# From Isolation Time to Node Resilience: Impact of Cascades in D2D-based Social Networks

Sigit Aryo Pambudi  Wenye Wang
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27606
Email: {sapambud,wwang}@ncsu.edu

Cliff Wang
Army Research Office
Research Triangle Park, NC 27709
Email: cliff.wang@us.army.mil

***Abstract*—The ever-increasing traffic demand from social networking service (SNS) users and recent progress in device-to-device (D2D) technology have empowered a new D2D-based SNS paradigm, which enables multimedia content exchange via short-range wireless networking. In this paradigm, a small node failure may trigger a collection of rapidly-spreading isolation events called cascade-of-failures. Unlike existing works that studied the outcome of cascading failures from the spatial and probabilistic perspectives, *this paper sheds light on the temporal properties of the cascade-of-failures.* To do this, we introduce a maximum isolation time that quantifies the steps needed until the last node is isolated by the cascades, and then show that it scales non-monotonically to the fraction of initial survivors (non-failure nodes) and increases logarithmically with the network size. Then, we use the result to further analyze a node resilience metric, which is the likelihood that a node does not become isolated before its social networking session is finished. These findings, which are validated using numerical simulations, provide a temporal perspective of network performance that is valuable in the design of D2D-based SNSs yet still missing in the literature.**

## I. Introduction

The ever-increasing popularity of online social networking services (SNSs) such as Facebook and Twitter and the enhanced mobile devices' multimedia capability–made possible by multi-core processors and high definition displays–have pushed more than one billion users to access SNSs on-the-go [1]. This, combined with recent advances in the short-range device-to-device (D2D)-based wireless technologies (such as, Apple iOS's Multipeer Connectivity framework [2]), paved way for a new paradigm of providing SNS access via D2D communications, briefly referred as *D2D-based SNS* [3]. Such a paradigm has two key advantages: (i) low-power, yet high-speed communication thanks to the short transmission range [4], and (ii) the ability to *de-congest* LTE and WiFi networks via traffic offloading. By 2015, several D2D-based SNS apps (i.e., FireChat and Jott) have gained momentum rapidly, especially among millions of teenagers and students [5].

Despite its potential usefulness and popularity, little is known regarding the performance of D2D-based SNSs, especially under *failures*. Here, failure is the inability of a node to interact (i.e., exchange social contents) via a D2D-based SNS that may be caused by either malware infection and/or hardware breakdown [6]. In the literature, there are extensive

This work is supported in part by Army Research Office under grant number W911NF-15-2-0102 and NSF Award number CNS-1527696.

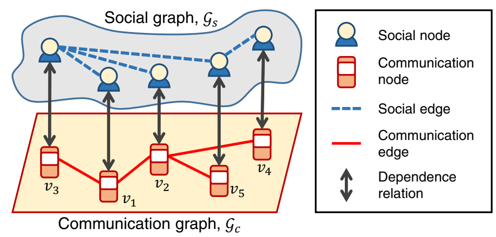



Fig. 1. Illustration of D2D-based SNS's structure.

studies regarding node failures in either SNS- or D2D-only network. For example, trace-driven simulations were used to investigate the characteristics of malware infection in online SNS [7], while the D2D network connectivity under misbehaving nodes was analyzed in [8]. However, these methods for analyzing SNS and D2D networks *separately* cannot be applied directly to D2D-based SNS that consists of co-existing D2D (communication) and SNS (social) networks with interdependence links, as depicted in Fig. 1.

In view of the co-existing communication and social networks, a D2D-based SNS can be regarded as a set of *interdependent networks* [9]. In such networks, an initial node failure may trigger another set of isolated nodes over the next time steps, which is referred as the *cascade-of-failures* [9]. As a result, each node has a corresponding *isolation time*. Although an important metric for analyzing the D2D-based SNS's resilience against node failures [3], little has been known regarding the isolation time.

Hence, a few questions remain unanswered. First, as the initial survival probability $\rho_0$ increases, there are less initial node failures, which result in fewer isolated nodes due to the cascade-of-failures [9]. We ask, *how will this impact the isolation time?* Second, when $\rho_0$ is constant, the number of isolated nodes due to the cascades will increase as the network size $N$ becomes higher. Thus, *how will isolation time be increased/decreased with respect to a unit growth in $N$?* Third, the isolation time directly affects the node resilience $\Psi_n$, which quantifies the D2D-based SNS's ability to provide satisfactory networking performance from users' perspective [3]. So, *what is the impacts of the above-mentioned relation between the isolation time versus $\rho_0$ and $N$ to $\Psi_n$?*

To answer the aforementioned questions, we define a *maximum isolation time* $T_{max}$ that serves as an upper bound to the node isolation time. To proceed, we describe the D2D-

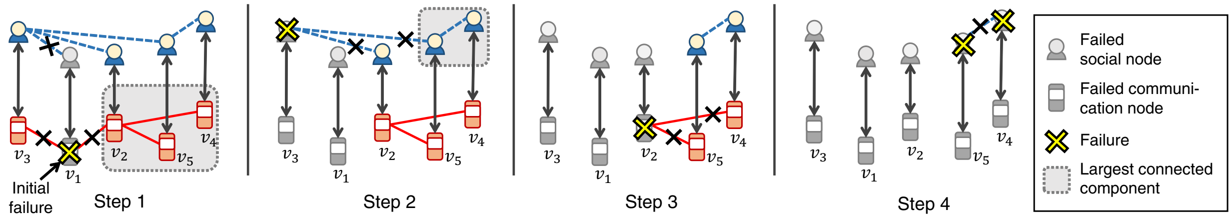



Fig. 2. Illustration of a cascade-of-failures process in D2D-based SNS.

