There and Back Again: Detecting Regularity in Human Encounter Communities

Matthew J. Williams, *Member, IEEE,*
Roger M. Whitaker, *Member, IEEE,* and Stuart M. Allen, *Member, IEEE*

Abstract—Detecting communities that recur over time is a challenging problem due to the potential sparsity of encounter events at an individual scale and inherent uncertainty in human behaviour. Existing methods for community detection in mobile human encounter networks ignore the presence of temporal patterns that lead to periodic components in the network. Daily and weekly routine is prevalent in human behaviour and can serve as rich context for applications that rely on person-to-person encounters, such as mobile routing protocols and intelligent digital personal assistants. In this article, we present the design, implementation, and evaluation of an approach to decentralised periodic community detection that is robust to uncertainty and computationally efficient. This alternative approach has a novel periodicity detection method inspired by a neural synchrony measure used in the field of neurophysiology. We evaluate our approach and investigate human periodic encounter patterns using empirical datasets of inferred and direct-sensed encounters.

Index Terms—social networks, temporal community detection, periodic patterns, physical proximity, human encounter networks

1 Introduction

Human encounter networks are inherently time-varying. Encounters represent instances where two or more individuals move into proximity. The need to understand and detect temporal encounter patterns has been recognised across many fields, including network science [1], epidemics on complex networks [2], [3], opportunistic network protocol design [4], [5], [6], and anticipatory mobile computing [7]. Much of this interest has been driven by the widespread adoption of mobile phones, which have become important proxies for human activity [8], providing the large-scale empirical data required to study human encounter patterns and enabled new applications that rely on mobility patterns as context in automated decision making. This trend is exemplified by the recent rise of intelligent digital personal assistants such as Google *Now*[1] and Apple *Siri*[2].

Weekly and daily routine have a strong influence on human behaviour, but these encounter patterns have received relatively little attention. Periodic encounter patterns have been shown to exist not only on a pairwise basis [9], [10], but also for communities [11] and connected components [1], [12]. The timings of encounters, however, are not perfectly consistent due to the time-variant uncertainty inherent in human mobility [13]. In this article, we introduce an algorithm for the detection of periodic encounter behaviour that is tolerant to uncertainty in the timing of encounter events. We adapt a regularity detection method [14] that is inspired by neural synchrony [15], [16], which allows similar repeated events to be captured without mandating that such events are precisely identical repetitions. Additionally the approach does not require discretisation of time into 'bins' as seen in previous approaches such as PEC-D (periodic encounter community detection) [11] and Habit [5]. Such discretisation results in loss of temporal resolution and is more susceptible to noise due to jitter.

The algorithm we develop also supports periodic encounter detection for delay tolerant and opportunistic network protocols, which have exploited community structure [17], [18] and models of temporal behaviour [4], [5], [10]. Decentralised methods are necessary in these settings due to their lack of infrastructure and open up the opportunity for new services that capitalise on human mobility [19], [20]. We also note that in cellular networks, decentralised approaches help to conserve bandwidth at the handset. Furthermore, given the highly sensitive nature of mobility data [21], avoiding the use of a centralised architecture can alleviate privacy concerns.

Generalising the model of *periodic encounter communities* (PECs) in [11], we refer to the communities we study in this article as *regular encounter communities* (RECs). We define a REC as a community of individuals that encounter at similar (rather than precise) times each week. Although we refer specifically to weekly patterns in this article, our approach can be applied, without loss of generality, to any other pre-defined period. The pattern detection method we develop is based on regularity detection [14], that finds regions of the week where the timing of the community's encounters are consistent. In particular, we present a novel inter-event interval analysis approach that allows individual nodes to extract their regular encounters with each node they meet. Decentralised detection of the broader RECs that these nodes may belong to then follows by an opportunistic

- *M.J. Williams is with the School of Computer Science, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK and the School of Biosciences, University of Exeter, Exeter EX4 4QE, UK. E-mail: m.j.williams@cs.bham.ac.uk.*
- *R.M. Whitaker and S.M. Allen are with Cardiff School of Computer Science & Informatics, Cardiff University, Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, UK. E-mail: {r.m.whitaker, stuart.m.allen}@cs.cardiff.ac.uk.*

1. http://www.google.com/landing/now/
2. http://www.apple.com/ios/siri/

sharing and incremental construction approach.

Using this technique, we explore regular encounter communities through two empirical datasets. REALITY is a dataset of Bluetooth encounters among staff and students during the 2004-2005 Reality Mining experiment at the Massachusetts Institute of Technology (MIT) [22]. DARTMOUTH is a dataset of encounters inferred from the WLAN access point co-locations of staff and students at Dartmouth College campus [23]. Our analysis considers the properties and prevalence of regular encounter communities, as well as a comparison of REC detection performance to PEC detection. We also consider the role of RECs in opportunistic peer-to-peer information spreading. As with previous work in this area, we assume a setting where low-level implementation details, such as lack of a global clock and asymmetric encounter sensing, are handled by a lower layer. In practice, the method introduced in this paper can be extended with a low-level communication protocol.

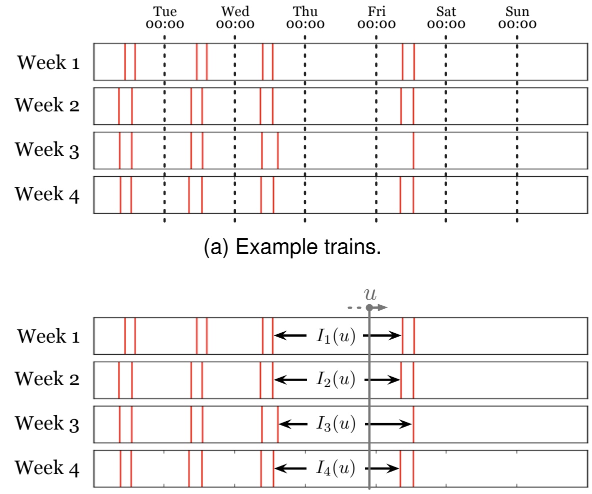
The article is organised as follows. We begin by presenting a method to extract regular encounters on a pairwise basis in Section 2. For each node, these extracted regular encounters are required for the next step of the decentralised regular encounter detection algorithm, which is presented in Section 3. In this section the method by which nodes incrementally share and construct their regular encounter communities is described. We then detail the datasets to be used in this article in Section 5, followed by experiments and results in Section 6. We discuss our results with respect to related work in Section 7 and conclude this article in Section 8.

2 REGULAR ENCOUNTER EXTRACTION

We refer to regularity as the behaviour of two individuals in consistently meeting one another at a particular time over a repeated window. To extract the regular encounters occurring between between a pair of individuals we introduce the **window size** parameter, denoted by ω. Consequently this parameter governs the scale at which regular patterns are to be detected. Although human mobility follows cycles at multiple scales [9], [11], [12] (e.g., daily, weekly, biweekly, and yearly), we adopt weekly regularity ($\omega = 7$) as this captures any weekday patterns that reoccur over multiple weeks.

A common representation of encounters is as a sequence of instantaneous events, where each event has zero duration. This is often how encounters are collected in practice, for example using periodic Bluetooth sensing [22]. The event-based nature of the encounter data we consider precludes the application of methods that require continuous, densely-sampled data, such as nonlinear time series analysis, harmonic analysis, and recurrence quantification analysis. Additionally, human event-based data may be sparse (the majority of individuals are unlikely to encounter one another more than a few times a day) making it challenging to capture structural properties.

These issues motivate alternative approaches and we adopt a neural coding analysis technique [14], [15], [16] that has originally been applied to look for patterns in neuron firing behaviour. Although this appears an unrelated domain, the commonality concerns event characterisation



(a) Example trains.

(b) Trains annotated to indicate $I_1(u)$, $I_2(u)$, $I_3(u)$, and $I_4(u)$ for a particular offset u.

Fig. 1. Example encounter trains for a pair of individuals over four weeks. Window width $\omega = 7$ days.

and consequently work in this area is of much interest but does not appear to have received attention. Importantly, this technique does not require discrete intervals (also called *data binning*) of data in the temporal dimension, and therefore it avoid changing the resolution of encounters.

The objective of regular encounter extraction is to identify the subset of encounters between a pair of individuals according to the chosen ω parameter. To formalise this we make the following definition:

Definition 1. Given two individuals, v and w, the **chronology** of encounters between v and w is denoted by the ordered sequence of encounter times $\mathcal{S}_{v,w} = \{t_i \,|\, i = 1, \ldots, L\}$, where L is the number of v's encounters with w. We assume that the chronology for two individuals is symmetric; i.e., we have $\mathcal{S}_{v,w} = \mathcal{S}_{w,v}$.

2.1 Instantaneous inter-event interval (IEI) irregularity

We assume that the encounter events in the chronology

$$\mathcal{S}_{v,w} = \{t_i \,|\, i = 1, \ldots, L\},$$

are offsets from some arbitrary origin, giving values $t_i \in (0, T_{max}]\ \forall\, i = 1, \ldots, L$. We segment each chronology into N continuous windows of duration ω, referring to each such window as an *encounter train*. The absolute times of encounter are considered modulo ω, thus translating the events in each encounter train into the interval $(0, \omega]$. We assume T_{max} and ω are chosen such that $\omega N = T_{max}$. L_n is used to denote the number of encounters in the n^{th} train and the sequence of offsets for the events in train n is denoted by

$$U^n = \{u_i^n \,|\, i = 1, \ldots, L_n\}.$$

An example of a four-week chronology and its corresponding encounter trains are shown in Figure 1a. Four encounter trains are present and offset from the start of each week can be clearly seen. Every point that isn't an event itself in an encounter train u falls between two events, which may not necessarily be in the same encounter train.

To examine this further we introduce the *inter-event interval* (IEI) as follows.

