University of Dayton
eCommons

Computer Science Faculty Publications | Department of Computer Science

3-2012

On Superposition of Heterogeneous Edge Processes in Dynamic Random Graphs

Zhongmei Yao
University of Dayton, zyao01@udayton.edu

Daren B. H. Cline
Texas A & M University - College Station

Dmitri Loguinov
Texas A & M University - College Station

Follow this and additional works at: http://ecommons.udayton.edu/cps_fac_pub

 Part of the Databases and Information Systems Commons, Numerical Analysis and Scientific Computing Commons, OS and Networks Commons, Other Computer Sciences Commons, and the Theory and Algorithms Commons

eCommons Citation

Yao, Zhongmei; Cline, Daren B. H.; and Loguinov, Dmitri, "On Superposition of Heterogeneous Edge Processes in Dynamic Random Graphs" (2012). *Computer Science Faculty Publications*. Paper 4.
http://ecommons.udayton.edu/cps_fac_pub/4

This Conference Paper is brought to you for free and open access by the Department of Computer Science at eCommons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

On Superposition of Heterogeneous Edge Processes in Dynamic Random Graphs

Zhongmei Yao
University of Dayton
Dayton, OH 45469
Email: zyao@udayton.edu

Daren B.H. Cline
Texas A&M University
College Station, TX 77843
Email: dcline@stat.tamu.edu

Dmitri Loguinov*
Texas A&M University
College Station, TX 77843
Email: dmitri@cs.tamu.edu

***Abstract*—This paper builds a generic modeling framework for analyzing the edge-creation process in dynamic random graphs in which nodes continuously alternate between active and inactive states, which represent churn behavior of modern distributed systems. We prove that despite heterogeneity of node lifetimes, different initial out-degree, non-Poisson arrival/failure dynamics, and complex spatial and temporal dependency among creation of both initial and replacement edges, a superposition of edge-arrival processes to a live node under uniform selection converges to a Poisson process when system size becomes sufficiently large. Due to the convoluted dependency and non-renewal nature of various point processes, this result significantly advances classical Poisson convergence analysis and offers a simple analytical platform for future modeling of networks under churn in a wide range of degree-regular and -irregular graphs with arbitrary node lifetime distributions.**

I. Introduction

Many contemporary distributed systems (e.g., P2P [7], [23] and wireless ad-hoc networks [25], [26]) and large data-centric applications (e.g., Google MapReduce [5], Amazon EC^2 [6]) can be modeled as decentralized graphs in which nodes rely on communication services and computational power of other hosts in the system. An important characteristic of these networks is their robustness to node churn, which includes topology randomization, operation under failure, and dynamic link rewiring to accommodate arriving/departing entities.

Previous research has analyzed multiple avenues for understanding and improving such networks, which includes ensuring connectivity [11], balancing load [27], reducing graph diameter [17], improving resilience [14], [31], mitigating channel interference [1], understanding routing mobility [25] and flooding [26], and optimizing capacity [4], [9], [19]. While relying on a separate model for each study is acceptable in certain cases, the field of dynamic distributed systems has reached sufficient maturity that calls for a unifying foundation that can explain the limiting behavior of the aggregate edge process of the system and pave the way for rigorous analysis of diverse application-specific metrics.

This paper considers a system of n nodes, each of which alternates between dead/alive (i.e., ON/OFF) states that represent the online/offline behavior of the corresponding user or host. Upon joining the network, each node i creates a routing table consisting of $k_i \geq 1$ initial out-degree neighbors selected from the pool of available participants. Upon failure of neighbors, each node replaces them with other live users in order to keep its routing table up-to-date, where all selections are made uniformly randomly.

The first half of this paper focuses on the edge-creation process $\{W_i(n,t)\}_{t\geq 0}$ of each user i, which is the number of edges injected from i into the system in $[0,t]$, including those in response to failure of neighbors. Note that it is important to differentiate between the *initial* edges created when i arrives in the system and *replacement* edges generated when i repairs failed links. The latter type creates complicated dependencies between the nodes and poses the main challenges we address later in the paper.