based SNS using a self-consistent equation, which specifies the fraction of non-isolated (functional) nodes between consecutive time steps. Then, we calculate $T_{max}$ by finding the number of steps until the equation converges, and show that $T_{max}$ grows logarithmically with $N$ but is non-monotonic to $\rho_0$. By combining the bounds of node resilience metric $\Psi_n$ in [3] with the scaling order of $T_{max}$, we derive the upper bound of $\Psi_n$'s scaling order versus $\rho_0$ and $N$. Finally, using numerical simulations, we demonstrate that the scaling laws of $T_{max}$ and $\Psi_n$ are tight and valid.

The rest of this paper is outlined as follows. Section II describes the D2D-based SNS model and our problem formulation. Section III analyzes the scaling order of $T_{max}$ and $\Psi_n$, which are validated numerically in Section IV. Finally, Section V concludes this paper.

## II. Network Model and Cascading Failures

In this section, we discuss the D2D-based SNS model and how cascade-of-failures are formed in such a network. Then, we present a maximum isolation time metric and state our main objective of analyzing the scaling order of this metric.

### A. Model of D2D-Based Social Networking Service

We consider a wireless communication network with the set of $N$ nodes denoted as $\mathcal{V}_c = \{v_1, \ldots, v_N\}$. Let $\mathcal{E}_c$ be the set of *communication edges*, where edge $e(i,j)$ is in $\mathcal{E}_c$ if nodes $v_i, v_j \in \mathcal{V}_c$ can communicate directly with each other via short-range D2D technology such as Bluetooth peer-to-peer and/or WiFi Direct [2]. Then, the inter-node wireless connection can be described by a *communication graph* $\mathcal{G}_c := (\mathcal{V}_c, \mathcal{E}_c)$. An example of the communication graph is depicted in the lower part of Fig. 1. We assume that store-and-forward mechanism [10] is employed so that nodes without direct communication edge (i.e., node pair $v_3$ and $v_5$) can still exchange packets via multi-hop communications (e.g., via $(v_3, v_5)$-path $\{e(v_3,v_1), e(v_1,v_2), e(v_2,v_5)\}$). To ensure that multi-hop communication is enabled between all node pairs, we assume $\mathcal{G}_c$ is *connected*.[1]

The D2D communication network (graph) is used to provide an SNS, which is a networking service that facilitates content exchange between users with common interest(s). We assume that each communication node is used by and corresponds to exactly one SNS user such that the set of social users is denoted as $\mathcal{V}_s = \mathcal{V}_c$. Let $\mathcal{E}_s$ be the set of *social edges*, where $e(i,j)$ is in $\mathcal{E}_s$ if users $v_i$ and $v_j$ are socially-related, that is, they have a common interest. For example, $v_i$ and $v_j$ can belong to the same chat room in Jott D2D-based social messaging app [5]. Then, the social relation between nodes can be described by a *social graph* $\mathcal{G}_s := (\mathcal{V}_s, \mathcal{E}_s)$, which is exemplified in the upper part of Fig. 1. In this figure, social edges (e.g., $e(3,5) \in \mathcal{E}_s$) capture the possible social content endpoints, while the contents are physically delivered by paths described by the communication edges (e.g., $(v_3, v_5)$-path $\{e(3,1), e(1,2), e(2,5)\}$). As a result, their graphs jointly capture how contents are delivered, and we refer their combination as a *D2D-based SNS graph* $\mathcal{G} := (\mathcal{G}_c, \mathcal{G}_s)$.

Note that the D2D-based SNS graph $\mathcal{G}$ can be varying over time due to node mobility and new social neighbor formation. However, by dividing the time into slots, the snapshot of $\mathcal{G}$ at each slot is a static, non-time-varying graph.

### B. Model of Cascading Failures

Due to the open nature of the wireless medium and the device capability limits, D2D-based SNS nodes are susceptible to malicious software infection as well as hardware malfunction [6]. These impairments render the nodes to *failure*, i.e., the inability to exchange social contents through the network. In this subsection, we discuss how a few node failures can trigger a sequence of time-increasing node failures called *cascade-of-failures*.

Let the network lifetime be divided into regularly-spaced discrete time steps $t \in \mathbb{N}_0$. Denote $\mathcal{I}_0 \subset (\mathcal{V}_c = \mathcal{V}_s)$ as the set of failed nodes at an initial time $t = 0$, in which node $v_i$ is included in $\mathcal{I}_0$ according to a random uniform selection with an *initial failure probability* of $p_f := \mathbb{E}[\frac{|\mathcal{I}_0|}{N}] \in (0,1]$. Every node in $\mathcal{I}_0$ has its social and communication edges removed, becomes isolated from other nodes, and eventually is unable to exchange social contents via the D2D-based SNS.

