# Unstructured P2P
# Link Lifetimes Redux

**Zhongmei Yao**[*], Daren B.H. Cline[+], and Dmitri Loguinov[+]

[*] The University of Dayton
[+] Texas A&M University

INFOCOM, Turin, Italy
4-17-2013

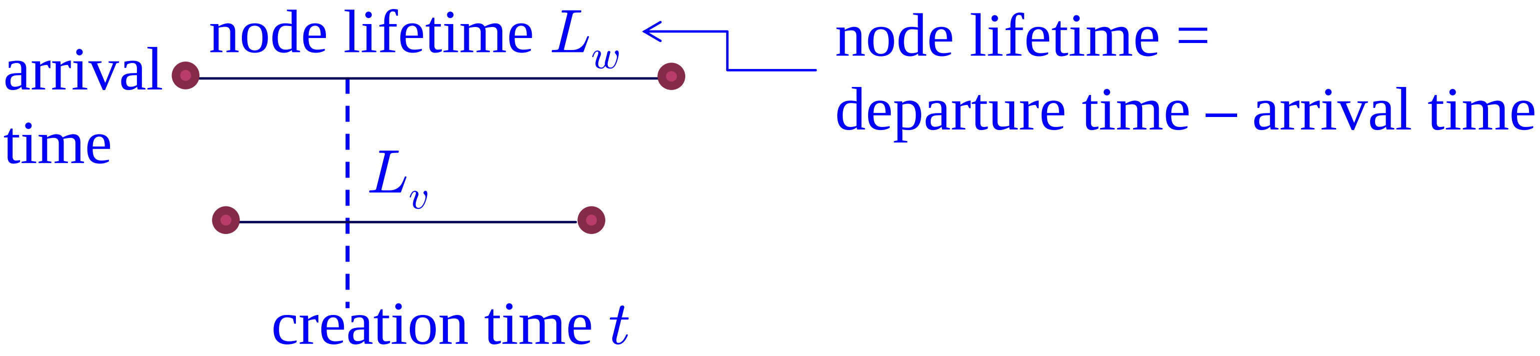
# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# Introduction

- A (virtual) link connects two end-points, *w* and *v*

  - Peer *w* is the initiator of this link
  - Node *v* is the recipient
  - Link is created at time $t$:



arrival time

node lifetime $L_w$

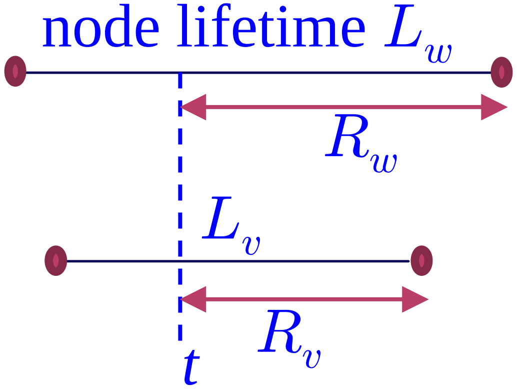
node lifetime = departure time – arrival time

$L_v$

creation time $t$

- Link lifetime determines how long a connection remains alive for forwarding pkts in highly dynamic P2P networks

# Introduction 2



- Two sides of a coin
  - From *w*'s perspective, link lifetime is equal to the remaining lifetime $R_v$ of node *v* since creation time *t*
  - From *v*'s point of view, it should be $R_w$
- Two link durations for each link $w \rightarrow v$
  - $R_v$ is termed out-link lifetime $V$, where *v* is an out-neighbor of *w*
  - $R_w$ is called in-link lifetime $W$
  - The link remains alive for min($V$,$W$)
- The transient node degree distribution depends on individual variables $V$ and $W$
  - Count how many out- and in-neighbors that a node has, as this node keeps alive in the system during its lifetime

# Motivation

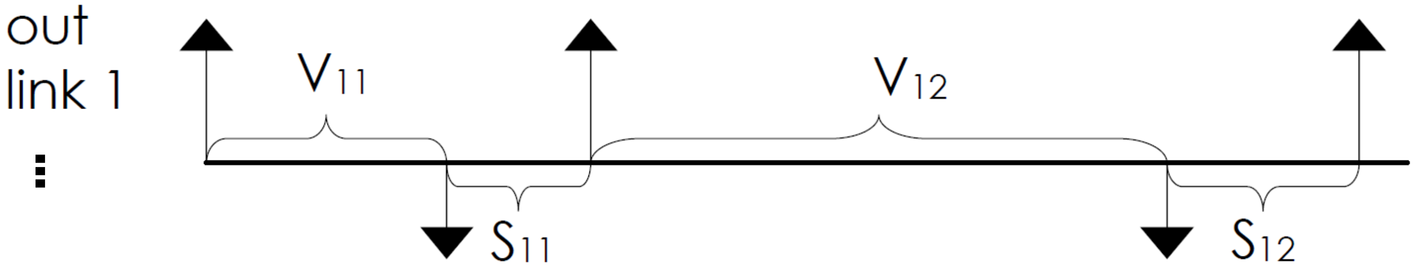
- In/out-link lifetimes are determined by
  - Node lifetime distribution $F_L(x)$ (X. Wang 2009)
  - Neighbor selection strategy
- Out-link lifetimes $V$ have been addressed under
  - Uniform selection (Yao 2006, Leonard 2007)
  - Max-age (Tan 2007, Yao 2009) and age-proportional (Yao 2009)
- Prior work focuses only on active systems
  - Failed neighbors are replaced with new ones
- Open issues
  - Properties of in-link lifetimes $W$ under *non-uniform* selection?
  - Since neighbor search requires substantial network resources, what is the performance of a *passive system*?
  - System performance is determined by in/out degree. What is the in/out degree distribution under *non-uniform* selection?

# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# Unifying Neighbor Selection Model

- Network model:
  - Consider $n$ participating nodes, where each node is either online or offline
  - Each user $i$ creates $k_i \geq 1$ outbound links upon arrival
- Active systems
  - Broken outbound connections are detected/repaired in random duration $S_{ij}$



- Passive systems
  - Only restrict neighbor search to the $k_i$ initial out-links
  - Failed neighbors are never replaced

# Unifying Neighbor Selection Model 2

- Implement arbitrary age-biased selection using a general neighbor preference function $p(x)$
  - User $w$ assigns non-negative weight $p(x)$ to a live user with age $x$
  - The probability that $w$ connects to a live peer $v$ is denoted by

$$c_N(v) = P(w \rightarrow v \mid A_1, \ldots, A_{N-1})$$

age of a live user

$N$: number of users currently alive

- <u>Assumption 1</u> (General Age-Biased Neighbor Selection):
  - As $N$ increases, $c_N(v)$ satisfies

$$\sum_{v=1}^{N} \left| c_N(v) - \frac{p(A_v)}{E[p(A)]N} \right| \rightarrow 0$$

  - The connection probability to $v$ is asymptotically proportional to $p(A_v)$, the weight assigned to $v$

# Preference Functions

- Assumption 1 covers:
  - Category 1, where $c_N(v)$ is proportional to function $p(x)$:

$$c_N^1(v) = \frac{p(A_v)}{\sum_{i=1}^{N} p(A_i)}$$

    - Uniform selection: $p(x) = 1,\ c_N(v) = 1/N$
    - Age-proportional: $p(x) = x,\ c_N(v) = A_v / \sum_{i=1}^{N} A_i$
  - Category 2, where $w$ randomly selects $m > 1$ users from the system into a set and then picks $s$-th order statistic (of the sampled ages) to identify the best neighbor

$$p_2(x) = m \binom{m-1}{s-1} F_A^{s-1}(x)(1 - F_A(x))^{m-s}$$

    - Max-age: $s = m$
    - Min-age: $s = 1$

$F_A(x)$: the distribution of user ages

# Preference Functions 2

- Property of heavy-tailed lifetime distributions
  - Older nodes have stochastically longer remaining lifetimes $R$
  - The age-proportional $p(x) = x$becomes unbounded in $x$ and in the extreme, users with large age (if there are only a few) may be overloaded
- The max-age $p(x) = m(F_A(x))^{m\text{-}1}$ is interesting
  - It favors older peers
  - It is viable (simply adjust the parameter $m$)
  - Unlike age-proportional, it is bounded in $x$
  - However, it is computation challenging – no simple closed-form results on metrics of interest
- We propose an approximation to max-age:
  - Weights are either 1 or 0: assign weight 1 to any user whose age is no smaller than $x_0$; otherwise assign 0
  - The step-function $p(x) = 1_{x>=x0}$

# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# Out-link Lifetimes

- Theorem 1: Under Assumption 1 and $n \to \infty$, the tail distribution of out-link lifetimes $V$ is

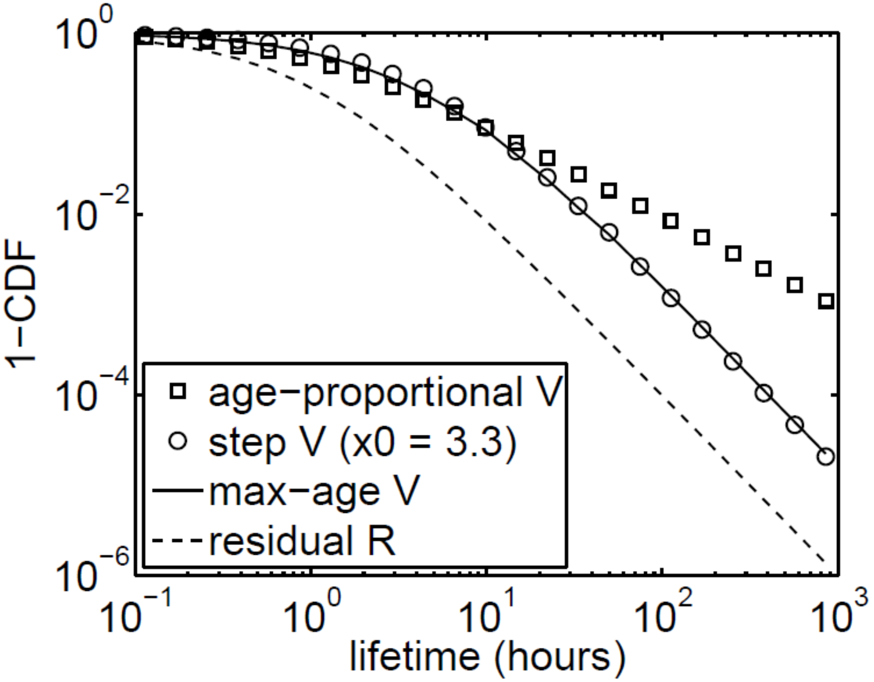
$$\bar{F}_V(x) = \frac{E[p(A-x)]}{E[p(A)]} = \frac{E[p(A-x) | A \geq x]}{E[p(A)]} \bar{F}_A(x)$$

