On the Analysis of Periodic Mobility Behaviour

Zhiyu Wang, Mario A. Nascimento and M.H. MacGregor
Department of Computing Science
University of Alberta, Canada
{zhiyu1, mn, macg}@cs.ualberta.ca

ABSTRACT

Encounters between devices can be very useful for message routing in mobile networks where there is no fixed infrastructure. In this context, understanding user mobile behaviour is essential to design effective and efficient network protocols. This paper presents a generic methodology to model and find periodic encounter patterns. Using this methodology and the Nokia-MDC datasets, we are able to find strong weekly and daily periodic behaviours that last up to a few months. The experimental results show that combining periodic behaviour from different network types, e.g., Bluetooth, GSM, and WLAN networks, is capable of providing communications to a large-scale network. Furthermore, we also investigate whether the network formed by periodic encounters has a small-world structure.

Keywords

Mobility, Periodicity, Connectivity, Small-world

1. INTRODUCTION

Current wireless services mostly depend on network infrastructure. In many scenarios, ad-hoc networks can provide new opportunities for communication by removing the need for network infrastructure. However, communications in such networks could be unreliable because there may not be a complete path between the source and the destination(s) from time to time. Therefore, we seek to explore useful characteristics and to discover new techniques to allow communication in ad-hoc networks.

Mobile devices are carried by people and/or animals. They typically do not move randomly and often exhibit periodically repeated movements. For example, people often use the same route daily to commute to/from work. As a result, each mobile node may encounter a set of other nodes periodically, for instance at approximately the same time every weekday. Such periodic encounters can be used to establish communication routes in mobile networks.

This material was prepared for the Mobile Data Challenge 2012 (by Nokia) Workshop; June 18-19, 2012; Newcastle, UK.

Thus, in this paper we propose a methodology to find periodic behaviour within the Nokia-MDC datasets [7] and examine connectivity among mobile nodes with periodic behaviour. Nonetheless, our methodology is applicable to other mobile traces as well. In summary, this paper has three main contributions: 1) we observe strong weekly and daily periodic behaviour within the datasets; 2) we evaluate the persistence of detected periodic behaviour; 3) we also discuss why, unlike other networks formed by periodic encounters, the one yielded by the NOKIA-MDC datasets does not lead to a small-world.

The remainder of this paper is structured as follows. A brief introduction about the datasets and their preparation is discussed in Section 2. In Section 3, we present detection techniques and experimental results. Section 4 presents discoveries regarding the small-world structure. Finally, conclusions are presented in Section 5.

2. DATASETS AND PREPARATION

The Nokia-MDC datasets include mobility datasets for different network types. The three types of datasets are Bluetooth, GSM and WLAN. Bluetooth traces track direct encounters between mobile nodes. Both GSM and WLAN traces record associations between mobile nodes and either access points or cellular towers. Even though access points and cellular towers are stationary, they can link mobile nodes. that never directly encounter. Therefore, we also include stationary devices as nodes in the network where associations are considered to be encounters as well.

In the Nokia-MDC datasets, encounters are grouped by user, where each encounter has a corresponding timestamp. From that, we need to define the duration of each encounter. In Figure 1, mobile node A encounters nodes B and C at timestamp T_i. At the next timestamp, A only meets node C. We assume that A stops seeing B right before T_{i+1}. Therefore, the encounter for node pair {A,B} starting at T_i has a duration of $T_{i+1} - T_i$ time units. Similarly, the encounter for node pair {A,C} starting at T_i has a duration of $T_{i+2} - T_i$ time units. Using this approach, there are two complete encounters, {A, B} and {A, C}, starting at T_i.

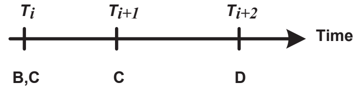


Figure 1: Encounter history of mobile node A

Given a granularity, e.g., 1 hour or 1 day, a mobile trace can be broken into consecutive disjoint intervals. Then we create an encounter series for every unique pair of nodes where the length of the series equals the number of pairs of intervals after partitioning the trace at a chosen granularity. For a pair of nodes, if they have an encounter during a specific interval, we record a 1 in that interval, otherwise, a 0 is recorded.

