Delay Analysis of Epidemic Schemes in Sparse and Dense Heterogeneous Contact Networks

Pavlos Sermpezis, and Thrasyvoulos Spyropoulos, *Member, IEEE*

Abstract—Epidemic algorithms have found their way into many areas of computer science, such as databases and distributed systems, and recently for communication in Opportunistic or Delay Tolerant Networks (DTNs). To ensure analytical tractability, existing analyses of epidemic spreading predominantly consider homogeneous contact rates between nodes. However, this assumption is generally not true in real scenarios. In this paper we consider classes of contact/mobility models with heterogeneous contact rates. Through an asymptotic analysis we prove that a first-order, mean value approximation for the basic epidemic spreading step becomes exact in the limiting case (large network size). We further derive simple closed form approximations, based on higher order statistics of the mobility heterogeneity, for the case of finite-size networks. To demonstrate the utility of our results, we use them to predict the delay of epidemic-based routing schemes and analyze scenarios with node selfishness. We validate the analytic results through extensive simulations on synthetic scenarios, as well as on real traces to demonstrate that our expressions can be useful also in scenarios with significantly more complex structure. We believe these results are an important step forward towards analyzing the effects of heterogeneity (of mobility and/or other characteristics) on the performance of epidemic-based algorithms.

Index Terms—epidemic algorithms, heterogeneous mobility, opportunistic networks, performance analysis

✦

1 INTRODUCTION

Epidemic spreading is probably one of the most popular bio-inspired principles that have made their way into computer engineering. Epidemic algorithms and variants (e.g. gossip) have been used for communication in distributed systems, synchronization of distributed databases, content searching in peer-to-peer systems, etc. Recently, epidemic-based schemes have also been proposed for routing and data dissemination in Opportunistic Networks [1]. *Opportunistic* or *Delay Tolerant Networks* (DTNs) are envisioned to support communication in case of failure or lack of infrastructure (disaster, censorship, rural areas), but also to enhance existing wireless networks (e.g., offload cellular traffic, novel applications) [2], [3].

In Opportunistic Networks, when two mobile devices, e.g. smartphones, laptops, etc., are in proximity (*in contact*) they can exchange data or information using local wireless communication (e.g. Bluetooth or WiFi). Message dissemination can be end-to-end or content-centric, yet neither the existence nor the knowledge of an end-to-end path is assumed. Nevertheless, over a sequence of node encounters, messages can get copied to many nodes as well as forwarded over multiple hops. In the *epidemic routing* case [4], any node that has a message (referred to as "infected") will forward it to any node encountered that does not have it yet (referred to as "susceptible"). While this guarantees that every node in the network will eventually receive the message, it comes with a high resource overhead. Numerous variants have been proposed to improve the resource usage of epidemic routing while maintaining good performance (see [1], [5], [6] for a detailed survey).

Since the mobility process of nodes involved (e.g. humans or vehicles carrying the devices) is, in most cases, not deterministic, the performance of epidemic-based algorithms heavily depends on the underlying contact patterns between nodes. To this end, epidemic algorithms have been extensively studied through both simulations and analytical models. While simulations with state-of-the-art synthetic models or real mobility traces can provide more reliable predictions for the *specific* scenario tested, analytical models can give quick, qualitative results and intuition, answer "what-if" questions, and help optimize protocols (e.g. choosing the number of copies in [7], or gossip probability [8]).

For the sake of tractability, state-of-the-art analytical models for epidemic spreading mainly rely on simple mobility assumptions. Such examples are Random Walk or Random Waypoint, where node mobility is stochastic and independent, identically distributed (IID) (see e.g. [8], [9], [10]). Nevertheless, numerous studies of real mobility traces [11], [12], [13], [14] reveal a different picture. One key finding is that contact rates between different pairs of nodes can vary widely. Furthermore, many pairs of nodes may never meet. This puts in question the accuracy and utility of these models' predictions. Yet, departures from these assumptions [12], [15], [16] seem to quickly increase complexity and/or limit the applicability of results.

This raises the question: can we derive *useful and accurate* closed form expressions for the performance of epidemic schemes, under more generic mobility assumptions? To this end, in this paper, we make contributions along the following directions:

- **Model.** We shed light on the factors that increase the complexity of the analysis for epidemic schemes when departing from the homogeneous mobility model. To cope

---

- *P. Sermpezis is with the Institute of Computer Science, FORTH, Greece, and the Department of Mobile Communications, EURECOM, France.*
  E-mail: sermpezis@ics.forth.gr
  T. Spyropoulos is with the Department of Mobile Communications, EURECOM, France.
  E-mail: spyropou@eurecom.fr

with complexity, we propose using a model (e.g., similar to [17], [18]) that extends the homogeneous model by incorporating heterogeneity in a probabilistic way. We discuss how such a model (i) remains quite *generic* (i.e., protocol independent) and *analytically tractable*, as the homogeneous model, and, at the same time, (ii) allows to study effects of *heterogeneous mobility* patterns. Moreover, we show how to extend the model towards capturing more complex characteristics (sparseness, node selfishness, etc.).

- **Analysis.** We use the heterogeneous model to analyse the delay of epidemic spreading: (i) We formally prove, through an *asymptotic analysis*, that a first-order approximation becomes exact in the limiting case. We also provide intuition about *when* a prediction using a single mobility parameter (e.g., similarly to the homogeneous model) can be safely used in heterogeneous networks, and *how* its accuracy is affected by the network parameters. (ii) For *finite network sizes*, we derive simple, closed form approximations for the epidemic delay. The expressions we provide involve only the 1^{st} and 2^{nd} moments of the contact rates distribution, which renders them easily applicable in performance evaluation, protocol design, etc.
- **Applications.** To demonstrate how our framework can be used in practice, we (i) derive closed form expressions for the delay of various *epidemic based protocols*, and (ii) show how further complex characteristics of opportunistic networks, like *heterogeneous node selfishness*, can be captured. Further application-related issues are presented along with simulation results, and highlight the usefulness of the heterogeneous model compared to previous modeling approaches.

The structure of the paper is as follows. In Section 2 we introduce the network model. Then, we commence our analysis by discussing the main steps and challenges (Section 3.1). We derive results for the delay of epidemic spreading for the asymptotic case (Section 3.2) and finite-size networks (Section 3.3), and extend our analysis for the case of sparse networks (Section 3.4). Applications of our basic results are presented in Section 4, where we calculate the message delivery delay of routing protocols (Section 4.1), and analyze scenarios of social selfishness (Section 4.2). In Section 5.1, we validate all our results against various synthetic simulation scenarios, and show that their accuracy is significant. Moreover, in Section 5.2, we test our theory against real traces, capturing node mobility and respective contacts, and find that useful levels of accuracy can still be achieved even for scenarios that are known to entail considerably more complexity. Finally, we discuss related work (Section 6) and conclude our paper (Section 7).

As a final note, while our focus here will be in the domain of Opportunistic and Delay-Tolerant Networking, we believe our framework has more general applicability to epidemic algorithms in many different contexts, as long as the contact process between nodes fits our contact classes. For example, one could imagine the probabilistic spread of malware over an email or chat network [19], where the spread is possible when two nodes decide to communicate. For this reason, we present our analysis in the general context of epidemic spreading. We focus on opportunistic networks when we discuss applications and compare our results to real scenarios.

2 NETWORK MODEL

Let assume a network $\mathcal{N}$ with N nodes, where two nodes can *contact* each other (i.e., exchange data, interact, etc.). Depending on the scenario considered, a contact might be two smartphones coming within (Bluetooth) transmission range, two users sending a message (or posting) to each other in an online social network (e.g. Facebook, Twitter), two people meeting each other, etc.

Since nodes can contact only during their *contact events*, a model needs only (from a performance analysis point of view) to capture the sequence of these contact events between nodes. Moreover, different node pairs might (i) contact regularly or never contact each other, (ii) contact with different frequency (rate) or with different contact processes. To this end, we can model a network $\mathcal{N}$ and the contact processes between nodes, in a generic way, as follows:

Definition 1 (Heterogeneous Contact Network)**.** *The sequence of the contact events between each pair of nodes* $\{i,j\}$ *is given by a random point process with rate* $\lambda_{ij} > 0$*. Some nodes might never come in contact, in which case* $\lambda_{ij} = 0$.

Remark: From the above definition, it can be seen that a Heterogeneous Contact Network $\mathcal{N}$ can be equivalently represented by its *contact matrix* $\mathbf{\Lambda}$,

$$\mathbf{\Lambda} = \{\lambda_{ij}\}$$

To simplify analysis, it is commonly assumed that *all* inter-contact times are *independent and exponentially distributed* with the same rate $\lambda_{ij} = \lambda$ [8], [9], [20], i.e., the contact process between each pair of nodes $\{i,j\}$ is a Poisson process with rate λ. However, *homogeneous* rates assumption, $\lambda_{ij} = \lambda$, is rather strong. Study of real traces has provided strong evidence that contacts between different pairs of nodes are in fact largely *heterogeneous*, with some pairs never meeting each other and others meeting much more frequently [12], [13], [21], which is also consistent with our intuition.

Hence, our goal in this paper is to raise the *homogeneity* assumption, and study the performance of epidemic spreading under the more realistic case of heterogeneous contact patterns. Specifically, we consider a *Heterogeneous Contact Network* (Def. 1) and we further make the following assumptions, which are a trade-off between realism and usability:

Assumption 1. *The contact process between a pair follows a Poisson process with rate* λ_{ij}.

Assumption 2. *The contact rates* λ_{ij} *are independently drawn from an arbitrary distribution* $f_\lambda(\lambda)$*, with finite mean value* μ_λ *and variance* σ_λ^2.

The first assumption (Poisson contact process) is common, and necessary in order to use a Markovian framework for the epidemic spreading. The majority of previous studies

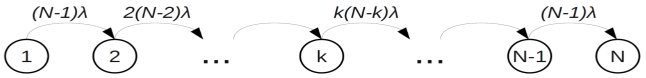


Fig. 1. Markov Chain for epidemic spreading over a *homogeneous* network with N nodes

in Opportunistic / Delay Tolerant Networks [9], [12], [16], [22], [23] assume Poisson contact processes, whereas the few studies we are aware of that assume more general inter-contact times [24], [25], [26] only deal with asymptotic behavior (i.e. whether delay is finite or infinite) [24], [25], or single-copy cases [26]. Additionally, although inter-contact time intervals cannot be expected to be always and exactly independent and exponentially distributed [27], [28], [29], there exist studies of real traces that do claim this to be the case [13], [30], or show that (aggregate) intermeeting times exhibit an exponential tail [31], [32]. Finally, findings of two recent analytic studies are consistent with our model: (a) even if the aggregate inter-contact time distribution is non-exponential (as suggested in [24]), individual pair contacts might still be exponential but with different rates [21]; and (b) even if the actual contact times are not Poisson, we can, under certain conditions, use a Markov Chain based analytical framework, as a good approximation [33].

In *Assumption 2*, we raise the homogeneity assumption and we allow different pairs to contact each other more frequently than others. While this contact class is far from exhaustive, $f_\lambda(\lambda)$ can describe a significantly broader range of scenarios. For example, large σ_λ^2 values imply that the contact frequencies between different pairs are very heterogeneous, e.g., some pairs will rarely contact each other while others much more often. An $f_\lambda(\lambda)$ symmetric around μ_λ (e.g., uniform distribution) implies a balanced number of high and low contact rates, while a right-skewed $f_\lambda(\lambda)$ (e.g., Pareto) describes a network with most pairs having large inter-contact times, but few contacting very frequently. As a comparison, we mention that the previous homogeneous models correspond to *one* function $f_\lambda^{HOM}(\lambda) = \lambda_0 = const.$, whereas the heterogeneous model we use allows an *infinite* function space for $f_\lambda(\lambda)$.

3 ANALYSIS

In this section we use the model of Section 2 to analyse the delay of epidemic spreading. We first present the basic steps of the analysis, and discuss the challenges when considering heterogeneous networks (Section 3.1). Then, we derive results for the delay prediction that become exact in the asymptotic case (Section 3.2), as well as, approximations for finite-size networks (Section 3.3). Finally, we extend our results for, the more generic case of sparse networks (Section 3.4).

3.1 Preliminaries (or, *what makes analysis difficult*)

3.1.1 Epidemic Spreading Delay

Epidemic spreading. Having a contact model, allows to analytically investigate how fast "messages" (i.e., data packets, or files)[1] can be exchanged between nodes in a network $\mathcal{N}$. During a contact event, a message currently on one of the nodes could be forwarded to ("infect") the other node as well. In the basic epidemic scheme (called *epidemic routing* in the context of DTNs [4]), a message starts from a source node, and a message transfer occurs at every contact opportunity involving a node with the message and one without it. We are usually interested in calculating the *spreading delay*, i.e. the time until either all nodes (broadcast), a percentage of them (multicast) or a specific destination node (unicast) get infected.

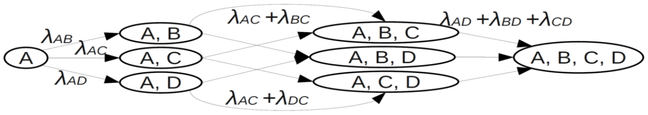


Fig. 2. Markov Chain for epidemic spreading over a *heterogeneous* network with 4 nodes

Homogeneous model. In the homogeneous case (i.e., previous models [8], [9], [20]), where all inter-contact times are i.i.d. and exponentially distributed with rate $\lambda_{ij} = \lambda$, the spreading delay is easy to derive. Specifically, one can model epidemic spreading with a pure-birth Markov chain, as depicted in Fig. 1, where a state k denotes the number of "infected" nodes (i.e., nodes with the message). In this homogeneous network, it is easy to show that the *step delay* $T_{k,k+1}$ (i.e. the time to move from state k to state $k+1$) is exponentially distributed with rate $k(N-k)\lambda$. Its expected value is then given by $E[T_{k,k+1}] = \frac{1}{k(N-k)\lambda}$, and, therefore, one could straightforwardly calculate the expected spreading delay.

