Trajectory Data Analysis using Complex Networks

Igo Ramalho Brilhante
Federal University of Ceará
Department of Computing
Fortaleza, Brazil
igobrilhante@lia.ufc.br

Jose Antonio Fernandes de Macedo
Federal University of Ceará
Department of Computing
Fortaleza, Brazil
jose.macedo@lia.ufc.br

Chiara Renso
KDDLab
National Research Council - CNR
Pisa, Italy
chiara.renso@isti.cnr.it

Marco Antonio Casanova
PUC-Rio
Department of Informatics
Rio de Janeiro, Brazil
casanova@inf.puc-rio.br

ABSTRACT

A massive amount of data on moving object trajectories is available today. However, it is still a major challenge to process such information in order to explain moving object interactions, which could help in revealing non-trivial behavioral patterns. To that end, we consider a complex networks-based representation of trajectory data. Frequent encounters among moving objects (trajectory encounters) are used to create the network edges whereas nodes represent trajectories. A real trajectory dataset of vehicles moving within the City of Milan allows us to study the structure of vehicle interactions and validate our method. We create seven networks and compute the clustering coefficient, and the average shortest path length comparing them with those of the Erdős-Rényi model. Our analysis shows that all computed trajectory networks have the small world effect and the scale-free feature similar to the internet and biological networks. Finally, we discuss how these results could be interpreted in the light of the traffic application domain.

Categories and Subject Descriptors

E.1 [**Data**]: Data Structures—*Graphs and networks*

General Terms

Algorithms

Keywords

Complex Network, Mobility, Trajectory

1. INTRODUCTION

The explosion of personal positioning devices, like GPS-enabled smartphones or vehicles tracking systems, enabled the collection and storing of a huge amount of tracking data. People wearing these devices leave traces of their movements in the form of sequences of spatio-temporal positions, called trajectories. Although the management of trajectory data dates back to the 1990s, when the first proposals for moving object databases came out, the challenging approaches towards the analysis and understanding of the movement complexity represented in the users tracks is being faced only recently [18]. Even more challenging is the aspect of moving object interaction. How and how much do these moving objects interact? How do the *encounters* among moving entities globally characterize the movement of a moving community? Is there a specific law explaining the interactions of moving individuals? Is the movement of people in vehicles (e.g. cars in a road network) differs from people free movement and/or multi transportation trajectories? How do the individual movements of independent entities influence a crowd's movement pattern?

Inspired by these questions, current work poses a first step in experimenting the complex network analysis techniques applied to a trajectory dataset. The main aim here is to formalize interactions between moving objects as edges in a graph and study the behavior of this graph in terms of complex networks. The approach presented in this paper can be placed between the discipline of mobility data analysis and complex networks, thus exploiting complex network properties to understand mobility of users. Also the challenging and innovative aspect of this experiment is that the network we computed is based on moving objects interactions, which is different from classical complex networks experiments, which focuses on objects that are "static", from the point of view of the spatial position.

The contributions of the present paper are twofold. The first contribution is a method for devising a complex network from a trajectory dataset, hereinafter called trajectory network. The aim of this method is to define specific steps for processing trajectory data in order to build and analyze the trajectory network. The second contribution is an algorithm for building a trajectory network given a trajectory dataset (set of spatio-temporal points). Indeed, this is the first work on analyzing trajectory interactions through complex network technique.

The proposed method has been evaluated using a real GPS dataset from vehicles moving in the City of Milan. All gen-

IDEAS11 2011, September 21-23, Lisbon [Portugal]
Editors: Bernardino, Cruz, Desai

erated trajectory networks from this dataset presented the small world effect and the scale-free feature similar to the internet and biological networks. However the interpretation of these features is an open issue, therefore we will discuss possible interpretations and exploitations of them.

This paper is structured as follows. Section 2 introduces some basic definitions and related work. Section 3 presents the methods and algorithms used to build the trajectory complex network, whereas Section 4 reports experimental results carried on a complex network of vehicle trajectories. Section 5 draws conclusions and future work.

2. BASIC CONCEPTS AND RELATED WORK

2.1 Basic Definitions

A *trajectory* can be defined as the spatio-temporal evolution of a moving object [15]. This evolution is typically represented as a sequence of sample points, representing the spatio-temporal positions detected by a tracking device, such as GPS tools or WI-FI sensors. More formally, a *trajectory* T of an object O is represented as: $\mathrm{T}_O = < x_1, y_1, t_1, \ldots, x_n, y_n, t_n >$ where n is the number of sample points recorded during the movement of the object O, x_i, y_i represent the spatial coordinates of the sample point and t_i the timestamp.