Definition 2. The **instantaneous inter-event interval function** $I^n(u)$ denotes the IEI for the n^{th} train at time offset $u \in (0, \omega]$; formally, instantaneous IEI is defined for three cases:

$$I^n(u) = u_1^n \quad \text{if} \quad 0 < u \leq u_1^n\,,$$
$$I^n(u) = \omega - u_{L_n}^n \quad \text{if} \quad u_{L_n}^n < u \leq \omega\,,$$

and

$$I^n(u) = \min(u_i^n \mid u_i^n \geq u) - \max(u_i^n \mid u_i^n < u)$$

if

$$u_1^n < u \leq u_{L_n}^n.$$

Figure 1b shows the example train from Figure 1a annotated with example instantaneous IEI values at a particular offset u. We define two further instantaneous measures: for time offset u, the instantaneous mean $\mu(u)$ is given by

$$\mu(u) = \frac{1}{N} \sum_{n=1}^{N} I^n(u)$$

and the instantaneous standard deviation $\sigma(u)$ is given by

$$\sigma(u) = \left(\frac{1}{N-1} \sum_{n=1}^{N} \left(I^n(u) - \mu(u)\right)^2 \right)^{1/2} .$$

Using these two instantaneous measures we can evaluate the dispersion in IEI values at a particular time offset, which represents the degree of dissimilarity in the timings of events across the N trains at that offset.

Definition 3. The **coefficient of variation** $c_{\text{var}}(u)$ provides a measure of dispersion in the IEI values at time offset u,

$$c_{\text{var}}(u) = \frac{\sigma(u)}{\mu(u)} .$$

$c_{\text{var}}(u)$ is a unitless measure and normalised against the mean, which enables comparison between the dispersion in collections of large IEI values and collections of small IEI values. $c_{\text{var}}(u) = 0$ indicates perfect regularity at offset u, and higher values indicate less consistency in the encounters between v and w at offset u. We also refer to $c_{\text{var}}(u)$ as the instantaneous IEI irregularity at offset u.

2.2 Computing instantaneous IEI irregularity

Efficient computation of instantaneous IEI irregularity is valuable. The instantaneous inter-event interval function $I^n(u)$ is obtainable from pre-computing inter-event interval (IEI) values. To achieve this we insert dummy event times 0 and ω into each train and compute the difference between neighbouring events in each train $n = 1, \ldots, N$. The arrival of an event in a train indicates a change in the instantaneous IEI for the period up to (but not including) the next event in that train. In other words, given two consecutive events u_i^n and u_{i+1}^n in train n, the instantaneous inter-event interval $I^n(u)$ during $u \in (u_i^n, u_{i+1}^n]$ is $u_{i+1}^n - u_i^n$.

It is necessary to know intervals of time where there are no intervening events in any train. These intervals therefore represent durations where $c_{\text{var}}(\cdot)$ is constant. More formally, given an interval $(u, u']$ such that no events appear between time offsets u and u' in any of the N trains, we need only compute the instantaneous coefficient of variation once for the interval $(u, u']$. To support this we define a master train to represent the event offset times taken from all trains.

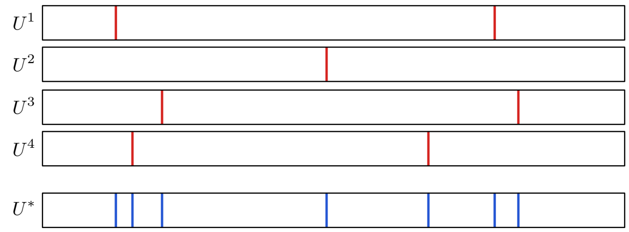


Fig. 2. Four example visit trains (U^1, U^2, U^3, and U^4) and their corresponding master train U^*.

Definition 4. A **master train**, denoted U^*, is the set of events $U^* = U^1 \cup \ldots \cup U^N$. For convenience we label the master train and its ordered events as $U^* = \{u_1^*, \ldots, u_L^*\}$.

An example of a collection of visit trains and the corresponding master train is depicted in Figure 2.

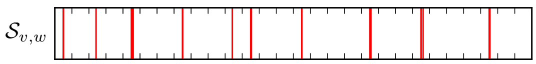
The master train conveniently describes the intervals during which there are no intervening events; specifically, there are no intervening events between offsets u_{i-1}^* and u_i^* for each $i = 2, \ldots, L$. We therefore only compute a new $c_{\text{var}}(\cdot)$ at each event in the master train. For each $i = 2, \ldots, L$ a coefficient of variation $c_{\text{var}}(u_i^*)$ is computed which yields the constant coefficient of variation for the interval $(u_{i-1}^*, u_i^*]$. In addition, we also calculate the coefficient of variation at the first event (i.e., u_1^*) to give the dispersion over $(0, u_1^*]$, and at ω to give the dispersion over $(u_L^*, \omega]$.

The approach described above has linear time complexity. Computing the coefficient of variation at a given time offset requires a look up of one instantaneous IEI value from each train, making the time complexity linear in the number of trains N for fixed L. To understand the reciprocal case (i.e., fixed N and varying L) we consider the effect of adding an event to a chronology. The result is one additional interval in the master train, requiring one additional calculation of the coefficient of variation. Therefore, assuming fixed N, the algorithm is linear in the number of events L.

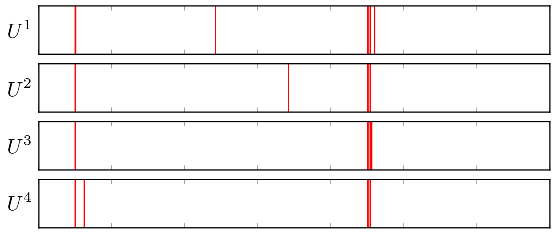
2.3 Extracting regular encounters

To distinguish events that repeatedly occur from irregular events we consider the dispersion in IEI values, as measured by the instantaneous coefficient of variation $c_{\text{var}}(\cdot)$ from Definition 3. We define a *dispersion threshold* θ beneath which encounters are classified as regular: that is an event at offset u is regular if and only if $c_{\text{var}}(u) \leq \theta$. The set of all such events across all N trains between nodes v and w are referred to as the *regular set* $R(\mathcal{S}_{v,w})$. Figure 3 provides an example of each stage in obtaining the regular events for a chronology.

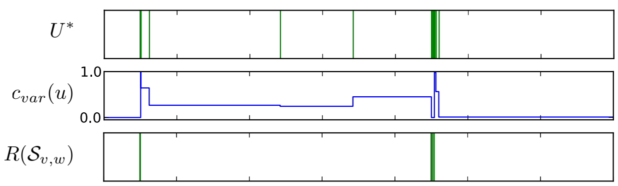
As noted in Section 2.2, computation of IEI measures such as coefficient of variation are linear in complexity. The algorithm for obtaining $R(S_{v,w})$ is therefore more computationally efficient than using PSE-Miner [24] for the extraction of local periodic communities (as shown in [11]), which



(a) An example chronology $\mathcal{S}_{v,w}$ depicting encounters between individuals v and w over 28 days.



(b) Four event trains U^1, U^2, U^3, and U^4 each of duration $\omega = 7$ days, corresponding to chronology $\mathcal{S}_{v,w}$.



(c) Master train U^* and the instantaneous coefficient of variation $c_{\text{var}}(u)$ at each event offset $u \in U^*$. Filtering by dispersion threshold $\theta = 0.2$ yields regular events $R(\mathcal{S}_{v,w})$.

Fig. 3. Pipeline for obtaining the regular events from an example chronology.

is polynomial [24]. We note that the efficiency improvement is a result of a single period of regularity (controlled by ω) being set *a priori*.

3 The REC detection problem

A regularity set $R(\mathcal{S}_{v,w})$ represents the regular encounters between a given pair of nodes v, w. For example, if two people met one another at the same time every seven days, then this time of week would be classified as a regular encounter. At a broader scale, the encounters among a whole community of individuals may also share the same regularity, giving rise to a *regular encounter community* (REC). To consider the representation of encounters among all individuals we introduce the following definition.

Definition 5. The **static encounter graph** $G^* = (V^*, E^*)$ is the graph representing all encounters and nodes appearing within the period of time represented by the interval $(0, T_{max}]$. That is, $\{v, w\} \in E$ if and only if individuals $v, w \in V$ were in proximity at least once.

The static encounter graph G^* corresponds to the aggregation of all encounters over the period time selected for analysis. Although this is a time-naive representation, by combination with regular encounter extraction we can identify groups of individuals that collectively share periodic encounters.

To exclude trivial cases we consider only the pairs of nodes that encounter each other a minimum number of times L_{min} in each window. Formally, given a window size ω we induce the subgraph $G = (V, E)$ of $G^* = (V^*, E^*)$ where the inclusion of each edge $\{v, w\} \in E^*$ in E is determined by the chronology $\mathcal{S}_{v,w}$ meeting the following condition: each of the N windows obtained through partitioning $\mathcal{S}_{v,w}$ by ω is such that $L_n \geq L_{min}$ for all $n = 1, \ldots, N$.

3.1 Problem definition

When comparing the regular events $R(\mathcal{S}_{v,w})$ among the edges of a prospective community it is prudent to permit a small degree of uncertainty around the timing of regular events, corresponding to jitter around the regular events identified in $R(\cdot)$. This is controlled by parameter ϕ which we use to map a regular event at time u to an interval $(u - \phi, u + \phi]$.

Definition 6. Let $R(\mathcal{S}_{v,w})$ be the set of regular events for the chronology of encounters $\mathcal{S}_{v,w}$ between individuals v and w, constructed with window size ω. Given a **jitter tolerance** of ϕ, the **regularity mask** $R_{v,w}$ is defined as the set

$$\bigcup_{u \in R(\mathcal{S}_{v,w})} (u - \phi, u + \phi]$$

with values falling outside $(0, \omega]$ being wrapped around.

