Temporal Networks Based on Human Mobility Models: A Comparative Analysis With Real-World Networks

DJIBRIL MBoup¹, CHERIF DIALLO¹, and HOcine cherry²

¹LACCA Laboratory, Gaston Berger University, Saint-Louis PB 234, Senegal
²LIB EA 7534, University of Burgundy Franche-Comté, 21078 Dijon, France

Corresponding author: Djibril Mboup (mboup.djibril@ugb.edu.sn)

ABSTRACT Mobility is a critical element for understanding human contact networks. In many studies, the researchers use random processes to model human mobility. However, people do not move randomly in their environment. Their interactions do not depend only on spatial constraints but on their temporal, social, economic, and cultural activities. The topological structure of the physical and/or proximity contact networks depends, therefore, entirely on the mobility patterns. This paper performs an extensive comparative analysis of real-world temporal contact networks and synthetic networks based on influential mobility models. Results show that the various topological properties of most of the synthetic datasets depart from those observed in real-world contact networks because the randomness of some mobility parameters tends to move away from human contact properties. However, it appears that data generated using Spatio-Temporal Parametric Stepping (STEPS) mobility model reveals similarities with real temporal contact networks such as heavy-tailed distribution of contact duration, frequency of pairs of contacts, and the bursty phenomenon. These results pave the way for further improvement of mobility models to generate meaningful artificial contact networks.

INDEX TERMS Temporal networks, contact networks, proximity networks, time varying graphs, human mobility networks, human dynamics.

I. INTRODUCTION
The evolution of contact network topology is of prime interest for understanding internal dynamics such as information diffusion, opinion, rumor spreading, and disease propagation. Most often, it depends on human mobility patterns. For instance, the number of person-to-person contacts for each individual between very close time intervals at an airport or in a hospital can be an essential factor in spreading the virus. Indeed, the spread of a virus can be fast and difficult to stop because of human mobility [1], [2]. Modeling human mobility is not an easy task. Nowadays, researchers try to integrate various characteristics of human behavior into synthetic mobility models: heavy-tailed distribution of jump length and waiting time [3], [4], super-diffusive behavior at shorter Spatio-temporal scales, and sub-diffusive behavior in large scales [5], exploration and preferential return [4], [5], burstiness in human activities and so on [6]. Artificial mobility models are crucial in the modeling of human displacement. They are essential in different research areas such as transportation, urban planning, opportunistic networks. In default of ground-truth data, they provide case studies of simulation scenarios in the real world. In this context, mobility models can be used to study the contact networks between individuals.

In this paper, we examine the main properties of human contact networks based on real-world data compared to artificial contact networks based on synthetic mobility data. Our goal is to better understand the similarities and dissimilarities between these two types of networks. For this purpose, we consider real contact datasets from the Sociopattern project [7]–[9], the Copenhagen Networks Study (CNS) [10], and the traces generated by four synthetic mobility models (Random Waypoint [11], [12], Gauss-Markov model [13], [14], Truncated Lévy Walk [15] and Spatio-Temporal Parametric Stepping [16]). To perform a comparative study...
of the topological properties of these two types of datasets, we build time-varying graphs called temporal networks.

This work is a starting point for new developments on the design of efficient mobility models able to generate meaningful artificial contact networks. Our main contributions are threefold:

- First, we perform a comparative analysis of real-world temporal contact networks with each other to get a clear idea about their main common properties.
- Second, we investigate the similarities of real-world temporal networks with synthetic temporal networks generated from the different mobility models under study.
- Finally, we try to identify the relationship that exists between the mobility model features and the topological properties of their related temporal contact networks.

This paper is organized as follows. First, we give a brief review of the recent works (Section II) before exploring the mobility models and the temporal network’s properties used in this study (Section III). The section IV describes the real data and methods used for the experimental setup and reports comparative analysis of their topological properties, while the section V reproduces the same work in synthetic contact datasets. As for the section VI, it evaluates the similar and dissimilar properties between real contact networks and synthetic contact networks with a comparative analysis. Finally, we summarize the main findings and discuss the future direction of this work in section VI, and then conclude.

II. RELATED WORKS

Temporal contact networks can be built from different data sources: survey [17], [18], Call Data Records (CDR) [3], [19], Radio Frequency Identification (RFID) [7] GPS [20], [21], Location-Based Social Network (LBSN) [22], traces of mobility models [23], [24] and so on. These contacts can be physical, social interactions or close-range of proximity. The topological properties of these networks are often influenced by the characteristics that govern human behavior.

In the literature, many studies deal with contact networks using synthetic or real data sets. Using mobile phone datasets from D4D (Data for Development), Blondel et al. [19] studied the static and dynamic properties of contact networks. They found that clustering and topological overlap are associated with strong persistence. Moreover, human mobility greatly impacts burstiness activities, and social ties [19]. Unfortunately, mobile phone datasets are noisy and present low resolution related to the distance of the antenna towers. This distance ranges from meters to kilometers. Furthermore, the positions of individuals are sometimes partially detected.

Other studies use GPS data to track physical interactions. In [25], Liqiang et al. use GPS traces to study the time-ordered path, the connectedness, the temporal efficiency, and the reachability in vehicular ad hoc networks (VANETs). They define a contact when the Euclidean distance of two vehicles is smaller than the wireless communication range \( R \) during a time window. Recently, in the context of COVID 19, we witness the development of numerous contact tracing apps using GPS signals or wireless to capture the historical data of contacts [26]–[28]. GPS data exhibit high spatial and temporal resolution. However, there are not always available because of privacy issues and data protection policies.

Barrat and colleagues [29] collected public contact datasets from RFID devices. These datasets called Sociopatterns are collected in a limited scale i.e. in close environment (building, school, hospital). They allow us to study epidemic spreading and temporal behavior of human activities. Using these datasets, Starnini et al. show that the contact duration and the inter-events time are compatible with heavy-tailed distributions. The heterogeneity of contacts is relevant to the diffusion process [30].

Synthetic traces of mobility models always serve as a proxy to test the robustness of routing protocols in Opportunistic networks. Scellato et al. [31] use temporal network’s properties to analyze the robustness of networks based on various models. They compare the effectiveness of the Erdos-Reyni model, Markov model, and networks based on random mobility models, particularly the Random Waypoint Model (RWP) and Random Waypoint Group Model (RWPG). The authors define robustness as the ratio between the temporal efficiency after the damage and the temporal global efficiency. This measure helps to understand the performance of wireless communication in mobile networks. However, it does not explore the human behavior characteristics in temporal contact networks.

In a previous work, we performed a preliminary study about the influence of distance proximity on an epidemic spreading process using synthetic mobility datasets [24]. Results show that data generated from mobility models present an excellent opportunity to study temporal contact networks. To our knowledge, there is no prior comparative study between synthetic temporal networks and real temporal networks.

III. BACKGROUND

A. MOBILITY MODELS

The majority of contributions on the mobility modeling issue come from the mobile ad hoc networks (MANETs) research area. Trying to mimic human motion, these models do not incorporate all the characteristics of human behaviors. Various mobility models have been developed based on random walks and human motion features. This section gives a brief review of synthetic mobility models used to generate temporal networks of contacts.

1) RANDOM WAYPOINT MOBILITY MODEL

Johnson and Maltz [11] proposed the Random Waypoint mobility model (RWP). It is a popular mobility model used for simulations in MANETs for testing routing protocols. This random walk model computes the mobile position according to its speed, direction, and pause time at each stage of its travel. At every time step, the mobile chooses
a random destination called waypoint (at position \((x_i, y_i)\)) in \(U([X_{\text{min}}, X_{\text{max}}])\) with a velocity \(v_i\) that is uniformly distributed in the interval \([v_{\text{min}}, v_{\text{max}}]\). At destination, the mobile observes a pause time \(t_p\) ranging from \(t_{p_{\text{min}}}\) to \(t_{p_{\text{max}}}\) before moving to another location at position \((x_{i+1}, y_{i+1})\). Note that RWP is a random walk mobility model if the pause times are null. Figure 1 illustrates the mobility of RWP model. The mobile moves in straight lines and turns sharply in different directions. Hyytia et al. [32] studied the spatial distribution of nodes. They found that nodes tend to concentrate near the center of the area while disconnected nodes are likely to locate near the border.

FIGURE 1. Movement of a mobile following the RWP mobility with speed chosen randomly in \([0.1\, \text{ms}^{-1}, 1.\, \text{ms}^{-1}]\) and \(t_{p_{\text{min}}} = 0, t_{p_{\text{max}}} = 5\, \text{s}\).

