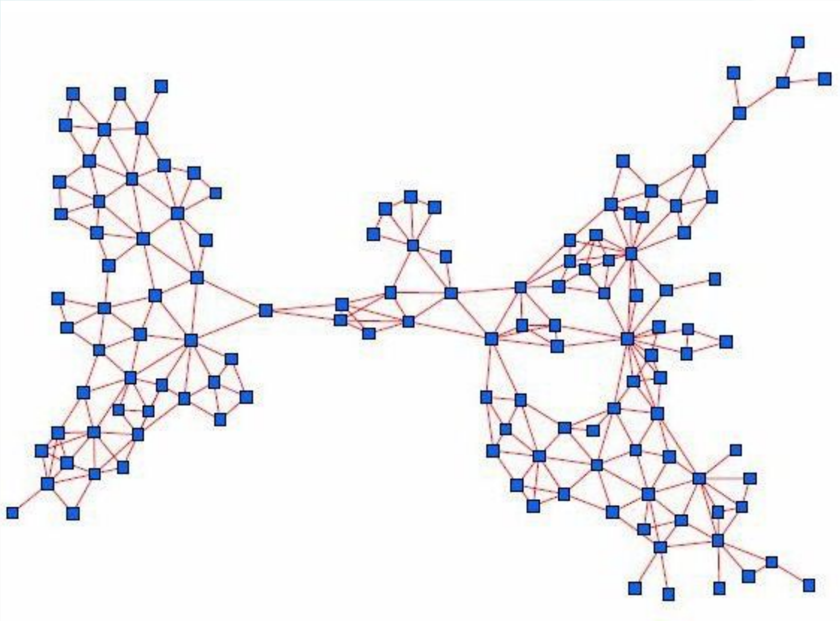
Popularity versus similarity in growing networks

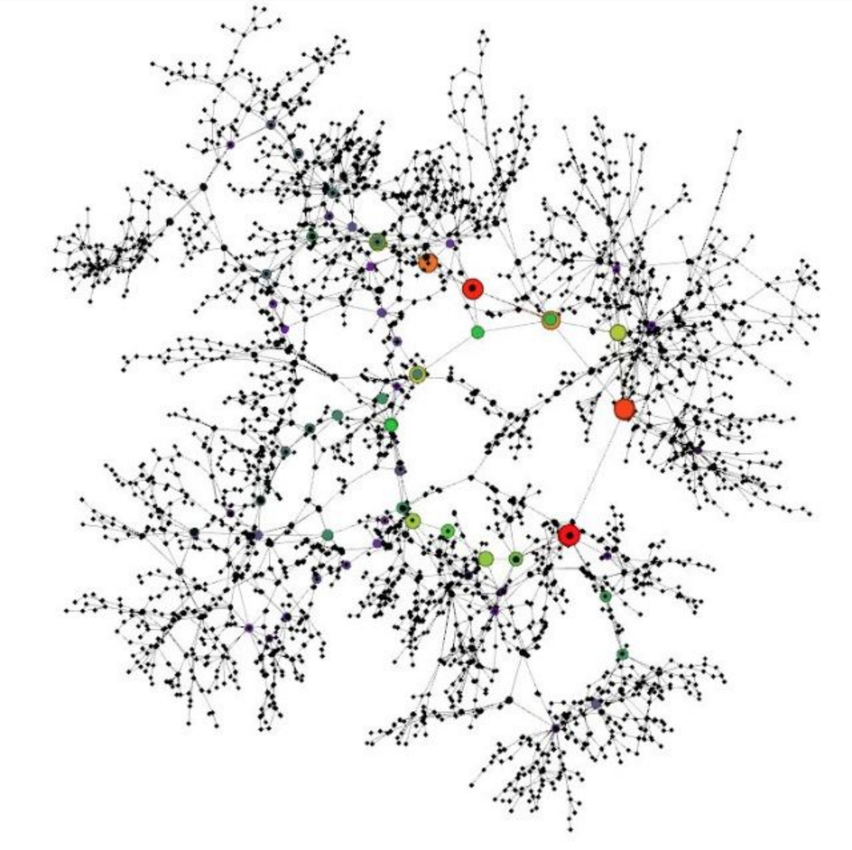
Reporter：Zhongjing Yu

Outline

- Introduction
- Basic model ($model_1$)
- $Model_2$
- $Model_2$'
- $Model_3$
- Expriment
- Conclusion

Introduction





How to generate the network?

"popularity" is considered as the most factor in PA

Introduction

- Basic idea : "similarity" is also an important factor for generating network. Develop a framework where new connections optimize certain trade-offs between popularity and similarity.
- Introduction of a measure of attractiveness that would somehow balance popularity and similarity.

