Social and Information Networks 2: Evolution

Internet Analytics (COM-308)

Prof. Matthias Grossglauser
School of Computer and Communication Sciences

EPFL

Overview

- Herding and “watching thy neighbor”
  - Information cascades: why imitating your friends makes sense – and how it can lead to surprising group behavior
  - Heavy-tailed degree distributions: “the rich get richer” applied to networks
- Observing network properties
  - The importance of the observer
  - Example: your friends are more popular than you!

Watching thy Neighbor?

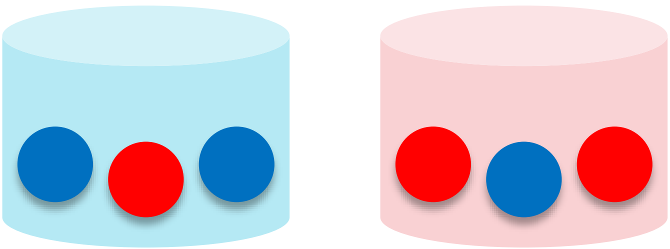
- Human decision-making:
  - Primary private information…
  - But also influenced by what decisions taken by others
- Reason:
  - Primary information: often too voluminous, noisy, not trustworthy,…
  - By imitating others, piggyback on their effort to interpret primary information
- Question:
  - Macro behavior of such systems?

Herding and information cascades

- Assumptions:
  - Decision: choose a restaurant, adopt new technology, political position, fashion,…
  - Sequential, and each person can observe choices made earlier by others
  - Each person has some private information to help guide decision: favorite food, taste,…
  - Private information not observed by others (can't see what others "know"), but decisions/actions are (can see what others "do")

Herding: how it can go wrong

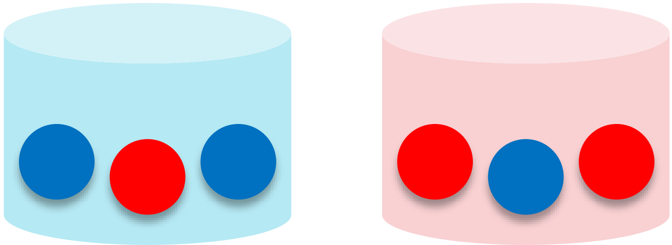
- Urn with 3 balls
  - A priori distribution (blue/red majority) = (0.5,0.5)
  - majority blue: 2 blue + 1 red
  - majority red: 2 red + 1 blue



- A group of people take turns:
  - Draw a ball from the urn at random
  - Check the color of the ball privately, put it back in urn
  - Announce their **guess** (blue/red majority) to everybody
  - Receive **reward** for correct guess

Herding: how it can go wrong

- Experiment at U. Virginia (1997) with students:
  - Subjects not told how to behave, just rules & payoff
  - Runs of 15 steps
  - Result: most runs (41/56) resulted in a cascade, ie, everyone after a while guesses the same thing
- Model for two extremes:
  - ❶Each individual is altruistic: do what allows others to make best guess
  - ❷Each individual is selfish = tries to make best guess for himself



Urn model: altruistic (➊)

- Every person:
  - Selects a ball at random (with replacement)
  - Announces the color of the ball to everybody as their guess, even if previous information suggests a different guess
- As $n \to \infty$, majority color of urn is equal to color most frequently observed
  - Consequence of law of large numbers
- After a few “sacrifices”, everybody could produce best guess
  - Sacrifice in the sense that individuals might be forced to say red (color of their ball) even if previous information suggests blue majority

Urn model: selfish (❷)

- Sequential decision-making
  - Selfish guess: use previous public and new private information to maximize own reward
  - Observed color remains private
- First individual:
  - Blue ball: announce guess(1) = blue
  - Red ball: announce guess(1) = red
  - Public guess of first fully reveals private information
- Second individual:
  - If color(2) = guess(1): announce this color
  - If color(2) $\neq$ guess(1): does not matter (assume color(2))
  - Public guess of second fully reveals private information

Urn model: selfish (❷)

- Third individual:
  - If guess(1) ≠ guess(2): announce guess(3) = color(3)
  - If guess(1) = guess(2):
    - Announce guess(3)=guess(2)=guess(1), regardless of color(3)
  - Why is this?
    - Person 3 knows that guesses 1+2 reveal perfect information
    - Therefore, regardless of color(3), guess(1)=guess(2) dominates guess
- Fourth,…,∞th individual:
  - If guess(1) = guess(2):
    - Announce guess(i) = guess(2)=guess(1), regardless of color(i)

Urn model: (❷) leads to cascade

- If guess(1) = guess(2) were both wrong, then all future guesses are wrong!
- This happens with prob. 1/9
- Even though each individual is using available information in the best way to make a guess
- But long-term total payoff would be maximized if everyone made their private information public → ie, early subjects "take one for the team"

Information cascade: suboptimal decision

- Cascade: sequential decisions
- Individual:
  - Efficiency gain by observing others' decisions
- Global behavior:
  - Primary information can "wash out"
  - Suboptimal or random decisions
- Might these be cascades:
  - Stock market gyrations, "flash crash"
  - Inexplicable shifts in popularity of {restaurants, clubs, celebrities,…}
  - Fashion, style, celebrity,…
  - …

Herding in networks

- Observation:
  - Degree distributions in networks often resemble power laws
- Power law:
  - $P(D > d) \propto d^{-\gamma}$
- Most distributions have "light tails":
  - $P(D > d) \propto e^{-\alpha d}$ (or lighter/bounded)
  - Exponential, Geometric, Gaussian, Poisson, …

