Heterogeneous Renewal Churn Model for Unstructured P2P Networks

Zhongmei Yao, Daren B.H. Cline, Xiaoming Wang, and Dmitri Loguinov

***Abstract*—Previous analytical results on the resilience of unstructured P2P systems have not explicitly modeled heterogeneity of user churn (i.e., difference in online behavior) or the impact of in-degree on system resilience. To overcome these limitations, we introduce a generic model of heterogeneous user churn, derive the distribution of the various metrics observed in prior experimental studies (e.g., aggregate lifetime distribution of all users visiting the system, joint distribution of session time of currently alive peers, and residual lifetime of a randomly selected user), derive several closed-form results on the transient behavior of in-degree, and eventually obtain the joint in/out degree isolation probability as a simple extension of the out-degree model in [14].**

I. Introduction

Peer-to-peer (P2P) networks have recently emerged as an efficient and highly resilient platform for large-scale distributed applications. One of the fundamental problems in comprehending how these systems behave is the analysis of their properties during *user churn*, which is a general term describing arrival/failure of individual nodes in the system and repair algorithms applied by surviving users to counteract the effects of abrupt departures. Unlike other distributed systems where failures may be considered rare or abnormal, most P2P networks constantly remain in the state of churn and embrace frequent failures as part of their normal operation.

While many metrics of a system (e.g., search latency, path existence probability, efficiency of routing, message overhead, file popularity) affect its usefulness to the user, one commonly studied problem in the literature is the ability of P2P networks to stay connected in the face of random failures [2], [6], [8], [10], [11], [12], [14], [16], [18], [21], [25], [27]. It may be argued that compromised connectivity is one of the most fundamental byproducts of churn that directly affects routing efficiency and other metrics observed by the user. However, before resilience and performance of P2P networks can be fully understood, a good model of churn is required since even today most analytical models that consider churn [11], [14], [18], [21] do not completely capture the inherent heterogeneity of users, the impact of in-degree on the resilience of the system, or the behavior of P2P networks under non-exponential lifetimes.

A. Churn Model

We start the paper with a goal of modeling churn in P2P systems and striking a balance between model complexity and its applicability. Our focus for the time being lies with networks in which neighbor selection occurs only during the initial join into the system and failure of an existing neighbor. This framework is commonly used in *unstructured* P2P implementations (such as the ultra-peer layer of Gnutella) where replacement takes place as a reactive response to failed keep-alive messages. Certain DHTs in which neighbor pointers do not have to change after arrival of new users (e.g., Randomized Chord [20]) fall into this category as well.[1] This type of churn was originally formalized in [14], where Leonard *et al*. equipped users with random lifetimes L_i that determined the duration of their presence in the system and modeled neighbor replacement using random delays S_i that included the timeouts to detect each failure and protocol delays to actually obtain a new neighbor. This model, however, treated P2P users equally in their online characteristics (i.e., all user lifetimes were drawn from the same distribution) and did not consider their offline behavior as having any impact on churn.

Heterogeneity of lifetimes is a fundamental property of P2P systems where some users consistently spend substantial periods of time in the system and others very little [28]. This observation prompts the question of *whether P2P systems can indeed be modeled using a single homogeneous lifetime distribution without sacrificing model accuracy*? In addition to lifetimes, churn is characterized by the distribution of offline durations, which together with lifetimes define the *availability* of each user [3], [26], i.e., the average fraction of time a user is logged in. It is therefore important to understand how offtimes contribute to the dynamics of the system and which peer characteristics affect local graph-theoretic properties (e.g., distribution of in and out-degree at each time t, probability that a given neighbor is alive, isolation probability within a lifetime) of each user.

To answer these questions, we offer a generic churn model that captures the *heterogeneous* behavior of end-users, including their difference in online habits and diversity of offline "sleep time." We view each user as an alternating renewal process that is ON when the user is logged in and OFF otherwise, where online/offline durations of each user i are respectively drawn from distributions $F_i(x)$ and $G_i(x)$. This approach creates a system of *heterogeneous* users, each with its own profile of behavior that stays constant during the peer's recurring participation in the network [28].

Armed with this model, we obtain the aggregate lifetime distribution $F(x)$ of all users who have joined the system, the

Z. Yao, X. Wang, and D. Loguinov are with the Department of Computer Science, Texas A&M University, College Station, TX 77843 USA (email: {mayyao, xmwang, dmitri}@cs.tamu.edu).

D. Cline is with the Statistics Department at Texas A&M University, College Station, TX 77843 USA (email: dcline@stat.tamu.edu).

[1] Analysis of DHTs in which arriving users actively replace existing neighbors is covered in [31].

lifetime distribution $J(x)$ of the users currently online, and the residual lifetime distribution $H(x)$ of a randomly selected user in the network. Our results show that all three metrics are weighted functions of individual lifetime distributions $F_i(x)$, where $H(x)$ is additionally dependent on the number of users currently in the network, the probability that a given user is picked by joining peers, and the conditional residual lifetimes of neighbors chosen by the selection method. The model for $H(x)$ is extremely complex and generally intractable unless neighbor selection is performed *uniformly among currently participating users*[2], in which case we show that $H(x)$ can be directly obtained from $F(x)$. This is an important conclusion that demonstrates that instead of measuring n individual lifetime distributions, where n is the total number of users participating in the system, one can measure lifetimes of joining users to obtain $F(x)$, which is then *sufficient to entirely model the effect of churn* on P2P graphs.

We also revisit the observation of [28] that the users already present in Gnutella and BitTorrent networks exhibit larger average lifetimes than those joining the system. We show that this effect is a consequence of $J(x)$ being the *spread* [30] of distribution $F(x)$, which allows us to prove that random users currently in the system have stochastically larger lifetimes than random arriving users *regardless of the shape of distributions* $F_i(x)$ and $G_i(x)$. We additionally show that while $F(x)$ may appear to be heavy-tailed as observed in practice [5], [9], [17], it is possible that individual lifetime distributions $F_i(x)$ may *all* be exponential, or contain a mix of exponential and heavy-tailed distributions. Occurrence of this effect depends on random availability of each user and shows that conclusions on the individual habits of peers may not be drawn from their aggregate behavior $F(x)$.

B. Local Resilience Model

In the second half of the paper, we tackle the issue of node resilience to isolation (i.e., loss of all neighbors within a user's lifetime) in the presence of churn under the assumption that neighbors provide *mutual* resilience benefits to each other (i.e., both outgoing and incoming edges increase resilience of peers). Prior work [14], [15] showed that many P2P graphs stayed connected if and only if they did not develop isolated nodes during churn. As in [14], we assume that each joining node selects k out-degree neighbors and replaces failed neighbors with existing users in the system after some random search delay. With this setup, Leonard *et al*. [14] derived that individual node isolation probabilities were functions of $\rho = E(L_i)/E(S_i)$, where $E(L_i)$ was the mean lifetime of homogeneous users and $E(S_i)$ was the mean search (i.e., node-replacement) delay. However, despite its importance, this approach only modeled the *out-degree* of each user and did not consider the increased resilience arising from additional *in-degree* edges arriving in the background to each user during its stay in the system.

We overcome this shortcoming and build a complete closed-form model characterizing the evolution of in-degree in unstructured systems under the assumption of uniform neighbor selection. We first show that under our churn model, the edge arrival process to each *live* user tends to a Poisson process when system size becomes sufficiently large, which is consistent with recent observation of this phenomenon in certain real networks [28]. We then derive the expected transient in-degree as a function of $F(x)$, *including cases with non-exponential peer lifetimes*, and show that users who stay online longer quickly accumulate non-trivial in-degree and become much more resilient to isolation over time. This outcome was intuitively expected as it makes sense that current unstructured P2P networks have been designed such that users with more contribution to the system (i.e., longer lives) become better connected over time and provide more search capabilities to their neighbors. In contrast, the original model of [14] showed that P2P users became progressively more susceptible to isolation as their age increased.

We finish the paper by combining the in and out-degree isolation models into a single approximation that clearly shows the contribution of in-degree to the resilience of the graph. Denoting by ϕ the isolation probability of a user (i.e., loss of all neighbors within its lifetime) and by ϕ_{out} the same metric with only the out-degree being considered [14], we show that for exponential $F(x)$ the following holds as search delays become asymptotically small (i.e., tend to zero):

$$\phi = \frac{1 - e^{-2k}}{2k}\phi_{out}, \tag{1}$$

where k is the initial number of neighbors obtained by each arriving user. This result illustrates that the amount of improvement from the in-degree amounts to approximately a factor of $2k$ reduction in the isolation probability. We also observe from our closed-form Markov-chain model that for non-negligible search delays, ratio ϕ_{out}/ϕ is often much larger than implied by (1), which suggests that (1) may be a worst-case upper bound on ϕ. We finish the paper with examples that demonstrate this effect.

This paper is organized as follows. Section II discusses related work. We introduce our churn model in Section III and study its properties in section IV. Section V models edge arrival to individual users, section VI discusses the in-degree of each user, section VII discusses joint in/out-degree isolation, and section VIII concludes the paper.

II. Related Work

One of the first models of churn was proposed in [21], which assumed an unstructured P2P system with Poisson arrivals and departures that could be modeled as an $M/M/1$ queue. Neighbor replacement in this system was in direct response to failures and was assumed to be instantaneous, where the possibilities for replacement were limited to the nodes currently alive in a certain centralized cache. The paper showed that under user churn the graph remained connected and exhibited a logarithmic diameter, both with high probability.

Later models of churn [18] and recently [11] assumed a DHT-like system in which repair algorithms were run *independently* of user failures and at exponentially distributed intervals (i.e., as Poisson processes). This approach modeled

[2]This can be implemented by picking users from uniformly random subsets of cached nodes or using special random walks on the graph [33].

the consistency check algorithm in Chord, which periodically verified the successor list and corrected invalid pointers. These models assumed homogeneous exponential lifetimes and Poisson arrival/departure processes with no way of generalizing their results to non-exponential system dynamics.

A different approach was undertaken in [14], where neighbor replacements were explicitly initiated in response to failed links. In this setup, each joining user randomly selected k neighbors from the graph and then monitored their online presence using keep-alive messages.[3] Once the failure of an existing neighbor was detected, a uniformly random replacement was sought from among the currently alive users in the system. Detection and replacement delays were also random, but explicitly non-zero. Under these conditions, the paper showed that each user became isolated with probability no larger than $\phi_{out} = k\rho/(1+\rho)^k$, where ρ was the ratio of the average lifetime to the average replacement delay, for all lifetime distributions with an exponential or heavier tail. This result was later generalized in [15] to show that the probability of non-partitioning in many P2P networks converged as $n \to \infty$ to that of avoiding isolation for each online user.

III. Churn Model

To understand the dynamics of churn and performance of P2P systems, we start by creating a model of user behavior and specifying assumptions on peer arrival, departure, and selection of neighbors. The focus of this section is to formalize recurring user participation in P2P systems in a simple model that takes into account heterogeneous browsing habits and explains the relationship between the various lifetime distributions observable in P2P networks.

A. Basics

Consider a P2P system with n participating users, where each user i is either alive (i.e., present in the system) at time $t \geq 0$ or dead (i.e., logged off). This behavior can be modeled by an ON/OFF right-continuous process $\{Z_i(t)\}$ for each i:

$$Z_i(t) := \begin{cases} 1 & \text{user } i \text{ is alive at time } t \\ 0 & \text{otherwise} \end{cases}, \quad 1 \leq i \leq n. \tag{2}$$

This framework is illustrated in Fig. 1, where parameter m stands for the cycle number and random variables $L_{i,m} > 0, D_{i,m} > 0$ are durations of user i's ON (life) and OFF (death) periods, respectively. The figure also shows the *residual process* $R_i(t)$, which is the duration of user i's remaining online presence from time t conditioned on the fact that it was alive at t.

We next make several modeling assumptions about this system and explain how users generate their online/offline durations.

Assumption 1: Set $\{Z_i(t)\}_{i=1}^n$ consists of mutually independent, alternating renewal processes.

[3]Gnutella, for example, sends a ping message every 3 seconds and detects link failure when TCP declares the connection aborted, which happens after several (e.g., 5 in Windows) subsequently failed retransmission attempts.

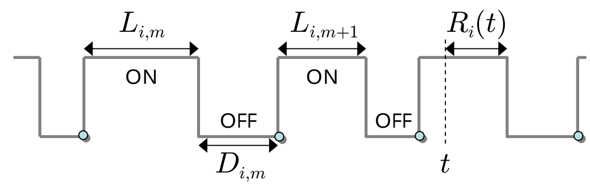


Fig. 1. Process $\{Z_i(t)\}$ depicting ON/OFF behavior of user i.

To elaborate, we restrict ON durations $\{L_{i,m}\}_{m=1}^\infty$ of user i to independent random variables (r.v.) with a general cumulative distribution function (CDF) $F_i(x)$ and OFF durations $\{D_{i,m}\}_{m=1}^\infty$ to independent r.v. with another CDF $G_i(x)$. This assumption also implies that the two sequences $\{L_{i,m}\}_{m=1}^\infty$ and $\{D_{i,m}\}_{m=1}^\infty$ are independent. We leave discussion of the more general case of correlated ON/OFF cycles to future work. Mutual independence in Assumption 1 additionally states that users do not synchronize their arrival or departures and generally exhibit uncorrelated lifetime characteristics (e.g., users simultaneously present in the system with multiple identities are not very common and have no large-scale impact on the dynamics of the network).

While Assumption 1 is a good start and allows certain results below to hold, asymptotically large systems require additional constraints on how users select their distributions $F_i(x), G_i(x)$. We next suppose that there are $\mathcal{T} \geq 1$ user types in the system representing different behavior (e.g., desktop peers that stay in the system for days is one type, while laptop users that frequently disconnect is another). Before the network starts to evolve, each user randomly decides on its type, which then remains fixed for all $t > 0$.

Assumption 2: (a) There exists some set $\mathcal{F}$ of distinct pairs of non-lattice CDFs defining non-negative random variables:

$$\mathcal{F} := \left\{ \left(F^{(1)}(x), G^{(1)}(x)\right), \ldots, \left(F^{(\mathcal{T})}(x), G^{(\mathcal{T})}(x)\right) \right\},$$

where $\mathcal{T} \geq 1$ is a fixed number of user types and CDFs $F^{(j)}(x) > 0$ and $G^{(j)}(x) > 0$ for all $x > 0$, all $j = 1, \ldots, \mathcal{T}$. Further, each mean $l^{(j)} := \int_0^\infty (1 - F^{(j)}(x))dx$ and $d^{(j)} := \int_0^\infty (1 - G^{(j)}(x))dx$ satisfies $0 < l^{(j)}, d^{(j)} < \infty$ for all types $j = 1, \ldots, \mathcal{T}$;

(b) The pair of ON/OFF duration CDFs $(F_i(x), G_i(x))$ of each user i, $i = 1, \ldots, n$, is independently drawn from set $\mathcal{F}$, where type j is selected with probability (w.p.) $p_j \geq 0$ and $\sum_{j=1}^{\mathcal{T}} p_j = 1$;

(c) Defining $\mathcal{S}$ to be set of selections made by each user and conditioning on $\mathcal{S}$, Assumption 1 holds.

Assumption 2(a) uses $\mathcal{T}$ as the “diversity” factor of user behavior (e.g., $\mathcal{T} = 1$ reduces the system to a network of homogeneous users) and mandates that all average online/offline durations are both positive and finite. Part (b) allows for bias in the selection process and lets certain user types be more popular than others. Part (c) ensures that once users have chosen their types (i.e., ON/OFF duration CDFs), the system evolves as described in Assumption 1. Note that Assumption 1 is more general and includes Assumption 2 as a special case.