Basic model($Model_1$)

- The model is simply as follows:

1. initially the network is empty.

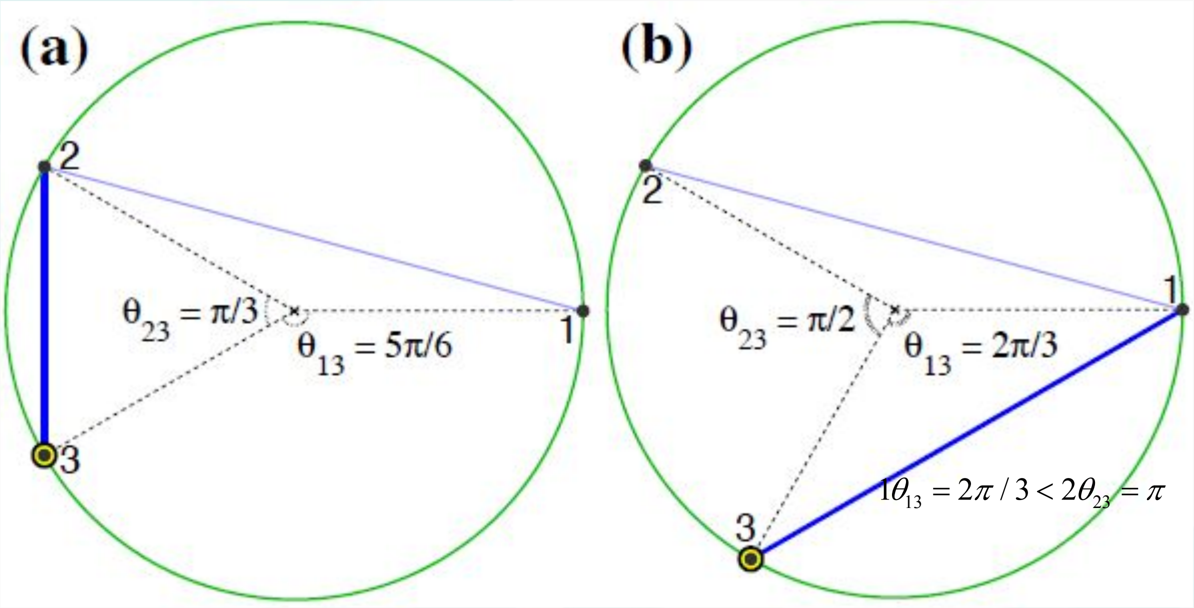
2. at time $t \geq 1$, new node t appears at a random angular position θ_t ,on the circle.

3. new node t connects to a subset of existing nodes s, s<t, consisting of the m nodes with the m smallest values of product $s\theta_{st}$,where m is a parameter controlling the average node degrees $\bar{k} = 2m$, and θ_{st} is the angular distance between nodes s and t .

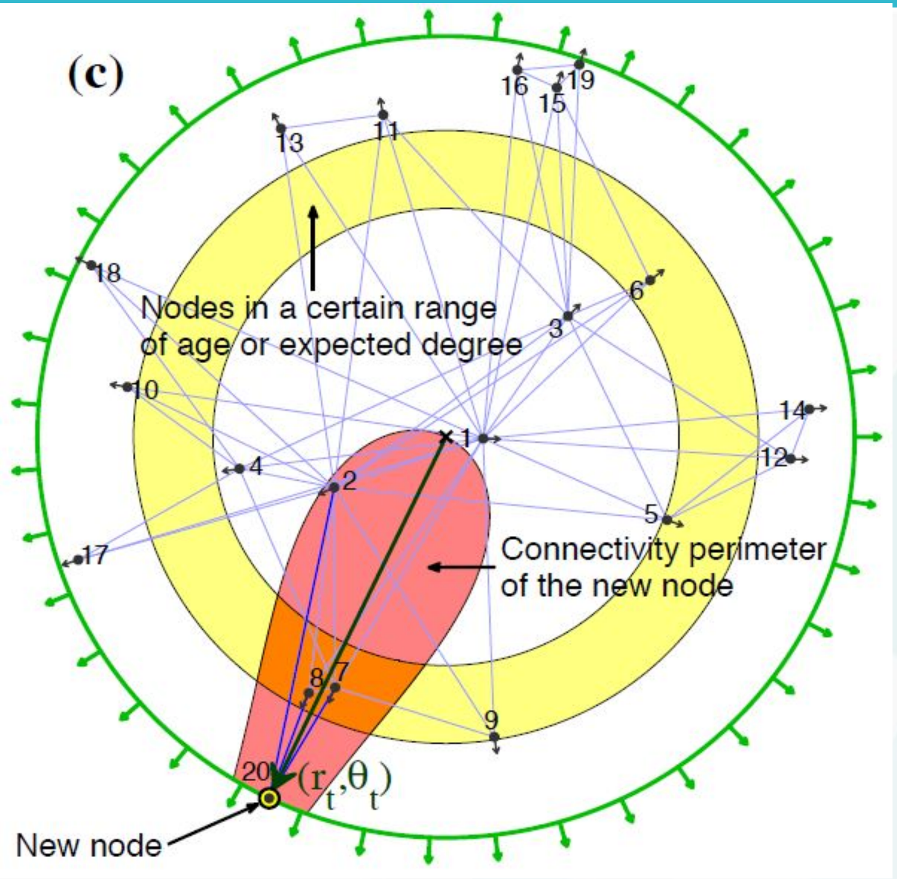
- At time ***t*** , mapping birth time *t* to its radial coordinate via ***r_t = ln t***
- Hyperbolic distance between two nodes at polar coordinates (r_s, θ_s) and (r_t, θ_t) is approximately