To proceed, let $\mathcal{R}_c(t) \subset \mathcal{V}_c$ and $\mathcal{R}_s(t) \subset \mathcal{V}_s$ be the sets of *residual* nodes at time $t \geq 1$. Moreover, denote $\mathcal{F}_c(t) \subset \mathcal{R}_c(t)$ and $\mathcal{F}_s(t) \subset \mathcal{R}_s(t)$ as the sets of *functional* nodes, i.e., those who still belong to the largest connected component (LCC), of $\mathcal{G}_c$ and $\mathcal{G}_s$, respectively. Then, the following steps occur.

1) In this example, node $\mathcal{I}_0 = \{v_1\}$ fails at an initial time $t = 0$. In the first step, the set of residues become $\mathcal{R}_c(1) = \mathcal{V}_c \setminus \mathcal{I}_0$ and only nodes $\mathcal{F}_c(1) = \{v_2, v_4, v_5\}$ that belong to the LCC of the communication graph remain functional (see Step 1 in Fig. 2). In contrast, node $v_2$ fails and has its communication edges removed.
2) Being unable to communicate with others, node $v_2$ also loses its social capability and social edges. We refer such a causal relation as a communication-to-social (C→S) dependence. As a result, $\mathcal{R}_s(2) = \mathcal{F}_c(1)$ while only nodes $\mathcal{F}_s(2) = \{v_4, v_5\}$ remain functional, as indicated by Step 2 in Fig. 2.

[1] $\mathcal{G}_c$ is connected if there is a $(v_i, v_j)$-path between any pair of its nodes.

3) In the next step, node $v_2$, which became isolated in $\mathcal{G}_s$ and cannot reaches its social neighbors, decides to disconnect from the communication graph, inducing a social-to-communication (S→C) dependence [3]. As a consequence of $v_2$'s disconnection, the residual nodes $\mathcal{R}_c(3) = \{v_4, v_5\}$ are completely detached from each other, thus leaving $\mathcal{G}_c$ with no LCC (i.e., $\mathcal{F}_c(3) = \emptyset$).

Notice that the number of functional nodes, $|\mathcal{F}_s(t)|$, quantifies how many users can still utilize the D2D-based SNS for content exchange. To quantify the worst-case value of this metric, which depends on the randomly-selected initial failures $\mathcal{I}_0$, we define the following.

*Definition 1:* Let $s(t) := \mathbb{E}[\frac{|\mathcal{F}_s(t)|}{N}]$ be the mean fraction of functional nodes at time $t \geq 1$. Then, the *minimum portion of functional nodes* is defined as

$$s_\infty := \lim_{t\to\infty} s(t) = \min_{t\geq 1} s(t). \tag{1}$$

### C. Problem Formulation

We have defined the minimum portion of functional nodes, $s_\infty$, that is the likelihood that a randomly-selected node becomes isolated by the cascading failures. This metric, however, does not provide a complete picture regarding the cascade process, which occurs in steps and results in a distinct set of isolated nodes at each step. To be specific, the isolation time, which diminishes the users' satisfaction toward their content exchange experience [3], may be different from one node to another. In this subsection, we define the upper bound of these distinct isolation times as our performance metric and then state our objective of analyzing its scaling laws. We start by adopting the following definition from [3].

*Definition 2: Isolation time* of node $v$ is the smallest even time steps in which $v$ becomes isolated, denoted as

$$T_v := \inf\{t \geq 1 : v \in \mathcal{F}_s(t-2), v \notin \mathcal{F}_s(t)\}. \tag{2}$$

We choose even (two) time steps as the granularity to ensure that $T_v$ captures both the C→S and S→C dependencies, which occur alternatively every two time steps. Moreover, we say that a node is *affected* if it becomes isolated at the end of a cascade-of-failures. Since analyzing the closed-form of $T_v$ for all nodes $v \in (\mathcal{V}_c = \mathcal{V}_s)$ is difficult, we examine its upper bound as follows.

*Definition 3: Maximum isolation time* is the largest expected number of steps until an affected node is isolated, i.e.,

$$T_{max} := \mathbb{E}[\sup\{TI_v : v \notin \lim_{t\to\infty} \mathcal{F}_c(t)\}]. \tag{3}$$

To map $T_{max}$ to the continuous time, consider the S→C dependence of Step 3 in Fig. 2. In this step, $v_4$ becomes isolated in $\mathcal{G}_c$ if its communication neighbor does not respond within a pre-determined timeout period $T_c$. For example, $T_c$ is equal to (`hello_interval*allowed_hello_loss`) milliseconds when AODV routing protocol [10] is used. In the next step, $v_4$ is already aware that its social neighbor, $v_5$, cannot be reached and instantly becomes isolated in $\mathcal{G}_s$. As a result, two consecutive time steps amount to $T_c$ milliseconds due to the S→C dependence only. For simplicity, we assume that one time step in $T_{max}$ has an effective length of $\frac{T_c}{2}$.

Before using $T_{max}$ for problem formulation, let $x$ be a variable that represents each of the following parameters.

