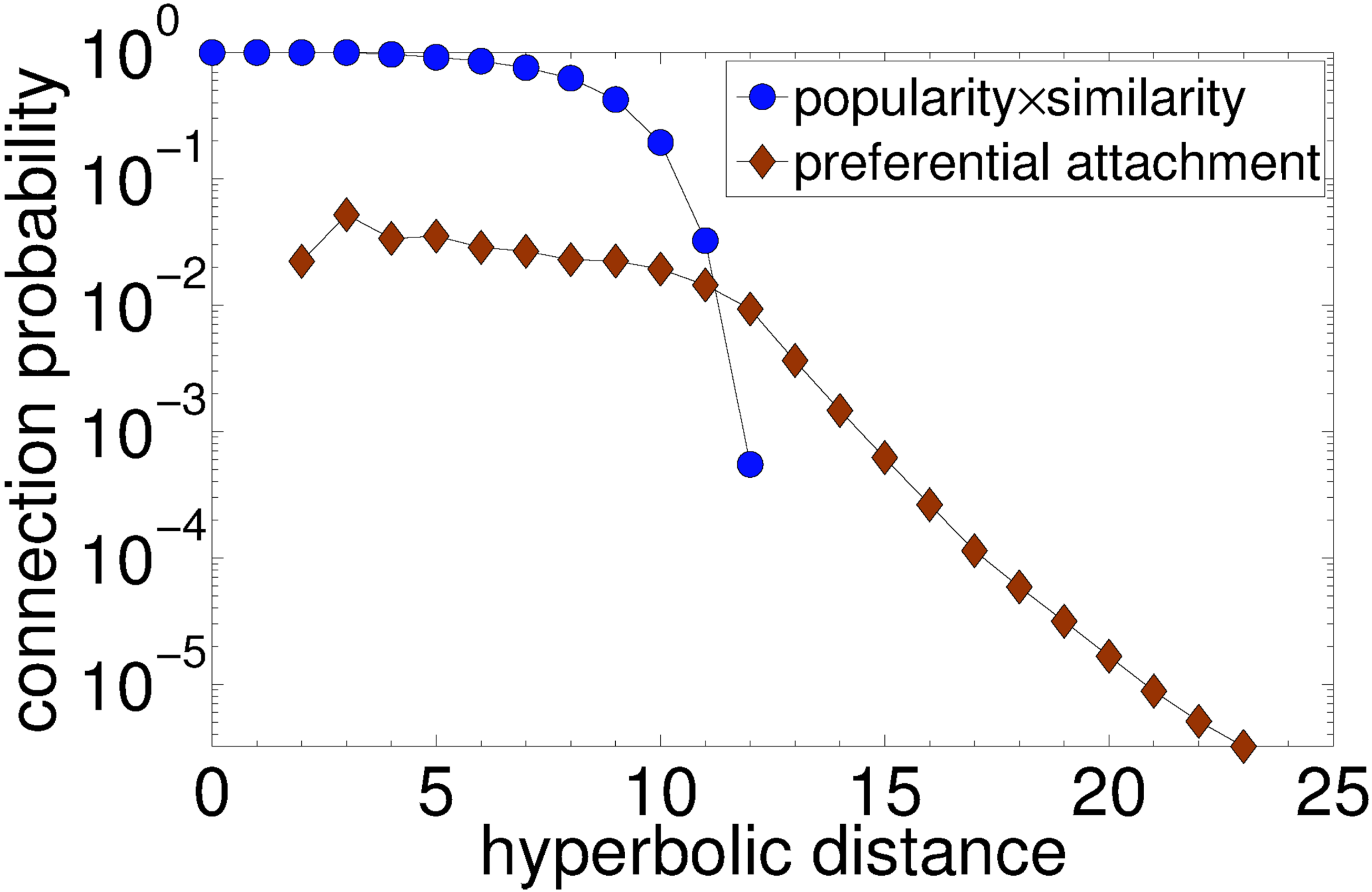
attraction probability
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
$\Pi(k) \sim k$
popularity×similarity
preferential attachment
theory
10^0
10^1
10^2
10^3
10^4
degree