$$x_{st} = \frac{1}{2}\operatorname{arccosh}\left(\cosh 2r_s \cosh 2r_t - \sinh 2r_s \sinh 2r_t \cos\theta_{st}\right)$$
$$\approx r_s + r_t + \ln(\theta_{st}/2), \quad \text{where } \theta_{st} = \pi - |\pi - |\theta_s - \theta_t||.$$

$$x_{st} = r_s + r_t + \ln(\theta_{st}/2) = \ln(st\,\theta_{st}/2)$$

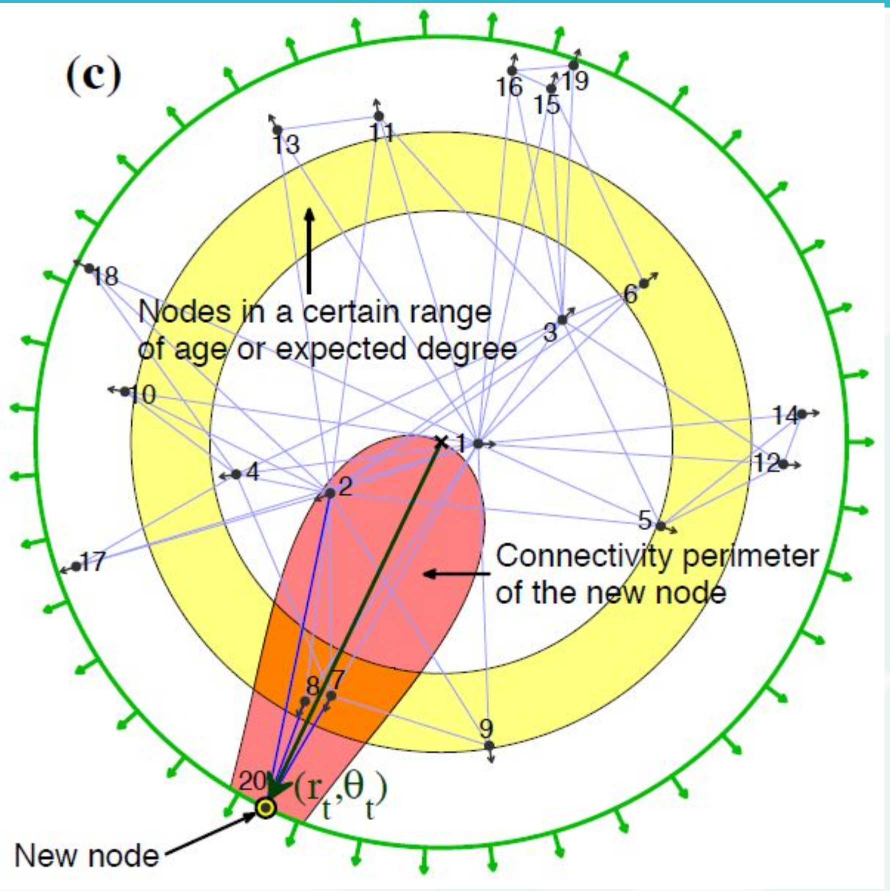


The new circled node *t* in the yellow annulus connects to ***m*** old nodes s minimizing ***s*** θ_{st} . In (a), (b) , t=3 and m=1. In (a), node 3 connects to node 2 because $2\theta_{23} = 2\pi/3 < 1\theta_{13} = 5\pi/6$ In (b), similarly. Node3 connects to node 1 because $1\theta_{13} = 2\pi/3 < 2\theta_{23} = \pi$



An optimization-driven network with m=3 is simulated for up to 20 nodes . The radial(popularity) coordinate of new node t=20 is $\boldsymbol{r_t}$ **= *lnt***, shown by the long thick shape marks the set of points located at hyperbolic distances less than $\boldsymbol{r_t}$ from the new node.

