Vessel's Identity Graph and Analysis

Alexandre Amouriq, Maryam Maslek Elayam, and Cyril Ray

Naval Academy Research Institute, Brest, France
{alexandre.amouriq,maryam.maslek_elayam,cyril.ray}@ecole-navale.fr

Abstract. Maritime mobility data is an inexhaustible source of information that can be used to enable the navigators and surveillance stations to have a comprehensive understanding of the maritime situation in real-time. Adequate models and analysis of historical maritime data can also favour advanced knowledge extraction for a deeper understanding of trajectories, routes and worldwide maritime networks along space and time. In this paper, we are interested in the modelling of maritime patterns of life. Having a focus on the life cycle of the ship's identity, we propose a graph model and its implementation on the Neo4j graph database. Several Cypher queries are applied to illustrate the potential of the model.

Keywords: Patterns of life · graph-based identity model · maritime mobility data.

1 Introduction

The surveillance of global maritime traffic is a major stake in our modern societies. The benefits are numerous and include, for example, the fight against trafficking and fraud, navigation safety, protection of the sea and its resources (particularly fishing) or the understanding of international trade.

Nowadays, several location systems exist such as the Automatic Identification System (AIS). This collaborative international system allows a number of ships (estimated at around 380,000) to communicate their position and identity by automatic radio communication. The volume of this data grows steadily and it results in difficulties of summarising and analysing information by a human.

Analysing AIS data, in real-time, and also by considering historical data over several years allows obtaining a wealth of information on the characteristics and habits of ships in order to understand their behaviours: ports they visit, maritime routes they use, their identity cycle, ships they meet, etc.

The identity of a ship is often variable. It may change throughout its entry into service depending on its owner, charterer, technical modifications. This information transmitted by the ship's transponder is affected by incorrect parameterisation and voluntary misappropriation of its identity or behaviour. In this work, the design of a vessel's identity representation aims to define a model of identity and activity cycle of a vessel. Its analysis allows to know the life of ships, their activity cycles but also the detection of abnormal and regular behaviours, threats to maritime navigation and identification of non-compliance with international rules or conventions (e.g. embargo).

Vessel's identity model and the related analysis are based on the processing and studying of historical AIS data. The addition of complementary data sources (geographic data, spatial index, list of ships, etc.) has been also addressed in order to correlate AIS data and thus make them more relevant and easily understandable. The proposed model is based on an annotated graph structure and is implemented in a graph database (Neo4j). Various analyses are carried out either by Cypher queries or by an algorithm implemented in Python.

The following of the paper is organised as follows: the next section proposes some related works and introduces the notion of patterns of life. Section 3 defines main components of the proposed model. Section 4 presents the real dataset and the graph-based database model implemented in Neo4J, as well as some query results. Finally, Section 5 concludes the paper.

2 Related works

The analysis and the processing of data from automated surveillance technologies like the Automatic Identification System (AIS), Synthetic Aperture Radar (SAR), satellite images and coastal radars is an emerging topic. Once integrated with contextual information [12] (coastlines, nautical charts, fishing areas, maritime protected areas, sea state and weather conditions, ...) positioning data can be analysed in order to understand and predict vessel's mobility [4, 14, 16], monitor and forecast maritime traffic [8, 7, 9], uncover activities or risks for the environment [11], living resources and navigation and so on. The research is expeditiously progressing, and many analysis techniques are applied to attain the above major goals using maritime navigational data.

Most of the prior works mainly concentrate on semi instant positioning data of ships, mostly stay limited to short term predictions and can be vulnerable to noisy data. The concept of Pattern of Life (PoL), already considered in urban domain [6], aims to understand human mobility patterns based on sequences of places (instead of studying raw geographic coordinates) which encode, at a coarse resolution, most daily activities. This is in general done addressing the problem of automatic place labelling from mobility data. Such an approach specifically used for understanding a subject's habits also suits to the maritime domain.

Maritime Pattern of Life (MPoL) aims at understanding and monitoring the behavioural patterns of a vessel in order to analyse its navigational habits and it is used for potentially predicting its future actions. MPoL addresses the understanding of navigation data at an aggregated level using a graph model [10] and provides a better representation of ship routes, more relevant for understanding trajectories, activity patterns, and relationships between them at different spatio-temporal scales.