We can obtain a regularity mask for two or more edges by obtaining the intersection of the multiple regularity masks. This represents the regions of regularity that are shared by the encounter patterns at all the edges involved. For example, given three nodes v, w, and x, the regularity mask $R_{x,v} \cap R_{x,w} \cap R_{v,w}$ represents the subset of regular regions common to all three regularity masks between nodes, as depicted in Figure 4.

The commutative and associative properties of the intersection operator are beneficial when developing decentralised solutions for REC detection, simplifying the process of incrementally combining RECs through knowledge sharing between nodes. This means that a node does not need to know the original regularity masks from which an intersection was computed, and the node can apply additional intersection operations without the order in which they are applied affecting the result.

In terms of notation, the empty regularity mask is denoted by $\emptyset$ and indicates that no regular regions were shared among the corresponding pair of individuals or collection of edges. We write $R_1 \subseteq R_2$ to denote the relationship of regularity mask R_1 being a subset or equal to regularity mask R_2. Finally, we use $|R|$ to denote the length of a regularity mask R. $|R|$ represents the overall duration of the regular regions in R (the Lebesgue measure of R) as follows.

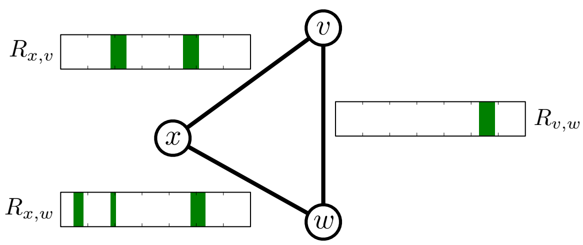
Definition 7. The **regularity length** $|R|$ of a regularity mask R of window size ω is

$$\int_0^{\omega} f(u)\, du$$

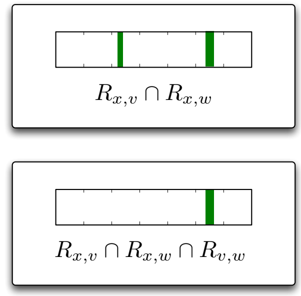
where $f(u) = 1$ if $u \in R$ and $f(u) = 0$ otherwise.

We now formally introduce the concept of a regular encounter community.

Definition 8. Denoted as the pair $\mathcal{R} = \langle C, R \rangle$, a **regular encounter community** (or **REC**) in an encounter graph



(a) Example pairwise regularity masks among three nodes. A regularity mask for a pair of nodes is drawn alongside the edge between those two nodes.



(b) Regularity mask intersections of two example edge subsets of the masks shown in Figure 4a.

Fig. 4. Regularity mask intersections from the encounter patterns among three nodes. Green bars indicate regions of regularity represented by a regularity mask. Figure 4a shows an example static encounter graph, annotated with the pairwise regularity masks for each pair of nodes. Figure 4b demonstrates the regularity masks obtained by intersecting masks associated with the example static encounter graph.

G is an encounter component $C = (V, E)$ in G along with a non-empty regularity mask R where

$$R \subseteq \bigcap_{\{v,w\} \in E} R_{v,w} \ .$$

Given a REC $\langle C, R\rangle$ with community $C = (V, E)$ constructed with jitter threshold ϕ, the regularity mask R encodes information about the locations of regular spikes in the chronologies corresponding to edges in E. In particular, a window offset u, such that $u \in R$, indicates that for each edge $\{v, w\} \in E$ there exists at least one regular encounter x in the regular set $R(\mathcal{S}_{v,w})$ where $|u - x| \leq \phi$.

It is possible that one REC may contain another REC and such subsumption can take two forms. A REC $\mathcal{R} = \langle C, R\rangle$ is *structurally subsumed* when there exists a connected community C' that contains C (i.e., $C \subset C'$) such that the regularity mask R is still valid for community C'. Structural subsumption means that we could add additional edges to encounter component C and re-computing the intersection of the edges' regularity mask would result in R. The counterpart to structural subsumption is *temporal subsumption*, which represents the case where there exists a REC with the same community but a larger regularity mask. Specifically a REC $\mathcal{R} = \langle C, R\rangle$ is temporally subsumed when there exists a regularity mask R' that is valid for the edge set of C and $R \subset R'$. We combine these two subsumption cases with the following definitions.

Definition 9. Let $\mathcal{R}_1 = \langle C_1, R_1\rangle$ and $\mathcal{R}_2 = \langle C_2, R_2\rangle$ be two RECs. We say that $\mathcal{R}_1$ is **subsumed** by $\mathcal{R}_2$ if and only if $R_1 \subseteq R_2$ and $C_1 \subseteq C_2$. We denote this subsumed-or-equal relationship by $\mathcal{R}_1 \preceq \mathcal{R}_2$, and write $\mathcal{R}_1 \prec \mathcal{R}_2$ to denote the case where $\mathcal{R}_1 \preceq \mathcal{R}_2$ and $\mathcal{R}_1 \neq \mathcal{R}_2$.

Definition 10. A REC $\mathcal{R}_1$ is **maximal** if and only if there is no REC $\mathcal{R}_2$ such that $\mathcal{R}_1 \prec \mathcal{R}_2$.

With knowledge of the maximal RECs, all other RECs are redundant. An algorithm or protocol for extracting RECs should seek to obtain maximal RECs. We formulate the REC detection problem as follows.

Definition 11. The **regular encounter community detection problem** requires identifying all maximal regular encounter communities that exist among the chronologies of a population of individuals.

4 DECENTRALISED REC DETECTION

In this section we present a decentralised algorithm for the regular encounter community detection problem. This REC detection algorithm is a variant of the opportunistic construction approach originally introduced in [11] where pairs of nodes share and combine their local knowledge of the RECs they belong to while they are in communication range. Through repeated opportunistic construction, nodes obtain more information on the structure of the globally maximal RECs they belong to.

To facilitate this, we assume each node v maintains a knowledge base of K_v of RECs it detects over time. When a node v encounters a node w, it receives K_w. It is the task of v to update its own knowledge base K_v by pairwise combining the RECs in K_v with those in K_w. This requires *compatibility* and *combination* rules to compare and aggregate RECs from two knowledge bases, and a local REC mining algorithm to initially populate a knowledge base with a node's local RECs.

The local REC miner extracts all maximal local RECs at a given node. The locality of a node is the set of chronologies that are incident at the node which corresponds to the encounter data that a node can directly observe.

4.1 Compatibility and combination rules

Compatibility and combination rules are required for opportunistic reconstruction, allowing a node to compare a local REC to a REC held by an adjacent node and, if they are compatible, to form a single new REC.

Definition 12. Two RECs $\langle C_1, R_1\rangle$ and $\langle C_2, R_2\rangle$ with encounter communities $C_1 = (V_1, E_1)$ and $C_2 = (V_2, E_2)$ are **compatible** for node v if all of the following hold:

1) Relevance to v: $v \in V_1$ and $v \in V_2$;
2) Structural overlap: $E_1 \cap E_2 \neq \emptyset$;
3) Intersection of regular regions: $R_1 \cap R_2 \neq \emptyset$.

Given two compatible RECs $\langle C_1, R_1 \rangle$ and $\langle C_2, R_2 \rangle$, we construct a new REC $\langle C', R' \rangle$ where $C' = C_1 \cup C_2$ and $R' = R_1 \cap R_2$. The intersection of the two RECs' regularity masks is the only regularity that both the RECs share, and therefore the only maximal (i.e., non-subsumed) regularity mask valid for the combined encounter community. We note that in the case $R_1 = R_2$, both RECs will be subsumed by the new REC, and therefore the knowledge base will require pruning.

4.2 Mining local RECs

We consider the task of mining all maximal local RECs for a node v. If we let $\mathcal{N}_v$ denote the set of nodes which v has encountered a minimum number of times, our task is to find all maximal RECs in the tree graph rooted at v and consisting of nodes $\{v\} \cup \mathcal{N}_v$.

There are $2^{|\mathcal{N}_v|} - 1$ connected subgraphs within this tree. For comparison, if we were to use a brute-force approach to mining local RECs the algorithm would need to construct each of these communities and check if each is a valid REC. The task of checking whether a local connected subgraph at v is a valid REC is straightforward. Let W be a subset of neighbours $W \subseteq \mathcal{N}_v$. By inducing a subgraph from the set of nodes $\{v\} \cup W$ we obtain an encounter community $C = (V, E)$. The regularity mask intersection for C is given by

$$R = \bigcap_{\{v,w\} \in E} R_{v,w} \ .$$

We can therefore obtain a REC $\langle C, R \rangle$ if the condition $R \neq \emptyset$ is met.

This brute-force approach requires checking of each non-empty subset of $\mathcal{N}_v$ and is computationally expensive. Furthermore, this approach requires an additional step to check whether each REC it generates is subsumed by RECs it has previously generated. However there are features of RECs that allow us to build a local miner that is more efficient than the brute-force approach.

Algorithm 1 presents a more-efficient approach, which exploits the properties of RECs to avoid unnecessary or redundant computation. Firstly, this algorithm avoids re-checking re-orderings of the same subset of neighbour nodes. Through the commutative property of regularity mask intersection we know that once a subset of neighbours has been checked, any re-orderings of that same subset will yield the same result. Secondly, improvement is made by using the property that if a particular neighbour subset $W \subset \mathcal{N}_v$ results in an empty regularity mask intersection, then any neighbour subset W' where $W \subset W' \subset \mathcal{N}_v$ will also result in an empty regularity mask intersection, allowing us to prune the search space.