- **Initial survival probability**, denoted as $\rho_0 := 1 - p_f$. It is clear that the minimum portion of functional nodes $s_\infty$ increases with respect to $\rho_0$ [9]. We aim to investigate whether such a monotonic relation versus $\rho_0$ also applies to $T_{max}$.
- **Network size**, $N$. Intuitively, the cascade-of-failures needs to reach more nodes as $N$ is enlarged, thus increasing $T_{max}$. However, it is unclear how $T_{max}$ will grow with respect to $N$.

To proceed, we present the following definition from [11].

*Definition 4 (Scaling order):* Let $T_{max}(x)$ be the maximum isolation time as a function of $x$. Then, (i) $T_{max}(x) = O(f(x))$ if there exist constants $c, x_0 > 0$ such that $T_{max}(x) \leq cf(x)$ for all $x \geq x_0$, (ii) $T_{max}(x) = \Omega(f(x))$ means $f(x) = O(T_{max}(x))$, while (iii) $T_{max}(x) = \Theta(f(x))$ if $T_{max}(x) = O(f(x))$ and $T_{max}(x) = \Omega(f(x))$.

Then, given Definitions 3 and 4, our research question can be stated as follows: *"What is the scaling order of the maximum isolation time versus $\rho_0$ and $N$?"*

## III. Analysis of Scaling Order

In this section, we answer our research question by deriving the scaling order of the maximum isolation time. Then, we analyze how such a scaling order affects the D2D-based SNS's capability to provide resilient network service.

### A. Scaling Order of Maximum Isolation Time

The maximum isolation time $T_{max}$ quantifies the smallest number of steps until less than one node becomes isolated, and can be restated as $T_{max} = \inf\{t \geq 1 : s(t-2) - s(t) < \frac{1}{N}\}$. This is proportional to the time until the mean fraction of residual nodes $s(t)$ converges to $s_\infty$, where $s(t)$ is the solution to [3, Proposition 1]. By denoting

$$\rho_{cr} := \inf\{\rho_0 : s_\infty > 0\} \tag{4}$$

as the *critical fraction* of the initial residues, there are two distinct regimes for such a convergence point [9]: (i) when $\rho \geq \rho_{cr}$ (i.e., $\rho_0$ is *supercritical*), then $s_\infty$ is larger than zero (non-trivial), while (ii) $s_\infty$ is always trivial in the *subcritical* regime ($\rho_0 < \rho_{cr}$). In this section, we analyze the scaling order of $T_{max}$ for these two distinct regimes. Our result is presented as follows.

*Theorem 1:* The maximum isolation time scales as

$$T_{max} = \begin{cases} \Theta\left(\frac{\log(N_c)}{\sqrt{\rho_0 - \rho_{cr}}}\right), & \text{for } \rho_o \geq \rho_{cr}, \\ \Theta\left(\frac{1}{\sqrt{\rho_{cr} - \rho_0}}\right), & \text{for } \rho_o < \rho_{cr}. \end{cases} \tag{5}$$

*Proof:* **Supercritical regime ($\rho_0 \geq \rho_{cr}$)**. The maximum isolation time is equivalent to the amount of time until $s(t)$ converges to $s_\infty$. According to [3, eq. (1)], such a convergence is achieved when the following equation is satisfied.

$$s'(t) = \rho_0 \times g_c(\rho_0 \times g_s(s'(t-2))), \tag{6}$$

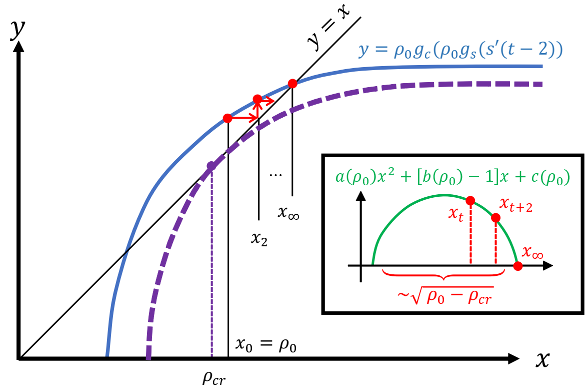



Fig. 3. Illustration for the proof of Theorem 1 when $\rho_0 \geq \rho_{cr}$.

where $s'(t) := \mathbb{E}[\frac{|\mathcal{R}_s(t)|}{N}]$ is the mean fraction of residual nodes at time $t \geq 1$. To determine the instance when (6) is satisfied, we plot the right side of (6) with $s'(t-2) = x$ as the solid blue curve in Fig. 3 and the left side with $s'(t) = x$ as the black diagonal plot. The blue plot intersects with the diagonal plot $y = x$ at two non-trivial points, since we are considering the supercritical regime ($\rho_0 \geq \rho_{cr}$). To find $T_{max}$, we start by setting $s'(0) = \rho_0$ (see point $x_0$ in Fig. 3). Let an *iteration* be a one-step calculation of $s'(t+2)$ from $s'(t)$ by using (6), which is represented as a horizontal red arrow from the blue to the black curve in Fig. 3. Then, $T_{max}$, which is the number of iterations until $s'(t-2)$ and $s'(t)$ in (6) converges to each other, is proportional to the number of arrows from $x_0$ until the blue and black plots meets at $x_\infty$.