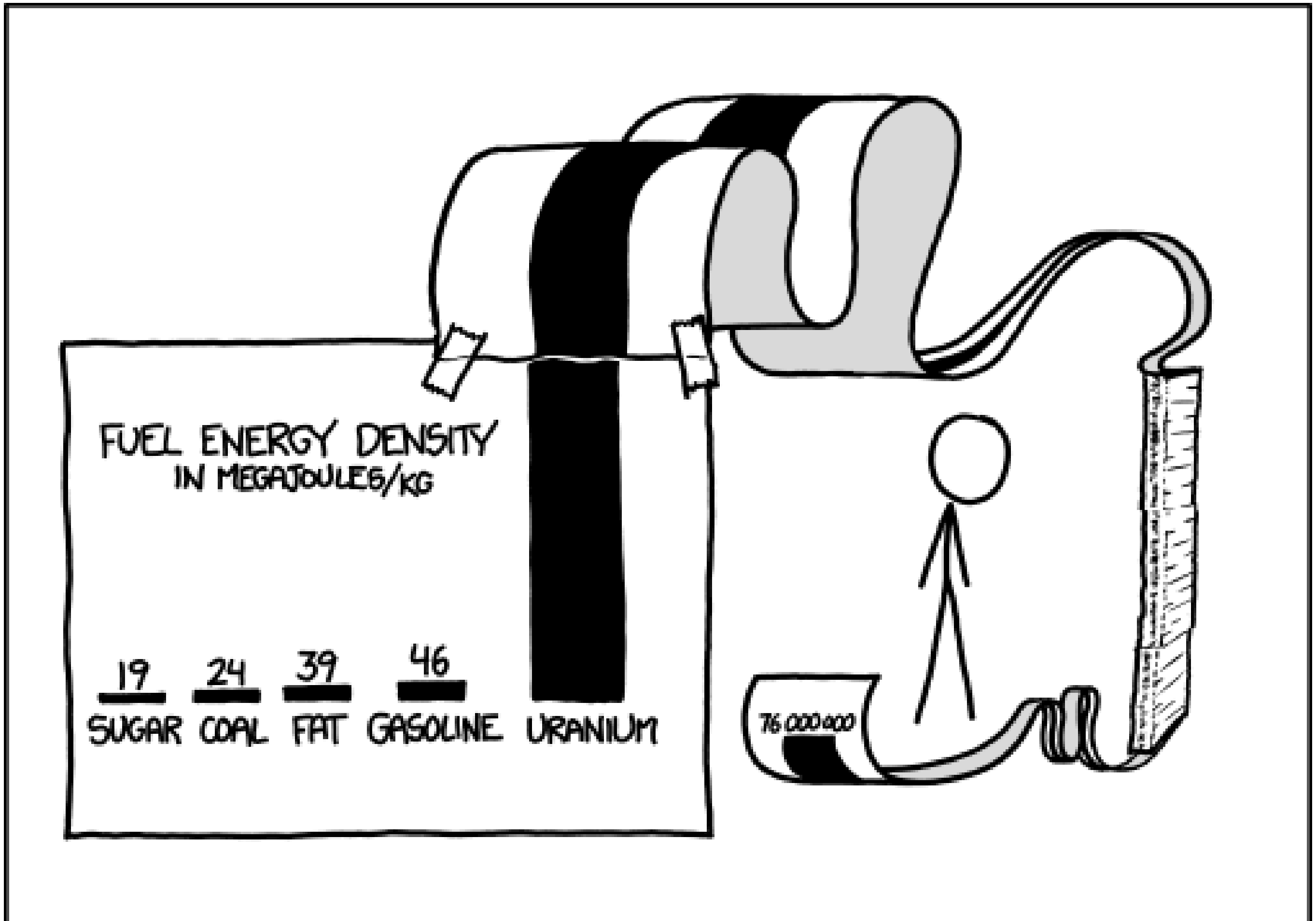
Pareto (β, γ) distribution

- Support: $d \in [\beta, \infty)$
- CCDF (Complementary Cumulative Distribution Function):
  - $P(D > d) = 1 - F_D(d) = \begin{cases} \left(\frac{d}{\beta}\right)^{-\gamma}, & d \geq \beta \\ 1 & \text{otherwise} \end{cases}$
  - γ: exponent, also called "Pareto index"
- Moments:
  - $E\left[D^k\right] = \begin{cases} \frac{\beta^k \gamma}{\gamma - k}, & k < \gamma \\ \infty & \text{otherwise} \end{cases}$

Numerical comparison exp/power

- Distribution of human height:
  - Mean = 178 cm
  - Stddev = 8 cm
- Compare tails: how tall are extremely tall people?
  - What is d^* such that $P(D > d^*) = 10^{-9}$
- Normal $N(178\text{cm}, (8\text{cm})^2)$:
  - d^* = 226 cm
- Pareto: choose β, γ s.t. first and second moments match data
  - $\gamma \cong 23, \beta \cong 170\text{cm}$
  - d^* = 420 cm !!
- Assumption very important for extremal values!