B. Properties

We next explain the ON/OFF distributions commonly considered in this paper and obtain basic properties of the system. The first lifetime CDF is exponential $F_i(x) = 1 - e^{-\mu_i x}, \mu_i > 0$, with mean $1/\mu_i$. The second one is shifted Pareto

$$F_i(x) = 1 - (1 + x/\beta_i)^{-\alpha_i}, \quad \alpha_i > 1, \beta_i > 0, \tag{3}$$

with mean $\beta_i/(\alpha_i - 1)$. Offline distributions $G_i(x)$ do not affect our analysis and are kept general. For convenience of notation, define the mean lifetime of each user $l_i := E(L_{i,m})$ and the mean offline duration $d_i := E(D_{i,m})$. Denote the reciprocal of the mean ON/OFF cycle length of user i by

$$\lambda_i := (l_i + d_i)^{-1}, \tag{4}$$

which is the time-averaged arrival rate of the user into the system. We easily obtain from Smith's theorem that the asymptotic *availability* of each user i, i.e., the probability that it is in the system at an arbitrary instance t, is given by

$$a_i := \lim_{t\to\infty} P(Z_i(t) = 1) = \frac{l_i}{l_i + d_i}. \tag{5}$$

The final metric related to our churn model is the distribution of the number of users in the system. Denote by $N(n,t) := \sum_{i=1}^{n} Z_i(t)$ the number of users in the network at time t and notice that it is also a random process that fluctuates with time. Since many P2P properties of interest require stationarity, our analysis below is frequently confined to limiting distributions when network age $t \to \infty$, which we call *equilibrium*.

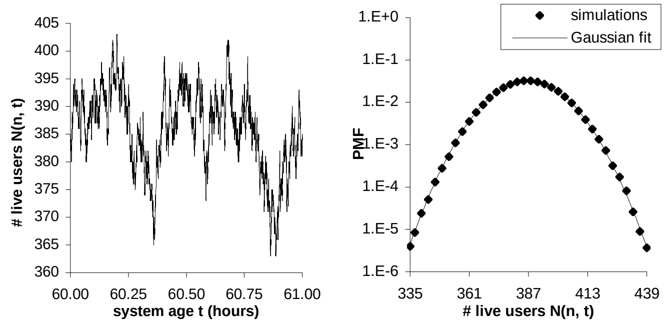
Define Z_i to be a Bernoulli r.v. with the equilibrium distribution of $Z_i(t)$, i.e., $P(Z_i = 1) = a_i$, where a_i is in (5). Further define $N(n) := \sum_{i=1}^{n} Z_i$, which is a r.v. with the equilibrium distribution of $N(n,t)$. Since Z_i's are independent and identically distributed under Assumption 2, the next lemma directly follows from the central limit theorem.

Lemma 1: Under Assumption 2, we have that as $n \to \infty$,

$$\frac{N(n) - \mu_n}{\sigma_n} \xrightarrow{D} \mathcal{N}(0,1), \tag{6}$$

where $\mu_n := E[N(n)] = na$, $\sigma_n^2 := \mathrm{var}[N(n)] = na(1-a)$, $a := \sum_{j=1}^{\mathcal{T}} p_j a^{(j)}$, $a^{(j)} := l^{(j)}/(l^{(j)} + d^{(j)})$ is the availability of user type j, and $\mathcal{N}(0,1)$ denotes a standard normal r.v.

We next show simulations explaining this result and its accuracy in systems with *finite* age and size. We generate a network of n users whose arrival/departure follows the introduced churn model. The system evolves for at least 50 virtual hours before being examined. We start by generating $\mathcal{T} = 1{,}000$ pairs of means $l^{(j)}$ and $d^{(j)}$, which are drawn randomly from two Pareto distributions with $\alpha = 3$ as described next. For mean ON durations, we use $\beta = 1$ and obtain $E(l^{(j)}) = 1/2$ hour; for mean OFF durations, we use $\beta = 2$ and get $E(d^{(j)}) = 1$ hour. We study three cases throughout the paper: 1) heavy-tailed system $\mathcal{H}$ with $F^{(j)}(x) \sim \mathrm{Pareto}(3, 2l^{(j)})$ and $G^{(j)}(x) \sim \mathrm{Pareto}(3, 2d^{(j)})$; 2) very heavy-tailed system $\mathcal{VH}$ with $F^{(j)}(x) \sim \mathrm{Pareto}(1.5, l^{(j)}/2)$ and $G^{(j)}(x) \sim \mathrm{Pareto}(1.5, d^{(j)}/2)$; and 3) exponential system $\mathcal{E}$ with $F^{(j)}(x) \sim \exp(1/l^{(j)})$ and $G^{(j)}(x) \sim \mathrm{Pareto}(3, 2d^{(j)})$, where notation Pareto$(\alpha_i, \beta_i)$ refers to (3). The actual pairs $(F_i(x), G_i(x))$ are selected uniformly randomly from $\mathcal{F}$.



(a) evolution (b) PMF

Fig. 2. Sample path and distribution of $N(n,t)$ in system $\mathcal{H}$ with $n = 1000$ users. The Gaussian fit is from Lemma 1 after 10^6 iterations.

Fig. 2(a) shows one example for the evolution of system size $N(n,t)$ as a function time t. Part (b) of the figure shows the PMF (probability mass function) of $N(n,t)$ at $t \gg 0$ and a Gaussian fit from Lemma 1, confirming its accuracy.

C. Aggregate Lifetimes

Prior measurement studies [28], [29] sampled lifetimes of all joining users over some long period of time to characterize the dynamics of P2P systems. We are now interested in what metric they estimated and how it can be expressed in our notation. For each instance of user i being present in the system during interval $[0,t]$, place its ON duration $L_{i,m}$ into set $S_i(t)$ and define $S(t) = \cup_{i=1}^{n} S_i(t)$. Then let $F(n,t,x)$ be the CDF of values collected in set $S(t)$ (i.e., the probability that the obtained lifetimes are less than or equal to x). Finally, define $F(n,x) := \lim_{t\to\infty} F(n,t,x)$ to be the *aggregate lifetime distribution* of the system and $l(n)$ to be its mean (both exist from Assumption 2).

Our next result shows that $F(n,x)$ a weighted average of individual lifetime distributions, where the weights are biased toward those peers who frequently join and leave the system since their sessions constitute the majority of overall peer arrival into the system.

Theorem 1: With Assumption 1 and any finite $n \ge 1$:

$$F(n,x) = \sum_{i=1}^{n} b_i F_i(x), \quad l(n) = \sum_{i=1}^{n} b_i l_i, \tag{7}$$

where $b_i := \lambda_i / \sum_{j=1}^{n} \lambda_j$ and λ_i is defined in (4).

Proof: For large t, set $S(t)$ contains approximately

$$f_i(t) = \frac{\lfloor t\lambda_i \rfloor}{\sum_{j=1}^{n} \lfloor t\lambda_j \rfloor} \tag{8}$$

lifetime variables from user i. Bounding this metric, we have:

$$b_i - \frac{1}{\sum_{j=1}^{n} t\lambda_j} \le f_i(t) \le \frac{t\lambda_i}{\sum_{j=1}^{n} t\lambda_j - n}, \tag{9}$$

where $b_i = \lambda_i / \sum_{j=1}^{n} \lambda_j$. Sending t to infinity in (9), it immediately follows that the proportion of r.v.'s from user i in $S(t)$ converges to $\lim_{t\to\infty} f_i(t) = b_i$. Therefore, the probability that the value of variable in set $S(t)$ is no larger than fixed $x \geq 0$ converges to:

$$\lim_{t\to\infty} F(n,t,x) = \lim_{t\to\infty} \sum_{i=1}^{n} P(L_i \leq x) f_i(t) = \sum_{i=1}^{n} P(L_i \leq x) \lim_{t\to\infty} f_i(t), \quad (10)$$

showing that the time limiting distribution exists.

Recalling that each $l_i < \infty$ by Assumption 1-b), we integrate the tail distribution $1 - F(n,x)$ for finite n to obtain:

$$E(L(n)) = \int_0^\infty \left(1 - \sum_{i=1}^{n} b_i F_i(x)\right) dx = \sum_{i=1}^{n} b_i \int_0^\infty (1 - F_i(x)) dx,$$

which leads to desired results in (7). ■

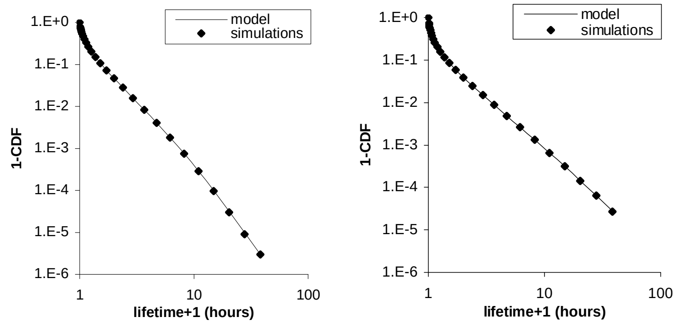
Observe from (7) that the expected time that users stay in the system is equal to the mean system population $\sum_i \lambda_i l_i = \sum_i a_i$ divided by the aggregate user arrival rate $\sum_i \lambda_i$, which is consistent with Little's law.

Theorem 1 holds under the more general Assumption 1 as long as n is finite; however, to guarantee that the sums in (7) converge as n increases, one requires Assumption 2. We show this analysis later in the paper. In the meantime, we state similar results for aggregate offline durations.

Corollary 1: With Assumption 1 and any finite $n \geq 1$, the CDF of aggregate offline durations is $G(n,x) := \sum_{i=1}^{n} b_i G_i(x)$ and the its mean is $d(n) := \sum_{i=1}^{n} b_i d_i$.

We verify (7) in simulations and discuss several implications of this result. Two typical simulations are presented in Fig. 3 for exponential and heavy-tailed lifetimes, both of which show that the model is very consistent with simulation results. Both figures are on log-log scale and plot $1 - F(n,x)$ vs. $1 + x$ to make the shifted Pareto distribution in (3) appear as a straight line. Notice in Fig. 3(a) that system $\mathcal{E}$ produces an appearance of a heavy-tailed aggregate distribution $F(n,x)$ *even though all individual $F_i(x)$ are exponential*. This can be explained as follows. It is well-known [7] that for a hyper-exponential distribution in the form of (7) and *any* desired distribution $W(x)$ with a monotonic PDF (probability density function), there exists a set of weights $\{b_1, \ldots, b_n\}$ such that (7) converges to $W(x)$ as $n \to \infty$. Given numerous possibilities for the arrival-rate set $\{\lambda_1, \ldots, \lambda_n\}$ in practice, it is possible that one can observe a nicely shaped Pareto, Weibull, or other distribution $F(n,x)$, which is produced by a mixture of exponential $F_i(x)$. It may therefore be premature to conclude that Pareto $F(n,x)$ measured experimentally [5], [26] necessarily reveals the true nature of individual user behavior.

While our current conclusion shows that one cannot characterize the lifetimes or availability of individual peers by observing their aggregate behavior, the next question we seek to answer is *whether the aggregate behavior $F(n,x)$ can be used to characterize the parameters of a single user selected from the system randomly?*



(a) system $\mathcal{E}$ (b) system $\mathcal{H}$

Fig. 3. Comparison of simulation results of $F(n,x)$ to model (7) in a graph with $n = 1000$ nodes. System evolved to age 10^5 hours.

IV. Characteristics of Selected Users

Suppose v picks a random currently-alive user i as a potential neighbor. Our primary goal is to understand the properties of i in terms of two metrics: its remaining online duration and its current session length.

A. Definitions

Let $R_i(t)$ denote the remaining life of a given user i at time t, i.e., the remainder of the current ON cycle illustrated in Fig. 1. Variable $R_i(t)$ is important since it determines how long this neighbor will remain online *after it has been selected*. The equilibrium residual lifetime distribution

$$H_i(x) := \lim_{t\to\infty} P(R_i(t) \leq x | Z_i(t) = 1)$$

can be written in terms of $F_i(x)$ [30]:

$$H_i(x) = \frac{1}{l_i} \int_0^x (1 - F_i(u)) du, \quad x \geq 0. \quad (11)$$

Next, define $R(n,t)$ to be the residual lifetime of the user *randomly selected* from among $N(n,t) \geq 1$ users that are alive. Denote by $H(n,x)$ the equilibrium distribution of $R(n,t)$ conditioned on $N(n,t) \geq 1$:

$$H(n,x) := \lim_{t\to\infty} P(R(n,t) \leq x | N(n,t) \geq 1). \quad (12)$$

Our goal is to obtain an expression for (12). We start with the most general case where choices may be based on the lifetimes of potential neighbors and then proceed to the much-simpler case of uniform selection.

B. General Case

To understand the results that follow, denote by

$$S_i(t) := \begin{cases} 1 & \text{user } i \text{ is selected by } v \text{ at } t \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

the indicator process that shows whether user i is randomly selected at time t from among $N(n,t) \geq 1$ users currently in the system. Define

$$\pi_i(x) = \lim_{t\to\infty} P(S_i(t) = 1 | Z_i(t) = 1, R_i(t) \leq x, N(n,t) \geq 1) \quad (14)$$

to be the equilibrium probability that user i is selected given that it is alive, its residual is no larger than x, and the system contains at least one user. We next elaborate on this metric.

In systems where the residual lifetime distribution of a user does not affect its chance of being chosen, $\pi_i(x) = \pi_i$ is not a function of x. This holds only in cases when neighbor selection is *independent* of the lifetimes (or ages) of selected users (e.g., this model was used in [14]). Examples that satisfy this condition include uniform selection, selection based on content similarity or random hashing space, age-independent popularity, etc. On the other hand, selection based on the age of potential neighbors or random walks (which depend on the in-degree of each user, which in turn depends on age) do not fall into this category (e.g., [32]).

Under uniform selection, each user i is picked with probability (conditioning on i being alive):

$$\begin{aligned}\pi_i(x) = \pi_i &= \lim_{t\to\infty} E(S_i(t)|Z_i(t)=1, N(n,t)\ge 1)\\ &= E\Big(\frac{1}{N(n)}|Z_i=1\Big),\end{aligned} \tag{15}$$

since $Z_i = 1$ implies $N(n) \ge 1$.

For stationary random walks, $\pi_i(x)$ becomes the limiting version of expectation $E(d_i(t)/\sum_{m=1}^{N(n,t)} d_m(t)|Z_i(t) = 1, R_i(t) \le x, N(n,t) \ge 1)$, where $d_i(t)$ is node degree of user i at time t. For content-based selection, assume that each user shares w_i files with others and that each peer is selected to be a neighbor proportionally to its “content utility” w_i. Then, the selection probability in (14) may be equal to $E(w_i/\sum_{m=1}^{N(n)} w_m|N(n) \ge 1)$.

As must be evident, the general model above can implement quite complex rules for choosing neighbors; however, tractability of the resulting distribution $H(n,x)$ is questionable for all except the simplest cases. Below, we first derive $H(n,x)$ for the most generic case and show that it can be expressed as a sum of weighted individual residual distributions, where the weights are biased towards users with large availability a_i and high probability $\pi_i(x)$ of being selected. We later simplify this expression for uniform selection.

Lemma 2: Given Assumption 1 and finite $n \ge 1$:

$$H(n,x) = \frac{1}{P(N(n)\ge 1)}\sum_{i=1}^{n} a_i \pi_i(x) H_i(x), \tag{16}$$

where $\pi_i(x)$ is given by (14) and $H_i(x)$ by (11).

Proof: Define:

$$q_i(x,t) = P(R_i(t)\le x, S_i(t)=1, Z_i(t)=1|N(n,t)\ge 1).$$

Recalling the additivity rule for disjoint events, (12) reduces to $H(n,x) = \lim_{t\to\infty}\sum_{i=1}^{n} q_i(x,t)$. For ease of presentation, break $q_i(x,t)$ into a product of the following two terms using conditional probabilities:

$$\begin{aligned}a(x,t) &= P(S_i(t)=1|Z_i(t)=1, R_i(t)\le x, N(n,t)\ge 1),\\ b(x,t) &= P(Z_i(t)=1, R_i(t)\le x|N(n,t)\ge 1)\\ &= \frac{P(Z_i(t)=1, R_i(t)\le x)}{P(N(n,t)\ge 1)},\end{aligned} \tag{17}$$

where the last step is obtained by Bayes' theorem and by the fact that $Z_i(t) = 1$ implies $N(n,t) \ge 1$. It is now clear that $\lim_{t\to\infty} a(x,t) = \pi_i(x)$ defined in (14) and $\lim_{t\to\infty} b(x,t) = a_i H_i(x)/P(N(n) \ge 1)$, which leads to (16). ■

Next, we focus on $H(n,x)$ under uniform selection and leave analysis of other strategies to future work.