To proceed, we assume that $\rho_0$ is near $\rho_{cr}$ such that the right side of (6), $f(x) := \rho_0 g_c(\rho_0 g_s(s'(t-2)))|_{s'(t-2)=x}$, can be approximated into the quadratic form

$$f(x) \approx a(\rho_0)x^2 + b(\rho_0)x + c(\rho_0). \tag{7}$$

Let $f_{cr}(x) := f(x)|_{\rho_0=\rho_{cr}}$ be the right side of (6) that corresponds to the critical point when $\rho_{min}$ starts to be non-zero, and is plotted as a purple dashed curve in Fig. 3. Now, the vertical distance between the quadratic function (7) and the left side of (6) is approximately equal to the vertical distance between (7) and $f_{cr}(x)$, which amounts to $(\rho_0 - \rho_{cr})$. This implies the horizontal distance between the starting point $x_0$ and the destination (convergence) point $x_\infty$ is proportional to $\sqrt{\rho_0 - \rho_{cr}}$, as indicated by the inset in Fig. 3. Since $f(x)$ is convex, its derivative is proportional to $-\sqrt{\rho_0 - \rho_{cr}}$. Moreover, according to [3, Fig. 3], the communication and social graphs can be modeled as Erdös-Rényi (ER) graphs with Poisson degree distributions. Thus, $g_c$ and $g_s$ are exponential functions [3, Proposition 2] and the iterations converge to the solution of $x_\infty$ as $f' = \exp\{-T_{max}\sqrt{\rho_0 - \rho_{cr}}\}$, where $T_{max}$ is the number of steps (red arrows in Fig. 3). This iteration will stop when there is less than one isolated nodes, that is, when $s'(t-2) - s'(t) < \frac{1}{N}$. As a result, the stopping condition can be stated as $f' \approx \frac{1}{N_c}$ and we obtain $T_{max} \sim \frac{\log(N)}{\sqrt{\rho_0 - \rho_{cr}}}$.

**Subcritical regime ($\rho_0 < \rho_{cr}$).** In the subcritical regime ($\rho_0 < \rho_{cr}$), the left and right sides of (6) intersect each other at a trivial point $x = 0$. Similar to Fig. 3, one iteration of calculating $s'(t)$ from $s'(t-2)$ using (1) can be represented as a horizontal red arrow, but now the arrow points to the left and the convergence point $x_\infty$ is at zero. Thus, our objective is to find $T_{max}$ by counting the number of iterations (arrows) from $x_0 = s'(0) = \rho_0$ to $x_\infty = s'(\infty) = 0$.

We proceed by re-employing the quadratic approximation in (7) to the right side of (1), and examining the vertical distance between the black plot and the starting point $x_0 = \rho_0$. This is equal to a small vertical distance proportional to $(\rho_{cr} - x_0)^2$, plus the gap between $f(x)$ as the right side of (1) and $f_{cr}(x) := f(x)|_{\rho_0=\rho_{cr}}$. By following the steps in the supercritical regime, this gap is equal to $(\rho_{cr} - \rho_0)$. Since the black plot representing the left side of (1) has a unity gradient, then its horizontal distance to $x_0$ is proportional to $(\rho_{cr} - \rho_0)^2 + (\rho_{cr} - \rho_0)$. For every small horizontal distance of $dx_0$, the number of steps to be taken is proportional to $\frac{dx_0}{(\rho_{cr} - x_0)^2 + (\rho_{cr} - \rho_0)}$. Notice that the total horizontal distance to be traveled by all iterations is equal to $\rho_0$, so that the total number of required steps becomes

$$T_{max} \sim \int_{x_0=0}^{\rho_0} \frac{dx_0}{(\rho_{cr} - x_0)^2 + (\rho_{cr} - \rho_0)} = \frac{\pi/2}{\sqrt{\rho_0 - \rho_{cr}}}, \tag{8}$$

where (8) can be acquired after applying [12, Sec. 2.103] and assuming that $\rho_0$ is near $\rho_{cr}$. Finally, (5) can be obtained using Definition 4 and dropping the constant $\frac{\pi}{2}$ from (8). ∎

## B. Impact of Scaling Order to D2D-based SNS's Resilience

We have discussed that users may be isolated by the cascade-of-failures after $T_{max}$ time steps. On the other hand, these users may have an ongoing social content exchange, which needs to be finished within a certain *session time*. It is desirable that every user can enjoy a resilient network connection during his/her session time, in order to have a satisfactory networking experience. In this subsection, we study how the D2D-based SNS's ability to provide such a satisfactory experience is affected by the aforementioned scaling laws of $T_{max}$.

Let a *session* represents an exchange of social contents between a user pair. Denote a session time $ST_v \in (0, \infty)$ as the time required to complete a session initiated by node $v$, which is normalized to the length of one time step, $\frac{T_c}{2}$. Motivated by real-world peer-to-peer networks [13], we assume that $ST_v$ is Pareto (heavy-tailed)-distributed. Moreover, we consider the *worst-case* session time, that is, all the sessions under consideration start at $t = 0$, while each node will not start a new session if it has detected any failure in the network.

With the aforementioned definitions, our interest is to analyze the likelihood that D2D-based SNS users can sustain its ongoing session, after the occurrence of initial failure $\mathcal{I}_0$. To do this, we employ the following metric from [3].