Log-log plot

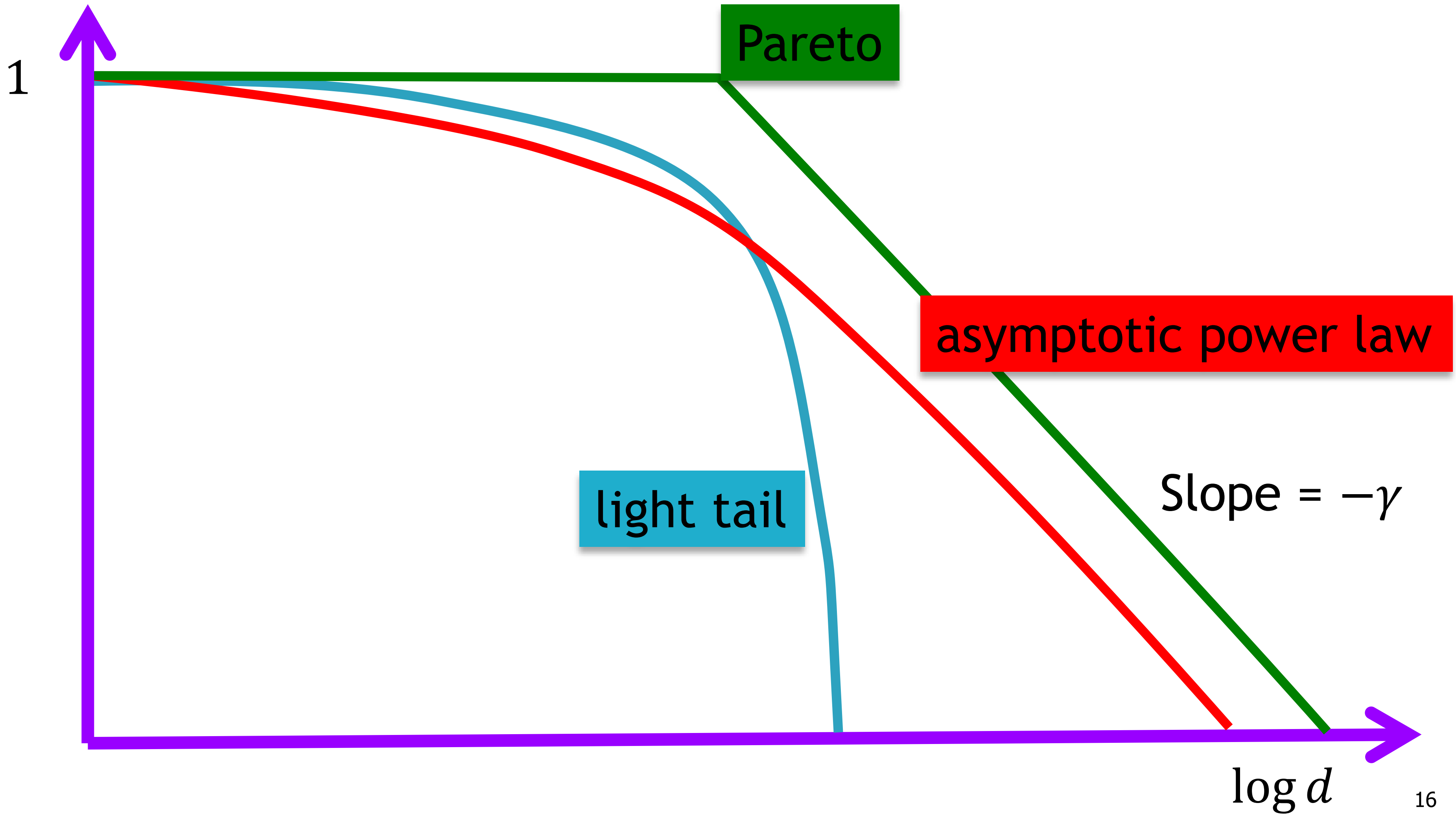


SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT *PROPERLY*.