2) GAUSS-MARKOV MOBILITY MODEL
The Gauss-Markov mobility model (GM) developed by Liang and Haas [13], tries to imitate the random process of Gaussian-Markov. It is more realistic than RWP. Indeed, the mobile can accelerate, slow down or turn gradually in any direction as shown in Figure 2. For each time step, the position of the mobile is given by the following equations [14]:

\[
x_i = x_{i-1} + v_{i-1} \cos \theta_{i-1}
\]

\[
y_i = y_{i-1} + v_{i-1} \sin \theta_{i-1}
\]

where \((x_i, y_i)\) and \((x_{i-1}, y_{i-1})\) are respectively the \(x\) and \(y\) coordinates of the mobile at position \(i\)-th and \((i-1)\)-th time step. \(v_{i-1}\) and \(\theta_{i-1}\) are respectively the speed and the direction of the mobile at \((i-1)\) time interval. Equations 3 and 4 give the expression of the velocity \(v_i\) and the direction \(\theta_i\) at \(i\)-th time step [14].

\[
v_i = \lambda v_{i-1} + (1 - \lambda) \mu v + \sqrt{1 - \lambda^2} w_{v_{i-1}}
\]

\[
\theta_i = \lambda \theta_{i-1} + (1 - \lambda) \mu \theta + \sqrt{1 - \lambda^2} w_{\theta_{i-1}}
\]

where \(\lambda \in [0, 1]\) is a tuning parameter keeping the randomness of the model. \(\mu_v\) and \(\mu_\theta\) are respectively the mean speed and direction values such that \(i \to \infty\), \(w_{v_{i-1}}\) and \(w_{\theta_{i-1}}\) are random variables following a Gaussian distribution with zero mean and unit standard deviation. Figure 2 shows a typical trajectory of a mobile moving according to the GM mobility model.

3) TRUNCATED LEVY WALK
Rhee et al. [15] introduced the Truncated Levy Walk (TLW) to reproduce important features observed in human mobility. These characteristics are the heavy-tailed distributions of the flight length and the waiting time and the super-diffusive behavior of the mean-squared displacement (MSD) i.e. \(MSD(t) \sim t^\gamma\) with \(\gamma > 1\) [4], [33]. In this model, the authors use the term ‘flight’, which can be defined as the longest straight-line from one location to another that a user performs, without any change of direction and pause time. Formally, a TLW is a random walk formed by a sequence of steps such that each step is represented by \(S = (l, \theta, t_f, t_p)\) such that \(l > 0\) is the flight length, \(\theta\) represents the direction of the flight. Finally, \(t_f\) and \(t_p\) are respectively the duration of the flight and the pause time. At each time step, the mobile chooses a random direction of angle \(\theta\) from a uniform distribution \(U([0, 2\pi])\), a finite duration \(t_f > 0\) picked up from a power-law distribution, a flight length \(l\) and a time pause \(t_p\) randomly chosen from the probability density \(p(l)\) and \(\Psi(t_p)\) following the Levy distribution with the coefficients \(\alpha\) and \(\beta\), respectively.

Figure 3 report a trajectory based on Truncated Levy Walk mobility model. One can see that the mobile often moves very often short distances and rarely long distances.

FIGURE 2. Trajectory of a mobile following a GM mobility trajectory with \(\mu_v = 1, \mu_\theta\) and the parameter \(\lambda = 0.2\).

FIGURE 3. Mobile trajectory following the TLW model with flight length exponent \(\alpha = -2.6\) and the exponent of pause time distribution \(\beta = -1.8\).
4) SPATIO-TEMPORAL PARAMETRIC STEPPING MODEL

Spatio-temporal parametric stepping model (STEPS) adds the preferential attachment in the choice of places to visit [16] to the power law property of the trip displacement of the individuals or jump length, in the pause time [3], [4] and in the duration of contacts and inter-contacts [34]. Initially, the mobile chooses a zone \( Z_0 \). Repetitively, it randomly selects a distance \( d \) depending on the following probability distribution

\[
P(D = d) = \frac{\zeta}{(1 + d)^\kappa}
\]  

(5)

with \( d \) a distance from \( Z_0 \), \( \kappa \) represents the exponent of the power law and \( \zeta \) a normalization constant. The mobile randomly selects a zone \( Z_i \) from the zones whose distances between \( Z_0 \) is \( d \) and a point in this zone where it will get at a speed varying in the range \([v_{min}, v_{max}]\). Inside \( Z_i \), it moves randomly, following a RWP motion. Note that the mobile stay in this zone during a pause time \( p_t \), picked in the power law distribution defined below.

\[
P(T = t) = \frac{\omega}{t^\tau}
\]  

(6)

where \( \tau \) is the degree of temporal preference of the mobile and \( \omega \) the constant of normalization.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image}
\caption{Movement of agent following the STEPS mobility with \( x = 2.5 \). The size of the zone is \( 10 \) m, the speed \( v_{RWP} \) is chosen in \([0.85, 1.38]\) (\( ms^{-1} \)). \( \nabla_{RWP} \in [1, 5] \), the speed inter zones is randomly picked in \([0.83, 3.6]\) (\( ms^{-1} \)) and the delay that agents staying inside zones \( Z_i \) is picked in \([20, 30]\) s.}
\end{figure}

Table 1 summarizes the main differences between the mobility models. Except for TLW, all the feature parameters of models are generic.

**B. TEMPORAL NETWORKS**

1) DEFINITION

A temporal network or time-varying graph can be represented as instantaneous sequences of events or snapshots of static graphs evolving over time. We used interchangeably these two terms in the following. By definition, a temporal network can be considered as a time-dependent sequence of contacts represented by \( C = (u_t, v_t, t, \delta t) \) where \( u_t \) and \( v_t \) are node pair at \( t \)th event, \( t \) and \( \delta t \) are respectively the time and the duration of the \( t \)th event. Ignoring the contact duration \( \delta t \), the temporal network is defined by the triplet \( (u_t, v_t, t) \) [35], [36] [37], [38]. Alternatively, we consider a time-varying graph to be a set of snapshots of graph \( G = \{G_1, G_2, \ldots, G_T\} \) where \( G_t = (V_t, E_t) \) is generated at each time interval \( t \in \mathbb{T} \), where \( \mathbb{T} \) is a measure of discrete time \( \mathbb{T} \subset \mathbb{N} \) and \( V_t = V \) [39]. Similarly, this definition does not take into account the duration that an event occurs. Otherwise, the temporal network can be seen as an extension of Multi-layer network such that each snapshot is considered to be a layer of level \( t \). Each of the properties of temporal networks uses at least one of these representations to be defined. Figure 5 shows that temporal networks can be represented as annotated graph or sequence of contacts. We now define the main properties of the local and global structure of temporal networks.

2) PROPERTIES

- **Temporal Path** Unlike static networks, the notion of path in temporal networks depends on the chronological order of connectivity, hence the name of time-respecting path. We can define the time-respecting path between a source node \( v \) and a target node \( w \) as a sequence of contacts such that:

\[(v_0, v_1, t_1), (v_1, v_2, t_2), \ldots, (v_{l-1}, v_l, t_l)\]

where \( v_0 = v \) and \( v_l = w \) and an ordered sequence of times such that \( t_1 < t_2 < \ldots < t_l \). In the literature, the term of time-respecting path is also called journey [36] or temporal path [39], [40]. Consequently, the concept of connectedness does not imply a symmetric or a transitive relationship. The term of strong connectivity is used when two nodes \( v \) and \( w \) are temporally connected in both directions.

