Following the right path: Using traces for the study of DTNs

G. Baudic*, T. Perennou, E. Lochin

ISAE-SUPAERO, Université de Toulouse, 10 avenue Edouard Belin, 31055 TOULOUSE Cedex 4, France

A B S T R A C T

Contact traces collected in real situations represent a popular material for the study of a Delay Tolerant Network. Three main use cases can be defined for traces: social analysis, performance evaluation and statistical analysis. In this paper, we perform a review on the technicalities of real trace collection and processing. First, we identify several factors which can influence traces during collection, filtering or scaling, and illustrate their impact on the conclusions, based on our experience with four datasets from the literature. We subsequently propose a list of criteria to be verified each time a trace is to be used, along with recommendations on which filters to apply depending on the envisioned use case. The rationale is to provide guidelines for researchers needing to perform trace analysis in their studies.

Keywords:
DTN
Real traces
Datasets
Trace collection
Statistical analysis

1. Introduction

Recent years have seen a major growth in the number of mobile devices, almost all providing network connectivity through NFC, Bluetooth or WiFi to name a few. This has made both opportunistic and Delay Tolerant Networks major topics of interest. Assessing the performance of such networks is a necessary step towards their deployment. So far, this task often relies on traces or datasets (we will use both terms indistinctly) including Contact Times and Intercontact Times between nodes, coming either from synthetic models or captured in real-life situations.

At first glance, traces such as the ones available at CRAWDAD seem to be the most realistic material usable for Delay Tolerant Network studies. There are however strong hypotheses captured into these traces: the capture setting (conference, campus...), the radio technology (Bluetooth, WiFi), the number of nodes, the total duration, the scope (standard working hours vs. round-the-clock recording), and the sampling period. Furthermore, because these data collection efforts are usually hard to perform, only few research teams have managed to provide such datasets. This led to some traces being used much more often than others. For the purpose of this study, we focus on datasets containing contact times between devices. GPS logs or network logs are consequently out of the scope of this paper.

In [2], we already studied the impact of some filtering techniques on the statistical analysis of contact datasets. We chose to focus on statistical distributions instead of network performance due to their use in some proposals in the literature [7,24], aiming at scaling the captured networks in terms of time scale and/or number of nodes. We found out that among the filters used, some had a major impact on distributions, while the contribution of others was much more limited.

In this paper, we try to go deeper in the study of traces. Hence, the work presented here considers all the steps in the life cycle of traces, from the real-life data collection to the statistical analysis. The various filters which can be applied in between are also discussed. The analysis of four existing datasets and their use in various research works provided us with several recommendations for future trace-based studies.

The remainder of the paper is organized as follows. In Section 2, we present the traces and identify use cases for which they can be exploited. Sections 3–5 then provide a detailed inventory of all factors which could influence subsequent results, when respectively collecting, filtering or scaling the trace. This in turn allows Section 6 to propose a list of parameters which need to be considered when working with a contact trace, and provide some advice. Finally, Section 7 concludes the paper.

2. Background

In this section, we present the contact datasets which will be covered by our study. We also identify three main use cases for contact datasets, based on the existing literature.

* Corresponding author.
E-mail addresses: gwilherm.baudic@isae.fr (G. Baudic), tanguy.perennou@isae.fr (T. Perennou), emmanuel.lochin@isae.fr (E. Lochin).

Table 1
Main characteristics of the datasets.

	Rollernet	MIT	Infocom '05	Humanet
Technology	Bluetooth	Bluetooth	Bluetooth	Bluetooth
Year	2006	2004	2005	2010
Device	iMote	phone	iMote	custom
Environment	urban	campus	conference	office
Duration (days)	0.125	284	3	1
Time span	NA	24/7	24/7	workday
Sampling period (s)	15	300	120	5
Internal nodes	62	89	41	56
Internal contacts	60,146	114,046	22,459	64,445
External contacts	72,365	171,466	5,757	64,531

2.1. Existing datasets chosen

We have selected four datasets to illustrate our study. Three of them were already used in [2] and in several previous works, while the fourth one represents a more recent experiment for which the authors made a careful study before choosing parameters. All are publicly available through the CRAWDAD archive website. An overview of their characteristics is presented in Table 1.

The first dataset is Rollernet [3], which was collected during a 3-h long rollerskating tour in Paris in 2006. 62 Bluetooth contact loggers (iMotes) were distributed to volunteers and staff members among approximately 2,500 participants. The second one comes from the Infocom 2005 experiment [22], which also relied on similar contact loggers. They were distributed among 41 participants of the student workshop of the conference, who also attended the rest of the conference afterwards for a total duration of three days. The third dataset, MIT Reality Mining [10], was collected through the use of an activity logging application embedded in mobile phones. These were lent to 100 MIT students over the course of the 2004-2005 academic year, representing almost 9 months of data. For this third dataset, we only considered the Bluetooth contact traces in our study (thus ignoring all information on the cellular network), and restricted the data to the 89 devices which effectively recorded contacts.

Finally, the Humanet dataset [4] was also produced using custom hardware and Bluetooth. It recorded contacts between 56 people in an office environment during workdays (no nights or weekends) for 6 weeks in 2010. However at the time of this writing, only 1 day has been published so far. We only considered contacts recorded between users when the devices were worn, thanks to the use of the mobility flag recorded in the trace.

As can be seen from Table 1, the four traces selected here offer a true diversity in terms of sampling periods, collection dates, trace lengths and captured environments. Although there are several other datasets available, the goal of the present work is to study the assumptions and not the datasets, which is why we decide to restrict ourselves to four of them.

Some studies of Delay Tolerant Networks also used WLAN traces, such as the Dartmouth dataset [15]. In this paper, we decide to focus on Bluetooth traces; however, most of the issues highlighted here also apply to WLAN traces.

2.2. Typical use cases for traces

We already mentioned that real traces are a straightforward solution for Delay Tolerant Network researchers willing to add some realism to their studies. More precisely, there are three major use cases for traces which can be identified from the literature: social and mobility inference, performance assessment and statistical analysis.