Given a set of neighbours S_0, a particular call to `REC-Generator` constructs $|S_0|$ neighbour sets, generates a local connected subgraph rooted at v from each one, and checks if each forms a valid REC. The function also recursively checks each neighbour set, with another neighbour node being introduced at each recursive call, until no more unused neighbour nodes remain. After the algorithm adds a node from S_0 to W_0 to produce W_1, if the neighbour set W_1 results in an empty regularity mask we know that any subsequent recursive calls adding another node to W_1 will also result in empty mask, and therefore no further recursion involving W_1 is necessary.

Finally, the algorithm prunes subsumed RECs during each recursive call to the function `REC-Generator`. Given that the algorithm begins with the largest possible regularity mask intersections (i.e., the regularity masks between v and each of its neighbours) and incrementally intersects these with other masks, the only subsumption the algorithm must check for is structural subsumption; that is, subsumption of REC $\langle C_1, R_1 \rangle$ by REC $\langle C_2, R_2 \rangle$ due to $C_1 \subset C_2$.

Algorithm 1 REC-LOCAL-MINER

Input: Node v whose maximal local RECs will be extracted
Input: The set $\mathcal{N}_v$ of all nodes neighbouring v
Output: The set L of all maximal RECs local to v

```
for all w ∈ N_v do
    Precompute R_{v,w}
end for
L ← REC-GENERATOR( v, {}, N_v )
```

Algorithm 2 EXTEND

function EXTEND(L_0, L_1)
Input: Two lists of RECs, L_0 and L_1
Output: List L_0 extended with RECs in L_1

```
for all R_0 ∈ L_0 do
    Discard R_0 from L_0 if ∃ R_1 ∈ L_1 such that R_0 ⪯ R_1
end for
for all R_1 ∈ L_1 do
    Discard R_1 from L_1 if ∃ R_0 ∈ L_0 such that R_1 ≺ R_0
end for
Add all RECs in L_1 to L_0
```

5 DATASETS

We explore the presence of RECs in real-world encounter networks through two datasets. In both datasets we regard encounter histories as streams of zero-duration encounter events. A summary of the base datasets is given in Table 1. We note that these benchmark datasets have emerged from an era where smartphone adoption in economically developed countries was less ubiquitous than it is today. The continued paradigm-shift in mobile device ubiquity highlights the potential for the proof-of-concept method demonstrated in this paper to be deployed in widespread real-world mobile encounter community detection.

5.1 REALITY: Bluetooth encounters in the MIT Reality Mining project

The 2004-2005 Reality Mining project carried out at the Massachusetts Institute of Technology (MIT) followed 100 subjects (staff and students) equipped with Bluetooth-enabled mobile phones and recorded information about their behaviour over a nine-month academic period [22]. These subjects were staff and students at MIT.

Among the data collected are Bluetooth sightings between subjects, with Bluetooth scanning carried out at five-minute intervals. The dataset also includes sightings with

Algorithm 3 REC-GENERATOR

function REC-GENERATOR(v, W_0, S_0)
Input: Node v
Input: A set W_0 of neighbour nodes
Input: A set S_0 of neighbour nodes not yet added to W_0
if $S_0 = \emptyset$ **then**
  return $\{\}$
end if
Create empty list L_0 to hold candidate RECs
Create set of neighbours S_1 as a copy of S_0
for all $s \in S_0$ **do**
  Remove s from S_1
  $W_1 \leftarrow W_0 \cup \{s\}$
  $R_1 \leftarrow \bigcap_{w \in W_1} R_{v,w}$
  if $R_1 \neq \emptyset$ **then**
    Generate a REC from neighbour set W_1***:***
    $E_1 \leftarrow \{\{v, w\} \mid w \in W_1\}$
    $V_1 \leftarrow W_1 \cup \{v\}$
    $C_1 \leftarrow (V_1, E_1)$
    $\mathcal{R}_1 \leftarrow \langle C_1, R_1 \rangle$
    $L_1 \leftarrow \{\mathcal{R}_1\}$
    EXTEND(L_0, L_1)
    Generate RECs with nodes remaining in S_1***:***
    $L_2 \leftarrow$ REC-GENERATOR(v, W_1, S_1)
    EXTEND(L_0, L_2)
  end if
end for
return L_0
end function

devices outside of the experiment. We ignore these encounter chronologies as we cannot guarantee whether the external device is a device that is reliably carried by another individual. 531,703 encounter events were recorded during the dataset's nine-month duration, giving an average of 19.7 encounters per subject per day. These encounters are distributed over 2,675 chronologies.

5.2 DARTMOUTH: access point co-location on Dartmouth College campus

Visits in the DARTMOUTH dataset are drawn from the use of wireless access points (APs) by staff and students at Dartmouth College campus in the United States [23]. Over 450 APs placed across the 800km^2 of campus provide wireless coverage for most of the area, serving roughly 5,000 undergraduates and 1,200 faculty. When staff and students with wireless-enabled electronic devices (such as laptops and mobile phones) access the campus network the AP used to do so is logged at a central server, thus providing a partial record of the users' movements across the campus. The dataset providers estimate that at least 75% of undergraduates owned portable laptops in the collection period.

From the four years of wireless traces available in the Dartmouth movement dataset we selected visits from a more-recent year (2003) for our experiments, as recent years are likely to feature more mobile devices (such as wireless-enabled smartphones or personal digital assistants), and thus provide a richer record of user mobility. Such devices did not become very common until recently, however.

TABLE 1
Summary of base encounter datasets. M denotes the number of chronologies and $\langle L \rangle$ denotes the mean number of encounters per chronology. A chronology $\mathcal{S}_{v,u}$ is only included in a dataset if v encountered u at least once in the duration of the dataset. Only active individuals are counted; that is, an individual is only counted if he/she was involved in at least one chronology.

	DARTMOUTH	REALITY
Area(s)	Dartmouth	MIT
Duration	365 days	270 days
Encounter type	Access point co-location	Bluetooth proximity
Encounter range	$\leq$ 40m	$\leq$ 10m
Individuals	7,173	100
Encounters	6,800,755	531,703
Encounters per day	18,632.2	1,969.3
Encounters per individual per day	2.6	19.7
M	897,996	2,675
$\langle L \rangle$	7.6	198.8

To prepare this dataset for our experiments we carried out a number of sanitisation and filtering steps. In particular, we found many cases where a user repeatedly visited with the same AP at a short interval (typically less than 15 minutes). These are artefacts of the WLAN AP protocol and are caused by the same device periodically re-associating with the same AP. When these re-associations are separated by less than 15 minutes we assume that the user has not moved and therefore we discard the repeat events. In addition to this, we also acknowledge that some devices may be stationary throughout their stay on campus. Although many devices are carried with the students and staff (such as laptops or smartphones), some individuals may have unportable devices (such as desktop computers) accessing the campus WLAN. Since our aim is to study human mobility, we wish to focus on the devices often carried by the user. To mitigate the effect of stationary devices we only included devices that visited at least five different APs.

Encounters were inferred from device co-location at the same access point. To generate an encounter event we identify occurrences of pairs of individuals visiting the same AP within 10 minutes of each other. An occurrence of a pair of individuals visiting the same location within 10 minutes of each other translates to one encounter event between those two individuals, with the time of the encounter taken as the midpoint between the two individuals' respective visit times. The resulting encounter dataset contains 6,800,755 encounters, 7,173 active individuals and 897,996 chronologies. A user in this dataset is on average involved in 2.6 encounters per day.

6 EXPERIMENTS AND RESULTS

Our first aim is to characterise the existence of RECs in the well-known REALITY and DARTMOUTH datasets. This is not previously known and offers a useful case study. Our second aim is to evaluate the information sharing potential of RECs, which is significant for emerging decentralised delay-tolerant communication architectures. To do so, we adapt the token sharing scenario introduced in [11] to RECs.

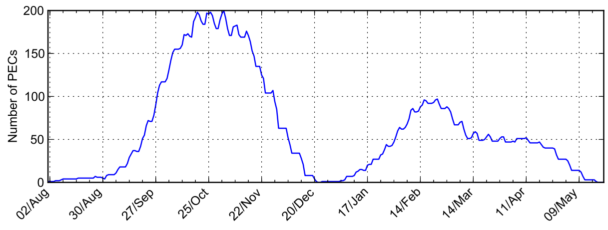


Fig. 5. Occurrences of periodic encounter communities (PECs) found in the REALITY dataset that have a period of seven days. The height of the curve at date t corresponds to $|\{ \langle C, S_{(i,p,n)} \rangle \in \mathcal{P}^* \,|\, i \leq t \leq i + (n-1)p \wedge p = 7\,\text{days} \}|$ where $\mathcal{P}^*$ is the set of all PECs extracted from REALITY with granularity $Q = 24$ hrs.

These results inform us on the speed with which nodes can discover their REC, and the speed of content sharing within the REC.

To assess the REALITY dataset we select a four-week period when students and faculty were in session (27th September to 25th October 2004) with window size $\omega = 7$ and $N = 4$. Similarly the duration selected for the DARTMOUTH is 7th April to 5th May 2003. Using techniques from [11], Figure 5 shows the variation in the number of seven-day-period PECs in the REALITY dataset. Trivial chronologies with fewer than two encounters each week are removed from both data sets, resulting in 33,484 REALITY encounters and $30,127$ DARTMOUTH encounters. A dispersion threshold $\theta = 0.2$ is applied for REC detection, with a jitter tolerance of $\phi = 30$ minutes.

The choice of θ and ϕ depends on a variety of factors. Dispersion threshold controls the point at which a time offset (e.g., hour of week) is considered regular for a pair of users. This depends on the underlying human behaviour among users, and also noise introduced by the encounter sensing technology (e.g., a particular implementation of Bluetooth hardware and software). With our datasets we found a range of $\theta = 0.15$ to 0.25 to give adequate results, although in practice a higher threshold can be selected for more-recent and reliable sensing technology. The choice of ϕ is application-driven: a narrower tolerance should be chosen for applications that rely on the precise timing of encounter events.