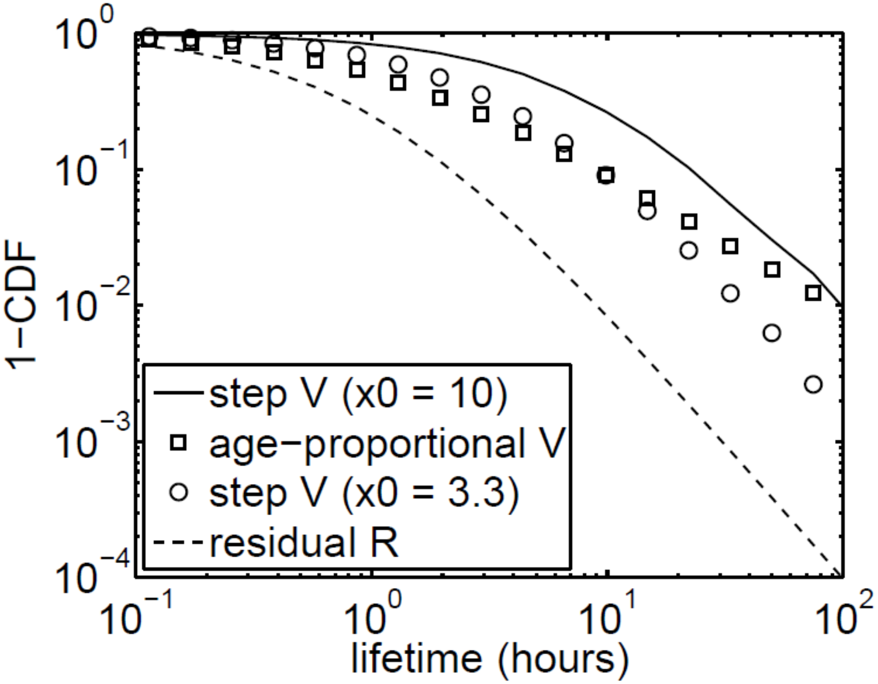
$$F_A(x) := P(A_i < x) = \frac{1}{E[L]} \int_0^x \bar{F}_L(y) dy$$

  - where $F_L(x)$ is the user lifetime distribution
  - Uniform selection: $F_V(x) = F_A(x)$
  - Age-proportional: $F_V(x) = \frac{1}{E[A]} \int_0^x \bar{F}_A(y) dy$
  - Step-function: $1 - F_V(x) = \frac{\bar{F}_A(x+x_0)}{\bar{F}_A(x_0)}$

# Out-Link Lifetimes 2

- Out-link lifetime tails for Pareto $L$ with shape $\alpha = 3$, $E[L] = 0.5$ hours



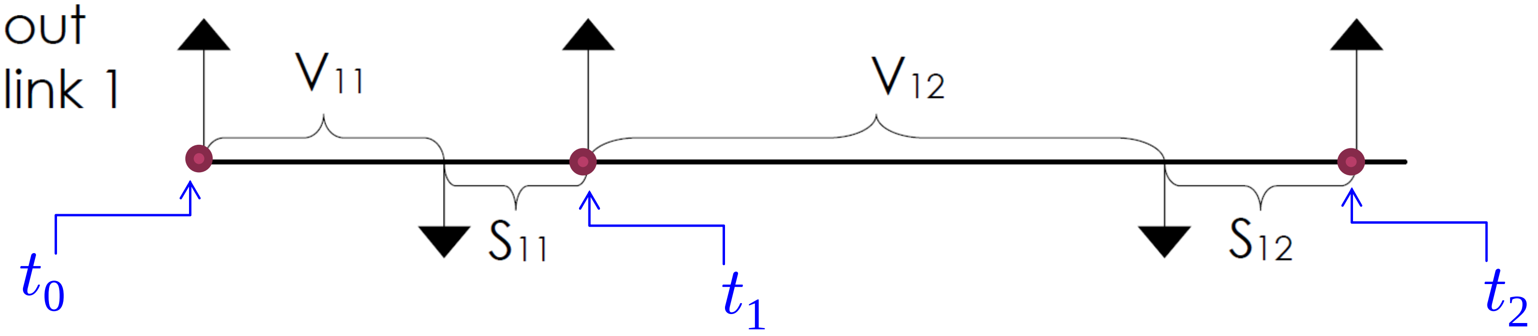


# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# Message Overhead

- Neighbor replacement (random walk or hop-limited flooding) consumes substantial network resources
- Let $t_j$ be the instance when a link gets $j$-th out-neighbor