Social and mobility inference. For this task, traces are typically processed in order to obtain social graphs, or at least links of various strengths (number of contacts, total duration of contacts, time of last contact, etc.) depending on the relationship between users. This in turn can allow to identify communities, or more simply to take routing decisions for protocols based on social proximity.

Performance assessment. It is rather commonplace for a paper proposing a new routing protocol to demonstrate its applicability on both synthetic and real traces. In this case, real traces are used to overcome the potential lack of realism of the mobility models behind the synthetic traces. It was however shown in [21] that because contact traces do not record actual available bandwidth and buffer occupancy, they can lead to overly optimistic performance results.

Statistical analysis. This was the main topic of [2], and is considered as a usual processing to go beyond the raw trace, for example by capturing the overall contact times distribution or node degree instead of individual values. This is also the approach used by [7,24] in order to extend a trace in time span and/or in number of nodes, as detailed in Section 5.

3. Collecting the trace

In this section, we provide a list of parameters which have to be set at the time of recording. This list will be useful to both practitioners willing to collect new traces and researchers planning to exploit existing datasets.

3.1. Communication hardware choice

Traces such as Humanet, Rollernet or those of the Haggle experiments have been produced with custom-made devices such as iMotes. The choice of custom devices was made by the Humanet researchers to be able to tweak Bluetooth parameters which were not adjustable on smartphones [5]. For real applications, which are highly likely to be deployed on off-the-shelf devices, the additional constraints of such devices would need to be taken into account. Constraints can come from the underlying operating system, the device usage patterns [23], or even, as mentioned before, the lack of control over some parameters. The MIT dataset [10] for example was produced using an app on smartphones, which were also used by participants on a daily basis to make phone calls or send text messages. The same year, researchers at the University of Toronto [23] also made a study on students equipped with PDAs running custom software. In these cases, exhausting the battery is not only harmful to the experiment itself, but also to the overall experience for the user. Note that this consideration would also be true for real applications. One parameter which can be adjusted in this case is the frequency to look for other devices, discussed next.

3.2. Sampling period

The traces considered in this paper have all been produced by probing at regular intervals for potential contacts. The choice of this frequency (called sampling period or sometimes granularity in the literature) can be determined by several factors. First, it is desirable to leave enough time for the other devices to respond to the probe. For this reason, the authors of [5] chose a value of 5 s after a careful study of response time for several smartphone models. Other limitations come from the portable nature of the devices used: a big sampling period has the advantage of saving energy (a concern already expressed in the previous subsection) and also memory [10]. Theoretical methods for choosing the optimal frequency can also be found in [18] and [20].

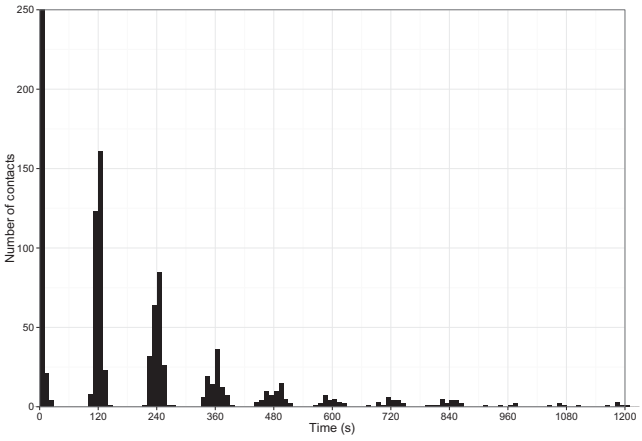


Fig. 1. Histogram of the contact times for the keynote in the Infocom 2005 dataset. The sampling period of 120 s is clearly visible between peaks. The peak at 0 s contains in fact 1575 contacts.

However, the choice of a big sampling period means that shorter contacts remain unseen, which may underestimate contact opportunities. To address this issue, the authors of [29] introduce a method to infer a *plausible* mobility from the contact traces, which can then be used to generate other traces with a higher temporal resolution. This approach was later criticized in [16], due to the high number of parameters which have to be set to properly capture the intrinsic mobility. We can also mention another solution of adaptive sampling algorithms, which vary the sampling period depending on the observed contacts or number of neighbors. Two examples are [11] and [28].

As an illustration, a close-up view of the distribution of contact times for the keynote contained in the Infocom 2005 dataset is provided in Fig. 1. According to the conference program, this keynote lasted for 2.5 h on the first morning of the conference. For clarity, we reduced the range of the x axis from 5000 to 1200 s, and of the y axis from 1600 to 250 contacts. Each bar has a width of 10 s. This plot only considers contact times for the keynote, but we also obtained similar shapes for other periods of this dataset (panels, sessions...). By using a histogram and linear scale instead of the much more common Complementary Cumulative Distribution Function log-log plots, the sampling period becomes more visible: the time interval between each peak is 120 s, which corresponds to the sampling period advertised for this dataset. Notice that these peaks would translate into steps on a Complementary Cumulative Distribution Function log-log plot. The fact that there are also other contacts whose durations are concentrated around these peaks comes from the desynchronization procedure, discussed in the next subsection. It should be noted here that the peak located at 0 s contains in fact 1575 contacts (instead of 250 as the scale choice might suggest).

3.3. Time synchronization

With a distributed data collection, it is crucial to have a common time reference to correctly detect encounters. Unfortunately, the intermittent connectivity which characterizes Delay Tolerant Networks makes this challenging. Interestingly, perfect time synchronization may be undesirable during data collection, as two Bluetooth devices scanning at the same time cannot discover each other. For this reason, the authors of [22] and [3] used a random dephasing between device clocks, respectively of $\pm$ 12 s and $\pm$ 5 s.

The authors of the Humanet dataset [4] used a nightly synchronization: each night, when the collection device was left at the office for charging and data uploading, it received a new timestamp from the central server [5]. During the day, perfect synchronization of scanning was avoided by using a random slave period $(3 + rand(1.5)$ s). On the contrary, authors of [22] and [3] performed a manual synchronization after the experiment, a process which is prone to errors: for example in [22], accuracy below 5 mn is not guaranteed.

3.4. Storage format

At the end of the experiment, data from all participating devices is gathered and merged. Then, the resulting files can be used for evaluation and/or shared with the community. As was already noticed in [17], there is no common format for all the traces. Some are available as text files [3,22], others as SQL database dumps [4,10].