Five illustrative examples of RECs from the REALITY dataset are presented in Figure 6. $\mathcal{R}_1$ is a simple REC of two nodes that is regular at many points during the week. $\mathcal{R}_2$ consists of one node that regularly meets three other nodes on Mondays at 12:00 and 21:30. $\mathcal{R}_3$ is a Friday 20:00 REC whose community is a path graph and has diameter four. Finally, $\mathcal{R}_4$ and $\mathcal{R}_5$ are an interesting example of mutual non-subsumption. The triad of subjects in $\mathcal{R}_4$ is the same triad in $\mathcal{R}_5$ and both RECs have Wednesday 21:30 among their regular times. Given that $\mathcal{R}_5$ has an additional node we may incorrectly regard $\mathcal{R}_4$ as being subsumed by $\mathcal{R}_5$. However, $\mathcal{R}_4$ is regular at a second time of week that $\mathcal{R}_5$ is not (i.e., Thursday 11:00). Interestingly the result is that $\mathcal{R}_4$ does not subsume $\mathcal{R}_5$ (due to $\mathcal{R}_5$ having an additional edge and node) and $\mathcal{R}_5$ does not subsume $\mathcal{R}_4$ (due to $\mathcal{R}_4$ having additional region of regularity).

6.1 Characterization of RECs in the REALITY dataset

210 REALITY RECs were detected in the four week period. Of the 50 nodes that remained in the dataset after removing those that did not have at least two encounters each week, 38 appear in one or more RECs. This indicates that RECs are prevalent in the REALITY dataset, with only 24% of nodes not exhibiting regular encounter behaviour.

To investigate the structural size of RECs a number of community properties are presented in Figure 7. Unsurprisingly, RECs containing a large number of nodes and edges are less likely. Most-common are RECs containing two or three individuals, which accounts for 64% of the RECs.

The diameters of RECs are also relatively small, but typically larger than the diameters of the PECs extracted from the same dataset. 63% of RECs were found to have a diameter of 2 or greater, whereas only between 28% and 16% (depending on granularity Q) of PECs had diameters in this range. We suspect that the reason for the larger diameter is that the RECs tend to occur as path and tree structures. Indeed, we found that only 4% of RECs contained one or more simple cycles, indicating that the large majority are tree graphs.

To investigate the structure of the communities further we consider the distribution of graph density in Figure 8. 76 of the RECs consist of two nodes. These two-node communities are omitted from the distribution since they are by definition complete graphs and therefore always have a density of 1. We see that complete graphs among RECs with three nodes or more are rare. 44% of communities have density between 0.6 and 0.7, many of which are triad communities missing one edge. Intuitively, we would expect that if a node v meets two nodes w and p at roughly the same time each week, then w is also likely to have met p at the same time. We therefore might expect triad RECs to tend to form cliques and therefore support transitivity [25]. However, our analysis finds that only three of the 59 triad RECs are transitive.

A number of factors may contribute to intransitivity in RECs. The range of Bluetooth allows for an individual to detect two other devices that are not in range of one another. This may occur even in the case where all three individuals are stationary; however, intransitivity is more likely when one or more devices are moving. A highly mobile central node may encounter two or more individuals in succession without simultaneously being in contact with all at the same time.

RECs such as these are interesting because they either represent an individual periodically repeating the same route that brings him/her into proximity with the same individuals, or an individual who periodically acts as a bridge between two nearly proximate nodes. This finding also indicates that RECs are not necessarily cliques, highlighting a difference between our definition and that of a meeting group [26].

Regularity masks represent the temporal structure of RECs. As noted in Section 3.1, the regularity mask for a REC asserts that all edges in the community have at least one regular event within $\phi = 30$ minutes of all times covered by the regularity mask. The longer the length $|R|$ of a mask R, the more points during the week the community

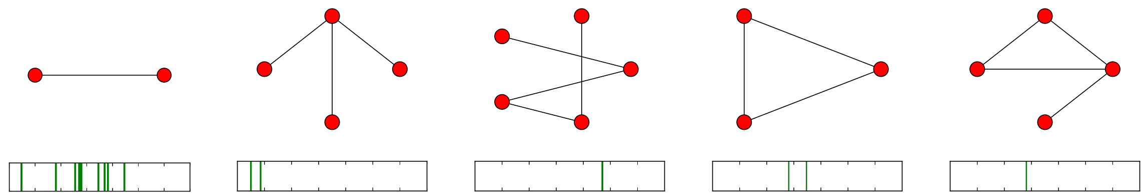


Fig. 6. Example RECs from the REALITY dataset. A regularity mask is depicted as a rectangle containing green bars. Each rectangle is separated into seven chunks, each representing a day of week beginning with Monday and ending with Sunday. Ticks denote midnight. Green bars indicate the time-of-week during which the REC is regular.

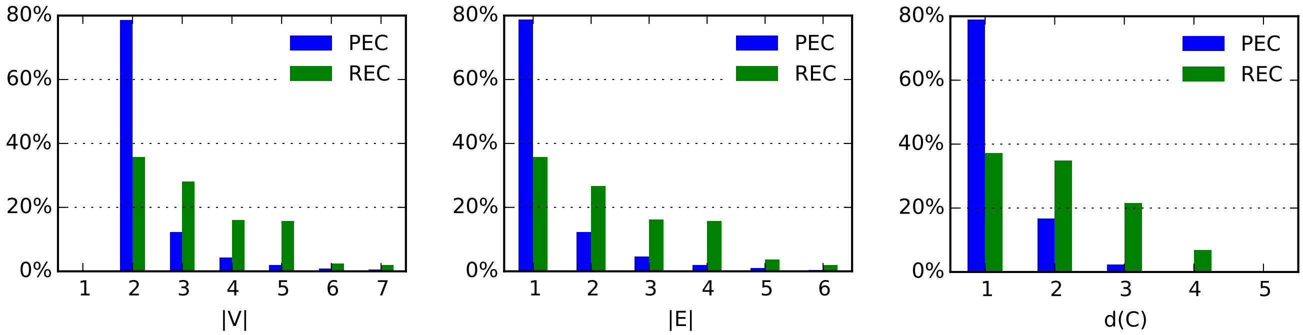


Fig. 7. Distributions of the number of nodes (denoted $|V|$), number of edges (denoted $|E|$), and diameters (denoted $d(C)$) of regular encounter communities (RECs) and periodic encounter communities (PECs) (with $Q = 4$ hrs) in the REALITY dataset.

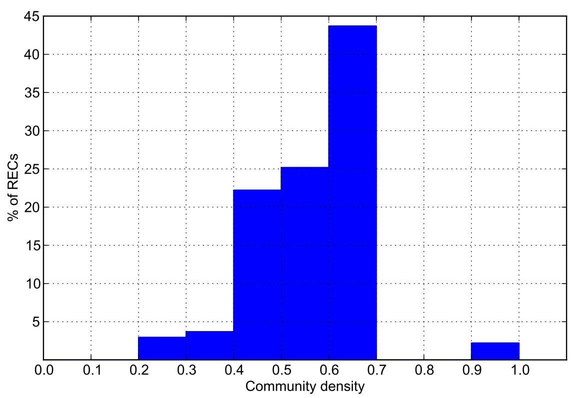


Fig. 8. Distribution of community density for REALITY RECs consisting of at least three nodes. For a REC with community $C = (V, E)$, community density is given by $\frac{2|E|}{|V|^2 - |V|}$.

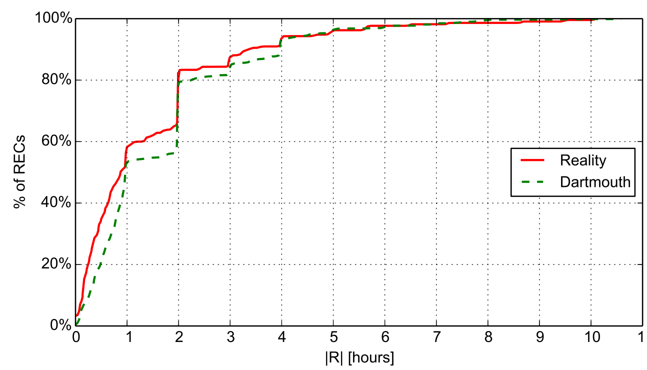


Fig. 9. Cumulative distribution of regularity mask durations in the REALITY (left) and DARTMOUTH (right) dataset. $|R|$ denotes the overall duration of a regularity mask R.

is regular for. Figure 9 shows the distribution of regularity mask lengths among the RECs extracted from REALITY. The average regularity mask length is 1.48 hours and the largest is 10.1 hours. Longer regularity mask lengths are associated with smaller diameters and, for the largest lengths, with two-node RECs. Regularity masks for RECs with more than three nodes are built from the intersection of the constituent edges, and are therefore almost always shorter in length than the two-node RECS that they subsume.