In the second half of this paper, we deal with edge arrival from $n-1$ users to a given live node v. We prove that despite user heterogeneity, non-Poisson arrival dynamics, irregular initial out-degree, and two types of dependent edge-creation processes (i.e., initial and replacement), the aggregate edge-arrival process from the system to each live node v converges in distribution as $n \to \infty$ to a Poisson process with fixed rate

$$\gamma = \frac{k+\theta}{l}, \tag{1}$$

where k is the *average* initial out-degree across all users, θ is the mean number of replacement neighbors that an existing user obtains during its ON duration, and l is the average uptime across all nodes.

II. Related Work

The first classical area that deals with dynamic graphs falls under the umbrella of mobile ad-hoc networks. Some of the recent work [3], [25] deals with routing under the uniform speed-limited mobility model, which allows these systems to be modeled in the framework of uniform neighbor selection studied in this paper. Additional results [26] explicitly rely on the Poisson assumption on edge arrival to each user, but do not formally establish the basis for it.

The second classical area of dynamic graphs studies the behavior of P2P networks under churn [2], [4], [8], [10], [12], [13], [16], [18], [22], [23], where measurement studies [28] have shown that P2P users are highly volatile (i.e., join and depart frequently). Traditional analytical work in P2P [13], [16], [23], [24] relies on a single homogeneous

*Supported by NSF grant CNS-1017766.

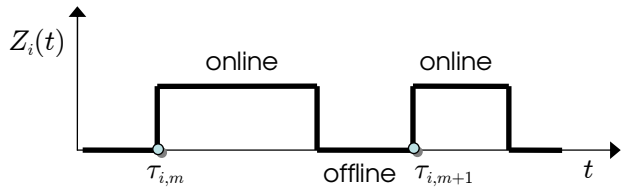


Fig. 1. Process $\{Z_i(t)\}$ depicts ON/OFF behavior of user $i = 1, \ldots, n$.

Poisson process that describes the arrival of all users. With the exception of [24], these studies additionally require that user lifetimes be exponentially distributed, which clearly does not hold in practice [28]. Other P2P modeling studies [14], [15], [30], [31] focus on local neighbor failure/replacement processes to understand how they affect user isolation and ultimately resilience of the system to partitioning.

Prior work has led to interesting insight into performance of random graphs under dynamic edge behavior; however, the question of whether these conclusions hold under more complex node activities remains open. It is our belief that the existing analysis would greatly benefit from the rigorous explanation of how random graphs under churn develop Poissonian dynamics and the specific closed-form expressions for the arrival rate established below.

III. General Edge-Creation Model

We start by formalizing the user join/departure model and the out-link model of dynamic networks. We use notation "node" and "user" interchangeably throughout the paper.

A. Overview

Each node becomes *active* when arriving in the system or *inactive* when departing from the system. Define $Z_i := \{Z_i(t)\}$ to be an ON/OFF right-continuous process on time interval $[0, \infty)$, indicating online/offline states of user i:

$$Z_i(t) := \begin{cases} 1 & \text{user } i \text{ is ON at time } t \\ 0 & \text{otherwise (OFF)} \end{cases}, \quad t \geq 0, \tag{2}$$

where $i = 1, 2, \ldots, n$, and n is the number of participating nodes. Define random variables $L_{i,m} > 0, D_{i,m} > 0$ to be the m-th durations of user i's ON and OFF periods, respectively. Denote by $\{\tau_{i,m}\}_{m=1}^{\infty}$ the arrival times of user i, as illustrated in Fig. 1, and by $\{M_i(t)\}$ the point process that counts the number of times i joins the system in the interval $[0, t]$:

$$M_i(t) := \sum_{m=1}^{\infty} 1_{\tau_{i,m} \in [0,t]}, \tag{3}$$

where $\tau_{i,m+1} = \tau_{i,m} + L_{i,m} + D_{i,m}$, for $m \geq 1$.