The **temporal path length** is the time interval between the first and the last contact along the time-respecting path i.e., \( t_l - t_1 + 1 \). The **latency** is the shortest temporal distance \( \tau_{ij} \) to go from \( i \) to \( j \) at time \( t \) following the time-respecting paths. Hence, the average **temporal path length** can be defined as

\[
\mathcal{L} = \frac{1}{N(N - 1)} \sum_{ij} \tau_{ij}
\]

(7)

Furthermore, this measure is important when we need to uncover the small-world effect in temporal networks. An alternative measure called temporal Efficiency noted by \( \mathcal{E} \) is proposed in the context of disconnected time-varying graphs:

\[
\mathcal{E} = \frac{1}{N(N - 1)} \sum_{ij} \frac{1}{\tau_{ij}}
\]

(8)

In real world situations, temporal network are known to present many disconnection with sparsity. Therefore, the measure of network density evolving in times, noted \( D_t \) is also used to characterize the topology of temporal networks.
TABLE 1. Mobility models: input parameters and human mobility characteristics.

| Models | Parameters                              | Type            | Human Mobility properties                             |
|--------|-----------------------------------------|-----------------|------------------------------------------------------|
| RWP    | $v_{\min}$, $v_{\max}$, $t_{p_{\min}}$, $t_{p_{max}}$ | Generic         | Randomness                                           |
| GM     | $\lambda, \mu_1, \mu_2$                | Generic         | Acceleration/Deceleration, smooth turn for any direction |
| TLW    | $\alpha, \beta, t_{max}, t_{p_{max}}$   | Traces based    | Truncated power law distributions of the flight length and the pause time, Nature super-diffusive of the mean squared displacement (MSD) |
| STEPS  | $\kappa, zone\_size, v_{rup}, t_{prup}, zone\_speed, zone\_time$ | Generic         | Truncated power law distributions of the jump length, the pause time, and the inter-contact; the preferential location choice. |

• **Temporal Reachability**
  The Reachability condition is time-dependent in temporal networks. The temporal reachability can be estimated as the delay to reach all or a proportion of nodes. In [41], Holme defines the two following concepts. The *reachability time* is the average of all pairs such that there exists
a time-respecting path connecting them. The reachability ratio is the percentage of the vertex-pairs that do have a time respecting path between them. Recently, Thompson et al. propose the measure of reachability latency which is the average duration to reach a portion $r$ of nodes in the temporal graph [42]. This formula is given by the following equation:

$$R_r = \frac{1}{TN} \sum_t \sum_i d^t_{i,k}$$  \hfill (9)

where $d^t_i$ is an ordered vector of size $N$, that contains the list of shortest temporal paths passing through node $i$ at time $t$. $k$ represents the $[kN]$-th element in the vector, that is the rounded product of the portion $r$, with $N$ the total number of nodes in the network. When $r = 1$, then 100% of nodes are reachable and it corresponds to the temporal diameter of the network defined by Eq. 10:

$$R_1 = \frac{1}{TN} \sum_t \sum_i \max_j d^t_{i,j}$$  \hfill (10)

- **Temporal Correlation Coefficient**

If two nodes are connected at time $t$ then there exists a non-negligible probability that they will share a link at time $t + \Delta t$. This characteristic is measured by the topological overlap of the neighborhood $C_t$ of nodes in the interval $[t, t + \Delta t]$. Its average value is called temporal correlation by Tang et al. [40]. The latter is given by the equation 12 and it can be interpreted as the concept of clustering coefficient in temporal networks.

$$C_t = \frac{1}{T-1} \sum_{i=1}^{T-1} \frac{\sum a_{ij}(t)a_{ij}(t+1)}{\left(\sum a_{ij}(t)\right)\left(\sum a_{ij}(t+1)\right)}$$  \hfill (11)

$$C = \frac{1}{N} \sum_{t=1}^{T-1} C_t$$  \hfill (12)

where $1 \leq t \leq T - 1$ and $a_{ij}(t) \in A^t$, the adjacency matrix of $G_t$. This measure quantifies the overall average probability for a link to persist across two consecutive snapshots of graphs. In other words, $C = 1$ if all snapshots are the same and larger value of $C$ is obtained where many links appear at both $t$ and $t+1$.

- **Fluctuability and Volatility**

Thompson et al. [42] propose the measure of Fluctuability to quantify the variability of connectivity in brain networks at the macroscopic level. It is defined as the ratio of the number of edges present in the aggregated matrix $A$ over the grand sum of the matrix $A_t$ corresponding to $G_t$:

$$F = \frac{\sum_i \sum_j U(a_{ij})}{\sum_{i=1}^{T} \sum_{j=1}^{N} a^t_{ij}}$$  \hfill (13)

with,

$$U(a_{ij}) = \begin{cases} 1, & \text{if } \sum_{t=1}^{T} a^t_{ij} > 0 \\ 0, & \text{if } \sum_{t=1}^{T} a^t_{ij} = 0 \end{cases}$$  \hfill (14)

From Eq. 13, one can deduct that $F$ reaches its maximum value ($F = 1$) if every edge is unique and occurs only once in time. It shows how connectivity patterns within the network fluctuate across time. Furthermore, the measure of fluctuability can be defined at the nodal level as follows:

$$F_i = \frac{\sum_j U(a_{ij})}{\sum_j a^t_{ij}}$$  \hfill (15)

The Volatility, suggested in [42], allows to quantify the temporal order of the connectivity not taken into account by the fluctuability. This global measure indicates how volatile the temporal network is over time. It is defined as

$$V = \frac{1}{T-1} \sum_{t=1}^{T-1} D(G_t, G_{t+1})$$  \hfill (16)

where $D(G_t, G_{t+1})$ is a distance function that measures the difference between two snapshots of graph $G_t$ and $G_{t+1}$. In [42], the authors use the Hamming distance. It allows to manipulate binary data. This measure is extensible at an edge level. It gives

$$V_{ij} = \frac{1}{T-1} \sum_{t=1}^{T-1} D(a^t_{ij}, a^{t+1}_{ij})$$  \hfill (17)

Figure 6 illustrates the main differences between fluctuability and volatility in a sequence of time step.

- **Burstiness**

The nature of inter-event time elapsed between two consecutive occurrences of events have been studied in different natural phenomena such as earthquake [43], neuronal firing [44], in/out-going phone call sequence of an individual [45], packets in network traffic [46], and other complex systems [47]. These phenomena have been studied as temporal stochastic processes. Previous studies used Poisson process as a reference model to study the characteristic inter-event or inter-contact time that corresponds to a random walk process. Barabasi and Goh [6] were the first authors to distinguish the heterogeneity of inter-event time in many dynamic systems. This Burstiness effect is characterized by the apparition of numerous activities of events over a short period followed by a significant interval of pause time before reappearing. Most empirical data exhibit the burstiness effect characterized by the heavy-tailed distribution of inter-events times [47].

In order to formally study the burstiness phenomenon in temporal networks, let us consider the sequence of events or contacts $\{e_1, e_2, \ldots, e_i \ldots\}$ that occurs in a sequence of time step $t_1, t_2, \ldots, t_i, \ldots$ . We define the $i$-th inter-event or inter-contact time by

$$\tau_i = t_{i+1} - t_i$$  \hfill (18)

1The Hamming distance is the sum of the difference between the adjacency matrices of two snapshots of graph $G_t$ and $G_{t+1}$.
FIGURE 6. Fluctuability and volatility: The temporal networks shown in panels A, B, and C all have the same numbers of nodes and edges, but they differ in fluctuability and volatility. (A) This network has low fluctuability ($F = 0.125$) and volatility ($V = 0.73$). (B) This network has the highest volatility ($V = 2.55$) of all three networks and a fluctuability ($F = 0.25$) equal to that of the network in panel C. (C) This network has lower volatility than B ($V = 1.27$) but equal fluctuability ($F = 0.25$) [42].

We obtain a sequence of inter-event time $ICT(\tau) = \{\tau_1, \tau_2, \ldots, \tau_i, \ldots\}$. It is easy to study the probability $P(\tau)$ distribution of $ICT(\tau)$, if we don’t take into account the temporal order of this sequence. The inter-event time is a Poisson process if the probability density function is given by:

$$P(\tau) = \lambda e^{-\lambda \tau},$$

(19)

where $\lambda$ is the event or contact rate per time interval. However, the bursty phenomenon is characterized in several data sets by the power-law distribution [47], [48].

$$P(\tau) \propto \tau^\alpha,$$

(20)

where $1 < \alpha < 3$. Alternatively, one can study the bursty activity by using the Burstiness coefficient ($B$) defined as a function of the coefficient of variation $CV = \sigma_\tau / \langle \tau \rangle$. $B$ is given by [6].

$$B = \frac{CV - 1}{CV + 1} = \frac{\sigma_\tau - \langle \tau \rangle}{\sigma_\tau + \langle \tau \rangle},$$

(21)

where $\sigma_\tau$ and $\langle \tau \rangle$ are the standard deviation and the mean of $\tau$, respectively. $B$ is equal to $-1$ for periodic sequence event, $\sigma_\tau = 0$. $B$ is 0 if the process is Poisson, with $\sigma_\tau = \langle \tau \rangle$. We observe burstiness where event sequence is far from Poisson process. We obtain $B > 0$ i.e $\sigma_\tau \rightarrow \infty$.