Although the distribution of regularity mask lengths tells us the overall durations which a REC is regular during the week, it does not describe the times of week where RECs are regular. We explore this further in Figure 10 which plots the times of week that are typically covered by a REC's regularity mask. The tallest peak is at 16:45 Friday, indicating that many RECs were regular at this time (in addition to any other times they may be regular). Smaller peaks also appear on each weekday at 11:00 and 16:45. These are significant times in the context of the REALITY dataset as they lie at the boundaries between MIT classes[3]. Exact class start and end times vary by day of week, but 11:00 is a common class start time and 16:30 is a common class end time. It is therefore likely that the 11:00 peaks correspond to subjects arriving at the same class each week, and that the 16:30 peaks correspond to subjects encountering one

3. MIT Fall 2004-2005 Class Schedule, via the Internet Archive: http://web.archive.org/web/20040917051038/http://web.mit.edu/registrar/www/schedules/csbindex.shtml

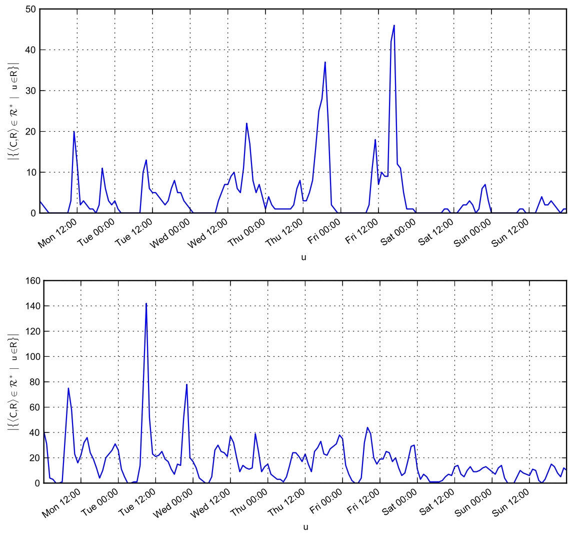


Fig. 10. Distribution of regularity masks belonging to REALITY (top) and DARTMOUTH (bottom) RECs throughout the week. Letting $\mathcal{R}^*$ denote the set of all RECs, the number of regularity masks that include the time of week u is given by $|\{\langle C, R\rangle \in \mathcal{R}^* \,|\, u \in R\}|$.

another during their commute from class or on arriving at their residences.

Figure 10 also shows that RECs are much less likely to form on weekends. The likelihood is that many students may choose to spend their weekends off campus, and therefore there is less encounter activity among participants. Also, when compared to weekdays, weekends have more erratic behaviour due to the lack of routine tasks such as timetabled lectures.

6.2 Comparison to REALITY and DARTMOUTH periodic encounter communities

RECs can be viewed as a noise-tolerant counterpart to periodic encounter communities (PECs) introduced in [11]. Although PECs and RECs are different definitions of community, it is interesting to consider whether there is any correspondence between the two.

To carry out our comparison we choose a four-week observation duration in each of the two datasets, as described in Table 2. The duration for REALITY is the same as discussed in Section 6. We select 7/Apr to 5/May 2003 for DARTMOUTH as this spans a period of uninterrupted term time for the students. Throughout this paper we consider RECs based on weekly regularity (i.e., we set $\omega = 7$ days). These are comparable to PECs that have a recurrence period of 7 days. Thus, we investigate whether the RECs in each dataset resemble any PECs that have a period of seven days. To do this, we check each PEC to see if its nodes appear in one or more REC. If the nodes of a particular seven-day-period PEC are a subset of the nodes of a REC, then we regard these two communities as being similar. For this analysis we set the PEC algorithm's temporal granularity parameter to $Q = 24$ hours, and thus each encounter snapshot represents one day.

In REALITY we find 82 PECs and 210 RECs. Our results find that 58.5% of PECs had a node set that also appeared in at least one REC. Reciprocally, we found that only 14.2% of RECs had a node set that appeared in at least one PEC. From these results we draw two conclusions. First, there is not a one-to-one correspondence between the RECs and PECs in the REALITY dataset. Second, the majority of RECs (85.8%) have no corresponding PEC, and therefore RECs capture more weekly encounter behaviour than PECs.

We find similar behaviour in DARTMOUTH, although PECs are rarer due to there being more noise in AP co-location data. In particular, we found only 12 PECs among the 428 nodes in the dataset, which contrasts with the dataset's 773 RECs. 50% of the PECs had a node set that also appeared in at least one REC, whereas 0.26% of RECs had a node set that appeared in at least one PEC.

6.3 Token broadcast in RECs

Token broadcast is used to measure the time required for a global maximal REC to be discovered by all its constituent nodes through decentralised opportunistic sharing and construction. This scenario is analogous to each node in a REC attempting to broadcast a token to each other, and therefore also represents the speed of information propagation.

6.3.1 Selecting encounters relevant to a REC

Each encounter between two nodes represents a token-sharing opportunity. When evaluating token broadcast for a particular REC, only encounters described by that REC's information are used for token sharing. Assuming the window size is one week, these are the encounters whose time-of-week offset lies within the REC's regularity mask (or close enough, according to the jitter tolerance parameter) and correspond to one of the edges in the REC's community.

More formally, consider a REC $\mathcal{R} = \langle C, R\rangle$, where $C = (V, E)$, constructed from chronologies of duration T_{max}, with window size ω and jitter tolerance ϕ. To determine the set of edges where token exchanges will occur at time $t \in (0, T_{max}]$ we consider whether the window offset $t \bmod \omega$ is within ϕ of any value in R. If so, an encounter at t that corresponds to an edge in E can be used for token exchange. The following function $f(t)$ formalises this concept and represents the mapping of the time $t \in (0, T_{max}]$ to the set of exchanges occurring at time t:

$$\begin{aligned} f(t) = \{\ \{v, w\} \mid \{v, w\} \in E \,\wedge \\ \exists\, t \in \mathcal{S}_{v,w},\ u \in R \ \text{ s.t.} \\ |u - (t \bmod \omega)| \leq \phi\ \}\,. \end{aligned}$$

To evaluate $\mathcal{R}$ we apply token exchanges described by $f(t)$ for each $t \in (0, T_{max}]$ in ascending order.

6.3.2 Broadcast time in REALITY *and* DARTMOUTH

Table 2 summarises the REALITY and DARTMOUTH datasets used in the following token broadcast experiments. We observe that RECs were less prevalent among nodes in DARTMOUTH than REALITY; in particular, 76% of REALITY nodes belonged to at least one REC, compared to 58% of DARTMOUTH nodes. This is likely due to the Reality Mining dataset being a closer representation of human encounters than the inferred Dartmouth WLAN encounters. REALITY's superior fidelity is due to it being direct-sensed data from subjects who are consciously participating in a study.

TABLE 2
Summary of datasets used in REC token broadcast experiments. Only nodes and edges that met the minimum number of encounters are included. Window width $\omega = 7$ days.

	REALITY	DARTMOUTH
Duration	27/Sept to 25/Oct 2004	7/Apr to 5/May 2003
Total nodes	50	428
Total edges	166	863
Total encounters	33,484	30,127
RECs	210	773
Nodes appearing in 1+ REC	38	247
Successful broadcasts	142	485
Average diameter	1.98	1.91

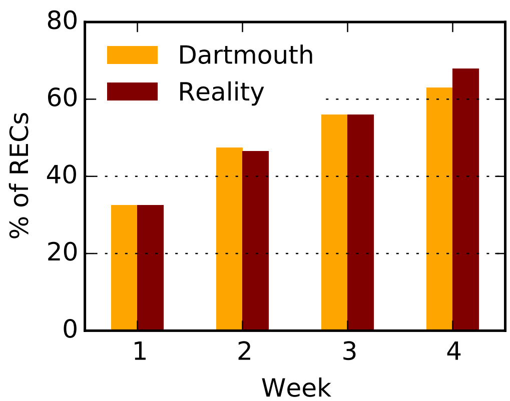


Fig. 11. Percentage of RECs that have reached full coverage (i.e., successful broadcast) at the end of each week. 32% of REALITY RECs did not reach full coverage. 37% of DARTMOUTH RECs did not reach full coverage.

There are a number of RECs which failed to reach full token coverage after four weeks. This contrasts with the token broadcast analysis of REALITY PECs, where each PEC successfully broadcast all its tokens by the last timestep of its periodic support set. Figure 11 depicts the number of RECs that reached full coverage by the end of each week. This reveals how many RECs are able to successfully broadcast after applying each week of exchanges corresponding to a regularity mask. Over the first three weeks the two datasets are almost identical in the increases in successful broadcast. By the end of week four, however, broadcast in 68% of REALITY RECs was successful, compared to 63% of DARTMOUTH RECs.

It was shown in [11] that community diameter has a strong influence on the speed of broadcast within PECs. We therefore investigate the extent to which diameter accounts for the higher failed broadcast rate in DARTMOUTH. In particular, the diameter and period of a PEC serve to limit the worst-case spreading behaviour.

RECs permit a degree of variation in the timing of the regular encounters and are tolerant to occasional missing encounters, and therefore the same strict limit does not apply; however, we do find that community diameter influences the speed of token propagation, as demonstrated in Figure 12.

We observe that all RECs with diameter one reached full coverage within 28 days and were the quickest to do so. Figure 12 shows that 68% of one-diameter REALITY RECs reached full coverage within seven days and all were at full coverage within 19 days. One-diameter RECs in the DARTMOUTH dataset have a similar full-coverage rate, with 65% reaching full coverage within seven days. These RECs are cliques and, as mentioned earlier, only require one exchange per edge to reach full coverage. This statistic also indicates that in 31% (REALITY) and 35% (DARTMOUTH) of the one-diameter RECs there was at least one pair of nodes that did not have an encounter within the REC's regularity mask in the first week.

For large diameters we see a significant reduction in both the number of successful broadcasts and the rate at which RECs reach full coverage.