Whenever arriving in the network, each user i creates k_i out-links to other nodes, which it continuously monitors and repairs as its neighbors become dead. Denote by $Y_i^c := \{Y_i^c(n,t)\}$ the alternating process representing the states of i's out-link c:

$$Y_i^c(n,t) := \begin{cases} 1 & \text{out-link } c \text{ of user } i \text{ is ALIVE at } t \\ 0 & \text{otherwise (DEAD)} \end{cases}, \tag{4}$$

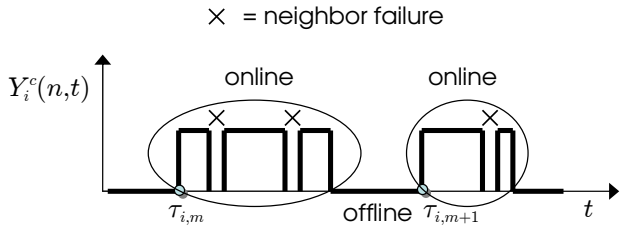


Fig. 2. Process $\{Y_i^c(n,t)\}$ indicates DEAD/ALIVE behavior of the c-th out-link of user i, for $c = 1, \ldots, k_i$.

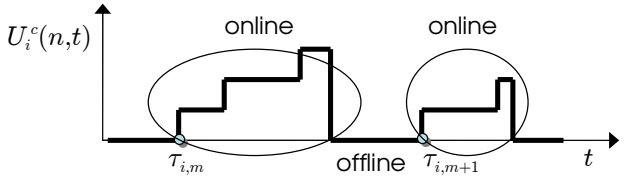


Fig. 3. Process $\{U_i^c(n,t)\}$ counts the number of DEAD→ALIVE transitions within the current ON/OFF cycle of i.

for $c = 1, \ldots, k_i$ and $t \geq 0$, where ALIVE means that the neighbor adjacent to this link is currently online and DEAD means that the neighbor has become failed and a replacement is being sought. If node i is offline at t, all of its links are considered DEAD.

As depicted in Fig. 2, user i's link c becomes ALIVE at arrival times $\{\tau_{i,m}\}_{m \geq 1}$ and then alternates between DEAD and ALIVE states during i's ON periods. Note that DEAD durations are *search delays*[1] for finding replacement neighbors and ALIVE durations of Y_i^c are *residual lifetimes* of selected neighbors with the exception of the very last ALIVE cycle before i leaves, which is terminated by i's departure rather than neighbor failure.

Whenever Y_i^c transitions from DEAD to ALIVE, user i creates an edge (i.e., performs one selection). Define *initial edges* to be those added when users arrive in the system (i.e., whenever Z_i transitions from OFF to ON) and *replacement edges* to be those added in response to neighbor failures. The collection of processes $\{Y_i^c\}_{i,c}$ contains elements with rather complex inter-dependency since multiple nodes may concurrently connect to the same neighbor and a given user i may select the same node v multiple times during different ON durations of i (see the introduction for more discussion).

Define $\{U_i^c(n,t)\}$ to be a right-continuous process that counts the number of transitions DEAD→ALIVE of Y_i^c *within* the current ON/OFF cycle of i up to time t, which is shown in Fig. 3. We assume $U_i^c(n, \tau_{i,m}) = 1$, use notation t^- to represent the instant just prior to t, and denote by

$$U_i^c(n, \tau_{i,m+1}^-) = \sup_{\tau_{i,m} \leq t < \tau_{i,m+1}} U_i^c(n,t) \tag{5}$$

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

[1]We assume that search delays are negligible compared to user lifetimes. The impact of search delays on graph dynamics is not studied in this paper, but can be found in, e.g., [14].