Similarly, the Local Variation is used alternatively to analyze the burstiness activity in temporal network data. This measure has been proposed by [49] to analyze the neuronal spike in the brain. The LV coefficient is defined as

$$LV = \frac{3}{n - 1} \sum_{i=1}^{n-1} \left( \frac{\tau_i - \tau_{i+1}}{\tau_i + \tau_{i+1}} \right)^2$$

(22)

where $3/(n - 1)$ is normalization factor.

IV. DATA-DRIVEN TEMPORAL NETWORKS

In this section, we present and analyze the topological structure of real-world temporal contact networks. The data are gathered from Sociopattern project and Copenhagen Networks Study (CNS).[^2][^3]

[^2]: [https://www.sociopatterns.org](https://www.sociopatterns.org)
[^3]: [https://figshare.com/articles/The_Copenhagen_Networks_Study_interaction_data/7267433](https://figshare.com/articles/The_Copenhagen_Networks_Study_interaction_data/7267433)
A. DATA

1) SOCIOPATTERN DATASETS
Sociopattern is a collaborative project that aims to collect longitudinal data from face-to-face interactions and proximity contact in various environments (school, hospital, conference, etc.). These data are collected with wearable radio frequency identification (RFID) badges. They are used in different real-world applications like transmission of infectious diseases, contact pattern detection, social network analysis, etc. In this study, we just focus on three datasets: Primary School (PS), Conference of Société Française d’Hygiène Hospitalière (SFHH) and the building of the Institut de Veille Sanitaire (InVS).

- **Primary School (PS)** contains proximity contacts between children and teachers [7]. It concerns 232 children and 10 teachers in a primary school in Lyon, France, and covers two days of school activities (Thursday, October 1st and Friday, October 2nd, 2009) [7]. Every row represents the active contacts during 20-second intervals of the data collection. Each line has the form “t i j Ci Cj”, where i and j are the anonymous IDs of the persons in contact, Ci and Cj are their classes, and the interval during which this contact was active is \([r – 20s, t]\).

- **Société Française d’Hygiène Hospitalière (SFHH) dataset**: These data are collected during two days from 405 participants at the SFHH conference in Nice, France (June 4-5, 2009) [8], [50]. The data collection is done almost under the same conditions as PS. The proximity distance between users is almost \(2\) meters. Each RFID signal is recorded after 20 seconds. The data were collected from 9 am to 9 pm the first day and from 8.30 am to 4.30 pm the second day.

- **Institut de veille sanitaire (InVS) dataset**: The building of InVS is composed of five departments that have a total of 145 candidates. Individuals are considered in contact when the distance is less than 1.5 meters. The records are taken from RFID badges after 20 seconds of contacts [9]. The data have been collected from June 24 to July 3, 2013.

2) COPENHAGEN NETWORKS STUDY (CNS)
This study traces the activities of 700 students for four weeks. During the experiment conducted by Lehman and colleagues [10] all the participants are at the Technical University of Denmark (DTU). The data collection, described in [10], includes multiple types of traces such as Bluetooth, phone call, SMS, Facebook. As we are more interested in physical contact and proximity networks, we focus on the Bluetooth dataset. Indeed, Bluetooth signals provide connectivity up to 10 meters. We use the method described in [10] to deduce the relationship between the Received Signal Strength (RSSI) and the physical distance. The contact information is recorded every five minutes.

Table 2 summarizes basic properties of the data under test.

| Datasets          | Size | Resolution | Distance | Devices   |
|-------------------|------|------------|----------|-----------|
| PS                | 242  | 20 sec     | ~1.5 m   | RFID      |
| SFHH              | 405  | 20 sec     | ~2 m     | RFID      |
| InVS              | 145  | 20 sec     | ~1.5 m   | RFID      |
| CNS               | 700  | 300 sec    | ~10 m    | Bluetooth |

Based on the real datasets described above, we define temporal contact networks using the following assumptions.

- we restrict the study to one day.
- For data collected with RFID devices, users are considered in contact if their proximity is less than 2 meters, and the timestamp is 20 seconds.
- For the CNS Bluetooth traces, to solve the distance constraint, we filter all RSSI greater than 80dBm to approximate 2 meters distance [10]. The time-step is the time between two Bluetooth signals (5 minutes).

B. TOPOLOGICAL PROPERTIES

1) NETWORK DENSITY
Figure 7 illustrates the variation of the density of contacts during the day. Before analyzing and comparing the network density, we did some data processing. For comparative purposes, it is convenient to normalize the time resolution to 5 minutes for all datasets. After normalization and removing some missing values, we project the timeline in the same window size, from 09:00h AM to 06:00h PM. In Figure 7, one can notice low densities for all real temporal networks due to their sparsity and disconnection. It appears that the max density of temporal graphs is less than \(3.10^{-3}\) in the PS network during the day. Moreover, the evolution of the time-dependent density observed is not homogeneous and exhibits the regularity of events in different time intervals during the day. This phenomenon is evident in the school place during the student’s recreation and descent time. For instance, in the PS network, the density of contacts increases between 10h-30 - 11h, 12h, and 16h. In the other environments (conference and workplace), higher density is observed at the break hours. In the CNS Bluetooth network, we also notice a higher density of contacts during class hours.

We use the Dynamic Time Warping (DTW) method [51], [52] to study the similarity and dissimilarity between the time series of network density. This method allows comparing two sequences by minimizing the effects of shifting and distortion in time by allowing “elastic” transformation of time series to detect similar shapes with different phases along the time [52]. Table 3 reports the DTW distances between the networks under study. It appears that the evolution of the density of the SFHH network is close to the InVS network. Moreover, the time-dependent density computed in PS and InVS temporal graphs are similar. We also find that the distance calculated with CNS time-dependent density is higher than the other temporal networks.
Figure 8 represents the accumulated cost matrix built from the local pairwise distances along the time series X and Y. The contour lines and colors show the cost values from the smallest to the biggest (from red to green). The warping path or the alignment path runs through the low-cost areas on the cost matrix to minimize the sum of distances between aligned elements (blue line). The optimal warping paths reported in Figure 8 measure how two time-dependent densities patterns are similar. Figure 8 (f), reveal the smallest warping path between SFHH and InVS. There is also a small warping path between PS and InVS (Figure 8 (c), and between PS and CNS (Figure 8 (b)).

In general, this measure does not give all the details observed in the temporal structure of the contact network. However, it provides an overview of the emergence of human behaviors. These results corroborate the cumulative distance reported in Table 3.

### TABLE 3. Dynamic time warping: The minimum cumulative distance measures between time-series.

| Distances       | Unnormalized | Normalized  |
|-----------------|--------------|-------------|
| d(PS, InVS)     | 0.068        | 2.775E-04  |
| d(PS, SFHH)     | 0.089        | 3.717E-04  |
| d(PS, CNS)      | 0.085        | 2.188E-04  |
| d(SFHH, InVS)   | 0.014        | 5.269E-05  |
| d(SFHH, CNS)    | 0.106        | 2.518E-04  |
| d(InVS, CNS)    | 0.080        | 1.877E-04  |

### 2) REACHABILITY LATENCY

The reachability latency defined by Thompson et al. [42] is an easy way to know the time to reach a given proportion of nodes. In this study, we calculate the $R_1$ and $R_{0.5}$ measures for the ratio $r = 1$ (i.e. the time to get 100% of nodes) and $r = 0.5$ (i.e the time to get 50% of nodes). Results are shown in the Table 12. Table 12 reports the $R_1$ and $R_{0.5}$ values in the various temporal networks. All nodes in the CNS Bluetooth network can be reached with the lowest $R_1$ value ($R_1 = 113.60TU$)\(^4\). Note that it takes almost one fifth of the $R_1$ value to reach 50% of the nodes ($R_{0.5} = 23.32TU$). The temporal proximity networks based on Sociopattern exhibit higher values of $R_1$. The smallest duration in which every node can join all other nodes in SFHH temporal network is greater than 800TU. Moreover, 50% of nodes can be reached approximately between 1/3 and 1/4 of the temporal diameter $R_1$.

The observed differences are linked to the network’s properties, such as the number of participants, the number of contacts per user, the contact duration, the inter-contact time, and the social relationship. It also depends on the mobility pattern of the individuals and the constraint of their spatial environment.

### 3) SMALL-WORLD IN TEMPORAL NETWORKS

A higher value of temporal efficiency and a higher value of temporal correlation than a randomized reference model (i.e., null models) [40] is characteristic of a small-world behavior in time-varying systems.