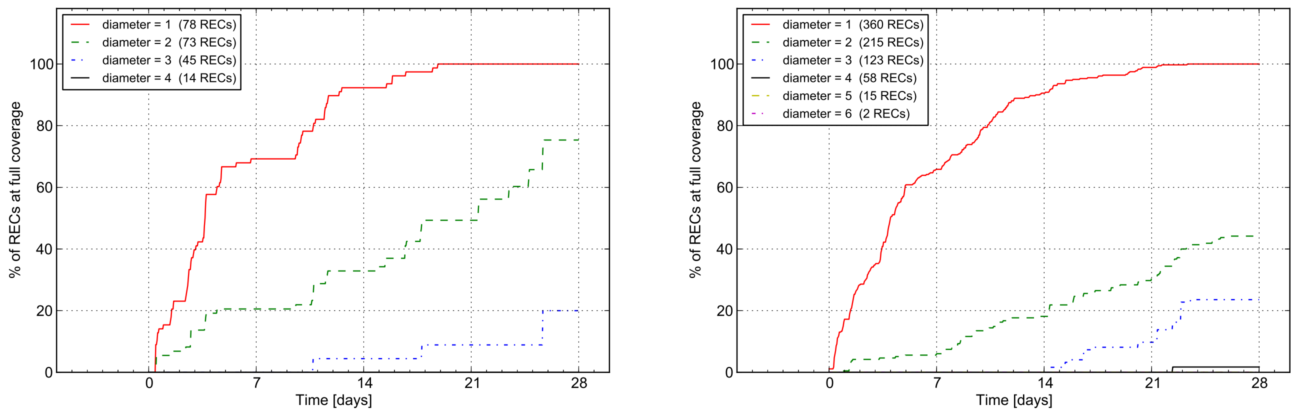
Despite no longer acting as a strict limit on broadcast behaviour, diameter is nevertheless important as large diameters indicate the presence of two nodes separated by a large number of hops. Rapid propagation across multi-hop paths such as these requires frequent and interleaved encounters among the intermediate nodes, which is rare among the RECs we have extracted. On the other hand, low-diameter RECs allow for rapid token broadcast. For example, if we consider a REC $\mathcal{R} = \langle C, R \rangle$ that is a clique, it has diameter $d(C) = 1$ and in the worst-case only requires one encounter at each edge in C before reaching full coverage.

7 RELATED WORK

Many different fields of research have modelled real-world human mobility, from universal rules [27], [28] through to models incorporating dynamic behaviour [4], [10]. Often static (non-dynamic) analysis is extended to incorporate temporal features [29]. Regularity of human encounters has particular relevance for opportunistic networks [30], where occasional short-range connections between devices are the basis for routing and content sharing. Wide ranging context have been utilised, including local network metrics (e.g., rate-of-change of connectivity in the CAR protocol [4]) and personal and location-based meta-data (e.g., HiBOp protocol [31]). The important distinction between frequency and regularity is addressed in [32], where it is found that transport network users tend to think of travel regularity as being related to destinations and time of travel, rather than amount of travel.

Collective periodic behaviour is often assessed by temporal changes to quantifiable metrics [33], [34]. These can be aggregated by time-of-day or time-of-week to analyse the daily or weekly profile of the measure [28], [34]. Work has highlighted periodic aggregate behaviour at this scale; for example, in [12] a wavelet decomposition of various time series of collective statistics shows strong daily and weekly periodicities in human encounter behaviour.

Link prediction has also been investigated. The Content Source Selection (CSS) algorithm in [10] is such an example, estimating the fraction of in-contact duration per hour as a way to model the time-of-day dependency of human encounter behaviour. The Habit protocol [5] uses a similar model, in which a regularity weight is computed as the frequency of encounters between two nodes at a given time-of-week. Moghaddam et al. [35] extend this concept further to develop a decentralised architecture for interest-based



(a) REALITY.

(b) DARTMOUTH.

Fig. 12. Percentage of RECs that have reached full coverage over time. RECs have been grouped to allow comparison of broadcast by diameter.

multicast communication between groups via opportunistic messaging. Knowledge-sharing over opportunistic networks has also been studied from an information-theoretic perspective [36], indicating the limits of information gain via interest-based communication. A number of methods have also been developed to assess predictability of visits [28], [37], [38] and encounters [39], and specifically predictability of WLAN traces [40].

Regularity in human encounters also relates to community detection. Most community detection methods are intended for offline analysis as surveyed in [41]. In relation to this paper, the most-relevant work concerns distributed community detection from Hui et al. [17], however the algorithm proposed considers aggregated graphs rather than any temporal or periodic trend in the encounter patterns. There is little prior research into *periodic* community behaviour, however. The closest example is the Habit communication protocol [5], which attempts to merge both multi-node encounter behaviour and periodicity.

Time-varying graphs have been a useful representation for temporal properties of networks, which represent a sequence of graphs capturing the time-varying nature of the network [42], [43]. This approach introduces discrete timesteps, which results in the loss of some information on the ordering of events within a particular timestep. In [44], [45] temporal analysis related to path length, network efficiency, and connected component size are presented and related to the spread of information within encounter networks. The concept of small-world networks [45] has also been applied in the context of dynamic graphs [46]. These analyses focus on temporal, but not periodic, encounter behaviour. This also applies to [1], which is nevertheless of interest as it considers temporal multi-nodal community behaviour. Other recent advances in the temporal modelling of communication networks has focused on the relationship between topology and pairwise dynamics [47]. Rather than a snapshot approach, these analyses consider patterns in the underlying event sequences, as we do in this work.

A unique example of encounter periodicity analysis is periodic subgraph mining, introduced in [24]. Rather than extending static networks, this approach adopts a data mining perspective, introducing *PSE-Miner*, a single-pass algorithm for extracting all periodic subgraphs embedded in a dynamic graph. Automatic detection of the recurrence period is particularly noteworthy. Dynamic graph approaches have been successfully applied in animal encounter networks [24], biological networks [46], and virtual networks [48]).

There have been wide-ranging work that has looked periodicity using hybrid methods and considering different scales. At the individual scale the data are more sparse and different methods are used. Methods include Markov models [49], [50], frequent sequence mining [51], and pattern matching (compression-based prediction, prediction by partial matching, and sampled pattern matching) [50]. These approaches do not model periodic temporal context, but a minority of sequence-based methods have been extended to incorporate periodic characteristics (e.g., [52]). Clauset and Eagle [33] have adapted the clustering coefficient and the node degree for a dynamic network. Analysis of the variation in encounters is also carried out in [53], where instead of looking at consecutive snapshots, the differences between encounters at the same time-of-day over different days is considered. Also at the individual scale is the *life entropy* metric, presented in [37] which measures the strength of the patterns in a user's high-level daily activities and hourly encounter rates. [22] provides insight into periodic encounter behaviour of friends and non-friends. At the collective scale, the evolution of the volume of interactions per unit time has been studied [53]. In [37] the same metric is examined, but at a finer granularity using hourly buckets instead of whole days. The temporal and periodic behaviour of other collective scale properties have been studied by Scellato et al. [12]. Wavelet decomposition reveals daily and weekly encounter cycles; i.e., multi-scale periodicity of human encounter patterns.

Signal processing techniques for modelling periodicity have also been applied, such as a Kalman filter for predicting pairwise encounters [4]. Other related literature includes methods for automatically detecting periodicities in location and encounter data. Periodica [54] does so by selecting the strongest frequency in the discrete Fourier transform in the binary sequence of visits of an individual to a location. PSE-Miner [24], on the other hand, is able to detect multiple periodicities in a stream of individual-to-individual encounters at the cost of extra computation time.

8 Conclusions

In this article we overcome disadvantages of a discrete-time representation for periodic encounter community (PEC) detection by defining communities in terms of inter-event interval patterns. The concept of a regular encounter community (REC) has been introduced, being tolerant to small variations in periodic encounter patterns and retaining the time-resolution of encounters. This generalises previous work [14] and overcomes the loss of temporal resolution and sensitivity to minor variations in the timing of periodic encounters. The decentralised construction approach introduced in [11] has been extended to allow nodes to self-detect their RECs in a distributed manner. Interestingly, this has been developed from an assessment approach for neuron synchronicity.

We have examined important benchmark data sets using our new analysis. Our results show that many individuals belong to one or more REC, making these an interesting feature for use in encounter-aware opportunistic forwarding protocols. Token broadcast analysis shows that diameter is an important factor in the propagation of information, a factor that is also important for periodic encounter communities. The requirement that the community's encounters must strictly repeat according to the identified period means that diameter acts as a hard limit on the PEC's broadcast time. This contrasts with RECs, whose tolerance to minor variations in weekly patterns permits occasional missing encounters. We found that due to this, and also due to the limited (four week) duration we allowed for propagation, not all REC will reach full token coverage.

In practice, REC construction would be faster and have a higher success rate by allowing communities to also use their irregular encounters for opportunistic REC construction; however, for our experiments we restricted our evaluation to encounters intrinsic to each REC so that we could investigate the propagation characteristics specific to the community.

When directly comparing PECs and RECs of the same periodicity (i.e., weekly), we found that REC detection identified over 2.5 times the number of encounter communities than PEC detection. Given that human mobility is not a strictly timed behaviour, it is not surprising that permitting an amount of uncertainty in encounter patterns allows us to capture more periodic communities, and result in more RECs being detected. Indeed, RECs were able to account for 58.5% of the communities extracted by PEC detection, and also were able to identify an additional 180 communities (of 210 overall RECs) that did not appear as PECs. These results provide insight and proof of concept for community detection in human encounter networks, where the presence of temporal patterns that lead to periodic components in the network are explicitly included.

Acknowledgments

This work has been supported by RECOGNITION grant 257756, an European Commission FP7 FET project.

References

[1] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora, "Components in Time-varying Graphs," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 2, p. 023101, 2012.

[2] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," *Nature Communications*, vol. 5, 2014.

[3] A. Barrat, C. Cattuto, V. Colizza, F. Gesualdo, L. Isella, E. Pandolfi, J.-F. Pinton, L. Ravà, C. Rizzo, M. Romano, J. Stehlé, A. E. Tozzi, and W. V. d. Broeck, "Empirical temporal networks of face-to-face human interactions," *European Physical Journal Special Topics*, vol. 222, no. 6, pp. 1295–1309, 2013.

[4] M. Musolesi and C. Mascolo, "CAR: context-aware adaptive routing for delay-tolerant mobile networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 2, pp. 246–260, 2009.

