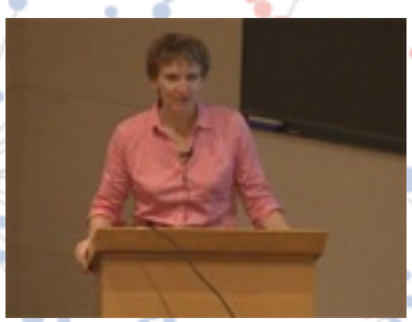
Homophily and the Emergence of a Glass Ceiling Effect in Social Networks

Claire Mathieu, CNRS and École Normale Supérieure



Chen Avin, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, Yvonne-Anne Pignolet

What is the glass ceiling effect?

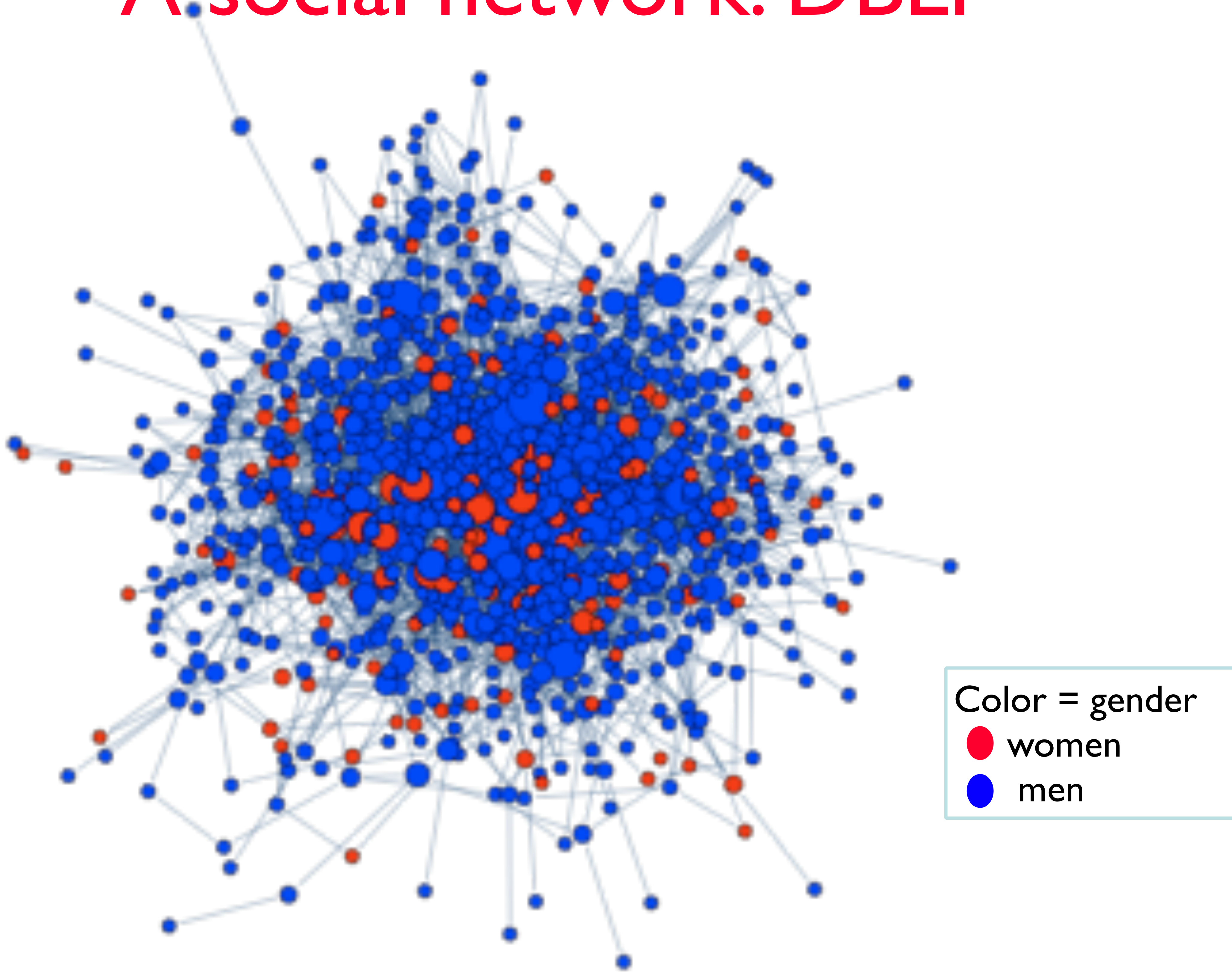
Federal Glass Ceiling Commission, US Gov, 1994:
"*the unseen, yet unbreakable barrier that keeps minorities and women from rising to the upper rungs of the corporate ladder, regardless of their qualifications or achievements*"

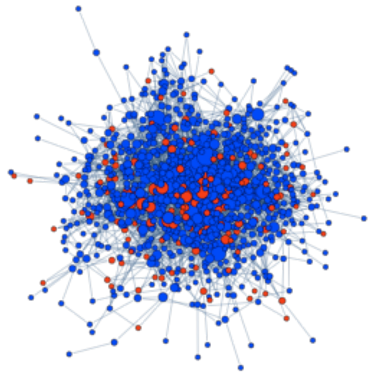
Merriam-Webster: "*an unfair system or attitudes that prevents some people (such as women or a certain race) from getting the most powerful jobs*"

What is our main goal?

Model glass ceiling in a social network, analyze it, compare with empirical data

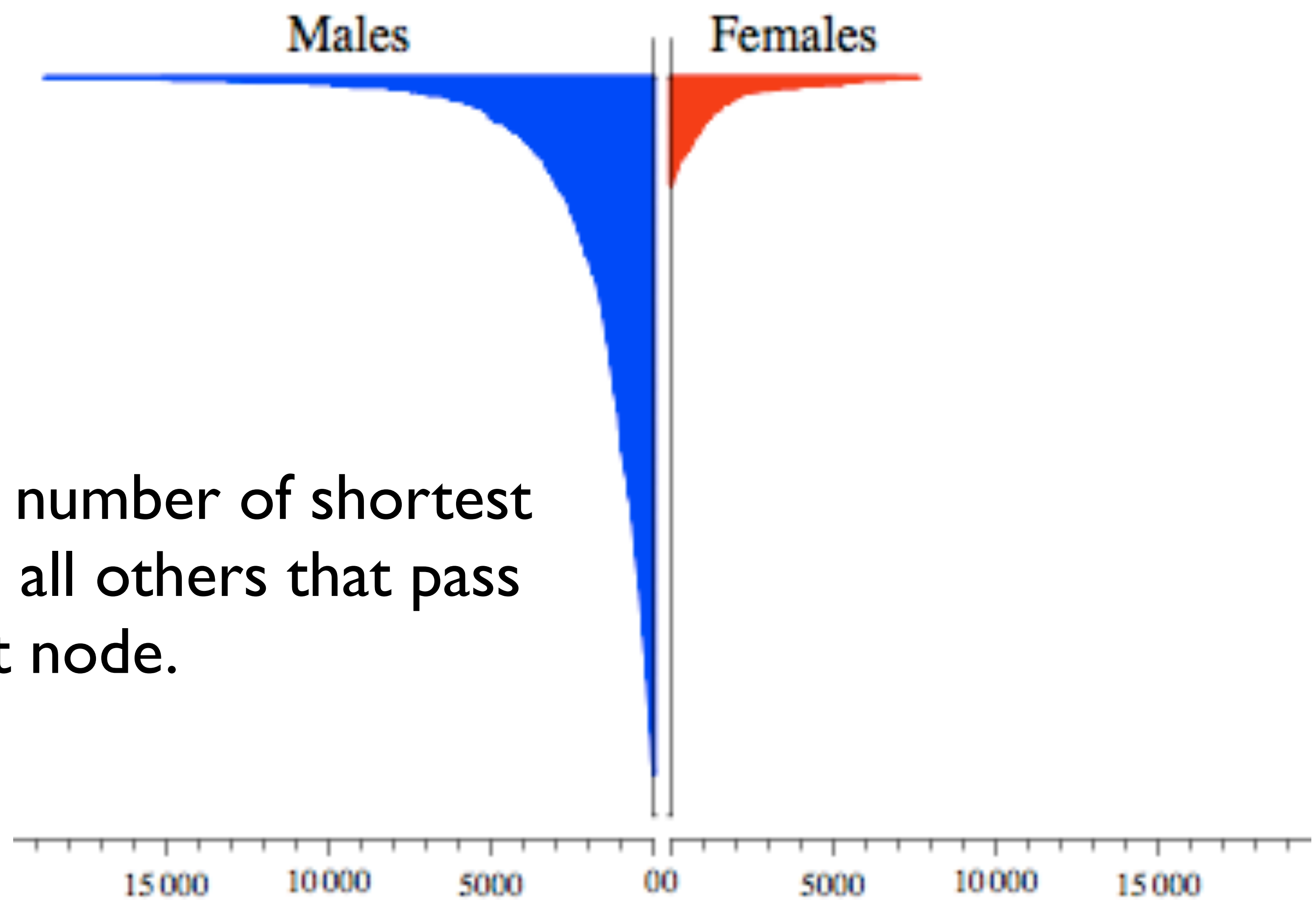
A social network: DBLP





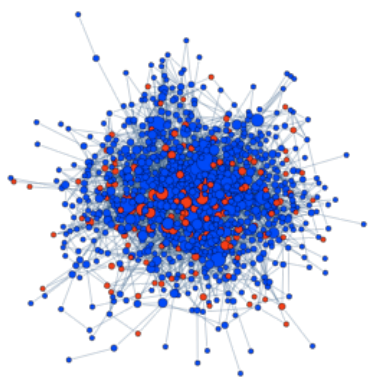
Are men and women different?