- Let $U(t)$ count the number of neighbor searches in $[0, t]$

$$U(t) := \sum_{j=0}^{\infty} \mathbf{1}_{t_j \in [0,t]}$$

  - $u(t) := E[U(t)]$ is the renewal function
  - In passive systems without replacement, $U(t) = 1_{t \geq 0}$

# Message Overhead 2

- The mean number of neighbor searches performed by a user during its lifetime $L$ is

$$kE[u(L)]$$

  - There are $k$ initial neighbors
  - The mean number of replacement neighbors per $L$ is thus

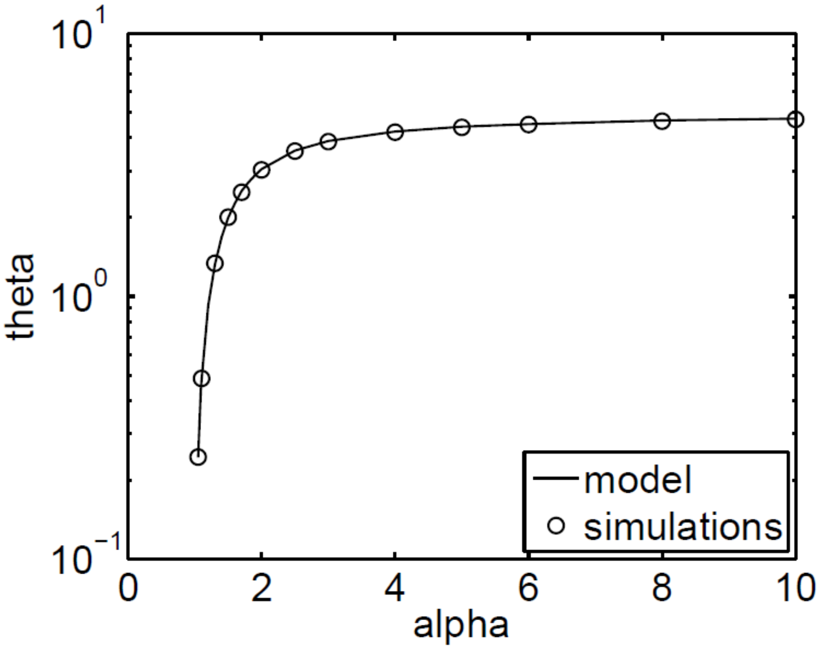
$$\theta := kE[u(L) - 1]$$

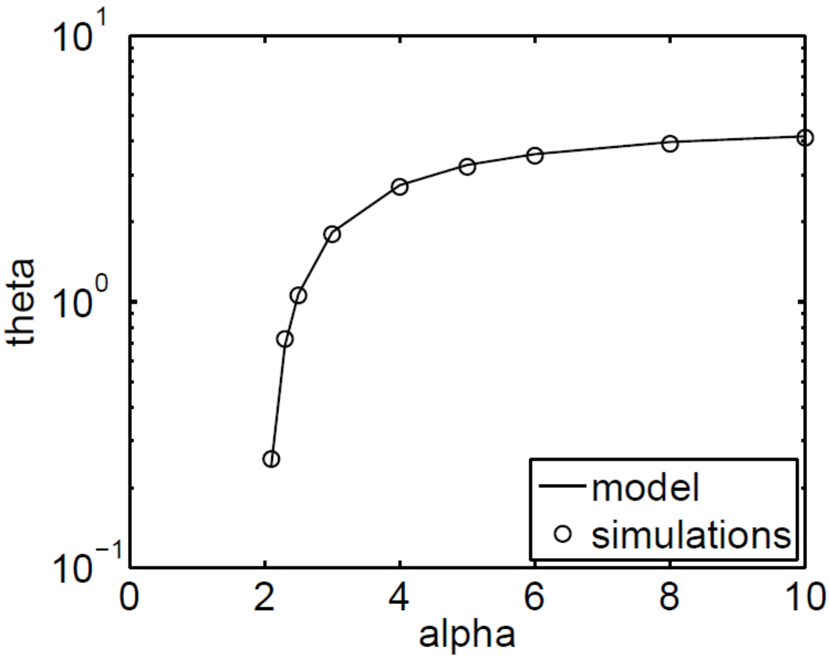
  - $\theta = 0$ in a passive system

- Theorem 3 (Message Overhead in Active Systems):
  - For exponential lifetimes, $\theta = k$ holds for all $p(x)$
  - For heavy-tailed $L$ and uniform selection, $\theta$ is always smaller than $k$ and eventually reducing to 0 as $R \to \infty$ (the system is driven by join overhead)
  - For light-tailed $L$ and uniform selection, $\theta$ is always larger than $k$ (the system is driven by edge failure)

