Large scale model for information dissemination with device to device communication using call details records

Rachit Agarwal, Vincent Gauthier, Monique Becker, Thouraya Toukabrigunes, Hossam Afifi

▶ **To cite this version:**

HAL Id: hal-01111385

https://hal.inria.fr/hal-01111385

Submitted on 4 Jul 2017

Large Scale Model for Information Dissemination with Device to Device Communication using Call Details Records

Rachit Agarwal[a], Vincent Gauthier[a,*], Monique Becker[a], Thouraya Toukabrigunes[a,b], Hossam Afifi[a]

[a]*Lab. CNRS SAMOVAR UMR 5157, Telecom SudParis/Institut Mines-Telecom.*
[b]*Orange Lab.*

Abstract

In a network of devices in close proximity such as *Device to Device* (D2D) communication network, we study the dissemination of public safety information at country scale. In order to provide a realistic model for the information dissemination, we extract a spatial distribution of the population of Ivory Coast from census data and determine migration pattern from the Call Detail Records (CDR) obtained during the *Data for Development* (D4D) challenge. We later apply epidemic model towards the information dissemination process based on the spatial properties of the user mobility extracted from the provided CDR. We then propose enhancements by adding latent states to the epidemic model in order to model more realistic user dynamics. Finally, we study dynamics of the evolution of the information spreading through the population.

Keywords: Device to Device communication, Human Mobility, Information Dissemination, Opportunistic networks.

1. Introduction

In communication networks, devices in each other's physical proximity can communicate. Information dissemination in such proximity based networks

---

*Corresponding author
**Email addresses of the authors are {rachit.agarwal, vincent.gauthier, monique.becker, hossam.afifi}@telecom-sudparis.eu and thouraya.toukabrigunes@orange.com

have been the focus of a lot of studies. Through this paper, we study the dissemination of public safety information at country scale. For the type of communication network, as a case study, we choose *Device to Device* ($D2D$) communication paradigms. This choice relates to the fact that the market of context-aware applications and location-based services has grown tremendously and operators have started to consider the deployment of $D2D$ communications as an underlay to the cellular networks. $D2D$ communication is defined as a short range communication between devices in physical proximity without any involvement of the network infrastructure. $D2D$ has many advantages like, autonomous communication, improved performance and spectrum reuse, low energy consumption and reduced load on the infrastructure. Moreover, other benefits of $D2D$ communication include direct communication between devices even when the traditional infrastructure is down, better connectivity in poor connectivity regions and increased average rate of successful message delivery. These above mentioned benefits of $D2D$ motivate us to choose $D2D$ communication paradigms as a Use Case. Nevertheless, our Use Case can also be applied to mobiles applications such as `Firechat` [1] that leverages direct connection between devices in close proximity to temporally connect users.

Moreover, in opportunistic networks, lot of studies have focus on the store-carry-and-forward paradigm and on the dynamics of the message delivery. The first attempt to model the process was represented via epidemic forwarding. Variations of the process includes the integration of *Social Network Analysis* measures to study the dissemination process [2, 3, 4, 5]. Further, modeling network performance through Markov Chains in opportunistic networks has also been the focus of several studies [6, 7, 8, 9]. However, none of the methods are scalable to the country scale since they rely on computing individual transmission probability. Dissemination methods based on mean field approach, however, are more scalable [10, 11] but rely upon homogeneous contact dynamics that is far from being valid. Thus, recent studies [12, 13, 14, 15] on dissemination process are based on heterogeneous flow within the *metapopulation*. This enabling us to study and create a heterogeneous population model. These studies offer

2

new perspectives on such models and enable us to study more easily country wide dissemination process.

Since, our interest lies in studying the dissemination process on a country-wide scale, we consider the population of the device to be spatially structured into subpopulation of well mixed individuals where the diffusion of safety information takes place. We use the official boundaries (termed as *subprefecture*) in *Ivory Coast* to define the subpopulation structure of our *metapopulation* model. In order to know the flow of people between the subprefectures, we use the Call Detail Records (CDR) provided by the Orange Labs during the Data for Development ($D4D$) Challenge [16]. For future reference, we refer the provided CDR as the $D4D$ dataset. The analysis of the movement pattern of the people in Ivory Coast reveals the flow of people between subprefectures. $D4D$ dataset enable us to monitor trajectories of people at unprecedented scale and derive key spatio-temporal patterns of the movement. Thus, the mobility is modeled as a network of interactions between each subpopulation (population within a subprefecture) where the connections correspond to the flow of people among them. Metapopulation[17, 18] model has been studied towards diffusion of information combined with human mobility where mobility pattern show strong heterogeneity [19, 20, 12]. Moreover, the outcome of the diffusion of the public safety information is regulated by the coupling of the mobility process and the diffusion process in each subpopulation. Thus, it is of utmost importance that the mobility model correctly describes realistic pattern of mobility.

However, a simple memoryless mobility model is far from being able to model recurring patterns such as commuting behavior or traveling behavior that are often observed in mobility data [19, 21]. To include this effect we use a non-markovian diffusive process to cope with the recurring behavior [13, 14]. In a realistic human population, a user is associated with a home location. Every time a user moves away from the home location, there is a probability to return to the home location from the new location. This return probability is also derived from the $D4D$ dataset. More details on how the return probability is computed is presented in the section 3.1.

3

The $D4D$ dataset includes the information about the user mobility and the association of a user to a community, termed as subprefecture. Further, in order to study the model with realistic population structure, we use the census data provided in [22] to populate each subprefecture (cf. Section 3.1).

Thus, in this paper, we use insights from the dissemination in the metapopulation model and the $D4D$ dataset to formulate our analytical model. Further, our model is divided into various stages. First, construct synthetic yet realistic population of the devices using census data [22] for the region of Ivory Coast. Then extract mobility information from the $D4D$ dataset. Then find mobility steady state population in each community. Finally, apply Susceptible(S)-Infected(I)-Recovered(R) type epidemic modeling (cf. Section 4.2) to model the information dissemination process at the countrywide scale.

In order to reduce the complexity of the dissemination process at country scale, we use the reactive-diffusive equation with homogeneous assumption to describe the subpopulation at the mechanistic level. We use a variation of the SIR epidemic model where Susceptible (S), Infected (I) and Recovered (R) states correspond to the state of a device 'not having the message', 'having the message and transmitting it' and 'having received the message but stop transmitting it due to message timeout (message time to live, TTL)', respectively. A comprehensive survey about the epidemic model is provided in [23, 24]. Recently, the authors of [25] showed that not only the community structure affects the dissemination process but the density of the communities also play an important role in the dissemination process. In metapopulation model it is often assumed that the population within a single subpopulation is well mixed. This is of course not generally true specially in a large subpopulation. To acknowledge this in a large subpopulation where some devices may never be in contact at a certain time, we include Latent states (E) in our model. Device in latent state account for devices that belong to a given subpopulation but are not participating in the diffusion process. To distinguish between devices that are Latent but Susceptible, Latent but Infected and Latent but Recovered, we subdivide E into three states: E_S, E_I and E_R. In our model only active devices in S, I
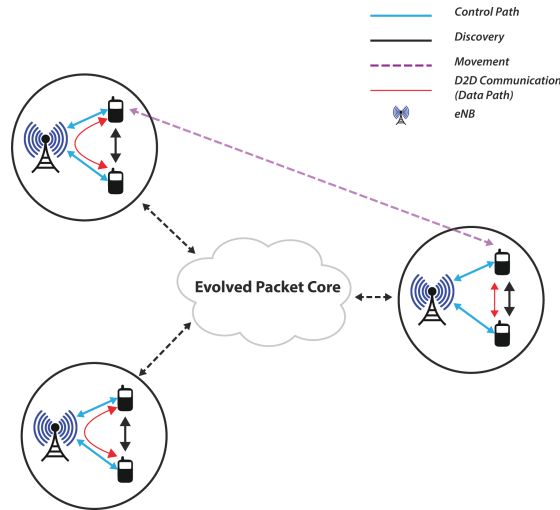
4

Figure 1: $D2D$ communication scheme.

and R states participate in the dissemination process. Thus, the novelty of the paper lies in the introduction of latent states and in the way we compute the number of interactions per device.