Source: xkcd #1162

Log-log distribution plot

- $\log P(D > d)$



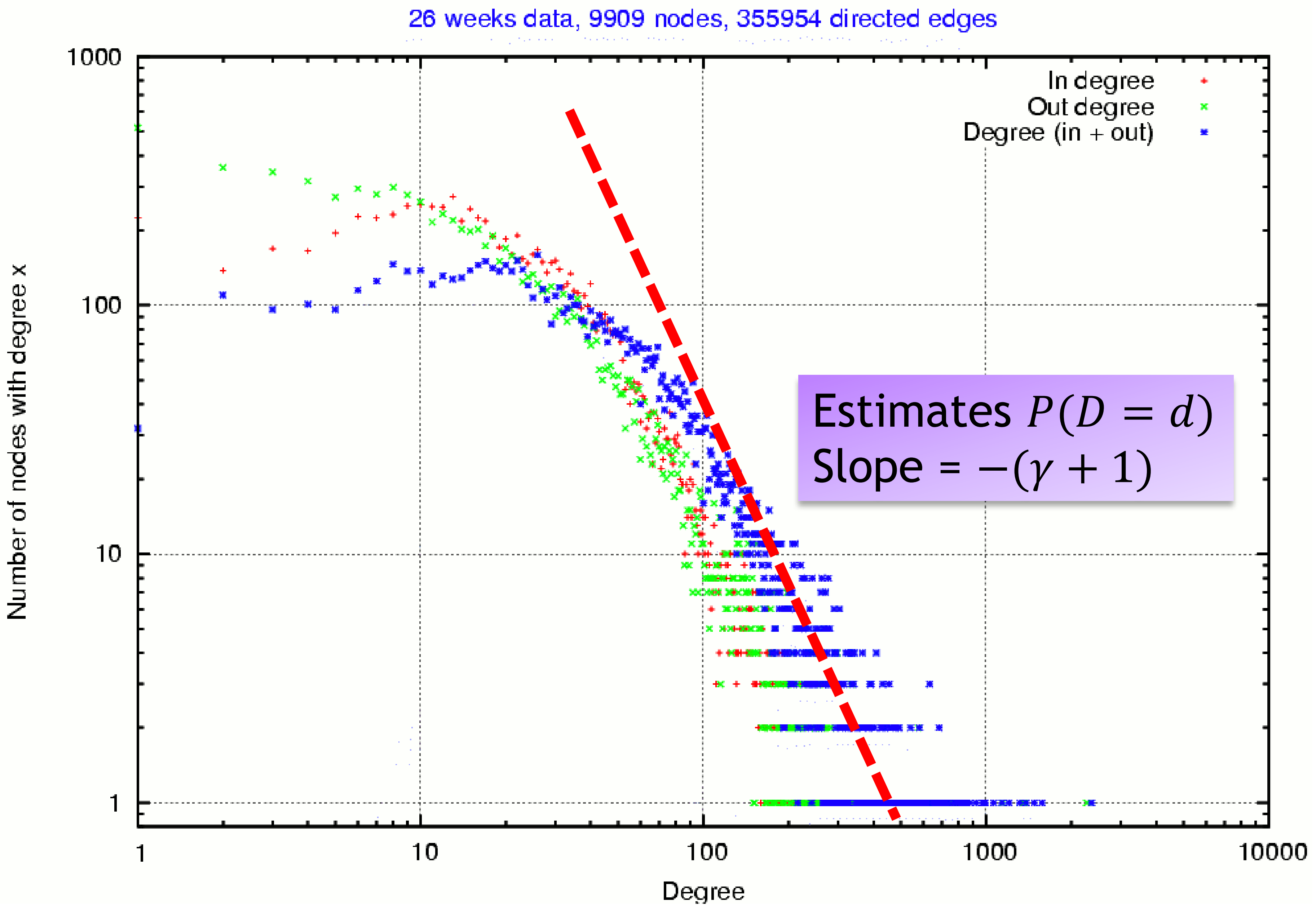
Examples of observed power laws

- File sizes on a computer
- Stock market crashes
- Sizes of cities
- Phone call length
- Wealth & income distribution
- Sizes of floods
- Popularity of web pages
- Word frequencies in prose
- Degree distribution in social networks
- …

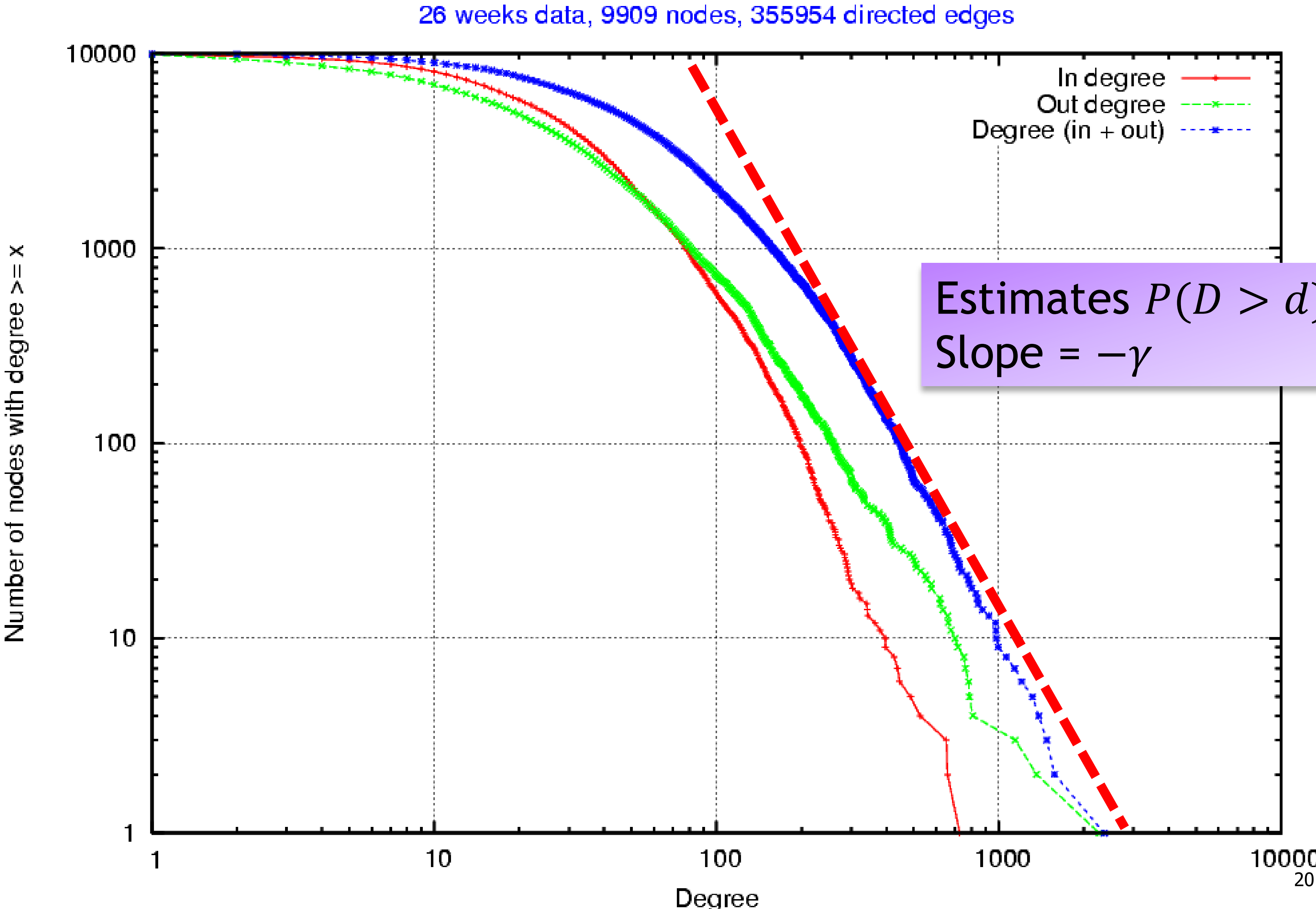
Why worry about the tail?

- Would you like to sit on a plane engineered under a Gaussian assumption for turbulence? 😲