The type of data stored also differs: if it is always feasible to find the identity of the two nodes involved, along with the start and end times of the contacts, contact and intercontact times sometimes have to be computed. Format of times also vary, ranging from Unix timestamps [10] to seconds from the beginning of the experiment [22]. Consequently, a researcher willing to use several traces would first have to decide for a common format, then do some scripting to get all the traces in this common format.

For access point records such as [15] (which were not used in this survey), the trace does not directly contain contacts between nodes, but only connections to access points; hence, a common assumption is to consider two nodes as in contact when they are both simultaneously connected to the same access point (see for example [9] or [25]).

4. Filtering the trace

In this section, we now place ourselves from the point of view of a researcher who would like to use real traces for his work, such as datasets available at CRAWDAD. Although it seems straightforward to use the raw data directly, there are in fact several filters which can be applied. Notice that this section can also be valuable for users of synthetic traces, because it lists several parameters of interest for simulations.

4.1. Defining intercontact times

A first factor which needs to be accounted for is the definition of an Intercontact Time, as we already pointed out in [2]: while most of the literature considers it to be the duration between the end of a contact and the beginning of the following one, some papers such as [24] consider the time difference between the beginning of two consecutive contacts. Because this latter definition aggregates both the contact time and the "usual" intercontact time, we believe it to be a reasonable choice when contacts are considered instantaneous, as is often the case in theoretical performance models. Due to these two definitions, special care must be taken when intercontact times have to be computed from contact start and end times, like we mentioned in the previous subsection; to the best of our knowledge, precomputed intercontact times available in some datasets already follow the usual definition.

4.2. Symmetrizing contacts

Bluetooth uses an asymmetric discovery procedure, which means that a recorded contact between node i and node j does not necessarily mean that j and i could also communicate; in fact, the authors of [30] claim that this is rather uncommon. However, some works choose to assume symmetry [26]. For social analysis, symmetry is likely to be desirable (if user A meets user

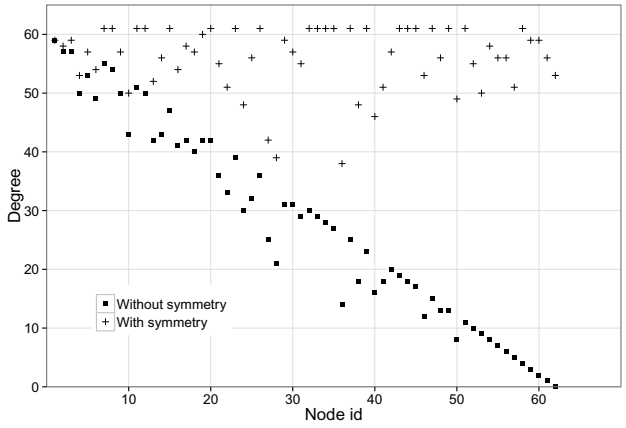


Fig. 2. Influence of the symmetry on the degree computation for the 62 internal nodes of the Rollernet dataset. In this example, forgetting the symmetry underestimates the degree, especially for the nodes having the biggest indexes.

B, then B also meets A). However, issues can appear with symmetric contacts when computing node degrees. In this case, researchers should not forget that a contact between nodes i and j means that j contributes to the degree of i, just as much as i contributes to the degree of j. We show the differences for the case of the internal nodes of the Rollernet dataset in Fig. 2. Here, forgetting that the raw dataset assumes symmetry leads to an underestimation of the computed degrees, especially for the nodes with the biggest indexes. Consider for example node 58 (out of 62), which degree jumps from 4 to 61 when symmetry is correctly taken into account. This leads the authors of [27] to conclude that node degrees exhibit a linear distribution, while the results presented here in Fig. 2 with symmetry clearly contradict these findings.

On the contrary, symmetry should preferably be avoided for network performance assessment, unless of course the connection is effectively symmetric.

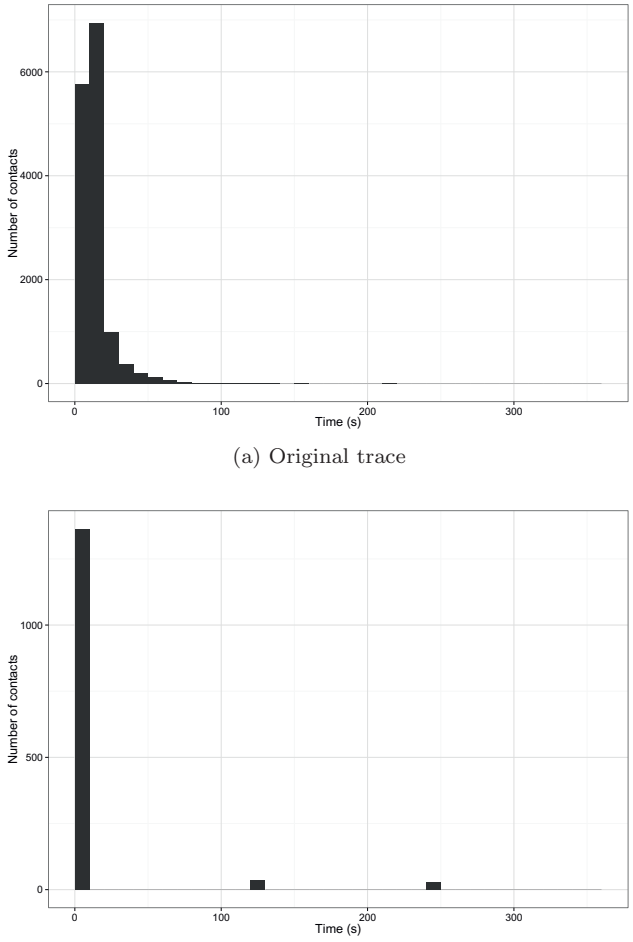
4.3. Merging contacts

In contact traces, a contact between nodes i and j lasts as long as probes from i get an answer from j. However, the authors of several datasets observed a large number of contacts which were only separated by one probe, and consequently decided in this case to merge such contacts. For the Infocom 2005 dataset, this choice is made to address a memory exhaustion problem caused by contacts with a specific brand of mobile phones [12].