The remainder of this paper is organized as follows. In Section 2 we first provide an overview of the $D2D$ scenario. In Section 3 we describe how the dataset provided by the $D4D$ organizers is used to get useful information. A detailed description of the model is then provided in Section 4 which is followed by the results obtained in Section 5. The paper finally concludes with Section 6.

2. Loosely controlled $D2D$ support for information dissemination

We assume the following $D2D$ scenario: Using a Public Warning System (PWS), an eNB (evolved Node B, a radio interface in the Long Term Evolution (LTE) network) has to broadcast a national emergency warning which should reach the maximum number of people in the country. However, in such situations, the operator's network is massively overloaded by a huge traffic of users trying to use their mobile devices at the same time. Also, the warning

5

message may not reach a part of the population due to network coverage issues in rural areas. Using $D2D$ communications, devices in the physical proximity can communicate directly with each other and exchange messages. This mechanism allows not only an extension to the broadcasting area but also enhances the traffic on the network infrastructure and avoids network saturation and waste of resources. A $D2D$ communication has two phases: the Neighbor Discovery phase and the Communication phase for the data exchange. Both phases can be based on either a direct approach or a network-assisted approach [26, 27, 28, 29, 30].

In the direct $D2D$ approach, devices discover other devices in their surrounding by exchanging presence beacons. Both the phases, the discovery and the communication phase, are done in an adhoc-like way without any assistance or control of the network infrastructure. The communicating devices then form a self-organized and a self-configurable network. This approach is flexible and highly scalable as it can adapt to an increasing number of connected devices and active $D2D$ links. This allows offloading the traffic of the core operator network to local $D2D$ communication. Besides, it is also a solution to avoid communication disruption when some $eNBs$ are down in a disaster situation like Earthquake or when an eNB fails. The users could still use their devices and be connected to the operator network using direct communication.

In the centralized $D2D$ approach, the operator network may fully control the discovery and communication phases (fully controlled $D2D$) or assist the devices during the whole $D2D$ process (loosely controlled $D2D$) by enabling authentication and QoS mechanisms. Nevertheless, this approach is less scalable than the direct approach as it consists of performance and load balancing issues at the Radio Access Network (RAN) when dealing with a huge number of $D2D$ connections. As depicted in Fig. 1 we shows that some $eNBs$ can allow the direct communications between devices that are in physical proximity while some do not.

In addition, a device can move from one eNB to another. Moreover, the device can also move within an associated eNB. In this paper, we assume that
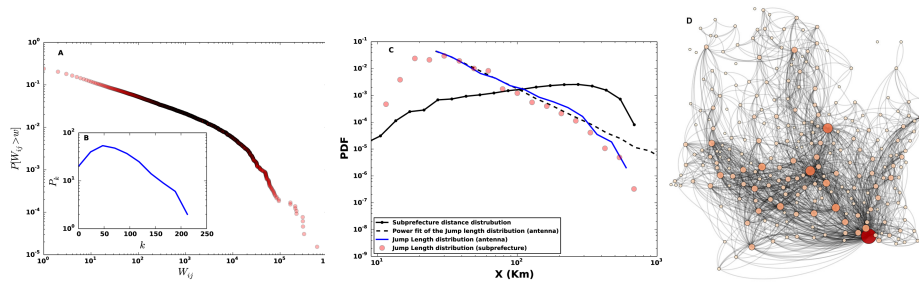
6

Figure 2: **a)** CCDF of flow distribution of the users obtained by aggregating all the transition made by user in the SET3 **b)** the degree distribution of the transition graph **c)** Transition graph **d)**

all $eNBs$ allow direct communication between devices that are in physical proximity and that a public warning message is generated in some area associated to an eNB. In the section 4, we are interested in investigating how fast dissemination of the warning message can be achieved through a large population in a $D2D$ environment where the devices are also mobile and links are intermittent. By fast dissemination we mean how, after the warning message is generated, a large set of devices can quickly receive the warning message.

3. Data

In this Section we provide detailed description of the $D4D$ dataset that we use and the methods we apply to extract useful information from the data. The dataset was collected by Orange Labs for the region of Ivory Coast. The data in the dataset contains the *Call Detail Records* of the calls made by users in the region of Ivory Coast [16]. The dataset contains both coarse grain mobility information as well as fine grain mobility information. The coarse grain mobility dataset is generated using subprefectures in Ivory Coast while fine grain dataset is generated using antenna locations. The dataset we use in this section is the course grain dataset and has the following structure. It contains the central location of the subprefecture area in the longitude and latitude format. The

7

dataset contains information of 500,000 users collected over 5 months from 5th December 2011 until 28th April 2012. Each tuple in dataset contains *subprefectureID* from where the call was made, *time* when the call was made and the *userID* of the users who made the call. Using aforementioned information we can estimate mobility pattern of a user. We decided mainly to use the coarse grain dataset, instead of the fine grain one, because it contains a larger user base.

Our first analysis (cf. Fig. 2a) of the dataset, in accordance with the previous analysis of [19, 20, 12], show that the flow distribution of the users movements display strong heterogeneities. The analysis highlights the fact that the flow of users toward some subprefectures is unevenly distributed (for example, the capital city subprefecture attracts more people than others. Cf. Fig. 2c where the graph depicts the transition structure between subprefectures in Ivory Coast. Here, a node is plotted according to their geographical coordinates and their size is proportional to the number of incoming transitions.). Secondly, we show that the choice of subprefecture as the base level for our metapopulation model introduces only a small bias in the distribution of the jump length (cf. Fig. 2b. A summary of various statistics is shown in Table 1). Note that, in Fig. 2b, the dashed line represents the distribution of the distances between subprefectures, the dots represent the distribution of jump lengths made by users when they make transitions between two subprefectures and the dashed line represents the power law fit of jump length distribution with parameter $\alpha = 2.51 \pm 0.0032$. Further, if we compare the distribution of jump length based on the subprefectures with the distribution of jump length made based on the position the antennas (cf. Fig. 2b), the former distribution is only slightly truncated at the head and at the tail as compared to the original distribution (cf. Fig. 2b). Note that the flow distribution at both subprefecture level and at antenna level is calculated using complete dataset.

Further, as an example, we track a user u (*userID*= 297412 in the dataset) (cf. Fig. 3a). Here, we show subprefectures reached by the user 297412. In the fig. 3a the width of the line represents the number of transitions made by the

8

Rank	Subprefecture	In-flow	Betweenness Centrality	Population Density (approx)	Area (approx)
1	Abidjan	0.1161	0.069256921	423.95	586.95
2	Bingerville	0.0371	0.010260404	423.95	257.66
3	Soubre	0.0323	0.017103843	54.39	1256.07
4	Meagui	0.0249	0.010316916	54.39	3491.10
5	Anyama	0.0245	0.030273783	423.95	695.02
6	Yamoussoukro	0.0240	0.043985509	67.19	1330.73
7	Songon	0.0216	0.022672720	423.95	623.31
8	Grand-Bassam	0.0180	0.011894960	52.56	124.98
9	Bonoua	0.0156	0.010142240	52.56	802.70
10	Tiassale	0.0156	0.030023932	57.24	2807.50

Table 1: Top 10 subprefectures ranked by their in-flow (per-capita). The table also lists the betweenness centrality of the subprefectures normalized by $N(N-1)$ (for directed graph) where N is the number of nodes in the graph, the respective population density ($people/km^2$) and area (km^2).

user between a pair of subprefecture and each dot represents the subprefecture location. The size of the subprefecture shows the time a user stays at a given subprefecture. The yellow subprefecture marks the home subprefecture of the user (cf. Def. 3).