Heterogeneous model. In the, more realistic, Heterogeneous Contact Network (see Def. 1), in order to calculate the expected step delay $E[T_{k,k+1}]$ at a step k (i.e. the time till one more node gets infected), we need to know *which exactly are the k infected nodes* (in contrast to the homogeneous case, where we only need to track the *number* of infected nodes). As an example, in Fig. 2, we present the Markov Chain of a message epidemic spreading in a heterogeneous network with four nodes, $\{A, B, C, D\}$. This Markov Chain is composed of 8 states (or 15 states, if we consider different source nodes), whereas the respective Markov Chain of an homogeneous network with 4 nodes would be composed of only 4 states. Hence, it becomes evident that the complexity increases quickly, even for this simple 4-node network. In a network with N nodes, the state space explodes with $2^N - 1$ total states, or, equivalently, with $\binom{N}{k}$ different states for *step k* (i.e. k infected nodes).

3.1.2 Basic Analysis: Step Delay

We present here the basic steps one needs to follow to calculate the *step delay* $T_{k,k+1}$ (and, consequently, the total epidemic

---

[1] In this paper, we consider "messages" in the context of an opportunistic network, i.e., a message denotes a data packet or a whole data file. Yet, a "message" can have a different meaning when considering different cases. For instance, a "message" could be a rumour or news in an online social network [34], a virus in a computer network [19], a disease in the physical world [35], etc.

spreading delay). As said above, the step delay $T_{k,k+1}$ is the time starting when the k^{th} node just received the message (i.e. *any* k nodes are infected) until the $(k+1)^{th}$ node receives it (i.e. *any* $k+1$ nodes are infected).

The step delay $T_{k,k+1}$ is a random variable, and the calculation of its expectation, $E[T_{k,k+1}]$, involves three sources of randomness:

- A network is initially created according to $f_\lambda(\lambda)$. In other words, $N(N-1)/2$ contact rates λ_{ij} are drawn independently from $f_\lambda(\lambda)$. The resulting graph or (symmetric) contact rate matrix $\mathbf{\Lambda} = \{\lambda_{ij}\}$ is a contact network *instance*.
- At step k, there are k nodes with the message. Hence, and conditioned on $\mathbf{\Lambda}$, $T_{k,k+1}$ is a random variable whose distribution will also depend on the actual set of k nodes that have the message, and their contact rates with the remaining nodes. Let $\mathbf{C}_k^m$ denote this set, where m is an integer indicating one of the $\binom{N}{k}$ possible sets of infected relays at step k.
- Finally, conditional on both the network instance $\mathbf{\Lambda}$ and $\mathbf{C}_k^m$, $T_{k,k+1}$ will also depend on the randomness of the inter-contact times involved.

Remark: For a clearer presentation of our analysis, in the remainder, we omit the notation $\mathbf{\Lambda}$, in the expressions that are conditional on $\mathbf{\Lambda}$. E.g., for simplicity we write $E[T_{k,k+1}]$ instead of $E[T_{k,k+1}|\mathbf{\Lambda}]$. Nevertheless, we stress here that our results hold for *any* network with a rate matrix $\mathbf{\Lambda}$.

The expected step delay $E[T_{k,k+1}]$ in a Heterogeneous Contact Network is given by the following Lemma, which we prove in Section 8.1.

Lemma 1. *The expected delay for the transition from step k to step $k+1$ is given by*

$$E\left[T_{k,k+1}\right] = \sum_{m=1}^{\binom{N}{k}} \frac{1}{S_k^m} \cdot P\{\boldsymbol{C}_k^m\} \tag{1}$$

where $P\{\boldsymbol{C}_k^m\}$ is the probability that $\boldsymbol{C}_k^m$ is the set of infected nodes at step k, and

$$S_k^m = \sum_{i \in \boldsymbol{C}_k^m} \sum_{j \notin \boldsymbol{C}_k^m} \lambda_{ij} \tag{2}$$

While keeping track in Eq. (1) of the probabilities $P\{\mathbf{C}_k^m\}$ and the rates S_k^m could be done recursively, the state space grows exponentially fast, so even numerical solutions [23] are infeasible beyond very simple problems. Instead, in the following, we choose a different approach to compute the expected delay $E[T_{k,k+1}]$: We prove that, in the limit of large N, all such starting states become statistically equivalent, and then collapse them (Section 3.2). Then we use our asymptotic results, to derive useful, simple, closed-form expressions for the spreading delay in finite-size and sparse networks (Sections 3.3 and 3.4, respectively).

3.1.3 A Note on Non-Poisson Contact Processes

Lemma 1 is derived under the assumption that the contact process of a node pair $\{i,j\}$ is given by a Poisson process with rate λ_{ij} (Assumption 1). However, it can be proved that Lemma 1 (and the following results) can be used as a good approximation also for non-Poisson cases. Here, we provide an initial discussion about the applicability of our analysis in such cases; however, a detailed investigation is out of the scope of the paper.

Specifically, let the pairwise contact processes to be *renewal processes* (i.e., a generalization of the Poisson process [36]) with *exponential tail*. This is a generic class of processes, which is also in agreement with observations in real traces [31], [32]. For this class of processes, the distribution of the inter-contact times t_{ij} between a node pair $\{i,j\}$ is given by [33]

$$P\{t_{ij} > x\} = \begin{cases} C_{ij} \cdot x^{-\alpha_{ij}} \cdot e^{-\beta_{ij} \cdot x} & for \;\; x \geq t_0^{ij} \\ 1 & for \;\; 0 < x < t_0^{ij} \end{cases}$$

where t_0^{ij} is the minimum inter-contact time for the node pair $\{i,j\}$, and C_{ij} a normalization constant. Then, following the methodology of [33], we can prove that the expected step delay $E[T_{k,k+1}]$ is approximately given by Eq. (1) with

$$S_k^m = \sum_{i \in \mathbf{C}_k^m} \sum_{j \notin \mathbf{C}_k^m} \beta_{ij}$$

In scenarios where contact processes do not exhibit an exponential tail, it is not possible to follow an analysis similar to Lemma 1. This is due to the fact that, after the first step ($k=1$), transitions between states are dependent on previous time points (which also differ between node pairs), and thus the epidemic spreading cannot be modeled with a (semi-) Markov chain. To our best knowledge, when contact processes are not Poisson or have not exponential tails, the performance of message spreading can be analysed only for (a) *single-hop* cases (i.e., the message is directly forwarded to the destination(s) from its source(s), without intermediate relay nodes) [37], or (b) *multi-hop single-copy* cases (i.e., at each time only one node has the message) [26].

3.2 Asymptotic Analysis

3.2.1 Theoretical Results

In this section, we derive our main result, Theorem 1, for the expected step delay $E\left[T_{k,k+1}\right]$ (Eq. (1)) in the limit of a large network.

To derive Theorem 1 we first need to make the following assumption.

Assumption 3 (***Full Mixing***)**.** *The contact rate distribution is defined in a positive, closed interval, i.e., $f_\lambda(\lambda), \lambda \in [\lambda_{min}, \lambda_{max}] \subseteq (0, \infty)$.*

With Assumption 3 we assume *for the moment* all contact rates to be non-zero. Later, in Section 3.4, we relax this assumption and allow some node pairs to never meet, i.e. $\lambda_{ij} = 0$.

We can now proceed and prove Lemmas 2 and 3, which we use later in the proof of Theorem 1. Let us first define the random variable S_k as

$$P\{S_k = S_k^m\} = P\{\mathbf{C}_k^m\} \tag{3}$$

and the random variable X_k as $X_k = \frac{S_k}{k(N-k)}$, i.e.

$$P\left\{X_k = \frac{S_k^m}{k(N-k)}\right\} = P\{\mathbf{C}_k^m\} \tag{4}$$

Lemma 2 gives the first two moments of the random variable S_k, and Lemma 3 shows how the random variable X_k converges as the network size N increases. The proofs of Lemmas 2 and 3 can be found in Appendix A and B, respectively.

Lemma 2. *The expectation and variance of the random variable S_k (Eq. (3)) at step k, are given by*

$$E[S_k] = k(N-k)\cdot\mu_\lambda\cdot(1-\epsilon_k)$$
$$Var\left[S_k\right] = k(N-k)\cdot\sigma_\lambda^2\cdot(1-\delta_k)$$

where $\epsilon_k = O\left(\frac{\lambda_{max}}{N}\right)$ *and* $|\delta_k| = O\left(\frac{\lambda_{max}^2}{N}\right)$.

Lemma 3. *As the network size N increases, the random variable X_k (Eq. (4)) converges as follows*

$$X_k \xrightarrow{m.s.} \mu_\lambda$$

where $\xrightarrow{m.s.}$ *denotes convergence in mean square.*

Using the above Lemmas, we prove (in Appendix C) Theorem 1, which suggests that in a large Heterogeneous Contact Network the expected step delay at a step k, can be approximated with arbitrary accuracy as follows

$$E[T_{k,k+1}] \approx \frac{1}{k(N-k)\mu_\lambda} \tag{5}$$

Theorem 1. *As the network size N increases, the relative error RE_k between the expected step delay $E[T_{k,k+1}]$ and the quantity $\frac{1}{k(N-k)\mu_\lambda}$ converges to zero*

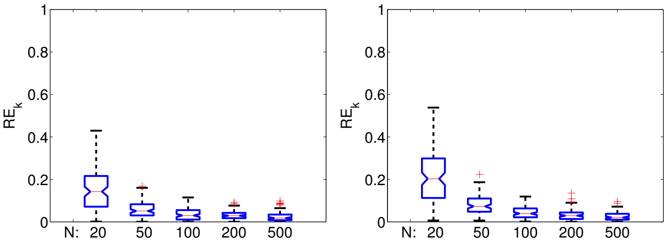
$$\lim_{N\to\infty} RE_k = \lim_{N\to\infty} \frac{E[T_{k,k+1}] - \frac{1}{k(N-k)\mu_\lambda}}{E[T_{k,k+1}]} = 0$$

3.2.2 Implications

Using the average contact rate μ_λ as a heuristic approximation (similarly to Eq. (5)), is not uncommon in related literature. For instance, in a heterogeneous network, one could calculate μ_λ, and then use a homogeneous model with $\lambda_{ij} = \mu_\lambda,\ \forall\{i,j\}$. However, depending on the contact patterns, network size, etc., this approximation can deviate considerably from the actual performance, making its applicability questionable. The contribution of the asymptotic analysis in this section is towards answering *if* it is reasonable to use this approximation and *when* it is correct. Specifically:

- Theorem 1 formally proves the asymptotic correctness of the approximation of Eq. (5) for a large class of contact networks (Heterogeneous Contact Networks). To our best knowledge, the correctness of a single parameter prediction (like, Eq. (5), or the similar expressions used by homogeneous models) has not been studied before under heterogeneous mobility assumptions.
- The analysis leading to Theorem 1 provides insights about the accuracy of the approximation Eq. (5), i.e., how close the quantity $\frac{1}{k(N-k)\mu_\lambda}$ is to the expected delay $E[T_{k,k+1}]$. As implied by Lemma 2 (see expressions for ϵ_k and δ_k), as well as, by Lemma 3 and Theorem 1 (correctness for large N):
    - (a) The approximation Eq. (5) becomes more accurate as the network size N increases.

  Moreover, the probabilistic convergence, in Lemma 3 and the proof of Theorem 1 (see Appendix C), indicates that distributions with lower heterogeneity converge faster, or, in other words:
    - (b) The approximation Eq. (5) becomes more accurate as the heterogeneity of the contact rates ($CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$) decreases.
- The discussion in Section 3.1.3, reveals in which other cases (not conforming strictly to our model) it is, or it is not, safe to use this approximation.



(a) spreading step: 20% (b) spreading step: 70%

Fig. 3. *Relative Step Error* RE_k for the step (a) $k = 0.2\cdot N$ (i.e. message spreading at 20% of the network) and (b) $k = 0.7\cdot N$. Each boxplot correspond to a different network size N (with $\mu_\lambda = 1$ and $CV_\lambda = 1.5$). In each box, the central horizontal (red) line is the median, the edges of the (blue) box are the 25^{th} and 75^{th} percentiles, the (black) whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as (red) crosses.

TABLE 1

Relative Step Error RE_k: Averaged over All Steps and over 100 Network Instances

	$N=20$	$N=50$	$N=100$	$N=200$	$N=500$
$CV_\lambda = 0.5$	4.3%	2.8%	2.7%	2.6%	2.5%
$CV_\lambda = 1$	10.6%	4.2%	3.1%	2.7%	2.6%
$CV_\lambda = 1.5$	22.4%	8.2%	4.6%	3.2%	2.6%
$CV_\lambda = 3$	126.7%	34.1%	15.3%	8.2%	3.8%

In the following section, we validate these insights with simulations, and quantify the accuracy of the results for different network sizes and contact patterns.