However, this strategy can merge two distant contacts, thus creating an artificially long contact which may not correspond to the reality. Although this may be desirable when inferring mobility or social relationships from the trace, it would lead to overestimating the transfer abilities of the network, thus biasing performance evaluations.

To illustrate this behavior, we applied such a filter to a synthetic trace. The trace was generated with 41 nodes on a 100 $\times$ 100 m^2 area, moving according to the Random Waypoint model (constant speed of 1 m/s, no pauses) during 2.5 h, the first one being discarded to allow convergence of the model. Node range was 10 m. The original sampling period was 1 s, a small value achieved thanks to the use of a simulator (namely, the ONE [13]) which would not be possible in reality. The final sampling period applied by the filter was 120 s. Note that the choice of parameters (node number, duration, sampling period) was directly inspired from paper sessions in the Infocom 2005 dataset.

The results are presented in Fig. 3. In the original trace (Fig. 3(a)), the longest contact recorded had a duration of 219 s.



(a) Original trace

(b) Filtered trace

Fig. 3. Distribution of the contact durations in the original and synthetic traces. Note the appearance of longer contacts after the filtering.

On the final trace however (Fig. 3(b)), we notice the appearance of 28 contacts with a duration of 240 s, which did not exist in the first trace, and even a 360-s contact. When evaluating performance, such long contacts would be seen as highly interesting data transfer opportunities, despite being only artifacts of the contact merging process.

One may however object that the previous study was done with a synthetic dataset instead of a real one. In fact, a similar study was also conducted in [6] for the Humanet dataset. Originally produced with a sampling period of 5 s, the authors processed the trace to virtually achieve sampling periods of 120 s as in [22] or 300 s as in [10]. Similarly to the present work, increasing the sampling period produced longer observed contacts, and dramatically reduced the total number of contacts recorded.

Like the previous filter, this one is sometimes already included in the traces available on repositories: this is the case for example for Rollernet (symmetry, contact merging) [3] and Infocom 2005 (contact merging) [22]. This can be problematic because for such datasets, there is no straightforward way of assessing the impact of these choices due to the unavailability of the raw, unfiltered data; worse, users of such traces may not even realize the implications of this filtering. On the contrary, the more recent Humanet dataset [4] does not include any filter in the released version; some are however explicitly applied by the authors for their subsequent studies [6].

4.4. Losing or removing extreme values

The range of values recorded in contact datasets can be quite broad, with intercontacts ranging for instance from a few seconds to a few days in the Infocom 2005 dataset. Hence, one may want to remove values which are outside a certain range. Some of them can indeed be seen as a kind of artifact of the collection process: in the case of the Infocom 2005 dataset, the longest intercontact times last for almost three days, which happens to be the length of the whole experiment. They could therefore be interpreted as pairs of nodes which saw each other only at the beginning and at the end of the experiment, thus indicating that the corresponding people would have been unlikely to meet at all without the experiment. It should also be noted that the disappearance of the longest values can be a side effect of restricting the length of the trace: it is impossible to record durations longer than the trace length.

However, extreme values can also be the smallest ones. Among these, one value is especially interesting: contacts with a duration of 0 second. They represent a large part of the raw Bluetooth traces studied here (75% of Rollernet, 52% of Infocom 2005, 43% of the MIT dataset and 55% of the Humanet dataset). It should be noted that these contacts did not really last for 0 s during the trace collection; instead, they were simply shorter than the sampling period and were only detected for one probe. This can be seen for example on Fig. 3 in Section 4.3, where the application of the sampling algorithm led to a majority of recorded contacts with a duration of 0 s while the original trace did not contain any (minimum duration was 1 s). Recall also the highest peak at 0 s in Fig. 1. Yet, some papers such as [24] appear to discard all contacts shorter than the sampling period (including the 0-s contacts), thus removing a substantial part of the data. On the contrary, the authors of [26] choose to include these 0-s contacts in their analysis by extending them to an arbitrary value of 1 s. This extension, despite being far from reality, is however necessary for some statistical tools due to the presence of logarithms in the formulas [2].

The question that is raised by these choices is about their impact on the intercontact times: if some contacts are deleted, then the surrounding intercontacts should be merged because they no longer separate anything. This would lead to overestimating the intercontact length.

4.5. Restricting contacts to device mobility

Referred to as the “human mask” in [6], this filter is used to discard contacts which happened while the device was not worn by the user (e.g., it was idle on the desk). If the use of this filter is desirable for inferring mobility from the contact traces, it should be avoided when evaluating network performance, since it may discard useful connections. Furthermore, applying this filter requires a way to tell when the device is worn and when it is not: in the case of the Humanet dataset [4], this was achieved through the collection of accelerometer data, a parameter which is unfortunately not present in all traces. For the MIT dataset [10], for which no accelerometer data was available, the authors tried to implement a so-called ’forgotten phone’ classifier to identify times when the phone was not with the user.

4.6. Filtering devices

Traces may not only record contacts between devices participating in the experiment (“internal” nodes), but also with other devices which were observed despite not collecting any data (“external” nodes).

The issue with the external nodes is that they can be detected, but not detect others. In other words, and unless symmetry of contacts is assumed, they will be able to receive data, but not to send it. Consequently, and for simplicity, contacts with these external nodes are sometimes simply discarded from the trace. The study of aggregated distributions of Contact Times and Intercontact Times led the authors of [12] and later of [2] to the conclusion that both categories of nodes exhibited similar behaviors.

Even when the nodes are restricted to the “internal” ones, further filtering may be needed. Some datasets ([4] and most of the traces in [22]) involve fixed nodes, typically placed by the researchers in strategic zones where people are highly likely to meet. Unlike the others, these nodes are not tied to a person. While these nodes are especially interesting for location inference, they are also known to dramatically improve the performance of Delay Tolerant Networks [1]; hence, they should be either properly acknowledged or discarded when evaluating network performance on such traces. For statistical inference or mobility modeling, we believe that both external and fixed nodes should also be discarded, because they will typically exhibit different connection patterns.