3.1. Interpreting mobility information

Let C be the set of subprefectures. Let $P \in \mathbb{R}^{|C| \times |C|}$ be a weighted matrix where $P_{ij} \geq 0$ represents the number of transition between subprefecture i and j. Note that a transition here represents the number of times users in subprefecture i moved to subprefecture j. Also note that $P_{ii} = 0$ and P is an aggregated matrix of transitions made by all the users tracked in the $D4D$ dataset.

Definition 1. *Let $\nu \in \mathbb{R}^{|C| \times |C|}$ be a row stochastic matrix where the element ν_{ij} estimates the conditional probability of moving from subprefecture i to sub-*

9

prefecture j given that it has moved out of subprefecture i. We define ν_{ij} $\forall i \neq j$ as

$$\nu_{ij} = \frac{P_{ij}}{\sum_{\forall k \in C} P_{ik}}. \tag{1}$$

where $\sum_{\forall j \in C} \nu_{ij} = 1$ and $\nu_{ii} = 0$. Note that $\nu_{ij} \in \mathbb{R}^{+}$.

Definition 2. *Let $\sigma \in \mathbb{R}^{|C|}$ where the element σ_i represents the rate of moving out of subprefecture i. We define σ_i $\forall i \in C$ as*

$$\sigma_i = \frac{\sum_{\forall k \in C} P_{ik}}{Time} \tag{2}$$

where $Time = 150 \times 24 \times 60$ minutes, the total duration in minutes over which SET3 was generated. Note that $\sigma_i \in \mathbb{R}^{+}$. Also note that when $\sigma_i \to \infty$ it means that the user does not stay in subprefecture i and the movement out of the subprefecture occurs instantaneously. However, when $\sigma_i \to 0$, it means that a user stays permanently in subprefecture i.

We now formulate the rate at which the user returns to its home subprefecture. For this we first define home subprefecture as follows.

Definition 3. *Let $U_u = (C_{t_1}^1, C_{t_2}^2, \ldots, C_{t_n}^n)$, be a list of subprefectures visited by user u and let $t_n - t_{n-1}$ be the time spent in subprefecture C^{n-1}. Let M_j^u be the total time spent by user u in subprefecture j. The home subprefecture k, of the user u is the subprefecture that satisfies*

$$k = \arg\max_{\forall j \in C} M_j^u.$$

Let $P^r \in \mathbb{R}^{|C| \times |C|}$ be an aggregated weighted matrix where $P_{ji}^r \geq 0$ represents the number of times users returns to their home subprefecture i from subprefecture j.

Definition 4. *Let $\zeta \in \mathbb{R}^{|C| \times |C|}$ where the element ζ_{ji} represents the rate of coming back to the home subprefecture i from another subprefecture j. We define ζ_{ji} $\forall i \in C$ as*

$$\zeta_{ji} = \frac{P_{ji}^r}{Time}. \tag{3}$$

10

where $Time = 150 \times 24 \times 60$ minutes, the total duration in minutes over which SET3 was generated. Note that $\zeta_{ji} \in \mathbb{R}^+$. Also note that when $\zeta_{ji} \to \infty$, it means that the device returns to its home community i instantaneously after visiting community j. However, when $\zeta_{ji} \to 0$, it means that a device never moves back to its home community.

In the dataset a user is associated to one and only one subprefecture at a given time. Using this information we determine the number of users in each subprefecture $P^T \in \mathbb{R}^{|T| \times |C|}$ at a given *time slot* $t \in T$ where T is the set of time slots in the dataset SET3.

Definition 5. *The user density matrix, $\varrho \in \mathbb{R}^{|T| \times |C|}$, with elements ϱ_{ti} is defined as*

$$\varrho_{ti} = \frac{P_{ti}^T}{A_i} \tag{4}$$

where A_i is the area of the subprefecture i.

In Fig. 3b we provide a visual representation for the density of users on a log scale for the period from 16th December 2011 starting at 00:00:00 until 31st December 2011 ending at 23:59:00. The Fig. 3b shows the variation of number of people in a subprefectures with respect to time. As the dataset is a CDR where the information is logged only when a user makes a call, this variation relates to the fact that the number of calls made vary over time. Thereby suggesting that people tend to not communicate during certain period of time and thus not connect with the existing network at that time. In a $D2D$ perspective this change causes the network structure to vary over time and it thus provides us the motivation to incorporate the changes in the density through the concept of Latent States explained in Section 4.3.

The information in the dataset is also used to infer the number of calls made in a subprefecture over time and also to analyze different properties related to the calls made. However, due to the scope of the paper we limit ourselves from providing these results.

We shall now describe our dissemination model using σ, ν and ζ in the next section.
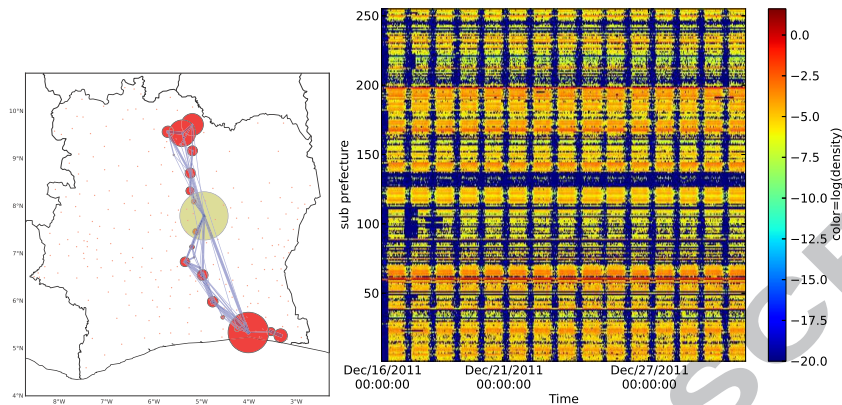
11

Figure 3: a) Trace of the user identified by number 297412, b) Density of the users in a subprefecture from 16th December 2011 until 31st December 2011 extracted from the dataset3.

4. Model

We represent our model in three layers. The bottom layer is the population layer where the population is spatially divided into communities. Note that we represent a person in the the population as a device. Above the population layer is the mobility layer where devices are allowed to jump from one community to another. Thus dissemination model is then applied on the top of the mobility model. The states in the dissemination model to which a device is associated to is either *Susceptible* or *Infected* or *Recovered*. We denote these states as S, I and R respectively.

4.1. Mobility Model

We consider a population of N devices distributed among $|C|$ subprefectures according to the census data [22]. Let $\rho_i = N_i/A_i$ be the density of devices in subprefecture i where N_i is the population in subprefecture i and $\sum_{i \in C} N_i = N$. In a $D2D$ context a subprefecture is the area where the population of devices is well mixed and can contact other devices when they are in close proximity to each other. Further, as argued by Watts *et al.*, community structure is evident in the population in a realistic scenario [18]. As each device is carried by humans

12

this allows us to state that each device closely follows the mobility pattern of humans. The movement of a device or the jump of a device from a community i to another community j occurs with a rate $\sigma_i \nu_{ij}$ where σ_i denotes the total rate out of community i and ν_{ij} denotes the conditional probability of going to community j given that there is a transition out of community i. We use $D4D$ SET3 as discussed in Section 3.1 to determine σ and ν. It was shown that the jump also depends on the nature of the community [31] and the activity pattern [32].

Further, we also consider rate of return to home community for each device [33, 34] and use ζ towards rate of return. Let N_{ii} be the number of devices having home community as i that are in community i and N_{ij} be the number of devices having home community as i that are in community j.