C. Uniform Selection

While (16) under uniform selection has a simpler shape

$$H(n,x) = \frac{1}{P(N(n)\ge 1)}\sum_{i=1}^{n} a_i \pi_i H_i(x), \tag{18}$$

the expectation in π_i in (15) remains to be expanded in closed-form. Our first auxiliary result establishes important properties of $E(1/N(n)|N(n) \ge 1)$.

Lemma 3: Given Assumption 2 and $N(n) \ge 1$, $\mu_n/N(n)$ converges to 1 in r-th mean for all $r \ge 1$:

$$\lim_{n\to\infty} E\Big(\Big|\frac{\mu_n}{N(n)} - 1\Big|^r\ |N(n)\ge 1\Big) = 0, \tag{19}$$

where $\mu_n = E(N(n))$ is the mean population.

Proof: For convenience, define A_n to a r.v. with the distribution equal to the conditional distribution of $N(n)/\mu_n$ *given* $N(n) \ge 1$. In what follows, we first show that $A_n^{-1} \xrightarrow{p} 1$ (i.e., convergence in probability) and then uniform integrability of $|A_n^{-r}|$ for all $r \ge 1$.

Recall from Assumption 2 that unconditional on user types, Z_i are i.i.d. with mean $a := \sum_{j=1}^{\mathcal{T}} p_j a_i$. Then, applying the weak law of large numbers for $N(n) = \sum_{i=1}^{n} Z_i$ leads to

$$\frac{N(n)}{\mu_n} = \frac{N(n)/n}{\mu_n/n} \xrightarrow{p} 1, \quad n\to\infty. \tag{20}$$

Meanwhile, by the Chernoff bound,

$$P(N(n)\ge 1) \ge 1 - \exp(-\mu_n(1-\mu_n^{-1})^2/2) \to 1, \tag{21}$$

as $n \to \infty$, since $\mu_n = \Theta(n)$ from (1). Using (20)-(21) yields

$$\begin{aligned}\forall\epsilon>0,\ P(|A_n - 1|\ge\epsilon) &= P\left(\left|\frac{N(n)}{\mu_n} - 1\right| \ge \epsilon|N(n)\ge 1\right)\\ &\le P\left(\left|\frac{N(n)}{\mu_n} - 1\right| \ge \epsilon\right)/P(N(n)\ge 1) \to 0,\end{aligned} \tag{22}$$

as $n \to \infty$, showing that $A_n \xrightarrow{p} 1$ as $n \to \infty$. Observing that $g(x) := x^{-1}$ is a continuous function for all $x > 0$, we then obtain that [4, pp. 112]

$$A_n^{-1} \xrightarrow{p} 1, \quad n\to\infty. \tag{23}$$

Our remaining step is to show that the following holds in order to prove uniform integrability of $|A_n^{-r}|$, for *all* $r \ge 1$:

$$\sup_{n\ge 1} E\left(|A_n^{-r}| 1_{|A_n^{-r}|>\alpha}\right) \to 0, \quad \alpha\to\infty, \tag{24}$$

which is the condition that ensures that a sequence converging in probability also converges in r-th mean [4, pp. 113].

To this end, note that given $N(n) \ge 1$, we have $A_n^{-r} \le \mu_n^r \le n^r$, $r \ge 1$. It is thus clear that for $n < \alpha^{1/r}$, $E(|A_n^{-r}| 1_{|A_n^{-r}|>\alpha}) = 0$. This leads to

$$\begin{aligned}\sup_{n\ge 1} E\left(|A_n^{-r}| 1_{|A_n^{-r}|>\alpha}\right) &= \sup_{n\ge\alpha^{1/r}} E\left(|A_n^{-r}| 1_{|A_n^{-r}|>\alpha}\right)\\ &\le \sup_{n\ge\alpha^{1/r}} \mu_n^r E(1_{|A_n^{-r}|>\alpha}),\end{aligned} \tag{25}$$

where $E(1_{|A_n^{-r}|>\alpha}) = P(|A_n^{-r}| > \alpha)$ will be examined next.

By the Chernoof bound, for all $n \geq 1$,

$$\begin{aligned} P(|A_n^{-r}| > \alpha) &= P(N(n) < \alpha^{-1/r}\mu_n \; |N(n) \geq 1) \\ &\leq P(N(n) < \alpha^{-1/r}\mu_n)/P(N(n) \geq 1) \\ &\leq \frac{\exp\Big(-\frac{\mu_n}{2}(1-\alpha^{-1/r})^2\Big)}{1-\exp(-\frac{\mu_n}{2}(1-\mu_n^{-1})^2)}. \end{aligned} \tag{26}$$

Upon combining (25) with (26) and noting that for $n \geq \alpha^{1/r}$, $\mu_n \to \infty$ as $\alpha \to \infty$, the following holds for all $r \geq 1$:

$$\sup_{n\geq 1} E\left(|A_n^{-r}|1_{|A_n^{-r}|>\alpha}\right) \leq \sup_{n \geq \alpha^{1/r}} \mu_n^r P(|A_n^{-r}| > \alpha) \to 0,$$

as $\alpha \to \infty$, showing that (24) holds.

Equipped with (23) and (24), applying Theorem 5 in [4, pp. 113] establishes this lemma. ■

In order to tackle the convergence of the sum in (18), our second auxiliary result shows that both $F(n,x)$ and $l(n)$ have limiting distributions.

Lemma 4: Under Assumption 2, the following sequences converge almost surely (a.s.) as $n \to \infty$:

$$F(n,x) \xrightarrow{a.s.} F(x) := \frac{\sum_{j=1}^{\mathcal{T}} p_j \lambda^{(j)} F^{(j)}(x)}{\lambda}, \tag{27}$$

$$l(n) \xrightarrow{a.s.} l := \frac{a}{\lambda}, \tag{28}$$

where $\lambda^{(j)} := 1/(l^{(j)} + d^{(j)})$ of type j, $\lambda := \sum_{j=1}^{\mathcal{T}} p_j \lambda^{(j)}$ is the average arrival rate, and a is the expected availability in (6). Furthermore, $F(x)$ is a proper CDF function and $0 < l < \infty$.

Proof: Re-writing (7), we get

$$F(n,x) = \frac{\sum_{i=1}^n \lambda_i F_i(x)}{n} \cdot \frac{1}{\frac{1}{n}\sum_{i=1}^n \lambda_i}.$$

Since $\{\lambda_i\}$, $\{F_i(x)\}$ are i.i.d. sequences under Assumption 2, both sample means $\frac{1}{n}\sum_{i=1}^n \lambda_i F_i(x)$ and $\frac{1}{n}\sum_{i=1}^n \lambda_i$ converge as $n \to \infty$ to their expected values w.p. 1 by the strong law of large numbers, which leads to (27). Using the same reasoning for $l(n)$, we obtain (28) and complete the proof. ■

Combining the last two lemmas, we have our main result.

Theorem 2: Given Assumption 2, $H(n,x)$ converges almost surely (a.s.) to the following as $n \to \infty$:

$$H(n,x) \xrightarrow{a.s.} H(x) := \frac{1}{l}\int_0^x (1 - F(u))\, du, \tag{29}$$

where $F(x)$ and l are given in (27)-(28).

Proof: Noting from the Chernoff bound that $P(N(n) \geq 1) \to 1$ as $n \to \infty$, (18) can be transformed into:

$$\lim_{n\to\infty} H(n,x) = \lim_{n\to\infty} \sum_{i=1}^n \frac{a_i H_i(x)}{n} \cdot n\pi_i, \tag{30}$$

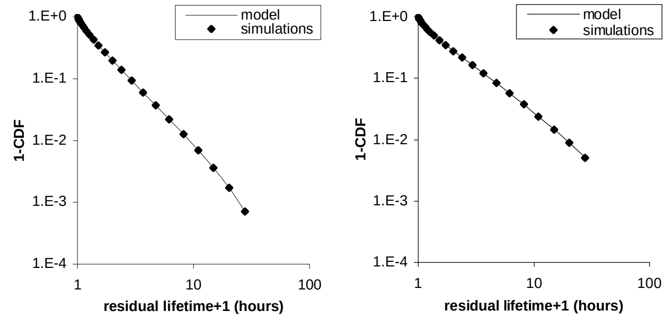
where π_i for uniform selection is given in (15).

We start with $n\pi_i$. Observing that

$$E\Big(\frac{\mu_n}{N(n)+1}|N(n) \geq 1\Big) \leq \mu_n \pi_i \leq E\Big(\frac{\mu_n}{N(n)}|N(n) \geq 1\Big)$$

and applying Lemma 3 to both bounds, we have

$$\lim_{n\to\infty} n\pi_i = \lim_{n\to\infty} \frac{n}{\mu_n} \cdot \mu_n \pi_i = \frac{1}{\sum_{j=1}^{\mathcal{T}} p_j a^{(j)}}, \quad a.s. \tag{31}$$



(a) system $\mathcal{E}$ (b) system $\mathcal{H}$

Fig. 4. Comparison of simulation results of $H(n,x)$ to model (33) in a graph with $n = 1000$ nodes. System age 500 hours and 10^5 iterations.

The second term in (30) simplifies to:

$$\begin{aligned} \sum_{i=1}^n \frac{a_i H_i(x)}{n} &= \frac{\sum_{j=1}^n \lambda_j}{n} \sum_{i=1}^n \Big(\frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \int_0^x (1-F_i(u))du\Big) \\ &\xrightarrow{a.s.} \sum_{j=1}^{\mathcal{T}} \big(p_j \lambda^{(j)}\big) \int_0^x (1-F(u))du. \end{aligned} \tag{32}$$

Combining (31)-(32) and recalling l in (28) yield (29). ■

Leveraging this theorem allows us to use the following approximation:

$$H(n,x) \approx \frac{1}{l(n)} \int_0^x (1 - F(n,u))\, du = \frac{\sum_{i=1}^n a_i H_i(x)}{\sum_{i=1}^n a_i}, \tag{33}$$

which we next examine in simulations with relatively small networks. As shown in Fig. 4 for the exponential and Pareto cases, simulation results of $H(n,x)$ match the model very well and also demonstrate that $\mathcal{E}$ may produce residual lifetime distributions that appear to be non-exponential. In practice, $n \geq 50$ is often sufficient to keep (33) very accurate (simulations omitted for brevity).

Note that (29) is very important since it shows that in practice one only needs to measure the aggregate lifetime distribution $F(x)$ and its mean l rather than each $F_i(x)$ and each user availability a_i in order to obtain the residual lifetime distribution of a uniformly selected neighbor. Assuming from measurement studies [5], [9], [17], that $F(x)$ is Pareto with $F(x) = 1 - (1 + x/\beta)^{-\alpha}$, (29) reduces to:

$$H(x) = 1 - (1 + x/\beta)^{-(\alpha-1)}. \tag{34}$$

Comparing (34) to $F(x)$, we see that residuals are stochastically larger than user lifetimes, which implies that a uniformly selected user is more reliable than new arrivals in terms of failure. For other neighbor selection strategies, it is important to realize that the distribution of residual lifetimes may be completely different from (29) and should be analyzed accordingly.

D. Lifetime of Users in the System

Denote by $J(n,x)$ the equilibrium lifetime distribution of users *currently* in the system conditioned on $N(n,t) \geq 1$. As observed in [28], distribution $J(n,x)$ is clearly different from $F(n,x)$; however, no closed-form analysis has been made available to date. The intuitional rationale behind this difference is that lifetimes of the peers observed in the system are biased towards larger values, which is commonly known as the *inspection paradox* [30]. Below, we formally derive $J(n,x)$ is as a simple function of $F(n,x)$ for $n \to \infty$.

Denote by $J_i(x) := \left(xF_i(x) - \int_0^x F_i(u)du\right)/l_i$ the CDF of the current ON cycle of user i given that it is "inspected" at $t \gg 0$, i.e., its spread [30]. Since $J(n,x)$ is the same as the lifetime distribution of a uniformly randomly selected user from the set of live peers, we reach the next result following the analysis in Theorem 2 for uniform selection.

Corollary 2: Given Assumption 2, the lifetime distribution $J(n,x)$ of living users converges a.s. as $n \to \infty$:

$$J(n,x) \xrightarrow{a.s.} J(x) := \frac{1}{l}\left(xF(x) - \int_0^x F(u)du\right), \quad (35)$$

where all parameters are the same as in Theorem 2.

The accuracy of (35) for finite n was confirmed in simulations, but is omitted here for brevity. Exponential lifetimes $F(x)$ imply that $J(x)$ is the Erlang(2) distribution with mean $2E(L)$. For Pareto $F(x)$, spread $J(x)$ has no closed-form expression, but is clearly more heavy-tailed than $F(x)$. The next result summarizes these observations, as well as those of [28], in more formal terms.

Corollary 3: With Assumption 2, spread distribution $J(x)$ is stochastically larger than $F(x)$ and the mean lifetime of a user currently alive in the system is double the mean residual lifetime of a uniformly selected user.

In conclusion, our results demonstrate that given heterogeneous users and uniform selection of neighbors, both metrics $H(x)$ and $J(x)$ can be reduced to the aggregate behavior $F(x)$ of joining users as long as $n \gg 1$. The rest of the paper shows that $F(x)$ in such systems can be additionally used to obtain the distribution of in-degree as a function of users' age and thus completely characterize local resilience of unstructured P2P networks.

V. Edge Arrival

Before analyzing node in-degree under uniform selection, we study the process of edge arrival into each user since this determines both the rate at which the user accumulates incoming neighbors and the stationary in-degree distribution. Our neighbor churn model prescribes that each joining user find k random out-degree neighbors and then continuously replace them as they fail, as in [14]. Define *initial edges* to be those added when users arrive in the system and *replacement edges* to be those added in response to neighbor failures.

Assumption 3: The number of neighbors k a user selects upon joining the system is a constant for all n.

This assumption often holds in unstructured P2P networks where individual users are unaware of system size (e.g., Gnutella) and some structured P2P networks with constant node degree (e.g., de Bruijn [19]).

A. Definitions

Considering the time-limiting behavior of the system (i.e., $t \to \infty$), the rest of the paper assumes that $Z_i := \{Z_i(t)\}_{t \geq 0}$ are *stationary* alternating renewal processes on time interval $[0,\infty)$, for $i = 1,2,\ldots,n$. Denote by $\{\tau_{i,m}\}_{m=1}^{\infty}$ arrival times of user i. Then, $\tau_{i,m+1} = \tau_{i,m} + L_{i,m} + D_{i,m}$, for $m \geq 1$. To ensure stationarity, let $\tau_{i,1} := L_i^e + D_i$ w.p. a_i and otherwise $\tau_{i,1} := D_i^e$, where L_i^e has the equilibrium distribution of $F_i(x)$ (i.e., $L_i^e \sim H_i(x)$ in (11)) and D_i^e has that of $G_i(x)$. Define $M_i(t)$ to be the number of arrivals of user i in interval $[0,t]$:

$$M_i(t) := \sum_{m=1}^{\infty} 1_{\tau_{i,m} \in [0,t]}, \quad (36)$$

whose expectation (due to stationarity) is $E(M_i(t)) = \lambda_i t$ for any $t \geq 0$, where λ_i is given in (4).