(from the viewpoint of DBLP)



Betweenness centrality = number of shortest paths from all vertices to all others that pass through that node.

DBLP - top coauthorship network



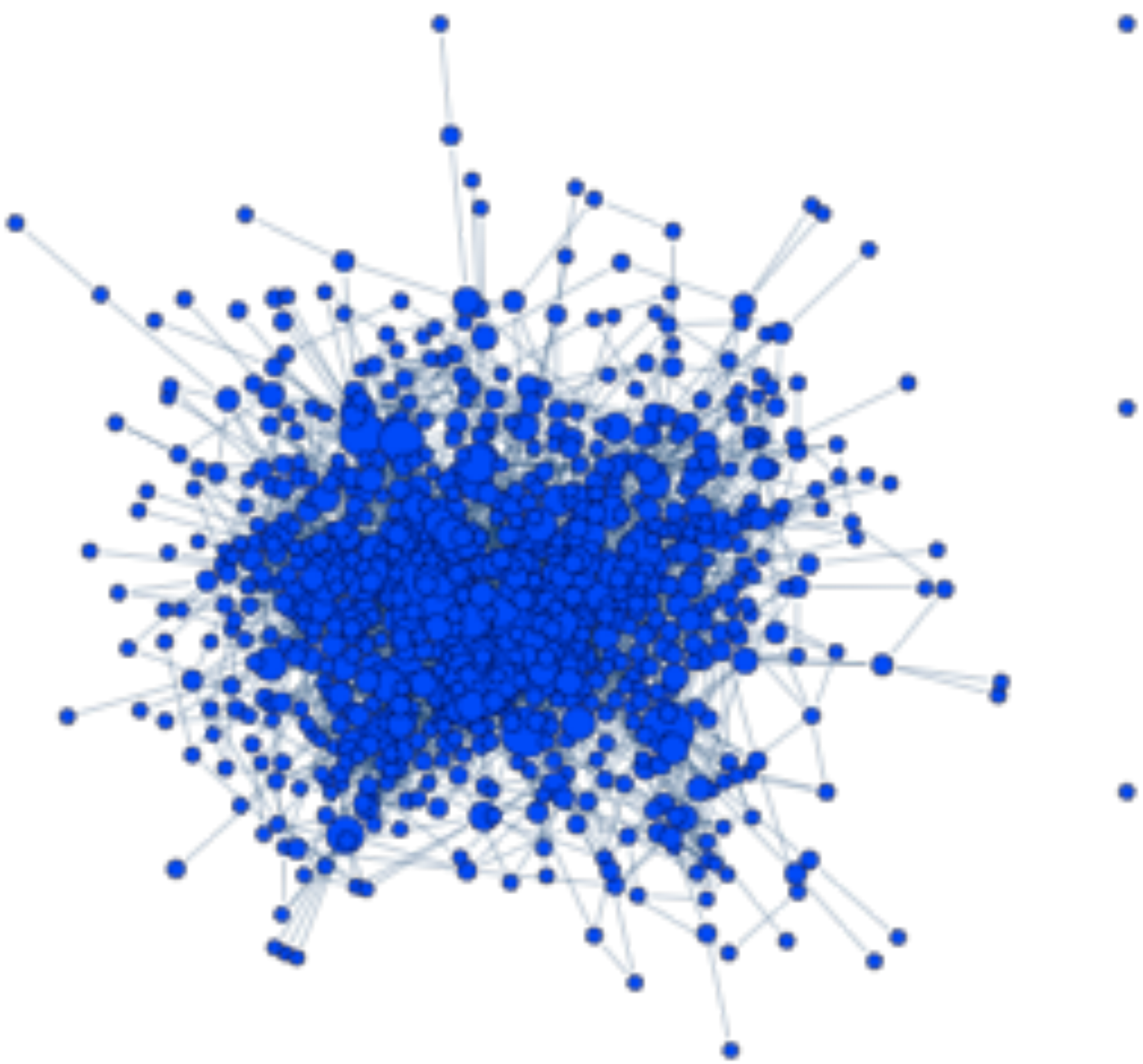
Are men and women different?

(from the viewpoint of DBLP)

Induced subgraphs

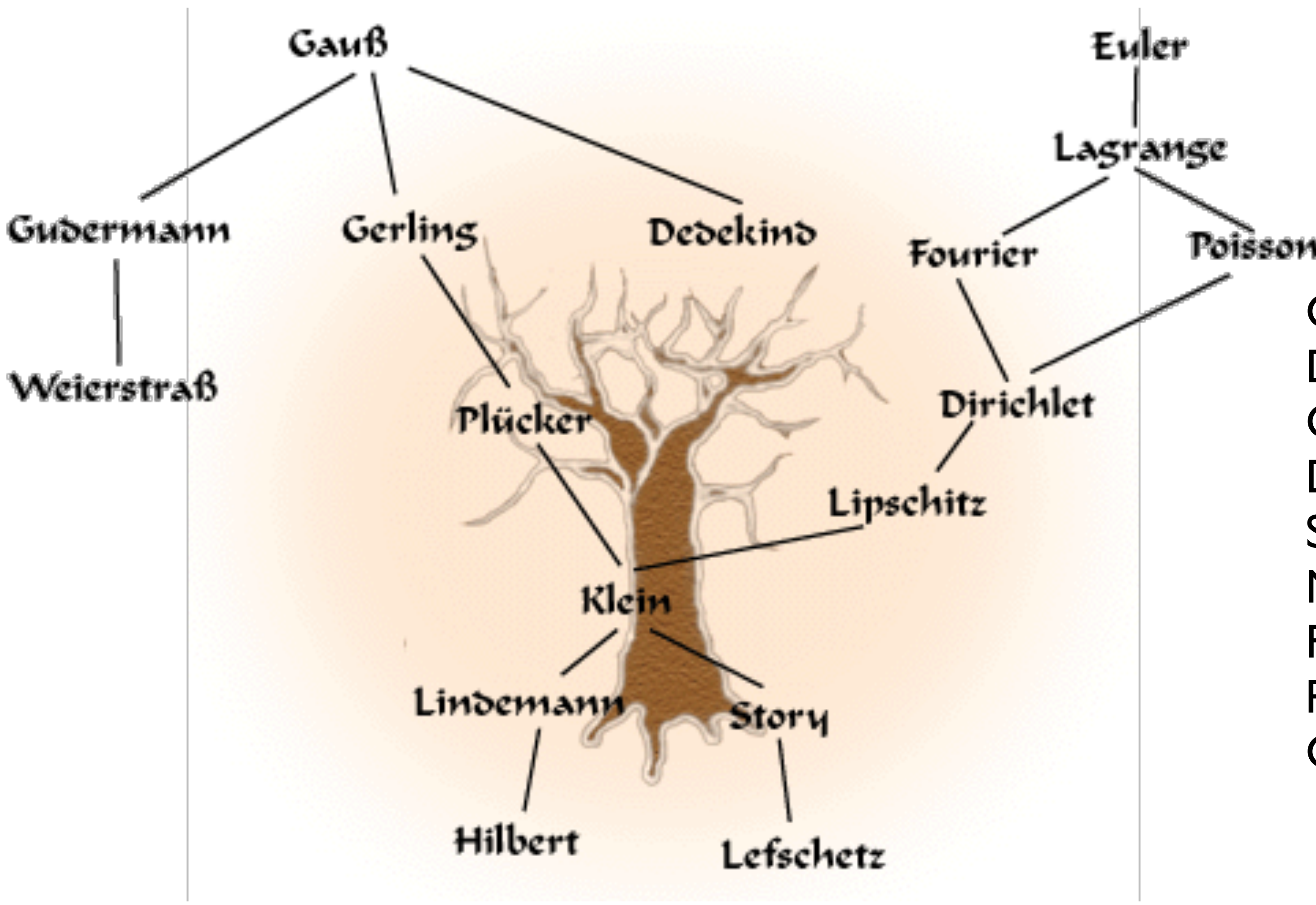
Female network

Male network



DBLP - top coauthorship network

Another social network: Student-Mentor (advisor) graph in academia



Poisson
Chasles
Darboux
Goursat
Darmois
Shutzenberger
Nivat
Flajolet
Puech
Claire Mathieu