A *complex network* is a network with thousands or millions of nodes whose structure is irregular, with non-trivial topology features [3]. The following features typically characterize complex networks:

- Clustering coefficient: represents the density of triangles in the network. Sparse random graphs have smaller clustering coefficients, while real-world networks typically have larger coefficients;
- Average shortest path length: is the average node-to-node distance. Random graphs exhibit a small average shortest path length as well as real-world networks;
- Power law distribution: is a distribution that follows a power law function, $p(x) = a * x^{-\alpha}$, such that $p(x)$ is the probability of occurrence of x, a is a constant of proportionality and α is the power law exponent.

Complex networks can be characterized by the so called "small world" property when the average number of edges between any two vertices is very small and the clustering coefficient is large [19]. Intuitively, this represents a short path between two edges. This is also known as the "six degrees of separation". Scale free networks are characterized by a degree distribution that follows a power law function. Intuitively, few nodes have many edges (the "hubs" or preferential attachment [1]), many nodes have few edges.

2.2 Mobility Analysis

With the increasing availabilities of trajectory datasets collected from GSM or GPS equipped devices we have the possibility of studying people behavior from their movement traces. Several application areas would benefit from an extensive study on people trajectories such as traffic management, public transportation, commercial advertising, security and police, hazard evacuation management, location based services and so on.

The task of analyzing large trajectory datasets can be carried out in four different directions. First, basic statistics may be applied to trajectory data mainly to discover the distributions of people presence and origin-destination matrices [4]; other studies focus on trajectory data mining, that is, on the application of data mining techniques to trajectory data [8]; other researches focus on representing and querying moving objects in database systems [13, 11]; finally, research originally coming from Physics studies mathematical models, such as complex networks, representing the general laws that describe human movement [9, 18].

Trajectory mining aims at finding correlations in large datasets of trajectory data, collected by personal positioning devices. Techniques include: (1) clustering discovery - finding groups of objects moving together; (2) sequential pattern discovery - finding the most frequent sequences of places visited; (3) flock detection - extracting the convergence of people moving together for a certain amount of time [6, 8].

Several works have investigated how to model and query movement data efficiently, in database literature a new class of databaase systems were created, called moving object database [13, 11]. However, these works were not focused in modeling or querying trajectory data as a first class object. In addition, they did not aim at exploring moving objects interactions.

Trying to model the basic laws governing the human motion is the aim of a broad research area coming from Physics [9]. Their objective is to study the physical laws representing human movements. Social interactions is also in the scope of this research area. A typical example is the study of the spreading of cell phone viruses thru GSM phone calls [18].

2.3 Complex Network Analysis

As already presented in the previous section, a network is a set of items, called vertices (or nodes), with connections among them, called edges (or links). The study of networks (in the form of mathematical graph theory) is one the fundamental pillars of discrete mathematics. Networks have also been extensively studied in different domains, such as Social sciences, Physics, etc. However, recent years have witnessed a substantial new movement in network research, focusing on developing methods and techniques to gather and analyze networks far larger than previously possible. Indeed, this new motivation is due to the inability of humans to draw a meaningful picture of a million vertices by direct eye analysis.

Network has been used as a mechanism of analyses of a huge amount of data with a set of objects which have a relationship or a interaction between them. For instance, the studies in psychology where a node represents a person and an edge represents friendship or that they work together or simply that they know each other or even they have sexual relationship; the studies in biology where the focus is on the species in an ecosystem and a interaction between them, that is, an edge (directed) from species A to species B indicates that A preys B [14]. Therefore, networks offer a perspective of analyses basing on the relationships or interactions.

With respect to mobility analysis, [10] presents a graph-based approach to represent the trajectories by using representative points, a new set of points based on the original one, to generate a graph and find clusters of trajectories. However, they do not consider the properties of the complex network area such as clustering coefficient. On other hand, [12] analyses global cargo ship movements by building a complex network whose nodes represent the ports and links represent the ship traffic between two ports. Differently from our approach, they do not represent the nodes as trajectories and, besides, the points of the trajectories are not taken into account, but only the ports that the cargo ships passed.

Figure 1 shows a trajectory network used in our experiments (this complex network was plotted through the use of the tool Gephi [2]), where each pixel represents one node and red points in the center indicate clustered nodes (nodes with a high degree). Due to the size of this network (36,824 nodes and 306,572 edges) it is hard to apply a visual analysis, even though it may offer a view of the structure of the network identifying a very connected core, nodes with high degrees, surrounded by nodes with low degrees. This topological property is known as ”jellyïňĄsh” or ”octopus” model [16, 5]. Thus, one solution to this problem was the application of statistical methods to describe relevant features of the structures of the network.

3. COMPLEX NETWORK AND TRAJECTORY

This section presents one approach about how to create a complex network from trajectory data. This approach constructs a simple graph where each node represents a trajectory and each edge represents a relationship among the nodes. A relationship between two nodes is established when there is an encounter between two trajectories in space and time with a minimum frequency of meetings.

In this approach, a set S of trajectories is represented as a network (N,E) with the help of a similarity function f between trajectories in S and a threshold constant c. This network is called ***trajectory network***. The ***trajectory network*** (N,E) is constructed as follows: (1) each node in N represents a trajectory in S; (2) there is an edge between two nodes n and m iff $f(m,n) \geq c$, that is, m and n represent trajectories whose similarity is above the given threshold.