Fig. 4 shows a mobility model with only three communities where movement from one community i to another community j is represented with an edge defined by $\sigma_i \nu_{ij}$. The dotted line shows the return rate from another community and is marked with ζ_{ji} and dotted line and rate of moving out of a community is given by σ_i. Note that each community i in the above model consists of devices. Each community has associated number of devices. For different communities these devices are marked in different color (white for i, red for j, blue for k). The color of the lines also represents which community the devices are associated to. All other notations for movements are similarly defined. Moreover, this can be generalized to $|C|$ communities. This type of mobility model is a simple depiction of the inter-community movements with return rates.

Note that $\sum_{\forall i \in C} \sum_{\forall j \in C} N_{ij} = N$. Due to mobility, the number of the devices in a community i changes. As represented in Fig. 4 the travel pattern leads to [33] where the change in the number of devices in a community is defined as

$$
\left.
\begin{aligned}
\frac{dN_{ii}}{dt} &= \sum_{\forall j \in C, j \neq i} \zeta_{ji} N_{ij} - \sigma_i N_{ii} \\
\forall j \in C, j \neq i, \frac{dN_{ij}}{dt} &= \sigma_i \nu_{ij} N_{ii} - \zeta_{ji} N_{ij}
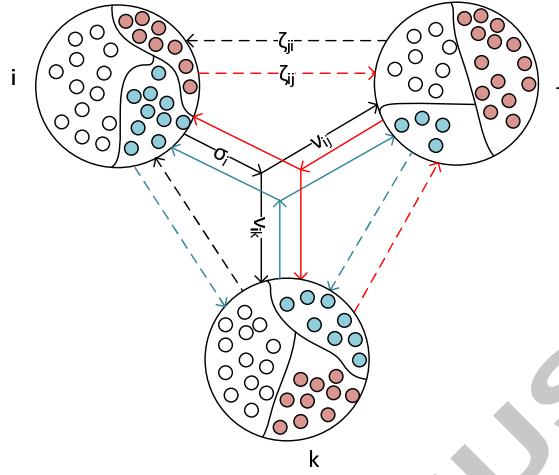\end{aligned}
\right\}
\tag{5}
$$

13

Figure 4: Mobility model with return rates.

The first term on the right hand side of the first equation in eq. 5, that is, $\sum_{\forall j \in C, j \neq i} \zeta_{ji} N_{ij}$, relates to the number of devices returning to their home community i from other communities while the second term of the same equation, that is, $\sigma_i N_{ii}$, refers to devices that are moving out of community i that have home community i. The second equation in eq. 5 can also be similarly explained. Note that similar equations apply to all other communities. Also note that the total number of such equations is $|C| \times |C|$ and $N_i = N_{ii} + \sum_{j \in C, j \neq i} N_{ij} = \sum_{j \in C} N_{ij}$, $\sum_{j \in C} N_i = N$.

In order to visualize the dissemination process, we first identify the steady state of the mobility model. Note that there are two models, mobility model and dissemination model. For the mobility model, we look for the steady state limit solution. Let N_{ii}^* and N_{ij}^* be the steady state population of devices having home location as i and are in community i and j respectively. Note that N_{ii}^* is a limit and will not change over time. Then, we use the differential equations in order to derive the transient solution for the dissemination process where we study the change in the number of devices in S, I and R state.

Using (5) we derive in (6) the number of devices per community at steady

14

state [33].

$$
\left.\begin{aligned}
N_{ii}^* &= N_i \left(\frac{1}{1 + \sigma_i \sum_{\forall k \in C, k \neq i} \frac{\nu_{ik}}{\zeta_{ki}}} \right) \\
\forall j \in C, j \neq i, N_{ij}^* &= N_i \left(\frac{\sigma_i \nu_{ij}}{\zeta_{ji} \left(1 + \sigma_i \sum_{\forall k \in C, k \neq i} \frac{\nu_{ik}}{\zeta_{ki}} \right)} \right)
\end{aligned}\right\}
\tag{6}
$$

4.2. Dissemination Model

We now describe the dissemination model. Let each device in a community be in either of the three states S, I and R. State S means that the device does not have the message. State I means that the device has the message and is transmitting it. State R means that the device has the message and is not transmitting the message as the TTL of the message is expired.

Let each device have an omnidirectional transmission range, r such that $r \in \mathbb{R}^+$. We assume the interactions among the devices to occur under mean field approximation where the expected number of contacts per device is given by $\langle k_i \rangle$.

Definition 6. *The expected neighborhood size for a device in a community i is defined as $\langle k_i \rangle = \rho_i^* \pi r^2$ where ρ_i^* is the population density at the mobility steady state and r the typical transmission range of a device.*

We assume that the devices in different communities have different β represented as β_i.

Definition 7. *In a community i the transition rate with which a message is received successfully is given by $\beta_i = -\log(1 - c_i)$ where c_i is the probability of correct transmission given a contact between two devices. It follows that the average number of newly people infected in community i per unit of time is $\langle k_i \rangle \frac{\beta_i}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji}$*

The variation in β across different community is because of many factors like, the density of the population and willingness of other devices to accept the transmitted information. This idea is analogous to the epidemic spreading in
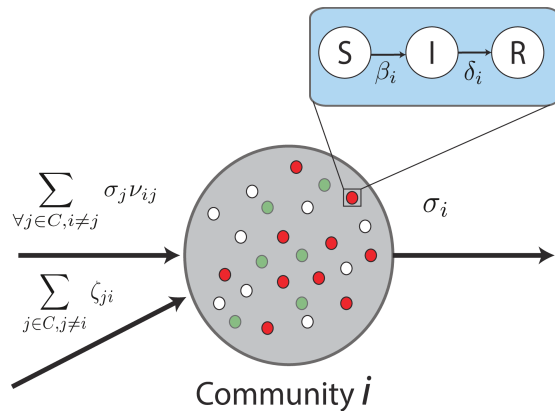
15

Figure 5: State diagram with states S, I and R with transition rates between states. The figure also shows the mobility transitions.

different communities which happens at different rates. Thus, devices in each community have different contact rates.

Definition 8. $1/\delta_i$ *is the expected time a device having a information will spread it to his neighborhood. In another word δ_i is the rate with which devices in community i change their state from I to R.*

In this scenario, the nodes that did not receive the public safety information yet (node in state I inside community i), will receive the information at rate $\langle k_i \rangle \beta_i$ and devices that have already received the information will stop sending it at rate given by δ_i (cf. Fig 5). Note that, this process occurs independently in each community. Due to the energy constraints, a device can only transmit information for a given period of time after which the device stop transfer the information. This can be understood as the rate with which devices change their state from I to R.

Similar to the explanation of N_{ii} and N_{ij}, let S_{ii}, I_{ii} and R_{ii} be the number of devices in state S, I and R respectively, that have the home community as community i and are in the community i while S_{ij}, I_{ij} and R_{ij} be the number of devices in state S, I and R respectively, that have the home community as community i and are in the community j. As we ap-

16

ply the dissemination model on top of the mobility model where we approximated the number of devices in each community by the steady state limit,

$$N_i^* = \sum_{\forall j \in C} N_{ij}^* = \sum_{\forall j \in C} [\mathbf{E}[S_{ij}] + \mathbf{E}[I_{ij}] + \mathbf{E}[R_{ij}]].$$