Consider : popularity fading



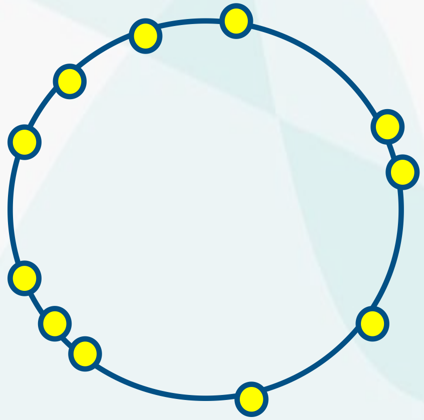
- The closer the node to the centre , the more popular it is : the higher its degree, and the more new connections it attracts. Therefore to model popularity fading, and let all nodes drift away from the centre.

Let all nodes drift away from the centre

the radial coordinate of node s at time t>s is increasing as

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

and parameter $\beta \in [0,1]$, It changes the power-law exponent to $\gamma = 1 + 1/\beta \geq 2$ .If $\beta = 1$,the nodes do not move and $\gamma = 2$. If $\beta = 0$, all nodes move with the maximum speed, always lying on the circle of radius r_t , while the network degenerates to **a random geometric graph** growing on the circle.



Compare : attraction probability

Compare: attraction probability:

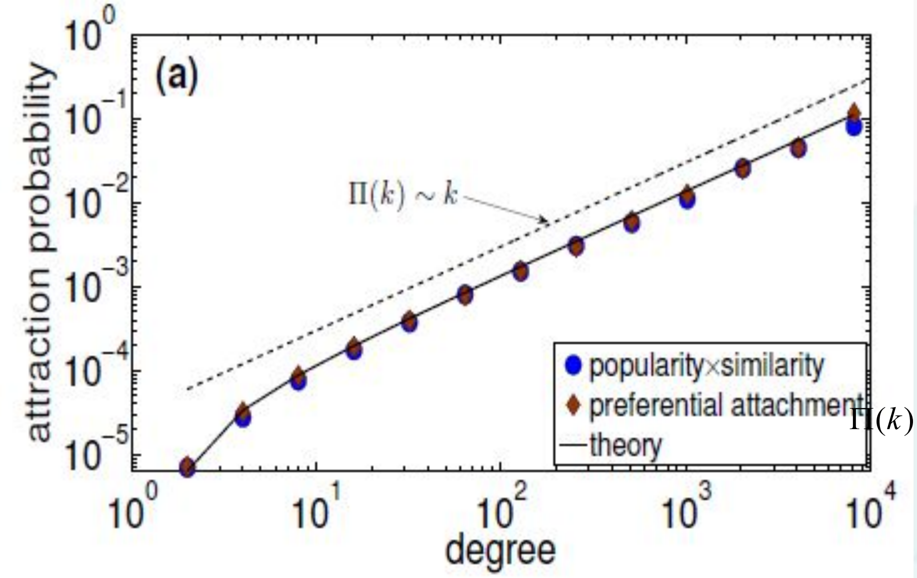
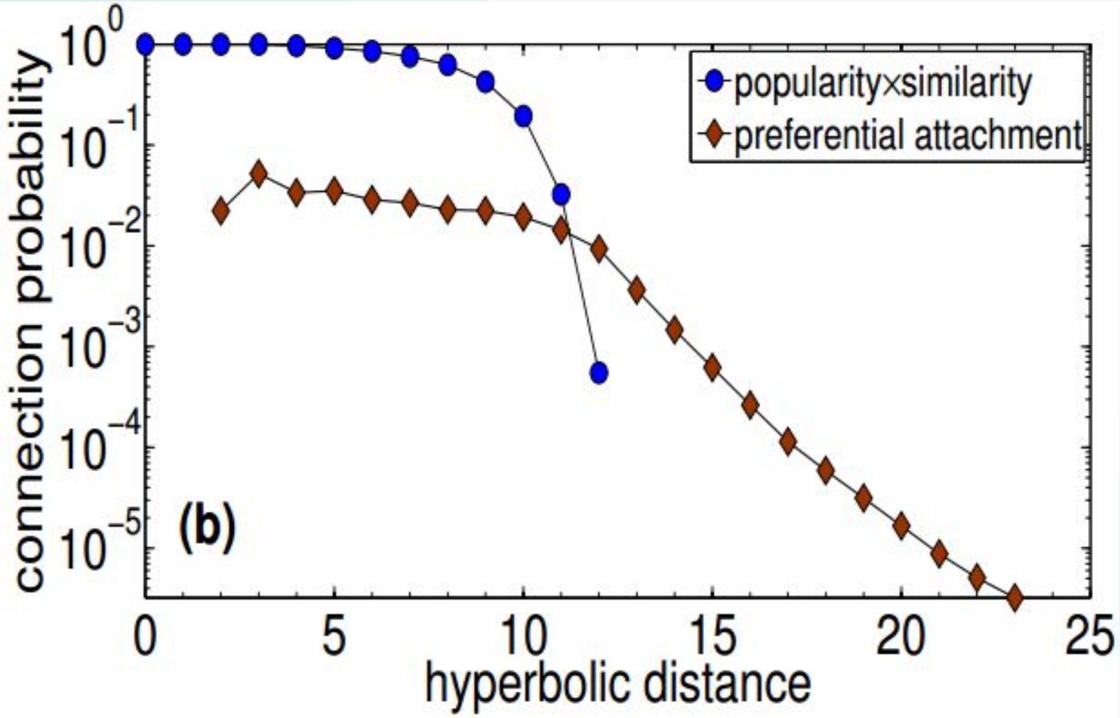