In what follows, we will not distinguish between trajectories in S and nodes in N.

Let f be the similarity function for trajectories used to construct the trajectory network, and let s, t and k respectively be the spatial, temporal and frequency parameters of f. Given a trajectory T, the spatial and temporal parameters induce a buffer $B[s,t](T)$ around T. Given two trajectories T and U, we then define:

Definition 1. meet (or collide): T and U meet iff $B[s,t](T)$ and $B[s,t](U)$ overlap.

Definition 2. encounter: there is an encounter between T and U iff T and U meet more than k times.

Note that, as defined above, two trajectories meet independently of their direction. In addition, our definition of encounter does not constrain that two moving objects must be present in the same place at the same time instant.

Our methodology for analyzing trajectory data using this network approach encompasses 3 steps:

1. Build the trajectory network;
2. Analyze trajectory network features;
3. Identify relevant trajectories within the trajectory network.

In the rest of this section, we analyse networks constructed using the specified similarity function.

3.0.1 Step 1 - Build Trajectory Network.

Intuitively, we define that two trajectories are similar iff they are within a certain distance of each other (spatial threshold) within a given time interval (temporal threshold) for a certain number of times (frequency parameter). The values of these parameters obviously depend on the application domain under study. For example, in the traffic management domain, we can establish that two vehicles meet with a minimum frequency of meetings (as few as 2-5 times, depending on the density of the dataset). This representation describes the interactions (given a spatio-temporal threshold and a minimum frequency) between vehicles regardless of the direction of their trajectories. But in other applications, where the study of the flows of vehicles is important, the direction should also be taken into account.

Algorithm 1 Compare two trajectories by their positions

Input: Trajectories $T1$ and $T2$, Temporal Distance Td, Spatial Distance Sd
Output: $frequency$
1: $frequency \leftarrow 0$
2: **for all** position p in $T1$ **do**
3:   **for all** position q in $T2$ **do**
4:     $spatial \leftarrow spatialDifference(p, q)$
5:     $temporal \leftarrow temporalDifference(p, q)$
6:     **if** $spatial \leq Sd$ and $temporal \leq Td$ **then**
7:       $frequency \leftarrow frequency + 1$
8:     **end if**
9:   **end for**
10: **end for**
11: **return** $frequency$

The main idea of the algorithm for building trajectory network (see Algorithm 2) is to compare each position (time, latitude, longitude) of a trajectory to all other trajectories (Algorithm 2 - line 6 to 13). The comparison between two trajectories (see Algorithm 1) compares all points of these two trajectories (Algorithm 1 - line 2 to 10). Each comparison takes into account geographic position - latitude and longitude - and timestamp and the temporal and spatial distance thresholds (Algorithm 1 - line 4 and 5), respectively Td and Sd variables. When the comparison satisfies the thresholds Td and Sd (Algorithm 1 - line 6), then the frequency variable is increased by one unit (Algorithm 1 - line 7). After all comparisons (Algorithm 2 - line 9), we have computed the frequency of encounters of a trajectory with respect to another trajectory. If the frequency is greater than the value of the input parameter Frequency, denoted by variable f (Algorithm 2 - line 10), then an edge is created between two trajectories (Algorithm 2 - line 11).

Some improvements can be done by using spatial index structures or spatial database systems to take advantage of the index structure and the data management language as well.

3.0.2 Step 2 - Analyze Trajectory Network Features.

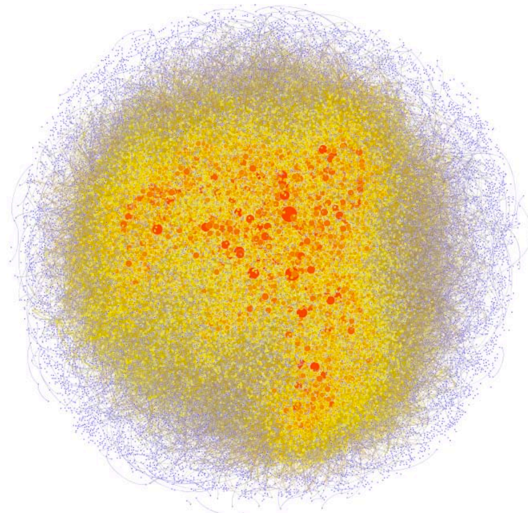


Figure 1: A plotted complex network composed by 36,824 nodes and 306,572 edges generated in our experiments.