*Definition 5: Node resilience* is the likelihood that a randomly-chosen node $v \in (\mathcal{V}_c = \mathcal{V}_s)$ does not become isolated and remains active within its session time, that is,

$$\Psi_n := \Pr\{ST_v < \tau_v\}. \tag{9}$$

To analyze the node resilience, we present the following.

*Corollary 1:* [3] Let a D2D-based SNS with Pareto-distributed $ST_v$ with density $Pr\{ST_v = x\} = \frac{\gamma x_m^\gamma}{x^{\gamma+1}}$, shape parameter $\gamma > 0$, and scale $x_m > 0$. Then, the node resilience metric is bounded by

$$s_\infty + (1-\rho_0+s_\infty)(1-x_m^\gamma) \leq \Psi_n \leq (\rho_0+s_\infty)(\frac{x_m}{\tau_{max}})^\gamma - \rho_0. \quad (10)$$

Since the bounds in (10) depends on the minimum portion of functional nodes $s_\infty$, we derive the following result.

*Lemma 1:* Consider a D2D-based SNS with $\mathcal{G}_c$ and $\mathcal{G}_s$ modeled as ER graphs. In the supercritical regime ($\rho_0 > \rho_{cr}$), the final portion of functional scales according to

$$s_\infty = \Theta(\rho_0). \quad (11)$$

*Proof:* **(Sketch of proof)** According to realistic D2D-based networking trace [14], graphs $\mathcal{G}_c$ and $\mathcal{G}_s$ can be modeled as ER graphs with mean communication and social degrees of $\langle k \rangle_c$ and $\langle k \rangle_s$, respectively [3, Fig. 3]. As a result, the minimum portion of functional nodes can be stated as [9]

$$s_\infty = \rho_0(1-e^{-\langle k \rangle_c s_\infty})(1-e^{-\langle k \rangle_s s_\infty}). \quad (12)$$

Let $s_\infty^*$ be the smallest non-trivial (positive) $s_\infty$ that corresponds to the critical initial fraction $\rho_{cr}$. Denote the right side of (12) as $f(\rho_0) := \rho_0(1-e^{-\langle k \rangle_c s_\infty})(1-e^{-\langle k \rangle_s s_\infty})$ and let $g(\rho_0) := f(\rho_0)|_{s_\infty = s_\infty^*}$. It is clear that $f(\rho_0) \geq g(\rho_0)$. Moreover, by setting $c^* := 1/g(\rho_{cr})$, we can show that $c^* g(\rho_0) \geq 1 \geq f(\rho_0)$. These results indicate that $s_\infty = \Theta(\rho_0(1-e^{-\langle k \rangle_c s_\infty^*})(1-e^{-\langle k \rangle_s s_\infty^*}))$, which can be reduced to (11) by omitting the constant exponential terms. ■

Notice that $s_\infty$ as the solution to (12) is trivial (zero) for $\rho_0 < \rho_{cr}$. Thus, it is not necessary to analyze $s_\infty$'s scaling order for this regime.

With Corollary 1 and Lemma 1 at hand, we are ready to analyze the scaling order of $T_{max}$ as follows.

*Theorem 2:* The scaling order of the node resilience $\Psi_n$ versus the initial survival probability $\rho_0$ satisfies

$$\Psi_n = \begin{cases} O\left(\frac{\rho_0}{(\rho_0-\rho_{cr})^{\gamma/2}}\right), & \text{for } \rho_0 \geq \rho_{cr}, \\ O\left(\frac{\rho_0}{(\rho_{cr}-\rho_0)^{\gamma/2}}\right), & \text{for } \rho_0 < \rho_{cr}, \end{cases} \quad (13)$$

while the scaling order versus the number of nodes $N$ for the supercritical regime ($\rho_0 \geq \rho_{cr}$) satisfies

$$\Psi_n = O\left(\log(N)\right). \quad (14)$$

*Proof:* **Scaling order versus $\rho_0$.** To prove the scaling order in (13), we vary $\rho_0$ and fix the remaining network parameters as follows. Firstly, we consider the subcritical regime ($\rho_0 < \rho_{cr}$) in which, according to (4), $s_\infty$ is trivial (zero). Notice that Theorem 1 says there exists $c_1 > 0$ such that $T_{max} \leq c_1(\rho_0 - \rho_{cr})^{-\gamma/2}$ for any large $\rho_0$. Plugging this inequality to the right side of (10) obtains

$$\Psi_n \leq c_1 \times \rho_0(\rho_0 - \rho_{cr})^{-\gamma/2} - \rho_0, \quad (15)$$

which can easily be reduced to the second line of (13) by discarding the constant factor and applying Definition 4.

Secondly, we consider the supercritical regime ($\rho_0 \geq \rho_{cr}$). From Lemma 1, we know that there exist $c_2, \rho_{0,2} > 0$ such that $s_\infty \leq c_2 \times \rho_0$ for $\rho_0 \geq \rho_{0,2}$. Moreover, according to Theorem 1, there also exist $c_3, \rho_{0,3} > 0$ in that $T_{max} \leq c_3 \frac{\log(N)}{\sqrt{\rho_0 - \rho_{cr}}}$ for any $\rho_0 \geq \rho_{0,3}$. These inequalities can be substituted into the upper bound of (10) to get

$$\Psi_n \leq c_3(1+c_2) \times x_m \log(N) \times \left(\frac{\rho_0}{\sqrt{\rho_0 - \rho_{cr}}}\right)^\gamma - \rho_0. \quad (16)$$

By dropping the constant term $c_3(1+c_2) \times x_m \log(N)$, noticing that $\rho_0$ grows slower than $(\rho_0/\sqrt{\rho_0 - \rho_{cr}})^\gamma$, and applying Definition 4, we obtain the first line of (13).