3.2.3 Validation

To verify our results, we test them against simulations. We use a simulator that creates network instances belonging to the class of Heterogeneous Contact Networks, we run Monte Carlo simulations of epidemic spreading, and calculate the mean step delay (further details for the simulation methodology are given in Section 5). In Table 1, we present the values for the relative error RE_k (Theorem 1) in simulation scenarios of different network sizes N and contact rates heterogeneity

CV_λ. The values in Table 1 correspond to the relative error RE_k averaged over all the steps k of the epidemic process and over 100 different network instances $\mathbf{\Lambda}$ with equivalent characteristics (N, f_λ). It can be seen that in networks with higher heterogeneity $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$ (and, thus, larger ranges $[\lambda_{min}, \lambda_{max}]$, since we set the mean rate $\mu_\lambda = 1$) the errors are larger, as our theory predicts (cf. remark "(b)" in Section 3.2.2). However, as the network size increases, the errors for all scenarios become very small (cf. remark "(a)" in Section 3.2.2).

The decrease of the relative errors can be observed also in Fig. 3, where we present the distribution (i.e., boxplots) of the values of RE_k over the different network instances. The distribution/boxplots of RE_k allows to verify that Theorem 1 holds for *every network instance* (in contrast to Table 1, where we present average values over different network instances). Also in Fig. 3, the relative errors do not correspond to averaged (over different steps) values, but we present the RE_k at the steps that correspond at the 20% (e.g. in the scenario with $N = 100$, we present the relative errors in the step $k = 20$) and 70% of the spreading process, in Fig. 3(a) and Fig. 3(b), respectively. As it can be observed, in later steps the error is slightly larger, which is expected, due to the accumulation of errors from all previous steps. Nevertheless, for large network sizes, the error diminishes for every step considered.

3.3 Finite-Size Networks

The asymptotic analysis and results of the previous section can be used to predict accurately the spreading delay in large networks. We now study *how the approximation for the step delay can be improved*, and derive simple, closed-form approximations for *finite size* networks.

Specifically, from Theorem 1, as discussed in Section 3.2.2, the quantity $\frac{1}{k(N-k)\mu_\lambda}$ can be used as a predictor for the step delay, and the prediction error converges to 0 as networks get larger. For finite cases though, this error might not be negligible. This motivates us to investigate how to improve the approximation for the step delay.

To this end, we derive in Result 1 a *second order* approximation for the expected step delay.

Result 1. *In a Heterogeneous Contact Network the expected step delay can be approximated by*

$$E[T_{k,k+1}] = \frac{1}{k(N-k)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{k(N-k)}\right) \qquad (6)$$

In Section 8.2, we provide the methodology we use to prove Result 1, as well as, an extensive discussion and useful insights relating to its accuracy.

More accurate approximations than Result 1 can be derived by taking into account higher (than the 1^{st} and 2^{nd}) moments of the contact rates distribution $f_\lambda(\lambda)$; more details are given in Section 8.2.

3.4 Sparse Networks

To derive the results of the previous sections, we have so far required *all* pairs of nodes to meet with non-zero rate, i.e. $\lambda_{ij} > 0$ (*Full Mixing* - Assumption 3). Even though under this assumption we can capture the characteristics of many contact environments, scenarios where some pairs of nodes *never contact* cannot be adequately described. This is an undesirable restriction. Intuition as well as studies of real traces suggest that many pairs of nodes in fact never meet. To this end, we now relax this assumption and extend our results to a special case of sparse contact networks, described as:

Assumption 4 (***Poisson Mixing***)**.** *For each pair of nodes i and j the following holds: (i) with probability p_s they regularly contact with rate λ_{ij}, according to Assumptions 1 and 2; (ii) with probability $1 - p_s$ they never contact each other, i.e. $\lambda_{ij} = 0$.*

In other words, we now first create a *Poisson (or Erdös-Renyi) graph* [20] between nodes. We then assign rates λ_{ij}, as before, but only to the existing links. With the parameter p_s, we can now capture arbitrarily sparse scenarios, where each node meets only a percentage of all nodes[2].

The following corollary suggests that the previous analysis is valid also in the case of a sparse network (Assumption 4) and the results hold by just modifying the values of μ_λ and σ_λ^2.

Corollary 1. *Under a Heterogeneous Poisson Mixing Contact Network (Assumption 4), the theoretical results for a Heterogeneous Full Mixing Contact Network (Assumption 3), are modified by substituting the moments of the contact rate distribution (μ_λ and σ_λ^2) with the expressions*

$$\mu_{\lambda(p)} = p_s \cdot \mu_\lambda \qquad (7)$$

$$\sigma_{\lambda(p)}^2 = p_s \cdot \left[\sigma_\lambda^2 + \mu_\lambda^2(1 - p_s)\right] \qquad (8)$$

Corollary 1 can be proved similarly to the theoretical results of Section 3.2. In Appendix D, we present a sketch of this proof comprising the main analytical arguments and differences compared to the analysis for the full-meshed network case.

A further utility of Corollary 1 is that it provides the guidelines for analysing additional properties of *Heterogeneous Contact Networks*. We demonstrate such an example in Section 4.2.

Remark: More complex models than the Poisson graph can be used to capture in a more accurate way the structure of a network. However, in these cases the complexity of the analysis increases as well. We refer the interested reader to [38] for such an example, where graphs with *arbitrary degree distributions* are considered, i.e. some nodes (e.g. because they are more mobile or social) contact regularly many other nodes, while some nodes *ever* contact with only a few of other nodes.

4 APPLICATIONS

In this section we turn our attention to some applications of our model and analysis for opportunistic networking. We first

---

[2] We *do* assume that the probability p_s is large enough for connectivity to be achieved. In practice, the theory of Poisson graphs tells us that connectivity can be achieved with arbitrary low p_s as long as N is large enough (with percolation occurring at an average degree as low as 1 in the limit) [20].

use our results for the individual epidemic steps to derive expressions for the delivery delay of three epidemic-based routing protocols (Section 4.1). As a second application, we demonstrate how our model can be used to capture further heterogeneous characteristics (other than mobility), namely the node selfishness, in opportunistic networking scenarios and how our analysis can be extended for these cases (Section 4.2).

The analytical methods we provide in the remainder of the section can be used for the analysis of the performance of other communication mechanisms in a range of scenarios (e.g., content-delivery schemes for mobile networks, information spreading mechanisms for Online Social Networks, etc.) [39], as well as for the investigation of the effects of heterogeneous social characteristics in opportunistic networks [18], [40].

4.1 Delivery Delay of Routing Protocols

We first use the basic building blocks of our analysis to predict the delivery ("end-to-end") delay for *epidemic* routing [4], *2-hop* routing [9] and *Spray and Wait* routing [7].

In the following, we briefly present the mechanism of these schemes:

Epidemic routing: In *epidemic routing*, the source node forwards the message to every node it meets, rendering this node a relay for the message. Similarly, each relay forwards the message to every "uninfected" node it meets. The first node (source or relay) to meet the destination delivers the message.

2-hop routing: In the 2-hop routing scheme, the source sends the message to every node it meets, like in epidemic routing. However, other nodes receiving the message can only give it directly to the destination, when and if they encounter it.

Spray and Wait (SnW) routing: In the Spray and Wait scheme[3], the source generates L copies of the message and when it meets another node, it gives to it half of the messages it holds at that time (if it holds more than one). The same mechanism applies when a relay node with more than one copies meets another node without the message. Eventually there would be L nodes (including the source) holding the message. If the message is not delivered to the destination before the L message copies are spread (*spray phase*), it will be delivered the first time any of the L nodes with the message meets the destination (*wait phase*).

The analysis for the delivery delay for each of the above protocols comprises the following steps:

(i) The *protocol's step delay* $E[T_{k,k+1}^{(protocol)}]$ is calculated by taking into account the number of possible infections and the delay till the next infection, which we compute based on Result 1 [4].

(ii) The delivery delay is calculated by summing the protocol's step delays till the message delivery, given that the delivery takes place at step k

$$E[T^{(protocol)}|k] = \sum_{i=1}^{k-1} E[T_{i,i+1}^{(protocol)}]$$

(iii) The probability of delivering the message at a step k is calculated, $P(k)$.

(iv) The expected delivery delay is given by the average

$$E[T^{(protocol)}|k] = \sum_{k} E[T^{(protocol)}|k] \cdot P(k)$$

Table 2 gives the approximate closed-form expressions for the delivery delay (corresponding to the approximation of Result 1 for the step delay). The detailed derivations of the formulas can be found in [41].

4.2 Beyond Mobility: Social Selfishness

In this section, we predict the performance in networks where some nodes are not willing to relay messages of other nodes. In particular, we consider the generic case where this unwillingness (or *selfishness*) is heterogeneous among different node pairs and related to their *social* characteristics. We show that even in this complex case (where both mobility and social heterogeneity are considered), our model/analysis can be used to derive simple predictions that require only one or two (1^{st} or 2^{nd} order approximation, respectively) parameters.

In our analysis, we have assumed that all nodes can exchange messages when they contact. However, due to limited resources (battery, bandwidth etc.) and/or privacy concerns, some nodes might be reluctant to cooperate and forward messages of other nodes. Although incentive mechanisms [42], [43] can be applied, resource and privacy constraints might still exist and, therefore, nodes might forward messages only from/to some of the other nodes.

This node selfishness is not expected to be always homogeneous. Specifically, since nodes in an opportunistic network are mobile devices handled by people, it is more probable that two nodes will cooperate if they have a strong social tie (e.g. friends or acquaintances), because they trust each other and/or have cooperated in the past [40], [44]. Additionally, studies from sociology [45], social media [46] and pervasive social networks [21] have shown that the strength of social ties can be predicted by the contact rate and, specifically, the more frequent two nodes contact, the stronger their social tie probably is.

Hence, combining the previous observations, i.e. (a) the level of cooperation is related to social ties and (b) the strength of the ties is related to the contact rate, we can model the node selfishness in a probabilistic way, which allows us to predict its effect on the network performance.

Definition 2 (Social Selfishness)**.** *A pair of nodes i and j either can exchange messages in every contact event with probability p_{ij} or can never exchange messages with probability $1 - p_{ij}$. The probability p_{ij} depends on the meeting rate between these nodes, i.e. λ_{ij}, and is described by the relation:*

$$p_{ij} = p(\lambda_{ij}), \qquad p_{ij} \in [0, 1] \tag{9}$$

---

[3] Here we describe the *binary* Spray and Wait, which is the scheme with the lowest expected delivery delay among all the SnW-based schemes (e.g. *source SnW*) [7].

[4] For instance, at step k, in epidemic routing there are $k(N-k)$ possible infections, while in 2-hop routing the number of the possible infections are $N-1$. Hence, this gives a *protocol's step delay* equal to $E[T_{k,k+1}^{(epidemic)}] = \frac{1}{k(N-k)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{k(N-k)}\right)$ and $E[T_{k,k+1}^{(2-hop)}] = \frac{1}{(N-1)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{(N-1)}\right)$ for *epidemic* and *2-hop* routing, respectively.

TABLE 2
Approximative expressions for the Expected Delivery Delay of different routing protocols.

Epidemic	$E[T_D^{(epid)}] \approx \frac{1}{N \cdot \mu_\lambda} \cdot \left(\ln(N) + CV_\lambda^2 \cdot \frac{1.65 \cdot N + 2 \cdot \ln(N)}{N^2} \right)$
2-hop	$E[T_D^{(2-hop)}] = A_{N-1} \cdot \sum_{k=1}^{N-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} \approx \frac{\sqrt{\frac{\pi}{2}}}{\sqrt{N} \cdot \mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N} \right)$
SnW, L copies	$E[T_D^{(SnW)}] \leq A_{N-1} \cdot \sum_{k=1}^{L-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} + (L \cdot A_{N-1} + A_L) \cdot \frac{(N-1)!}{(N-1)^L \cdot (N-L-1)!}$
	where $A_m = \frac{1}{m\mu_\lambda} \cdot \left[1 + \frac{CV_\lambda^2}{m} \right]$

The above model is very generic, since any valid function $p(\cdot)$ can be selected to describe a given setting. It can also capture scenarios where node cooperation is determined by a protocol (rather than social factors), as in the case of [47] where decisions for message exchanges depend on nodes' contact rates.

Under the social selfishness model of Def. 2, we can still use our basic analysis for the message delivery delay. As is proved in [48] (by applying similar arguments to the Poisson graph case), we just need to substitute μ_λ and σ_λ^2 in the expressions of our results with their *effective* values given in the following result:

Result 2. *The mean value, $\mu_{\lambda-eff}$, and the variance, $\sigma_{\lambda-eff}^2$, of the effective contact rates in a Heterogeneous Contact Network with contact rate probability function f_λ (μ_λ, σ_λ^2) and Social Selfishness (Def. 2), are given by*

$$\mu_{\lambda-eff} = E[\lambda \cdot p(\lambda)] \tag{10}$$

$$\sigma_{\lambda-eff}^2 = E[\lambda^2 \cdot p(\lambda)] - (E[\lambda \cdot p(\lambda)])^2 \tag{11}$$

where the expectations are taken over the p.d.f. f_λ.

Remark: The case where the selfishness is homogeneous (independent of the contact rate), i.e. $p(\lambda) = p_0$, is equivalent to a Heterogeneous Poisson Mixing Network with $p_s = p_0$.

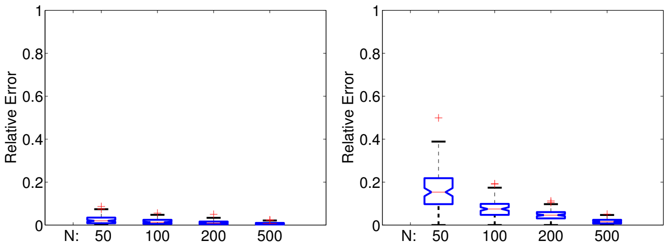
The effects of social selfishness in opportunistic networks are extensively investigated in [48]. Hence, and due to space limitations, in the model validation section (Section 5) we provide a few, indicative results for social selfishness scenarios, and we refer the interested reader to [48] for further results and applications.