Given a monitoring duration σ and a granularity τ, the encounter series for a pair of nodes {x,y}, is defined as:

$$s_{x,y} = \{d_1, ..., d_n\} \text{ where } n = \lceil \frac{\sigma}{\tau} \rceil \text{ and } d_j = \{0, 1\}\ \forall j \in [1, n]$$

In this paper, our analyses focus on 1 hour and 1 day granularity. Table 1 presents the summary of the datasets and encounter series.

Network Type	NN	NS	Hourly ANE	Daily ANE
Bluetooth	126235	179606	4.24	1.55
GSM	23644	45839	13.01	4.96
WLAN	126969	257491	8.86	3.59

NN=Number of Nodes
NS=Number of Series
ANE=Average Number of Encounters [1]

Table 1: Overall summary of all traces

3. PERIODICITY

Using binary encounter series as input, our objective is to find periodic behaviour. In this section, we present our detection techniques and observations on the Nokia-MDC datasets.

By examining the cumulated re-appearance probabilities of nodes to specific locations [4, 6, 10] and the cumulated probabilities of users to return to a certain location after a certain amount of time [1], previous work revealed periodic behaviour within user mobility. However, finding a route in a mobile network, using a store-and-forward approach, needs more than that. First, we need to find out which pairs of nodes exhibit periodic behaviours. Second, among encounters between two users, we need to know which encounters are periodic encounters, and which encounters are just occasional or even random encounters. A route may fail if an occasional or random encounter is selected and therefore should be avoided.

Recent studies aimed at detecting pair-wise periodic encounters. One study reported spectral analyses of periodicities of association patterns from access points [8]. Another presented spectral analyses of mobility traces to quantize the regularity and periodicity of access points nodal encounters [9]. The fundamental tool used in both works is Discrete Fourier Transformation. However, one of the main problems from using Discrete Fourier Transformation is the artifact that affects the accuracy of period detection. For example, the most significant period could be 7 days and 8 hours instead of integers as 7 days or 8 days. How to correctly interpret fractional period is itself a non-trivial (and a domain-dependent issue). In addition to spectral analyses, techniques from data mining can also retrieve locally/partially periodic behaviour that persists only for a very short period of time, whose long-term persistence is unknown [2]. Targeting at long-term periodic behaviour within real mobility traces, our methods extract the specific encounter patterns (both period and phase) for pairs of nodes that meet each other periodically whereas other work has stopped short of this.

Our fundamental tool is the auto-persistence function [11, 12]. It calculates the conditional probabilities of the different combinations of two values in a binary time series that are separated by a given lag k, i.e., that is k positions away in encounter series. There are, of course, four combinations of two binary values. We concentrate on the combination where two intervals k-lags away from each other have value 1, i.e., containing encounters. We define the auto-persistence function (APF) as:

$$APF(k) = P(d_{t+k} = 1 | d_t = 1) \tag{1}$$

where k is the lag between the two intervals, and t represents the interval number.

As an empirical counterpart to the auto-persistence function, the auto-persistence graph (APG) is defined as:

$$APG(k) = \frac{n}{n-k} \frac{\sum_{t=0}^{n-k} I\{d_{t+k} = 1, d_t = 1\}}{\sum_{t=0}^{n} I\{d_t = 1\}} \tag{2}$$

where n is the size of encounter series, k is the lag, and I is the binary indication function, whose value equals 1 if and only if the condition is satisfied.

As an example, Figure 2 shows the APG for an encounter series derived from WLAN traces with a 1 day granularity. Three peaks can be found at lags 7, 14 and 21. These lags represent exactly 7 days, 14 days and 21 days, which appear at multiples of the minimum lag of 7 days. This repetition in the APG comes from the underlying periodicity in the encounter series. If an encounter series has periodic behaviour with length p, then this encounter series also has periodic behaviour with length $2p$, $3p$ and so on.