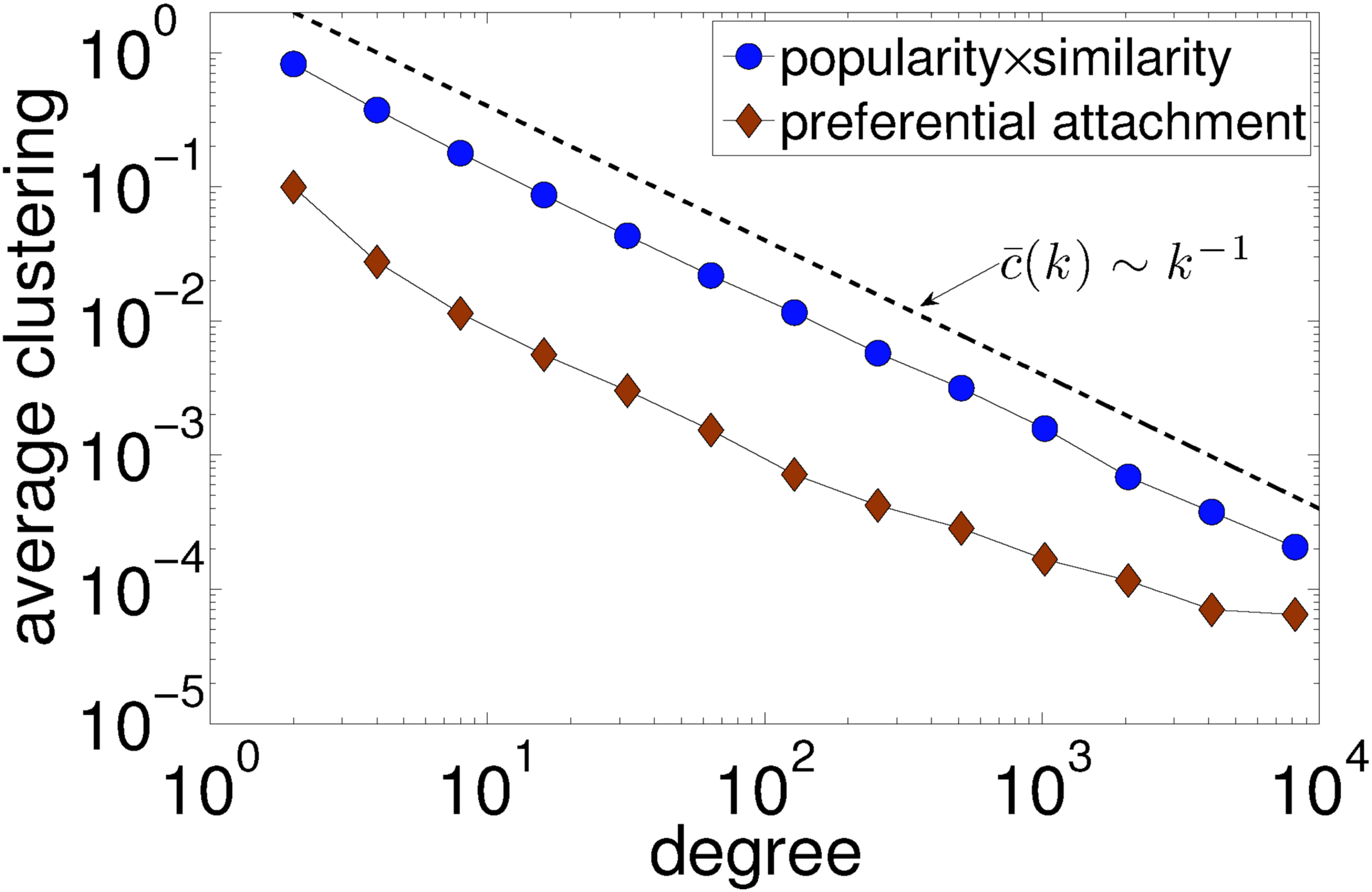
degree distribution
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
popularity×similarity
preferential attachment
theory
$P(k) \sim k^{-2.1}$
10^0
10^1
10^2
10^3
10^4
degree



Nodes in a certain range
of age or expected degree
Connectivity perimeter
of the new node
New node
(r_t,θ_t)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20