Finally, the number of selections that i performs for link c across all of its ON durations in $[0,t]$ is given by

$$W_i^c(n,t) := U_i^c(n,t) - U_i^c(n,0) + \sum_{m=1}^{M_i(t)} U_i^c(n,\tau_{i,m}^-), \quad (6)$$

where $M_i(t)$ is the number of arrivals of i in $[0,t]$. Then, define the *edge-creation process* $W_i(n,t)$ of user i as the superposition of k_i processes W_i^c, i.e., $W_i(n,t) := \sum_{c=1}^{k_i} W_i^c(n,t)$.

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 live users in $[0,t]$.

B. Assumptions

We first split the system into $\mathcal{K}$ unique user types to model the heterogeneous nature (i.e., diversity) of user behavior (e.g., $\mathcal{K}=1$ reduces the system to a network of homogeneous users) and mandate that all average online/offline durations be both positive and finite.

Assumption 1: There exists a fixed set $\mathcal{F}$ of $\mathcal{K} \geq 1$ distinct user types, which are specified by pairs of continuous CDFs $(F^{(j)}(x), G^{(j)}(x))$ that define non-negative random variables whose means

$$l^{(j)} := \int_0^\infty (1 - F^{(j)}(x))dx \quad (7)$$

$$d^{(j)} := \int_0^\infty (1 - G^{(j)}(x))dx \quad (8)$$

satisfy $0 < l^{(j)}, d^{(j)} < \infty$ for all $j = 1, \ldots, \mathcal{K}$. Furthermore, $F^{(j)}(x) > 0$ and $G^{(j)}(x) > 0$ for all $x > 0$, all $j = 1, \ldots, \mathcal{K}$.

We next allow for bias in the selection process and let certain user types be more popular than others.

Assumption 2: Each Z_i is associated with a pair of CDFs $(F_i(x), G_i(x))$ that is independently drawn from set $\mathcal{F}$, where type j is selected with probability $p_j \geq 0$ and $\sum_{j=1}^{\mathcal{K}} p_j = 1$;

We finally ensure that once users have chosen their types (i.e., ON/OFF duration CDFs), $\{Z_i(t)\}_{i=1}^n$ evolve as independent stationary renewal processes.

Assumption 3: Defining $\mathcal{S}$ to be set of selections made by each user and conditioning on $\mathcal{S}$, set $\{Z_i(t)\}_{i=1}^n$ consists of mutually independent, stationary alternating renewal processes, where ON durations $\{L_{i,m}\}_{m=1}^\infty$ are independently identically distributed (i.i.d.) with CDF $F_i(x)$ and OFF durations $\{D_{i,m}\}_{m=1}^\infty$ are i.i.d. with distribution $G_i(x)$.

We next regulate the initial out-degree of each user, ensuring that all k_i are uniformly bounded as $n \to \infty$ and there is no dependency between Z_i and k_i.

Assumption 4: The initial out-degree k_i of each user i is independently drawn from some distribution $K(x)$ with mean k whose probability space is upper-bounded by a constant independent of n. Furthermore, given that users have set their k_i, the number of out-degree neighbors each user i connects to at arrival times $\tau_{i,m}$ is k_i for all $m \geq 1$.

For simplicity of notation, conditioning on user type, define $l_i := E[L_{i,m}|i\text{'s type}]$ and $d_i := E[D_{i,m}|i\text{'s type}]$ to be the mean ON/OFF durations of each user i, respectively. Let

$$\lambda_i := (l_i + d_i)^{-1} \quad (9)$$

be the arrival rate of user i into the system. To ensure stationarity, the first arrival time $\tau_{i,1}$ is defined as

$$\tau_{i,1} := \begin{cases} L_i^e + D_i & \text{with probability (w.p.) } a_i \\ D_i^e & \text{w.p. } 1 - a_i \end{cases}, \quad (10)$$

where L_i^e has the equilibrium distribution of $F_i(x)$ (i.e., $P(L_i^e < x) = (l_i)^{-1}\int_0^x (1 - F_i(u))du, x \geq 0$), D_i^e has the equilibrium distribution of $G_i(x)$, and $a_i := l_i/(l_i + d_i)$ is user i's availability.