Algorithm 2 Trajectory Network Generator

Input: Trajectory List $traj$, Temporal Distance Td, Spatial Distance Sd, Frequency f
Output: Trajectory Network TN

$n \leftarrow |traj|$
create a undirected graph TN
for all trajectory T in $traj$ **do**
  create a node in TN
end for
for $i = 1$ to n **do**
  $frequency \leftarrow 0$
  for $i = i + 1$ to n **do**
    $frequency \leftarrow compare(traj[i], traj[j], Td, Sd)$
    if $frequency \geq f$ **then**
      add edge $(traj[i], traj[j])$ in TN
    end if
  end for
end for
return TN

In our approach we are interested in identifying the existence of two very important network features: the power-law distribution and the small-world effect. The analysis of the distribution degree of network vertices allows identifying whether such distribution is highly skewed, meaning that it has a power-law distribution profile. In this case, we conclude that few trajectories have many encounters, while most of the trajectories have very few encounters. The discovery of such property can be useful, for example, to identify trajectories having a high degree of encounters, which means that this trajectory has passed through paths with a high number of moving objects. Besides, small-world property may help to identify a set of trajectories that represent hubs in the trajectory network.

The small-world effect feature determines the mean shortest path length between pairs of trajectories as well as if the network has a high clustering coefficient. Through this measure we can quantify how well connected the trajectories in the network are. Besides a high clustering coefficient indicates the presence of a transitivity property among high connected nodes. This information can be very useful, for example, when analyzing bus trajectories and their encounters we can verify how buses lines are connected and how easy is it to move through the city using bus lines.

There are several free tools for computing network features, which can be used in this step without requiring devising an algorithm for that. One example of such a tool is the Network Workbench [17] developed by Indiana University, Northeastern University, and University of Michigan. This tool provides several algorithms to calculate the properties of large complex networks.

3.0.3 *Step 3 - Identify relevant trajectories within trajectory network.*

The third step in our approach aims at analyzing trajectories that have greater relevance within the network. The relevant trajectories are those that possess a high degree of connectivity. These trajectories are plotted on a map for visual analysis, allowing the user to give an interpretation of the relevance of these trajectories in the geographic context 4.2. This type of analysis will help reducing the number of trajectories to be analyzed. Furthermore, we can restore back the spatial information, which was lost during the creation of the network. Visualizing trajectories that are very connected, which we hereinafter call hub trajectories, is useful for understanding entities moving in the high dense paths with respect to the amount of moving objects.

4. EXPERIMENTS

4.1 Experiments on the vehicles' movements in Milan city

In this paper, we propose a number of experiments using a mobility network that represents the trajectories of vehicles moving in the City of Milan, Italy (collected by GPS devices installed in the vehicles). We split the dataset into seven different files corresponding to the days of the week. The number of trajectories in each day of the week and their average length is depicted in Table 1. In this dataset, each trajectory corresponds to only one car.

Table 1: Information about trajectory dataset

Day	Number of trajectories	Average number of points per trajectory
Sunday	23535	8.290461
Monday	34812	8.927956
Tuesday	36824	9.206279
Wednesday	36023	9.467285
Thursday	35340	9.871647
Friday	33822	8.697179
Saturday	25576	7.746950

We can notice a high number of trajectories (i.e. vehicles) tracked each day; we also notice a decrease of moving vehicles during the weekend, as expected. The average number of sample points for each trajectory stays within 7 and 10. The low number of points per trajectory is due to a data cleaning process performed on the original data, which eliminated outliers and redundant points.

We start our experiments running Algorithm 2, Trajectory Network Generator, described in Section 3 with the following parameters:

- trajectory dataset: specifies the name of a trajectory dataset to be processed;
- spatio-temporal window for encounter: this parameter defines a temporal and spatial distance that is used to select trajectories that have a meeting in time and space;
- minimum frequency of encounters: this parameter specifies a minimum number of meetings between two trajectories.

We have generated 28 trajectory networks referring to the 7 days of the week with 4 different parameter configurations as presented in Table 2. We chose a spatial distance threshold of 300 meters, and two different temporal intervals of 15 and 30 minutes. These values have been chosen starting from dataset statistics. For example, since the trajectory sample points are quite far from each other, we choose a quite broad notion of "meeting" setting this parameter to some hundreds meters. Obviously, denser sample points will correspond to smaller distance. Analogous is the consideration for the temporal threshold. The values 3 and AVG for minimum amount of encounters was defined taking into account the average number of encounters computed from our data set. The AVG value is related to the average number of points per trajectory in Table 1 for each day. For instance, on Sunday we get 8 as value to this parameter. Table 2 below summarizes the 4 parameters combinations that were used to setup our 4 experiments on the 7 daily trajectories datasets.

Table 2: Four different parameter combinations

Experiment	Minimum Frequency	Spatial Distance (km)	Temporal Distance (min)
1	3	0.3	30
2	3	0.3	15
3	AVG	0.3	30
4	AVG	0.3	15

4.2 Computed Trajectory Network Features

Tables 3, 4, 5, and 6 show the trajectory network features computed from our experiments. The computation of the trajectory network features was accomplished through the use of the Network Workbench Tool [17]. Each table reports the number of nodes n (each node represents a trajectory), the edges m (an edge represents an encounter between two trajectories), the clustering coefficient C (represents the density of triangles in the network), the average shortest path l and the diameter d.