4.7. Pair discarding

To ensure statistical validity of the distribution fitting process, which will be discussed in more details in Section 5, we need to have a minimum number of samples for each pair. This lower bound can also be interesting when building social graphs, as a way to remove the less active pairs from the analysis which would translate into low-weighted edges and unnecessarily clutter the graph. The authors of [9] use a threshold of 4 contacts, while the value chosen in [26] is 9 contacts. It should be noted here that the assumption of symmetry mentioned earlier will also impact the number of pairs which would be discarded. Indeed, merging the contacts from node i to node j with those from j to i into a single pair will logically increase the total number of contacts recorded for this pair, eventually going over the threshold. Hence, the use of the symmetry filter should be properly acknowledged for such uses, as was done in [9,26].

The major issue of this filter is the fact that some pairs will be removed from the analysis, as if they did not exist. This is especially problematic for statistical analysis, because it means that such pairs being missing as inputs will typically be also missing in the data generated from this analysis. One possible solution could be to aggregate the data from all such pairs, and treat them as one big virtual pair, at the expense of ironing out any differences that may exist between the original pairs.

5. Scaling the trace

Once the trace has been adequately filtered, this may not be sufficient for the envisioned purposes. Indeed, datasets may contain time periods which exhibit different properties: some nodes may not be functioning properly [10], or the experimental conditions were different (such as a break during a rollerskating tour [3], holidays between school terms [10], or nights [22]). Sometimes, the dataset is simply too long for the envisioned experiment, so that only a subset of it suffices. This subset should however be chosen carefully: it would be unfortunate to try to study the dynamics of users in a session of the Infocom 2005 conference by taking a 1.5-h[1] long sample in the middle of the night, or studying the on-campus interactions of MIT students over a holiday week.

But the opposite can also happen: the trace can be too short in time or not contain enough nodes. Due to the inherent difficulties of setting up trace collections, and the resulting limited number of existing datasets, such limitations are commonplace. To

[1] This was the session length, as found in the conference program.

Table 2
Filter settings for the fitting study.

Filter	Value
Intercontact time definition	Usual
Symmetry	No
Contact merging	Yes, included in raw trace
Extreme values	0 s contacts extended to 1
Device mobility	NA (no data available)
Device filter	Internal nodes only
Trace length	Whole trace
Contacts per pair	≥ 3

address these shortcomings, two solutions have been proposed in the literature, namely the Community Trace Generator [7] and the Encounter-based MOdel [24].

5.1. Existing scaling approaches

These two proposals share a common approach: extracting characteristics from the trace as statistical distributions (and not as raw data), which are subsequently used for the generation of a new trace exhibiting the same statistical properties. CTG [7] captures the number of nodes along with the distributions of node degree, aggregated Contact Times and Intercontact Times. EMO [24] also captures the number of nodes and distribution of node degree, but contact and intercontact times are treated in a pairwise manner. More precisely, the pairwise empirical distributions are first fitted to a probability distribution (which needs to be the same across all pairs), then the empirical distribution of the *parameters* of these distributions is fitted to another probability distribution. For example, the pairwise contacts of the MIT dataset [10] are found to be exponentially distributed, with parameters of the exponential law following a normal distribution [24].

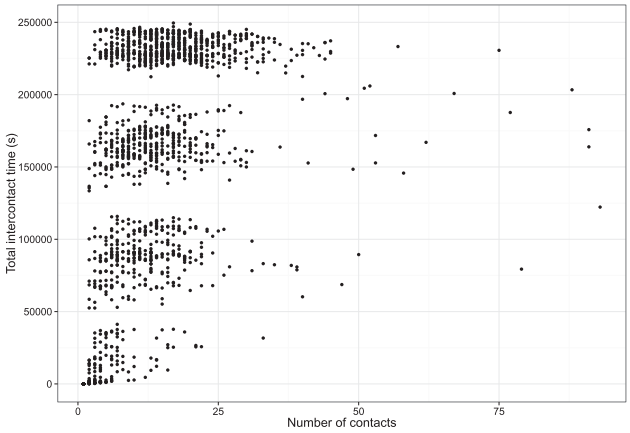
At this point, we would like to mention again the aggregated contact time distribution shown in Fig. 1. Because of the sampling process, contact durations are distributed close to the sampling periods, leading to ties (several samples with the same values) and an intermittent shape, with long runs of time values for which no contact is recorded. This is in sharp contrast with the continuous character typically displayed by usual statistical distributions.

5.2. Statistical fitting issues

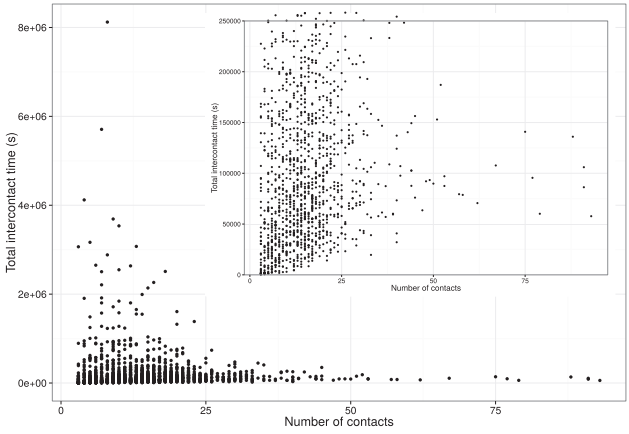
A first problem comes from the distribution fitting process used, which strives to use the same statistical distribution to fit the (inter-)contacts of all the pairs. In [9] and later in [26], it was shown for four different traces (Infocom 2005, Rollernet, MIT and Dartmouth) that the Intercontact Times of some pairs could be modeled equally well by several distributions, while for some other pairs none of the three distributions considered (namely Pareto, exponential and log-normal) gave a satisfactory result. Interestingly, the distribution which was able to represent the largest number of pairs was for the four traces the log-normal distribution, despite the variety of environments captured.

We applied this statistical fitting approach (one distribution to model all the pairs) to the pairwise Intercontact Times of the Infocom 2005 dataset [22]. We provide the list of parameters for the filters introduced in the previous section in Table 2.