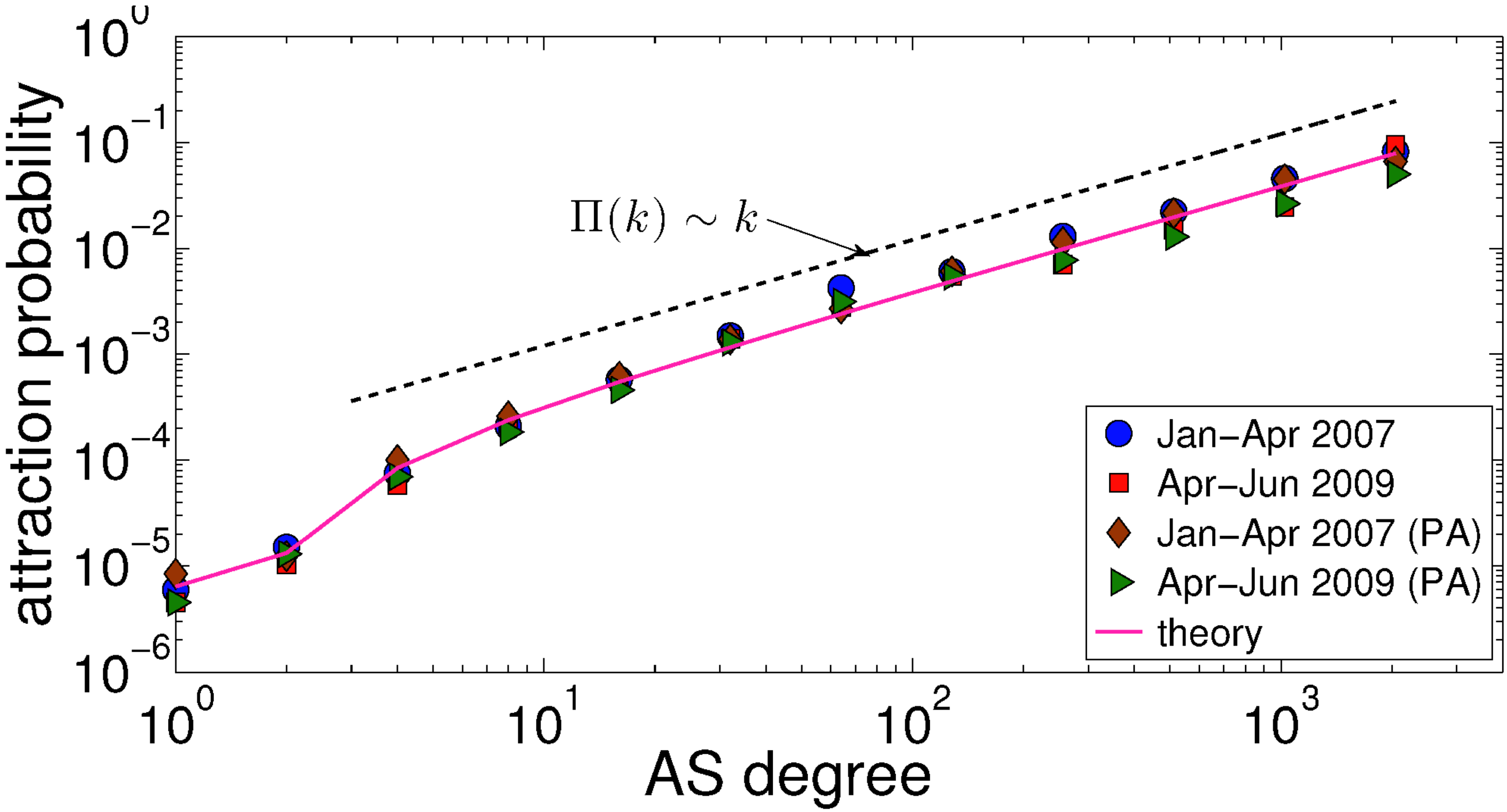
popularity×similarity
preferential attachment
connection probability
hyperbolic distance
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
0
5
10
15
20
25