Finally, we would like to note that Result 2 further highlights the contributions of the proposed mobility model and analysis. As we showed, using our framework (i) it becomes possible to analyse the effects of various heterogeneous characteristics (not restricted to mobility), and (ii) derive simple, closed form expressions even for very complex scenarios. In [18], [37], we have similarly used this model, to incorporate other heterogeneous characteristics observed in opportunistic networks, and derive simple performance predictions.

5 MODEL VALIDATION

5.1 Synthetic Simulations

In order to validate the accuracy of our predictions for the message delivery delay under different routing schemes, we first compare them against simulations of various *synthetic* scenarios belonging to the class of Heterogeneous Contact Networks. We use Monte Carlo simulations to examine the accuracy of our various analytical expressions (i) in finite size networks and (ii) as a function of other parameters of interest (e.g. statistics of the contact rates generating function f_λ).



(a) $\mu_\lambda = 1, CV_\lambda = 1$ (b) $\mu_\lambda = 1, CV_\lambda = 3$

Fig. 4. Relative Error $\frac{|ET_{sims} - ET_{theory}|}{ET_{theory}}$ between the *simulated expected delivery delay of epidemic routing* and the *theoretical approximation*. Each boxplot corresponds to a different network size N. Box-plots show the distribution of the Relative Error for 100 different network instances of the same size.

In each simulation, we create a network of N nodes and a contact pattern by generating a $N \times N$ matrix $\mathbf{\Lambda} = \{\lambda_{ij}\}$. Each entry λ_{ij} characterizes the contact process of the pair of nodes *i* and *j*: it is zero with probability $1 - p_s$ (in Poisson mixing), otherwise it takes values drawn from a chosen distribution f_λ with mean μ_λ and variance σ_λ^2 ($CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$). Then for the pairs that meet ($\lambda_{ij} > 0$) we generate a sequence of contact events with exponentially distributed intercontact times with rate $\lambda_{ij} > 0$.

For every network instance $\mathbf{\Lambda}$, we run 1000 message spreading simulations, choosing randomly the source and destination nodes, and calculate the average delivery delay. We have considered scenarios with contact rate distributions f_λ with varying heterogeneity (CV_λ). Without loss of generality and for a clearer comparison we set the average contact rate equal to the unit, i.e. $\mu_\lambda = 1$.

5.1.1 Prediction accuracy vs. network instances

We first investigate the accuracy of our theoretical results in different network instances (i.e., with different contact matrices $\mathbf{\Lambda}$) of the same type (i.e., same network size N, and contact rate distribution $f_\lambda(\lambda)$).

Fig. 4 shows the distribution of the relative error between theoretical and simulation results for the expected delay of epidemic routing: $\frac{|ET_{sims} - ET_{theory}|}{ET_{theory}}$, where ET_{sims} is the average delay observed in simulations in a network, and ET_{theory} is the corresponding theoretical prediction given in Table 2. We consider two set of scenarios with different contact rates heterogeneity CV_λ. Presenting the results in box-plots, allows to observe the accuracy of the approximation over the different network instances (i.e., different $\mathbf{\Lambda}$) of the same size. It can be seen that the accuracy of our prediction is significant, even for small networks and *for every network instance*, when the heterogeneity is not high (Fig. 4(a)). Although the accuracy decreases with heterogeneity (Fig. 4(b)), for networks larger than a hundred nodes, the relative errors are less than 10% for the majority of ($N = 100$ and $N = 200$) or for all ($N = 500$) the network instances.

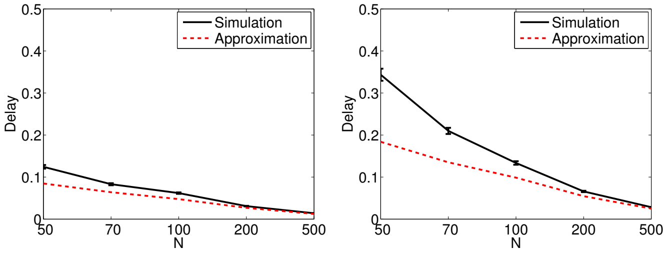
These results verify the insights stemming from our analysis (see, e.g., Section 3.2.2), and quantify the accuracy of our predictions. They indicate that our approximation can be safely used in networks of large size and/or moderate heterogeneity (since they are accurate for every network instance). In more challenging cases, i.e., small and very heterogeneous networks, they are still accurate on average, however, there might exist a small percentage of network instances where the error can be more than 10%.

5.1.2 Prediction accuracy vs. network parameters

We now test our results under various network parameters (N and CV_λ). In each case, we compare our predictions against the average simulation results values over 100 different network instances.

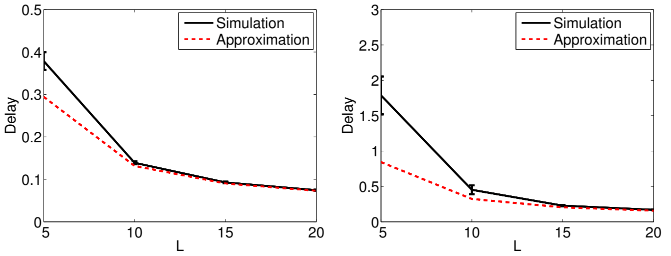
In Fig. 5 we present the simulation results (continuous lines; with confidence intervals) and theoretical predictions (dash lines) for the expected delivery delay of epidemic routing. We simulated networks of different sizes (from very small $N = 50$ to large $N = 500$) and *highly heterogeneous* contact patterns, i.e., $CV_\lambda = 3$. As shown, our predictions are close to the simulated results in most scenarios, and become more accurate as the network size increases. Moreover, comparing Fig. 5(a) and Fig. 5(b), we can see that in the Poisson-mixing case a certain accuracy can be achieved in larger networks compared to the Full-mixing case. This is due to the fact that introducing sparseness in a network, increases the heterogeneity of the contact patterns. Therefore, the theoretical approximation becomes more accurate for larger networks (see also the commentary related to the convergence of the approximation, in Section 3.2.2). Finally, we would like to stress here that in the presented scenarios the contact patterns are very heterogeneous[5]; in scenarios with low/moderate heterogeneity, our predictions are more accurate (i.e., Fig. 5 could be considered a "worst-case" for our predictions).

Fig. 6 shows simulation results for the case of Spray and Wait routing, for scenarios with different number of copies L, in networks with $N = 500$ nodes and $CV_\lambda = 1.5$. In all scenarios, simulation results are averaged over 100 network instances. As it can be seen, the predictions of our approximate expressions become more accurate as the number of copies L increases. This is because the effect of the number of copies L on the accuracy is similar to this of the network size N in the epidemic routing case (i.e., the convergence of the approximation is achieved earlier; see also the derivation of the expressions of Table 2 in Section 4.1 and [41]).



(a) Full-mixing (b) Poisson-mixing

Fig. 5. Delivery Delay of *epidemic routing* in (a) Full-mixing and (b) Poisson-mixing ($p_s = 0.5$) networks with highly heterogeneous contact patterns $CV_\lambda = 3$.



(a) Full-mixing (b) Poisson-mixing

Fig. 6. Delivery Delay of *Spray and Wait* routing in (a) Full-mixing and (b) Poisson-mixing ($p_s = 0.5$, $CV_\lambda = 0.5$) networks of size $N = 500$ and contact rates heterogeneity $CV_\lambda = 1.5$.

5.1.3 Homogeneous vs. Heterogeneous model

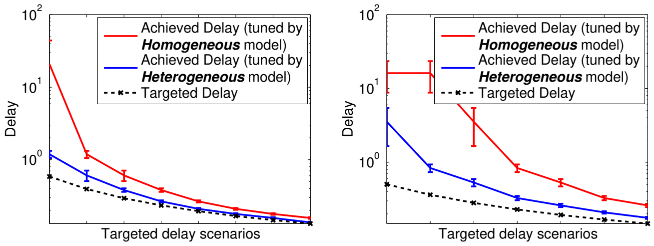
In the theoretical predictions we presented in the previous sections, we use the *second order* approximation (see Result 1 and Table 2), which takes into account the heterogeneity of contact rates, i.e., CV_λ. However, as discussed earlier, one could omit the heterogeneity (a common approach in literature), and use a homogeneous model with $\lambda_{ij} = \mu_\lambda$, $\forall\{i,j\}$ (see, e.g., Eq. (5)). Hence, a question that naturally follows, is: *what is the added value of using a heterogeneous model?*

Towards answering the above question, in this section, we compare the performance predictions under the heterogeneous (i.e. our approach) and homogeneous (i.e., previous approaches) models. Specifically, we consider two application cases, and demonstrate how considering the heterogeneity of the contact rates CV_λ (i.e. the second order approximation) can benefit the design and evaluation of routing protocols for opportunistic networking.

Tuning the SnW protocol. The SnW protocol [7] can

---

[5] As a comparison, in the mobility traces from real opportunistic networks we studied (see Section 5.2), we observed a much lower heterogeneity of the (non-zero) contact rates distribution, i.e., $CV_\lambda < 2$.



(a) $CV_\lambda = 1.5$ (b) $CV_\lambda = 2$

Fig. 7. Simulation results in scenarios where the copy selection of SnW, L_{target}, is based on the heterogeneous and homogeneous models. Networks with (a) moderate, $CV_\lambda = 1.5$, and (b) high, $CV_\lambda = 2$, contact rate heterogeneity.

deliver messages, without flooding the network and draining nodes' resources. Selecting the number of copies L affects the delivery delay: the more copies are used, the faster (on average) a message is delivered. Here, we consider a scenario, where the goal is to achieve a certain average delivery delay, or a *targeted delay*. The expression $E[T_D^{(SnW)}]$ of Table 2 gives the expected delay of SnW as a function of L. Hence, to achieve a targeted delay ET_{target}, one can use our expression and choose the lowest number of copies, under which SnW can achieve this ET_{target}, i.e,

$$L_{target} = min\left\{L : E[T_D^{(SnW)}] \leq ET_{target}, L \in [1, N-1]\right\}$$

In the case that a homogeneous model is used for predicting the performance, one could use a similar expression to this of Table 2 with $A_m = \frac{1}{m \cdot \mu_\lambda}$ (i.e., the term containing CV_λ is omitted). Since the predictions $E[T_D^{(SnW)}]$ derived by a heterogeneous and a homogeneous model differ, the selections for the number of copies L_{target} can possibly differ as well.

In Fig. 7, we present the simulation results of the copy selection process under the two models. The dashed line is the targeted delay ET_{target}, and the continuous lines denote the delay that was achieved by tuning the SnW protocol, i.e., selecting the L_{target}, with the two models. Some important observations are: (a) When tuning SnW using the predictions of the heterogeneous model, the achieved delay is closer to the targeted delay, than in the homogeneous model case. (b) As expected, in more heterogeneous scenarios (i.e., Fig. 7(b)) the difference between the achieved and targeted delays is larger (the accuracy of the approximations decreases). However, using the heterogeneous model gives results that are still relatively close to the targeted delay, whereas in the homogeneous model case the results are much farther (note the logarithmic scale of the y-axis). (c) It becomes more crucial to use a heterogeneous model for large values of ET_{target}. This is because the number of required copies L_{target} decreases, and considering the second order term in the delay approximation significantly improves the predictions.

Social Selfishness. As discussed in Section 4.2, in some scenarios nodes do not always forward messages of other nodes, and this behavior can be related to their social/mobility characteristics. We consider here a scenario, where contact rates are drawn from a uniform distribution[6]

$$f_\lambda(\lambda) = \frac{1}{\lambda_{max} - \lambda_{min}}, \quad \lambda \in [\lambda_{min}, \lambda_{max}]$$

and selfishness of nodes is characterised by the function

$$p(\lambda) = \frac{c}{\lambda}, \quad c = \lambda_{min}$$

In this scenario, nodes who contact less frequently (small λ) have a higher probability to cooperate (large $p(\lambda)$). The intuition behind this selfishness policy is to avoid consuming too much resources by continuously exchanging messages with frequently contacted nodes; more detailed discussions about this type of selfishness can be found in [47], [48].

Using the expressions of Result 2, we can calculate

$$\mu_{\lambda-eff} = c$$
$$\sigma^2_{\lambda-eff} = c \cdot \mu_\lambda - c^2$$

and use them, in our expressions of Table 2 to predict the performance of routing protocols. Similarly to the previous application case, we compare here the predictions that can be made using our heterogeneous model and a homogeneous model (with $\lambda_{ij} = \mu_{\lambda-eff}, \ \forall i, j$), and compare them with simulation results in Fig. 8.

Specifically, in Fig. 8(a) we present scenarios where messages are delivered using SnW routing with different number of copies L. As it can be seen, our predictions are always accurate, while the predictions of the homogeneous model are less accurate and their accuracy decreases for lower number of copies L (for the same reasons as before: less copies, higher importance of the heterogeneity). In Fig. 8(b) we present scenarios with varying heterogeneity: we keep λ_{min} equal to 1 and vary the values of λ_{max}; higher values of λ_{max} denote a network with more heterogeneity. Messages are delivered under the SnW protocol with $L = 10$ copies. This plot demonstrates clearly the added value of the heterogeneous model approach. When heterogeneity is low, both models lead to similar, and accurate, predictions. As the heterogeneity increases, the prediction of the homogeneous model becomes less accurate (up to 35% error for $\lambda_{max} = 9$), whereas the accuracy of the heterogeneous model predictions remains the same ($\sim 1\%$).