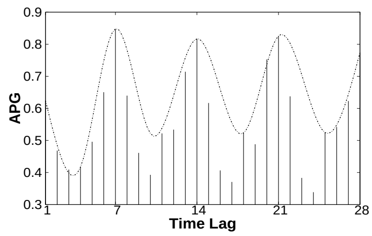


Figure 2: APG for a real encounter series

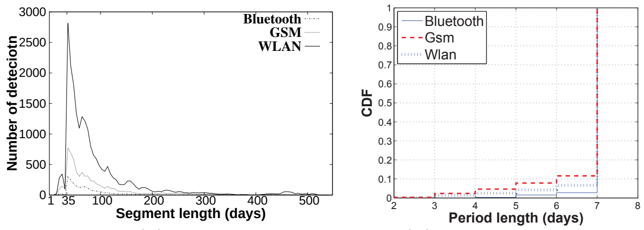
3.1 Periodic behaviour

The duration of the Nokia-MDC datasets is 13128 hours, which covers 548 days. After transforming a mobility trace into encounter series for every encountering node pair, we obtain two sets of encounter series. One with 548 values is the daily encounter series, and the other with 13128 values is the hourly encounter series.

We applied our detection techniques to both daily and hourly encounter series. Unfortunately, none of the encounter series exhibits periodic behaviour for the full duration of the series. However, this does not mean that there is no periodic behaviour that lasts for shorter periods. We will discuss the reason for this shortly.

To discover short-term periodic behaviour, we break the traces into disjoint segments with the same length. For example, each daily encounter series could be broken down into 19 segments where each segment is 30 days in length.

[1]ANE=$\frac{1}{NS}\sum_{s_{x,y}\in S}\sum_{d_i\in s_{x,y}} d_i$ where S contains all encounter series.



(a) daily (b) 35-day segments

Figure 3: Detected periodic behaviour

For daily and hourly encounter series, we apply our detection techniques to segments with length varying from 1 day to 548 days or 24 hours to 13128 hours with a 1 day or 24 hour increment.

As shown in Figure 3(a), 35-day segments in the daily encounter series have the largest number of detected periodic encounters. Similarly, even though not shown here, we observed that 120-hour segments in the hourly encounter series return the largest number of detected periodic series. To understand these periodic behaviours in daily encounter series, Figure 3(b) provides more details about detected periodic behaviour by presenting the Cumulative Distribution Function (CDF) of their period. The results indicate that there is a strong weekly encounter behaviour in Figure 3(b) in daily encounter series. Similarly, a 24 hour pattern could also be seen in the hourly encounter series.

Theoretically, our detection techniques require the length of each segment to be a minimum of four times the period length. For example, if there is a 7-day periodic behaviour, our techniques need a minimum of 28-day segment to detect such periodicity. With a 28-day segment, the APG can calculate correlations within encounter series from lag 1 up to lag 14 where two peaks can be observed at lag 7 and 14. However, because the APG stops at lag 14, we do not know whether the curve at lag 14 is going to climb or drop at the next lags, i.e., lag 15, 16 and so on. In other words, we can not decide whether the peak at lag 14 is caused by the repetition of underlying periodicity. Therefore, instead of using 28-day segments, our detection techniques requires longer segments. Because 35 days is the next duration greater than 28 that perfectly accommodate 7-day periodicities, 35-day segments have the largest number of detected periodic encounters even though they do not match to any commute schedule.

3.2 Persistence of periodic behaviour

Once the period length is known, we can extract periodic encounters from the original segments using alignment. For alignment, each segment is split into disjoint sub-segments where the length of each sub-segment equals the period length previously detected. Then, we stack all sub-segments vertically to examine the probability that encounters appear at a particular column. When the probability of encounters in a column is greater than a user-defined threshold, we consider that encounter as a periodic encounter. After finding all periodic encounters, we examine how long those periodic encounters reliably appear in the future. We call this time range the projected persistence.

After extracting periodic encounters, we evaluate their projected persistence by using the remaining encounter series right after the segment. For example, if the very first

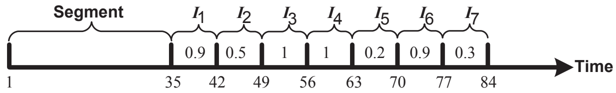


Figure 4: Projected persistence evaluation

35-day segment has periodic encounters, we use the remaining 513 days to evaluate the persistence of detected periodic encounters. Similarly, if the second 35-day segment has periodic encounters, we use the remaining 478 days to evaluate persistence. We first partition the remaining encounter series into disjoint intervals, I, whose size equals the previously detected period. To evaluate projected persistence, we introduce two thresholds: 1) the matching probability of periodic encounters within an interval, θ, and 2) the acceptable number of intervals without periodic encounters, Δ. Given an interval, θ measures the probability that this interval has periodic encounters. For example, if detected periodic encounters take place every Monday and Friday, and a given interval only has an encounter on Monday, then $\theta = 1/2 = 0.5$ for the given interval. Because of unexpected events, periodic encounters may get interrupted. Therefore, Δ is used to tolerate the number of intervals that suffer from missing periodic encounters because of interruptions.