[5] A. J. Mashhadi, S. B. Mokhtar, and L. Capra, "Habit: leveraging human mobility and social network for efficient content dissemination in MANETs," in *Proc. 2009 IEEE Int. Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2009, pp. 214–219.

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

[7] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: a survey of the state of the art and research challenges," *ACM Computing Surveys*, vol. 47, no. 3, pp. 47:1–47:29, Apr. 2015.

[8] R. M. Whitaker, M. Chorley, and S. M. Allen, "New frontiers for crowdsourcing: the extended mind," in *48th Hawaii Int. Conf. on System Sciences (HICSS)*, 2015, pp. 1635–1644.

[9] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proceedings of the National Academy of Sciences*, vol. 110, no. 34, pp. 13 774–13 779, 2013.

[10] L. McNamara, C. Mascolo, and L. Capra, "Media sharing based on colocation prediction in urban transport," in *Proc. 14th ACM Int. Conf. on Mobile Computing and Networking*. ACM, 2008, pp. 58–69.

[11] M. J. Williams, R. M. Whitaker, and S. M. Allen, "Decentralised detection of periodic encounter communities in opportunistic networks," *Ad Hoc Networks*, vol. 10, no. 8, pp. 1544–1556, 2012.

[12] S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, "On nonstationarity of human contact networks," in *Proc. 2nd Workshop on Simplifying Complex Networks for Practitioners (SIMPLEX 2010)*, 2010.

[13] J. McInerney, S. Stein, A. Rogers, and N. R. Jennings, "Breaking the habit: Measuring and predicting departures from routine in individual human mobility," *Pervasive and Mobile Computing*, vol. 9, no. 6, pp. 808–822, 2013.

[14] M. J. Williams, R. M. Whitaker, and S. M. Allen, "Measuring individual regularity in human visiting patterns," in *Proc. 2012 ASE/IEEE Int. Conf. on Social Computing (SocialCom)*, 2012.

[15] T. Kreuz, D. Chicharro, R. G. Andrzejak, J. S. Haas, and H. D. I. Abarbanel, "Measuring multiple spike train synchrony," *Journal of Neuroscience Methods*, vol. 183, no. 2, pp. 287–299, 2009.

[16] T. Kreuz, D. Chicharro, M. Greschner, and R. G. Andrzejak, "Time-resolved and time-scale adaptive measures of spike train synchrony," *Journal of Neuroscience Methods*, vol. 195, no. 1, pp. 92–106, 2011.

[17] P. Hui, E. Yoneki, S.-Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proc. 2nd ACM/IEEE Int. Workshop on Mobility in the Evolving Internet Architecture*. ACM, 2007, pp. 1–8.

[18] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576–1589, 2011.

[19] S. M. Allen, M. J. Chorley, G. B. Colombo, and R. M. Whitaker, "Opportunistic social dissemination of micro-blogs," *Ad Hoc Networks*, vol. 10, no. 8, pp. 1570–1585, Nov. 2012.

[20] S. M. Allen, M. J. Chorley, G. B. Colombo, E. Jaho, M. Karaliopoulos, I. Stavrakakis, and R. M. Whitaker, "Exploiting user interest similarity and social links for micro-blog forwarding in mobile opportunistic networks," *Pervasive and Mobile Computing*, vol. 11, pp. 106–131, Apr. 2014.

[21] L. Rossi, M. Williams, C. Stich, and M. Musolesi, “Privacy and the City: User Identification and Location Semantics in Location-Based Social Networks,” in *Ninth International AAAI Conference on Web and Social Media*, Apr. 2015.
[22] N. Eagle, A. Pentland, and D. Lazer, “Inferring Friendship Network Structure by using Mobile Phone Data,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15 274–15 278, 2009.
[23] T. Henderson, D. Kotz, and I. Abyzov, “The changing usage of a mature campus-wide wireless network,” *Computer Networks*, vol. 52, no. 14, pp. 2690–2712, 2008.
[24] M. Lahiri and T. Berger-Wolf, “Periodic subgraph mining in dynamic networks,” *Knowledge and Information Systems*, vol. 24, no. 3, pp. 467–497, 2009.
[25] P. W. Holland and S. Leinhardt, “A method for detecting structure in sociometric data,” *American Journal of Sociology*, vol. 76, no. 3, pp. 492–513, 1970.
[26] E. Yoneki and D. Greenfield, “Inferring significance of meeting groups in human contact networks,” in *Proc. 2010 European Conference on Complex Systems (ECCS 2010)*, 2010.
[27] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabási, “Understanding individual human mobility patterns,” *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
[28] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
[29] P. Holme, “Analyzing temporal networks in social media,” *Proc. IEEE*, vol. 102, no. 12, pp. 1922–1933, 2014.
[30] L. Pelusi, A. Passarella, and M. Conti, “Opportunistic networking: data forwarding in disconnected mobile ad hoc networks,” *Communications Magazine, IEEE*, vol. 44, no. 11, pp. 134–141, 2006.
[31] C. Boldrini, M. Conti, and A. Passarella, “Exploiting users’ social relations to forward data in opportunistic networks: the HiBOp solution,” *Pervasive and Mobile Computing*, vol. 4, no. 5, pp. 633–657, 2008.
[32] N. Lathia and L. Capra, “How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives,” in *Proc. 2011 Int. Conf. on Ubiquitous Computing (UbiComp)*. ACM, 2011, pp. 291–300.
[33] A. Clauset and N. Eagle, “Persistence and periodicity in a dynamic proximity network,” in *Proc. DIMACS Workshop on Computational Methods for Dynamic Interaction Networks*, 2007.
[34] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, “Uncovering individual and collective human dynamics from mobile phone records,” *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 22, p. 224015, 2008.
[35] S. Moghaddam and A. Helmy, “Interest-aware Implicit Multicast (iCast): Opportunistic Mobile Data Dissemination Without Per-group Management,” in *Proc. 9th ACM Workshop on Mobility in the Evolving Internet Architecture*. ACM, 2014, pp. 55–60.
[36] M. ElSherief, T. ElBatt, A. Zahran, and A. Helmy, “An Information-Theoretic Model for Knowledge Sharing in Opportunistic Social Networks,” in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 2015, pp. 446–451.
[37] N. Eagle and A. Pentland, “Reality mining: sensing complex social systems,” *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.
[38] J. McInerney, S. Stein, A. Rogers, and N. R. Jennings, “Exploring periods of low predictability in daily life mobility,” in *Proc. Workshop on the Nokia Mobile Data Challenge, in Conjunction with 10th Int. Conf. on Pervasive Computing (PerCom)*, 2012.
[39] M. Musolesi and C. Mascolo, “Evaluating context information predictability for autonomic communication,” in *Proc. 2006 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2006, pp. 495–499.
[40] J. Kim and A. Helmy, “Analysing the Mobility, Predictability and Evolution of WLAN Users,” *Int. J. Autonomous and Adaptive Communications Systems*, vol. 7, pp. 169–191, 2014.
[41] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.
[42] P. Holme and J. Saramäki, “Temporal networks,” *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012.
[43] M. J. Williams and M. Musolesi, “Spatio-temporal networks: reachability, centrality and robustness,” *Royal Society Open Science*, 2016, to Appear.
[44] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical Review Letters*, vol. 87, no. 19, p. 198701, 2001.
[45] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-world networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
[46] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, “Small-world Behavior in Time-varying Graphs,” *Physical Review E*, vol. 81, no. 5, p. 055101, 2010.
[47] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, “Small but Slow World: How Network Topology and Burstiness Slow Down Spreading,” *Physical Review E*, vol. 83, no. 2, Feb. 2011.
[48] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, “Characterising temporal distance and reachability in mobile and online social networks,” *SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 118–124, 2010.
[49] D. Ashbrook and T. Starner, “Learning significant locations and predicting user movement with GPS,” in *Proc. 6th Int. Symposium on Wearable Computers*, 2002, pp. 101–108.
[50] L. Song, D. Kotz, R. Jain, and X. He, “Evaluating next-cell predictors with extensive Wi-Fi mobility data,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1633–1649, 2006.
[51] G. Yavas, D. Katsaros, O. Ulusoy, and Y. Manolopoulos, “A data mining approach for location prediction in mobile environments,” *Data and Knowledge Engineering*, vol. 54, no. 2, pp. 121–146, 2005.
[52] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, “A hybrid prediction model for moving objects,” in *Proc. IEEE 24th Int. Conf. on Data Engineering (ICDE)*, 2008, pp. 70–79.
[53] A. Miklas, K. Gollu, K. Chan, S. Saroiu, K. Gummadi, and E. de Lara, “Exploiting social interactions in mobile systems,” in *Proc. 2007 Int. Conf. on Ubiquitous Computing (UbiComp)*, 2007, pp. 409–428.
[54] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, “Mining periodic behaviors for moving objects,” in *Proc. 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2010, p. 1099.

Matthew J. Williams received the BSc degree in computer science from Cardiff University, UK in 2008. He received the PhD in computer science degree from the same institution in 2013. He is currently a Postdoctoral Research Fellow at the University of Birmingham and visiting researcher at University College London. His research interests include spatio-temporal complex systems, human mobility and dynamics, and mobile communication networks.

Roger M. Whitaker is Professor of Mobile and Biosocial computing at Cardiff University, UK. His research concerns in smartphones, networks and human nature. He acts as an associate editor/editorial board member for journals including Social Network Analysis and Mining, Telecommunication Systems and Wireless Networks.

Stuart M. Allen is a Professor and Head of the School of Computer Science & Informatics at Cardiff University UK. He has a BSc in Mathematics from Nottingham University UK and a PhD in Graph Theory from the University of Reading UK. His research interests are in the application of social structures to problems in mobile networks.

Statutory notices: This document is the postprint version of the article. The publisher version can be accessed through the following DOI: 10.1109/TMC.2016.2599169