**Scaling order versus $N$.** To prove (14), we vary $N$ and re-use the inequality in (16). By noticing that the factors $c_4 := c_3(1+c_2) \times x_m \left(\rho_0/\sqrt{\rho_0 - \rho_{cr}}\right)^\gamma$ and $\rho_0$ are constant with respect to $N$, we obtain $T_{max} \leq c_4 \log(N) - \rho_0$, which can be reduced to (14) by applying Definition 4. ■

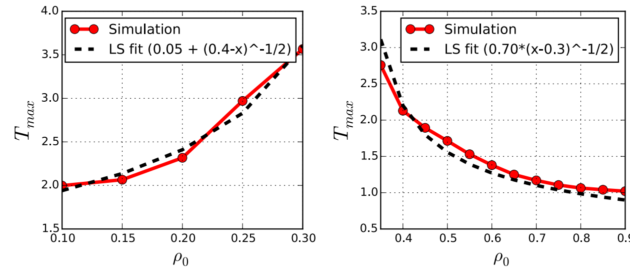
Note that the "Big-O" $O(\cdot)$ results in (13) and (14) only indicate the scaling order *upper bound* of $\Psi_n$. In the following, we will show that such an upper bound is *tight*.

## IV. NUMERICAL EVALUATION

Next, we validate the scaling order analyses in Section III. To do this, we set the network size to $N = 10^3$ and use two ER graphs with average degrees of $\langle k \rangle_c = 6$ and $\langle k \rangle_s = 10$ for the communication and social graphs, $\mathcal{G}_c$ and $\mathcal{G}_s$, respectively. These values are selected because (i) they result in a minimum portion of functional nodes, $s_\infty$, similar to realistic network traces [14] (see [6, Fig. 7]), and (ii) they ensure that $\mathcal{G}_c$ and $\mathcal{G}_s$ are connected.[2] Using these parameters, we perform numerical simulations in `Python`, where `NetworkX` framework is used to generate $\mathcal{G}_c$ and $\mathcal{G}_s$. In our numerical results, each point is an average over 50 unique D2D-based SNS graph realizations, where 100 different initial failure sets are generated for each realization.

### A. Scaling Order of Maximum Isolation Time

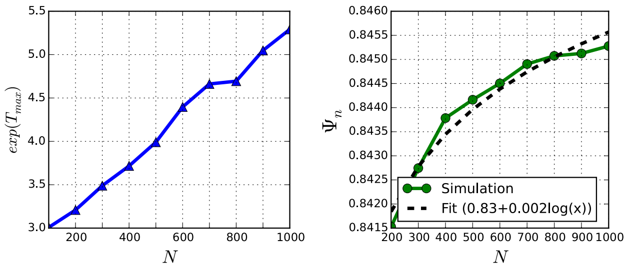
We start by plotting $T_{max}$ versus the initial survival probability $\rho_0$ in Figs. 4(a) and 4(b). From these figures, we observe that the numerical results (the red solid plots) matches the scaling order analysis in Theorem 1 (see the black dotted lines for the least-square fit). More importantly, the maximum isolation time is a convex *increasing* function




(a) Subcritical regime. (b) Supercritical regime.

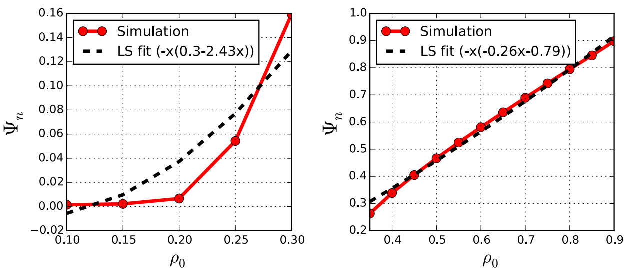
Fig. 4. Scaling order of $T_{max}$ versus initial survival probability, $\rho_0$.

[2]An ER graph with $N$ nodes is *almost surely* connected if it mean degree $\langle k \rangle$ satisfies $\langle k \rangle > (1-\epsilon)\frac{N-1}{N}\ln(N)$, where $\epsilon > 0$ is a small constant.




(a) Maximum isolation time, $T_{max}$. (b) Node resilience metric, $\Psi_n$.

Fig. 5. Scaling order versus network size, $N$.




(a) Subcritical regime. (b) Supercritical regime.

Fig. 6. Scaling order of $\Psi_n$ versus initial survival probability, $\rho_0$.

of $\rho_0$ in the subcritical regime ($\rho_0 < \rho_{cr}$) and a convex *decreasing* function when $\rho_0 \geq \rho_{cr}$. Thus, *determining the critical fraction of initial residues* (i.e., $\rho_{cr} \approx 0.3$ in Fig. 4) *is equivalent to finding* $\rho_0$ *that corresponds to the highest* $T_{max}$. Note that $\rho_{cr}$ is an important critical point [9], below which the cascading failures will disconnect the network completely.