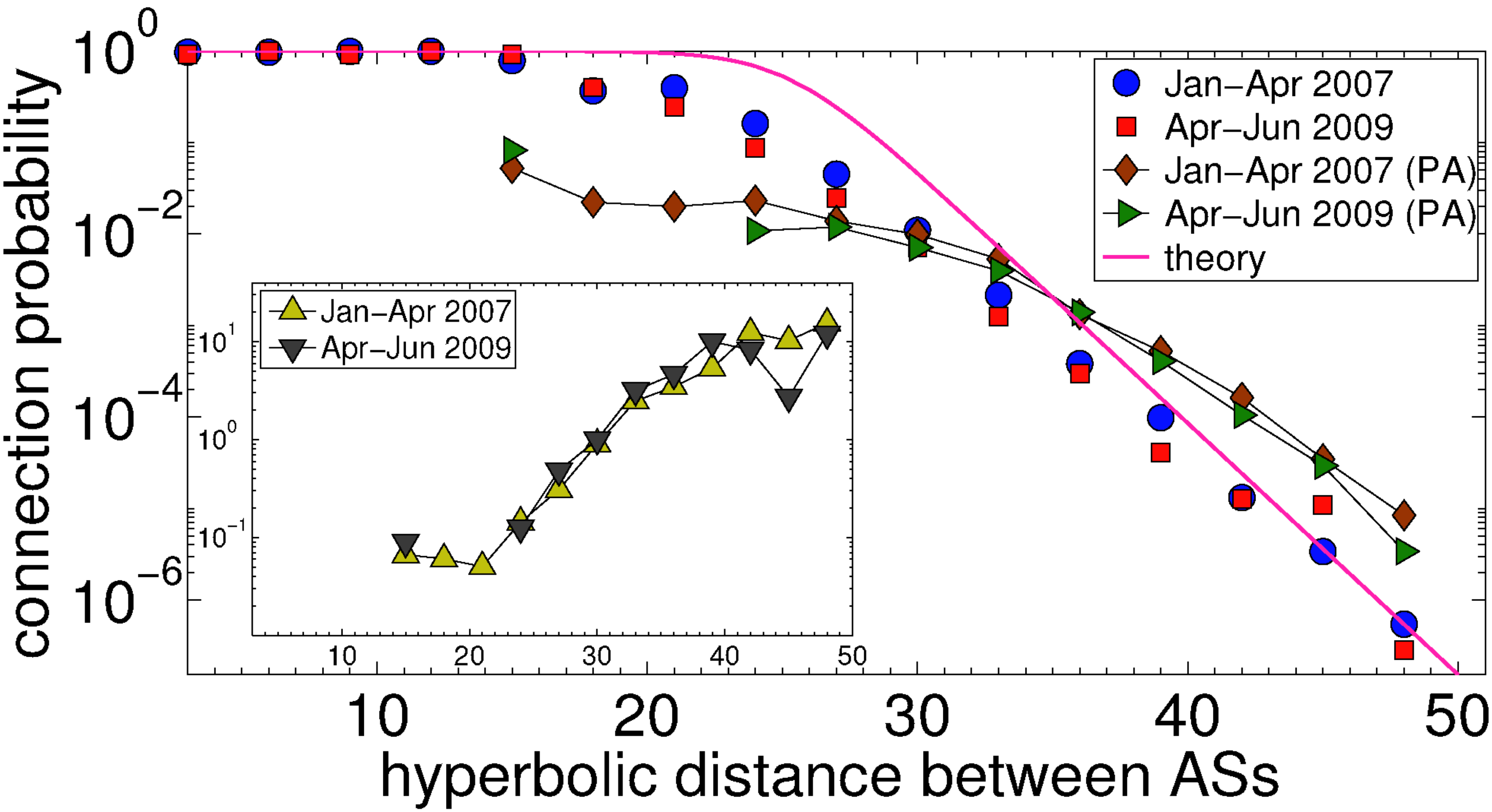
popularity×similarity
preferential attachment
$\bar{c}(k) \sim k^{-1}$
average clustering
degree
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^0
10^1
10^2
10^3
10^4