Based on the state of the art of Microcanonical Randomized Reference Models (MRRs) defined by Gauvin et al. [53], we generate events permutation null models such that the time stamps of events are randomly permuted between $t_i$ and $t_f$ among all timelines.

- **Temporal Efficiency**: Table 5 reports the temporal efficiency values for the networks under study. These values are relatively small in Sociopattern contact networks. They range between $E = 0.001$ and $E = 0.0025$. The CNS Bluetooth network exhibit a higher value of temporal efficiency ($E = 0.010$). It means that temporal paths are shorter in the CNS Bluetooth than in the Sociopattern proximity networks. Except for the CNS Bluetooth network, the temporal efficiency measured in real contact networks are less important than their corresponding null models (Table 5). For instance, the temporal efficiency measured in PS contact networks is $E = 0.0038$ against $E_{Random} = 0.0082$, and $E = 0.0017$ against $E_{Random} = 0.0024$ in InVS contact network.

- **Topological Overlap and Temporal Correlation**: Figure 9 shows the violin plot as the comparison of the distribution of the topological overlap measured in the real networks. One can observe that almost 75% of the topological overlap values in InVS and SFHH networks are highly concentrated around the median with some extreme values. In the PS network, the topological overlap distribution appears to be multimodal. The 98% of the topological overlap values PS are less than 0.2. The distribution of CNS Bluetooth network values has long tail with a positive skewness. The max value is greater than 0.7. Moreover, the Table 5 allows us to compare the temporal correlation (noted by $TC$) computed in real temporal contact networks (PS, InVS, SFHH, CNS Bluetooth) and their corresponding null models. The results reported in the Table 5 prove that there are higher temporal correlation values compared the randomized reference models. The CNS Bluetooth network has the highest value $TC = 0.073$ against $TC_{Random}$, followed by the PS network ($TC = 0.049 > TC_{Random} = 6.489E - 6$). One can conclude that the small-world property is more probable in the CNS Bluetooth network than other temporal networks.

### TABLE 4. Reachability latency with ratio $r = 1$ and $r = 0.5$.

| Datasets      | $R_1$ (ratio=1) | $R_{0.5}$ (ratio=0.5) |
|---------------|----------------|----------------------|
| CNS Bluetooth | 113.60         | 23.32                |
| PS            | 595.19         | 175.15               |
| InVS          | 674.72         | 157.92               |
| SFHH          | 824.55         | 249.35               |

\(^4\)Time Unit.
that the temporal correlation coefficients measured in these temporal networks are very small compared to the datasets studied in [40] and [54].

4) OCCURRENCE OF PAIR OF CONTACTS
The occurrence of pair of contacts is an essential characteristic in temporal networks. It measures how many times an
individual A meet an individual B. Commonly, its distribution is no homogeneous. Indeed, most individuals maintain strong relationships with a limited number of individuals, while few individuals have a high number of contacts. To fit the empirical distributions, we use the methodology described in [55] implemented by Jeff Alstott in the powerlaw Python package. Figure 10 presents the distribution of pair of contacts in a log-log scale for the network under test. Table 6 gives the expressions of heavy-tailed distributions used to fit the empirical distributions. The Kolmogorov-Smirnov test (KS) measures the goodness of fit of the various theoretical distribution. According to KS values reported in Table 7, the power law (PL) distribution suggests that some individuals spend more time with few individuals while the vast majority of interactions with others occur in a short delay.

### TABLE 6. Some heavy tailed distributions.

| Distribution                | Expression of $p(x)$ | Abbreviation |
|----------------------------|-----------------------|--------------|
| Power Law                  | $C x^{-\alpha} e^{-\lambda x}$ | PL           |
| Truncated Power Law        | $C x^{-\alpha} e^{-\lambda x}$ | TPL          |
| Stretched Exponential      | $C x^{\beta-1} e^{-\lambda x^{\beta}}$ |             |
| Log-Normal                 | $\frac{1}{2\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ | S Exp        |

5) **CONTACT DURATION DISTRIBUTION**

The contact duration between users is an essential element in an epidemic process because it increases the probability of infection. It is the reason why we choose to estimate its distribution in the temporal proximity graphs. Figure 11, show that the best fit for the distribution of contact duration is obtained with the power law and the truncated power law. One can see that the power law is the best fit for SFHH and CNS Bluetooth data. The truncated power law (TPL) distribution outperforms the power law in the PS and InVS data. The KS-tests reported in the Table 8 confirm these empirical results. Furthermore, these results corroborate precedent studies carried out on other real contact data [56], [57]. Indeed, the power-law distribution suggests that some individuals spend more time with few individuals while the vast majority of interactions with others occur in a short delay.
6) INTER-CONTACT TIME DISTRIBUTION

The inter-contact time between pairs of nodes is the duration of elapsed time between two successive links. Figure 12 presents the distribution of the inter-contact times for the temporal graphs of PS, InVS, SFHH, CNS Bluetooth in a log-log scale. It shows that the empirical distribution are heavy-tailed. We can see that power-law and truncated power-law present to be the best fits of inter-contact time distribution for the Sociopatterns datasets. Their exponents $\alpha$ values are between 1 and 2. However, the log normal distribution reveals to be a better fit for CNS Bluetooth. Results of the KS goodness of fit test confirming these results are reported in the Table 9.

7) BURSTINESS

Burstiness is measured by the burstiness coefficient $B$ or the local variation $LV$. Hence, we compute these two coefficients in the pairs of contacts of each temporal graph. Typical values

![Figure 10: Distribution of occurrence of pair of contacts in a Log-log scale plot for the temporal networks primary school (PS), Institut de veille sanitaire (InVS), Société Française d’Hygiène Hospitalière (SFHH) and CNS Bluetooth network.]

| Datasets     | Exp | Log-normal | PL | S Exp | TPL |
|--------------|-----|-------------|----|-------|-----|
| PS           | 0.345 | 0.103       | **0.064** | 0.109 | 0.094 |
| InVS         | 0.399 | 0.098       | **0.053** | 0.105 | 0.090 |
| SFHH         | 0.403 | 0.133       | **0.050** | 0.134 | 0.108 |
| CNS Bluetooth| 0.425 | 0.091       | **0.051** | 0.097 | 0.076 |

![Table 7: Kolmogorov Smirnov (KS-test) values for distribution of occurrence of pair of contacts measured in real contact datasets. The results showed the power law (PL) as the best fit, but the truncated power law (TPL) reveals to be close to the PL distribution.]

| Datasets     | Exp | Log-normal | PL | S Exp | TPL |
|--------------|-----|-------------|----|-------|-----|
| PS           | 0.093 | 0.083       | 0.110 | 0.095 | **0.072** |
| InVS         | 0.182 | 0.094       | 0.082 | 0.103 | **0.077** |
| SFHH         | 0.229 | 0.114       | **0.070** | 0.312 | 0.091 |
| CNS Bluetooth| 0.202 | 0.100       | **0.069** | 0.107 | 0.085 |

![Table 8: Kolmogorov Smirnov (KS-test) values for distribution of contacts duration measured in real contact data. These distribution are heavy tailed and the PL and TPL show the best fit of contact duration.]

TABLE 7.

| Datasets     | Exp | Log-normal | PL | S Exp | TPL |
|--------------|-----|-------------|----|-------|-----|
| PS           | 0.345 | 0.103       | **0.064** | 0.109 | 0.094 |
| InVS         | 0.399 | 0.098       | **0.053** | 0.105 | 0.090 |
| SFHH         | 0.403 | 0.133       | **0.050** | 0.134 | 0.108 |
| CNS Bluetooth| 0.425 | 0.091       | **0.051** | 0.097 | 0.076 |

TABLE 8.

TABLE 9.

6) INTER-CONTACT TIME DISTRIBUTION

The inter-contact time between pairs of nodes is the duration of elapsed time between two successive links. Figure 12 presents the distribution of the inter-contact times for the temporal graphs of PS, InVS, SFHH, CNS Bluetooth in a log-log scale. It shows that the empirical distribution are heavy-tailed. We can see that power-law and truncated power-law present to be the best fits of inter-contact time distribution for the Sociopatterns datasets. Their exponents $\alpha$ values are between 1 and 2. However, the log normal distribution reveals to be a better fit for CNS Bluetooth. Results of the KS goodness of fit test confirming these results are reported in the Table 9.