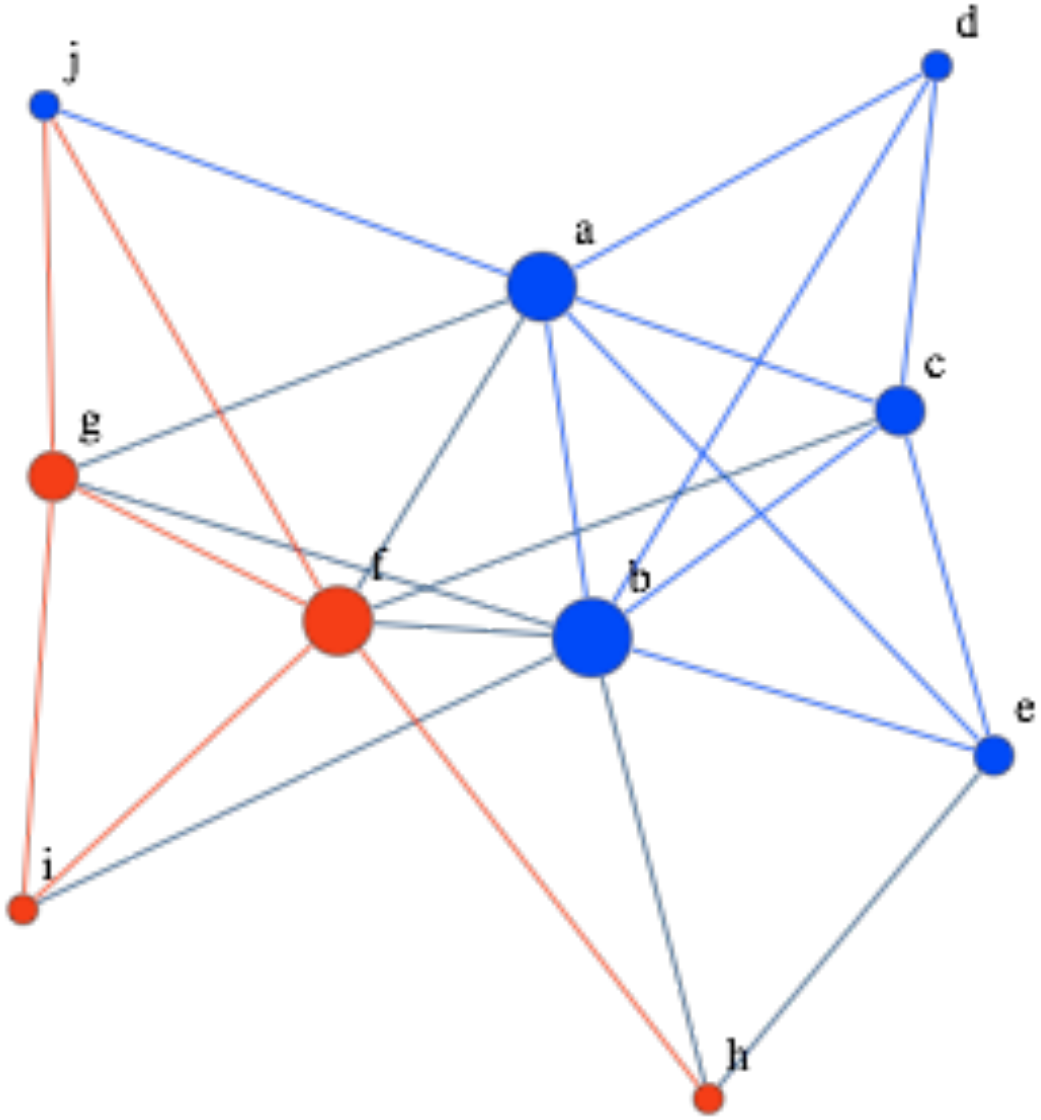
Hermite
Tannery
and Chatelet

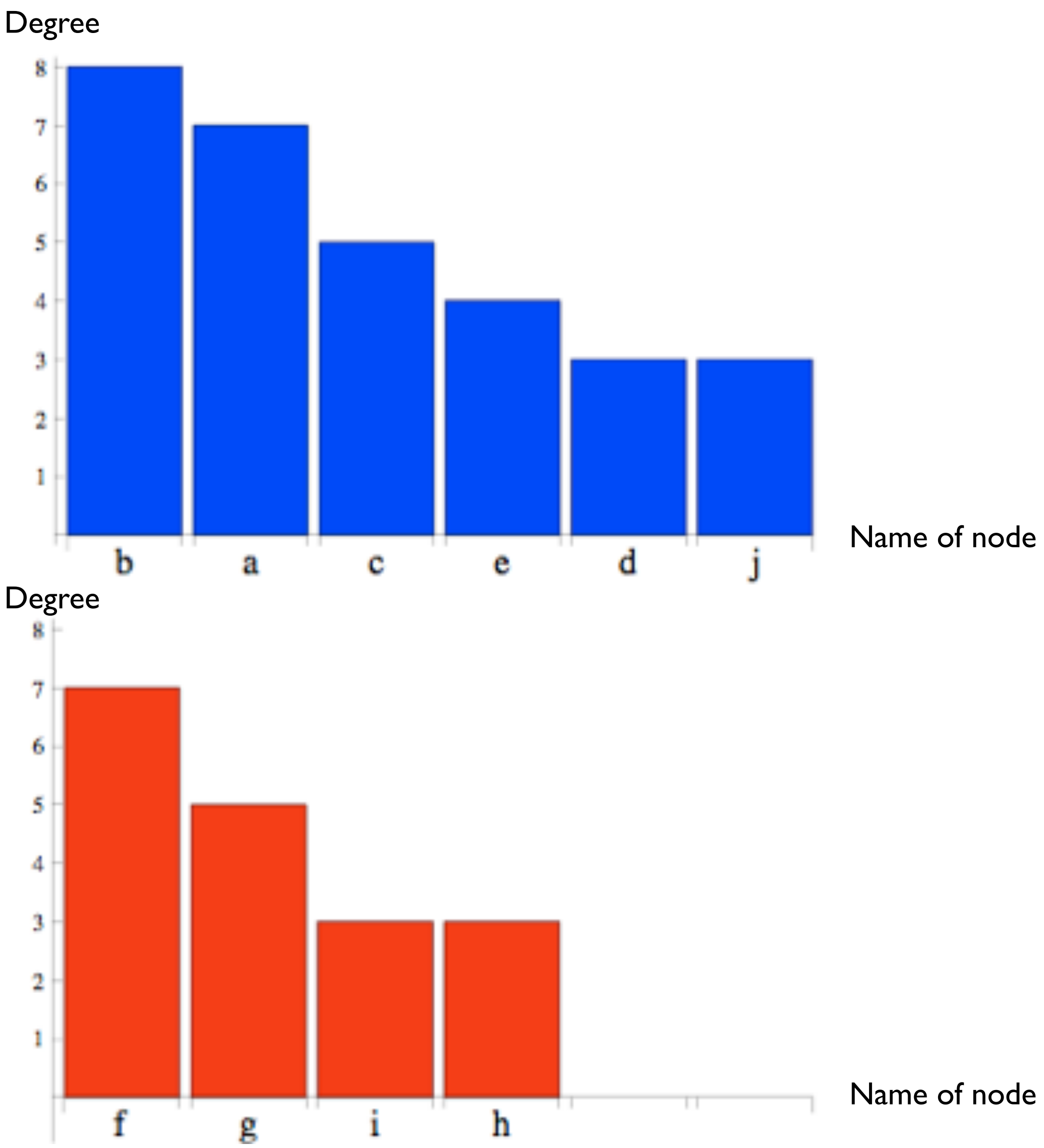
Chasles
H.Newton
Moore
Veblen
Franklin
Perlis
Manna
and Vuillemin

- people: nodes
- edge: advisor-advisee, mentor-mentee

From DBLP, extract gender and Student-Mentor graph

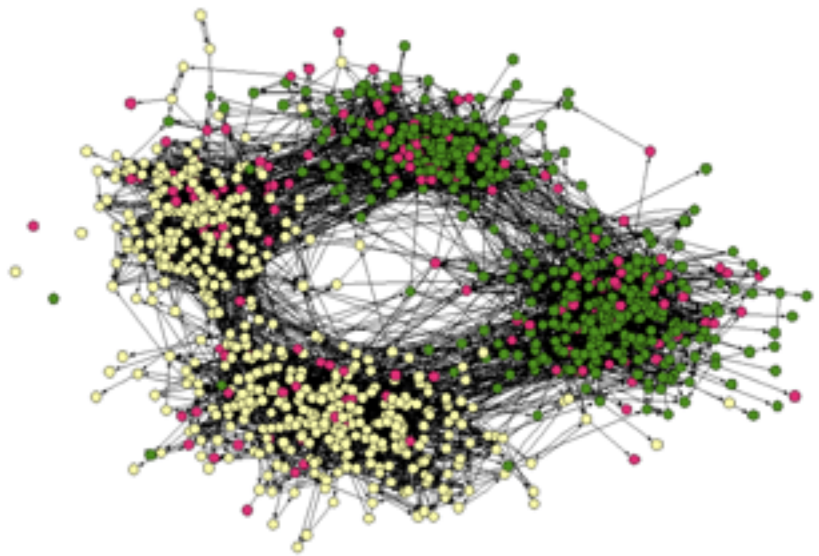
“Most powerful”, “upper rungs”: high degree





Social concepts used for the model

- "rich get richer" - preferential attachment
- homophily - tendency of individuals to associate and bond with similar individuals.
- minority



Rich Get Richer Lyrics

"Rich Get Richer" was written by Jerry A. Lang.