Fig.(a) illustrated that the probability $\Pi(k)$ that an existing node of degree *k* attracts a connection from a new node is the same **linear function** of *k* in the described model and in PA. and those models that will be introduced are same power laws. And prove that the exponent γ of this power law approaches 2.

$$\Pi(k) = m\frac{k-m+A}{(m+A)t}, A = (\gamma-2)m, \gamma = 1+\frac{1}{\beta}$$

Compare :connection probability

Compare: connection probability :



in PA the probability of their connection is lower. on other hand, nodes that are far apart are never connected in the optimization model, whereas they can be connected in PA.

model$_2$

- Consider :

  Strongest clustering VS Weaken clustering

  ➔the closest nodes VS farther nodes.

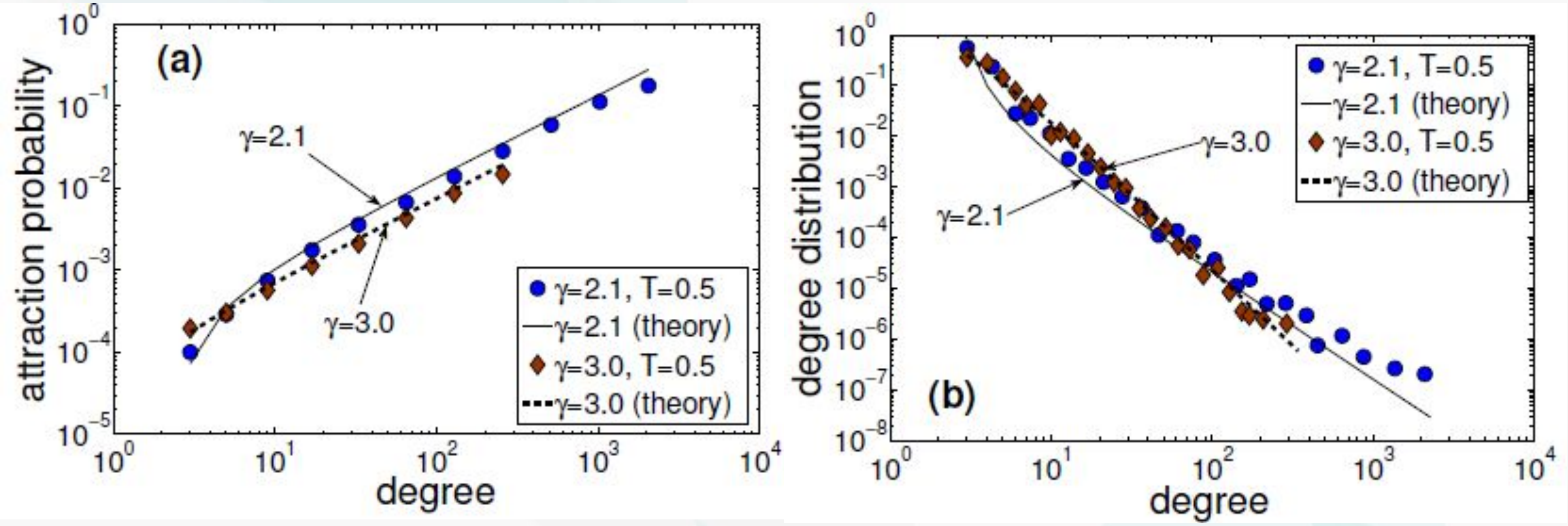
- Modify step(3) follows : new node *t* picks a randomly chosen node *s* , s<t , and given that it is not already connected to it, it connects to it with probability $p(x_{st}) = 1/\{1 + \exp[(x_{st} - R_t)/T]\}$ $R_t \approx r_t$, where parameter $T \geq 0$ is the network temperature, and connect to nodes lying within distance $R_t \approx r_t$ Node *t* repeats this step until it gets connected to *m* nodes