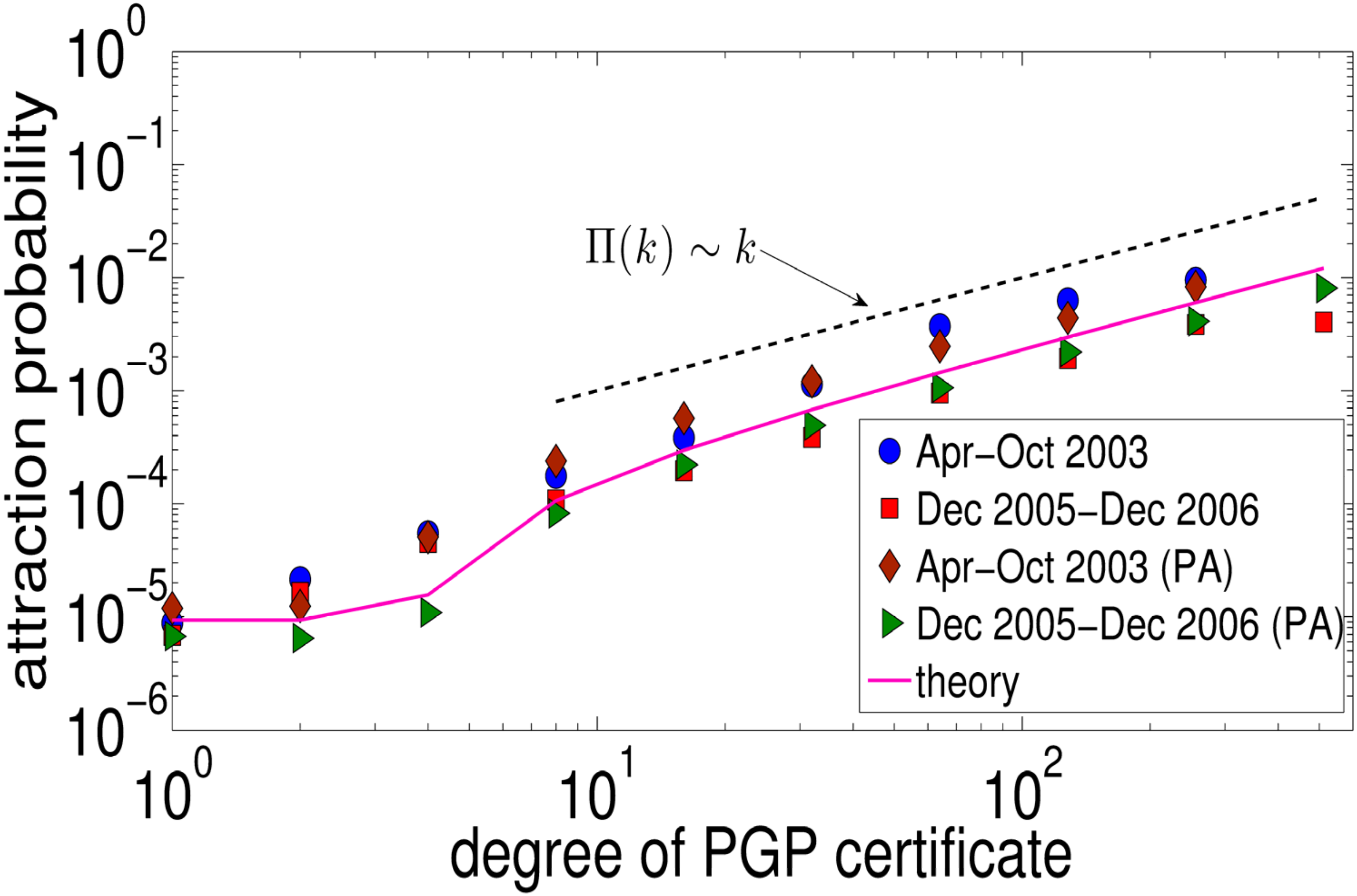
Validation

- Take a series of historical snapshots of a real network
- Infer angular/similarity coordinates for each node
- Test if the probability of new connections follows the model theoretical prediction