# Message Overhead 3

- The mean number of replacement neighbors sought during user lifetime in systems with different shape parameters
  - Pareto $L$ with mean $E[L] = 0.5$ hours
  - Left: uniform selection
  - Right: age-proportional selection





# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# In-Link Arrival Process

- The rate at which a user generates outgoing edges is:

$$\lambda = \frac{k+\theta}{E[L]}$$

  - Recall that in passive systems, $\theta = 0$
  - In active systems, $\theta \geq 0$
- Consider the aggregate edge-arrival process to a live user $v$ from the rest of the system
- Theorem 6 (Edge-Arrival Process):
  - Under Assumption 1, the arrival process of in-links to a live user $v$ converges in distribution to a non-homogeneous Poisson process where the edge-arrival rate to $v$with age $x$ is proportional to $p(x)$:

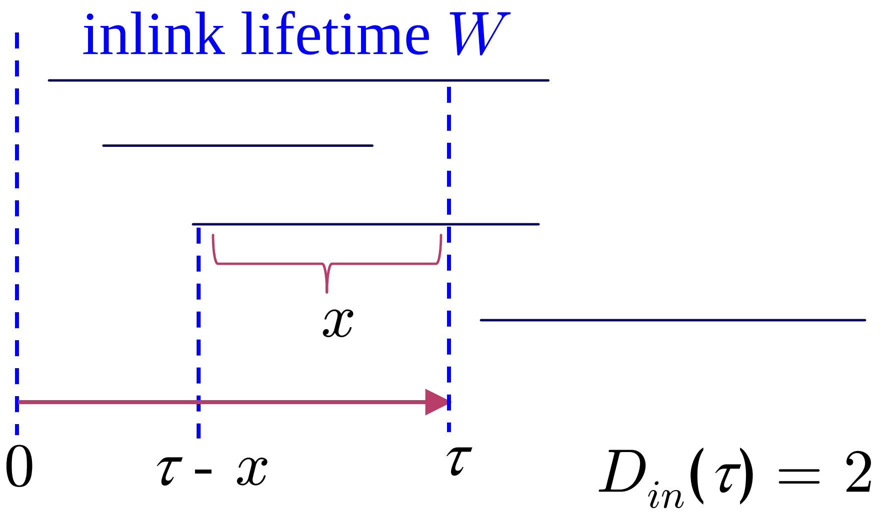
$$\lambda(x) := \lambda \frac{p(x)}{E[p(A)]}$$

# In-Degree

- Theorem 7 (In-degree):
  - For a fixed age $\tau \geq 0$, in-degree $D_{in}(\tau)$ of a live user $v$ converges in distribution to a Poisson random variable with mean:

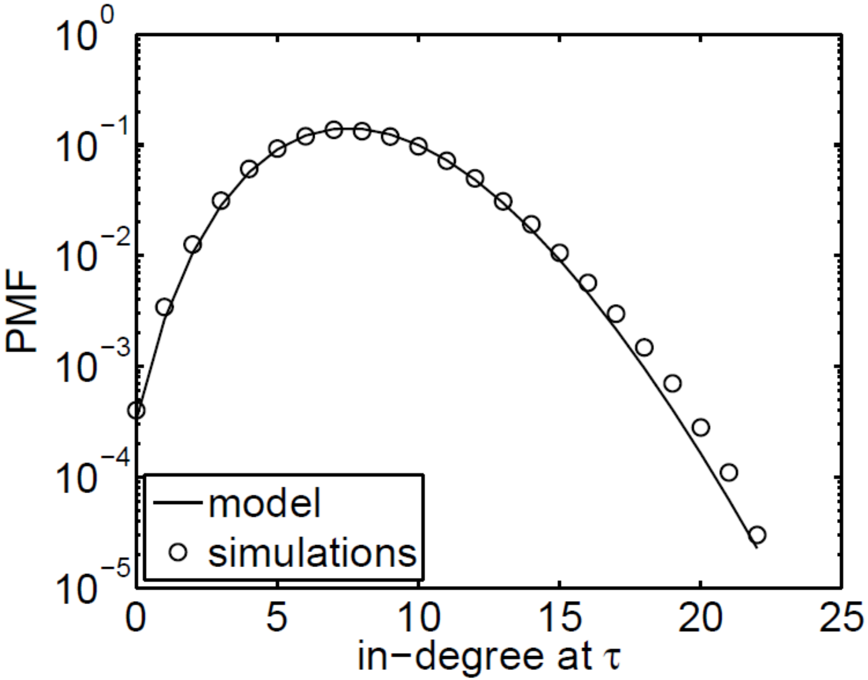
$$\nu(\tau) = \int_0^{\tau} \bar{F}_W(x)\lambda(\tau - x)dx$$