Table 3: Experiment 1 (3 | 0.3 km | 30 minutes)

Day	n	m	C	l	d
Sunday	23535	229938	0.441	5.028	17
Monday	34812	515866	0.445	4.615	19
Tuesday	36824	590821	0.456	4.537	16
Wednesday	36023	587151	0.455	4.498	16
Thursday	35340	587399	0.457	4.393	16
Friday	33822	447244	0.437	4.776	19
Saturday	25576	210073	0.424	5.512	19

Table 4: Experiment 2 (3 | 0.3 km | 15 minutes)

Day	n	m	C	l	d
Sunday	23535	116671	0.418	5.691	22
Monday	34812	264897	0.417	5.196	22
Tuesday	36824	306572	0.432	5.099	19
Wednesday	36023	305338	0.432	5.058	20
Thursday	35340	307179	0.431	4.936	20
Friday	33822	229983	0.413	5.371	21
Saturday	25576	106424	0.398	6.363	22

We can notice that all 28 trajectory networks generated in our experiments are highly clustered networks. To arrive at this conclusion, we generated a Erdős and Rényi (ER) model [7] network with the same size, in nodes, as the networks generated in experiment 1 (Table 3), and with the number of edges proportional in order to have a similar network in size and structure. Table 7 shows the size and the computed features of the ER networks. They are generated randomly and characterized by their low average shortest path length as well as low clustering coefficient. Comparing Tables 3 and 7 , one may verify that the level of clustering of the trajectory networks is much higher than that of the corresponding ER network. This means that we have a set of trajectories that have a high number of encounters. In combination, they present low values for average shortest path length (less than 6), and this has a clear interpretation as small world property [19].

Table 5: Experiment 3 (AVG | 0.3 km | 15 minutes)

Day	n	m	C	l	d
Sunday	23535	27789	0.431	6.574	22
Monday	34812	78204	0.449	6.014	22
Tuesday	36824	82795	0.487	6.052	17
Wednesday	36023	86811	0.488	5.995	20
Thursday	35340	92753	0.490	5.629	24
Friday	33822	63898	0.443	6.168	18
Saturday	25576	28633	0.399	8.065	25

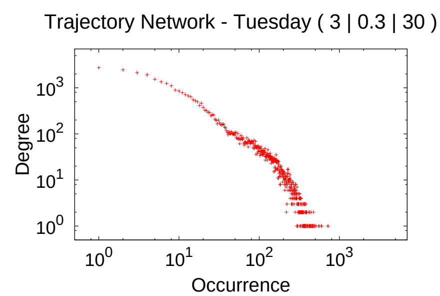
Table 6: Experiment 4 (AVG | 0.3 km | 30 minutes)

Day	n	m	C	l	d
Sunday	23535	58901	0.464	5.906	21
Monday	34812	161632	0.480	5.328	18
Tuesday	36824	168823	0.511	5.319	20
Wednesday	36023	175310	0.510	5.267	17
Thursday	35340	185919	0.513	5.021	18
Friday	33822	132142	0.471	5.486	21
Saturday	25576	60438	0.431	6.614	23

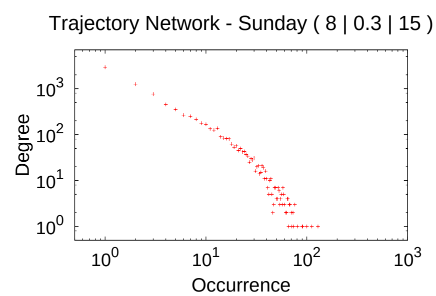
Also notice the presence of many trajectories that have few encounters and few trajectories that have a large number of encounters. This characterizes a phenomenon known as "the rich get richer". This analysis is done by plotting a curve that shows the correlation between the amount of nodes versus their respective degrees. Two graphs were created (Figures 2(a) and 2(b) showing the distribution of trajectories of Tuesday and Sunday, corresponding to the largest and the smallest networks generated by our experiments. In both graphs we have a power law curve. By contrast, Figure 2(c) illustrates the degree distribution of the corresponding ER networks. Note that the node degrees of the ER networks do not follow a power law distribution, whereas the node degrees of the trajectory networks do. In fact, the trajectory networks have nodes that are hubs [1]. This is a important result since all trajectory networks in our experiment presents a small world and power law feature similar to the internet and biological networks.
Hereinafter, we analyze each trajectory network feature separately, namely clustering coefficient, average path length, and diameter. Figure 4(a) shows the clustering coefficient feature computed for all 28 trajectory networks. In this Figure, we can observe that the clustering coefficient increases on Tuesday, Wednesday and Thursday. This is consistent with the dataset statistics (see Figure 3) where we notice an increase of trajectories moving on Tuesday, Wednesday and Thursday compared to Friday (for that particular week the friday was just before Easter holidays) and the weekend. Intuitively, this reflects the fact that we have more vehicle's