No extreme values are discarded. As shown in Table 2, 0-s contacts are however extended to 1 s, so no intercontact has to be modified. We used the same statistical fitting procedure as in [2] (Maximum Likelihood estimation for the parameters, and Kolmogorov–Smirnov test to determine the best candidate). The choice of distributions to test is also the same: exponential, log-normal or Pareto with $x_{\min} = 1$. We found out as in [9,26] that



(a) Original trace



(b) Synthesized trace

Fig. 4. Sums of pairwise intercontact times, for both the original dataset and the synthetic one obtained through statistical analysis. Each dot represents a pair. Some pairs exhibit much longer durations in the synthetic trace than in the original one, and the four-group structure disappears. Notice the scale change between the two graphs.

the statistical distribution able to represent the largest number of pairs was log-normal. Then, we used the estimated parameters found for each pair to generate the same number of synthetic Intercontact Times for each pair as in the original trace. The results in terms of total Intercontact Times can be found on Fig. 4. On these plots, each dot represents a node pair. When comparing the original trace with the generated one, we notice that the totals for some of the pairs are much higher as the original ones (even exceeding the initial trace length), implying that longer intercontacts have been generated. Furthermore, the dots for the original trace are grouped into four regions. Closer inspection of the intercontact times for each pair reveals that this structure is caused by a few very long intercontacts (having durations of at least one night). These long (but rare) intercontacts, which could be seen as extreme values, are lost after the fitting process.

Even the fitting process itself can hide many details. A first example is the Pareto law. In [19], the authors mention two definitions for it, which they name “Pareto0” and “Pareto” depending on their ability to accept values arbitrarily close to 0. When only the Pareto law is used, it is also possible to find several definitions for the exponent. Another issue with power laws is the fact that they have a lower bound, usually written $x_{\min}$. In [2], we already mentioned that there are two possibilities to estimate this lower bound: arbitrarily setting it to a value such as the sampling

period [24], or using a mathematical estimator such as the one proposed in [8]. In both cases, data below the lower bound will be discarded, as if the filter presented in Section 4.4 had been applied. This makes comparison between distributions somewhat trickier, because other statistical distributions (like, for example, the exponential one) do not have any lower bound in their usual definition, and will therefore try to fit all the data. Once the various distributions to test have been chosen, a tool to help decide which one best models experimental data is needed. As we mentioned in [2], there are several possibilities. The simplest is to plot the data with an adequate scale and perform a graphical fitting, as in [12]. Other authors used statistical tools, such as the Kolmogorov–Smirnov test [24] or the Cramer-von Mises test [9,26]. Indeed, according to the authors of [8], statistical tests should always be favored over graphical fitting because the latter approach is too error-prone.

5.3. Validation issues

These statistical fitting approaches have another drawback: they are validated by showing their ability to reproduce aggregated distributions of contact and intercontact times. The major advantage of aggregated distributions over their pairwise counterparts is that they are much more compact (only one distribution to consider) [9], which made them a common approach in the literature for validation. However, the authors of [14] claim that validation should be made on parameters which are *not* used as inputs of the algorithm: they validate their mobility model (which is based on hotspots, speed distribution of users and transitions between hotspots) by its ability to correctly estimate the number of people at a hotspot across the workday.

Furthermore, aggregated distributions are known to hide many details. In [9], it was found that both exponential and log-normal pairwise distributions could lead when aggregated to the observed power laws. The authors of [19] also show that several other pairwise distributions can lead to power laws when aggregated, and provide a list of cases when using only the aggregated distribution is in fact not correct. Another issue was shown in [25]. In this work, the authors select three mobility models, and tune their parameters to reproduce the empirical aggregated distributions of Contact Times and Intercontact Times found in real datasets. When performing simulations with both the real traces and the synthetic ones derived from them, they find out that synthetic traces always provide much more optimistic performance results than the real ones.

6. Recommendations

In Section 3, we reviewed the parameters to be considered at the time of data collection, while Section 4 provided an overview of the filters applied for trace exploitation. Then, Section 5 presented two approaches to scale existing traces and their limitations. Based on these findings, we can now list the various parameters which have to be considered when using a trace.

6.1. Production

First, the hardware and radio technology used for collection must match the ones which are envisioned for the application. The sampling period must also be chosen carefully, taking into account both what is planned for the application and the limitations of the radio technology, as detailed in [5]. More precisely, the sampling period must consider the time to find other devices, but also the acceptable impact in terms of energy consumption. However, energy is no longer a major concern: while the team behind the MIT dataset [10] tried in 2004 to make cell phones batteries to last for more than 2 days without charging, it is now not uncommon to charge a smartphone on a daily basis. Based on the conclusions from [6] and our complementary analysis on contact merging, a realistic trace should also be as precise as possible, thus requiring a small sampling period. Therefore, a trade-off between energy and precision is needed. We believe that trace collection efforts should focus on precision, at the expense of frequent charges of the experimental devices; on the contrary, real application deployments should aim for a pleasant user experience, which means limiting their footprint on energy consumption.

Issues on time synchronization must definitely be considered; based on existing literature, we would recommend random slave periods [5] instead of clock desynchronization, due to the smaller resulting errors.

If the collected data is to be made public, it is desirable to perform as little filtering as possible on the data, which will then allow other researchers to apply their own filters depending on their goal. Regarding these filters, their application should always be properly and precisely described, especially when they are incorporated in released data (as has already been the case in some datasets for symmetry, intercontact definition or contact merging). The aim here is to make results easier to reproduce and compare, a concern which was the original motivation for [2].

6.2. Filtering

A summary of recommendations on the filters can be found in Table 3, for the three common uses of traces introduced in Section 2.2: social structures and mobility inference, performance evaluation, and statistical analysis. The filters are presented in the same order as they were introduced in the previous sections. For each filter, we detail for each use case if it can, must or must not be applied. In this context, *can* means that we believe it is possible to apply this filter, provided that the researcher is well aware of it and of its consequences.

For example, the length filter can be used in all cases to isolate a part of the trace exhibiting particular properties (workdays, coffee breaks...). The same observation applies to the Intercontact Time definition: using the alternative definition is not wrong by itself, except when simultaneously considering non-zero contact times. Regarding symmetry, we also believe that it can be applied if the connection was really symmetric, but must be avoided otherwise.