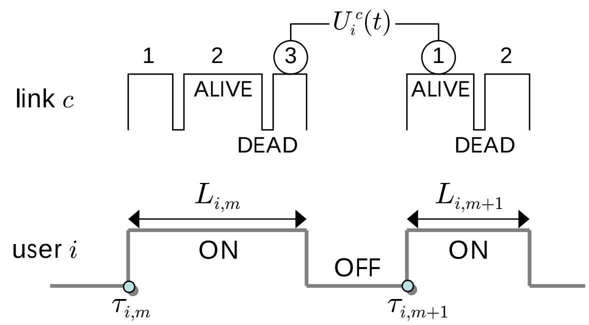


Fig. 5. Process $\{Y_i^c(n,t)\}$ indicates DEAD/ALIVE behavior of the c-th out-link of user i. Process $\{U_i^c(n,t)\}$ counts the number of DEAD→ALIVE transitions within the current ON cycle of i.

Recall that in our resilience model, each user i has k out-degree neighbors, which are either dead (i.e., a replacement is being sought) or alive at any given time. Let $Y_i^c := \{Y_i^c(n,t)\}$ be an alternating process indicating the state of i's link c:

$$Y_i^c(n,t) := \begin{cases} 1 & \text{out-link } c \text{ of user } i \text{ is ALIVE at } t \\ 0 & \text{otherwise (DEAD)} \end{cases}, \quad (37)$$

for $c = 1,\ldots,k$, where neighbors connected by link c are chosen from existing peers. If i is offline at t, all of its links are considered dead. The out-degree of user i at time t is simply $\sum_{c=1}^{k} Y_i^c(n,t)$. Whenever Y_i^c transitions from DEAD to ALIVE, user i delivers one edge into the system (i.e., performs one selection). Thus, processes $\{Y_i^c\}_{i,c}$ determine the edge-generation rate of individual users.

As illustrated in Fig. 5, link c becomes ALIVE at arrival times $\{\tau_{i,m}\}_{m\geq 1}$ and then alternates between DEAD/ALIVE states during i's ON periods. Note that ALIVE durations of Y_i^c are neighbor residual lifetimes and DEAD durations are search delays for finding replacement neighbors, with the exception of the very last ALIVE cycle in each ON period, which is terminated by i's departure rather than neighbor failure. To save space, we assume that search delays are negligible compared to $L_{i,m}$ and do not explicitly model their effect on the in-degree process.[4]

The figure also shows right-continuous process $\{U_i^c(n,t)\}$, which is the number of transitions DEAD→ALIVE of Y_i^c within the current ON cycle up to time t. We assume

[4]Search delays can be accommodated by changing link residual lifetime $R(t)$ to $R(t) + S(t)$, where $S(t)$ is the search delay at t.

$U_i^c(n,\tau_{i,m})=1$ for all $m\geq 1$, use notation t^- to represent the instant just prior to t, and denote by

$$U_i^c(n,\tau_{i,m+1}^-)=\sup_{\tau_{i,m}\leq t<\tau_{i,m+1}} U_i^c(n,t) \tag{38}$$

the number of selections for link c in the m-th ON cycle.

Denote by $\{\delta_{i,z}^c\geq 0\}_{z=1}^{\infty}$ random times at which Y_i^c becomes ALIVE (i.e., an edge is generated by i and delivered to some target peer). Define $W_i^c(n,t):=\sum_{z=1}^{\infty}1_{\delta_{i,z}^c\in[0,t]}$ to be the number of selections for link c in $[0,t]$, which is

$$W_i^c(n,t)=\sum_{m=1}^{M_i(t)}U_i^c(n,\tau_{i,m}^-)-U_i^c(n,0)+U_i^c(n,t). \tag{39}$$

Finally, denote by $W_i(n,t):=\sum_{c=1}^{k}W_i^c(n,t)$ the number of edges delivered by i into the system in $[0,t]$. Observe that $\sum_{i=1}^{n}W_i(n,t)$ is the number of out-degree edges generated by n users in $[0,t]$, which is the same as the number of in-degree edges received by living users in $[0,t]$.

B. Edge Creation Process

This subsection deals with the rate of edge generation from each user. We first turn to the issue of uniform integrability (UI) of the number of edges a user creates in interval $[0,t]$ for each $t>0$.

Lemma 5: Given Assumptions 2-3, uniform selection, and each user i and $t>0$, collections $\{(W_i(n,t))^r\}_{n\geq 1}$ are uniformly integrable in n, for any $r>0$.

Proof: The key aspect of this proof is to show that $W_i(n,t)$ defined in (39) is stochastically smaller than some integrable r.v. $\widehat{U}$, which is independent of n. This automatically implies UI of all r.v. stochastically smaller than $\widehat{U}$ [24, pp. 183]. The major impediment to achieving this is that uniform selection allows repeated connections to the same user (during different lifetimes of this user), which creates dependency of residuals of acquired neighbors. While for $n\to\infty$ this dependency diminishes, our analysis in this proof takes it into account and creates a foundation that will be used in the derivations that follow in the next section.

To proceed, call a *new* neighbor of user i if it is different from any previous selections that i has made for link c; otherwise, a *reconnected* neighbor. Denote by $H^{(j)}(x):=(l^{(j)})^{-1}\int_0^x(1-F^{(j)}(u))du$ the residual CDF for user-type j and by $\widehat{H}(x):=\max_{1\leq j\leq\mathcal{T}}H^{(j)}(x)$. Since users are independent and stationary, it is clear that residuals of i's new neighbors follow $H^{(j)}(x)$ for some j and that $\widehat{H}(x)$ represents the worse-case scenario for residuals of new selections.

Next, fix time $s>0$ from when a new neighbor (call user u) of user i departed to the current time. Define $B^{(j)}(s,x)$ to be the CDF of u's residual lifetime for type j at time s, given that u is ON at s, which represents the residual CDF of i's reconnected neighbors. Our main task is to show that $B^{(j)}(s,x)$ is non-trivial for all s, i.e.,

$$\exists x_0>0 \text{ such that } \sup_{s>0}B^{(j)}(s,x_0)<1. \tag{40}$$

Let $S_{u,0}=0$ and $S_{u,m}=S_{u,m-1}+D_{u,m}+L_{u,m}$ be the m-th random departure time of u since u was selected as i's new neighbor. We start by considering the probability $C^{(j)}(s,x)$ that u is ON at time s and has a residual greater than x, i.e.:

$$C^{(j)}(s,x):=\sum_{m=1}^{\infty}P(S_{u,m-1}+D_{u,m}\leq s<s+x<S_{u,m}).$$

It is not hard to see from the above definition that under the assumption that $F^{(j)}(x)>0$ and $G^{(j)}(x)>0$ for all $x>0$ and all $j=1,\ldots,\mathcal{T}$, there exists some $x_0>0$ such that $1-F^{(j)}(x_0)>0$ and $C^{(j)}(s,x_0)>0$ for all $s>0$. This also implies that $C^{(j)}(s,0)>0$ for all $s>0$. Then, conditioning on u being ON at s, $B^{(j)}(s,x)$ is given by

$$B^{(j)}(s,x)=1-\frac{C^{(j)}(s,x)}{C^{(j)}(s,0)}. \tag{41}$$

This leads to that $B^{(j)}(s,x_0)<1$ for all $s>0$. Our next step is to examine $B^{(j)}(s,x_0)$ for $s\to 0$ and $s\to\infty$, respectively.

Observe that $C^{(j)}(s,x)$ can be transformed into

$$\begin{aligned}C^{(j)}(s,x)&=\sum_{m=1}^{\infty}P(S_{u,m}>s+x|S_{u,m-1}+D_{u,m}\leq s)\\&\qquad\cdot P(S_{u,m-1}+D_{u,m}\leq s)\\&\geq\sum_{m=1}^{\infty}P(L_{u,m}>x+s)P(S_{u,m-1}+D_{u,m}\leq s)\\&\geq P(L_{u,1}>x+s)C^{(j)}(s,0),\end{aligned} \tag{42}$$

from which we obtain

$$\liminf_{s\downarrow 0}\frac{C^{(j)}(s,x_0)}{C^{(j)}(s,0)}\geq 1-F^{(j)}(x_0)>0. \tag{43}$$

Combining (41) and (43) and noting that $C^{(j)}(s,x_0)$ is a continuous function of s and that $B^{(j)}(s,x_0)\to H^{(j)}(x_0)<1$ as $s\to\infty$, we establish (40).

Now, define $\widehat{B}^{(j)}(x):=\sup_{s>0}B^{(j)}(s,x)$, $x>0$. It is easy to see that $\widehat{B}^{(j)}(x)$ is monotone and bounded by 1. Moreover, since $\widehat{B}^{(j)}(x)\geq H^{(j)}(x)$, $\lim_{x\to\infty}\widehat{B}^{(j)}(x)=1$. Therefore, after adjust it to be right continuous (which can only increase its value), $\widehat{B}^{(j)}(x)$ is a non-trivial CDF on $[0,\infty)$. It is then clear that $\widehat{B}(x)$ defined below is non-trivial:

$$\widehat{B}(x):=\max_{1\leq j\leq\mathcal{T}}\hat{B}^{(j)}(x)\geq\hat{H}(x). \tag{44}$$

It follows from (44) that both new and reconnected neighbors have residual lifetime CDFs no more than $\widehat{B}(x)$. Denote by $\widehat{U}:=\{\widehat{U}(t)\}$ a pure renewal process whose cycle lengths follows distribution $\widehat{B}(x)$. It is known that given non-trivial $\widehat{B}(x)$, $\widehat{U}(t)$ has all moments finite [23, pp. 186]. Noting that $U_i^c(n,\tau_{i,m}^-)$'s depend on $Z_i(t)$ only through the values of i's ON durations and denoting by

$$L'_{i,m}:=\begin{cases}\tau_{i,1}\wedge t & m=0\\ L_{i,m} & 1\leq m<M_i(t)\\ t-\tau_{i,M_i(t)} & m=M_i(t),m\geq 1\end{cases},$$

we get $W_i^c(n,t) \leq^{st} \sum_{m=0}^{M_i(t)} \widehat{U}(L'_{i,m})$. Since $m \leq M_i(t)$ implies $L'_{i,m} \leq t$, it follows that

$$E[(W_i^c(n,t))^r] \leq E\Big[\Big(\sum_{m=0}^{M_i(t)} \widehat{U}(t)\Big)^r\Big] \\ = E[(M_i(t)+1)^r]E[\widehat{U}(t))^r] < \infty, \quad (45)$$

for all $r > 0$ and all n, i. It is then clear that $E[(W_i(n,t))^r] \leq k^{r-1}E[\sum_{c=1}^{k}(W_i^c(n,t))^r] < \infty$, indicating UI of $\{(W_i(n,t))^r\}_{n\geq 1}$ for all $r > 0$. ■

Armed with Lemma 5, we next focus on bounding the distribution of the number of available users, among which user i search for neighbors during its lifetimes, and then develop a closed-form result on the edge creation rate from each user for $n \to \infty$.

For convenience, denote by $\delta_{i,z}$ the z-th time at which user i makes a selection, across all links of i, and by $I_{i,z}^j$ the indicator that user i selects peer j for its z-th connection[5]. Observe that selection time $\delta_{i,z}$ is determined by Z_i and residual lifetimes of i's *previous* selections $z' < z$. Define $\mathcal{H}_{i,t} := \sigma(\{Z_i(s)\}_{s\leq t}, \{W_i(n,s)\}_{s\leq t})$ to be the σ-field consisting of the information observed by user i up to time t, including its connection times but not including which users were selected. For any stopping time δ, denote by

$$\mathcal{H}_{i,\delta} = \{A : A \cap \{\delta \leq t\} \in \mathcal{H}_{i,t} \text{ for all } t \geq 0\}, \quad (46)$$

the usual σ-field representing information up to and including the stopping time. Further, let $\mathcal{G}_{i,z} =: \sigma(\{I_{i,z'}^j\}_{z'<z, j\neq i})$, which is the selection information *prior* to the z-th selection. Finally, define $\mathcal{F}_{i,z} := \sigma(\mathcal{H}_{i,\delta_{i,z}} \cup \mathcal{G}_{i,z})$, which is all the information observed by i up to time $\delta_{i,z}$ *except* the z-th and later selections.

Note that the number of available users for the z-th selection of user i, $N_n^i(\delta_{i,z}) := \sum_{j=1,j\neq i}^{n} Z_j(\delta_{i,z})$, is dependent on $\mathcal{F}_{i,z}$, since $\mathcal{F}_{i,z}$ provides information that helps predict whether the previously selected users are alive at $\delta_{i,z}$. The next lemma examines properties of $N_n^i(\delta_{i,z})$, given the previously selections and the connection times observed by user i.

Lemma 6: Assuming the same as in Lemma 5, we have

$$E\Big(\frac{1}{\max(N_n^i(\delta_{i,z}),1)} \;|\mathcal{F}_{i,z}\Big) \leq E\Big(\frac{1}{\max(N(n-1)-z,1)}\Big),$$
$$E\Big(\frac{1}{\max(N_n^i(\delta_{i,z}),1)} \;|\mathcal{F}_{i,z}\Big) \geq E\Big(\frac{1}{N(n)+z}\Big), \quad (47)$$

where $N(n) \sim \text{Bernoulli}(n,a)$ and a is the user availability in (6).

Proof: Denote by $X_{i,z}$ the number of (distinct) users that were selected by user i *prior* to the z-th selection and by $Y_{i,z}$ the number of previously *unselected* users that are alive at $\delta_{i,z}$. Since $X_{i,z} < z$, the number of available users for the z-th selection of user i is bounded by

$$Y_{i,z} \leq N_n^i(\delta_{i,z}) \leq Y_{i,z} + z. \quad (48)$$

We next consider the distribution of $Y_{i,z}$, conditioned on $\mathcal{F}_{i,z}$. At i's 1^{st} selection, the number of previously unselected users $Y_{i,1}$ is distributed as $N(n-1)$ due to user independency. Given that *some* user j is connected at $\delta_{i,1}$,

$$P\Big(\sum_{u\neq i,j} Z_u(\delta_{i,1}) = k \;|I_{i,1}^j = 1\Big) \\ = P\Big(\sum_{u\neq i} Z_u(\delta_{i,1}) = k+1 \;|Z_j(\delta_{i,1}) = 1\Big) \\ = P(N(n-1) = k+1 | N(n-1) \geq 1), \quad (49)$$

upon noting that $\delta_{i,1}$ depends only on Z_i, which indicates that $\sum_{u\neq i,j} Z_k(\delta_{i,1})$ has the distribution of $N(n-1)-1$, conditioned on $N(n-1) \geq 1$. However, $\sum_{u\neq i,j} Z_u(\delta_{i,1}+t)$ is not stationary since conditional on j having been selected, it does not have the stationary distribution at $t=0$. Instead, as t increases, it is converging monotonically to its stationary distribution, i.e., that of $N(n-2)$. Noting that the second selection time $\delta_{i,2}$ depends only on Z_i and Z_j, we then obtain that the number of unselected users $Y_{i,2} = \sum_{u\neq i,j} Z_u(\delta_{i,2})$ has a distribution between those of $N(n-1)-1$, conditioned on $N(n-1) \geq 1$, and $N(n-2)$; i.e., given $\mathcal{F}_{i,2}$, $Y_{i,2}$ is stochastically bounded by

$$N(n-1) - 1 \leq^{st} Y_{i,2} \leq^{st} N(n-2) \leq^{st} N(n). \quad (50)$$

Likewise, given that there were $z-1$ users selected by i prior to the z-th selection and conditioned on $N(n-1) \geq z-1$,

$$N(n-1) - (z-1) \leq^{st} Y_{i,z} \leq^{st} N(n). \quad (51)$$

Combining (51) with (48) and using the simplification $X^{-1}1_{X\geq 1} + 1_{X=0} = (\max(X,1))^{-1}$ lead to (47). ■

Motivated by Lemma 6, define the following bounding sequences for analyzing each selection of user i:

$$\overline{b}_n := nE\Big(\frac{1}{\max(N(n-1)-\sqrt{n},\ 1)}\Big),$$
$$\underline{b}_n := nE\Big(\frac{1}{N(n)+\sqrt{n}}\Big). \quad (52)$$

Recalling (31) and $a = \sum_{j=1}^{\mathcal{T}} p_j a^{(j)}$, we obtain

$$\lim_{n\to\infty} \underline{b}_n = \lim_{n\to\infty} \overline{b}_n = 1/a. \quad (53)$$

The main result of this subsection is presented next.