Table 7: Random Graph - Erdős and Rényi [7]

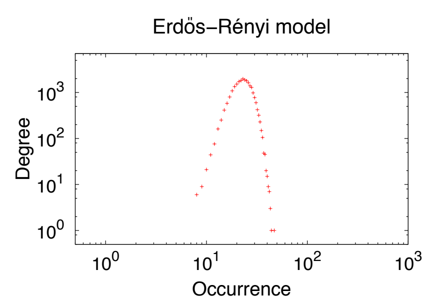
Day	n	m	C	l	d
Sunday	23535	278105	0.0010	3.539	5
Monday	34812	606707	0.0010	3.262	4
Tuesday	36824	679044	0.0009	3.221	4
Wednesday	36023	649475	0.0010	3.237	4
Thursday	35340	625219	0.0009	3.251	4
Friday	33822	572781	0.0009	3.283	4
Saturday	25576	328251	0.0009	3.483	5



(a) Degree distribution of Table 3 at row 3



(b) Degree distribution of Table 5 at row 1



(c) Degree distribution of Table 7 at row 1

Figure 2: Degree distribution

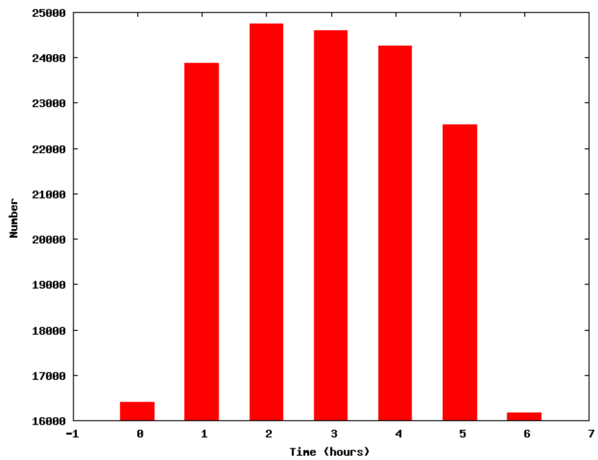


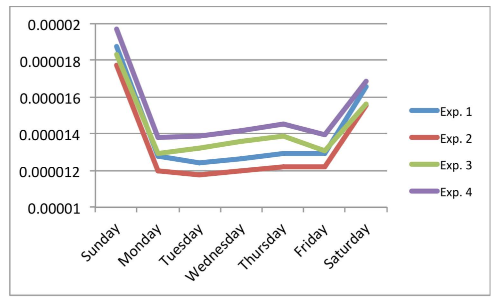
Figure 3: Plot showing the number of trajectories for each day of the week

encounters during these central days, that reduce when approaching to Friday and the weekend. From these trajectory networks, we deduce that Tuesday, Wednesday and Thursday are the days with more encounters than the other days of the week and this is consistent with dataset statistics. Figure 4(b) shows the average shortest path length feature for all 28 trajectory networks. We observe that in all trajectory networks the average path length increases during the weekend and reduces during the working days. Particularly, Thursday is the day with the lowest average shortest path length in all trajectory networks. Since the shortest path length characterizes the distance between two nodes, then when we have more encounters among vehicles we decrease the average path length. Thursday is the day with the largest number of encounters, which increases the probability of having traffic congestion. This result is coherent with the clustering coefficient feature graph (Figure 4(a)), where the highest clustering coefficient is also associated with Thursday.

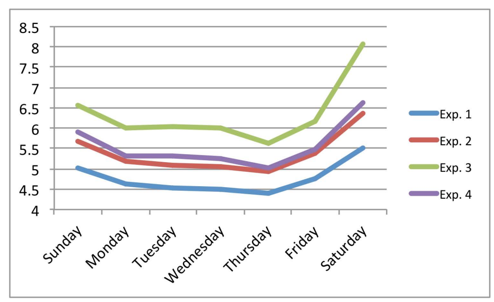
The diameter feature (the largest distance between two nodes) is illustrated in Figure 4(c). In this graph we observe that there is no explicit correlation among diameter features from all the trajectory networks. This behavior is justified by the fact that the diameter feature is very sensitive to trajectory network topology. Besides, given the size of the trajectory networks, which are around 5 orders of magnitude, the variation in diameter along each day of the week (showed in Figure 4(c)) is insignificant.

Figures 5(a), 5(b), 5(c), 5(d), 5(e), 5(f) and 5(g) show the hub trajectories for each day of the week. These trajectories are projected on GoogleMaps. Particularly, we note that a part of the road segment, the A51 highway, is presented in five of seven high connected trajectories (indicated in the figure by an arrow). This fact suggests that this road segment has a high concentration of cars, possibly because this is a ring highway that goes towards the Milano Linate Airport. In Figure 5(b) (Monday) we observe that the most connected trajectory represents a moving object that goes back and forth between the airport and Milan's downtown, probably a shuttle service, which corroborates to increase the amount of encounters to this trajectory.