For the other filters, the situation is slightly different. Merging contacts is recommended for social and mobility inference if one does not want to minimize the links between nodes in the network, as also found in [6]. On the contrary, overestimating data transfer opportunities will be detrimental to performance assessment and statistical analysis. We recommend filtering (or modifying) extreme values only for statistical analysis, at the very least when a large percentage of 0-s contacts is present in the trace. Similarly, we recommend the use of the mobility filter only for social or mobility inference, because it would underestimate transfer opportunities in the two other use cases considered. Regarding the choice of devices to consider, a study on social or mobility inference should typically ignore fixed or external nodes because they do not represent social interactions. Fixed nodes may however be useful for localization. For performance assessment, the choice of nodes is left to the researcher, as long as it is properly documented. For statistical analysis, the choice of nodes should also be carefully made in order to avoid the capture of unwanted behaviors. One may object that conclusions of [2,12] indicated similar trends among both internal and external nodes; however, these conclusions did not consider fixed nodes, and were drawn based on the study of aggregated distributions instead of pairwise ones. Finally, ensuring the validity of the statistical analysis requires a

Table 3
Recommendations for the use of filters. + = MUST, - = MUST NOT, o = CAN.

	Social/mobility inference	Performance assessment	Statistical analysis
Intercontact time definition	o	o	o
Symmetry	o	o	o
Contact merging	+	-	-
Extreme values	-	-	+
Device mobility	+	o	o
Device filter	+	o	+
Contacts per pair	-	-	+
Trace length	o	o	o

lower bound on the number of contacts for each pair, a condition which we believe is less necessary otherwise.

It should also be noted that there can be some interplay between the filters: for example, symmetry will increase the number of contacts for a given pair, and, due to its impact on Intercontact Times, also modify the range of values; choosing only a portion of the dataset will cause all (inter-)contacts longer than this new duration to disappear; removing the shortest or longest values directly influences the number of values of the node pair in question, as does the mobility filter or the trace length. This list is not exhaustive, and represents an even stronger motivation to properly document the filters used in a study.

However, one may argue that if a researcher has enough control on the experiment to correctly set all the parameters mentioned above, it may be desirable to skip the trace collection part and directly proceed to application deployment. But even in this situation, trace collection can be a reasonable choice, since a recording of all contacts will be available for subsequent trials. Availability of this data would increase the repeatability of the deployment. It also allows to exploit the trace with various filter combinations, and then possibly for multiple use cases. For example, it would be possible to increase the sampling period without compromising the validity of the collected data.

This raises the question of representativity of traces. More precisely, real applications are unlikely to be limited to simply probing their surrounding for other nodes, which is what trace collection efforts have been doing so far; instead, they will also involve data transfer between devices (of images, news items, chunks of large files...). Assessing the difference in terms of recorded contacts between both approaches is outside the scope of this work, but would represent a highly valuable result for the community. A first step in this direction has been made in [21], where the authors recorded message exchanges in addition to the contacts between nodes.

6.3. Scaling

We mentioned that available traces do not always exactly match the situations one is willing to study. In Section 5, we presented some solution proposals to this problem. We do not see any issue in taking only a subset of an existing dataset, provided the choice of time period is carefully conducted (i.e., not mistaking a conference session with a lunch or a social event) and precisely described. If already present in the trace, Intercontact Times should be recomputed to reflect the time span change. However, based on the various limitations previously outlined, we would strongly advise against upscaling (extending the number of nodes and/or the time span), unless new tools are created to address current issues of the existing approaches. Based on the findings expressed in Section 5, these new tools should be able to work at the pair level (and not only with aggregates [7]), and be able to use several statistical distributions to model them instead of only one. Definitions of the statistical tools (distributions and goodness-of-fit tests) should be properly provided to avoid any ambiguities when building upon the results. Furthermore, the quality of such tools should not be judged on their ability to accurately reproduce aggregated distributions [25], but rather on pairwise distributions or network performance results.

7. Conclusion

Contact traces collected during field trials are the most straightforward way for researchers to introduce some realism in their work. They are mainly used in the context of Delay Tolerant Networks for social inference, performance assessment or statistical analysis. In this work, we show that the conditions captured by traces can be quite far from reality, due to the high number of other factors which can influence the collection process (sampling period, device characteristics...). We also listed filters which have been applied to traces in various research works, and proposed recommendations on their use depending on the envisioned experiments. For all the filters, researchers need to be aware of their existence, especially when these are already incorporated in the files available in public repositories. Finally, two existing approaches aiming to extend datasets while retaining their intrinsic characteristics are presented. We found out that there are in fact several hidden limitations (such as modeling all node pairs with the same distribution, or validating with aggregated measurements) which may prevent them to truly achieve their goal.

Because of the large number of factors which have to be considered when using real traces for performance assessment, we have come to the conclusion that synthetic traces should be considered with more attention. Unlike real datasets, synthetic models have the advantage of offering full control over their parameters. However, measuring their conformance to reality is still a challenge, considering that this is often done by comparing them with real traces.

Acknowledgments

The authors are grateful to the anonymous reviewers for valuable comments that helped greatly improve the paper. They would also like to thank Victor Ramiro for his comments and questionings, and Jeremie Bigot for pointing out some issues of the scaling procedures.

References

[1] N. Banerjee, M.D. Corner, D. Towsley, B.N. Levine, Relays, base stations, and meshes: enhancing mobile networks with infrastructure, in: ACM MobiCom '08, 2008, pp. 81–91, doi:10.1145/1409944.1409955.
[2] G. Baudic, T. Pérennou, E. Lochin, Revisiting pitfalls of DTN datasets statistical analysis, in: ACM CHANTS '14, 2014, pp. 73–76, doi:10.1145/2645672.2645683.
[3] F. Benbadis, J. Leguay, CRAWDAD dataset upmc/rollernet (v. 2009-02-02), 2009, (Downloaded from http://crawdad.org/upmc/rollernet/20090202). http://dx.doi.org/10.15783/C7ZK53
[4] J.M. Cabero, V. Molina, I. Urteaga, F. Liberal, J.L. Martin, CRAWDAD dataset tecnalia/humanet (v. 2012-06-12), 2012, (Downloaded from http://crawdad.org/tecnalia/humanet/20120612). http://dx.doi.org/10.15783/C74G60

[5] J.M. Cabero, V. Molina, I. Urteaga, F. Liberal, J.L. Martin, Acquisition of human traces with Bluetooth technology: challenges and proposals, Ad Hoc Netw. 12 (2014) 2–16, doi:10.1016/j.adhoc.2012.05.007.