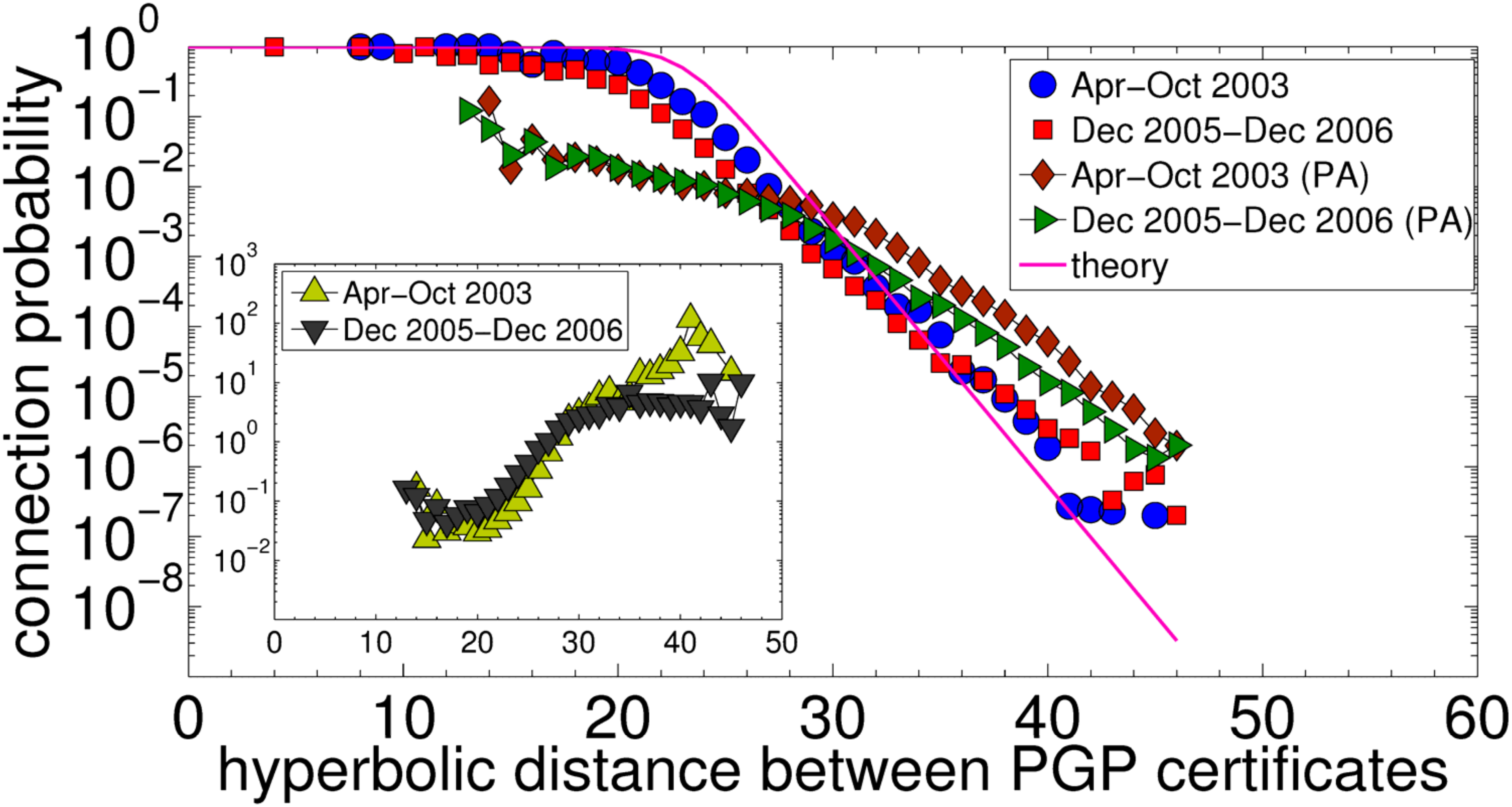


attraction probability
AS degree
$\Pi(k) \sim k$
Jan−Apr 2007
Apr−Jun 2009
Jan−Apr 2007 (PA)
Apr−Jun 2009 (PA)
theory
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{0}
10^{1}
10^{2}
10^{3}