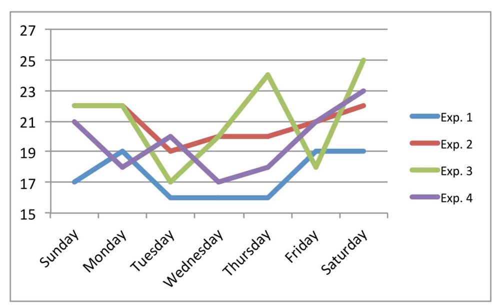
We can also remark that on Monday, Tuesday, Thursday,



(a) Clustering Coefficient



(b) Average Shortest Path Length



(c) Diameter

Figure 4: Network properties of tables 3, 4, 5 and 6

Friday and Saturday (respectively Figures 5(b), 5(c), 5(e), 5(f) and 5(g)) the corresponding hub trajectory has some movement within the airport. This observation reveals that the airport is a hot spot (with respect to encounters of moving entities) in Milan's city. Beyond this, we can observe that ring highways appear on six of the seven hub trajectories, which reinforces that this is a road that is very much used by vehicles. Indeed, several analyses can be done by using hub trajectories, such as comparing several hub trajectories of the same trajectory network, or analyze relevant parts of hub trajectories. Due to time and space limitations we left this investigation for future work.

5. CONCLUSION

In this paper we have presented a theoretical analysis and experimental results for moving object trajectories on complex networks. Our motivation comes from the desire to understand the influence of moving objects trajectories interaction on the traffic dynamics. We have defined a method for devising trajectory network from a dataset of moving object trajectories. In addition, we have built 28 trajectory networks from a real dataset of trajectories of vehicles. We have computed three network features (i.e. clustering coefficient, average shortest path length and diameter), for each trajectory networks and compare them.

Our analysis reveals that all trajectory networks are scale free network, presenting small world and power law features. Our results have practical implications for investigating moving objects interactions from complex network perspective. Although we have provided basic methods for building trajectory networks and analyzing their features, future investigation is needed in order to define how to interpret such features taking into account the application domain knowledge.

Comparing to existing data mining and statistical methods, our proposed approach provides another method for analyzing trajectories from the potential *interaction* perspective. Besides, complex network technique is adequate to analyze relationships among a large set of entities by computing topological features of the graph. Although building a trajectory network is time consuming, computing its properties is not. Thus, we believe that this technique can open new opportunities in mining the network structure of interactions between a large number of moving object trajectories.

The future research focus is on further analyzing the interactions between trajectories and space (i.e. landmarks, point of interest,etc), or between trajectories and time (i.e. hush hours, week-end, etc), or between trajectories and events (i.e. soccer match, festival, etc), to name a few.

6. ACKNOWLEDGEMENTS

This work was partly supported by CNPq, under grants 473110/2008-3, 557128/2009-9, 483552/2009-7 and 308247/2009-4, by FAPERJ under grant E-26/170028/2008, by FUNCAP under grant GPF 2151/22, CAPES under grant NF 21/2009. This work was also supported by CAPES through the fellowship granted to the master student Igo Ramalho Brilhante.

7. REFERENCES

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286, 1999.

[2] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[3] S. Boccaletti, V. Latora, Y.Moreno, M. Chaves, and D.-U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, (424):175–308, 2006.

[4] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti. Real-time urban monitoring using cellular phones: a case-study in rome. *IEEE Transactions on Intelligent Transportation Systems*, 2010.

[5] F. Chung and L. Lu. *Complex Graphs and Networks*, volume 107 of *CBMS*. American Mathematical Society, 2006.

[6] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7:240–252, June 2008.

[7] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[8] F. Giannotti and D. Pedreschi, editors. *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery.* Springer, 2008.

[9] M. González, C. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, (453):479–482, 2008.

[10] D. Guo, S. Liu, and H. Jin. A graph-based approach to vehicle trajectory analysis. *J. Locat. Based Serv.*, 4:183–199, September 2010.

[11] R. H. Güting, M. H. Böhlen, M. Erwig, C. S. Jensen, N. a. Lorentzos, M. Schneider, and M. Vazirgiannis. A foundation for representing and querying moving objects. *ACM Transactions on Database Systems*, 25(1):1–42, Mar. 2000.

[12] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society, Interface / the Royal Society*, 7(48):1093–103, July 2010.

[13] L. V. Nguyen-Dinh, W. G. Aref, and M. F. Mokbel. SECONDO: A Platform for Moving Objects Database Research and for Publishing and Integrating Research Implementations. *Data Engineering*, 33(2):46–55, 2010.

[14] S. L. Pimm. *Food Webs.* University of Chicago Press, second edition, 2002.