Lemma 7: With Assumptions 2-3, uniform selection, and each $t \geq 0$, we have:

(1) Residuals $\{R(n,\delta_{i,z})\}_{z\geq 1}$ of selected neighbors at random times $\delta_{i,z}$ converge in distribution as $n \to \infty$ to i.i.d. r.v.'s with CDF $H(x)$ in (29);
(2) Unconditioned on user types, the mean number of selections of each user i in $[0,t]$ is

$$\lim_{n\to\infty} E(W_i(n,t)) = (k+\theta)\lambda t, \quad (54)$$

where λ is the user-arrival rate in (28), and $\theta := k\sum_{r=1}^{\infty}\int_0^{\infty} H^{*r}(x)dF(x)$ is the mean number of replacement edges generated per lifetime, and $F(x)$ is the lifetime CDF in (27).

Proof: We prove each of the statements in sequence.

[5] As i makes k selections simultaneously at the beginning of its lifetimes, we order these k selections according to the link index c.

1) Residuals: Let $J_{i,z}$ be the indicator that a previous selection is again selected by i at time $\delta_{i,z}$. Our first task is to compute the expected number of repeat selections made by i in $[0,t]$:

$$E(\sum_z J_{i,z}1_{\delta_{i,z}\le t}) = E(\sum_z J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}}) + E(\sum_z J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)\ge\sqrt{n}}). \quad (55)$$

Since at most z users were chosen by i prior to $\delta_{i,z}$,

$$\begin{aligned} &E(J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}} \mid \mathcal{F}_{i,z}) \\ &\le E\Big(\frac{z}{\max(N_n^i(\delta_{i,z}),\ 1)} \mid \mathcal{F}_{i,z}\Big)1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}} \\ &\le \frac{\overline{b}_n}{n}z1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}} \le \frac{\overline{b}_n}{\sqrt{n}}1_{\delta_{i,z}\le t}, \end{aligned} \quad (56)$$

by Lemma 6, where $\overline{b}_n$ is in (52). It follows that

$$\begin{aligned} &E\Big(\sum_z J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}}\Big) \\ &\le E\Big[\sum_z E\big(J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)\le\sqrt{n}} \mid \mathcal{F}_{i,z}\big)\Big] \\ &\le \frac{\overline{b}_n}{\sqrt{n}}E[W_i(n,t)] \to 0, \quad n\to\infty. \end{aligned} \quad (57)$$

On the other hand,

$$E\Big(\sum_z J_{i,z}1_{\delta_{i,z}\le t}1_{W_i(n,t)>\sqrt{n}}\Big) \le E(W_i(n,t)1_{W_i(n,t)>\sqrt{n}}),$$

which goes to 0 as $n\to\infty$ by recalling UI of $W_i(n,t)$.

It is now clear that $E(\sum_z J_{i,z}1_{\delta_{i,z}\le t})$ in (55) approaches 0. From the Markov's inequality, we then have

$$P\Big(\sum_z J_{i,z}1_{\delta_{i,z}\le t}\ge 1\Big) \le E\Big(\sum_z J_{i,z}1_{\delta_{i,z}\le t}\Big) \to 0, \quad (58)$$

showing that the probability that user i selects any other user more than once goes to 0. Therefore, residuals $\{R(n,\delta_{i,z})\}_{1\le z\le w}$ of i's first w selections are asymptotically independent (due to the independence of different users), which in turn implies that $\{R(n,\delta_{i,z})\}_{z\ge1}$ are asymptotically independent. Moreover, since these residual times are independent of Z_i, we reach that $R(n,\delta_{i,z})$'s have the same limiting CDF (29) of $R(n,t)$ selected at time t.

2) Regenerative Property: Let $N_i^{(j)}(n-1,t)$ be the number of live users, other than user i, with type j. For fixed n,

$$\bar{N}_i(n-1,t) = (N_i^{(1)}(n-1,t),\dots,N_i^{(\mathcal{T})}(n-1,t))$$

is a stationary process independent of $Z_i(t)$. Consequently, $(U_i^1(n,t),\dots,U_i^k(n,t))$ is regenerative (as a process in t). As $\{U_i^c(n,\tau_{i,m}^-)\}_{m\ge2}$ have the same distribution, it follows from Smith's theorem for regenerative processes that the mean number of selections made by i for link c in $[0,t]$ is

$$E(W_i^c(n,t)|i\text{'s type}) = \lambda_i t E(U_i^c(n,\tau_{i,2}^-)|i\text{'s type}), \quad (59)$$

where $\lambda_i t = t/(l_i+d_i)$ is the mean number of arrivals of i in $[0,t]$.

The following is directly obtained from renewal theory. Denote by $\{U(s)\}_{s\ge0}$ a pure renewal process with cycle lengths $R_r\sim H(x)$ for $r\ge1$. The mean number of renewals in $[0,s]$ is

$$E(U(s)) = \sum_{r=0}^{\infty}P(U(s)>r) = 1+\sum_{r=1}^{\infty}H^{*r}(s). \quad (60)$$

It follows from part (1) that for $n\to\infty$, $U_i^c(n,\tau_{i,m+1}^-)$ for link c has the same distribution of $U(L_{i,m})$, where lifetime $L_{i,m}\sim F_i(x)$ is independent of $\{U(s)\}$. This leads to

$$\begin{aligned} &\lim_{n\to\infty}E(U_i^c(n,\tau_{i,2}^-)|i\text{'s type}) = E(U(L_{i,2})|i\text{'s type}) \\ &= \int_0^\infty E[U(x)]dF_i(x) = 1+\sum_{r=1}^{\infty}\int_0^\infty H^{*r}(x)dF_i(x), \end{aligned}$$

which is finite given the mean lifetime $l_i<\infty$. Combining the above with (59) and unconditioning on user types lead to

$$\begin{aligned} &\lim_{n\to\infty}E(W_i(n,t)) = k\lim_{n\to\infty}E(W_i^c(n,t)) \\ &= kt\sum_{j=1}^{\mathcal{T}}p_j\lambda^{(j)}\Big(1+\sum_{r=1}^{\infty}\int_0^\infty H^{*r}(x)dF^{(j)}(x)\Big). \end{aligned}$$

By (27), the above yields (54). ■

The result in (54) shows that as $n\to\infty$, each user brings k initial edges and an average of $\theta<\infty$ replacement edges per ON/OFF cycle into the system and that the asymptotical edge creation rate from each user approaches $(k+\theta)\lambda$. We leverage the fact that $(k+\theta)\lambda$ is a constant in the next subsection.

C. Edge Arrival Process

Now, given a set of n participating users, our approach is to set aside a *given* user v (so the values of $Z_v(\cdot)$ are given) and examine edge-arrivals to this user, from $n-1$ other peers under uniform selection.

Define $\xi_{n,i}(t) := \sum_{z=1}^{W_i(n,t)}I_{i,z}^v$ to be the number of edges delivered by user i to node v in $[0,t]$, for $i\ne v$, where $I_{i,z}^v$ is the indicator that i selects v at time $\delta_{i,z}$. Then, the edge arrival process from the system to user v is

$$\xi_n(t) := \sum_{i=1,i\ne v}^{n}\xi_{n,i}(t) = \sum_{i=1,i\ne v}^{n}\sum_{z=1}^{W_i(n,t)}I_{i,z}^v. \quad (61)$$

Denote by $\bar{W}_n(t) = n^{-1}\sum_{i=1}^n W_i(n,t)$ the number of selections made by a user in $[0,t]$, averaged over n users. In what follows, we first focus our attention on the properties of $\bar{W}_n(t)$ and then utilize them to analyze process $\xi_n := \{\xi_n(t)\}$.

Lemma 8: Given Assumptions 2-3 and uniform selection,

$$\lim_{n\to\infty}E[(\bar{W}_n(t))^r] = [(k+\theta)\lambda t]^r, \quad (62)$$

for each t and all $r>0$. Moreover, $\bar{W}_n(t)\to(k+\theta)\lambda t$ in distribution as a stochastic process (i.e., the functional convergence in distribution).

Proof: Note that $W_i(n,t)$'s, $1\le i\le n$, are identically distributed, but are dependent. It is easy to get from (54) that

$$E(\bar{W}_n(t)) = n^{-1}nE(W_i(n,t)) = (k+\theta)\lambda t. \quad (63)$$

Next, we compute the covariance of $W_i(n,t)$ and $W_j(n,t)$ for any pair of users $i\ne j$. By an extension of the discussion about residuals in Lemma 7, residuals $\{R(n,\delta_{u,z})\}_{u=i,j,\ z\geq 1}$ of users selected by i,j in $[0,t]$ are asymptotically independent and thus $W_i(n,\cdot)$ and $W_j(n,\cdot)$ are asymptotically independent stochastic processes on $[0,t]$. Moreover, since $W_i(n,t)W_j(n,t) \leq (W_i(n,t))^2 + (W_j(n,t))^2$, $W_i(n,t)W_j(n,t)$ is uniformly integrable. It then follows that

$$\lim_{n\to\infty} E(W_i(n,t)W_j(n,t)) = \lim_{n\to\infty} E(W_i(n,t))E(W_j(n,t)),$$

showing $\operatorname{cov}(W_i(n,t),W_j(n,t)) \to 0$. This in conjunction with the fact that all pairs $(W_i(n,t),W_j(n,t))$ have the same distribution yields

$$\begin{aligned}&\operatorname{var}(\bar{W}_n(t))\\ &= \frac{(n-1)\operatorname{cov}(W_i(n,t),W_j(n,t))}{n} + \frac{\operatorname{var}(W_i(n,t))}{n},\end{aligned} \tag{64}$$

which approaches 0 as $n\to\infty$ for each t.

By (63)-(64), $\bar{W}_n(t) \to (k+\theta)\lambda t$ in distribution for each t. Moreover, UI of $\{(W_i(n,t))^r\}_{n\geq 1, i\leq n}$ shown in Lemma 5 implies that $(\bar{W}_n(t))^r$ is uniformly integrable, which then establishes (62). Finally, since $\bar{W}_n(t)$ is right-continuous and has left-hand limits, it follows immediately that we have functional convergence in distribution. ■

Equipped with Lemmas 5-8, we are ready to establish the result on the edge arrival process to v, conditioned on Z_v.

Theorem 3: Under Assumptions 2-3 and uniform selection, the point process ξ_n defined in (61) converges in distribution as $n\to\infty$ to a non-homogeneous Poisson process ξ with local rate $\gamma Z_v(t)$, where $Z_v(t)$ is a deterministic function of t,

$$\gamma := (k+\theta)/l, \tag{65}$$

θ is in (54), and l is the mean lifetime given in (28).

Proof: We set ξ to be a non-homogeneous Poisson process with finite local rate $\gamma Z_v(t)$. It has been shown in [22, Proposition 3.22] that ξ_n converges in distribution to ξ under the following constraints:

(1) The probability that no point occurs exactly at time t is $P(\xi_n(\{t\})=0)\to 1$ as $n\to\infty$;
(2) $\forall t>0$: $\lim_{n\to\infty} E(\xi_n(t)|Z_v) = E(\xi(t)|Z_v) = \gamma\int_0^t Z_v(u)du$; and
(3) $\forall t>0 : \lim_{n\to\infty} P(\xi_n(t)=0|Z_v) = P(\xi(t)=0|Z_v) = \exp(-E(\xi(t)|Z_v))$.

We establish these conditions next.

1) Continuity: This condition is trivially met since the first, and thus the remaining, arrival times of any user i have an absolutely continuous distribution, which is ensured by stationarity and non-lattice lifetime distributions.

2) Mean Convergence: This step is to compute the conditional intensity measure $E(\xi_n(t)|Z_v) = (n-1)E(\xi_{n,i}(t)|Z_v)$.

Recalling the UI of $(W_i(n,t))^r$, we have

$$\begin{aligned}&nE(\xi_{n,i}(t)1_{W_i(n,t)>\sqrt{n}} \mid Z_v)\\ &\quad\leq nE(W_i(n,t)1_{W_i(n,t)>\sqrt{n}} \mid Z_v)\\ &\quad\leq E((W_i(n,t))^3 1_{W_i(n,t)>\sqrt{n}} \mid Z_v) \to 0,\end{aligned} \tag{66}$$

as $n\to\infty$. On the other hand, applying Lemma 6 yields

$$\begin{aligned}&nE\left(\xi_{n,i}(t)1_{W_i(n,t)\leq\sqrt{n}} \mid Z_v\right)\\ &= nE\bigg(\sum_z E(I^v_{i,z}1_{\delta_{i,z}\leq t}1_{W_i(n,t)\leq\sqrt{n}} \mid \mathcal{F}_{i,z}, Z_v) \mid Z_v\bigg)\\ &\leq nE\bigg(\sum_z \frac{\overline{b}_n}{n} Z_v(\delta_{i,z})1_{\delta_{i,z}\leq t}1_{W_i(n,t)\leq\sqrt{n}}) \mid Z_v\bigg)\\ &= \overline{b}_n E\bigg(\sum_{z=1}^{W_i(n,t)} Z_v(\delta_{i,z})1_{W_i(n,t)\leq\sqrt{n}} \mid Z_v\bigg).\end{aligned} \tag{67}$$

where $\overline{b}_n$ is in (52). Likewise, we obtain the lower bound:

$$\begin{aligned}&nE\left(\xi_{n,i}(t)1_{W_i(n,t)\leq\sqrt{n}} \mid Z_v\right)\\ &\quad\geq \underline{b}_n E\bigg(\sum_{z=1}^{W_i(n,t)} Z_v(\delta_{i,z})1_{W_i(n,t)\leq\sqrt{n}} \mid Z_v\bigg).\end{aligned} \tag{68}$$

Next, using Lemma 8 and noting that $\bar{W}(n,t)$ is asymptotically independent of Z_v, we get that conditioned on Z_v, $\int_0^t Z_v(u)\,\bar{W}(n,du) \to \lambda(k+\theta)\int_0^t Z_v(u)\,du$ in distribution and that

$$E\Big(\int_0^t Z_v(u)\bar{W}(n,du)\ |Z_v\Big) \to \lambda(k+\theta)\int_0^t Z_v(u)\,du.$$

Meanwhile, we readily reach

$$\begin{aligned}&E\bigg(\sum_{z=1}^{W_i(n,t)} Z_v(\delta_{i,z})1_{W_i(n,t)>\sqrt{n}}\ |Z_v\bigg)\\ &\quad\leq E\big(W_i(n,t)1_{W_i(n,t)>\sqrt{n}}|Z_v\big) \to 0, \quad n\to\infty,\end{aligned}$$

by UI of $W_i(n,t)$. It then follows that for each i,

$$\begin{aligned}&\lim_{n\to\infty} E\bigg(\sum_{z=1}^{W_i(n,t)} Z_v(\delta_{i,z})1_{W_i(n,t)\leq\sqrt{n}})\ |Z_v\bigg)\\ &\quad= \lim_{n\to\infty} E\bigg(\sum_{z=1}^{W_i(n,t)} Z_v(\delta_{i,z})\ |Z_v\bigg)\\ &\quad= \lim_{n\to\infty} E\Big(\frac{1}{n}\sum_{i=1}^{n}\int_0^t Z_v(u)W_i(n,du)\ |Z_v\Big)\\ &\quad= \lambda(k+\theta)\int_0^t Z_v(u)\,du.\end{aligned} \tag{69}$$

Finally, combining (67), (68) and (69) and using (53) yield

$$\begin{aligned}&\lim_{n\to\infty} nE(\xi_{n,i}(t)1_{W_i(n,t)\leq\sqrt{n}} \mid Z_v)\\ &\quad= \frac{\lambda(k+\theta)}{a}\int_0^t Z_v(u)\,du = \gamma\int_0^t Z_v(u)\,du,\end{aligned} \tag{70}$$

where γ is in (65). Together this result and (66) establish the desired result.