You might have more than me,
I'm not so blind I can't see,
The rich getting richer, who says that it has to be?
Money talks, money screams,
Middle class lies, American dreams,
The rich get richer, the poor get poorer,
Who says it has to be.

Model glass ceiling in a social network: Biased Preferential Attachment (BPA) - $G(n, r, \rho)$

1. New node (student) arrives. **Minority** parameter:
   - Red with probability $r < 1/2$
   - Blue with probability $1 - r$
2. Selects a neighbor (advisor) by **preferential attachment** $\frac{\delta_t(v)}{\sum_{i \in V(t)} \delta_t(i)}$
3. Checks neighbor's color. **Homophily** effect:
   Same color: accept
   Different color: accept with probability ρ, otherwise go to step 2.

of (red, blue) nodes with degree at least k: $\mathrm{top}_k(\mathtt{R}), \mathrm{top}_k(\mathtt{B})$

- **Tail glass ceiling:** $G(n)$ exhibits tail glass ceiling effect for the red nodes if:

$$\lim_{G \to \infty} \frac{\mathrm{top}_k(\mathtt{R})}{\mathrm{top}_k(\mathtt{B})} \longrightarrow 0$$

  while k=$k(n)$ is s.t.

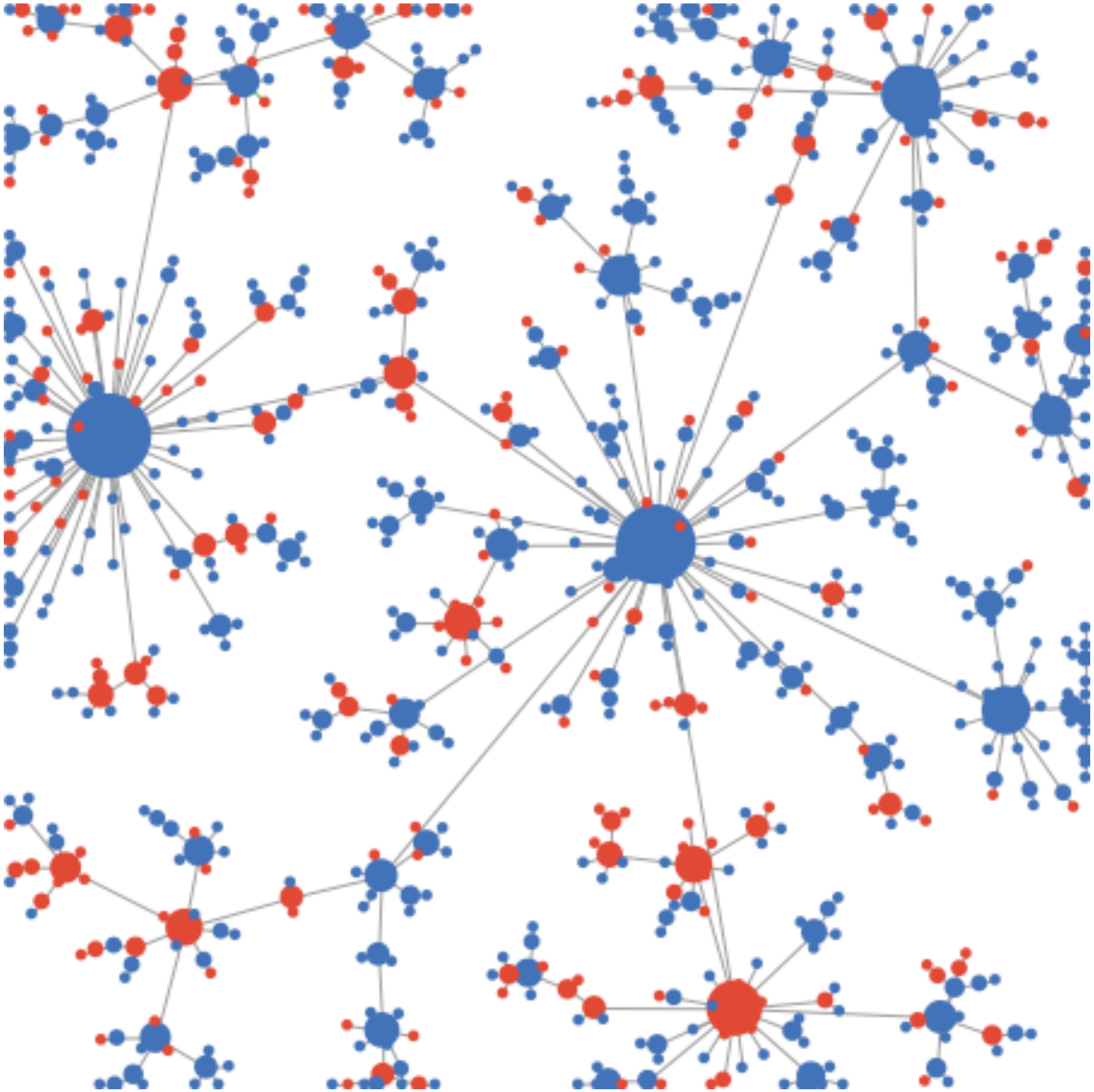
$$\mathrm{top}_k(\mathtt{B}) \to \infty$$

- **Moment glass ceiling:** ratio of expected square degree of red vs. of blue vertices

Main Theorem

- If $0 < r < ½$ and $0 < \rho < 1$ then: **glass ceiling**
- If $r = ½$ (no minority) or $\rho = 1$ (no homophily) then: **no glass ceiling**
- If new vertex selects advisor uniformly (no preferential attachment) then: **no moment glass ceiling**.

I.e.: three assumptions → glass ceiling ,
any two assumptions → no glass ceiling



r=30, ρ=0.71

Proving the glass ceiling effect

Analyze it: Proof outline

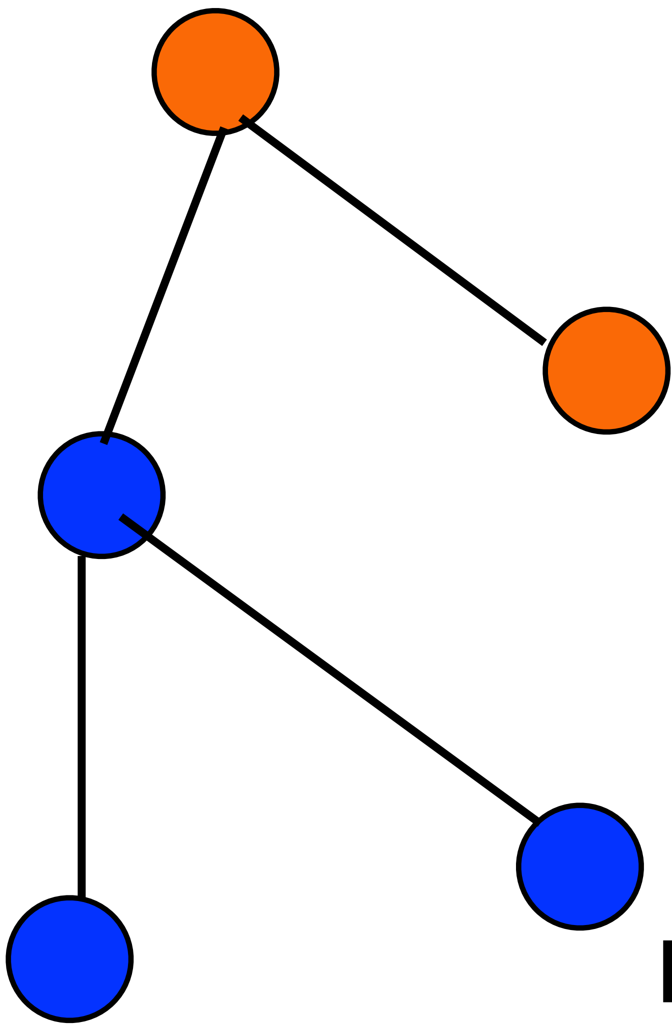
1. $\sum_{v \text{ red}} \text{degree}(v)/\#(\text{red nodes})$: fast convergence in expectation

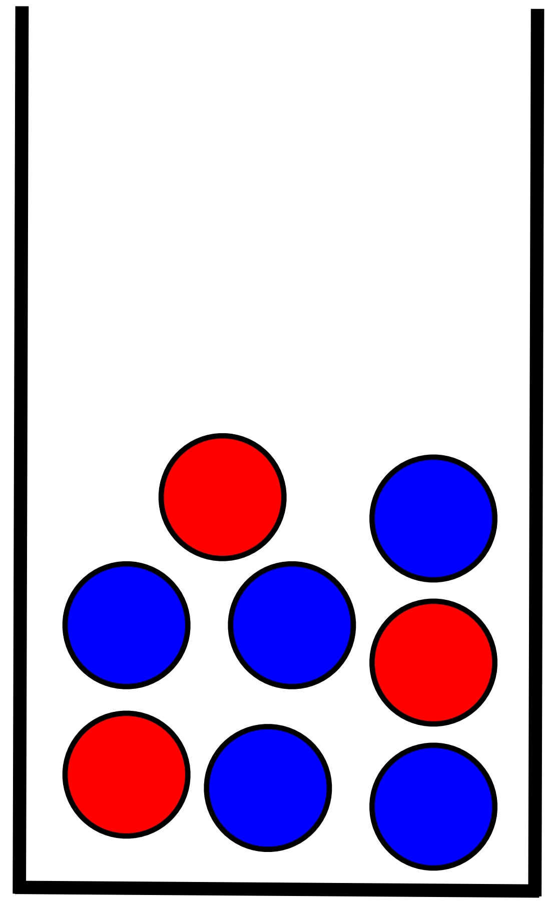
2. $\sum_{v \text{ red}} \text{degree}(v)/\#(\text{red nodes})$: fast convergence with high probability