C. Properties

In this subsection, we focus our attention on the properties of aggregate user lifetimes and residual lifetimes of selected neighbors, upon which we compute the rate of edge-creation from each user in the next section.

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). Given n participating users, 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. By [30, Theorem 1], the aggregate lifetime CDF $F(n,x)$ and its mean $l(n)$ for any finite $n \geq 1$, are respectively given by

$$F(n,x) = \sum_{i=1}^n b_i F_i(x), \quad l(n) = \sum_{i=1}^n b_i l_i, \quad (11)$$

where $b_i := \lambda_i / \sum_{j=1}^n \lambda_j$ and λ_i is defined in (9). The asymptotic results on $F(n,x)$ and $l(n)$ are given below.

Lemma 1: Under Assumptions 1-3, the following sequences converge almost surely (a.s.) as $n \to \infty$:

$$F(n,x) \xrightarrow{a.s.} F(x) := \frac{\sum_{j=1}^{\mathcal{K}} p_j \lambda^{(j)} F^{(j)}(x)}{\lambda}, \quad (12)$$

$$l(n) \xrightarrow{a.s.} l := \frac{a}{\lambda}, \quad (13)$$

where $\lambda^{(j)} := 1/(l^{(j)} + d^{(j)})$ is the arrival rate of type j, $\lambda := \sum_{j=1}^{\mathcal{K}} p_j \lambda^{(j)}$ is the average arrival rate across all user types, $a := \sum_{j=1}^{\mathcal{K}} p_j a^{(j)}$, and $a^{(j)} := l^{(j)}/(l^{(j)} + d^{(j)})$. Further, $F(x)$ is a proper CDF function and $0 < l < \infty$.

Now, suppose node v picks a random currently-alive user i as a potential neighbor. Denote by $R_i(t)$ the remainder of the current ON cycle of user i, as illustrated in Fig. 4, and by $H_i(x)$ its CDF:

$$H_i(x) := P(R_i(t) \leq x | Z_i(t) = 1), \quad (14)$$

which is invariant in t due to stationarity of Z_i. It is known from [29] that for $x \geq 0$: $H_i(x) = (l_i)^{-1}\int_0^x (1 - F_i(u))du$.

Next, define $R(n,t)$ to be the residual lifetime of the user *uniformly randomly selected* from among peers that are alive at time t. Let $N(n,t) := \sum_{i=1}^n Z_i(t)$ be the system population

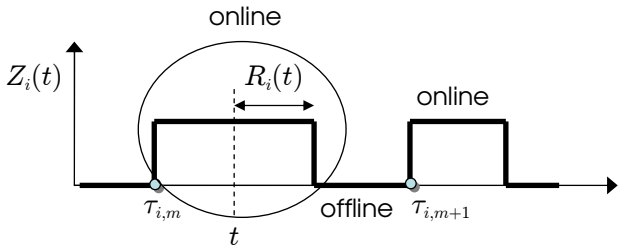


Fig. 4. Residual $R_i(t)$ is the remaining online duration of user i until it departs, conditioned on $Z_i(t) = 1$.

at that time and denote by $H(n, x)$ the distribution of $R(n, t)$ conditioned on at least one user being alive:

$$H(n, x) := P(R(n, t) \le x | N(n, t) \ge 1). \tag{15}$$

The next lemma shows that under uniform selection, the residual CDF $H(n, x)$ can be reduced to a simple function of the aggregate lifetime CDF $F(x)$.