[6] J.M. Cabero, I. Urteaga, V. Molina, F. Liberal, J.L. Martin, Reliability of Bluetooth-based connectivity traces for the characterization of human interaction, Ad Hoc Netw. 24 (Part A) (2015) 135–146, doi:10.1016/j.adhoc.2014.08.010.

[7] R. Calegari, M. Musolesi, F. Raimondi, C. Mascolo, CTG: a connectivity trace generator for testing the performance of opportunistic mobile systems, ACM ESEC/FSE07, 2007.Dubrovnik, Croatia.

[8] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703, doi:10.1137/070710111.

[9] V. Conan, J. Leguay, T. Friedman, Characterizing pairwise inter-contact patterns in delay tolerant networks, in: Autonomics '07, 2007, pp. 19:1–19:9.

[10] N. Eagle, A.S. Pentland, CRAWDAD dataset mit/reality (v. 2005-07-01), 2005, (Downloaded from http://crawdad.org/mit/reality/20050701). http://dx.doi.org/10.15783/C71S31

[11] A. Hess, E. Hyytia, J. Ott, Efficient neighbor discovery in mobile opportunistic networking using mobility awareness, in: Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on, 2014, pp. 1–8, doi:10.1109/COMSNETS.2014.6734890.

[12] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: ACM WDTN '05, 2005, pp. 244–251.

[13] A. Keränen, J. Ott, T. Kärkkäinen, The ONE simulator for DTN protocol evaluation, in: ICST Simutools '09, 2009, pp. 55:1–55:10, doi:10.4108/ICST.SIMUTOOLS2009.5674.

[14] M. Kim, D. Kotz, S. Kim, Extracting a mobility model from real user traces, in: IEEE INFOCOM 2006, 2006, pp. 1–13, doi:10.1109/INFOCOM.2006.173.

[15] D. Kotz, T. Henderson, I. Abyzov, J. Yeo, CRAWDAD dataset dartmouth/campus (v. 2009-09-09), 2009, (Downloaded from http://crawdad.org/dartmouth/campus/20090909). http://dx.doi.org/10.15783/C7F59T

[16] A.K. Monfared, M.H. Ammar, E.W. Zegura, Plausible mobility inference from wireless contacts using optimization, in: ACM CHANTS '13, 2013, pp. 7–12, doi:10.1145/2505494.2505501.

[17] M. Musolesi, C. Mascolo, Mobility models for systems evaluation - a survey, Middleware for Network Eccentric and Mobile Applications, Springer-Verlag, 2008.

[18] A. Nayebi, G. Karlsson, Beaconing in wireless mobile networks, in: IEEE WCNC 2009, 2009, pp. 1–6, doi:10.1109/WCNC.2009.4917610.

[19] A. Passarella, M. Conti, Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks, IEEE Trans. Mobile Comput. 12 (12) (2013) 2483–2495, doi:10.1109/TMC.2012.213.

[20] S. Qin, G. Feng, Y. Zhang, How the contact-probing mechanism affects the transmission capacity of delay-tolerant networks, IEEE Trans. Veh. Technol. 60 (4) (2011) 1825–1834, doi:10.1109/TVT.2011.2131693.

[21] N. Ristanovic, G. Theodorakopoulos, J.-Y. Le Boudec, Traps and pitfalls of using contact traces in performance studies of opportunistic networks, in: IEEE INFOCOM 2012, 2012, pp. 1377–1385.

[22] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD dataset cambridge/haggle (v. 2009-05-29), 2009, (Downloaded from http://crawdad.org/cambridge/haggle/20090529). http://dx.doi.org/10.15783/C70011

[23] J. Su, A. Chin, A. Popivanova, A. Goel, E. de Lara, User mobility for opportunistic ad-hoc networking, in: IEEE WMCSA 2004, 2004, pp. 41–50, doi:10.1109/MCSA.2004.29.

[24] F. Tan, Y. Borghol, S. Ardon, EMO: A statistical encounter-based mobility model for simulating delay tolerant networks, in: IEEE WoWMoM 2008, 2008, pp. 1–8, doi:10.1109/WOWMOM.2008.4594848.

[25] G. Thakur, U. Kumar, A. Helmy, W.-j. Hsu, On the efficacy of mobility modeling for DTN evaluation: analysis of encounter statistics and spatio-temporal preferences, in: Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, 2011, pp. 510–515, doi:10.1109/IWCMC.2011.5982586.

[26] P. Tournoux, J. Leguay, F. Benbadis, V. Conan, M. Dias de Amorim, J. Whitbeck, The accordion phenomenon: analysis, characterization, and impact on DTN routing, in: IEEE INFOCOM 2009, 2009, pp. 1116–1124.

[27] P. Vieira, A. Costa, J. Macedo, A Comparison of opportunistic connection datasets, in: 2012 Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), 2012, pp. 66–73, doi:10.1109/EIDWT.2012.52.

[28] W. Wang, V. Srinivasan, M. Motani, Adaptive contact probing mechanisms for delay tolerant applications, in: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, in: MobiCom '07, ACM, New York, NY, USA, 2007, pp. 230–241, doi:10.1145/1287853.1287882.

[29] J. Whitbeck, M.D. de Amorim, V. Conan, M. Ammar, E. Zegura, From encounters to plausible mobility, Pervasive Mobile Comput. 7 (2) (2011) 206–222, doi:10.1016/j.pmcj.2010.11.001.

[30] E. Yoneki, The importance of data collection for modelling contact networks, in: Computational Science and Engineering, 2009. CSE '09. International Conference on, vol. 4, 2009, pp. 940–943, doi:10.1109/CSE.2009.332.