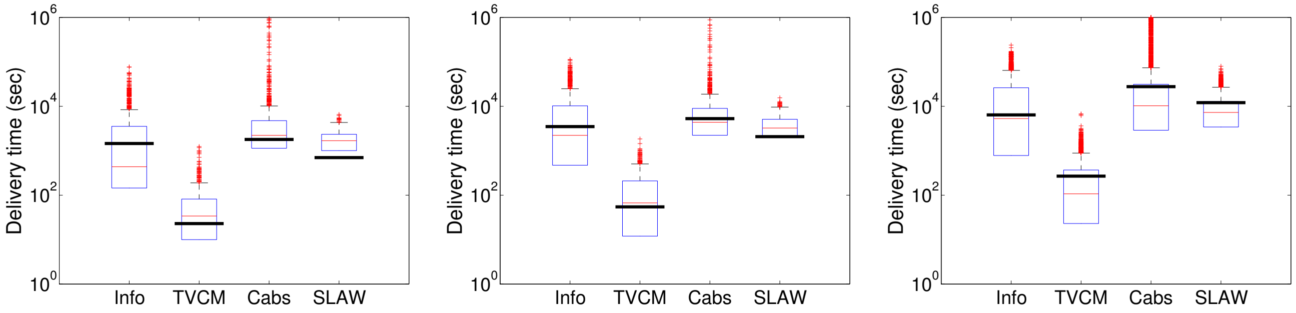
The above results are indicative of the usefulness of the proposed model in scenarios with heterogeneous selfishness. Performance predictions can be significantly improved, and this can help towards a more careful, or even optimal, protocol design, as is discussed in more detail in [48].

5.2 Real Mobility Traces

The simulation results presented in Section 5.1 show that our analytical predictions achieve significant accuracy even in finite-size networks belonging to the Heterogeneous Contact Networks (Def. 1) mobility class. While these contact classes are rather broad, whether they capture "real" scenarios, and to

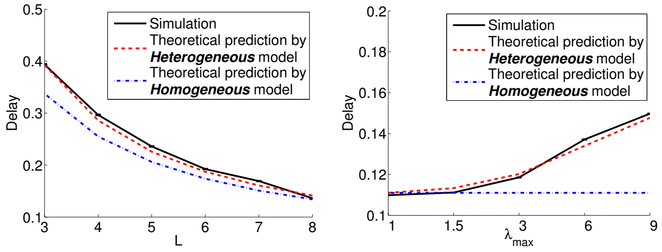
---

[6] Note that in this set of simulations (in contrast to all previous scenarios) the average contact rate is *not* always equal to 1: $\mu_\lambda = \frac{\lambda_{min} + \lambda_{max}}{2}$.



(a) epidemic routing (b) 2-hop routing (c) SnW routing

Fig. 9. Box-plots of the message delivery delay under (a) epidemic, (b) 2-hop routing, and (c) SnW (with $L = 6$ copies) routing. On each box, the central horizontal line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as crosses. The thick (black) horizontal lines represent the theoretical values predicted by our model.



(a) Delay vs. L (b) Delay vs. heterogeneity

Fig. 8. Message delivery delay of SnW routing in social selfishness scenarios with (a) varying number of copies L, and (b) heterogeneity of contact rates.

what extent, depends on the application setting, contact scenario, etc. To this end, in this section, we evaluate our analytical results against simulations on scenarios with real/realistic mobility patterns.

In the context of opportunistic networking, some mobility traces collected in real experiments and/or networks do exist. Arguably, the size of most of them is small and they represent each only a single instance of the random mobility process at play, often with a number of measurement complications and errors. Nevertheless, it is of interest to see how our performance predictors behave in some of these scenarios, and whether they can capture the quantities of interest (even if qualitatively), despite the considerably higher complexity (e.g. community structure) of such scenarios, and departures from the assumptions for which our predictors are designed.

To this end, we use the following sets of real mobility traces:

- *Cabspotting* [49], which contains GPS coordinates from 536 taxi cabs collected over 30 days in San Francisco.
- *Infocom* [50], which contains traces of Bluetooth sightings of 78 mobile nodes from the 4 days iMotes experiment during Infocom 2006.

We also generated mobility traces with two recent mobility models that have been shown to capture well different aspects of real mobility traces,

- TVCM [51], scenario composed of 4 equal communities of 26 nodes each; nodes mainly move within their community and have a small probability of going outside it.
- SLAW [52], scenario with 100 nodes, $30m$ transmission range; the other parameters are set as the default from [52].

In order to compare with analysis, we parse each trace and estimate the mean contact rate for all pairs $\{i, j\}$. We then produce estimates for the 1^{st} and 2^{nd} moments of these rates, $\hat{\mu}_\lambda$ and $\hat{\sigma}^2_\lambda$, and use them in our analytical expressions.

Fig. 9 shows the message delay under epidemic, 2-hop routing, and SnW routing. Source and destination are chosen randomly in different runs and messages are generated in random points of the trace.

The first thing to observe is that delay values span a wide range of values for different source-destination pairs. This implies a large amount of heterogeneity in the “reachability” of different nodes. Our analytical predictions are shown as thick dark horizontal lines. As it can be seen, our result is in most cases close to the median and in almost all cases between the 25^{th} and 75^{th} percentile of the delay observed in both the real traces and mobility models.

It is somewhat remarkable that our delay predictors are close to the actual results (qualitatively or even quantitatively in some cases) in a range of real or realistic scenarios; studies of these scenarios reveal considerable differences to the much simpler contact classes for which our results are derived. We should also be careful not to jump to generalizations about the accuracy of these results in all real scenarios, as we are aware of situations that could force our predictors to err significantly. Nevertheless, we believe these results are quite promising in the direction of finding simple, usable analytical expressions even for complex, heterogeneous contact scenarios.

6 RELATED WORK

Models for epidemic spreading of deceases [20] and/or computer malware [19], were early derived, based on the well known SIR model, and studied widely. In DTNs, efforts to analyze the performance of epidemic routing and other protocols also abound. Stochastic analyses, like the one in [9], define a Markov chain as in Fig. 1, in order to give closed form results for epidemic and 2-hop routing. Fluid models [8], [10], [53], take an approach similar to the SIR model in biology, and define the number of messages in the network as a continuous function (of time). Then, ordinary differential equations (ODEs) are used to derive expressions for the total delay, delivery probability etc. While these models provide closed form results and thus can be used in tuning protocol parameters (e.g. gossiping probability [8], number of copies [7], TTL [54], they all assume a homogeneous network with a common meeting rate for every pair of nodes.

Recent studies on real network traces [11], [12], [13], [14] suggest that the homogeneity assumption is not true. To overcome this limitation, a number of works has introduced heterogeneity in contact network models, by allowing different meeting rates for each node pair [12], [14], [15], [16], [22], [26]. Yet, most of these works use the heterogeneous model to design new, better protocols (e.g. multicast [12] or unicast [14]) that take heterogeneity into account, but do not analyze performance. One exception is [16], but only for the cases of direct transmission and 2-hop routing. To our best knowledge, the work closer to this paper is that of [22], where a very generic contact graph is considered. However, due to the large generality of the contact model, only upper bounds for the delay can be provided.

In our work, while we allow arbitrary link rates between nodes, as in [12], [16], [22], we restrict the underlying contact graph model, in order to derive closed-form expressions. We validated our results with synthetic simulations for the targeted contact classes, as previous work did [7], [8], [9], but also demonstrate their applicability in real networks.

As a final note, in theoretical biology and epidemiology, there are many studies trying to capture heterogeneity and model it in the context of complex networks [55], [56], using different levels of mixing [57] or stratified populations [58]. Yet, the majority of these works focus on deriving thresholds above which the epidemic will spread and their results usually consider infinite time.

7 CONCLUSIONS

In this paper, we have considered classes of heterogeneous contact models that can be used to describe opportunistic networking scenarios or even scenarios of different contexts, where nodes interact with each other (e.g. P2P networks, OSNs, etc.). We derived both asymptotic results and simple closed form approximations for epidemic spreading. From the validation of the model against synthetic models and realistic traces we can conclude that: (a) simple delay expressions, that can be used for performance prediction and protocol optimization, exist not only for the homogeneous contact case; (b) performance predictions that are accurate qualitatively, and (somewhat more surprisingly) sometimes quantitatively also, can be made even for a number of real scenarios, despite the highly more complex structure of the latter. Our methodology can be applicable in more generic contact graph structures as well (e.g. as we did in [38] for configuration model [20] contact graphs). However, we think that a limit exists, probably related to the size of the min-cut of the contact graph, beyond which are techniques are not applicable and other methods are needed (e.g. [22]).

8 PROOFS OF THEORETICAL RESULTS

8.1 Proof of Lemma 1

Proof. At a step k, let i and j be two nodes, where $i \in \mathbf{C}_k^m$ and $j \notin \mathbf{C}_k^m$. Let us denote as t_{ij} the first time (after the k^{th} node received the message) that nodes i and j contact each other. As the next message exchange will take place when any of the nodes with the message contacts any of the nodes without it, the step delay is given by $T_{k,k+1} = \min_{i\in\mathbf{C}_k^m\, j\notin\mathbf{C}_k^m}\{t_{ij}\}$. Moreover, since t_{ij} are independent, exponentially distributed random variables with rate λ_{ij}, $T_{k,k+1}$ is also exponentially distributed with rate $\sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}$:

$$t_{ij} \sim exp(\lambda_{ij}) \quad \Rightarrow \quad T_{k,k+1} \sim exp\left(\sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}\right) \tag{12}$$

and, thus [36]

$$E\left[T_{k,k+1}|\mathbf{C}_k^m\right] = \frac{1}{\sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}} \tag{13}$$

Using the properties of conditional expectation, we get the expected delay for the transition from step k to step $k+1$:

$$\begin{aligned} E\left[T_{k,k+1}\right] &= \sum_{m=1}^{\binom{N}{k}} E\left[T_{k,k+1}|\mathbf{C}_k^m\right]\cdot P\{\mathbf{C}_k^m\} \\ &= \sum_{m=1}^{\binom{N}{k}} \frac{1}{\sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}}\cdot P\{\mathbf{C}_k^m\} \end{aligned} \tag{14}$$

Denoting $S_k^m = \sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}$ proves the Lemma. ☐

8.2 Proof or Result 1

From Eq. (1) we can express $E[T_{k,k+1}]$ as

$$E\left[T_{k,k+1}\right] = \sum_{m=1}^{\binom{N}{k}} \frac{1}{S_k^m}\cdot P\{\mathbf{C}_k^m\} = E\left[\frac{1}{S_k}\right] \tag{15}$$

where S_k is defined in Eq. (3). Since we do not know the probabilities $P\{\mathbf{C}_k^m\}$ (i.e. the exact distribution of S_k), it is not possible to calculate the quantity $E\left[\frac{1}{S_k}\right]$. However, $E\left[\frac{1}{S_k}\right]$ is the expectation of a function of S_k (i.e. the function $g(x) = x^{-1}$), and thus we can approximate it by using the *Delta method* [59], where the expectation of a function of a random variable (i.e. $E[g(S_k)] \equiv E\left[\frac{1}{S_k}\right]$) is approximated using the Taylor expansion of the function and the first moments of the random variable (i.e. $E[S_k]$, $Var\left[S_k\right]$, etc.).

The calculation of these moments though, still depends on the knowledge of the probabilities $P\{\mathbf{C}_k^m\}$, and exact expressions cannot be found. Hence, to proceed further and be able to derive useful results, we approximate the first two central moments of S_k, by neglecting the terms ϵ_k and δ_k in the expressions of Lemma 2, i.e.

$$E[S_k] \approx k(N-k) \cdot \mu_\lambda \tag{16}$$

$$Var[S_k] \approx k(N-k) \cdot \sigma_\lambda^2 \tag{17}$$

These approximations, as Lemma 2 implies, become more accurate as (a) the size of the network N increases, or (b) the heterogeneity of the contact rates $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$ decreases (see also, Section 3.2.2). To further support this argument, we present some initial simulation results. Table 3 and Fig. 10 (in a similar way to Section 3.2.3) show the relative errors between the quantity $E[X_k]$ and the approximation we consider, $k(N-k)\mu_\lambda$. As it can be seen, the approximation is relatively accurate even for moderate network sizes.

Now, using the *Delta method* and the expressions of Eq. (16) and Eq. (17), we derive a *second order* approximation for the expected step delay: To compute $E\left[\frac{1}{S_k}\right] = E[g(S_k)]$, at first we express the function $g(S_k) = \frac{1}{S_k}$ as a Taylor series expansion, centered at $E[S_k]$, the mean value of S_k.

$$\begin{aligned} T_g(S_k) &= \sum_{n=0}^{\infty} \frac{g^{(n)}(E[S_k])}{n!}(S_k - E[S_k])^n \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n (S_k - E[S_k])^n}{(E[S_k])^{n+1}} \end{aligned} \tag{18}$$

We can approximate $g(S_k)$ by taking the first m terms of the Taylor series. That will result in:

$$g(S_k) \quad \approx \quad \sum_{n=0}^{m} \frac{(-1)^n}{(E[S_k])^{n+1}}(S_k - E[S_k])^n \tag{19}$$

An approximation for the mean value of $g(S_k)$ follows after taking the expectation of both sides in the last equation.

$$E[g(S_k)] \approx \sum_{n=0}^{m} \frac{(-1)^n}{(E[S_k])^{n+1}} M_n \tag{20}$$

where $M_n = E[(S_k - E[S_k])^n]$ is the n^{th} central moment.