$$\overline{N(R_t)} = tP(t) = \frac{2T}{\sin T\pi} e^{-(r_t - R_t)} \frac{1}{1-\beta} \left(1 - e^{-(1-\beta)r_t}\right).$$

$$R_t = r_t - \ln\left[\frac{2T}{\sin T\pi} \frac{\left(1 - e^{-(1-\beta)r_t}\right)}{m(1-\beta)}\right]$$

where $\overline{N(R_t)}$ is the average number of existing nodes lying within R_t

model$_2$



The plot also shows the corresponding theoretical predictions

- Degree distribution (PA, $model_1$ and $model_2$) :

result :

$$\Pi_{\mathrm{Model}_2}(s,t) = m\frac{P(s,t)}{P(t)} = m\frac{e^{-r_s(t)}}{\int_1^t e^{-r_i(t)}di} = \Pi_{\mathrm{Model}_1}(s,t) = \Pi_{\mathrm{PA}}(s,t).$$

this means that for fixed *m* and $\beta = \frac{1}{\gamma - 1}$,the degree distribution and link attraction probability in $Model_2$ are the same as in $Model_1$

setting R_t with $\overline{N(R_t)} = m$.$Model_2$ becomes identical to $Model_1$

$$R_t = r_t - \ln\left[\frac{2}{\pi}\frac{\left(1-e^{-(1-\beta)r_t}\right)}{\overline{N(R_t)}(1-\beta)}\right].$$

About parameter T ($p(x_{st}) = 1/\{1 + \exp[(x_{st} - R_t)/T]\}$) :

Temperature T : parameter controlling clustering in the network.

- T = 0 : $p(x_{st})$ is either 0 or 1 depending on whether distance x_{st} is less or greater than R_t.
- $T \geq 1$: clustering gradually decreases to zero
- $T \to \infty$: the model degenerates either to growing random graphs, or remarkably, to standard PA.

model$_2$'

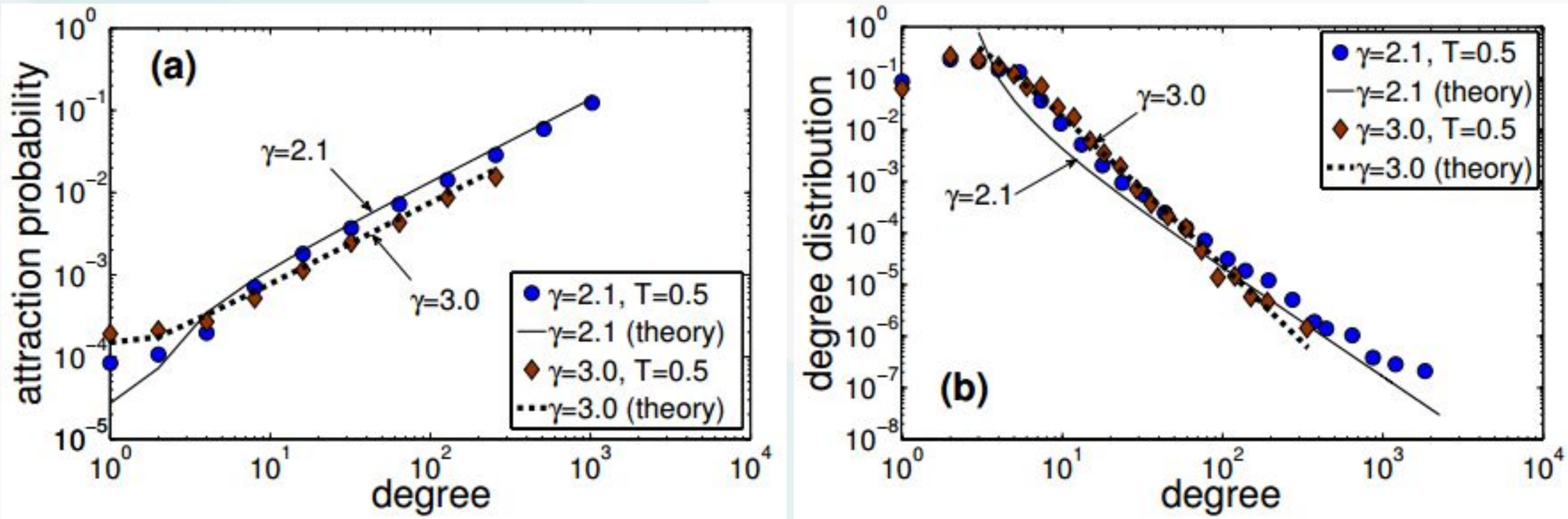
looks at every existing node *s*, s<t,only once and connects to it with probability $p(x_{st})$ given by