3) Probability Convergence: Next, we focus on the probability that conditioned on Z_v, there is no edge-arrival from the system to v in $[0,t]$, i.e., $P(\xi_n(t)=0|Z_v)$.

Define $\hat{\xi}_n(t)$ to be the number of peers that select user v at least once in $[0,t]$:

$$\hat{\xi}_n(t) := \sum_{i=1,i\neq v}^{n} 1_{\xi_{n,i}(t)\geq 1}. \tag{71}$$

Denote by $\{i_1,\ldots,i_r\}$ a subset of $r>1$ distinct users with each $i_q \neq v$ and by $f(n,r)$ the sum of probabilities that particular subsets of size r are among the users that select v at least once:

$$f(n,r) := E\Big(\sum_{\{i_1,\ldots,i_r\}} \prod_{q=1}^{r} 1_{\xi_{n,i_q}(t)\geq 1} \mid Z_v\Big), \qquad (72)$$

where there are $\binom{n-1}{r}$ terms in the summation. Observe that

$$P(\xi_n(t)=0 \mid Z_v) = P(\hat{\xi}_n(t)=0 \mid Z_v), \qquad (73)$$

i.e., the probability that no user selects v in $[0,t]$, which can be transformed by the inclusion-exclusion principle into:

$$\sum_{r=1}^{2p-1}(-1)^r f(n,r) \leq P(\hat{\xi}_n(t)=0|Z_v)-1 \leq \sum_{r=1}^{2p}(-1)^r f(n,r),$$

for any integer $p\geq 1$. Our task is to show that

$$\lim_{n\to\infty} f(n,r) = \frac{1}{r!}\Big(\gamma\int_0^t Z_v(u)\,du\Big)^r, \qquad (74)$$

for all integer $r\geq 1$, from which the desired result follows:

$$\begin{aligned}
&\lim_{n\to\infty} P(\hat{\xi}_n(t)=0 \mid Z_v)\\
&\quad = 1+\sum_{r=1}^{\infty}\frac{(-1)^r}{r!}\Big(\gamma\int_0^t Z_v(u)\,du\Big)^r = e^{-\gamma\int_0^t Z_v(u)\,du}.
\end{aligned}$$

In the rest of the proof, we first derive $n^r E(\prod_{q=1}^r \xi_{n,i_q}(t)|Z_v)$ and then $n^r E[(\prod_{q=1}^r \xi_{n,i_q}(t) - \prod_{q=1}^r 1_{\xi_{n,i_q}(t)\geq 1}) \mid Z_v]$.

We start by focusing on a particular subset $\{i_1,\ldots,i_r\}$ of $r<\sqrt{n}$ users and their selection times δ_{i_q,z_q}. Considering the first $\sqrt{n}/r-1$ selections of each user i_q, we have

$$\begin{aligned}
g(n,r) &:= E\Big(\prod_{q=1}^{r}\sum_{z_q=1}^{\sqrt{n}/r-1} I^v_{i_q,z_q} 1_{\delta_{i_q,z_q}\leq t} \mid Z_v\Big)\\
&= \sum_{1\leq q\leq r}\sum_{z_q=1}^{\sqrt{n}/r-1} E\Big(\prod_{q=1}^{r} I^v_{i_q,z_q}1_{\delta_{i_q,z_q}\leq t} \mid Z_v\Big). \qquad (75)
\end{aligned}$$

where $E(\prod_{q=1}^r I^v_{i_q,z_q} \mid Z_v)$ is the probability that r *different* users select v at times δ_{i_q,z_p}. Define $\delta'_1<\cdots<\delta'_r$ to be the ordered selection times δ_{i_q,z_q}. Let I'_q by the indicator that user v is selected at δ'_q. Denote by $\mathcal{F}'_q$ the information observed by users $i'_1,...,i'_q$ including the values of $i'_1,...,i'_q$, but not including the selection at δ'_q. It is then clear that, given $\mathcal{F}'_q$ and the history of Z_v, I'_q is conditionally independent of δ'_s for $s>q$, upon which we derive the next inequality.

Notice that given each $z_q\leq\sqrt{n}/r-1$, the total number of selections made by users $\{i_1,\ldots,i_r\}$ is no more than $\sqrt{n}-r$. As in Lemma 6, excluding $\sqrt{n}-r$ previous selections and the r users under consideration, we obtain

$$\begin{aligned}
&E\Big(I'_q\prod_{q<s\leq r} Z_v(\delta'_s) \mid \mathcal{F}'_q, Z_v\Big)\\
&\quad\leq E\Big(\frac{Z_v(\delta'_q)}{\max(N(n-1)-\sqrt{n},\ 1)}\prod_{q<s\leq r} Z_v(\delta'_s) \mid \mathcal{F}'_q, Z_v\Big)\\
&\quad= \frac{\overline{b}_n}{n}E\Big(\prod_{q\leq s\leq r} Z_v(\delta'_s) \mid \mathcal{F}'_q, Z_v\Big). \qquad (76)
\end{aligned}$$

Using (76) and taking conditional expectations recursively in reverse time lead to

$$\begin{aligned}
E\Big(\prod_{q=1}^{r} I^v_{i_q,z_q} \mid Z_v\Big) &= E\Big(\prod_{q=1}^{r} I'_q \mid Z_v\Big)\\
&= E\Big[\Big(\prod_{q=1}^{r-1} I'_q\Big)E\Big(I'_r \mid \mathcal{F}'_r, Z_v\Big) \mid Z_v\Big]\\
&\leq E\Big[\Big(\prod_{q=1}^{r-1} I'_q\Big)\frac{\overline{b}_n}{n}E\Big(Z_v(\delta'_r)|\mathcal{F}'_r, Z_v\Big) \ \mid Z_v\Big]\\
&= \frac{\overline{b}_n}{n}E\Big(\prod_{q=1}^{r-1} I'_q\ Z_v(\delta'_r) \mid Z_v\Big)\\
&\leq\cdots\leq\Big(\frac{\overline{b}_n}{n}\Big)^r E\Big(\prod_{q=1}^{r} Z_v(\delta'_q) \mid Z_v\Big)\\
&= \Big(\frac{\overline{b}_n}{n}\Big)^r E\Big(\prod_{q=1}^{r} Z_v(\delta_{i_q,z_q}) \mid Z_v\Big). \qquad (77)
\end{aligned}$$

Then, applying (77) and setting $n_0 := \sqrt{n}/r-1$, (75) yields

$$\begin{aligned}
g(n,r) &\leq \sum_{\substack{z_q\leq n_0\\ 1\leq q\leq r}}\Big(\frac{\overline{b}_n}{n}\Big)^r E\Big(\prod_{q=1}^{r} Z_v(\delta_{i_q,z_q})1_{\delta_{i_q,z_q}\leq t} \mid Z_v\Big)\\
&= \Big(\frac{\overline{b}_n}{n}\Big)^r E\Big(\prod_{q=1}^{r}\sum_{z_q=1}^{n_0} Z_v(\delta_{i_q,z_q})1_{\delta_{i_q,z_q}\leq t} \mid Z_v\Big). \qquad (78)
\end{aligned}$$

Likewise, we have the lower bound:

$$g(n,r) \geq \Big(\frac{\overline{b}_n}{n}\Big)^r E\Big(\prod_{q=1}^{r}\sum_{z_q=1}^{n_0} Z_v(\delta_{i_q,z_q})1_{\delta_{i_q,z_q}\leq t} \mid Z_v\Big). \qquad (79)$$

Meanwhile, we have

$$\begin{aligned}
&E\Big(\Big|\prod_{q=1}^{r}\sum_{z=1}^{W_{i_q}(n,t)} Z_v(\delta_{i_q,z}) - \prod_{q=1}^{r}\sum_{z=1}^{n_0} Z_v(\delta_{i_q,z})1_{\delta_{i_q,z}\leq t}\Big| \mid Z_v\Big)\\
&\quad\leq 2E\Big(\prod_{q=1}^{r}\sum_{z=1}^{W_{z_q}(n,t)} Z_v(\delta_{i_q,z})1_{W_{z_q}(n,t)>n_0}|Z_v\Big)\\
&\quad\leq 2E\Big(\prod_{q=1}^{r} W_{z_q}(n,t)1_{W_{z_q}(n,t)>\sqrt{n}/r-1}|Z_v\Big)\to 0. \qquad (80)
\end{aligned}$$

Putting (78)-(80) together leads to

$$\begin{aligned}
\lim_{n\to\infty} n^r g(n,r) &= \frac{1}{a^r}\lim_{n\to\infty} E\Big(\prod_{q=1}^{r}\sum_{z=1}^{n_0} Z_v(\delta_{i_q,z})1_{\delta_{i_q,z}\leq t} \mid Z_v\Big)\\
&= \frac{1}{a^r}\lim_{n\to\infty} E\Big(\prod_{q=1}^{r}\int_0^t Z_v(u)\,W_{i_q}(n,du) \mid Z_v\Big). \qquad (81)
\end{aligned}$$

By the similar technique used in the proof of Lemma 8, $\{\int_0^t Z_v(u)\,W_{i_q}(n,du)\}_{1\leq q\leq r}$ are asymptotically independent and converge jointly in distribution, conditional on Z_v. Since

$\{\prod_{q=1}^{r}\int_0^t Z_v(u)\, W_{i_q}(n,du)\}_{n>r}$ is uniformly integrable,

$$\lim_{n\to\infty} E\Big(\prod_{q=1}^{r}\int_0^t Z_v(u)\, W_{i_q}(n,du) \mid Z_v\Big)$$
$$= \lim_{n\to\infty}\Big[E\Big(\int_0^t Z_v(u)\, W_i(n,du) \mid Z_v\Big)\Big]^r$$
$$= \Big(\lambda(k+\theta)\int_0^t Z_v(u)\, du\Big)^r, \tag{82}$$

upon recalling (69). Substituting (82) into (81) yields

$$\lim_{n\to\infty} n^r g(n,r) = \Big(\gamma\int_0^t Z_v(u)du\Big)^r. \tag{83}$$

Next, similar to (66) and (80),

$$n^r E\Big(\Big|\prod_{q=1}^{r}\xi_{n,i_q}(t) - \prod_{q=1}^{r}\sum_{z=1}^{n_0} I^v_{i_q,z}1_{\delta_{i_q,z}\le t}\Big| \mid Z_v\Big)$$
$$\le 2n^r E\Big(\prod_{q=1}^{r}\xi_{n,i_q}(t)1_{W_{i_q}(n,t)>\sqrt{n}/r-1} \mid Z_v\Big) \to 0. \tag{84}$$

Combining (83)-(84) establishes

$$\lim_{n\to\infty} n^r E\Big(\prod_{q=1}^{r}\xi_{n,i_q}(t) \mid Z_v\Big) = \Big(\gamma\int_0^t Z_v(u)du\Big)^r. \tag{85}$$

Our remaining task is to show that

$$h(n,r) := n^r E\Big(\prod_{q=1}^{r}\xi_{n,i_q}(t) - \prod_{q=1}^{r} 1_{\xi_{n,i_q}(t)\ge 1} \mid Z_v\Big) \tag{86}$$

converges to 0 as $n\to\infty$. Observe that for each i,

$$E\left(\xi_{n,i}(t) - 1_{\xi_{n,i}(t)\ge 1}\right) \le E\left[(\xi_{n,i}(t))^2 - \xi_{n,i}(t)\right]$$
$$= E\Big(2\sum_{z<z'} I^v_{i,z}I^v_{i,z'}1_{\delta_{i,z}\le t}1_{\delta_{i,z'}\le t}\Big). \tag{87}$$

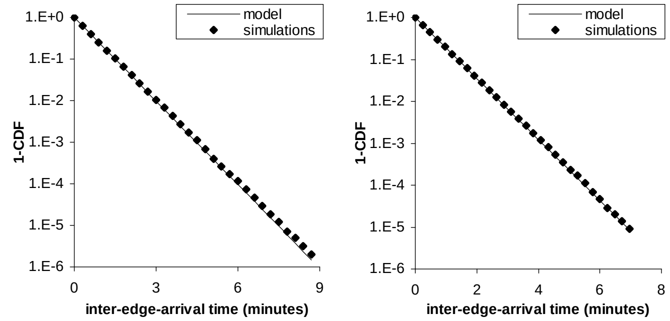
Using (87), (86) can be transformed into

$$h(n,r) \le n^r E\Big[\sum_{p=1}^{r}\Big((\xi_{n,i_p}(t))^2 - \xi_{n,i_p}(t)\Big)\prod_{\substack{q=1\\ q\ne p}}^{r}\xi_{n,i_q}(t) \mid Z_v\Big]$$
$$= rn^r E\Big[\Big((\xi_{n,i_1}(t))^2 - \xi_{n,i_1}(t)\Big)\prod_{q=2}^{r}\xi_{n,i_q}(t) \mid Z_v\Big]$$
$$= 2rn^r E\Big[\sum_{z_1<z_1'\le W_{i_1}(n,t)} I^v_{i_1,z_1}I^v_{i_1,z_1'}\prod_{q=2}^{r}\xi_{n,i_q}(t) \mid Z_v\Big].$$

As in (77), applying the upper bound to the above yields

$$2rn^r E\Big[\sum_{z_1<z_1'<\sqrt{n}/r-1} I^v_{i_1,z_1}I^v_{i_1,z_1'}\prod_{q=2}^{r}\xi_{n,i_q}(t) \mid Z_v\Big]$$
$$\le \frac{2rn^r\overline{b}_n^r}{\sqrt{n}/r-1}E\Big(\prod_{q=1}^{r}\xi_{n,i_q}(t) \mid Z_v\Big) \to 0, \tag{88}$$

where the last step is obtained recalling (85). It is then easy to reach that $h(n,r)\to 0$ as $n\to\infty$. Combining this result



(a) system $\mathcal{H}$ with $n = 1000$ (b) system $\mathcal{VH}$ with $n = 5000$

Fig. 6. Distribution of edge inter-arrival delays approaches exponential with rate γ in (65) for $n = 1000$, $k = 10$, and $\theta = 10$ using 10^9 iterations.

with (85), (72) reduces to

$$\lim_{n\to\infty} f(n,r) = \lim_{n\to\infty}\binom{n-1}{r}E\Big(\prod_{q=1}^{r}1_{\xi_{n,i_q}(t)\ge 1} \mid Z_v\Big)$$
$$= \lim_{n\to\infty}\frac{n^r}{r!}E\Big(\prod_{q=1}^{r}\xi_{n,i_q}(t) \mid Z_v\Big) = \frac{1}{r!}\Big(\gamma\int_0^t Z_v(u)\, du\Big)^r,$$

demonstrating that (74) holds, which completes the proof. ■

Theorem 3 indicates that when user v is alive (i.e., $Z_v(\cdot) = 1$), the instantaneous rate of edge-arrival to v is the constant γ; otherwise it is 0. Therefore, the edge arrival process to v is a Poisson process whose rate varies according to Z_v, from which the next corollary follows immediately.