7) BURSTINESS

Burstiness is measured by the burstiness coefficient $B$ or the local variation $LV$. Hence, we compute these two coefficients in the pairs of contacts of each temporal graph. Typical values
for a bursty phenomenon are $B \to 1$ or $LV$ in the range from 1 to 3. Figure 13 reports on the y-axis the measures of $B$ values and on the x-axis LV values. For PS, InVS, and SFHH contact networks, most values of $B$ are in the interval (0, 1]. Half of the data points of the bursty coefficient in the CNS Bluetooth network are almost concentrated around 0, but most values are positives. Analyzing the burstiness phenomenon along the y-axis shows that the LV values computed in the PS network are close to 1. However, it is not always true for the CNS Bluetooth networks. To sum up, these results confirm the burstiness behavior of the data, especially in Sociopatterns.

8) FLUCTUABILITY AND VOLATILITY

Table 10 reports the fluctuability and volatility values measured in the four temporal networks. It shows low fluctuability in PS and CNS networks ($F \simeq 0.10$). SFHH temporal contact network has the highest fluctuability value ($F \simeq 0.12$). Overall, these values are small. Consequently, we have a lower temporal diversity of edge connectivity. In other words, contacts between individuals do not change frequently. The volatility measures vary in the opposite direction of fluctuability. The PS network with the lowest fluctuability has the highest volatility ($V = 17.10^{-4}$). The SFHH contact network shows very small volatility ($V \simeq 17.10^{-5}$). Despite the higher dynamic of edge/contact between individuals, the rate of connectivity changes much slower through time in real proximity networks than in brain networks [42].

V. SYNTHETIC CONTACT NETWORKS

A. SYNTHETIC DATASETS BASED ON MOBILITY MODELS

1) SYNTHETIC TEMPORAL NETWORKS

We use traces generated by the mobility models described in section III-A (Random Waypoint, Gauss-Markov, Truncated

---

**TABLE 9.** Real temporal contact networks: Kolmogorov Smirnov (KS-test) values for the distribution inter-contact time.

| Datasets          | Exp | Log-normal | PL  | S Exp | TPL  |
|-------------------|-----|------------|-----|-------|------|
| PS                | 0.525 | 0.048       | 0.027 | 0.059 | 0.049 |
| InVS              | 0.690 | 0.100       | 0.082 | 0.155 | 0.114 |
| SFHH              | 0.609 | 0.061       | 0.045 | 0.099 | 0.073 |
| CNS Bluetooth     | 0.131 | 0.029       | 0.060 | 0.031 | 0.032 |

---

**FIGURE 11.** Distributions of contact duration in a Log-log scale for the temporal networks primary school (PS), Institut de veille sanitaire (InVS), Société Française d’Hygiène Hospitalière (SFHH) and CNS Bluetooth network.
Levy Walk, Spatio-Temporal Parametric Stepping). These models are used as a proxy to build temporal networks based on the contacts and proximity between agents. A node is associated with a trace. There is a link between two nodes when the distance between the associated traces is lower than a threshold distance denoted $d_{th}$. The configuration of the temporal contact network changes at every time-step of the mobility models. The parameter values used for each model are reported in Table S1 of the Supplementary Materials. We conduct a comparative analysis of the topological characteristics of the temporal networks associated with the mobility models under investigation. For each experiment, we choose 200 nodes following one of the mobility models defined above. We notice that nodes move in grid area of $100 \times 100$. For TLW and STEPS mobility models, we reproduce the same condition suggested by their authors.

### B. TOPOLOGICAL PROPERTIES

#### 1) DENSITY

Figure 14 represents the evolution of the density of each temporal graph obtained from traces of the synthetic mobility models. We define the timeline as the time-step of the mobility models. In Figure 14, we observe higher density values (max density $D \sim 0.0018$) in the temporal network generated by RWP, compared to the other graphs. Therefore, we can deduce that the time-varying graph of RWP is more connected than GM, TLW, and STEPS. This higher connectivity can be justified because mobile agents following the RWP model tend to concentrate often in the middle of the simulation space over time. This observation corroborates previous observations made by Bettstetter in [12]. The density of the GM, TLW, and STEPS synthetic temporal networks exhibit many fluctuations with different peaks. In contrast, the temporal density evolution in the RWP network tends to be more homogeneous.

The dynamic Time Warping (DTW) describes in Section IV-B1 is used to compare the density of the synthetic temporal graphs. Results reported in the table 11 show the accumulated DTW distance. It shows a high similarity between TLW, STEPS, and GM networks. The DTW distances between RWP networks and the other synthetic graphs show that the patterns found in the RWP density series are far
FIGURE 13. Burstiness: Bursty coefficient and Local Variation measured in real contact datasets.

from the other series. Alternatively, the warping path shown in Figure 15 allows us to better understand this observation. Figures 15a, 15e, and 15f show the optimal warping path with low costs. Therefore, these results confirm the similarity between the patterns exhibit by the time-dependent density measure in TLW, STEPS, and GM temporal graphs. Nevertheless, we don’t have the same optimal alignment between the RWP density, and the other densities computed in TLW, STEPS, and GM networks (Figures 15b, 15c, and 15d).

TABLE 11. Dynamic time warping: The minimum cumulative distance measures between densities of temporal synthetic networks.

| Distances          | Unnormalized | Normalized  |
|--------------------|--------------|-------------|
| $d(RWP, TLW)$      | 0.0299       | 1.495E-4    |
| $d(RWP, GM)$       | 0.0286       | 1.431E-4    |
| $d(RWP, STEPS)$    | 0.0330       | 1.461E-4    |
| $d(TM, TLW)$       | 0.0061       | 3.083E-05   |
| $d(TM, STEPS)$     | 0.0057       | 2.891E-05   |
| $d(TM, STEPS)$     | 0.0052       | 2.612E-05   |

2) REACHABILITY LATENCY

Table 12 reports the $R_1$ and $R_{0.5}$ values in temporal contact networks originating from synthetic data. The STEPS-based temporal network has the lowest temporal diameter ($R_1 = 10.14$ TU). Its $R_{0.5}$ value is the half of $R_1$. Therefore, a dynamic process spread faster in STEPS temporal networks. We get the with the RWP based temporal network model the higher reachability latency $R_1$ value ($R_1 = 45.33$ TU). However, one can observe that it takes around one-fifth of this time to reach 50% of the nodes (5.91 TU). All nodes are reached in 21.19 TU in the GM network, and it takes almost half the time to reach 50% of nodes (9.91 TU). Therefore, the relationship between $R_1$ and $R_{0.5}$ is not necessarily linear. Overall, it takes less time to reach all the nodes in these synthetic networks than in real temporal networks.

TABLE 12. Reachability latency measured in mobility-based datasets with ratio $r = 1$ and $r = 0.5$.

| Datasets | $R_1$(ratio=1) | $R_{0.5}$(ratio=0.5) |
|----------|----------------|----------------------|
| RWP      | 45.33          | 5.91                 |
| GM       | 21.19          | 9.91                 |
| TLW      | 35.30          | 12.5                 |
| STEPS    | 10.14          | 5.01                 |

3) SMALL-WORLD

We also investigate the small-world behavior in the temporal proximity networks generated using the RWP, GM, TLW,
and STEPS mobility models. As reported in Section IV-B3, we use the null models as reference to study the small-world property in synthetic temporal contact networks. We also use the same method to generate randomized reference null models [53].

- **Temporal Efficiency:** Table 13 reports its values in the networks based on the mobility model dataset. STEPS temporal network has the best score ($E \approx 0.17$). It is followed by the network originating from the RWP mobility model ($E \approx 0.14$). This result suggests that, on

**FIGURE 14.** Series of time-dependent density measured in different synthetic networks: RWP, GM, TLW and STEPS.

**FIGURE 15.** Cost matrix heatmap and optimal warping path between time-dependent network density computed in the synthetic temporal contact networks. Each Figure measures the pattern similarity between series of time-dependent network densities. The contour lines (and colors) show the cost values from the smallest to the biggest (from red to green). The blue line represents the optimal time warping path. (a) Warping path between TLW and STEPS network densities. (b) Warping path between RWP and STEPS network densities. (c) Warping path between RWP and TLW network densities. (d) Warping path between RWP and GM network densities. (e) Warping path between GM and STEPS network densities. (f) Warping path between GM and TLW network densities.
average, the temporal shortest paths are shorter in STEPS networks compared to the alternatives. Compared to its random null model, the synthetic contact networks based on STEPS model exhibit low temporal efficiency: $\bar{E} = 0.170 < \bar{E}_{\text{Random}} = 0.899$ (see Table 13). Same remark is valid for the other synthetic networks according to the results reported in Table 13 (RWP: $\bar{E} < \bar{E}_{\text{Random}}$, GM: $\bar{E} < \bar{E}_{\text{Random}}$, TLW: $\bar{E} < \bar{E}_{\text{Random}}$).