3. Distribution of degrees of red nodes and of degrees of blue nodes

$\sum_{v \text{ red}} \text{degree}(v)/\#(\text{red nodes})$

From graph to urn





Generate new ball, red or blue

(r, 1-r)

Pick existing ball uniformly from urn

Same color - add a copy

Different color -

with probability ρ add a copy

with probability $1-\rho$ retry

Ball = half-edge,
color = color of endpoint

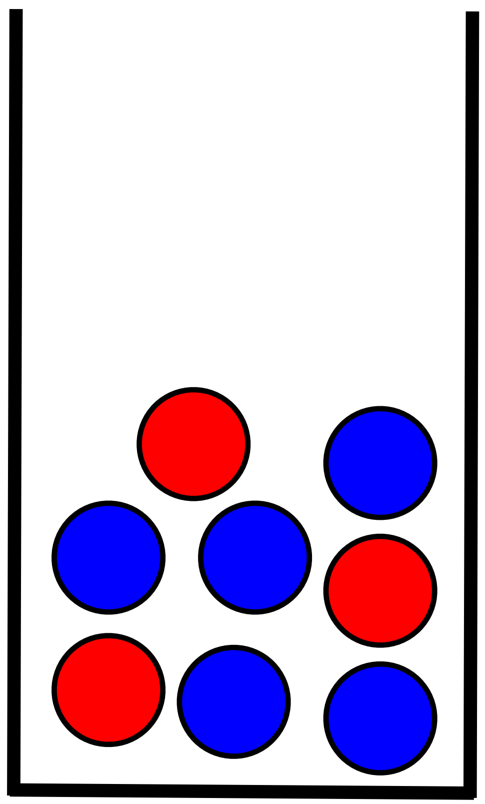
$$\sum_{v \text{ red}} \text{degree}(v) \sim \#(\text{red balls})$$

$$\#(\text{red nodes}) \sim rt$$

$$\#(\text{balls}) \sim 2t$$

$$\text{Analyze } \frac{\#(\text{red balls})}{\#(\text{balls})}$$

step 1- expected fraction of red balls (1/3)



Generate new ball, red or blue

(r, 1-r)

Pick existing ball uniformly from urn

Same color - add a copy

Different color -

with probability ρ add a copy

with probability 1-ρ retry

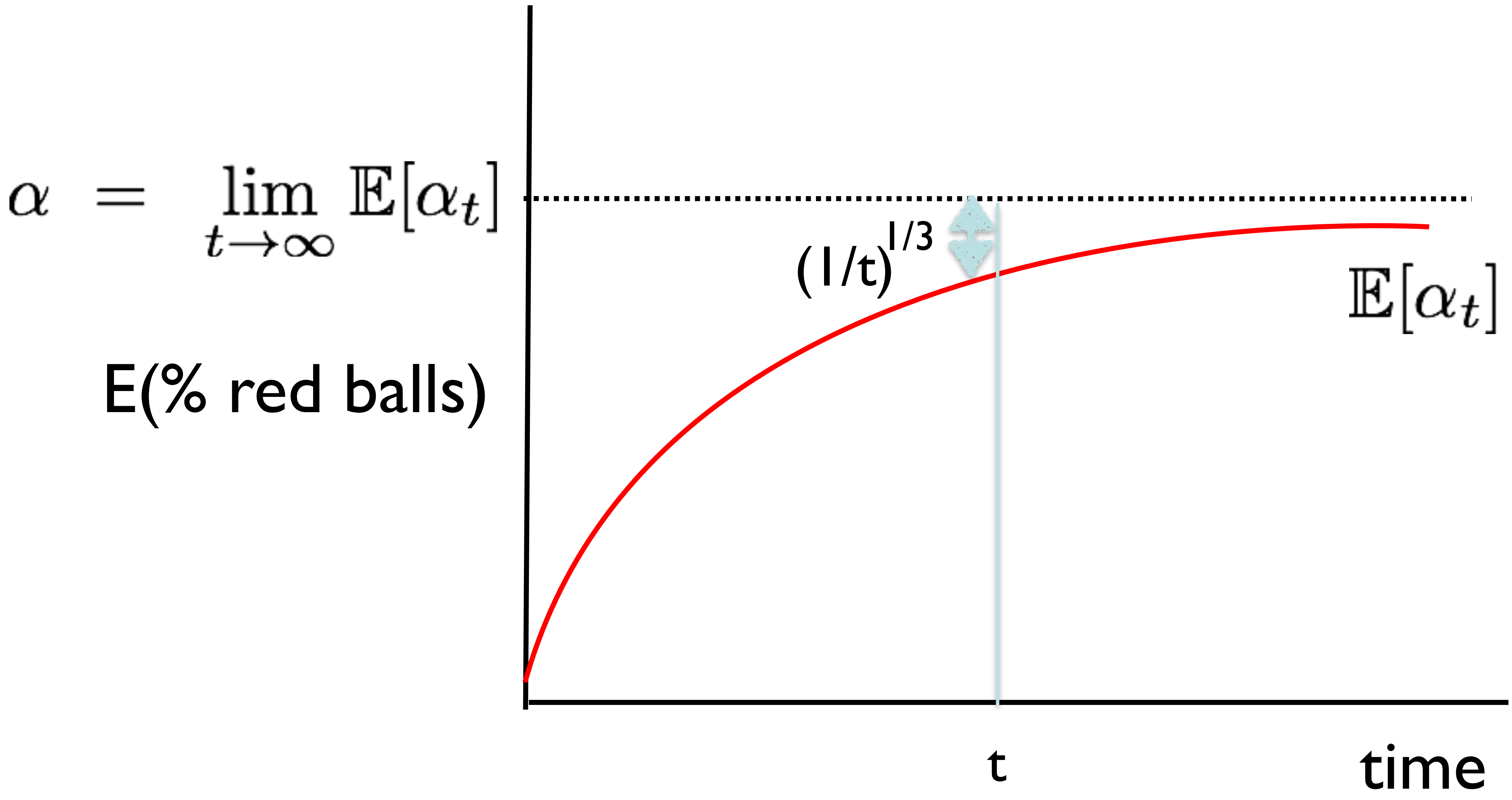
Let $X_t = \#(\text{red balls})$

Let $\alpha_t = \#(\text{red balls})/\#(\text{balls})$

$$X_{t+1} - X_t = \begin{cases} 2 & \text{w.p. } \frac{r\alpha_t}{1-(1-\alpha_t)(1-\rho)} \\ 0 & \text{w.p. } \frac{(1-r)(1-\alpha_t)}{(1-\alpha_t)(1-\rho)} \\ 1 & \text{otherwise} \end{cases}$$

$$\implies \mathbb{E}(\alpha_{t+1} | \alpha_t)$$

Rate of convergence of expectation (3/3)



step 2 - High probability convergence (1/3)

Martingale def'n: B_i is a martingale if

$$\mathbb{E}(B_{i+1}|\text{history up to time } t) = B_i$$

Doob martingale: If N_i = number of new red balls at time i then

$B_i = \mathbb{E}(N_0 + \ldots + N_t | N_0, \ldots, N_i)$ is a martingale.

$B_0 =$ Expected number of red balls at time t

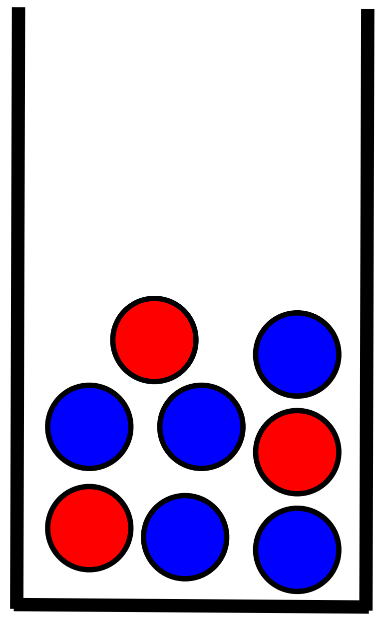
$B_t =$ Actual number of red balls at time t

Azuma's inequality: Martingale B_i and bounded change $|B_i - B_{i-1}|$ implies tail exponential decay

Application to our problem: $|B_i - B_{i-1}|$ is $O(\sqrt{t/i})$

$$\mathbb{P}(|B_t - B_0| \geq x) < 2e^{-\frac{x^2}{2\sum_i c_i^2}}$$

step 2 - High probability convergence (2/3)



Generate new ball, red or blue

(r, 1-r)

Pick existing ball uniformly from urn

Same color - add a copy

Different color -

with probability ρ add a copy

with probability $1-\rho$ retry

Application to our problem: |B(i)-B(i-1)| is O(sqrt(t/i)).