a) $\Delta = 0$

θ threshold	35-day	120-hour
0.1	54.68 days	563.75 hours
0.3	39.9 days	469.39 hours
0.5	29.01 days	252.54 hours
0.7	22.28 days	124.94 hours
0.9	14.59 days	64.77 hours

b) $\theta = 0.9$

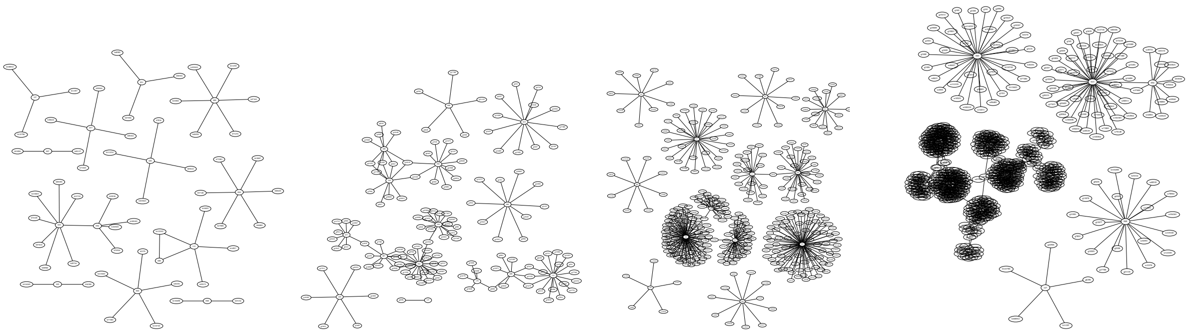
Δ threshold	35-day	120-hour
0	14.59 days	64.77 hours
1	17.29 days	81.61 hours
2	23.41 days	103.78 hours
3	29.17 days	138.76 hours

Table 2: Projected persistence

Using Figure 4 as an example, let us assume that from the very first 35-day segment, we detected periodic encounters repeating every 7 days. Since the period length is 7 days, we partition the remaining encounter series into intervals whose size equals 7 days and calculate θ for each interval. Given thresholds at $\theta = 0.9$ and $\Delta = 0$, only I_1 satisfies the θ threshold; therefore, the projected persistence is 7 days. If thresholds are $\theta = 0.9$ and $\Delta = 1$, by treating I_2 as the tolerated non-periodic interval, I_4 is the interval furthest in the future for which the thresholds are satisfied, therefore, the projected persistence is 28 days. Similarly, if thresholds are $\theta = 0.9$ and $\Delta = 2$, the projected persistence is 42 days.

Table 2 shows the average projected persistence for both 35-day and 120-hour segments. On one hand, as θ increases, the average persistence decreases because we are looking for intervals with highly persistent and stable periodic encounters. On the other hand, as Δ increases, we have longer projected persistence because we are more tolerant of interruptions in the patterns.

This explains why we did not detect any long-term periodic behaviour in the full-length traces: their persistence gets interrupted by unexpected events from time to time. As shown in Table 2, for a certain value of Δ, periodic behaviour can last up to a few months on average, i.e., 35+29.17=64.17



(a) Bluetooth network (b) GSM network (c) WLAN network (d) Integrated network

Figure 5: Connectivity in different network types

days (segment length plus projected persistence).

Due to human factors, the limitation of hardware features, services running behind different network platforms and other causes, the persistence of periodic behaviour could be affected. For example, even though a person stays at the same place, it does not mean that he/she will always connect to the same cellular tower. However, determining the causes of such short-term periodic behaviour is beyond the scope of this paper.

3.3 Network connectivity

We have also built graphs based solely on periodic encounter series. The connectivity graph $G = \langle V, E \rangle$ is an undirected graph, which is defined as follows. Given a set of encounter series with periodic behaviour, P, V is the set of nodes participating in P. For a pair of nodes u and v, the edge $e_{u,v} \in E$ if and only if $s_{u,v} \in P$.