- **Topological overlap and Temporal Correlation:**
  Figure 16 represents the distribution of the synthetic networks in violin plots. STEPS temporal network, exhibit a bimodal distribution. The modal values are around 0.3 and 0.5. Indeed, 95 % of the topological overlap values are smaller than 0.5 in STEPS temporal contact network. In TLW, we get large values concentrated around the mean that is greater than 0.9. In the violin plot associated with the topological overlap computed in GM network, we observe a symmetric distribution that is close to Gaussian. The vast majority of values are in the range [0.6, 0.8]. From the topological overlap values, we can deduct the temporal correlation as the average. Hence, table 16 shows the comparison of temporal correlation measured in temporal graphs generated from mobility models and their corresponding null models. Unlike temporal efficiency, it appears that the temporal correlation values measured in synthetic temporal networks are higher compared to their null model counterparts.

According to the results obtained, one cannot confirm the small world phenomenon in temporal networks based on these synthetic mobility models.

### Table 13. Small world behavior in synthetic temporal contacts networks: temporal correlation and temporal efficiency measures.

| Datasets | $TC$ | $\bar{E}$ | $T\bar{C}_{\text{Random}}$ | $\bar{E}_{\text{Random}}$ |
|----------|------|----------|-----------------|-----------------|
| RWP      | 0.887| 0.141    | 0.043           | 0.648           |
| GM       | 0.718| 0.096    | 0.020           | 0.542           |
| TLW      | 0.956| 0.076    | 0.119           | 0.369           |
| STEPS    | 0.388| 0.170    | 0.005           | 0.899           |

Overall, whatever fit, all the empirical distributions are non-homogeneous with heavy tails.

### Table 14. Kolmogorov Smirnov (KS-test) values for distribution of occurrence of pair of contacts in synthetic contact networks.

| Datasets | Exp | Log-normal | PL | S Exp | TPL |
|----------|-----|------------|----|-------|-----|
| RWP      | 0.053| 0.008      | 0.083| 0.012 | 0.016 |
| GM       | 0.036| 0.004      | 0.093| 0.015 | 0.016 |
| TLW      | 0.017| 0.037      | 0.177| 0.018 | 0.106 |
| STEPS    | 0.065| 0.017      | 0.096| 0.040 | 0.123 |

### Table 15. Kolmogorov Smirnov (KS-test) values for the distribution of contact duration in synthetic contact networks.

| Datasets | Exp | Log-normal | PL | S Exp | TPL |
|----------|-----|------------|----|-------|-----|
| RWP      | 0.0531| 0.0126     | 0.0861| 0.0188| 0.0224 |
| GM       | 0.0073| 0.0144     | 0.1069| 0.0070| 0.0362 |
| TLW      | 0.0266| 0.0191     | 0.1949| 0.0179| 0.1156 |
| STEPS    | 0.0398| 0.0195     | 0.0665| 0.0420| 0.0564 |

### 6) INTER-CONTACT TIME

Previous studies show that the heavy-tailed distribution behavior is characteristic of temporal networks originating from various mobility models. Abdulla and Simon [56] uncovered an exponential distribution of inter-contact time in synthetic networks using the RWP mobility model. The authors of STEPS [16] show that the inter-contact time distribution follows a power law with an exponential decay.

Figure 19 reports the empirical inter-contact time distribution and the theoretical distributions estimates to fit these data. The KS distances quantifying the distribution fit are reported in Table 16. It appears that the stretched exponential distribution is the best fit for RWP and TLW networks. The inter-contact time of GM model is well estimated by the log-normal distribution ($\mu = 2.07$ and $\sigma = 1.16$). Finally, the Power Law distribution appears as the best fit in the network related to the STEPS model with exponent $\alpha = 1.42$. Note that the log normal might be a good fit for all distributions with very close KS values as compared to the best fit.
To summarize, the results of our investigations confirm the general trend reported in the literature.

**7) BURSTINESS**

Figure 20 shows the measurements of B and LV in synthetic temporal networks. B tending to one is characteristic of bursty phenomenon. It is the case for temporal networks based on STEPS and GM. Indeed, there is a high number of values above zero. In RWP and TLW networks, several values of the bursty coefficient are positive. According to the Local variation reported in the y-axis, the STEPS network is the one that shows the burstiness. Most values associated with the RWP and TLW graphs are very close to 0, showing a periodic behavior (Figure 20). These results demonstrate two things. First, the burstiness highlighted by the local variation measure corroborates the power-law distribution of inter-contact observed in the STEPS network. Second, the local variation (LV) seems to be more effective than the burstiness coefficient (B) to uncover bursty phenomenon in temporal contact networks.

**8) FLUCTUABILITY AND VOLATILITY**

Table 17 reports those measures for the four synthetic networks under test. Temporal networks built from TLW and RWP mobility models exhibit small fluctuability values. Fluctuability is more important in STEPS than in GM temporal network ($F_{GM} = 0.17, F_{STEPS} = 0.31$). These measurements clearly show that the edge fluctuation is greater for STEPS than for the other networks. Volatility values are small in all the generated temporal graphs. Therefore, those slow changes occur in network connectivity over time. The STEPS network is the most dynamic.

**VI. COMPARATIVE ANALYSIS OF REAL AND SYNTHETIC TEMPORAL CONTACT NETWORKS**

In this comparative study, we focus on the main properties such as the density, the reachability, the efficiency, the
temporal correlation, the fluctuability, the volatility, the burstiness, the contact duration, the inter-contact events, the occurrence of pair of contacts to analyze the topological structure of the temporal networks based on proximity/contact datasets to understand the typical points of similarity and dissimilarity between real and synthetic temporal networks. However, there exists several measurements to characterize the topological properties of temporal networks \[35\], which are out of our study. We summarize our empirical results in the Table 18.

### A. NETWORK DENSITY

In general, the density of contact in real temporal networks follows circadian rhythms. One observes peaks of contacts during the day, especially during break times, a periodic decrease in the frequency of contacts during nights, and specific Points of Interest. It also appears that individuals tend to gather in clusters or communities during break times for discussions, food, or games.

Through our empirical analysis, we discover that the time-varying networks built on real data of contacts have in common different characteristics. We found that the temporal evolution of density shows the highly sparsity and disconnection in networks.

According to DTW analysis, we found that SFHH, InVS and PS have some similar patterns unlike CNS density. In our opinion, this fact is mainly due to the characteristics of human mobility to socio-environmental constraints. Because CNS Bluetooth datasets are collected in a large environment like campus, people does not follow any regularity mobility pattern during the day as observed in Sociopattern data.

Contrary to the real data, the synthetic contact networks are denser, but do not show the patterns of human behaviours. In these models, we can see that the agents have a very high mobility. The temporal connectivity depends intrinsically to the properties of the mobility model. We also discover that the patterns of the RWP network have strong dissimilarity with GM, TLW and STEPS synthetic networks.

### B. REACHABILITY LATENCY

However, the maximum and the minimum delay required to broadcast a message to the percentage of nodes is crucial in the context of delay-tolerant networks (DTN) or the delay such that an epidemic reaches a ratio of the population.

The reachability latency found in real contact networks is more important than those measured in synthetic datasets (Table 18). This high reachability latency proves the low connectivity and the small temporal paths discovered in real temporal networks compared to the synthetic temporal networks. Another promising finding is that there is not
necessarily a relationship between the reachability latency measured and the percentage of reachable nodes. In other words, the percentage to reach 50% nodes is not proportional to the percentage to get 100% of nodes.

C. SMALL-WORLD

Furthermore, results reveal that the temporal efficiency measured in real contact datasets are lower than their corresponding null models except for the CNS Bluetooth network. However, we find higher temporal correlation values compared to their randomized reference models. Hence, we can conclude that only CNS Bluetooth temporal networks exhibit the small-world effect.