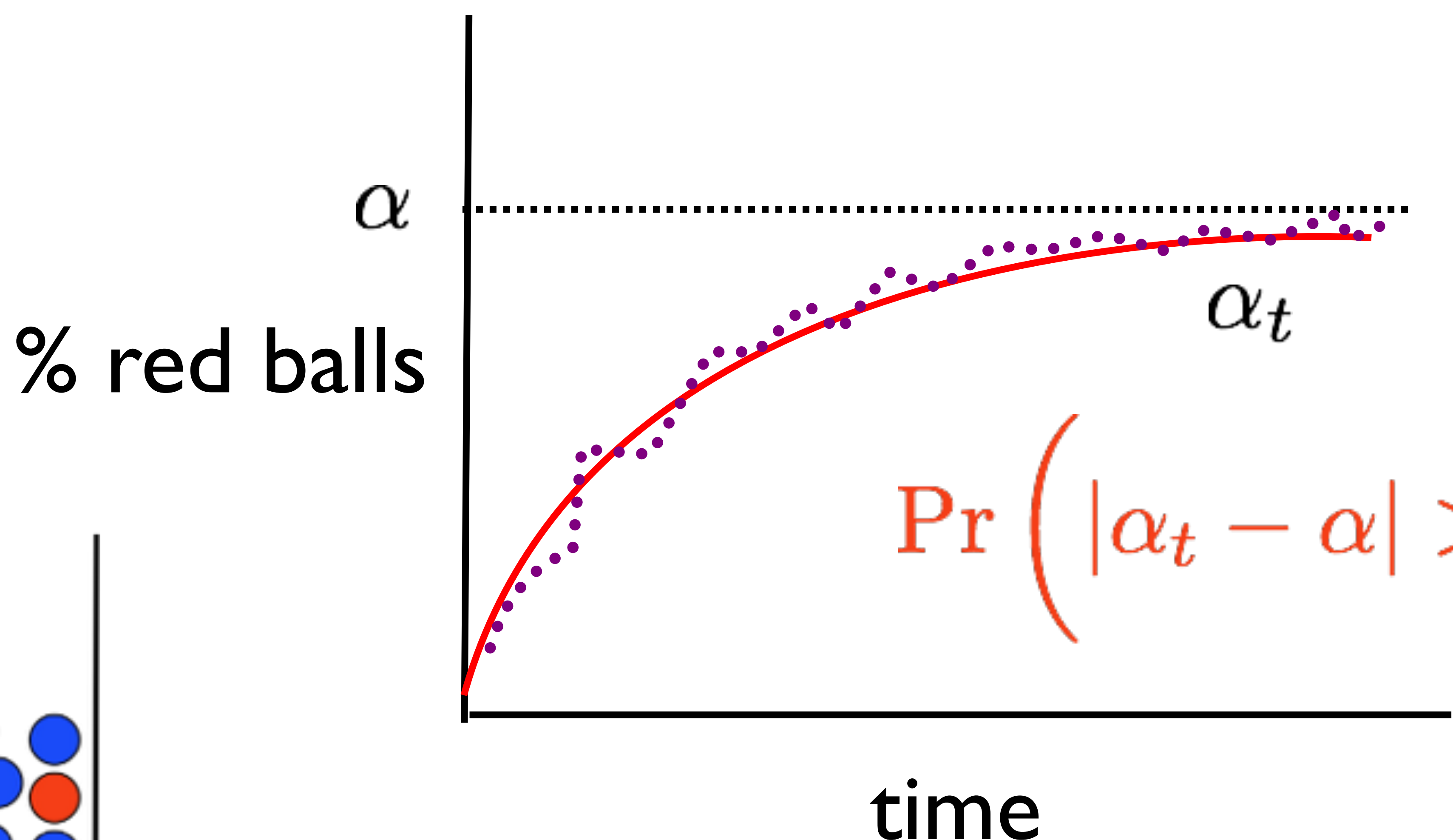
To compute |B(i)-B(i-1)|: analyze z(i)= expected number of additional red balls at times [i,t] given that there is one additional red ball at time i

Note: each additional red ball creates an **independent** effect on the color of its descendants (copies of the ball)

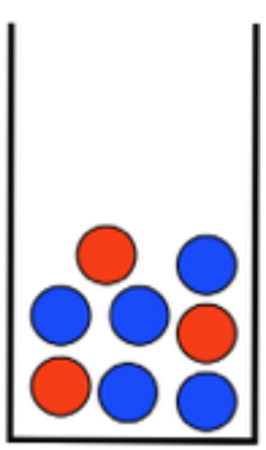
$$z_i = 1 + \frac{\gamma}{2(i+1)} z_{i+1} + \frac{\gamma}{2(i+2)} z_{i+2} + \cdots + \frac{\gamma}{2t} z_t$$

step 2- high probability convergence (3/3)

$$\alpha_t = \frac{R(t)}{2t} \qquad \alpha = \lim_{t\to\infty} \mathbb{E}[\alpha_t]$$



$$\Pr\left(|\alpha_t - \alpha| > \frac{1}{\sqrt{t}}\right) < \frac{1}{t^4}$$



step 3- Degree distribution (1/3)

extend analysis of degree distribution of
preferential attachment graph:

red node with degree k at time t

$$(m_{k,t}(R) = \#(\text{red nodes of degree } k \text{ at time t}))$$

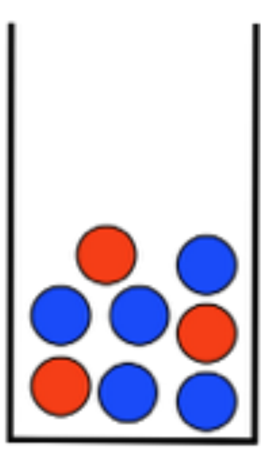
comes from

red node with degree k at time t-1

or

red node with degree k-1 at t-1

Homophily, *r*, *k* and % of red nodes play a role



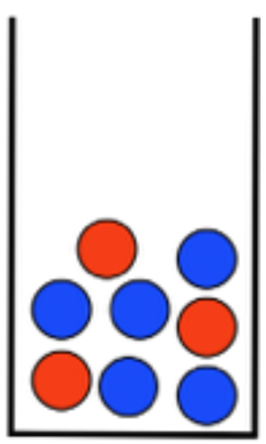
$$\mathbb{E}(m_{k,t}(R)|\text{history}) = \ldots m_{k,t-1}(R) \ldots m_{k-1,t-1}(R) \ldots$$

step 3- Degree distribution (2/3)

The expected degree distribution of both the blue nodes and the red nodes follows a power law with **different** parameters:

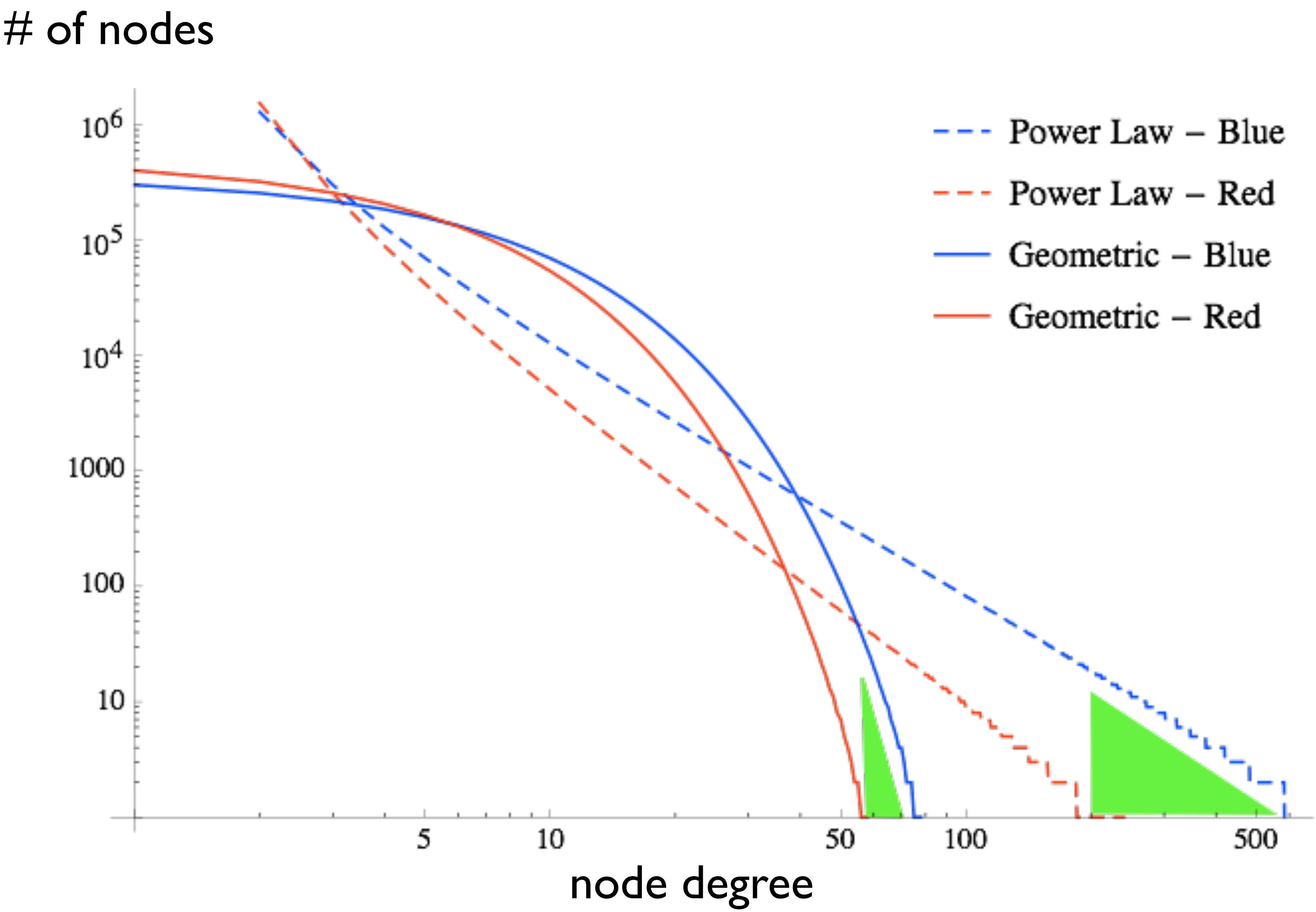
$$M_k(\mathtt{R}) \propto k^{-\beta(\mathtt{R})}$$