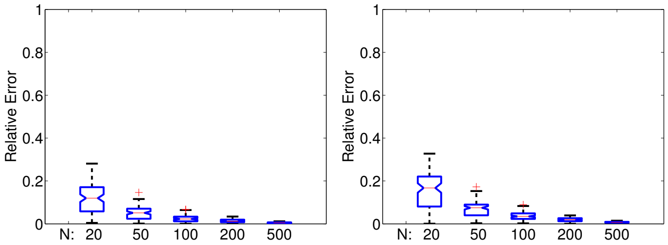
Considering $m = 2$ in Eq. (20) and using the expressions of Eq. (16) and Eq. (17) for the moments $M_0 = E[S_k]$ and $M_2 = Var[S_k]$, proves Result 1.

Remark: In the Delta method, different number of terms of the Taylor series can be taken into account, depending on the required accuracy (the more terms one considers, the more accurate the result). For example, taking only the first term ($m = 0$), we get the asymptotic expression, i.e. $E[T_{k,k+1}] = \frac{1}{k(N-k)\mu_\lambda}$. As a better approximation, we consider here the first three terms ($m = 2$) of the Taylor series, which involve the first two moments of S_k. Our choice for using the approximation that depends on the first two moments is a trade off between usability and expressibility of the result, and its accuracy.

TABLE 3

Relative error $\frac{E[X_k]-k(N-k)\mu_\lambda}{k(N-k)\mu_\lambda}$: Averaged over All Steps and over 100 Network Instances

	$N = 20$	$N = 50$	$N = 100$	$N = 200$	$N = 500$
$CV_\lambda = 0.5$	3.3%	1.3%	0.7%	0.3%	0.1%
$CV_\lambda = 1$	8.3%	3.0%	1.7%	0.9%	0.3%
$CV_\lambda = 1.5$	15.3%	6.6%	3.5%	1.9%	0.7%
$CV_\lambda = 3$	38.7%	21.6%	12.1%	7.1%	2.8%



(a) spreading step: 20% (b) spreading step: 70%

Fig. 10. Relative error $\frac{E[X_k]-k(N-k)\mu_\lambda}{k(N-k)\mu_\lambda}$ for the step (a) $k = 0.2 \cdot N$ and (b) $k = 0.7 \cdot N$. Box-plots correspond to 100 different network instances of the same size N, and mobility characteristics $\mu_\lambda = 1$ and $CV_\lambda = 1.5$.

REFERENCES

[1] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *Comm. Mag., IEEE*, vol. 44, no. 11, pp. 134–141, Nov. 2006.

[2] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. on Mob. Comp.,*, vol. 11, no. 5, 2012.

[3] M. Pitkänen, T. Kärkkäinen, and et al., "SCAMPI: service platform for social aware mobile and pervasive computing," *ACM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 503–508, Sep. 2012.

[4] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke University, Tech. Rep. CS-200006, 2000.

[5] T. Spyropoulos and A. Hossmann-Picu, *Opportunistic Routing (Chapter 11 in Mobile Ad Hoc Networking: The Cutting Edge Directions)*, second edition ed. John Wiley & Sons, 2012.

[6] T. Spyropoulos, R. N. Rais, T. Turletti, K. Obraczka, and A. Vasilakos, "Routing for disruption tolerant networks: taxonomy and design," *Wirel. Netw.*, vol. 16, no. 8, pp. 2349–2370, Nov. 2010.

[7] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the multiple-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 77–90, Feb. 2008.

[8] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[9] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, pp. 210–228, 2005.

[10] Z. J. Haas and T. Small, "A new networking model for biological applications of ad hoc sensor networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, pp. 27–40, 2006.

[11] W. Hsu and A. Helmy, "On nodal encounter patterns in wireless lan traces," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 1563–1577, 2010.

[12] W. Gao, Q. Li, B. Zhao, and G. Cao, "Social-aware multicast in disruption-tolerant networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1553–1566, 2012.

[13] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *Proc. ACM Autonomics*, 2007.

[14] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1576–1589, 2011.

[15] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Transactions on Mobile Computing*, vol. 8, pp. 1132–1147, 2009.

[16] C.-H. Lee and D. Y. Eun, "On the forwarding performance under heterogeneous contact dynamics in mobile opportunistic networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 6, 2013.

[17] A. Passarella and M. Conti, “Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks,” *IEEE Trans. on Mobile Computing*, vol. 12, no. 12, pp. 2483–2495, 2013.

[18] P. Sermpezis and T. Spyropoulos, “Modelling and analysis of communication traffic heterogeneity in opportunistic networks,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 11, pp. 2316–2331, 2015.

[19] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, “Epidemic spreading in real networks: An eigenvalue viewpoint.” in *SRDS*. IEEE Computer Society, 2003, pp. 25–34.

[20] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.

[21] A. Passarella, R. I. Dunbar, M. Conti, and F. Pezzoni, “Ego network models for future internet social networking environments,” *Computer Communications*, vol. 35, no. 18.

[22] A. Picu, T. Spyropoulos, and T. Hossmann, “An analysis of the information spreading delay in heterogeneous mobility dtns,” in *Proc. IEEE WoWMoM*, 2012.

[23] Y. Kim, K. Lee, N. B. Shroff, and I. Rhee, “Providing probabilistic guarantees on the time of information spread in opportunistic networks,” in *IEEE INFOCOM*, 2013.

[24] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, “Impact of human mobility on opportunistic forwarding algorithms,” *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[25] C. Boldrini, M. Conti, and A. Passarella, “The stability region of the delay in pareto opportunistic networks,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 1, pp. 180–193, 2015.

[26] ——, “Performance modelling of opportunistic forwarding under heterogenous mobility,” *Computer Communications*, vol. 48, pp. 56–70, 2014.

[27] A.-K. Pietiläinen and C. Diot, “Dissemination in opportunistic social networks: the role of temporal communities,” in *Proc. ACM MobiHoc*, 2012.

[28] W. Gao, G. Cao, T. La Porta, and J. Han, “On exploiting transient social contact patterns for data forwarding in delay-tolerant networks,” *IEEE Trans. on Mob. Computing*, vol. 12, no. 1, pp. 151–165, jan. 2013.

[29] G. Zyba, G. Voelker, S. Ioannidis, and C. Diot, “Dissemination in opportunistic mobile ad-hoc networks: The power of the crowd,” in *Proc. IEEE INFOCOM*, 2011.

[30] W. Gao, Q. Li, B. Zhao, and G. Cao, “Multicasting in delay tolerant networks: a social network perspective,” in *Proc. of ACM MobiHoc*, 2009.

[31] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, “Power law and exponential decay of intercontact times between mobile devices,” *IEEE Transactions on Mobile Computing*, vol. 9, no. 10, pp. 1377–1390, 2010.

[32] H. Cai and D. Y. Eun, “Crossing over the bounded domain: from exponential to power-law intermeeting time in mobile ad hoc networks,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1578–1591, Oct. 2009.

[33] A. Picu and T. Spyropoulos, “Dtn-meteo: Forecasting the performance of dtn protocols under heterogeneous mobility,” *IEEE/ACM Trans. on Networking*, vol. PP, no. 99, 2014.

[34] J. Leskovec, M. Mcglohon, C. Faloutsos, N. Glance, and M. Hurst, “Cascading behavior in large blog graphs,” in *In SDM*, 2007.

[35] M. E. J. Newman, “Spread of epidemic disease on networks,” *Phys. Rev. E*, vol. 66, Jul 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.66.016128

[36] S. M. Ross, *Introduction to Probability Models*, 9th ed. Academic Press, Elsevier, 2007.

[37] P. Sermpezis and T. Spyropoulos, “Effects of content popularity on the performance of content-centric opportunistic networking: An analytical approach and applications,” *IEEE/ACM Transactions on Networking*, no. 99, pp. 1–1, 2016.

[38] ——, “Information diffusion in heterogeneous networks: The configuration model approach,” in *Proc. IEEE NetSciCom (INFOCOM Workshops)*, 2013.

[39] ——, “Not all content is created equal: Effect of popularity and availability for content-centric opportunistic networking,” in *Proc. ACM MOBIHOC*, 2014.

[40] Y. Zhu, B. Xu, X. Shi, and Y. Wang, “A survey of social-based routing in delay tolerant networks: Positive and negative social effects,” *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–15, 2012.

[41] Pavlos Sermpezis, “Performance analysis of mobile social networks with realistic mobility and traffic patterns,” Ph.D. dissertation, Thesis, 2015.

[42] A. Mei and J. Stefa, “Give2get: Forwarding in social mobile wireless networks of selfish individuals,” *IEEE Trans. on Dependable and Secure Computing,*, vol. 9, no. 4, pp. 569–582, july-aug. 2012.

[43] B. B. Chen and M. C. Chan, “Mobicent: a credit-based incentive system for disruption tolerant network,” in *Proc. IEEE INFOCOM*, 2010.

[44] Q. Li, S. Zhu, and G. Cao, “Routing in socially selfish delay tolerant networks,” in *Proc. IEEE INFOCOM*, 2010.

[45] M. S. Granovetter, “The Strength of Weak Ties,” *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[46] E. Gilbert and K. Karahalios, “Predicting tie strength with social media,” in *Proc. CHI*, 2009.

[47] S. Ioannidis and A. Chaintreau, “On the Strength of Weak Ties in Mobile Social Networks,” in *Proc. ACM SNS Workshop*, 2009.

[48] P. Sermpezis and T. Spyropoulos, “Understanding the effects of social selfishness on the performance of heterogeneous opportunistic networks,” *Computer Communications, Elsevier, Volume 48*, 04 2014.

[49] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, “CRAWDAD data set epfl/mobility (v. 2009-02-24),” Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[50] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, “CRAWDAD data set cambridge/haggle (v. 2009-05-29),” Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle, May 2009.

[51] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, “Modeling spatial and temporal dependencies of user mobility in wireless mobile networks,” *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 5, pp. 1564–1577, 2009.

[52] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, “On the levy-walk nature of human mobility,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, 2011.

[53] A. Khelil, C. Becker, J. Tian, and K. Rothermel, “An epidemic model for information diffusion in manets,” in *Proc. of ACM MSWiM*, 2002. [Online]. Available: http://doi.acm.org/10.1145/570758.570768

[54] R. Lu, X. Lin, T. H. Luan, X. Liang, X. Li, L. Chen, and X. Shen, “Prefilter: An efficient privacy-preserving relay filtering scheme for delay tolerant networks.” in *Proc. IEEE INFOCOM*, 2012.

[55] M. J. Keeling, “The effects of local spatial structure on epidemiological invasions.” *Proc. R. Soc. B*, vol. 266, no. 1421, pp. 859–867, 1999.

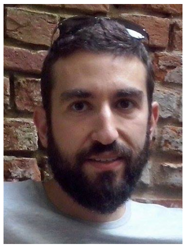
[56] M. E. J. Newman, “Spread of epidemic disease on networks,” *Phys. Rev. E*, vol. 66, p. 016128, Jul 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.66.016128

[57] F. Ball, D. Mollison, and G. Scalia-Tomba, “Epidemics with two levels of mixing,” *Annals of Applied Probability*, vol. 7, pp. 46–89, 1997.

[58] S. Rushton and A. Mautner, “The deterministic model of a simple epidemic for more than one community,” *Biometrika*, vol. 42, pp. 126–132, 1955.

[59] G. W. Oehlert, “A note on the delta method,” *The American Statistician*, vol. 46, no. 1, pp. 27–29, 1992.

[60] A. Karr, *Probability*, ser. Springer texts in statistics. Springer, 1993.



Pavlos Sermpezis received the Diploma in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece, and a PhD in Computer Science and Networks from EURECOM, Sophia Antipolis, France. He is currently a post-doctoral researcher at FORTH, Greece. His main research interests are in modeling and performance analysis for communication networks and protocols, and Internet routing.

Thrasyvoulos Spyropoulos received the Diploma in Electrical and Computer Engineering from the National Technical University of Athens, Greece, and a Ph.D degree in Electrical Engineering from the University of Southern California. He was a post-doctoral researcher at INRIA and then, a senior researcher with the Swiss Federal Institute of Technology (ETH) Zurich. He is currently an Assistant Professor at EURECOM, Sophia-Antipolis. He is the recipient of the best paper award in IEEE SECON 2008, and IEEE WoWMoM 2012.

APPENDIX A
PROOF OF LEMMA 2

Proof. The sum of the contact rates between infected nodes ($i \in \mathbf{C}_k^m$) and susceptible nodes ($j \notin \mathbf{C}_k^m$) at step k, is related to the respective sum of the previous step as follows

$$S_k^m = S_{k-1}^m - \sum_{i\in\mathbf{C}_{k-1}^m, j=n_k} \lambda_{ij} + \sum_{j=n_k, i\notin\mathbf{C}_k^m} \lambda_{ij} \tag{21}$$

where

(i) We denote as n_k the k^{th} infected node (i.e. the node infected at the transition between step $k-1$ and step k).

(ii) $\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k} \lambda_{ij}$ is the sum of the contact rates between the infected nodes at step $k-1$ and node n_k. These rates are included in the sum S_{k-1}^m, but are not included to the sum S_k^m (since at step k they belong to the set of contact rates between infected nodes), and, hence, we subtract them in Eq. (21).

(iii) $\sum_{i\notin\mathbf{C}_k^m, j=n_k} \lambda_{ij}$ is the sum of the contact rates between node n_k and the susceptible nodes at step k. These rates are included in the sum S_k^m, but are not included to the sum S_{k-1}^m (since at step $k-1$ they belong to the set of contact rates between susceptible nodes), and, hence, we add them in Eq. (21).