In a community i, let initially at time $t = 0$, all devices be in susceptible state. In order to study the dissemination process, let one community i at time $t = 0$ be the source of information, i.e., $S_{ii} = N_{ii}^* - \varepsilon$, $I_{ii} = \varepsilon$ and $\sum_{\forall j \in C} R_{ij} = 0$ where $0 < \varepsilon < N_{ii}$. Under mean field approximation, $\dfrac{1}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji}$ provides the fraction of interactions between devices in state S that have home community as i and are in community i with the devices in state I that have home community j and are in community i. As each device has a neighborhood $\langle k_i \rangle$ and the contact rate between devices is β_i, the number of devices in community i that will change their state from S to I is given by $\beta_i \dfrac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji}$. Let us consider the change in the number of devices in state S. Due to mobility, some devices in state S will return from other communities to community i while some devices in state S will move from community i to other communities. The number of devices moving from other communities to i is given by $\sum_{\forall j \in C, j \neq i} \zeta_{ji} S_{ij}$ while those moving from i to other communities is given by $\sigma_i S_{ii}$. Thus the rate of change in the number of devices in state S in community i that have home community i and are in community i in dt time is given by $-\beta_i \dfrac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} S_{ij} - \sigma_i S_{ii}$. Using such interpretation and above explanation the rate of change in the number of devices in state I and state R can be formulated. The rate equations for the change in number of devices in state S, I and R are given as

$$\begin{aligned}
\frac{dS_{ii}}{dt} &= -\beta_i \frac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} S_{ij} \\
&\quad - \sigma_i S_{ii} \\
\forall j \in C, j \neq i, \frac{dS_{ij}}{dt} &= -\beta_j \frac{\langle k_j \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} S_{ii} \\
&\quad - \zeta_{ji} S_{ij}
\end{aligned} \right\} \tag{7}$$

$$\begin{aligned}
\frac{dI_{ii}}{dt} &= \beta_i \frac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} I_{ij} \\
&\quad - \sigma_i I_{ii} - \delta_i I_{ii} \\
\forall j \in C, j \neq i, \frac{dI_{ij}}{dt} &= \beta_j \frac{\langle k_j \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} I_{ii} \\
&\quad - \zeta_{ji} I_{ij} - \delta_j I_{ij}
\end{aligned} \right\} \tag{8}$$

$$\begin{aligned}
\frac{dR_{ii}}{dt} &= \delta_i I_{ii} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} R_{ij} - \sigma_i R_{ii} \\
\forall j \in C, j \neq i, \frac{dR_{ij}}{dt} &= \delta_j I_{ij} + \sigma_i \nu_{ij} R_{ii} - \zeta_{ji} R_{ij}
\end{aligned} \right\} \tag{9}$$

Note that the total number of such equations in the system is $3 \times |C| \times |C|$ and the system is easy to solve.

For a single population model (when number of communities=1), the outbreak of epidemic will be reached if the basic reproduction number R_0 is reached. Under the homogeneous mixing of the population the basic reproduction number is

$$R_0 = \frac{\langle k \rangle \beta}{\delta} > 1$$

However, this did not take into account the mobility of an infected person. In the case of the metapopulation model, we need to take into account two thresholds. First, the local epidemic threshold, $R_0 > 1$, within each subpopulation. Second, the global invasion threshold, $R_* > 1$, that defines the travel rate of individual. More details on the two aforementioned thresholds for metapopulation model is available in [35].

18

4.3. Dynamic users model

Irrespective of whether a device is in state S, I or R, the capability of the device to transmit or receive also depends on whether the device is switched on or off. It is possible that a device has the information but has been switched off by its user. This hampers the transmission of the information from the device to other devices. Also, if a device is switched off, the device will not be able to receive the information from other devices. We call such state of a device as latent state and term them as E_S, E_I and E_R to represent latent states pertaining to each active state, S, I and R respectively. When a device is switched on, this will mark the transition in the state of the device from either E_S, E_I or E_R to S, I or R respectively. Further, to clarify the differences between the states, we list below the definitions of all the states

- **State** S - means that a device is ON and does not have received the message.

- **State** I - means that a device is ON, it has the message in his memory, the TTL of the message has not expired and the device is transmitting it.

- **State** R - means that the device is ON, it has received the message and the device stops transmitting it as the TTL of the message has expired.

- **State** E_S - means that a device is switched OFF and does not have the message.

- **State** E_I - means that a device is switched OFF, it has the message in its memory, the TTL of the message has not expired but the device cannot transmit the message as it is switched off. Also, the device can restart sending the message once it returns to the state I.

- **State** E_R - means that a device is OFF, it has received the message and the device stops transmitting it as the TTL of the message has expired.

Definition 9. *We define all the following parameters in* $\mathbb{R}^+$. *Let the transition rate to change the state from S to E_S in a community i be* μ_{S_i}. *Let the transition*
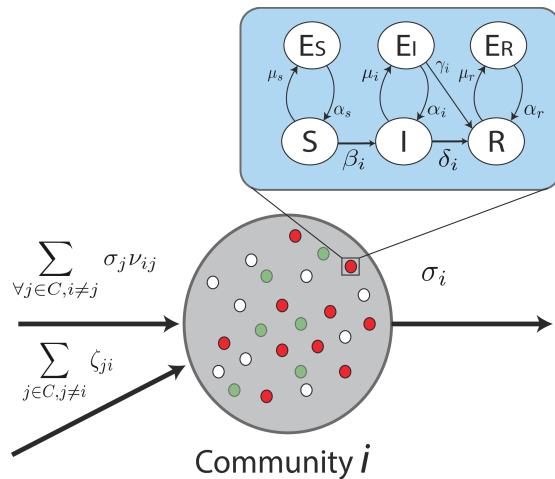
19

Figure 6: State diagram with states S, I and R and their latent states E_S, E_I and E_R respectively with transition rates between states. The figure also shows the mobility transitions.

rate to change the state from I to E_I in a community i be μ_{I_i}. Let the transition rate to change the state from R to E_R in a community i be defined as μ_{R_i}. The transition rate to change the state from E_S to S in a community i is defined as α_{S_i}. The transition rate to change the state from E_I to I in a community i is defined as α_{I_i}. The transition rate to change the state from E_R to R in a community i is defined as α_{R_i}.

Further, a device in state E_I can wake up and decide not to transmit information.

Definition 10. *The transition rate to change the state from E_I to R in a community i is defined as γ_i such that $\gamma_i \in \mathbb{R}^+$.*

Note that due to addition of three new states, $N_i^* = \sum_{\forall j \in C}[\mathbf{E}[S_{ij}] + \mathbf{E}[I_{ij}] + \mathbf{E}[R_{ij}] + \mathbf{E}[E_{S_{ij}}] + \mathbf{E}[E_{I_{ij}}] + \mathbf{E}[E_{R_{ij}}]$. Similar to initial condition as in the previous Subsection 4.2, we assume at time $t = 0$ $S_{ii} = N_{ii}^* - \varepsilon$, $I_{ii} = \varepsilon$ and $R_{ii} = E_{S_{ii}} = E_{I_{ii}} = E_{R_{ii}} = 0$, where $0 < \varepsilon < N_{ii}$.

Thus, the state diagram for this case is given by Fig. 6. This addition of latent states leads us to modify the equations from (7) - (9) for a community i

20

to equations (10) - (15). Note that, the total number of such equations in the system is $6 \times |C| \times |C|$.

Definition 11. *We borrow from [14] the idea of an heterogeneous model to compute the dwell time for the initial population currently in subprefecture i but having home location in subprefecture j. Let us assume that the return rate ζ_{ij} is $\zeta_{ij}^{-1} = \bar{\zeta}^{-1} \frac{d_j^{\chi}}{\langle d^{\chi} \rangle}$ where $\bar{\zeta}^{-1}$ is the average dwell time in our simulation and d_i is the degree of connectivity of the subprefecture i extracted from the transition matrix $\mathbf{P}$ between subprefectures, and $\langle d \rangle$ is the average degree of the subprefectures. The authors in [14] introduced $\chi < 0$ in order to take in account the fact that low degree subprefectures are generally peripheral node in the graph. A trip from a given origin thus may take multiple steps to reach the final destination marked by peripheral locations, and thus a longer dwell time at the destination may then account for longer travel time to peripheral locations.*

This assumption fits the case of the Ivory Coast where the peripheral subprefectures (subprefectures with low degree centrality) have a very poor road infrastructure (only trails) and as a result they are more difficult to reach. Thus taking more time to reach them.