Log-log distribution plot



Log-log cumulative plot

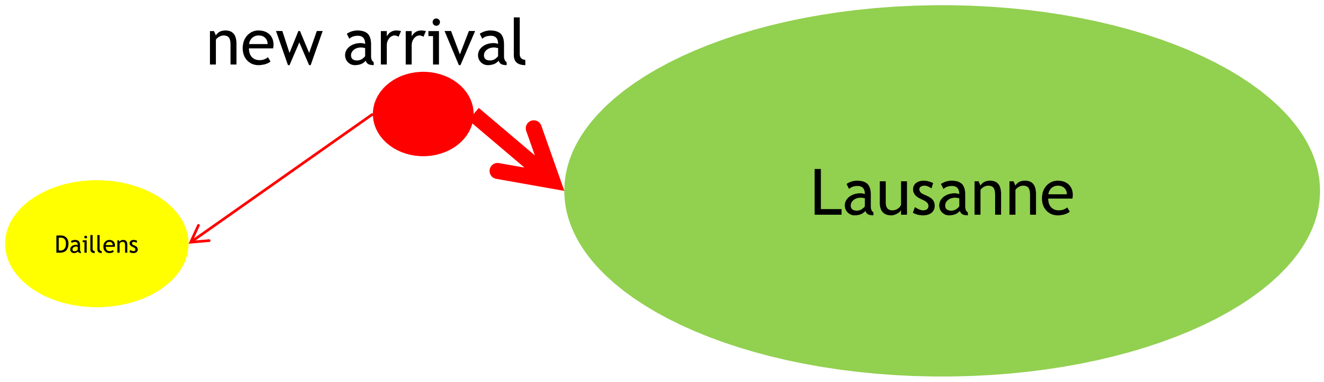


Estimates $P(D > d)$

Slope = $-\gamma$

One explanation: the rich get richer

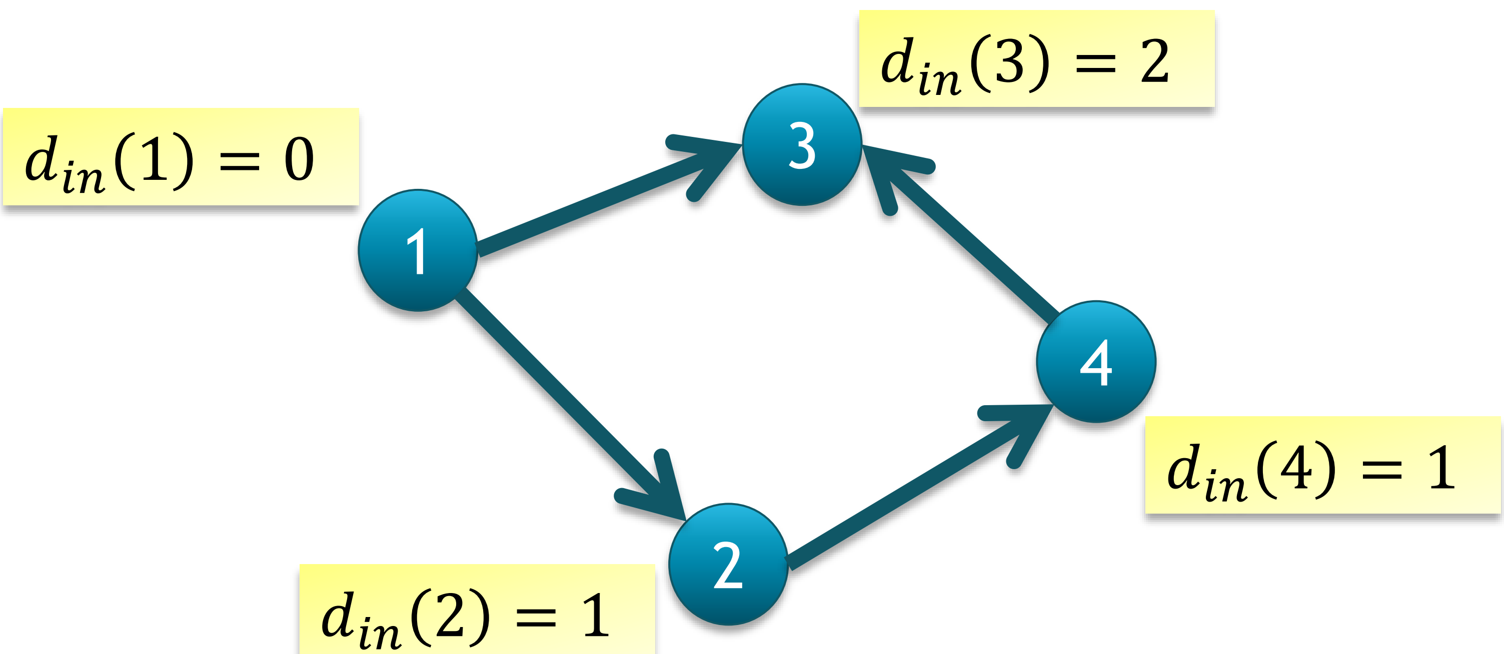
- New arrival in our region: move to Daillens or Lausanne?
  - More likely Lausanne, because **more people already there**
- City size distribution after many arrivals?



- Also: “the first million is the hardest” ;-)