Corollary 4: With Assumptions 2-3 and uniform selection, unconditioned on Z_v, the edge-arrival process ξ_n to user v converges in distribution to a Cox process directed by the random measure $\gamma\int_0^t Z_v(u)du$. Its intensity measure is

$$E\left(\xi_n(t)|v\text{'s type}\right) \to \gamma E\Big(\int_0^t Z_v(u)du\Big|v\text{'s type}\Big) = \gamma a_v t.$$

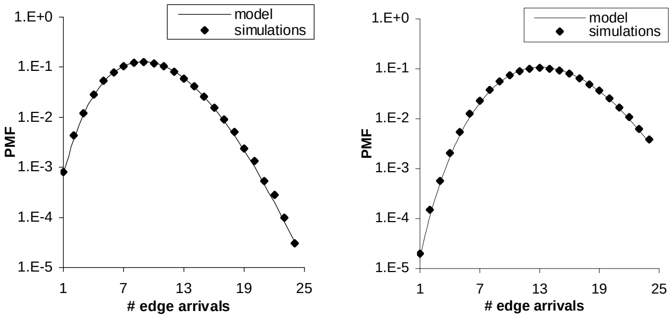
where a_v is the availability of v. Unconditional on user types,

$$E(\xi_n(t)) \to \gamma a t = \lambda(k+\theta)t. \tag{89}$$

where the parameters are the same as in (65).

Comparing (54) to (89), observe that the mean number of edges created by a user in $[0,t]$ is equal to $E[\xi_n(t)]$ delivered to a user, which is consistent with the symmetric nature of edge creation/arrival and uniform selection. Furthermore, Theorem 3 and Corollary 4 show that despite multiple user-types and non-Poisson user-arrival dynamics, the edge-arrival process to each user v is Poisson (modulated by Z_v). This allows us to obtain easy analytical results on edge arrivals and other related metrics (e.g. in-degree and joint in/out-degree).

Fig. 6 shows the distribution of edge inter-arrival delays to a single node obtained in simulations with two types of systems, given that the node is alive. Notice in the sub-figures that for finite n, the distribution of inter-arrival delay is approximately exponential with the rate given by (65). Additionally, Fig. 7 shows that the distribution of the number of edge arrivals to a node in an interval of size Δt approaches a Poisson distribution with the same rate γ in (65).



(a) system $\mathcal{H}$ with $\Delta t = 6$ min (b) system $\mathcal{VH}$ with $\Delta t = 9$ min

Fig. 7. Distribution of the number of edge arrivals to a node in the interval $[t, t+\Delta t]$ in a system with $n = 1000$ users, $k = 10$, and $\theta = 10$. The lines show Poisson fits with γ in (65) at $t = 500$ hours and after 10^5 iterations.

Finally, note that the Poisson result in Theorem 3 is not an *assumption* of the paper as in prior work [11], [18], [21], but rather a *consequence* of the churn model introduced earlier.

VI. In-Degree

We now focus on understanding how the in-degree of each *live* user changes with time. For the rest of the paper, we assume $n \to \infty$ and the edge arrival process to each user is Poisson with rate γ in (65) given that it appears in the system.

A. Expected In-Degree

In a stationary system, define $X_n(t)$ to be the in-degree of a random online user at age $t \geq 0$. In this section, we focus on transient and limiting properties of $X_n(t)$ under uniform selection of neighbors. We have the following result.

Theorem 4: Let $\{U(s)\}_{s\geq 0}$ be a pure renewal process with cycle length $R \sim H(x)$. Given that a user is alive in the system, its expected in-degree at fixed age $t \geq 0$ converges as $n \to \infty$ to a monotonically increasing function of age

$$E[X_n(t)] \to k \int_0^\infty \left(E[U(x) - U(x-t)]\right) dH(x), \tag{90}$$

where $E[U(x)] = \sum_{r=0}^{\infty} H^{*r}(x)$ for $x \geq 0$, $E[U(x)] = 0$ for $x < 0$, and $H(x)$ is in (29).

Proof: Assume that user v begins an ON period at time 0 (setting v's arrival time $\tau_{v,m} = 0$ and imaging the system started in the past). Suppose that v remains alive in $[0, t]$, where $t > 0$ is v's fixed age. Denote by $A_i(t)$ the age of any other alive user i, i.e., the duration from i's latest arrival time to the epoch t. Conditioned on v being ON in $[0, t]$, the age of user i at time t follows the equilibrium distribution: $P(A_i(t) \leq x | Z_i(t) = 1) = H_i(x)$, where $H_i(x)$ is in (11).

Denote by $\delta^c_{i,r}$, $r \geq 1$, the random times at which user i makes selections for link c within i's *current* ON duration, where $\delta^c_{i,1}$ is the i's latest arrival time (note that $A_i(t) = t - \delta^c_{i,1}$). Let $U(s)$ count the number of selections made by i through link c in interval $[\delta^c_{i,1}, \delta^c_{i,1} + s]$ and set $U(s) = 0$ for $s < 0$. Now, define $J^c_i(t)$ to be the indicator of i connecting v through link c by time t. Considering all selections of i by its age $A_i(t)$, we have

$$\begin{aligned} P(J^c_i(t) = 1) &= P(Z_i(t) = 1) P(i \text{ selects } v | Z_i(t) = 1) \\ &= a_i \sum_{r=1}^{\infty} P\left[U(A_i(t) - t) < r \leq U(A_i(t))\right] E\Big[\frac{1}{N^i_n(\delta^c_{i,r})}\Big], \end{aligned}$$

where $N^i_n(\delta^c_{i,r}) \geq 1$ conditioned on v being alive. Applying the similar method in (56) to $E[1/N^i_n(\delta^c_{i,r})]$, the above yields

$$\begin{aligned} &\lim_{n\to\infty} P(J^c_i(t) = 1) \\ &\quad = a_i \lim_{n\to\infty} \frac{1}{na} \sum_{r=1}^{\infty} P\left[U(A_i(t) - t) < r \leq U(A_i(t))\right] \\ &\quad = a_i \lim_{n\to\infty} \frac{1}{na} E[U(A_i(t)) - U(A_i(t) - t)], \end{aligned} \tag{91}$$

where $a = \sum_{j=1}^{\mathcal{T}} p_j a^{(j)}$ and $E[U(s)]$ is in (60) for $n \to \infty$.

The expected number of edges received by user v in $[0, t]$ from the entire system is then given by

$$\begin{aligned} \lim_{n\to\infty} E[X_n(t)] &= \lim_{n\to\infty} kE\bigg(\sum_{i=1, i\neq v} J^c_i(t)\bigg). \\ &= k \lim_{n\to\infty} \frac{1}{na} \sum_{i=1, i\neq v} a_i E[U(A_i(t)) - U(A_i(t) - t)], \end{aligned}$$

which leads to (90) by using $A_i(t) \sim H_i(x)$ and (29)-(30). ■

Model (90) can be written as

$$\lim_{n\to\infty} E[X_n(t)] = kE[U(R) - U(R - t)], \tag{92}$$

where $R \sim H(x)$ denotes an existing user i's age, $U(R)$ counts the number of edges created by i for link c by its age R (since i's latest arrival), and $U(R - t)$ corresponds the selections of i before v appeared in the system. Furthermore, as $t \to \infty$, (92) tends to $kE[U(R)]$, which provides a simple upper-bound at which the in-degree of each user saturates.

We next show that (92) can be expressed in simple closed-form for exponential lifetimes.

Theorem 5: For exponential lifetimes L and $n \to \infty$, the mean in-degree at failure $\theta = k$ and

$$E[X_n(t)] \to 2k(1 - e^{-t/E[L]}). \tag{93}$$

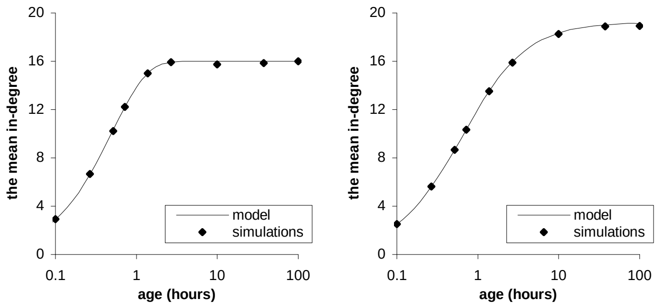
Furthermore, the distribution of $X_n(t)$ converges to a Poisson distribution with mean $2k(1 - e^{-t/E[L]})$ as $n \to \infty$.

Proof: Since $F(x)$ is exponential, $H(x) = 1 - e^{-x/E[L]}$. Then, the pure renewal process $\{U(s)\}$ with cycle length $R \sim H(x)$ is a Poisson process with a point at time 0. This leads to $E[U(x)] = 1 + x/E[L]$ for $x \geq 0$. We thus have

$$\begin{aligned} E[X_n(t)] &\to k\Big(\int_0^t \big(1 + \frac{x}{E[L]}\big)\, dH(x) + \int_t^\infty \frac{t}{E[L]}\, dH(x)\Big) \\ &= k\Big(H(t) + \frac{E[\min(L, t)]}{E[L]}\Big) \\ &= 2k(1 - e^{-t/E[L]}). \end{aligned} \tag{94}$$

The mean number of replacement edges per lifetime is $\theta = k\int_0^\infty x/E[L] dF(x) = k$, which is equal to the in-degree at failure:

$$\lim_{n\to\infty} \int_0^\infty E[X_n(t)] dF(t) = 2k \int_0^\infty P(L < x) dF(x) = k.$$



(a) exponential (93) with $\mu = 2$ (b) Pareto (90) with $\alpha = 3$

Fig. 8. Comparison of the model for $E[X_n(t)]$ to simulation results for $n = 2000$, $E[L] = 0.5$ hours, and $k = 8$ after 10^6 iterations.

Finally, for exponential L, the in-degree model reduces to an $M/M/\infty$ queuing model, where the arrival process is Poisson with rate γ shown in Theorem 3 and the duration in which an in-degree neighbor remains alive (i.e., service time in a queuing model) is exponential with rate $1/E[L]$ due to the memoryless properties of exponential distributions. The number of in-degree at time t (i.e., the queue length at t) thus follows a Poisson distribution [23, pp. 351]. ■

In (93), the mean in-degree of a node increases monotonically from $X_n(0) = 0$ when it arrives into the system to $E[X_n(\infty)] = 2k$ when its age tends to infinity. For the exponential case we directly use (93), while for the Pareto lifetime case with $\alpha > 2$ we numerically solve (92). Simulation results in Fig. 8 demonstrate that the models are very accurate and indeed saturate at predicted values $2k$ and $kE[U(R)]$ as age $t \to \infty$. Furthermore, if a node survives for more than 1 hour in the system, it develops an average of $12 - 15$ in-degree neighbors (depending on the distribution of L) and is unlikely to be isolated from the graph from that point on. It is also interesting to observe in the figure that the Pareto curve increases slower, but saturates at larger values, which suggests more resilience support for users with very large lifetimes. The saturation effect illustrated in Fig. 8 also shows that P2P implementations should cap user in-degree at values no smaller than the limit of (90) for $t \to \infty$. The corresponding upper bound in Gnutella (i.e., 30 in-degree neighbors) satisfies this condition for the two examples shown above.

VII. Joint In/Out-Degree Model

Analytical results in the previous section show that the early stage in a node's life in the network is actually risky from the isolation point of view as it must rely solely on its out-degree neighbors. However, once a node survives this early stage, it increases its resilience to isolation through constantly arriving incoming edges. In this section, we combine the in-degree and out-degree models to derive the *joint* isolation probability of an arriving user. We drop subscript n and assume $n \to \infty$.

A. Preliminaries

Denote by $X^*(t)$ the out-degree of a node v at given age t and define it to be *isolated* when its in-degree $X(t)$ and out-degree are simultaneously zero. Define *time to isolation* T to be the first-hitting time of both processes to state 0:

$$T = \inf\{t > 0 : X^*(t) = X(t) = 0 | X^*(0) = k, X(0) = 0\}.$$

Then the probability of node isolation is simply $\phi = P(T < L)$, where L is the random lifetime of node v. Unlike in the out-degree process, a node does not replace its incoming edges, which means that the in-degree and out-degree processes are independent of each other.

In the next subsections, we derive ϕ for systems with exponential user lifetimes and exponential search delays using two methods. The first approach provides an exact model using matrix algebra, while the second one shows an asymptotically accurate approximation that is available in simple closed-form.

B. Exponential Lifetimes (Exact Model)

Let pair $(X^*(t), X(t))$ be the joint process of out-degree and in-degree of a node at age t and (i, j) denote any admissible state of the joint process for $0 \le i \le k$ and $0 \le j < n$. Recall that edge arrival at any live node occurs according to a Poisson process with rate (65). Therefore, under uniform selection, incoming neighbors arrive to v at rate:

$$\gamma = \frac{k + \theta}{E[L]} = \frac{2k}{E[L]} \tag{95}$$

since $\theta = k$ for exponential L. The current in/out-degree neighbors of v fail at rate $\mu = 1/E[L]$ due to the memoryless property of exponentials. This directly leads to the next result.

Lemma 9: Given $L \sim \exp(\mu)$ and search times $S \sim \exp(\sigma)$ for finding replacement neighbors, the joint process $\{(X^*(t), X(t))\}$ is a homogeneous continuous-time Markov chain with a transition rate matrix $Q = (q_{uu'})$:

$$q_{uu'} = \begin{cases} i\mu & (i,j) \to (i-1, j) \\ (k-i)\sigma & (i,j) \to (i+1, j), \text{ for } i < k \\ j\mu & (i,j) \to (i, j-1) \\ 2k\mu & (i,j) \to (i, j+1) \\ -\Lambda_{ij} & (i,j) \to (i,j) \\ 0 & \text{otherwise} \end{cases}, \tag{96}$$

where u and u' represent any suitable states of the joint chain satisfying transition requirements on the right side of (96) and $\Lambda_{ij} = i\mu + (k - i)\sigma + j\mu + 2k\mu$.

It is convenient to treat $\{(X^*(t), X(t))\}$ as an absorbing Markov chain in order to derive the PDF of the first-hitting time T on state $(0, 0)$. Assuming $(0, 0)$ is an absorbing state, we can write Q in canonical form as:

$$Q = \begin{pmatrix} 0 & 0 \\ \mathbf{r} & Q_0 \end{pmatrix}, \tag{97}$$

where Q_0 is the rate matrix obtained by removing the rows and columns corresponding to state $(0, 0)$ from Q and $\mathbf{r}$ is a column vector of transition rates into state $(0, 0)$.

Applying [32, Theorem 2], we obtain the next result.

Lemma 10: For exponential lifetimes $L \sim \exp(\mu)$ and exponential search delays $S \sim \exp(\sigma)$, the probability of node isolation is given by:

$$\phi = \pi(0) V B V^{-1} \mathbf{r}, \tag{98}$$

TABLE I
EXACT MODEL (98) AND SIMULATIONS ($n = 2000$, $E[L] = 0.5$ HOURS)

$E[S]$	$k=6$		$k=8$	
min	Simulations	Model (98)	Simulations	Model (98)
6	3.63×10^{-6}	3.61×10^{-6}	2.80×10^{-8}	2.87×10^{-8}
18	3.15×10^{-5}	3.17×10^{-5}	5.91×10^{-7}	5.98×10^{-7}
30	6.04×10^{-5}	6.08×10^{-5}	1.48×10^{-6}	1.46×10^{-6}
42	8.38×10^{-5}	8.37×10^{-5}	2.30×10^{-6}	2.27×10^{-6}
60	1.06×10^{-4}	1.09×10^{-4}	3.27×10^{-6}	3.28×10^{-6}

where V is a matrix of eigenvectors of Q_0, $B = \text{diag}(b_j)$ is a diagonal matrix with $b_j = 1/(\mu - \xi_j)$, ξ_j is the j-th eigenvalue of Q_0, and $\pi(0) = (\pi_{(i,j)}(0))$ is the initial state distribution with $\pi_{(k,0)}(0) = 1$ and $\pi_{(i,j)}(0) = 0$ in all other pairs.

We verify (98) in simulations shown in Table I, which shows that our results are indeed very accurate. While (98) provides values ϕ that are smaller than isolation probability ϕ_{out} of the out-degree model [14] by several orders of magnitude, it is still unclear what impact in-degree has on the probability that a user gets isolated as its age increases and how large the improvement ratio ϕ_{out}/ϕ is. We study these issues below.