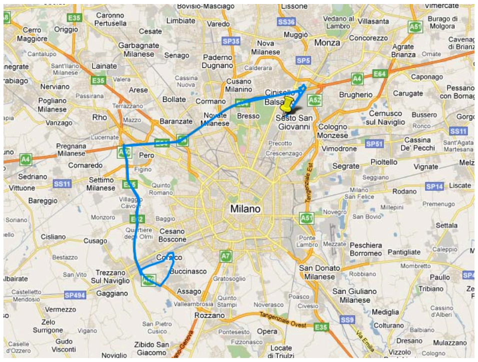
[15] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. F. de Macêdo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *Data Knowl. Eng.*, 65(1):126–146, 2008.

[16] S. L. Tauro and C. Palmer. A Simple Conceptual Model for the Internet Topology. *Power*, 1995.

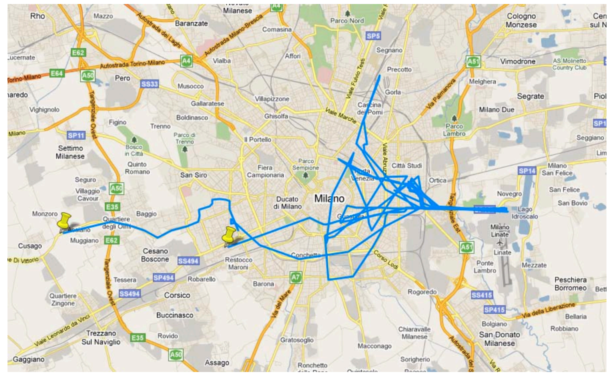
[17] N. Team. Network workbench tool. (2006). Indiana University, Northeastern University, and University of Michigan. http://nwb.slis.indiana.edu.

[18] P. Wang, M. González, C. Hidalgo, and A.-L.Barabási. Understanding the spreading patterns of mobile phones viruses. *Science*, (324):1071–1076, 2009.

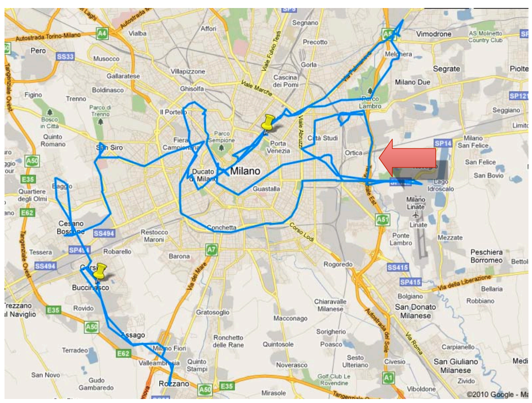
[19] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, (393), 1998.



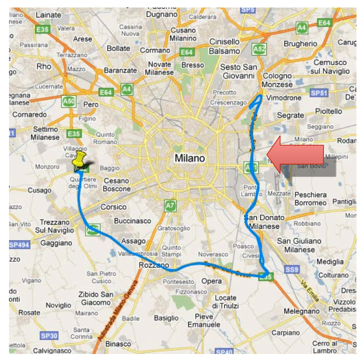
(a) Sunday



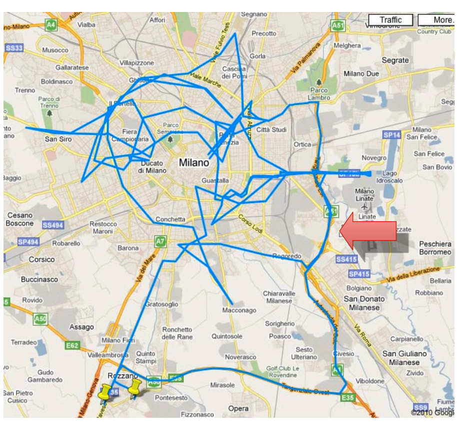
(b) Monday



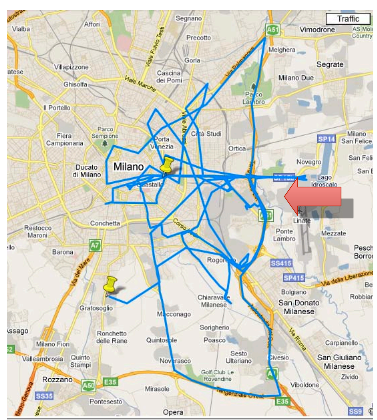
(c) Tuesday



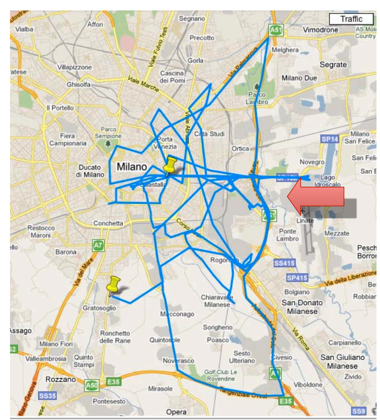
(d) Wednesday



(e) Thursday



(f) Friday



(g) Saturday

Figure 5: Trajectory plot of the days of the week