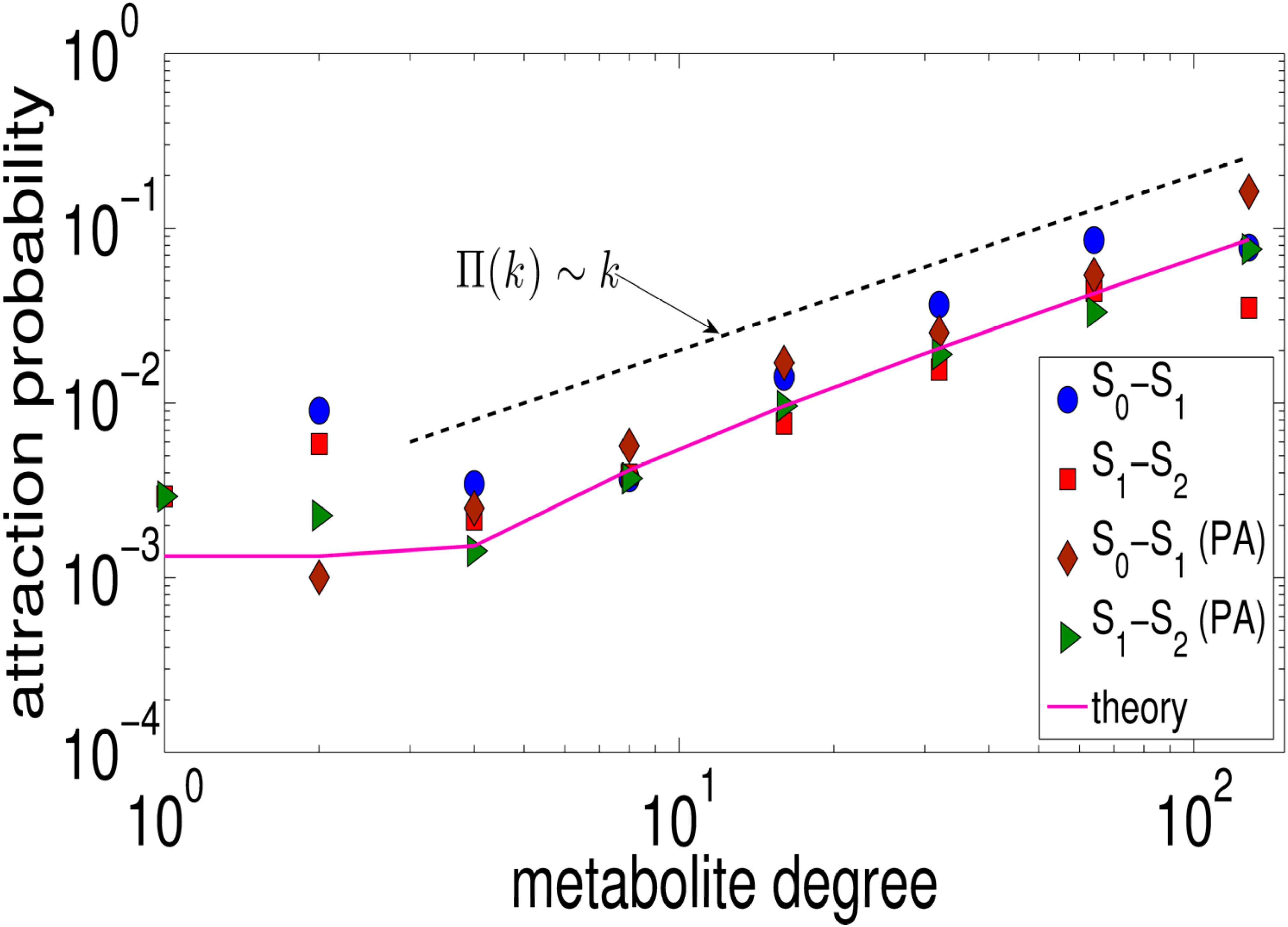

connection probability
hyperbolic distance between ASs
Jan−Apr 2007
Apr−Jun 2009
Jan−Apr 2007 (PA)
Apr−Jun 2009 (PA)
theory
Jan−Apr 2007
Apr−Jun 2009



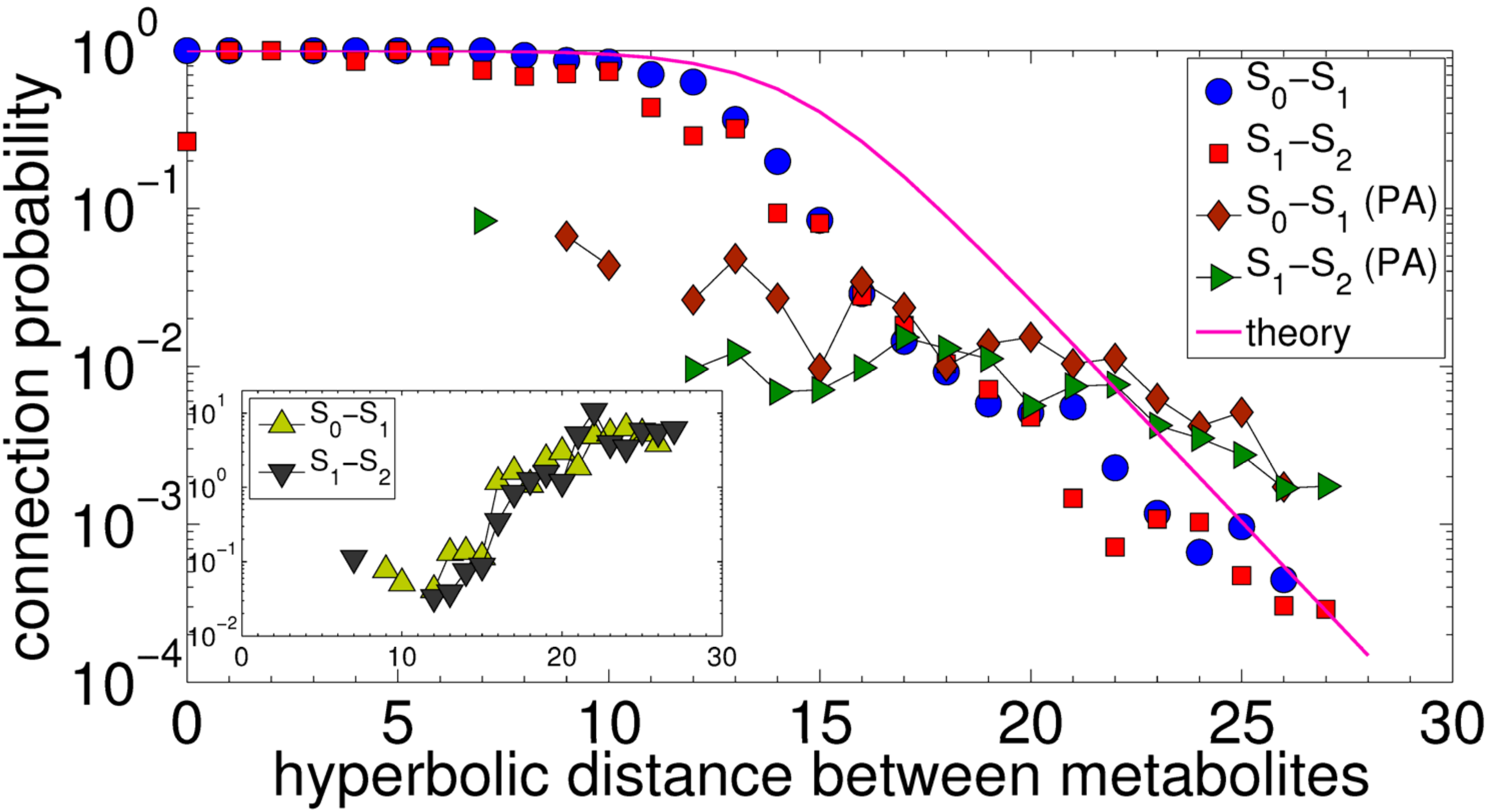
attraction probability
degree of PGP certificate
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^0
10^1
10^2
$\Pi(k) \sim k$
Apr−Oct 2003
Dec 2005−Dec 2006
Apr−Oct 2003 (PA)
Dec 2005−Dec 2006 (PA)
theory