5. Results

We use the Gillespie (τ-leap method [36, 37]) algorithm to solve the stochastic equations of the compartmental system of $255 \times 255 \times 6$ differential equations that describe the evolution of the system in each subprefecture (cf. Appendix A for detail explanation of the simulation setup). We use census data of the Ivory Coast in the year 1998 [22] to initialize the population of each subprefecture. Further, we compute surface area of each subprefecture through GIS mapping. Initially, we trigger a message inside a subprefecture *Akoboissue* ($I_{Akoboissue}(0) = \varepsilon$ and $\forall i \neq Akoboissue$ $I_i = 0$ where $\varepsilon = 0.0052$, i.e., $\varepsilon = \frac{100}{18959}$) in one simulation and in *Abidjan* in another ($I_{Abidjan}(0) = \varepsilon$ and $\forall i \neq Abidjan$ $I_i = 0$ where $\varepsilon = 4e - 4$, i.e., $\varepsilon = \frac{100}{248842}$). We define the transmission radius of a device to be $r = 100m$. Through the evolution of overall

21

$$\frac{dS_{ii}}{dt} = -\beta_i \frac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} S_{ij} - \sigma_i S_{ii} - \mu_{S_i} S_{ii} + \alpha_{S_i} E_{S_{ii}} \left.\right\}$$
$$\frac{dS_{ij}}{dt} = -\beta_j \frac{\langle k_j \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} S_{ii} - \zeta_{ji} S_{ij} - \mu_{S_j} S_{ij} + \alpha_{S_j} E_{S_{ij}} \left.\right\} \quad (10)$$

$$\frac{dI_{ii}}{dt} = \beta_i \frac{\langle k_i \rangle}{N_i^*} \sum_{\forall j \in C} S_{ii} I_{ji} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} I_{ij} - \sigma_i I_{ii} - \delta_i I_{ii} - \mu_{I_i} I_{ii} + \alpha_{I_i} E_{I_{ii}} \left.\right\}$$
$$\frac{dI_{ij}}{dt} = \beta_j \frac{\langle k_j \rangle}{N_j^*} \sum_{\forall q \in C} S_{ij} I_{qj} + \sigma_i \nu_{ij} I_{ii} - \zeta_{ji} I_{ij} - \delta_j I_{ij} - \mu_{I_j} I_{ij} + \alpha_{I_j} E_{I_{ij}} \left.\right\} \quad (11)$$

$$\frac{dR_{ii}}{dt} = \delta_i I_{ii} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} R_{ij} - \sigma_i R_{ii} - \mu_{R_i} R_{ii} + \alpha_{R_i} E_{R_{ii}} + \gamma_i E_{I_{ii}} \left.\right\}$$
$$\frac{dR_{ij}}{dt} = \delta_j I_{ij} + \sigma_i \nu_{ij} R_{ii} - \zeta_{ji} R_{ij} - \mu_{R_j} R_{ij} + \alpha_{R_j} E_{R_{ij}} + \gamma_j E_{I_{ij}} \left.\right\} \quad (12)$$

$$\frac{dE_{S_{ii}}}{dt} = \mu_{S_i} S_{ii} - \alpha_{S_i} E_{S_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{S_{ij}} - \sigma_i E_{S_{ii}} \left.\right\}$$
$$\frac{dE_{S_{ij}}}{dt} = \mu_{S_j} S_{ij} - \alpha_{S_j} E_{S_{ij}} + \sigma_i \nu_{ij} E_{S_{ii}} - \zeta_{ji} E_{S_{ij}} \left.\right\} \quad (13)$$

$$\frac{dE_{I_{ii}}}{dt} = \mu_{I_i} I_{ii} - \alpha_{I_i} E_{I_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{I_{ij}} - \sigma_i E_{I_{ii}} - \gamma_i E_{I_{ii}} \left.\right\}$$
$$\frac{dE_{I_{ij}}}{dt} = \mu_{I_j} I_{ij} - \alpha_{I_j} E_{I_{ij}} + \sigma_i \nu_{ij} E_{I_{ii}} - \zeta_{ji} E_{I_{ij}} - \gamma_j E_{I_{ij}} \left.\right\} \quad (14)$$

$$\frac{dE_{R_{ii}}}{dt} = \mu_{R_i} R_{ii} - \alpha_{R_i} E_{R_{ii}} + \sum_{\forall j \in C, j \neq i} \zeta_{ji} E_{R_{ij}} - \sigma_i E_{R_{ii}} \left.\right\}$$
$$\frac{dE_{R_{ij}}}{dt} = \mu_{R_j} R_{ij} - \alpha_{R_j} E_{R_{ij}} + \sigma_i \nu_{ij} E_{R_{ii}} - \zeta_{ji} E_{R_{ij}} \left.\right\} \quad (15)$$

percentage of devices in either of the three states, we show the effect on the dissemination of the public warning message across the population.

We first display as supplementary material a movie [38] that shows the diffusion process in Ivory Coast. We observe that the diffusion initially takes place in the subprefecture *Akoboissue* of the country. The information then spreads to major cities of Ivory Coast (as suggested by the mobility model cf. Fig. 2c) through subprefecture *Abidjan* (the economic capital), *Bouake* (the second largest city), *Yamoussoukro* (Political Capital) and *Soubre*. The information then spreads slowly to other cities in the Ivory Coast. Also note that, the south east side of the country is well known to be mostly a cocoa producing agricultural region [39]. It is also note worthy that the cocoa accounts for 25% of the economic GDP of the country. *Abidjan* being one major economic city in south
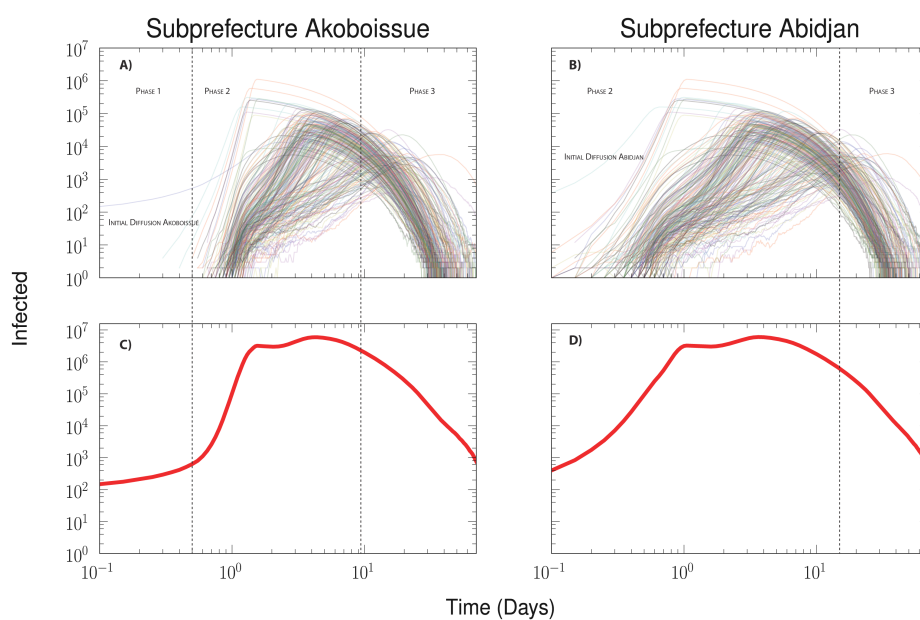
22

Figure 7: (**a**) and (**b**) Time evolution of the diffusion process across all subprefectures in the country for two different origins (sparsely populated and densely populated respectively) when latent states are not used. (**c**) and (**d**) show the overall fraction of infected devices at a given time for respective origins.

east of Ivory Coast, accounts for high mobility of people and is well connected, thus also resulting into initial diffusion. We also note that the diffusion of the information first spreads to south eastern part of the country, then to south western. It then takes a long time to spread over the northern part of the country. The authors of [40], who worked on the same dataset, brought out the fact that one factor that the northern part of the country is less diffusive is due to the consequence of socio-economic disparity inside the country. They also brought out that this disparity can be a consequence of the fact that the northern part of the country is still relatively "*disconnected from the main economic and political center of Ivory Coast*". Further, it should also be noted that the northern part of the country is mostly *Savanna* land with relatively less population.