C. Isolation with Increased Age

To better understand the impact of in-degree on ϕ, let us define the first hitting time T_{out} on state 0 of the out-degree process $\{X^*(t)\}$, i.e., $T_{out} = \inf\{t > 0 : X^*(t) = 0 | X^*(0) = k\}$. Analysis in [15] shows that $\{X^*(t)\}$ is a birth-death Markov chain and derives its CDF function $P(T_{out} < t)$ in matrix form. We use this result and the CDF of T derived in the proof of Theorem 10 to compare the distribution of isolation times in the joint in/out degree model with that studied in [15]. We plot the exact distributions of both T_{out} and T as functions of user age in Fig. 9. Notice in the figure that $P(T_{out} < t)$ increases almost linearly in time t indicating that users with large lifetimes have proportionally higher probabilities of isolation. In contrast, the curve of $P(T < t)$ becomes almost flat as time t increases beyond 0.5 hours showing that users with lifetimes in the range $[0.5, 200]$ hours exhibit almost the same isolation probabilities. In fact, once the initial $1/2$-hour period is over, isolation probability is orders of magnitude smaller than in the initial phase. As user age increases above 200 hours, the CDF of T slowly increases in time since $X(t)$ becomes saturated and can no longer keep up with the increased possibility of neighbor failure.

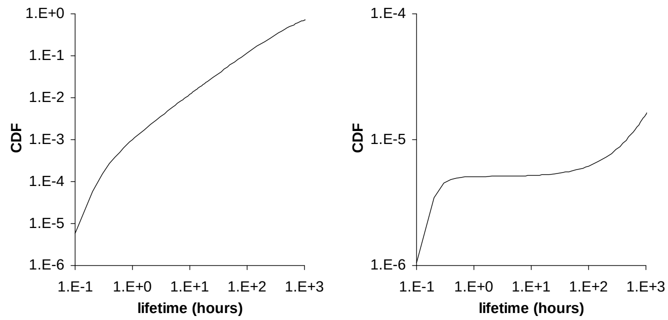
D. Exponential Lifetimes (Asymptotic Model)

Although (98) provides exact results for ϕ, it relies on numerical matrix algebra. Our next task is to obtain a simple closed-form solution to ϕ when the mean search delay $E[S] \to 0$.

Theorem 6: For $L \sim \exp(\mu)$ and $S \sim \exp(\sigma)$, isolation probability converges to the following as $E[S] \to 0$:

$$\phi = \frac{1 - e^{-2k}}{2k} \phi_{out}, \tag{99}$$

where $\phi_{out} = \rho k/(1+\rho)^k$ and $\rho = \sigma/\mu = E[L]/E[S]$.



(a) T_{out} (b) T

Fig. 9. The CDF of T_{out} and T for exponential lifetimes with $E[L] = 0.5$ hours, exponential search delays with $E[S] = 0.1$ hours, and $k = 6$.

TABLE II
CONVERGENCE OF (99) TO (98) FOR $E[L] = 0.5$ HOURS AND $k = 6$

$E[S]$	Exact model (98)	Approx. model (99)	Relative error
36 sec	8.721×10^{-10}	1.421×10^{-9}	62.91%
3.6 sec	1.498×10^{-14}	1.581×10^{-14}	5.57%
360 ms	1.589×10^{-19}	1.598×10^{-19}	0.55%
36 ms	1.600×10^{-24}	1.600×10^{-24}	0

Proof: See Appendix A. ■

It can be seen from (99) that by considering both in-degree and out-degree, the probability of node isolation is reduced by a factor of approximately $2k$ for non-trivial k. The reason for this relatively small improvement is that only a handful of users benefit from the in-degree in their isolation resilience since the majority of users depart very quickly and are unable to accumulate any in-degree neighbors. Nevertheless, analysis of this section has important consequences as it shows that *the most reliable users of the system (i.e., those with large lifetimes) extract huge benefits from the in-degree process and are thus allowed to continue providing services to others with much higher probability than possible with just the out-degree*.

To complete this section, Table II shows the relative approximation error of (99) and confirms its asymptotic accuracy. For large S, our numerical results from the exact model suggest that (99) provides an upper bound on the isolation probability, where ϕ_{out}/ϕ is 3-10 times larger than the $2k$ suggested by (99). For instance, for fixed $E[L] = 0.5$ hours and $k = 6$, ratio ϕ_{out}/ϕ is 39 when $E[S] = 2$ minutes and 120 when $E[S] = 6$ minutes.

VIII. CONCLUSION

This paper introduced a simple model of churn and developed numerous closed-form results describing the behavior of users including their joint and residual lifetime distributions, evolution of system size, transient in-degree, and isolation probability under the joint in/out degree model.

Future work involves modeling of non-uniform neighbor selection, measurement of real P2P dynamics, and analysis of asymptotically small networks.

REFERENCES

[1] D. Aldous and M. Brown, "Inequalities for Rare Events in Time-Reversible Markov Chains I," in *Stochastic Inequalities*, M. Shaked and

Y. L. Tong, Eds. Hayward, CA: Institute of Mathematical Statistics, 1992, vol. 22, pp. 1–16.

[2] J. Aspnes, Z. Diamadi, and G. Shah, “Fault-Tolerant Routing in Peer-to-Peer Systems,” in *Proc. ACM PODC*, Jul. 2002, pp. 223–232.

[3] R. Bhagwan, S. Savage, and G. M. Voelker, “Understanding Availability,” in *Proc. IPTPS*, Feb. 2003, pp. 256–267.

[4] A. A. Borovkov, *Probability Theory*. New York, NY: Gordon and Breach, 1998.

[5] F. E. Bustamante and Y. Qiao, “Friendships that Last: Peer Lifespan and its Role in P2P Protocols,” in *Proc. Intl. Workshop on Web Content Caching and Distribution*, Sep. 2003.

[6] B.-G. Chun, B. Zhao, and J. Kubiatowicz, “Impact of Neighbor Selection on Performance and Resilience of Structured P2P Networks,” in *Proc. IPTPS*, Feb. 2005, pp. 264–274.

[7] A. Feldmann and W. Whitt, “Fitting Mixtures of Exponentials to Long-tailed Distributions to Analyze Network Performance Models,” *Performance Evaluation*, vol. 31, no. 3-4, pp. 245–279, Jan. 1998.

[8] K. Gummadi, R. Gummadi, S. Gribble, S. Ratnasamy, S. Shenker, and I. Stoica, “The Impact of DHT Routing Geometry on Resilience and Proximity,” in *Proc. ACM SIGCOMM*, Aug. 2003, pp. 381–394.

[9] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, “Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload,” in *Proc. ACM SOSP*, Oct. 2003, pp. 314–329.

[10] M. F. Kaashoek and D. Karger, “Koorde: A Simple Degree-Optimal Distributed Hash Table,” in *Proc. IPTPS*, Feb. 2003, pp. 98–107.

[11] S. Krishnamurthy, S. El-Ansary, E. Aurell, and S. Haridi, “A Statistical Theory of Chord under Churn,” in *Proc. IPTPS*, Feb. 2005, pp. 93–103.

[12] S. S. Lam and H. Liu, “Failure Recovery for Structured P2P Networks: Protocol Design and Performance Evaluation,” in *Proc. ACM SIGMETRICS*, Jun. 2004, pp. 199–210.

[13] D. Leonard, V. Rai, and D. Loguinov, “On Lifetime-Based Node Failure and Stochastic Resilience of Decentralized Peer-to-Peer Networks,” in *Proc. ACM SIGMETRICS*, Jun. 2005, pp. 26–37.

[14] D. Leonard, Z. Yao, V. Rai, and D. Loguinov, “On Lifetime-Based Node Failure and Stochastic Resilience of Decentralized Peer-to-Peer Networks,” *IEEE/ACM Trans. Networking*, vol. 15, no. 3, pp. 644–656, Jun. 2007.

[15] D. Leonard, Z. Yao, X. Wang, and D. Loguinov, “On Static and Dynamic Partitioning Behavior of Large-Scale P2P Networks,” *IEEE/ACM Trans. Networking*, vol. 16, no. 6, pp. 1475–1488, Dec. 2008.

[16] J. Li, J. Stribling, T. M. Gil, R. Morris, and M. F. Kaashoek, “Comparing the Performance of Distributed Hash Tables under Churn,” in *Proc. IPTPS*, Feb. 2004, pp. 87–99.

[17] J. Li, J. Stribling, R. Morris, and M. F. Kaashoek, “Bandwidth-Efficient Management of DHT Routing Tables,” in *Proc. USENIX NSDI*, May 2005, pp. 1–11.

[18] D. Liben-Nowell, H. Balakrishnan, and D. Karger, “Analysis of the Evolution of the Peer-to-Peer Systems,” in *Proc. ACM PODC*, Jul. 2002, pp. 233–242.

[19] D. Loguinov, J. Casas, and X. Wang, “Graph-Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault Resilience,” *IEEE/ACM Trans. Networking*, vol. 13, no. 5, pp. 1107–1120, Oct. 2005.

[20] G. S. Manku, M. Naor, and U. Weider, “Know thy Neighbor’s Neighbor: the Power of Lookahead in Randomized P2P Networks,” in *Proc. ACM STOC*, Jun. 2004, pp. 54–63.

[21] G. Pandurangan, P. Raghavan, and E. Upfal, “Building Low-Diameter Peer-to-Peer Networks,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 6, pp. 995–1002, Aug. 2003.

[22] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*. Boston, NY: Springer-Verlag, 1987.

[23] S. I. Resnick, *Adventures in Stochastic Processes*. Boston, MA: Birkhäuser, 2002.

[24] S. I. Resnick, *A Probability Path*. Boston, MA: Birkhäuser, 2003.

[25] S. Rhea, D. Geels, T. Roscoe, and J. Kubiatowicz, “Handling Churn in a DHT,” in *Proc. USENIX Ann. Tech. Conf.*, Jun. 2004, pp. 127–140.

[26] S. Saroiu, P. K. Gummadi, and S. D. Gribble, “A Measurement Study of Peer-to-Peer File Sharing Systems,” in *Proc. SPIE/ACM Multimedia Computing and Networking*, vol. 4673, Jan. 2002, pp. 156–170.

[27] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications,” in *Proc. ACM SIGCOMM*, Aug. 2001, pp. 149–160.

[28] D. Stutzbach and R. Rejaie, “Understanding Churn in Peer-to-Peer Networks,” in *Proc. ACM IMC*, Oct. 2006, pp. 189–202.

[29] X. Wang, Z. Yao, and D. Loguinov, “Residual-Based Measurement of Peer and Link Lifetimes in Gnutella Networks,” in *Proc. IEEE INFOCOM*, May 2007, pp. 391–399.

[30] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, NJ: Prentice Hall, 1989.

[31] Z. Yao and D. Loguinov, “Link Lifetimes and Randomized Neighbor Selection in DHTs,” in *Proc. IEEE INFOCOM*, Apr. 2008.

[32] Z. Yao, X. Wang, D. Leonard, and D. Loguinov, “On Node Isolation under Churn in Unstructured P2P Networks with Heavy-Tailed Lifetimes,” in *Proc. IEEE INFOCOM*, May 2007, pp. 2126–2134.

[33] M. Zhong, K. Shen, and J. Seiferas, “Non-Uniform Random Membership Management in Peer-to-Peer Networks,” in *Proc. IEEE INFOCOM*, Mar. 2005, pp. 1151–1161.

APPENDIX A
PROOF OF THEOREM 6

Proof: We begin with obtaining the asymptotic distribution of the first-hitting time T_{out} onto state 0 of the out-degree process $\{X^*(t)\}$, followed by the derivation of the asymptotic distribution of T onto state $(0,0)$ of the join in/out-degree process, and finally the proof for (99).

To this end, using previous results in [1], we know that for Markov chain $\{X^*(t)\}$, the first hitting time of a rare event (e.g., state 0 of $\{X^*(t)\}$) behaves as an exponential random variable with rate $1/E[T_{out}]$:

$$P(T_{out} < t) = 1 - e^{-t/E[T_{out}]}, \text{ as } E[S] \to 0, \qquad (100)$$

where $E[T_{out}]$ is available in closed form [13]:

$$E[T_{out}] = \frac{E[S]}{k} (1+\rho)^k, \qquad (101)$$

for small mean search delay $E[S]$, where $\rho = E[L]/E[S]$. Observe that $E[T_{out}] \to \infty$ as $E[S] \to 0$. Thus by Taylor expansion (100) reduces to

$$P(T_{out} < t) = t/E[T_{out}], \text{ as } E[S] \to 0, \qquad (102)$$

showing that asymptotically T_{out} behaves like a uniform random variable. Taking the derivative of (102), we obtain the asymptotic PDF of T_{out}:

$$f_{T_{out}}(t) = 1/E[T_{out}], \text{ as } E[S] \to 0. \qquad (103)$$

It is then straightforward to obtain

$$\begin{aligned}\phi_{out} = P(T_{out} < L) &= \int_0^\infty P(L > t) f_{T_{out}}(t)dt \\ &= \frac{E[L]}{E[T_{out}]}, \quad E[S] \to 0. \qquad (104)\end{aligned}$$

Next, observe that user lifetime L is small compared to the value of the first hitting time T on state $(0,0)$. Therefore, $P(T < L)$ is mainly affected by the CDF $P(T < x)$ only for small x. Further, note that the probability that out-degree process $\{X^*(t)\}$ hits more than once on state 0 within interval $[0,x]$ for small x is negligible when $E[S] \to 0$ (i.e., $E[T_{out}] \to \infty$). Thus, based on the property of the first hitting time T_{out} and the probability that the in-degree is zero at epoch T_{out}, we obtain a simple formula for the asymptotic CDF of T:

$$P(T < x) = \int_0^x P(X(t) = 0) f_{T_{out}}(t)dt, \qquad (105)$$

as $E[S] \to 0$. By Theorem 5, we obtain the probability that node in-degree is zero at time t for exponential lifetimes:

$$P(X(t) = 0) = e^{-2k(1-e^{-\mu t})}. \qquad (106)$$

Substituting (103) and (106) into (105) yields

$$\begin{aligned} P(T<x) &= \frac{1}{E[T_{out}]}\int_0^x e^{-2k(1-e^{-\mu t})}dt \\ &= \frac{e^{-2k}}{\mu E[T_{out}]}\int_{-2ke^{-\mu x}}^{-2k}\frac{e^{-z}}{z}dz, \end{aligned} \tag{107}$$

for $E[S]\to 0$. Notice that

$$\int_{-2ke^{-\mu x}}^{-2k}\frac{e^{-z}}{z}dz = \int_{-2ke^{-\mu x}}^{\infty}\frac{e^{-z}}{z}dz - \int_{-2k}^{\infty}\frac{e^{-z}}{z}dz. \tag{108}$$

Substituting (108) into (107) and using $\mu = 1/E[L]$ and (104), we easily establish

$$P(T<x) = e^{-2k}(\mathrm{Ei}(2k) - \mathrm{Ei}(2ke^{-\mu x}))\phi_{out}, \tag{109}$$

where ϕ_{out} is given by (104) and $\mathrm{Ei}(x) = -\int_{-x}^{\infty} e^{-z}/z dz$ is the exponential integral.

Finally, integrating (109) using the PDF $f(x) = \mu e^{-\mu x}$ of user lifetimes, we have

$$\begin{aligned} \phi &= \int_0^{\infty} P(T<x)f(x)dx \\ &= e^{-2k}\Big(\mathrm{Ei}(2k) - \int_0^{\infty}\mathrm{Ei}(2ke^{-\mu x}))f(x)dx\Big)\phi_{out} \\ &= e^{-2k}\Big(\mathrm{Ei}(2k) - \frac{1}{2k}\int_0^{2k}\mathrm{Ei}(x)dx\Big)\phi_{out}. \end{aligned} \tag{110}$$

Observe that

$$\int_0^{2k}\mathrm{Ei}(x)dx = 1 - e^{2k} + 2k\mathrm{Ei}(2k). \tag{111}$$

Substituting (111) into (110), we easily obtain (99). ■