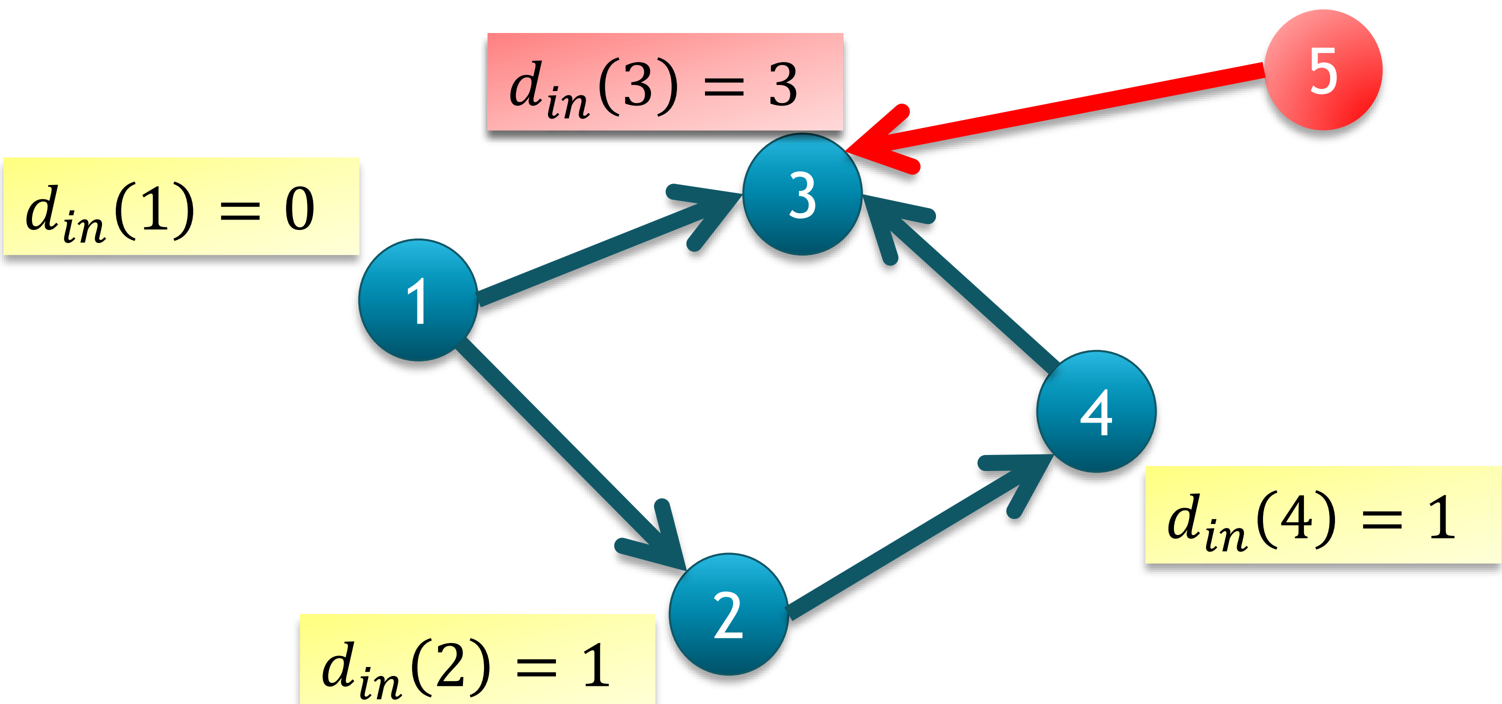
Preferential attachment in growing nets

- Growth model: nodes arrive one by one and join the existing network
  - Directed graph
  - In-degree $d_{in}(v)$ measures "popularity" and "attractiveness" of a node



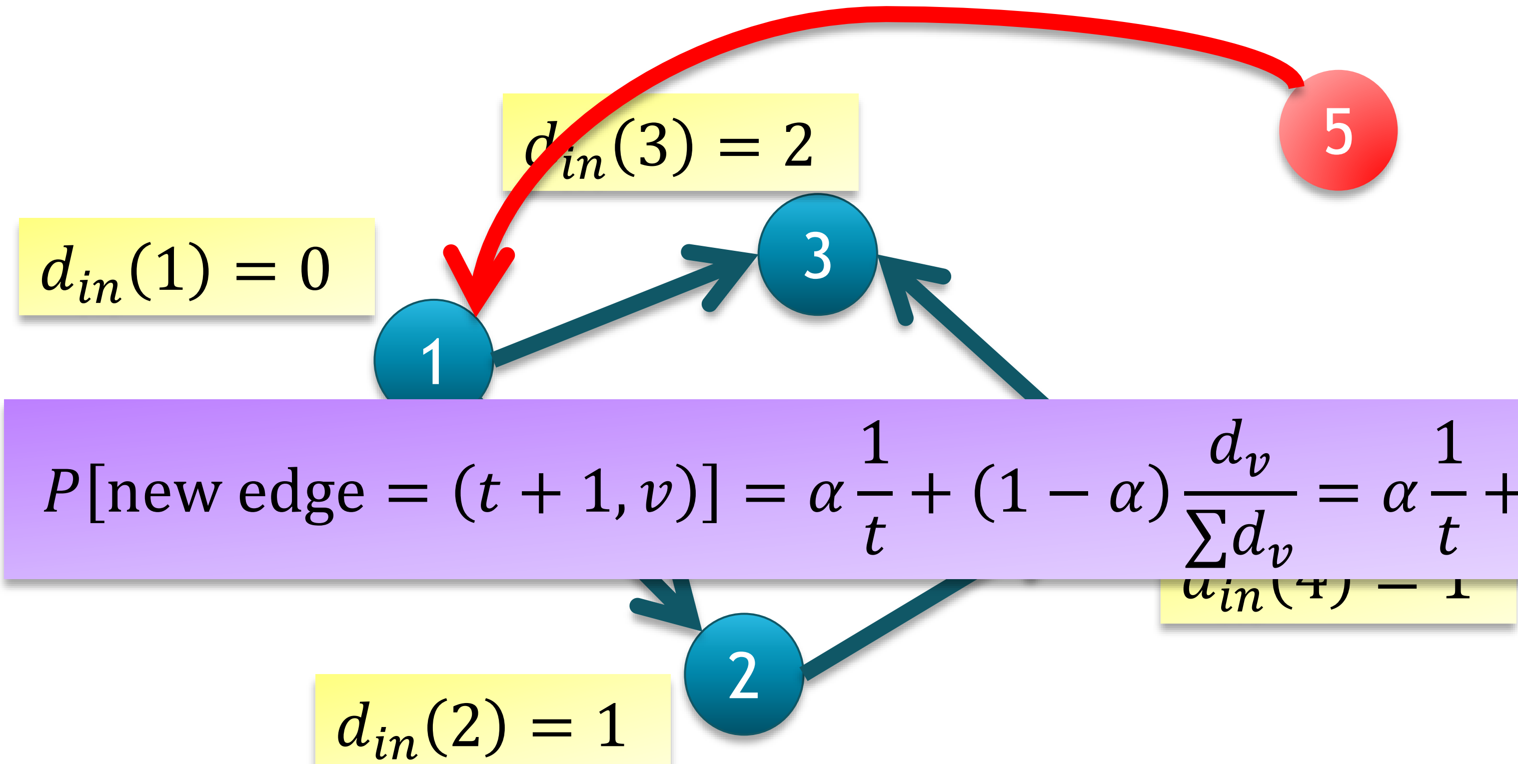
Preferential attachment

- Preferential attachment: new node creates one edge
- Prob. of connecting to v is $\propto d_{in}(v)$
  - Intuition: high-degree easier to meet; more popular; more useful;…