connection probability
hyperbolic distance between PGP certificates
Apr−Oct 2003
Dec 2005−Dec 2006
Apr−Oct 2003 (PA)
Dec 2005−Dec 2006 (PA)
theory
Apr−Oct 2003
Dec 2005−Dec 2006



attraction probability
metabolite degree
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^0
10^1
10^2
$\Pi(k) \sim k$
S_0–S_1
S_1–S_2
S_0–S_1 (PA)
S_1–S_2 (PA)
theory



connection probability
hyperbolic distance between metabolites
S_0-S_1
S_1-S_2
S_0-S_1 (PA)
S_1-S_2 (PA)
theory

Popularityfisimilarity optimization

- Explains PA as an emergent phenomenon
- Resolves all major issues with PA
- Generates graphs similar to real networks across many vital metrics
- *Directly* validates against some real networks
  - Technological (Internet)
  - Social (web of trust)
  - Biological (metabolic)

PSO compared to PA

- PA just ignores similarity, which leads to severe aberrations
  - Probability of similar connections is badly underestimated
  - Probability of dissimilar connections is badly overestimated
- If the connection probability is correctly estimated, then one immediate application is
  - prediction of new links, or
  - network evolution prediction

Bottom line

- PA is a degenerate (infinite-temperature) regime with similarity/homophily factors reduced to nothing but noise
- If we take these factors into account, then
  - We can predict large-scale growth dynamics of real networks with a remarkable accuracy
  - This growth dynamics has seemingly nothing to do with PA (optimization vs. randomness)
  - Yet if one looks only at degree-based statistics, there is no difference

- M. Boguñá, F. Papadopoulos, and D. Krioukov,
  Sustaining the Internet with Hyperbolic Mapping,
  Nature Communications, v.1, 62, 2010

- D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá,
  Hyperbolic Geometry of Complex Networks,
  Physical Review E, v.82, 036106, 2010,
  Physical Review E, v.80, 035101(R), 2009

- F. Papadopoulos, D. Krioukov, M. Boguñá, and A. Vahdat,
  Greedy Forwarding in (Dynamic) Scale-Free Networks Embedded in Hyperbolic Metric Spaces,
  INFOCOM 2010,
  SIGMETRICS MAMA 2009

- M. Boguñá, D. Krioukov, and kc claffy,
  Navigability of Complex Networks,
  Nature Physics, v.5, p.74-80, 2009