In Fig. 7, we investigate the sensibility of our model to the placement of the initial diffusion. Through Fig. 7 we show the time evolution of the diffusion process across the country (in all 255 subprefectures) using the model without latent states for two different origins *Akoboissue* and *Abidjan*. Fig. 7a and Fig. 7c show the case when the origin of the diffusion is located in a sparsely populated area (Akoboissue) while Fig. 7b and Fig. 7d show the case when the diffusion process originates from a dense area (the capital of the country - Abidjan). Through Fig. 7a and Fig. 7c we observe three distinct phases, phase 1: the messages are spreading only inside the initially infected subprefecture, phase 2: most of the subprefectures get infected, phase 3: the overall number of infected devices starts to decrease, however the infection continues in the peripheral subprefectures. Here it is observed that the diffusion reaches the major city of the country one day after the initial spreading (Phase 1), and further diffuses into the rest of the country in the time interval between 2 to 10 days (Phase 2). Finally, it reaches the northern part of the country after 10 days (Phase 3). However in contrast, via Fig. 7b and Fig. 7d, it is noted that when the diffusion is initiated in a denser area, the diffusion process progresses quickly throughout the country. Here phase 1 is almost non existent and the infection spreads to other neighboring subprefectures rapidly. Further, Fig. 7 also support that the subprefectures in the north receive messages during
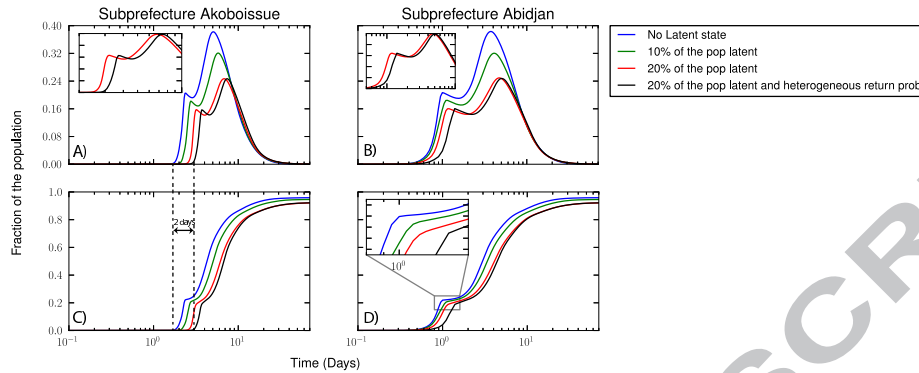
24

Figure 8: Time evolution of the number of infective device in Ivory Coast with different percentage of population in latent state. (**a**) when the diffusion starts from a sparsely populated area, (**b**) when the diffusion starts from a dense area, (**c**) Cumulative fraction of the population infected the initial infection is initiated from a sparsely populated area, (**d**) Cumulative fraction of the population infected when the initial infection is initiated from a densely populated area.

the phase 3 (long time after the densely connected subprefecture of the south received the message). From the Fig. 7d we also observe that the first peak of infection takes place just less than one day after the initial diffusion started as compare to day and half in Fig. 7c.

It is also important to note that the diffusion process is enhanced if initial broadcast of the message started in the subprefecture that had high betweenness centrality. In this case the diffusion of the message is more quickly to a large portion of other subprefectures since its average distance to other subprefectures is relatively small (cf. Fig 7c and Fig. 7d). In the Fig. 7b and Fig. 7d the initial diffusion takes place inside the subprefecture of Abidjan which is both the most connected and also has the highest betweenness centrality (Table 1).

Finally from the Fig. 8 we show the time evolution of the number of infective device in Ivory Coast when the latent state concept was added. Here, we temporally disable a given fraction of the population from being able to receive or transmitting message. In the Fig. 8, the blue curve refers to the model without latent state. The green and the red curves refer to the case when 10%

25

and 20% devices are in latent state respectively. While the black curve reflects the evolution when 20% of the population is latent and heterogeneous return probability (cf. Def. 11) is used. Here, for heterogeneous case, we assume that the dwell time of people has an exponential distribution with a mean of $\frac{1}{2}$ day. The salient fact about this simulation is that removing temporally a fraction of the population from being active has for obvious consequence of delaying the peak of infection (cf. Fig. 8a and Fig. 8c). However, this effect is mitigated in the second case when we initiate the diffusion process from Abidjan (cf. Fig. 8b and 8d). Due to its strategic position inside the graph, in Abidjan (cf. Fig. 2c) (highest centrality, highest connectivity, denser area cf. Table 1) the removal of small fraction of the active population has almost no effect on the speed of the dissemination process (however the height of peak of infection remains attenuated as compared to the no latent case). This stems from the fact that in the case when the initial diffusion takes place in *Akoboissue*, at the initial state the rate of migration (which is very low) of people carrying the message to neighboring (neighbors based on the flow of people, cf. Fig. 2c) subprefectures is slowed down by people in inactive states (latent states), therefore it takes more time for neighboring prefecture to reach the critical threshold that enables them to sustain the diffusion process themselves.

6. Discussion and Conclusion

In this paper, we presented a large scale data driven model for information dissemination in an heterogeneous network of relations (metapopulation) between subpopulation. We used the data provided by the $D4D$ organizers to determine the movement probabilities for the movement of the devices from one community to another. To realize the information dissemination process we used SIR type epidemic model with addition of latent states E_S, E_I and E_R introducing dynamic users's behaviors. Thus, the paper presents two main contributions, first the introduction of latent states to account for variable density and second, how mobility in a network with community structure could be

26

applied to achieve large scale dissemination in a dynamic $D2D$ based communication network.

7. acknowledgement

Appendix A. Simulation Setup

During the simulation of the spreading process using the "τ-leap" method we modeled the departure rate as a Poisson process with specific mean for each subprefecture according to figure A.9. The dwell time in each latent states is modeled as an exponential time of half a day in the generic case and according to Definition 11 in the heterogeneous case. The rates μ_s, μ_i, μ_r are modeled as a Poisson process defined as function of the mean fraction of the population we wanted to put in latent states (either 10% or 20% of the total population).

In our simulation, the diffusion process starts with a population repartition that is already in steady state. Nonetheless, one question remains, does the repartition of the population at steady state $D^{(2)}$, remains consistent with the initial repartition of the population $D^{(1)}$ (census data). In order to answer this question, we compare distributions quantitatively by calculating the symmetrized *Kullback-Leibler (KL)* divergence between the two partitions. The KL divergence between distribution $D^{(1)} = (D_i^{(1)})$ and $D^{(2)} = (D_i^{(2)})$ is defined as:

$$DIV_{KL}\left(D^{(1)}||D^{(2)}\right) = \frac{1}{2}\left(\sum_i D_i^{(1)} \log \frac{D_i^{(1)}}{D_i^{(2)}} + \sum_i D_i^{(2)} \log \frac{D_i^{(2)}}{D_i^{(1)}}\right)$$

We find that the KL divergence of the order 0.001, meaning, there is only a marginal difference between the population distribution at initial state and at the steady state. To conclude our model of the population (model at steady state) does not deviate from original distribution of the population (census).
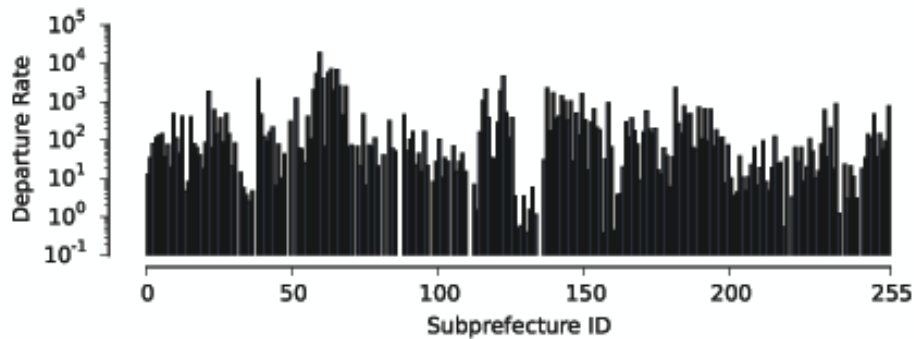
27

Figure A.9: Departure rate from each subprefecture

However, it remains that the overall mobility of persons inside the country is highly skewed towards some prefectures (cf. Fig. 2).