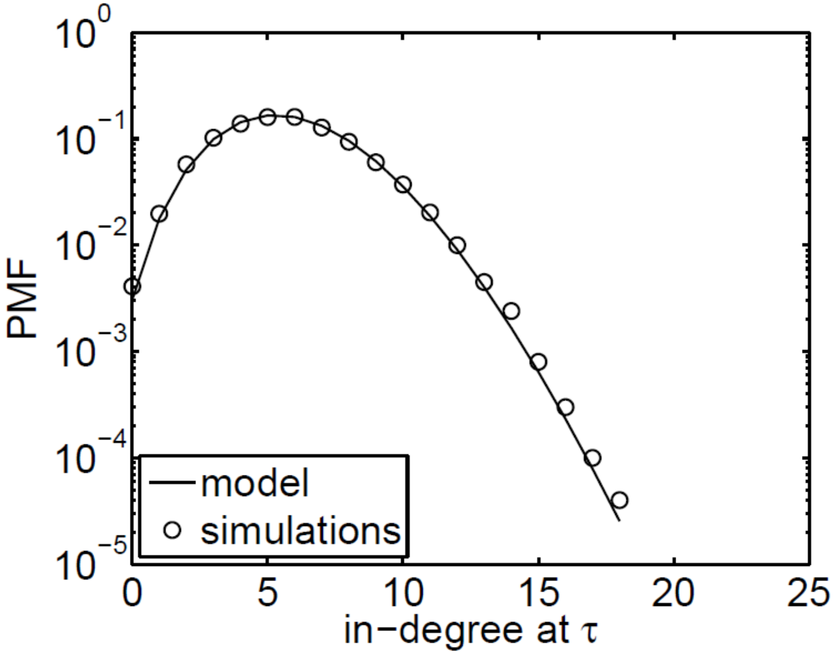
$F_W(x)$: in-link lifetime distribution



# In-Degree 2

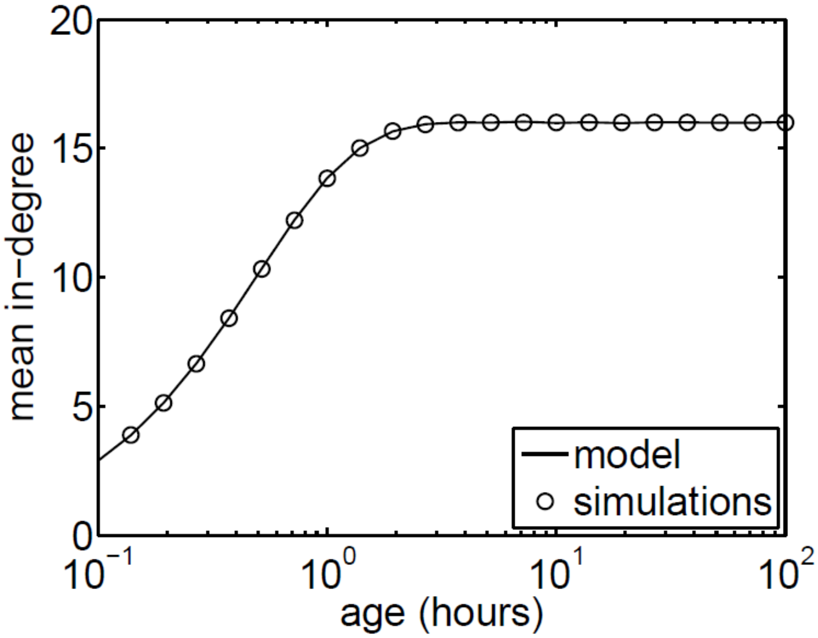
- The in-degree at $\tau$ follows a Poisson distribution
  - $\tau$ = 1 hour, Pareto alpha = 3 and $E[L]$ = 0.5 hours, $k$ = 8 (active systems)
  - Left: max-age with $m$ = 5
  - Right: age-proportional