Preferential attachment

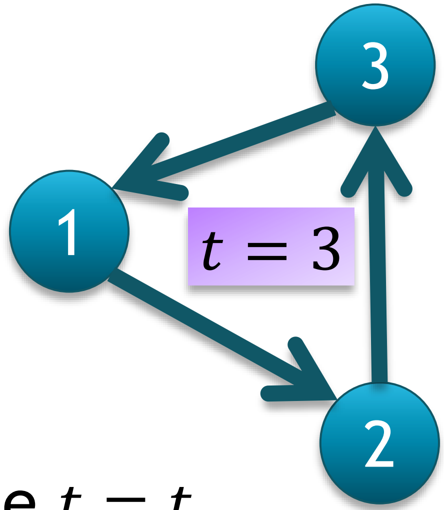
- Node with in-degree 0 never gets “started”
- Need another assumption:
  - With prob. α, new node connects uniformly at random
  - With prob. $(1-\alpha)$, preferential attachment



$$P[\text{new edge} = (t+1, v)] = \alpha\frac{1}{t} + (1-\alpha)\frac{d_v}{\sum d_v} = \alpha\frac{1}{t} + (1-\alpha)\frac{d_v}{t}$$

Preferential attachment: analysis

- Evolution of this system:
  - Graph structure only matters through in-degrees
  - Markov chain $\{X_j(t)\}$: # nodes with in-degree = j at time t
- Initialize appropriately:
  - For example:



  - Total # of nodes and edges at time $t = t$
  - Notation: $X_j \coloneqq X_j(t)$

Preferential attachment: analysis

- Drift:
  - $P\big(X_j(t+1){=}X_j(t)+1\big) = \alpha\frac{X_{j-1}}{t} + (1-\alpha)(j-1)\frac{X_{j-1}}{t}$

    prob. of selecting a degree-$(j-1)$ node uniformly at random

    prob. of selecting a degree-$(j-1)$ node with prob. ∝ degree

  - $P\big(X_j(t+1){=}X_j(t)-1\big) = \alpha\frac{X_j}{t} + (1-\alpha)j\frac{X_j}{t}$

    prob. of selecting a degree-j node uniformly at random

    prob. of selecting a degree-j node with prob. ∝ degree

Preferential attachment: analysis

- Combined drift (pretend $X_j, t \in \mathbb{R}$)
  - $$\frac{dX_j}{dt} = \frac{\alpha(X_{j-1} - X_j) + (1-\alpha)((j-1)X_{j-1} - jX_j)}{t}$$
- Assume as $t \to \infty$, degree sequence converges $y_j \triangleq \frac{X_j}{t}$, then solve for y_j:
  - y_j: fraction of nodes with degree j,i.e., empirical distribution
  - $y_j = \alpha(y_{j-1} - y_j) + (1-\alpha)((j-1)y_{j-1} - jy_j)$
  - $$\frac{y_j}{y_{j-1}} = \frac{\alpha + (1-\alpha)(j-1)}{1 + \alpha + (1-\alpha)j} = 1 - \frac{2-\alpha}{1+\alpha+(1-\alpha)j}$$
  - Asymptotically for large j, this is $\cong 1 - \frac{2-\alpha}{1-\alpha} j^{-1}$

Preferential attachment: analysis

- Note that $\left(\frac{j}{j-1}\right)^{-(\gamma+1)} = \left(1-\frac{1}{j}\right)^{\gamma+1} \sim 1-\frac{\gamma+1}{j}$
- So $\gamma = \frac{2-\alpha}{1-\alpha} - 1 = \frac{1}{1-\alpha}$
- Putting together:
  - $\frac{y_j}{y_{j-1}} = \left(\frac{j}{j-1}\right)^{-(\gamma+1)}$, hence
- $y_j \propto j^{-(\gamma+1)}$: **asymptotic power law**
- The stronger the preferential attachment (α smaller), the "heavier" the tail of the degree distribution (γ smaller)
- Arguments can be made rigorous

Network effects and "winner-takes-all"

- Other examples of "rich-get-richer" phenomena:
  - Facebook vs {friendster, sixdegrees, xing,…}
  - Android vs iPhone
  - Technology standards: BluRay,…
- Metcalfe's Law:
  - The value of a network is proportional to n^2
  - Because the value to an individual is proportional to n
- Lock-in
  - Being early is very important

Observer: Friendship Paradox

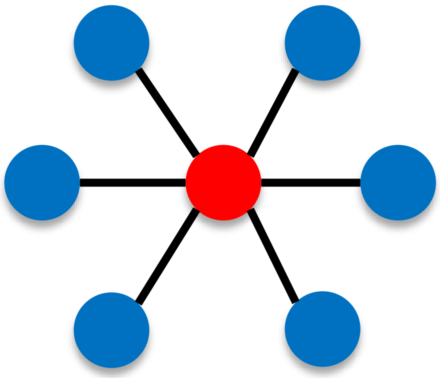
- “Your friends have more friends than you”
- Experiment:
  - Get on facebook and compute the average # friends of your friends
  - How does this compare to your own # friends?