$$p(x_{st}) = \frac{1}{1 + \left(X(s,t)\frac{\theta_{st}}{2}\right)^{\frac{1}{T}}}, \quad \text{where } X(s,t) = e^{(r_s(t)+r_t-R_t)}$$

$$x_{st} = r_s + r_t + \ln(\theta_{st}/2) = \ln(st\,\theta_{st}/2)$$

$$p(x_{st}) = 1/\{1 + \exp[(x_{st} - R_t)/T]\}$$

model$_2$'



Plot(a) shows the probability $\Pi(k)$ that an existing node of degree k attracts a link in network grown up to $t = 10^4$ nodes according to model$_2$' with T=0.5 and $\gamma = 2.1, 3.0$ .Each new node connects on average to m=3 existing nodes. The theoretical predictions are given by $\Pi(k) = m\frac{k-m+A}{(m+A)t}$. when $k \geq m$, and when k<m are given by the formula $\Pi(k) = m\frac{A}{(m+A)t}$.

Plot(b) shows the distribution P(k) of nodes degree in same networks. The theoretical predictions are given by $P(k) = (\gamma-1)\frac{\Gamma[(m+1)(\gamma-2)+1]\Gamma[k+m(\gamma-3)]}{\Gamma[m(\gamma-2)]\Gamma[k+m(\gamma-3)+\gamma]}$

model$_3$: the fitness model

- The main motivation behind the fitness model is that the popularity of a node does not depend only on its birth time, but also on its ability(fitness) to compete for links.
- The following attraction probability is introduced :

$$\Pi(k_{\eta_s}(t)) = m\frac{\eta_s\left(k_{\eta_s}(t) - m + A\right)}{\int_1^t \eta_i \left(\overline{k_{\eta_i}(t)} - m + A\right) di},$$

and thus have:

$$\Pi_{\text{fitness}}(s,t) = m\frac{\eta_s\left(\overline{k_{\eta_s}(t)} - m + A\right)}{\int_1^t \eta_i \left(\overline{k_{\eta_i}(t)} - m + A\right) di}.$$

model$_3$

- Now drifts away by increasing its radial coordinate using the formula :

$$r_s = \beta(\eta_s) r_s + (1 - \beta(\eta_s)) r_t - \ln \frac{\eta_s}{\eta_{\max}}$$

Parameter $\beta(\eta_s)$ is some function of the fitness of node s, η_s, and therefore its value can be different for different nodes.

- **using same methods :**

$$\Pi_{\mathrm{Model}_3}(s,t) = m\frac{e^{-r_s(t)}}{\int_1^t e^{-r_i(t)}di} = m\frac{\eta_s\left(\frac{s}{t}\right)^{-\beta(\eta_s)}}{\int_1^t \eta_i\left(\frac{i}{t}\right)^{-\beta(\eta_i)}di},$$

$$\overline{N(R_t)} = \frac{2T}{\sin T\pi}e^{-(r_t-R_t)}\frac{1}{\eta_{max}t}\int_1^t \eta_i\left(\frac{i}{t}\right)^{-\beta(\eta_i)}di,$$

$$R_t = r_t - \ln\left[\frac{2T}{\sin T\pi}\frac{\frac{1}{\eta_{max}t}\int_1^t \eta_i\left(\frac{i}{t}\right)^{-\beta(\eta_i)}di}{m}\right] \quad \text{for } \overline{N(R_t)} = m.$$