Lemma 2: Given Assumptions 1-3 and uniform selection, $H(n, x)$ defined in (15) converges a.s. to

$$H(n, x) \xrightarrow{a.s.} H(x) := \frac{1}{l} \int_0^x (1 - F(u))\, du, \tag{16}$$

as $n \to \infty$, where $F(x)$ and l are given in (12)-(13).

Recalling that ALIVE durations of Y_i^c are residual lifetimes of selected users and applying (16), we next study edge-creation processes $\{W_i(n, t)\}_{i=1}^n$.

IV. Edge Creation

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. Note that the z-th selection time $\delta_{i,z}$ of user i is determined by the ON/OFF process Z_i and residual lifetimes of i's all *previous* selections $z' < z$. This indicates that the history observed by i up to time $\delta_{i,z}$ helps i predict whether the previously selected peers are alive at $\delta_{i,z}$. The model for the number of users available for selection at time $\delta_{i,z}$ is rather intricate due to the dependence on i's past selections.

In the following, we first deal in Lemma 3 with uniform integrability (UI) of set $\{(W_i(n, t))^r\}_{n\ge1}$, where $W_i(n, t)$ is the number of edges created by i in interval $[0, t]$, and then prove in Lemma 4 that residuals $R(n, \delta_{i,z})$ of selected neighbors at random times $\delta_{i,z} \le t$ are asymptotically independent. These two critical results make the distribution of $R(n, \delta_{i,z})$ tractable and lay the foundation for studying the edge-creation process in Theorem 1.

A. Preliminaries

We start by addressing UI of set $\{(W_i(n, t))^r\}_{n\ge1}$.

Lemma 3: With Assumptions 1-4 and uniform selection, for each user i and all $t \ge 0$, collections $\{(W_i(n, t))^r\}_{n\ge1}$ are uniformly integrable in n, for any $r > 0$.

Lemma 3 is needed to show that $E[(W_i(n, t))^r]$ converge as $n \to \infty$, for all $r > 0$, and that the intensity measure $E[W_i(n, t)]$ of each edge-creation process is finite.

Lemma 4: Under the same assumptions as in Lemma 3, residuals $\{R(n, \delta_{i,z})\}_{z\ge1}$ of selected neighbors at random times $\delta_{i,z}$ are asymptotically independent and converge in distribution as $n \to \infty$ to i.i.d. r.v.'s with CDF $H(x)$ in (16).

The rationale behind Lemma 4 can be explained as follows. Intuitively, *as the pool of users available for selections becomes larger*, the probability that user i selects *any* other peer *more than once* in interval $[0, t]$ diminishes. In other words, for sufficiently large n, it is more likely that the set of neighbors that i connects to in $[0, t]$ contains *distinct* peers. Therefore, given the assumption that each user's arrival/departure times are independent of any other nodes', the CDF of residuals selected at $\delta_{i,z}$ approaches that of residual $R(n, t)$ of a randomly selected user in (16).

For simplicity of notation, denote by $\{U(t)\}_{t\ge0}$ a pure renewal process (with a point at time 0) whose cycle lengths follow $H(x)$ in (16). The expected number of renewals (i.e., points) in interval $[0, x]$ is given by [21]

$$E[U(x)] = 1 + \sum_{r=1}^{\infty} H^{*r}(x), \tag{17}$$

where $H^{*r}(x)$ is the r-fold convolution of $H(x)$.

Using the regenerative property of process $\{U_i^c(n, t)\}$ defined in (5), we next show that $E[W_i(n, t)]$ can be reduced to a function of $E[U(t)]$ in (17).