Next, we consider the supercritical regime ($\rho_0 \geq \rho_{cr}$) and verify the scaling order of $T_{max}$ versus the network size, $N$. To do this, we set $\rho_0 = 0.65$ and plot $\exp(T_{max})$ versus $N$ in Fig. 5(a). The nearly-strait resulting plot suggests that the scaling order of $\Theta(\log(N))$ in Theorem 1 is valid. This means $T_{max}$ increases *sub-linearly* with respect to $N$ and *the cascade-of-failures propagate faster in large networks.*

### B. Scaling Order of Node Resilience Metric

We proceed by validating the scaling order of the node resilience metric, $\Psi_n$, in Theorem 2. To calculate this metric, we generate Pareto-distributed session times with shape parameter $\gamma = 2$ and scale $x_m = 1$, which results in a finite mean session time of $\mathbb{E}[ST_v] = 2$ time steps. We consider the supercritical regime by setting $\rho_0 = 0.85$ and plotting the corresponding $\Psi_n$ in Fig. 5(b). Observe that the least-square fit of the scaling order in (16) matches with the numerical result, indicating that $\Psi_n$ *increases logarithmically* with respect to $N$. Such an increase is because the cascade-of-failures will reach the last user at a later step when $N$ is large, thus giving him/her more time to finish a session.

Finally, we verify the scaling order of $\Psi_n$ versus $\rho_0$, presented in (13). From the plots in Figs. 6(a) and 6(b), which respectively depict the sub- and super-critical regimes, we observe that the numerical and analytical results agree with each other. Unlike $T_{max}$ in Fig. 4, the node resilience is both a concave (when $\rho_0 < \rho_{cr}$) and a convex (when $\rho_0 \geq \rho_{cr}$) function that *monotonically increase* with respect to $\rho_0$. This is because users are less likely to be disconnected and can enjoy stable network connection when $\rho_0$ is high.

*Remark 1:* The matching numerical and analytical results Figs. 5(b) and 6 indicate that the scaling order upper bound in Theorem 2 is *tight*. In addition, these figures imply that $\Psi_n$ can be enhanced by raising $\rho_0$ (i.e., by installing anti-malware and/or using reliable hardware) and increasing $N$ (i.e., allowing new nodes to join the D2D-based SNS).

## V. Conclusion

In this paper, we studied the temporal impact of cascades due to initial node failures in a D2D-based SNS. To do so, we analyzed the scaling order of the maximum isolation time, $T_{max}$, which quantifies the time until the last node becomes disconnected by a cascade-of-failures. Then, we used the result to study the scaling order of the node resilience metric $\Psi_n$, which characterizes the likelihood that a node can finish an ongoing content exchange before being disconnected by the cascade-of-failures. Through analytical and numerical results, we showed that $T_{max}$ and $\Psi_n$ can both be increased by expanding the network size, while the latter also improves as the number of initially-failed nodes is reduced.

## References

[1] Statista, "Statistics and facts about mobile social media." https://www.statista.com/topics/2478/, 2017. Accessed: 2017-03-30.

[2] A. Inc, "Multipeer Connectivity Framework Reference." https://developer.apple.com/library/ios/, 2016. Accessed: 2016-11-07.

[3] S. A. Pambudi, W. Wang, and C. Wang, "On the resilience of D2D-based social networking service against random failures," in *GLOBECOM'16*, pp. 1–6, Dec 2016.

[4] X. Lin *et al.*, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6727–6740, 2014.

[5] S. Buhr, "Messaging App Jott Is Blowing Up Among Junior High And High Schoolers." https://techcrunch.com/2015/06/08/page/2/, June 2015. Accessed: 2017-01-16.

[6] S. A. Pambudi, W. Wang, and C. Wang, "On the Root Cause of Dropout-Induced Contraction Process in D2D-Based Mobile Social Networks," in *ICCCN'17*, pp. 1–6, Aug 2017.

[7] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware propagation in online social networks: nature, dynamics, and defense implications," in *Proceedings of 6th ASIACCS*, pp. 196–206, ACM, 2011.

[8] F. Xing and W. Wang, "Modeling and analysis of connectivity in mobile ad hoc networks with misbehaving nodes," in *ICC'06*, vol. 4, pp. 1879–1884, IEEE, 2006.

[9] X. Huang *et al.*, "Robustness of interdependent networks under targeted attack," *Phys. Rev. E*, vol. 83, no. 6, p. 065101, 2011.

[10] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," tech. rep., 2003.

[11] Y. Xu and W. Wang, "Scheduling partition for order optimal capacity in large-scale wireless networks," in *15th ACM MobiCom*, pp. 109–120, ACM, 2009.

[12] D. Zwillinger, *Table of integrals, series, and products*. Elsevier, 2014.

[13] X. Wang, Z. Yao, and D. Loguinov, "Residual-based estimation of peer and link lifetimes in P2P networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 3, pp. 726–739, 2009.

[14] "CRAWDAD's thlab/sigcomm2009 (v. 2012-07-15)." Downloaded from http://crawdad.org/, July 2012.