Using 35-day segments, the connectivity graphs shown in Figure 5 represent connectivities among nodes with periodic behaviour from different network types. In this figure, connection graphs derived from Bluetooth, GSM and WLAN traces consist of several star shapes. In the Nokia-MDC datasets, we have a total of 38 participants whose encounters are recorded. As a result, the center of each cluster is one of the participants whereas leaf nodes are external MAC addresses whose encounter series is unknown, except for one encounter with a non-leaf node. All three graphs are composed of multiple isolated clusters where nodes from different network types cannot communicate with each other.

In real life there are various co-existing network types. For example, in the Nokia-MDC datasets we have cellular networks, wireless local area networks and Bluetooth encounters. Mobile nodes could exhibit different periodic behaviour within each network type. As shown in Figure 5(d), integrating periodic encounter series from different network types presents a better connected graph, which could substantially improve the effectiveness, and likely efficiency, of communication overall. Nonetheless, there still exist isolated clusters. How to provide services between different clusters in large-scale networks requires further research.

3.4 Modelling Periodicity

To utilize periodic encounter for routing in mobile networks, our detection techniques provide three important information: 1) the pair of mobile nodes who exhibits periodic behaviour, 2) the period length of each periodic behaviour and 3) the phases containing periodic encounters. If we combine detected periodicities from all mobile nodes in the network, a connectivity network, i.e., a graph model, can be created to represent all periodic encounters. During an encounter, messages can be exchanged between the two encountering nodes. Further, mobile nodes can relay messages among themselves serving as temporary storages. Therefore, the problem to be solved is to establish a route between a source and a destination (unicast), a set of destinations (multicast) or all destinations (broadcast) with respect to minimize network constraints such as the energy cost, the delivery delay and nodes' storage space, where only change to different constraints in the network is in the assignment of weights for the edges. With the connectivity network, all these problems can be converted to classical graph problems such as the shortest path problem, the minimum weighted Steiner tree problem and the minimum spanning tree problem. For instance, our initial proposal [13] for the unicast problem required a Shortest Path solution on a connectivity graph (that could be obtained by using the methodology presented in this paper).

4. SMALL-WORLD

In this section, we examine the properties of the connectivity graphs by focusing on Figure 5(d). In that graph, we have several clusters. Since some analyses can only be performed on connected graphs, we present results for the largest cluster in the graph.

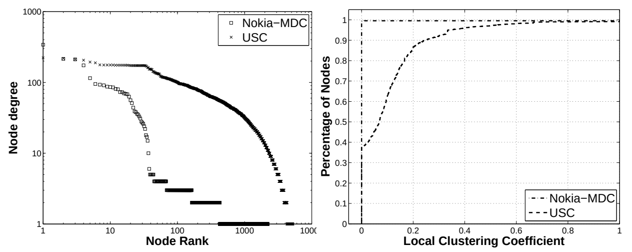
4.1 Degree Distribution

The largest cluster shown in Figure 5(d) has $N = 2025$ nodes. We rank nodes in descending order based on their degree. Plotting the sorted nodes in Figure 6(a), the first 38 nodes follow the power-law distribution. This number matches the total number of participants in the Nokia-MDC datasets whereas all other nodes are external nodes whose complete mobility is unknown. This strongly suggests that by building a network using nodes with periodic behaviour, we can obtain a scale-free network rather than a random network.

To show that networks formed by nodes with periodic behaviour can have a small-world structure, we present our studies of another dataset, the USC traces [5]. In that connectivity graph, the largest cluster contains 4961 nodes. As shown in Figures 6(a) and 6(b), the node degree distribution follows the power-law distribution. Therefore, they come from a scale-free network. We also calculate the average path length for that network. Its value at 3.34 is a good indication of the small-world structure, where 3.34 ≈