Now, we first split the sum $\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k} \lambda_{ij}$ in two terms[7]

$$\begin{aligned}\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k} \lambda_{ij} &= \lambda_{k-1}^{next} + \sum_{i\in\mathbf{C}_{k-1}^m, j=n_k, \lambda_{ij}\neq\lambda_{k-1}^{next}} \lambda_{ij} \\ &= \lambda_{k-1}^{next} + S_{k-1}^{next}\end{aligned} \tag{22}$$

where we denoted as λ_{k-1}^{next} the meeting rate between the next node to get the message (i.e. n_k) and the node who infected him, and

$$S_{k-1}^{next} = \sum_{i\in\mathbf{C}_{k-1}^m, j=n_k, \lambda_{ij}\neq\lambda_{k-1}^{next}} \lambda_{ij} \qquad (k-2 \text{ terms}) \tag{23}$$

We further denote

$$S_{next}^k = \sum_{i\notin\mathbf{C}_k^m, j=n_k} \lambda_{ij} \qquad (N-k \text{ terms}) \tag{24}$$

From Eq. (22), Eq. (23) and Eq. (24), Eq. (21) can be written as

$$S_k^m = S_{k-1}^m - \lambda_{k-1}^{next} - S_{k-1}^{next} + S_{next}^k \tag{25}$$

Based on the above recursive relation, in the remainder, we calculate the expectation and variance of S_k. Before proceeding, let us first define the following quantities (for $k = 1, \cdots, N-1$)

$$\mu_k = \frac{E[S_k]}{k(N-k)} \tag{26}$$

$$\sigma_k^2 = \frac{Var\left[S_k\right]}{k(N-k)} \tag{27}$$

Expectation

Taking the expectation in Eq. (25), gives

$$E[S_k] = E[S_{k-1}] - E[\lambda_{k-1}^{next}] - E[S_{k-1}^{next}] + E[S_{next}^k] \tag{28}$$

Now, we express the terms in the right side of Eq. (28), as follows:

(i) At first, by the definition of Eq. (26), we can write

$$E[S_{k-1}] = (k-1)(N-k+1)\cdot\mu_{k-1} \tag{29}$$

(ii) The probability that the node pair $\{x, y\}$, $x \in \mathbf{C}_{k-1}^m, y \notin \mathbf{C}_{k-1}^m$ (among all the node pairs $\{i, j\}$, $i \in \mathbf{C}_{k-1}^m, j \notin \mathbf{C}_{k-1}^m$), is the pair through which the message is spread at step $k-1$ (i.e. y is the k^{th} node that is infected, and it is infected by node x), is proportional to its contact rate λ_{xy} (because inter-contact intervals are exponentially distributed). Hence, we can, equivalently, write

$$\begin{aligned}&P\{\lambda_{k-1}^{next} = \lambda_{xy} | x \in \mathbf{C}_{k-1}^m, y \notin \mathbf{C}_{k-1}^m\} \\ &\qquad = \frac{\lambda_{xy}}{\sum_{i\in\mathbf{C}_{k-1}^m}\sum_{j\notin\mathbf{C}_{k-1}^m}\lambda_{ij}} = \frac{\lambda_{xy}}{S_{k-1}^m}\end{aligned} \tag{30}$$

From Eq. (30), it is easy to see that the rate λ_{k-1}^{next} will be on average larger than the average rate between node pair $\{i, j\}$, $i \in \mathbf{C}_{k-1}^m, j \notin \mathbf{C}_{k-1}^m$, i.e.

$$E[\lambda_{k-1}^{next}] \geq \mu_{k-1} \tag{31}$$

Moreover, it holds (by definition) that

$$\lambda_{k-1}^{next} \leq \lambda_{max} \tag{32}$$

Combining the two above inequalities, we can express the expectation $E[\lambda_{k-1}^{next}]$ as

$$E[\lambda_{k-1}^{next}] = \mu_{k-1} + \epsilon_{k-1}^{next} \tag{33}$$

where

$$0 \leq \epsilon_{k-1}^{next} \leq \lambda_{max} - \mu_{k-1} \quad\Rightarrow\quad \epsilon_{k-1}^{next} = O\left(\lambda_{max}\right) \tag{34}$$

(iii) The sum S_{k-1}^{next} consists of $k-2$ independent random variables, whose mean value is slightly smaller than μ_{k-1} (the rate λ_{k-1}^{next} -see Eq. (33)- is not taken into account). Hence, we can write

$$E[S_{k-1}^{next}] = (k-2)\cdot\mu_{k-1} + \epsilon_{k-1}^* \tag{35}$$

where, it can be shown (considering the effect of not considering the rate λ_{k-1}^{next}) that

$$\epsilon_{k-1}^* \leq \frac{(\lambda_{max} - \mu_{k-1})\cdot(k-1)}{(k-1)(N-k+1)-1} \;\Rightarrow\; \epsilon_{k-1}^* = O\left(\frac{\lambda_{max}}{N-k+1}\right) \tag{36}$$

In Eq. (36) it is easy to see that $\epsilon_{k-1}^* \ll (k-2)\cdot\mu_{k-1}, \forall k$ (for large N). Therefore, in the remainder, we can ignore it[8] and, thus, we write

$$E[S_{k-1}^{next}] = (k-2)\cdot\mu_{k-1} \tag{37}$$

---

[7] In Eq. (22)-Eq. (25), the quantities λ_{k-1}^{next}, S_{k-1}^{next} and S_{next}^k correspond to the sets $\mathbf{C}_{k-1}^m$ and $\mathbf{C}_k^m$. We dropped the superscripts m to avoid notation complexity. Also, in the remainder, the expectations of these quantities are taken over all the possible values (for different m) that these quantities can take.

[8] Additionally, since in Eq. (28) we consider $\epsilon_{k-1}^{next} = O\left(\lambda_{max}\right)$ (Eq. (34)), we can omit ϵ_{k-1}^*, for which it holds $\epsilon_{k-1}^* = O\left(\frac{\lambda_{max}}{N-k+1}\right) \leq \epsilon_{k-1}^{next}$.

(iv) At each step k, the values of the rates between the uninfected nodes, i.e.

$$\lambda_k^{out} \in \{\lambda_{ij} : i \notin \mathbf{C}_k^m, j \notin \mathbf{C}_k^m\}$$

are independent of the spreading process. Thus, these rates are distributed with the initial contact rate distribution $f_\lambda(\lambda)$, which means that

$$E[\lambda_k^{out}] = E[\lambda] = \mu_\lambda \tag{38}$$

$$Var\left[\lambda_k^{out}\right] = Var\left[\lambda\right] = \sigma_\lambda^2 \tag{39}$$

Therefore, from Eq. (38) it follows that the expectation of the sum S_{next}^k, which consists of $N-k$ contact rates between nodes that are not infected in step $k-1$ (i.e. $\notin \mathbf{C}_{k-1}^m$), is equal to

$$E[S_{next}^k] = (N-k)\cdot E[\lambda_{k-1}^{out}] = (N-k)\cdot \mu_\lambda \tag{40}$$

Substituting in Eq. (28) the expressions we derived in (i)-(iv) (Eq. (29), Eq. (33), Eq. (37) and Eq. (40)), we get

$$\begin{aligned}
&E[S_k] = \\
&= (k-1)(N-k+1)\mu_{k-1} - (\mu_{k-1} + \epsilon_{k-1}^{next}) \\
&\quad - (k-2)\mu_{k-1} + (N-k)\mu_\lambda \\
&= (k-1)(N-k+1)\cdot\mu_{k-1} - (k-1)\cdot\mu_{k-1} - \epsilon_{k-1}^{next} + (N-k)\cdot\mu_\lambda \\
&= (k-1)(N-k)\cdot\mu_{k-1} - \epsilon_{k-1}^{next} + (N-k)\cdot\mu_\lambda \\
&= k(N-k)\cdot\left(\frac{(k-1)\cdot\mu_{k-1}+\mu_\lambda}{k} - \frac{\epsilon_{k-1}^{next}}{k(N-k)}\right)
\end{aligned} \tag{41}$$

or

$$E[S_k] = k(N-k)\cdot\left(\frac{(k-1)\cdot\mu_{k-1}+\mu_\lambda}{k} - \epsilon_k^{'}\right) \tag{42}$$

where

$$\epsilon_k^{'} = O\left(\frac{\lambda_{max}}{k(N-k)}\right) \tag{43}$$

Now, to calculate $E[S_k]$ for every step k, we start from the first step ($k=1$), where S_1^m is a sum of $N-1$ i.i.d. random variables λ_{ij} with mean value μ_λ (by the definition of the mobility class). Therefore,

$$E[S_1] = (N-1)\cdot\mu_\lambda \tag{44}$$

For the second step ($k=2$), substituting Eq. (44) in Eq. (42), gives

$$E[S_2] = 2(N-2)\cdot\mu_\lambda\cdot(1-\epsilon_2^{'}), \quad \epsilon_2^{'} = O\left(\frac{\lambda_{max}}{N-1}\right) \tag{45}$$

Finally, following the same process recursively, for $k = 3, 4, \ldots$, it can be shown that $\forall k$ it holds

$$E[S_k] = k(N-k)\cdot\mu_\lambda\cdot(1-\epsilon_k), \quad \epsilon_k = O\left(\frac{\lambda_{max}}{N-1}\right) \tag{46}$$

which proves the first item of Lemma 2.

Variance

The second item of Lemma 2, for the variance $Var\left[S_k\right]$, is proved following a similar methodology as above. Taking the variances in Eq. (25), gives [36][9]

$$\begin{aligned}
&Var\left[S_k\right] = \\
&\quad Var\left[S_{k-1}\right] + Var\left[\lambda_{k-1}^{next}\right] + Var\left[S_{k-1}^{next}\right] + Var\left[S_{next}^k\right] \\
&\quad - 2\cdot Cov\left[S_{k-1}, \lambda_{k-1}^{next}\right] - 2\cdot Cov\left[S_{k-1}, S_{k-1}^{next}\right]
\end{aligned} \tag{47}$$

We proceed similarly to the derivation of the expectation, and express the terms in the right side of Eq. (47), as follows:
(i) At first, by the definition of Eq. (27), we can write

$$Var\left[S_{k-1}\right] = (k-1)(N-k+1)\cdot\sigma_{k-1}^2 \tag{48}$$

(ii) Similarly to Eq. (37)-Eq. (36), it follows that

$$Var\left[S_{k-1}^{next}\right] = (k-2)\cdot\sigma_{k-1}^2 + \delta_{k-1}^* \tag{49}$$

and because δ_{k-1}^* is small, we ignore it:

$$Var\left[S_{k-1}^{next}\right] = (k-2)\cdot\sigma_{k-1}^2 \tag{50}$$

(iii) Making similar arguments as in Eq. (40) and using Eq. (39), it follows that

$$Var\left[S_{next}^k\right] = (N-k)\cdot\sigma_\lambda^2 \tag{51}$$

(iv) The covariance of two random variables is given by the expression

$$Cov\left[X, Y\right] = E[X\cdot Y] - E[X]\cdot E[Y]$$

Therefore, for the first covariance appearing in the sum of Eq. (47) we can write

$$Cov\left[S_{k-1}, \lambda_{k-1}^{next}\right] = E[S_{k-1}\cdot\lambda_{k-1}^{next}] - E[S_{k-1}]\cdot E[\lambda_{k-1}^{next}] \tag{52}$$

Since S_{k-1} is a sum of $(k-1)(N-k+1)$ independent random variables, of which one of them is the contact rate λ_{k-1}^{next}, it follows that

$$\begin{aligned}
&E[S_{k-1}\cdot\lambda_{k-1}^{next}] = \\
&\quad E[(\lambda_{k-1}^{next})^2] + [(k-1)(N-k+1)-1]\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}]
\end{aligned} \tag{53}$$

Substituting Eq. (53) in Eq. (52), and using the expression derived in Eq. (29), we get

$$\begin{aligned}
&Cov\left[S_{k-1}, \lambda_{k-1}^{next}\right] = \\
&= E[(\lambda_{k-1}^{next})^2] + [(k-1)(N-k+1)-1]\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&\qquad - (k-1)(N-k+1)\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= E[(\lambda_{k-1}^{next})^2] - \mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= \left(Var\left[\lambda_{k-1}^{next}\right] + (E[\lambda_{k-1}^{next}])^2\right) - \mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= Var\left[\lambda_{k-1}^{next}\right] + E[\lambda_{k-1}^{next}]\cdot\left(E[\lambda_{k-1}^{next}] - \mu_{k-1}\right)
\end{aligned} \tag{54}$$

---

[9] The covariances of independent variables are zero, and, thus, we do not include them in Eq. (47).