Map real network

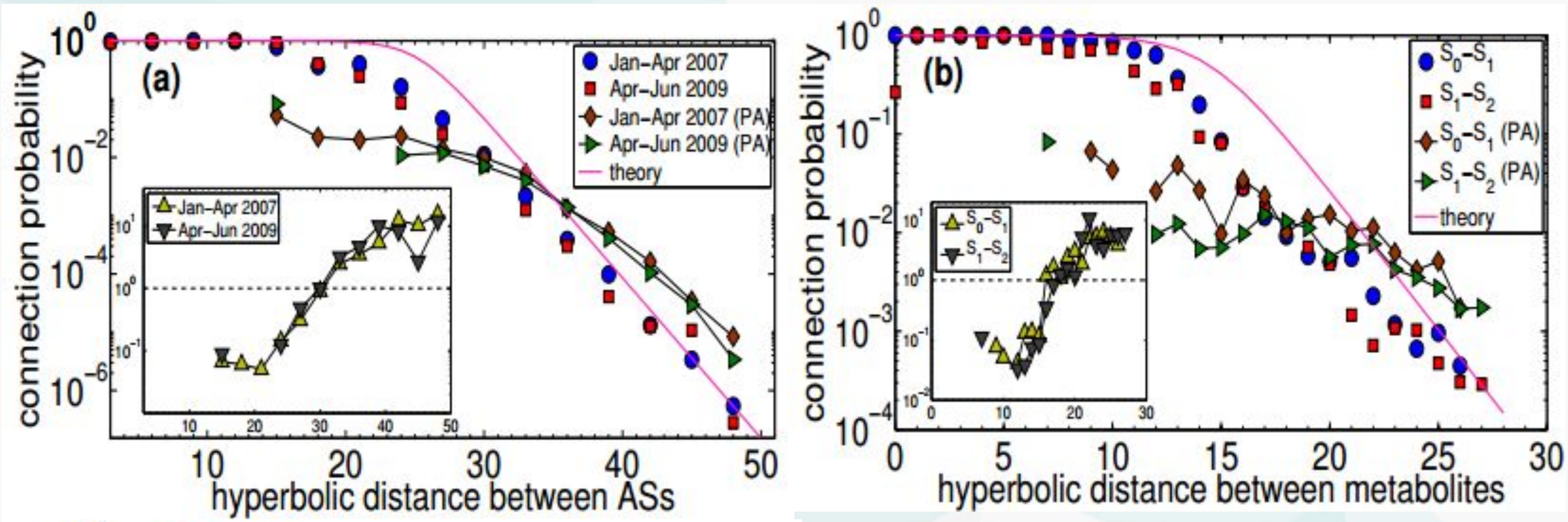
- Target: infer (r_i, θ_i),
- Given: adjacency matrix a_{ij} ,i,j=1,2,...,t. current degree k_i
- Method :

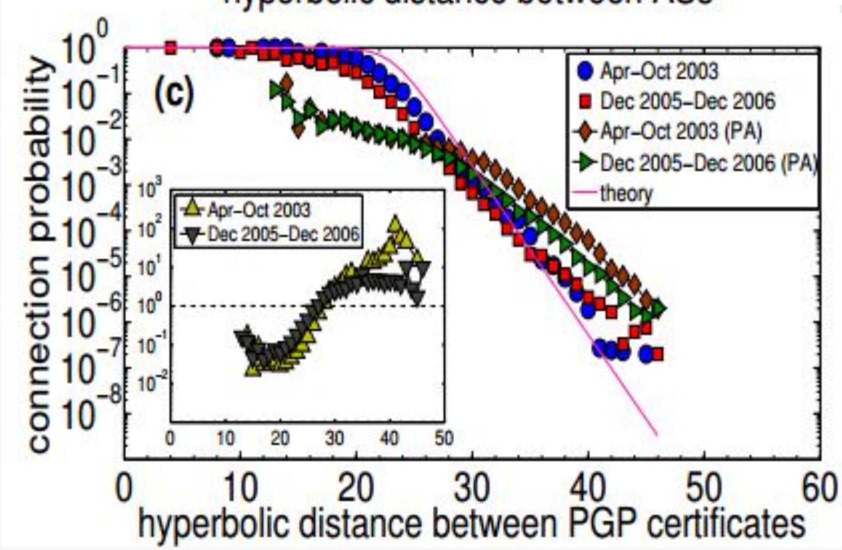
Current degree k_i ($k_i \sim e^{r_t - r_i}$) instead radial coordinate r_i to get x_{ij}

Then find t $\mathcal{L} = \prod_{i<j} p(x_{ij})^{a_{ij}} [1 - p(x_{ij})]^{1-a_{ij}}$ ates(θ_i) that maximize likelihood

where $p(x_{st}) = 1/\{1 + \exp[(x_{st} - R_t)/T]\}$ (MCMC)

Experiment





the insets of each plot show the ratio between the connection probabilities in PA emulations and the real networks, i.e., the ratios of the values shown by diamonds and circles.

Conclusion

- Provides a natural geometric explanation
- More accuracy
- Robust