Theorem 1: Assuming the same as in Lemma 3, the mean number of edges that user i generates in $[0, t]$ converges as $n \to \infty$ to

$$E[W_i(n, t) \,|\, i\text{'s type}, k_i] \to k_i \lambda_i t \int_0^{\infty} E[U(x)]\, dF_i(x), \tag{18}$$

where k_i is i's initial out-degree, λ_i is its arrival rate, $F_i(x)$ is its lifetime CDF, and $E[U(x)]$ is given in (17). Unconditionally,

$$\lim_{n\to\infty} E[W_i(n, t)] = (k + \theta)\lambda\, t, \tag{19}$$

where λ is the arrival rate shown in (13), k is the mean initial out-degree, $\theta := k \sum_{r=1}^{\infty} \int_0^{\infty} H^{*r}(x) dF(x)$ is the mean number of replacement neighbors that i selects in its lifetime, and $F(x)$ is the aggregate lifetime distribution in (12).

To understand (18), recall that the mean number of arrivals of user i in interval $[0, t]$ is simply $\lambda_i t$ due to stationarity of renewal process $\{Z_i(t)\}$. During each ON duration L_i, i generates $E[U(L_i)]$ edges on average for out-link c, for $c = 1, \ldots, k_i$. The total number of edges created in $[0, t]$ is thus $k_i \lambda_i t E[U(L_i)]$, which leads to (18) upon expanding $E[U(L_i)]$. Averaging (18) over user types and initial out-degree, we reach (19).

We are now ready to examine the process of incoming edge arrival to a given live user in the system.

B. Superposition of Edge Arrival Processes

We set aside a *given* user v (so the values of $Z_v(\cdot)$ are given) and examine edge arrival 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 $i \neq v$, where $I_{i,z}^v$ is the indicator that i selects v at time $\delta_{i,z}$. Then, the edge-arrival process from $n-1$ other nodes to user v is the superposition of processes $\xi_{n,i}$:

$$\xi_n(t) := \sum_{i=1, i\neq v}^{n} \xi_{n,i}(t) = \sum_{i=1, i\neq v}^{n} \sum_{z=1}^{W_i(n,t)} I_{i,z}^v. \qquad (20)$$

Invoking Lemmas 3-4 and Theorem 1 and applying the result on the convergence of random measures in [20, Proposition 3.22], we reach the main result of this paper.

Theorem 2: Under Assumptions 1-4 and uniform selection, conditioned on Z_v, the point process ξ_n defined in (20) 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,

$$\gamma := (k+\theta)/l, \qquad (21)$$

k and θ are in (19), and l is the mean lifetime given in (13).

The local rate $\gamma Z_v(t)$ in Theorem 2 means that when user v remains alive in the system (i.e., $Z_v(\cdot) = 1$), the instantaneous rate of edge arrival to v is the constant γ; otherwise it is 0 (i.e., no users connect to v when it is offline). This immediately implies that unconditioning Z_v, the edge-arrival process ξ_n to user v converges in distribution to a Cox process directed by the random process Z_v.

Theorem 2 states that despite multiple user-types, non-Poisson user-arrival dynamics, and complex edge-creation dependencies, the arrival of edges to any online user v is Poisson, which holds for *any* pair of online/offline distributions that fall under the assumptions introduced earlier. This allows future work to obtain relatively simple models for many important metrics (e.g., in-degree, routing load) that are interesting to network designers.

V. Conclusion

We introduced a generic framework for modeling user join/departure and edge arrival in distributed systems under churn, which allows tractable analysis of system dynamics as $n \to \infty$. Using this framework, we developed theoretical results on the edge-arrival process to each user and paved the way for future analytical work in this area that tackles load balancing, dynamic topology adaptation, routing distances, and other important issues.

References

[1] J. Ansari and P. Mähönen, "Channel Selection in Spectrum Agile and Cognitive MAC Protocols for Wireless Sensor Networks," in *Proc. ACM MobiWac*, Oct. 2010, pp. 83–90.

[2] M. Castro, M. Costa, and A. Rowstron, "Performance and Dependability of Structured Peer-to-Peer Overlays," in *Proc. IEEE DSN*, Jun. 2004, pp. 9–18.