$$M_k(\mathtt{B}) \propto k^{-\beta(\mathtt{B})}$$



$$\beta(\mathtt{B}) > \beta(\mathtt{R})$$

step 3- Degree distribution (3/3)



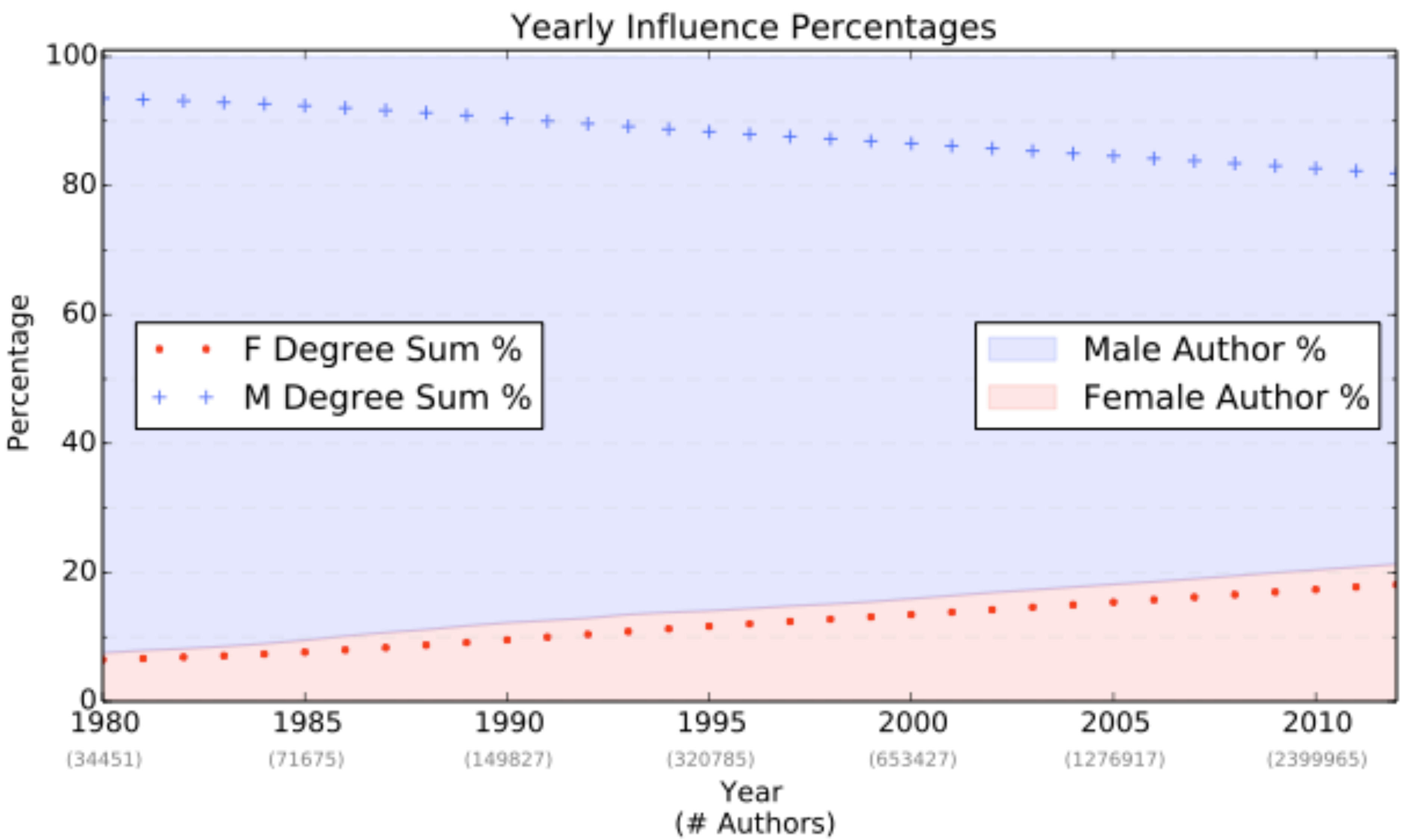
$$\lim_{G\to\infty} \frac{\mathrm{top}_k(\mathtt{R})}{\mathrm{top}_k(\mathtt{B})} \longrightarrow 0$$

Compare with empirical data :

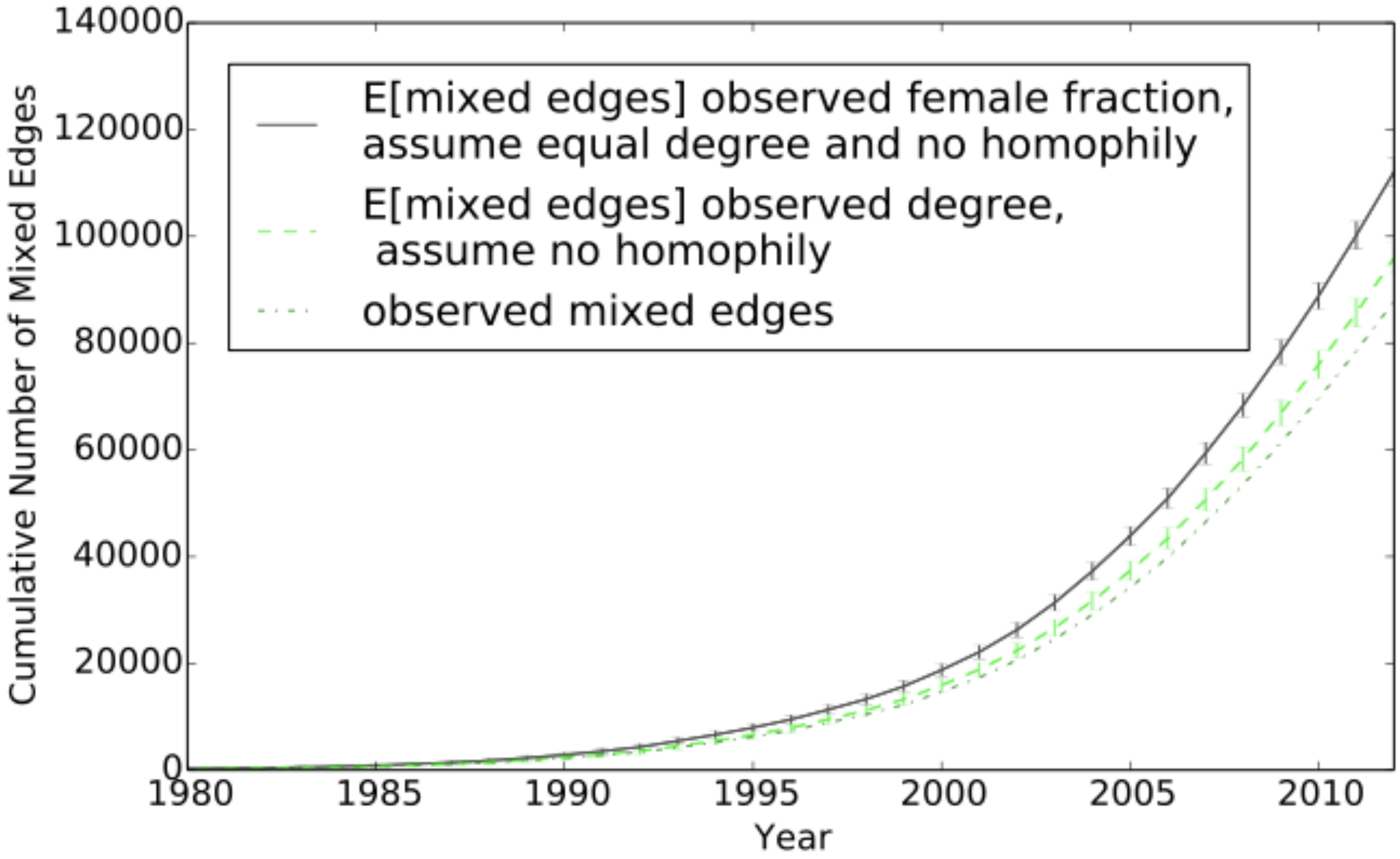
Any connection to real life? -

Student-advisor links in DBLP

Are women a minority in DBLP?