Compared to real temporal networks, the synthetic networks are generally characterized by an important temporal efficiency (Table 18). Nevertheless, the values are smaller than their null models. Moreover, we observe that the temporal correlation measured in temporal networks generated from mobility models (RWP, GM, TLW, STEPS) is more important compared to their corresponding null models. According these results, the synthetic temporal networks do not display the small-world property.
**D. Occurrence of Pair Contacts**

An important characteristic of real contacts in temporal networks is the occurrence of pair of contacts that quantifies the number of repeat contacts between pairs of nodes. This measure allows to understand how tied a relationship is for instance the friendship, co-worker relationships, the sexual contacts and so on. It appears that the distribution of the occurrence of pair of contacts in the real-world networks follow a power-law. This phenomenon is also valid for the STEPS network. However, the RWP, GM and STEPS are better fitted with a log-normal distribution, and they show closely this phenomenon.

**E. Contact Duration**

In our experimental results, we found that the contact duration (CD) of real contact networks is power law and truncated power law distributed. By comparing to the distribution obtained in synthetic datasets, the results don’t show this behaviour. Thus, the contact duration of RWP and STEPS networks are approximately log-normal distributed while GM and TLW obey the stretched exponential distribution. However, the contact duration of RWP and STEPS can be closely fitted to PL with an acceptable KS-test.

**F. Inter-Contact Time**

In this study, the analysis of distributions of inter-contact time confirms that most empirical datasets show power law (see Figure 12). The origin of this heavy tailed distribution is due to the burstiness activities of the contact events. The exponential distribution corresponds to the Poisson process of contact events. We find that only STEPS mobility model shows the PL distribution of inter-contact times. The GM is fitted with log-normal distribution, but it is close to power law. The RWP and TLW are fitted with stretched-exponential.

**G. Burstiness**

The burstiness behaviour characterized by the power law distribution of the inter-contact times can be easily deduced from the measurements of the burstiness coefficient ($B$) or the local variation ($LV$). The results obtained clearly indicate the burstiness effect in most of real-world networks. Except for STEPS and GM, the synthetic temporal contact networks do not exhibit the appearance of the burstiness behaviour.

**H. Fluctuability and Volatility**

These two measures quantify the dynamic of edge in temporal networks. In the experiment, we observe no significant fluctuability and volatility in real contact networks (see Table 18). It implies lower diversity and variability of edge connectivity. We notice that when fluctuability increases, volatility decreases. Except for the network generated using the STEPS mobility model, the fluctuability values are not as important in most synthetic contact network compared to real-world networks. However, the volatility values are more important.
than those measure in real contact networks. This fact can be due mobility pattern that follow the nodes in simulation area. To better understand, we plan to analyze the influence of each mobility parameter on these measures.

VII. CONCLUSION

This work investigates differences between real-world contact networks and synthetic temporal networks’ topological structure. Results show that overall synthetic networks deviate from real-world data. Therefore, one can conclude that mobility models are not sophisticated enough to describe human mobility behavior accurately. According to the randomness, contact data generated from the RWP model tend to move away from real data properties. Nevertheless, some results are encouraging. Indeed, it appears that the temporal contact networks generated using the STEPS mobility models exhibit the most similar properties to real-world temporal contact networks. Both share heavy-tailed distribution for the contact duration and the frequency of pairs of contacts. Furthermore, they exhibit the bursty phenomenon. Regarding fluctuability, one does not observe significant differences between real-world and synthetic temporal contact networks. Nevertheless, the variation of edges due to volatility is much slower in real networks. This work allowed us to discover the main properties of human behavior found in real contact networks and to compare them with those found in artificial contact networks to understand the gap that exists between them. Now, this study aims to pave the way to new developments on mobility models that consider more accurately characteristics of human behavior.

REFERENCES

[1] N. Oliver et al., “Mobile phone data and COVID-19: Missing an opportunity?” Mar. 2020, arXiv:2003.12347.
[2] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, “Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing,” Science, vol. 368, no. 6491, pp. 1–8, May 2020. [Online]. Available: https://science.sciencemag.org/content/368/6491/eabb6936
[3] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, “Understanding individual human mobility patterns,” Nature, vol. 453, no. 7196, p. 779, 2008, doi: 10.1038/nature06958.
[4] C. Song, T. Koren, P. Wang, and A.-L. Barabási, “Modelling the scaling properties of human mobility,” Nature Phys., vol. 6, no. 10, p. 818, 2010, doi: 10.1038/nphys1760.
[5] H. Barbosa-Filho, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomassini, “Human mobility: Models and applications,” Sep. 2017, arXiv:1710.00004.
[6] K.-I. Goh and A.-L. Barabasi, “Burstiness and memory in complex systems,” Oct. 2006, arXiv:physics/0610233.
[52] P. Senin, “Dynamic time warping algorithm review,” Dept. Inf. Comput.
Sci., Univ. Hawaii, Honolulu, HI, USA, Tech. Rep. CSDL, Jan. 2009.

[53] L. Gauvin, M. Géniois, M. Karsai, M. Kivelä, T. Takaguchi, E. Valdano, and
C. L. Vestergaard, “Randomized reference models for temporal networks,”
Jun. 2018, arXiv:1806.04032.

[54] K. Büttner, J. Salau, and J. Krieter, “Temporal correlation coefficient for
directed networks,” SpringerPlus, vol. 5, no. 1, p. 1198, Jul. 2016, doi:
10.1186/s40064-016-2875-0.

[55] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distributions
in empirical data,” Jun. 2007, arXiv:0706.1062.

[56] M. Abdulla and R. Simon, “The impact of intercontact time within oppor-
tunistic networks: Protocol implications and mobility models,” 2009.

[57] S. Hong, I. Rhee, S. J. Kim, K. Lee, and S. Chong, “Routing performance
analysis of human-driven delay tolerant networks using the truncated Levy
walk model,” in Proc. 1st ACM SIGMOBILE Workshop Mobility Mod-
els (MobilityModels), New York, NY, USA: Association for Computing
Machinery, 2008, pp. 25–32, doi: 10.1145/1374688.1374694.

DJIBRIL MBOUP received the master’s degree in computer science from the UFR Sciences
Appliquées et Technologies de l’ Information et de la Communication, the master’s degree in data
science from the African Institute of Mathematical Sciences (AIMS), in 2018, and the Ph.D. degree in
computer science from Gaston Berger University, St. Louis, Senegal. His research interests include
from complex networks to data science. He started his research, in January 2015, at the Faculty of
Applied Science and Technologies, UFR Sciences Appliquées et Technologies de l’Information et de la Communication. He has been working as a Data Scientist and a Big Data Engineer at Atos Senegal, since 2017. His research interests include mobility, internal, and external dynamic of complex network, and data analytic using machine learning and deep learning.

CHERIF DIALLO received the master’s degree in computer science (engineering) from Paul Sabatier
University, Toulouse, the master’s degree in computer science (information systems security and auditing) from University Paris 2, the master’s degree in computer science (scientific computing) from University Paris 6, and the Ph.D. degree in computer science from Telecom SudParis (Former INT of Evry, National Institute of Telecommunications). He has been a Professor of computer science at Gaston Berger University, St. Louis, Senegal, since 2012. He worked as a Consultant Engineer, specialist in telecom networks and information systems security, he has led several audit missions of telecom networks and innovative projects on behalf of very large French companies in such diverse industries. He has also directed various courses at Cheikh Anta Diop University, Dakar (UCAD), the Alioune Diop University of Bamby, the African Institute for Mathematical Sciences (AIMS), Senegal, and the Franco-American Academy of Management (AFRAM), Libreville, Gabon. He is recognized as a Professor and the Head of the IT Center at Gaston Berger University and a former member of the National Cryptology Com-
misson (CNC) of Senegal and UN-GGE.

HOCINE CHERIFI received the Ph.D. degree from the National Polytechnic Institute, Grenoble,
in 1984. He has been a Professor of computer science at the University of Burgundy, Dijon,
France, since 1999. Prior to moving to Dijon, he held faculty positions at the University of Rouen
Normandy and Jean Monnet University, France. He has held visiting positions at Yonsei Uni-
versity, South Korea, The University of Western Australia, Australia, the National Pingtung Uni-
versity, Taiwan, and Galatasaray University, Turkey. He has published more than 200 scientific papers in international refereed journals and conference proceedings. His recent research interests include computer vision and com-
plex networks. He held leading positions in more than 15 international conference organization (the general chair and the program chair), and he served in more than 100 program committee. He is the Founder of the International Conference on Complex Networks and Their Applications. Currently, he is a member of the Editorial Board of Computational Social Networks, PLOS One, IEEE Access, Journal of Imaging, Complex Systems, Quality and Quantity, and Scientific Reports. He is the Founding Editor-in-Chief of the Applied Network Science journal.

VOLUME 10, 2022