Remark: In the previous derivations we used the the expression that relates the second moment of a random variable with its variance and mean value, i.e.

$$Var\left[X\right] = E[x^2] - \left(E[x]\right)^2 \quad \Leftrightarrow \quad E[x^2] = Var\left[X\right] + \left(E[x]\right)^2$$

Substituting in Eq. (54) the expression of Eq. (33), gives

$$Cov\left[S_{k-1}, \lambda_{k-1}^{next}\right] = Var\left[\lambda_{k-1}^{next}\right] + \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right) \cdot \epsilon_{k-1}^{next} \quad (55)$$

(v) We, similarly, express the second covariance appearing in the sum of Eq. (47) as

$$Cov\left[S_{k-1}, S_{k-1}^{next}\right] = E[S_{k-1} \cdot S_{k-1}^{next}] - E[S_{k-1}] \cdot E[S_{k-1}^{next}] \quad (56)$$

The sum S_{k-1}^{next} consists of $k-2$ terms, which are also included in the sum S_{k-1}. For each term λ^* in S_{k-1}^{next}, there are $(k-1)(N-k+1)-1$ terms in S_{k-1} that are independent of λ^*. Hence, we can, successively, write for Eq. (56)

$$\begin{aligned}
&Cov\left[S_{k-1}, S_{k-1}^{next}\right] = \\
&= (k-2) \cdot E[S_{k-1} \cdot \lambda^*] - E[S_{k-1}] \cdot (k-2) \cdot E[\lambda^*] \\
&= (k-2) \left((\sigma_{k-1}^2 + \mu_{k-1}^2) + [(k-1)(N-k+1)-1]\mu_{k-1}^2\right) \\
&\quad - (k-1)(N-k+1)\mu_{k-1} \cdot (k-2) \cdot \mu_{k-1} \\
&= (k-2) \cdot \sigma_{k-1}^2 \qquad (57)
\end{aligned}$$

Substituting in Eq. (47) the expressions we derived in (i)-(v) (Eq. (48), Eq. (50), Eq. (51), Eq. (55) and Eq. (57)), we get

$$\begin{aligned}
&Var\left[S_k\right] = \\
&= (k-1)(N-k+1) \cdot \sigma_{k-1}^2 + Var\left[\lambda_{k-1}^{next}\right] + (k-2) \cdot \sigma_{k-1}^2 \\
&\quad + (N-k) \cdot \sigma_\lambda^2 - 2 \cdot \left(Var\left[\lambda_{k-1}^{next}\right] + \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right) \cdot \epsilon_{k-1}^{next}\right) \\
&\quad - 2 \cdot (k-2) \cdot \sigma_{k-1}^2 \\
&= (k-1)(N-k+1) \cdot \sigma_{k-1}^2 - (k-2) \cdot \sigma_{k-1}^2 + (N-k) \cdot \sigma_\lambda^2 \\
&\quad - \left(Var\left[\lambda_{k-1}^{next}\right] + 2 \cdot \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right) \cdot \epsilon_{k-1}^{next}\right) \\
&= (k-1)(N-k) \cdot \sigma_{k-1}^2 + (N-k) \cdot \sigma_\lambda^2 \\
&\quad - \left(Var\left[\lambda_{k-1}^{next}\right] + 2 \cdot \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right) \cdot \epsilon_{k-1}^{next} - \sigma_{k-1}^2\right) \\
&= k(N-k) \cdot \left[\frac{(k-1) \cdot \sigma_{k-1}^2 + \sigma_\lambda^2}{k}\right. \\
&\quad \left. - \frac{Var\left[\lambda_{k-1}^{next}\right] + 2 \cdot \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right) \cdot \epsilon_{k-1}^{next} - \sigma_{k-1}^2}{k(N-k)}\right] \qquad (58)
\end{aligned}$$

or

$$Var\left[S_k\right] = k(N-k) \cdot \left(\frac{(k-1) \cdot \sigma_{k-1}^2 + \sigma_\lambda^2}{k} - \delta_k^{'}\right) \quad (59)$$

where

$$|\delta_k^{'}| = O\left(\frac{\lambda_{max}^2}{k(N-k)}\right) \quad (60)$$

Following a recursive procedure (for $k = 1, 2, ...$) as previously, it can be shown that $\forall k$ it holds

$$Var\left[S_k\right] = k(N-k) \cdot \sigma_\lambda^2 \cdot (1-\delta_k), \quad |\delta_k| = O\left(\frac{\lambda_{max}^2}{N-1}\right) \quad (61)$$

which proves the second item of Lemma 2. □

APPENDIX B
PROOF OF LEMMA 3

Proof. Using Lemma 2 we can write for the random variable $X_k = \frac{S_k}{k(N-k)}$:

$$E[X_k] = \frac{E[S_k]}{k(N-k)} = \frac{k(N-k) \cdot \mu_\lambda (1-\epsilon_k)}{k(N-k)} = \mu_\lambda (1-\epsilon_k) \quad (62)$$

and

$$\begin{aligned}
Var\left[X_k\right] &= \frac{Var\left[S_k\right]}{\left(k(N-k)\right)^2} = \frac{k(N-k) \cdot \sigma_\lambda^2 \cdot (1-\delta_k)}{\left(k(N-k)\right)^2} \\
&= \frac{\sigma_\lambda^2 \cdot (1-\delta_k)}{k(N-k)} \qquad (63)
\end{aligned}$$

The variance of X_k, Eq. (63), can be written as

$$\begin{aligned}
Var\left[X_k\right] &= E\left[\left(X_k - E[X_k]\right)^2\right] = E\left[\left(X_k - \mu_\lambda \cdot (1-\epsilon_k)\right)^2\right] \\
&= E\left[\left(X_k - \mu_\lambda\right)^2\right] - \left(\mu_\lambda \cdot \epsilon_k\right)^2 \qquad (64)
\end{aligned}$$

where we used the expression of Eq. (62) for $E[X_k]$.

Combining Eq. (63) and Eq. (64), we get

$$E\left[\left(X_k - \mu_\lambda\right)^2\right] = \frac{\sigma_\lambda^2 \cdot (1-\delta_k)}{k(N-k)} + \left(\mu_\lambda \cdot \epsilon_k\right)^2 \quad (65)$$

and taking the limit, for $N \to \infty$, in both sides of Eq. (65), gives

$$\lim_{N\to\infty} E\left[(X_k - \mu_\lambda)^2\right] = \lim_{N\to\infty} \left(\frac{\sigma_\lambda^2 \cdot (1-\delta_k)}{k(N-k)} + (\mu_\lambda \cdot \epsilon_k)^2\right) = 0$$

Therefore, (by definition [60, Def. 5.3, p. 136]) it follows that

$$X_k \xrightarrow{m.s.} \mu_\lambda \quad (66)$$

where $\xrightarrow{m.s.}$ denotes convergence in mean square. □

APPENDIX C
PROOF OF THEOREM 1

Proof. Lemma 3 shows the convergence in mean square for X_k. Therefore, it follows directly that X_k converges in probability as well [60, p. 140-141]

$$X_k \xrightarrow{m.s.} \mu_\lambda \quad \Rightarrow \quad X_k \xrightarrow{p} \mu_\lambda \quad (67)$$

where $\xrightarrow{p}$ denotes convergence in probability.

Let us, now, define the random variable Y_k as $Y_k = \frac{1}{X_k} = \frac{k(N-k)}{S_k}$, with probability distribution

$$P\left\{Y_k = \frac{k(N-k)}{S_k^m}\right\} = P\{\mathbf{C}_k^m\} \quad (68)$$

Since (see Eq. (67)) $X_k \xrightarrow{p} \mu_\lambda$, it also holds that [60, Thm. 5.23, p. 148]

$$Y_k = \frac{1}{X_k} \xrightarrow{p} \frac{1}{\mu_\lambda} \quad (69)$$

Moreover, since each contact rate λ_{ij} takes values in the interval $[\lambda_{min}, \lambda_{max}]$, it is easy to see that

$$\frac{1}{\lambda_{max}} \leq Y_k \leq \frac{1}{\lambda_{min}} \tag{70}$$

Using Eq. (70) and the definition of *uniform integrability* [60, Def. 5.15, p. 142], it follows that Y_k is uniformly integrable $\forall N$ and $\forall k \in [1, N-1]$, i.e.

$$\lim_{\alpha \to \infty} \sup_N E\left[|Y_k|; \{|Y_k| > \alpha\}\right] = 0 \tag{71}$$

because $P\{|Y_k| > \alpha\} = 0$ for $\alpha > \frac{1}{\lambda_{min}}$.

Eq. (69) states that Y_k converges in probability to $\frac{1}{\mu_\lambda}$ and Eq. (71) that Y_k is uniformly integrable. Therefore, [60, Thm. 5.17, p. 144], it follows that Y_k converges in mean value to $\frac{1}{\mu_\lambda}$:

$$Y_k \xrightarrow{m.} \frac{1}{\mu_\lambda} \quad \text{or} \quad E[Y_k] = E\left[\frac{1}{X_k}\right] \to \frac{1}{\mu_\lambda} \tag{72}$$

Finally, the relative error RE_k can be written as

$$\begin{aligned} RE_k &= \frac{E[T_{k,k+1}] - \frac{1}{k(N-k)\mu_\lambda}}{E[T_{k,k+1}]} = \frac{E\left[\frac{1}{S_k}\right] - \frac{1}{k(N-k)\mu_\lambda}}{E\left[\frac{1}{S_k}\right]} \\ &= \frac{E\left[\frac{1}{X_k}\right] - \frac{1}{\mu_\lambda}}{E\left[\frac{1}{X_k}\right]} = \frac{E[Y_k] - \frac{1}{\mu_\lambda}}{E[Y_k]} \end{aligned} \tag{73}$$

Taking the limit in Eq. (73) for $N \to \infty$ and using Eq. (72), gives

$$\lim_{N \to \infty} RE_k = \frac{\frac{1}{\mu_\lambda} - \frac{1}{\mu_\lambda}}{\frac{1}{\mu_\lambda}} = 0 \tag{74}$$

□

APPENDIX D

SKETCH OF PROOF OF COROLLARY 1

As previously defined, $\mathbf{C}_k^m$ is the set of nodes with the message (the "infected" nodes) at step k. For each node $i \in \mathbf{C}_k^m$, we now define the set $\mathbf{D}_{\mathbf{C}_k^m}(i)$ as

$$\mathbf{D}_{\mathbf{C}_k^m}(i) = \{j : j \notin \mathbf{C}_k^m \textit{ and } \lambda_{ij} > 0\} \tag{75}$$

$\mathbf{D}_{\mathbf{C}_k^m}(i)$ is the set of the nodes j that have not received yet the message and can contact node i. In a full-mesh network (Assumption 3), the cardinality of the set $\mathbf{D}_{\mathbf{C}_k^m}(i)$ is $\|\mathbf{D}_{\mathbf{C}_k^m}(i)\| = (N-k)$, whereas in a sparse network $0 \leq \|\mathbf{D}_{\mathbf{C}_k^m}(i)\| \leq N-k$. In particular, for the case we consider here (Poisson graphs), the sizes $\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|$ are (approximately; and exactly in the limit of large N) binomially distributed[10] as

$$P\{\|\mathbf{D}_{\mathbf{C}_k^m}(i)\| = d\} = \binom{N-k}{d} \cdot (p_s)^d \cdot (1-p_s)^{(N-k)-d} \tag{76}$$

with

$$E\left[\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = (N-k) \cdot p_s \tag{77}$$

$$Var\left[\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = (N-k) \cdot p_s \cdot (1-p_s) \tag{78}$$

where the probability space is defined over all possible sets m at step k, and all nodes i.

Now, similarly to Eq. (2) and Eq. (3), we define

$$S_{k(p)}^m = \sum_{i \in \mathbf{C}_k^m} \sum_{j \in \mathbf{D}_{\mathbf{C}_k^m}(i)} \lambda_{ij} \tag{79}$$

and the random variable

$$P\{S_{k(p)} = S_k^m\} = P\{\mathbf{C}_k^m\} \tag{80}$$

In a full-mesh network, S_k^m is a sum of $k(N-k)$ terms λ_{ij}, and the moments of S_k are given by Lemma 2. In the Poisson graph case we consider here, $S_{k(p)}^m$ is a sum of $\sum_{i \in \mathbf{C}_k^m} \|\mathbf{D}_{\mathbf{C}_k^m}(i)\|$ terms, where the quantity

$$D_k^m = \sum_{i \in \mathbf{C}_k^m} \|\mathbf{D}_{\mathbf{C}_k^m}(i)\| \tag{81}$$

is a random variable as well.

Therefore, (i) taking into account that $S_{k(p)}^m$, as a sum of a random number (D_k^m) of i.i.d. random variables (λ_{ij})[11], (ii) making similar arguments as in the proof of Lemma 2, and (iii) neglecting terms $O\left(\frac{1}{N}\right)$ (see e.g. ϵ_k and δ_k in Lemma 2), it can be shown that the expectation and variance of $S_{k(p)}$

$$E[S_{k(p)}] = E[D_k^m] \cdot \mu_\lambda \tag{82}$$

$$Var\left[S_{k(p)}\right] = E[D_k^m] \cdot \sigma_\lambda^2 + \mu_\lambda^2 \cdot Var\left[D_k^m\right] \tag{83}$$

Then, since D_k^m is a sum of k independent random variables (Eq. (81)), whose expectations and variances are given by Eq. (77) and Eq. (78), respectively, it follows that [36]

$$E[D_k^m] = k \cdot E\left[|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = k(N-k) \cdot p_s \tag{84}$$

$$Var\left[D_k^m\right] = k \cdot Var\left[|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = k(N-k) \cdot p_s \cdot (1-p_s) \tag{85}$$

Substituting Eq. (84) and Eq. (85) in the expressions of Eq. (82) and Eq. (83), we get

$$E[S_{k(p)}] = k(N-k) \cdot p_s \cdot \mu_\lambda \tag{86}$$

$$\begin{aligned} Var\left[S_{k(p)}\right] &= k(N-k) \cdot p_s \cdot \sigma_\lambda^2 + \mu_\lambda^2 \cdot k(N-k) \cdot p_s(1-p_s) \\ &= k(N-k) \cdot p_s \left[\sigma_\lambda^2 + \mu_\lambda^2 \cdot (1-p_s)\right] \end{aligned} \tag{87}$$

Comparing Eq. (86) and Eq. (87) to the corresponding expressions of Lemma 2, we can observe the correspondence suggested in Corollary 1.

---

[10] This is because, by the definition of a Poisson graph, for each node $i \in \mathbf{C}_k^m$, there are $N-k$ other nodes $j \notin \mathbf{C}_k^m$, each of which is a neighbor of i with probability p_s and independently of all other links.

[11] Expressions for the statistic moments of sums of random number of random variables are given in [36].