As a conclusion, the variance of the mobility process is bounded and does not evolve during the simulation, since we already computed its steady state before the simulation of the spreading process started. The simulation code used in this paper is available at the following address [41].

References

[1] FireChat, http://opengarden.com/firechat, 2014.

[2] P. Hui, J. Crowcroft, E. Yoneki, Bubble rap: Social-based forwarding in delay-tolerant networks, Mobile Computing, IEEE Transactions on 10 (2011) 1576–1589.

[3] E. Daly, M. Haahr, Social network analysis for information flow in disconnected delay-tolerant manets, Mobile Computing, IEEE Transactions on 8 (2009) 606–621.

[4] A. Mtibaa, K. A. Harras, Caf: Community aware framework for large scale mobile opportunistic networks, Computer Communications 36 (2013) 180 – 190.

28

[5] N. Belblidia, M. D. de Amorim, L. H. M. Costa, J. Leguay, V. Conan, Part-whole dissemination of large multimedia contents in opportunistic networks, Computer Communications 35 (2012) 1786 – 1797. Smart and Interactive Ubiquitous Multimedia Services.

[6] R. Groenevelt, P. Nain, G. Koole, The message delay in mobile ad hoc networks, Performance Evaluation 62 (2005) 210 – 228. Performance 2005 24th International Symposium on Computer Performance, Modeling, Measurements and Evaluation.

[7] A. A. Hanbali, P. Nain, E. Altman, Performance of ad hoc networks with two-hop relay routing and limited packet lifetime (extended version), Performance Evaluation 65 (2008) 463 – 483. Innovative Performance Evaluation Methodologies and Tools: Selected Papers from ValueTools 2006.

[8] T. Spyropoulos, K. Psounis, C. Raghavendra, Efficient routing in intermittently connected mobile networks: The single-copy case, Networking, IEEE/ACM Transactions on 16 (2008) 63–76.

[9] C. Boldrini, M. Conti, A. Passarella, Performance modelling of opportunistic forwarding under heterogenous mobility, Computer Communications 48 (2014) 56 – 70. Opportunistic networks.

[10] X. Zhang, G. Neglia, J. Kurose, D. Towsley, Performance modeling of epidemic routing, Comput. Netw. 51 (2007) 2867–2891.

[11] Z. Haas, T. Small, A new networking model for biological applications of ad hoc sensor networks, Networking, IEEE/ACM Transactions on 14 (2006) 27–40.

[12] V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations., Journal of theoretical biology 251 (2008) 450–467.

[13] D. Balcan, A. Vespignani, Phase transitions in contagion processes mediated by recurrent mobility patterns., Nature physics 7 (2011) 581–586.

29

[14] C. Poletto, M. Tizzoni, V. Colizza, Heterogeneous length of stay of hosts'
     movements and spatial epidemic spread., Scientific reports 2 (2012).

[15] L. Wang, Z. Wang, Y. Zhang, X. Li, How human location-specific con-
     tact patterns impact spatial transmission between populations?, Scientific
     reports 3 (2013).

[16] V. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot,
     Z. Smoreda, C. Ziemlicki, Data for Development: The D4D Challenge on
     Mobile Phone Data, http://arxiv.org/pdf/1210.0137v1.pdf, 2012.

[17] J. Arino, P. Van den Driessche, Disease Spread in Metapopulation, Fields
     Institute Communications 48 (2006) 1–13.

[18] D. Watts, R. Muhamad, D. Medina, P. Dodds, Multiscale, Resurgent
     Epidemics in a Hierarchical Metapopulation Model, Proceedings of the
     National Academy of Sciences of the United States of America 102 (2005)
     11157–11162.

[19] M. González, C. Hidalgo, A. Barabási, Understanding individual human
     mobility patterns., Nature 453 (2008) 779–782.

[20] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel.,
     Nature 439 (2006) 462–465.

[21] C. Song, Z. Qu, N. Blumm, A. Barabási, Limits of predictability in human
     mobility., Science (New York, N.Y.) 327 (2010) 1018–1021.

[22] http://www.statoids.com/yci.html, 2000.

[23] H. W. Hethcote, The Mathematics of Infectious Diseases, SIAM Review
     42 (2000) 599–653.

[24] T. Britton, Stochastic Epidemic Models: A Survey, Mathematical bio-
     sciences 225 (2010) 24–35.

30

[25] H. Lund, L. Lizana, I. Simonsen, Effects of City-Size Heterogeneity on Epidemic Spreading in a Metapopulation: A Reaction-Diffusion Approach, Journal of Statistical Physics (2013).

[26] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, Z. Turanyi, Design Aspects of Network Assisted Device-to-Device Communications, IEEE Communications Magazine 50 (2012) 170–177.

[27] L. Lei, Z. Zhong, C. Lin, X. Shen, Operator Controlled Device-to-Device Communications in LTE-Advanced Networks, IEEE Wireless Communications 19 (2012) 96–104.

[28] 3GPP TR 22.803 Feasibility study for Proximity Services (ProSe), Technical Report, 2013.

[29] 3GPP TS 23.303 Architecture enhancements to Support Proximity Services (ProSe), Technical Report, 2014.

[30] C. Toukabri Gunes, S. Tsang Kwong, H. Afifi, Hybrid model for LTE Network-Assisted D2D communications, in: Proceedings of Ad-Hoc, Mobile and Wireless Networks (ADHOCNOW), 2014.

[31] B. Wang, L. Cao, H. Suzuki, K. Aihara, Safety-Information-Driven Human Mobility Patterns with Metapopulation Epidemic Dynamics, Scientific reports 2 (2012) 1–8.

[32] S.-Y. Liu, A. Baronchelli, N. Perra, Contagion Dynamics in Time-Varying Metapopulation Networks, Physical Review E 87 (2013) 032805(1–9).

[33] L. Sattenspiel, K. Dietz, A Structured Epidemic Model Incorporating Geographic Mobility among Regions, Mathematical Biosciences 128 (1995) 71–91.

[34] V. Belik, T. Geisel, D. Brockmann, Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases, Physical Review X 1 (2011) 011001(1–5).

[35] A. Apolloni, C. Poletto, J. Ramasco, P. Jensen, V. Colizza, Metapopulation epidemic models with heterogeneous mixing and travel behaviour, Theoretical Biology and Medical Modelling 11 (2014) 3.

[36] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, The Journal of Chemical Physics 115 (2001) 1716–1733.

[37] Y. Cao, D. T. Gillespie, L. R. Petzold, Efficient step size selection for the tau-leaping simulation method, The Journal of Chemical Physics 124 (2006).

[38] R. Agarwal, V. Gauthier, M. Becker, Diffusion process using mobility data available for D4D challenge, http://dx.doi.org/10.6084/m9.figshare.698176, 2013.

[39] Map of Côte d'ivoire economic activity, http://bit.ly/1nTWR3c, 2013.

[40] C. Andris, L. Bettencourt, Development, Information and Social Connectivity in Côte d'Ivoire, http://www.santafe.edu/media/workingpapers/13-06023.pdf, 2012.

[41] Code Repository of the paper, http://dx.doi.org/10.5281/zenodo.11759, 2014.

32