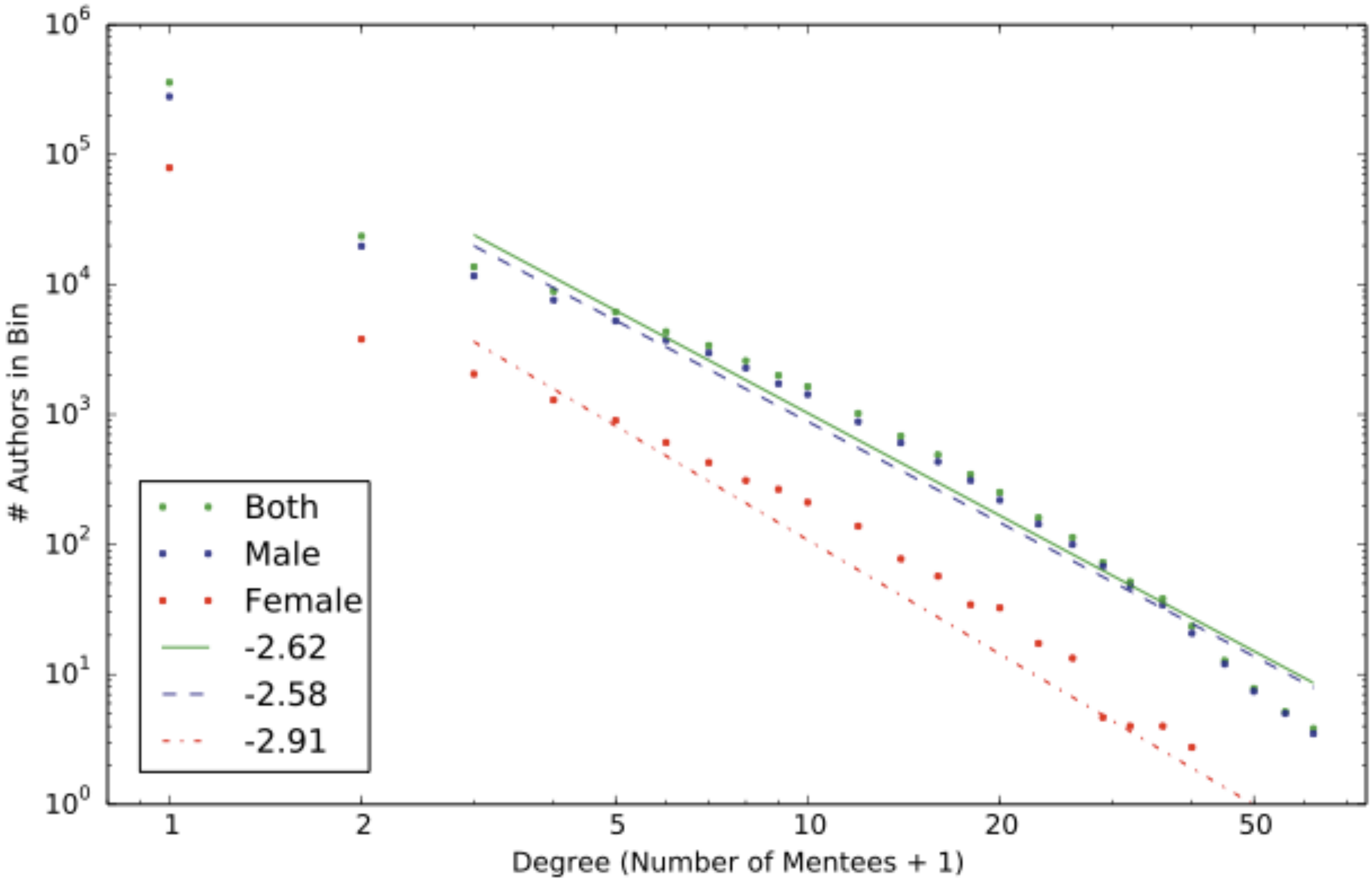


Is there homophily in DBLP?

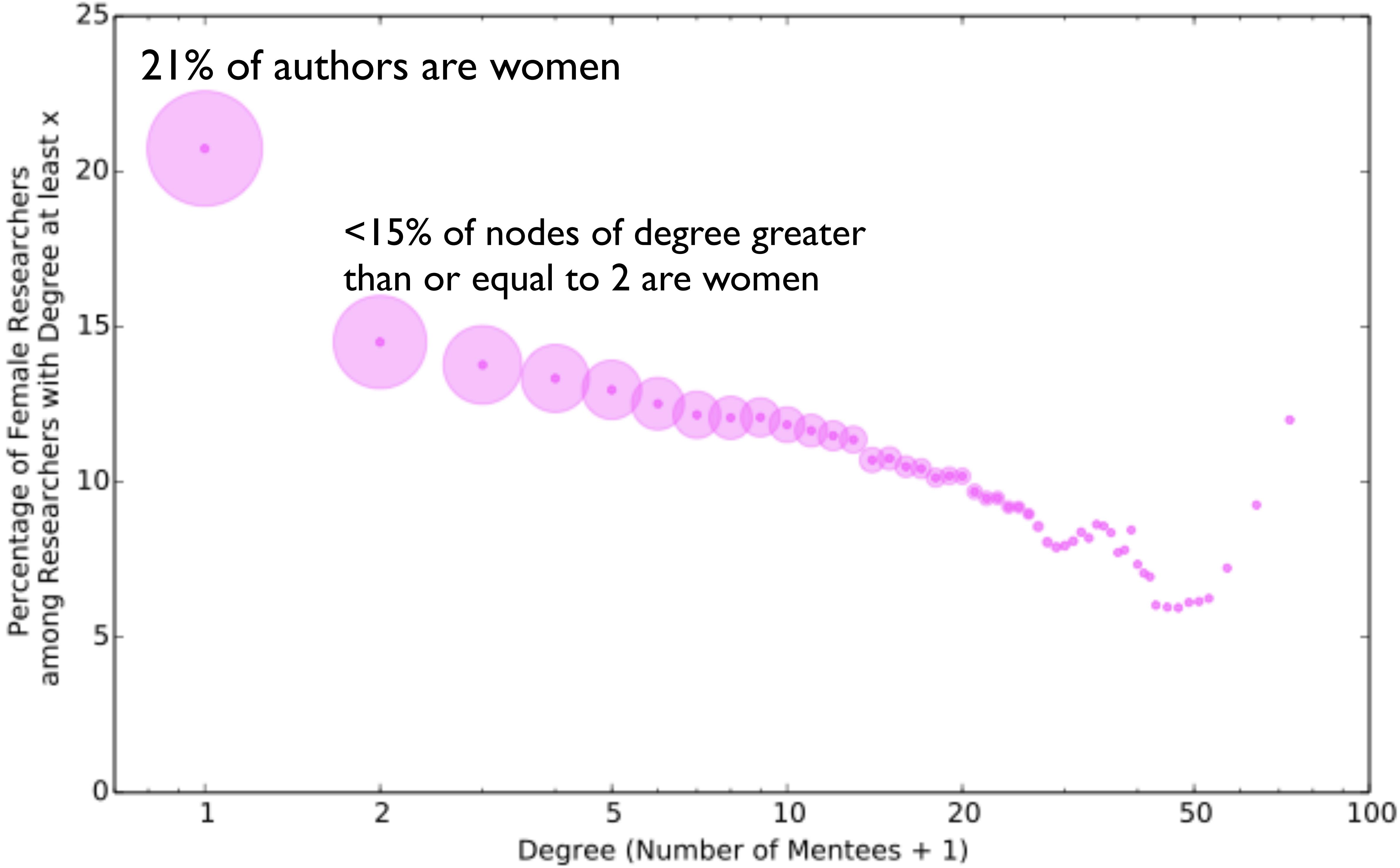


Homophily: unconditional or conditional on total degree

Is there preferential attachment in DBLP? Distribution of degrees



Is there a glass ceiling in DBLP?



Q: How does one extract gender from DBLP?

A: To assign a gender to the authors in DBLP we looked up their first name in our dictionary. If the probability of the name being female or male was over 90%, then the corresponding gender was assigned to the author.

Q: How does one extract the student-advisor network from DBLP?

A: A person was only considered as a potential mentor of a mentee if the difference in the number of years between the dates of their first articles exceeded four.

Notes

1. no minority?

2. no homophily?

3. no preferential attachment?

4. unequal rate and unequal qualification?

5. “leaky pipeline”?