Friendship Paradox

- Formally:
  - Social network = $G(V, E)$
  - d_v: degree of node v
  - $n = |V|$: number of nodes, $m = |E|$: number of edges
- Average number of friends: $\mu = \frac{\sum d_v}{n}$
- How to talk about average number of friends' friends?
  - Natural measure: $\frac{1}{n}\sum_{u \in V} \frac{1}{d_u} \sum_{v \in N_u} d_v$
  - Easier to analyze: degree "seen" by random edge

Friendship Paradox

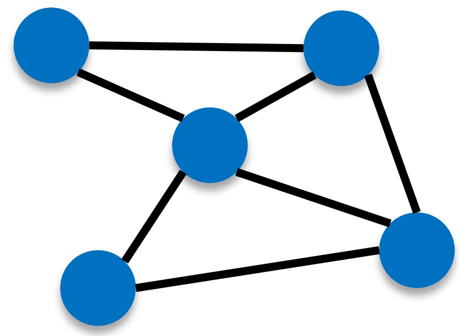
- Star network ($|V| = n$):



- Avg degree: $\frac{1}{n}\big((n-1) + (n-1)\big) = \frac{2(n-1)}{n} \to 2$
- Avg degree of neighbors:
  $\frac{1}{n}\big((n-1)^2 + 1\big) \to n$
- Degree of random edge: $\frac{1}{2}(n-1) + \frac{1}{2} = \frac{n}{2}$

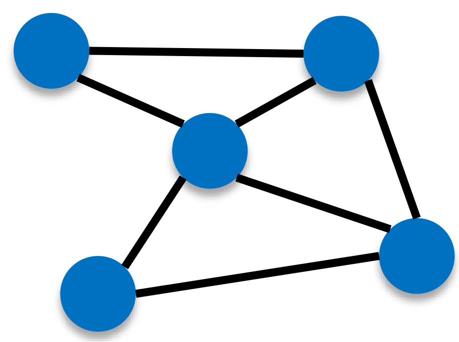
Sampling nodes vs sampling edges

- Average degree over nodes:
  - $\mu = \frac{\sum d_v}{n} = 2\frac{m}{n}$



"expected degree of a random person"

- Average degree over edges:
  - $\frac{\sum_{(u,v)\in E} d_v}{2m}$



"expected degree of a random friend of a random person"

Friendship Paradox

- Lemma:
  - $$\frac{\sum_{(u,v)\in E} d_v}{2m} = \mu\left(1 + \frac{\sigma^2}{\mu^2}\right)$$
  - Degree (empirical) variance:

$$\sigma^2 = \frac{1}{n}\sum_{v\in V} d_v^2 - \left(\frac{1}{n}\sum_{v\in V} d_v\right)^2$$
$$= \widehat{Var}[d_v]$$

Friendship Paradox

- Proof:
  - $\frac{\sum_{(u,v)\in E} d_v}{2m} =$
  - $= \frac{\sum_{v\in V} d_v^2}{2m} =$ (because v appears d_v times in sum over E)
  - $= \frac{\sum_{v\in V} d_v^2}{\mu n} =$ (because avg degree is $\mu = 2m/n$)
  - $= \frac{\sigma^2 + \left(\frac{1}{n}\sum_v d_v\right)^2}{\mu}$

Why is it important?

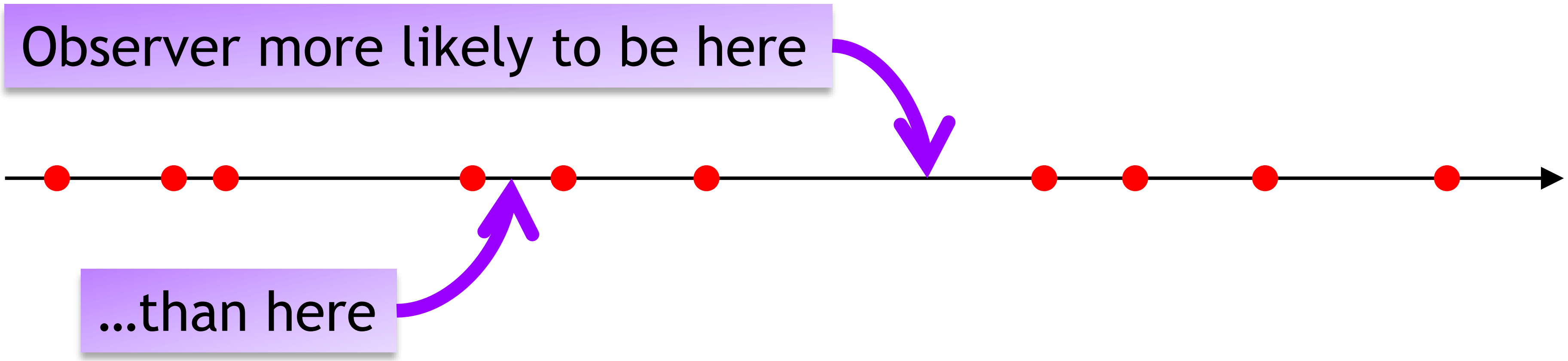
- Epidemiology:
  - Best protection for a population with a given budget?
  - Assume social network is not knowable globally
- Two strategies:
  - (a) immunize a random set of people
  - (b) immunize random friends of a random set of people
- Friendship Paradox:
  - (b) better than (a)!
  - Bias towards "higher-degree friends"
- Other applications:
  - Finding good monitors, trend-setters, etc.

The observer matters

- Other examples:
  - Occupancy distribution:
    - Suppose a train is full 50% of the time, and empty 50% of the time
    - Observer: train is full 100% of the time

The observer matters

- Waiting time:
  - Suppose buses arrive a Poisson(λ) point process
  - Average interarrival interval: $1/\lambda$
  - Observer point of view:
    - Residual time (until next bus): mean = $1/\lambda$
    - Since last bus: mean = $1/\lambda$
    - Mean observed interval length: $2/\lambda$!



Summary and lessons

- Herding
  - Following others' decisions: natural social mechanism, can lead to suboptimal global behavior
  - Information cascades: watching others can wash out primary information
  - Rich-get-richer: huge differences in {wealth/degree/influence/membership/…}, winner-takes-all markets
- Observing
  - Choice of observer – sampling bias
  - Paradox: average friend is more popular than average individual
- Next week:
  - Processes on networks: epidemics, sampling

References

- [D. Easley and J. Kleinberg, Networks, Crowds, and Markets (chapter 16), 2010]
- [Grossglauser & Thiran, COM-512: Models and Methods for Random Networks (class notes)]
- [Anderson & Holt, Information Cascades in the Laboratory, American Econ Review, 87(5), 1997]