4.2 The small-world network

The average node degree in the network is $d = 2.51$. To investigate the small-world structure, we find that the average path length, L, between any pair of nodes is 3.12. This average path length is approximately equal to the logarithm of the number of nodes in the network, $3.12 \approx log(2025) = 3.30$. This is a good indication that the network has a small-world structure [14]. In addition, we calculate the Clustering Coefficients (CC) for all nodes in the network. There are two types of CC, local and global, which measure the number of triangles with respect to the number of open triplets in the graph [3, 14]. In other words, CCs measure how close the graph is to a clique. Our calculations show that 99% of the nodes in the Nokia-MDC datasets are external nodes whose local CC equals zero as shown in Figure 6(b). Therefore, the global CC of the network equals the relatively small value of 0.0015. If we consider a node in a random graph, the probability that two of its neighbours are connected is equal to the probability that two randomly selected nodes are connected. Consequently the global CC of a random graph equals $\frac{d}{N}$. Since the global CC for the network in the Nokia-MDC datasets is very close to the one from a random graph, $\frac{2.51}{2025} = 0.0012$, we conclude that the network does not have a small-world structure even though the average path length still satisfies the small-world phenomenon. One possible reason is that the traces of external nodes are unknown where we may lose periodic encounter series among them. As a result, the graph still consists of multiple star-shaped components.



(a) Node degree rank (b) Local clustering coefficient

Figure 6: Comparison between Nokia-MDC and USC datasets in log-log scale

$log(4961) = 3.69$. In addition, given the average node degree of 19.44, its global CC, 0.139, is much greater than the one from a random graph $\frac{19.44}{4961} = 0.0039$. Therefore, we conclude the network formed by the periodic nodes in the USC traces have a small-world structure.

5. CONCLUSIONS

In this paper we investigate the problem of finding encounter patterns that could be used to further improve communications in a mobile network setting. We proposed techniques to determine periodic encounters among mobile devices that can be used in any mobility trace/dataset. Using the Nokia-MDC as a sample instance, we were able to identify strong weekly and daily patterns (with different lags and phases). In addition we were also able to determine how stable and persistent a pattern is, which explains why one is not able to find a pattern that lasts the whole duration of the mobility trace. To understand the properties of networks formed by periodic encounter series, our studies show that networks are scale-free networks, and some of them satisfy a small-world structure where messages between any pair of nodes can be delivered through a very small number of hops in the network.

6. ACKNOWLEDGEMENTS

Research partially supported by NSERC's DIVA Strategic Network.

7. REFERENCES

[1] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi. Understanding individual human mobility patterns. *NATURE*, 453:479, 2008.

[2] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. In *Proc. Int. Conf. on Data Engineering*, pages 106–115, 1999.

[3] S. Holland, P. W.; Leinhardt. Transitivity in structural models of small groups. *Comparative Group Studies*, (2):107–124, 1971.

[4] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE/ACM Transactions on Networking*, 17(5):1564 –1577, 2009.

[5] W. jen Hsu and A. Helmy. CRAWDAD data set usc/mobilib (v. 2008-07-24). Downloaded from http://crawdad.cs.dartmouth.edu/usc/mobilib, July 2008.

[6] W. jen Hsu and A. Helmy. On nodal encounter patterns in wireless lan traces. *IEEE Transactions on Mobile Computing*, 9(11):1563 –1577, 2010.

[7] J. K. et al. The mobile data challenge: Big data for mobile computing research. In *Proc Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing*, Newcastle, June 2012.

[8] M. Kim and D. Kotz. Periodic properties of user mobility and access-point popularity. *Personal Ubiquitous Computation*, 11:465–479, 2007.

[9] S. Moon and A. Helmy. Understanding periodicity and regularity of nodal encounters in mobile networks: A spectral analysis. In *Proceedings of IEEE Global Telecommunications Conference*, pages 1–5, 2010.

[10] C. Song, Z. Qu, N. Blumm, and A.-L. BarabÃąsi. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[11] Startz and Richard. Binomial Autoregressive Moving Average Models With an Application to U.S. Recessions. *Journal of Business and Economic Statistics*, 26(1):1–8, 2008.

[12] C. Wang and W. K. Li. On the autopersistence functions and the autopersistence graphs of binary autoregressive time series. *Journal of Time Series Analysis*, 32(6):639–646, 2011.

[13] Z. Wang, M. A. Nascimento, and M. H. MacGregor. Towards end-to-end routing for periodic mobile objects. In *Proceedings of DIVANet*, pages 61–68, 2011.

[14] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, (393):440–442, 1998.