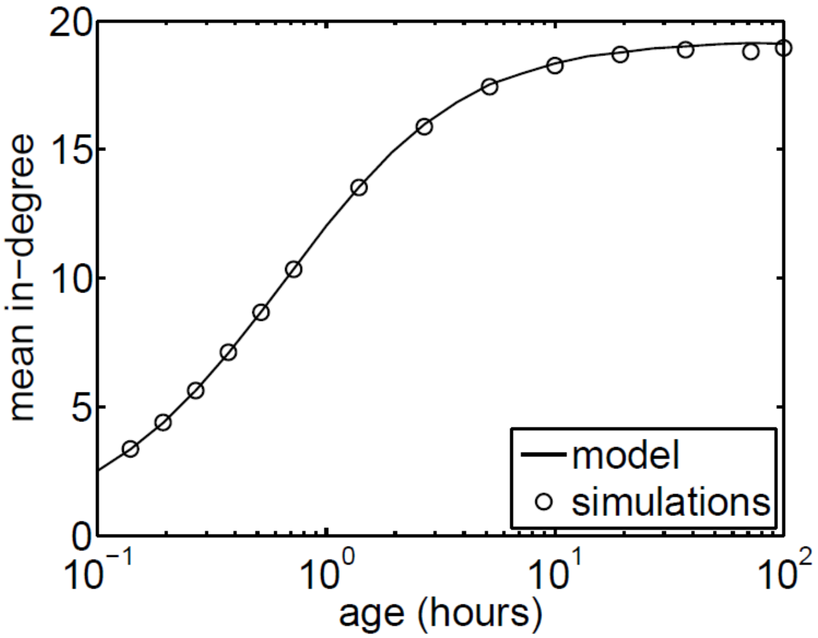


# In-Degree 3

- Mean in-degree under different $p(x)$



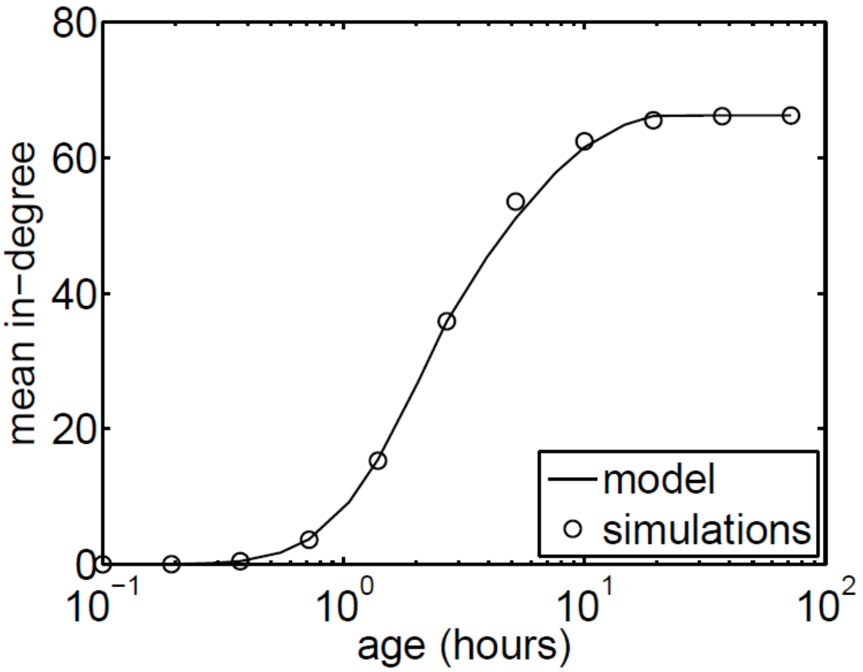
(a) uniform/exponential ($n = 2\text{K}$)



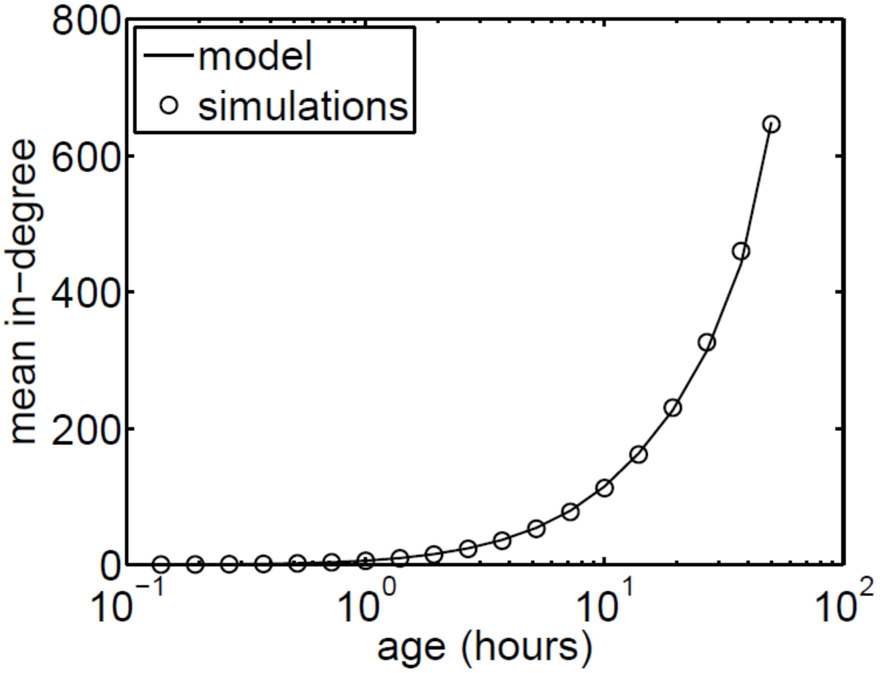
(b) uniform/Pareto ($n = 2\text{K}$)

# In-Degree 4

- Mean in-degree under different $p(x)$



(c) max-age/Pareto $(m = 5, n = 2\text{K})$



(d) age-prop/Pareto $(n = 15\text{K})$

# Agenda

- Introduction
  - Motivation
- Unifying neighbor selection model
  - Active and passive systems
  - General neighbor preference function
- Metrics
  - Out-link churn
  - Message overhead
  - In-link churn (see paper)
  - In-degree
  - Combined in/out-degree
- Conclusion