[3] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot, "The Diameter of Opportunistic Mobile Networks," in *Proc. ACM CoNEXT*, Dec. 2007.

[4] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, "Making Gnutella-like P2P Systems Scalable," in *Proc. ACM SIGCOMM*, Aug. 2003, pp. 407–418.

[5] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in *Proc. USENIX OSDI*, Dec. 2004, pp. 137–150.

[6] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: Amazons Highly Available Key-Value Store," in *Proc. ACM SOSP*, Oct. 2007, pp. 205–220.

[7] Gnutella. [Online]. Available: http://en.wikipedia.org/wiki/Gnutella.

[8] B. Godfrey, S. Shenker, and I. Stoica, "Minimizing Churn in Distributed Systems," in *Proc. ACM SIGCOMM*, Sep. 2006, pp. 147–158.

[9] S. Govindasamy, D. Bliss, and D. Staelin, "Spectral Efficiency in Single-Hop Ad-Hoc Wireless Networks with Interference Using Adaptive Antenna Arrays," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1358–1369, Sept. 2007.

[10] 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.

[11] P. Gupta and P. Kumar, "Critical Power for Asymptotic Connectivity," in *Proc. IEEE CDC*, Dec. 1998, pp. 1106–1110.

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

[13] 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.

[14] 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.

[15] D. Leonard, Z. Yao, X. Wang, and D. Loguinov, "On Static and Dynamic Partitioning Behavior of Large-Scale Networks," in *Proc. IEEE ICNP*, Nov. 2005, pp. 345–357.

[16] 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.

[17] 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.

[18] P. Maymounkov and D. Mazieres, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," in *Proc. IPTPS*, Mar. 2002, pp. 53–65.

[19] C. Reddy, D. Leonard, and D. Loguinov, "Optimizing Capacity-Heterogeneous Unstructured P2P Networks for Random-Walk Traffic," in *Proc. IEEE P2P*, Sep. 2009, pp. 41–50.

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

[21] S. Resnick, *Adventures in Stochastic Processes*. Birkhäuser, 2002.

[22] S. Rhea, D. Geels, T. Roscoe, and J. Kubiatowicz, "Handling Churn in a DHT," in *Proc. USENIX ATC*, Jun. 2004, pp. 127–140.

[23] 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.

[24] G. Tan and S. Jarvis, "Stochastic Analysis and Improvement of the Reliability of DHT-based Multicast," in *Proc. IEEE INFOCOM*, May 2007, pp. 2198–2206.

[25] D. Tschopp, S. Diggavi, and M. Grossglauser, "Hierarchical Routing over Dynamic Wireless Networks," in *Proc. ACM SIGMETRICS*, Jun. 2008, pp. 73–84.

[26] M. Vojnovic and A. Proutiere, "Hop Limited Flooding over Dynamic Networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 685–693.

[27] X. Wang and D. Loguinov, "Load-Balancing Performance of Consistent Hashing: Asymptotic Analysis of Random Node Join," *IEEE/ACM Trans. Networking*, vol. 15, no. 4, pp. 892–905, Aug. 2007.

[28] X. Wang, Z. Yao, and D. Loguinov, "Residual-Based Estimation of Peer and Link Lifetimes in P2P Networks," *IEEE/ACM Trans. Networking*, vol. 17, no. 3, pp. 726–739, Jun. 2009.

[29] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*. Prentice Hall, 1989.

[30] Z. Yao, D. Leonard, X. Wang, and D. Loguinov, "Modeling Heterogeneous User Churn and Local Resilience of Unstructured P2P Networks," in *Proc. IEEE ICNP*, Nov. 2006, pp. 32–41.

[31] Z. Yao, X. Wang, D. Leonard, and D. Loguinov, "Node Isolation Model and Age-Based Neighbor Selection in Unstructured P2P Networks," *IEEE/ACM Trans. Networking*, vol. 17, no. 1, pp. 144–157, Feb. 2009.