# Combined In/Out-Degree

- In passive systems, the mean in-degree at $\tau$ is

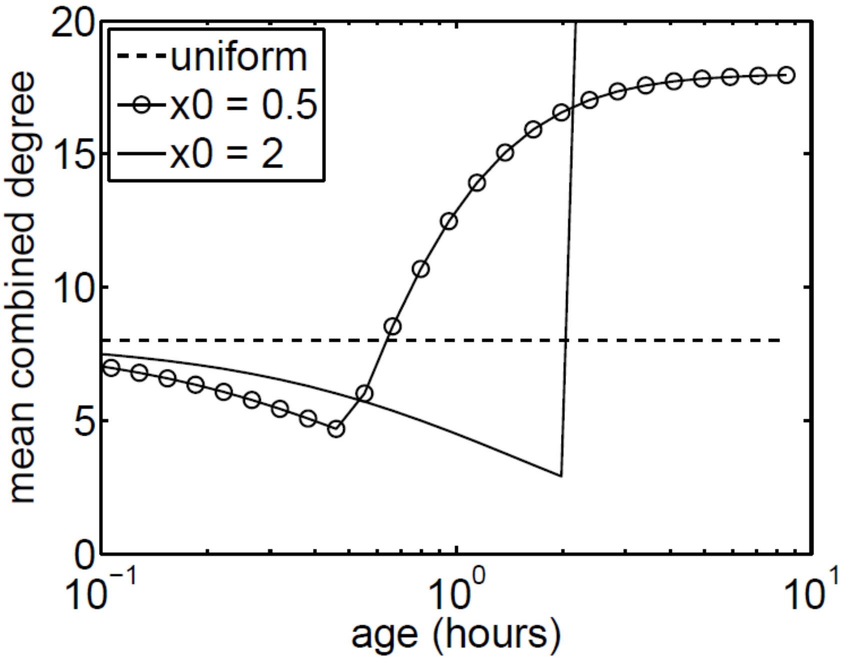
$$\nu(\tau) = \frac{kE[p(\tau - A)]}{E[p(A)]}$$

  - Step-function: $\nu_{step}(\tau) = \frac{kF_A(\tau - x_0)}{1 - F_A(x_0)}$
  - Age-proportional: $\nu_{age}(\tau) = k\left(\frac{\tau}{E[A]} - F_Z(\tau)\right)$
  - The mean in-degree in active systems is even bigger
  - This indicates that unbounded functions $p(x)$ are not unsuitable in both active and passive systems
- The mean out-degree at $\tau$ in passive systems is

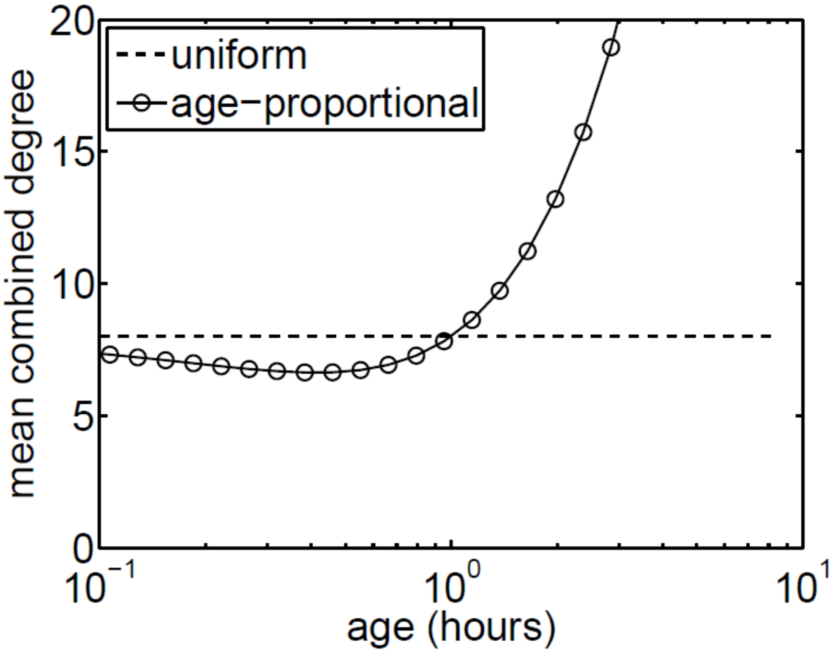
$$E[D_{out}(\tau)] = k\bar{F}_V(\tau)$$

# Combined In/Out-Degree 2

- Combined expected degree in passive systems under Pareto lifetimes with alpha = 3 and $k = 8$
  - A more aggressive $p(x)$ results in a more heavy-tailed $V$
  - This occurs at the expense of lowering resilience of in-links and increasing the in-degree (thus workloads) of high-age peers



(a) step-function



(b) age-proportional

# Conclusion

- We introduced a novel unifying neighbor selection model
  - Under this umbrella, we examined both passive and active systems
  - Analyzed both uniform and non-uniform neighbor selection strategies in unstructured P2P networks
- We analyzed metrics that are important to such systems
  - Resilience of out/in-links
  - Message overhead for searching neighbors
  - Edge-arrival process to a live user under general age-based selection
  - Transient in-degree process
  - Combined in/out-degree
- We offered practical guidelines for balancing the various tradeoffs and selection system parameters