3 Vessel's Identity Model

Maritime traffic data represents a large volume of worldwide data. Data issued from embedded location systems, like AIS, provides monitoring of the navigation and helps operational decision making. However, it is annoying to analyse positions and contextual data for each vessel at every point. Then an abstraction is made, in space and time, in order to capture the main behaviours of vessel and maritime networks. Abstraction on historical AIS data allows a knowledge extraction that helps to understand and analysing ship's identity, its habits, its states, its functionalities and its specificities... This section presents a formal view of the vessel's identity and a graph model that makes it possible to represent mobile object identity and in space and time.

A ship is a mobile object that crosses the maritime environment (space) at different periods of time, under specific status while performing several activities (fishing, transporting goods, transporting people, transporting crude oil and gas,...). This determines its identity and then its patterns of life. Therefore we can define a pattern of life such as:

Definition 1. *(Pattern of Life) The pattern of life of a ship is defined as the sequence of navigational statuses the ship performs, ordered by time. Let N be the number of statuses of a ship defined by its identity Id. Its pattern of life is defined as follows:*

$$PatternOfLife(Id) = status_1, status_2, status_3, status_4, ..., status_N \tag{1}$$

Statuses (e.g., moored, sailing, engaged in fishing, ...) are provided by position report systems like the AIS. They change depending on space and time. Such statuses help to express the variety of activities a ship can perform. In addition, each status is carried out within a precise and particular environment defined by a location, a speed, a draft or even certain weather conditions. Vessel's patterns of life are used to define vessel activities such as being in port, moving, fishing ... All these components bring semantic information to the Vessel's Identity Model as they add specifications to a ship's identity, to the actions a ship can achieve and to its navigation history.

Statuses that made a pattern of life can be assorted together to form activities. A set of activities made up a pattern cycle. This can be used for the activity analysis at sea for each vessel but also for the detection of anomalous activities.

Definition 2. *(Pattern Cycle) The pattern cycle of a ship is the sequence of activities that a ship performs, ordered by time. Let M be the number of activities that a ship Id achieves. Each activity i is defined by a set of statuses ordered by time, and as the combination is specific to the activity i. A pattern cycle is defined as:*

$$\begin{aligned} PatternCycle(Id) = activity_1(status_{(1,1)}, status_{(1,2)}, ..., status(1,k)), \\ activity_2(status_{(2,1)}, status_{(2,2)}, ..., status_{(2,l)}), ..., \\ activity_M(status_{(M,1)}, status_{(M,2)}, ..., status_{(M,m)}) \end{aligned} \tag{2}$$

while $task_{(i,j)} \subset C(PatternOfLife)$

Graph model makes it possible to represent interconnected data as a set of nodes and edges that are relationships between nodes. In order to represent the diversity and complexity of the vessel's identity, the graph model includes several types of nodes that are annotated with characteristics of ship's statuses, activities, credentials, main areas of trajectories, ports... The aim is to integrate various data types to analyse vessel's identities and behaviours. The relationships between these nodes represent the time period during which the vessel performs some behaviour, identity...

Definition 3. *(Identity Graph Model) The identity graph model IG is the sequence of edges and nodes such as:*

$$IG = \{v_1, e_1, v_2, e_2, ..., e_{N-1}, v_N\} \tag{3}$$

while v_i is the node i annotated by the activity i (or status i) and e_i is the edge linking the nodes i and $i+1$ and annotated by the time interval $[t_{ij}, t_{ik}]$ such as t_{ij} and t_{ik} correspond, respectively, to the starting and ending times when the mobile object had the identity i or performs the activity i.

Figure 1 presents the schema of the graph model as different types of nodes and edges. The idea behind the graph structure is such that a *Ship* (blue node), attributed by its MMSI (supposed to be unique), is from a home *Country* (yellow node), had a *Ship_ID*, a *State* that is the status (orange node), assigned by a *Licence*, has dimensions that are *Ship_CONF* (red node) and achieves a *Travel* (green node) to a specific destination. Each AIS signal has a signature characterised by temporal features (e.g. time before modulation of the signal) integrated into the model (purple node). This may confirm the ship's identity [1].

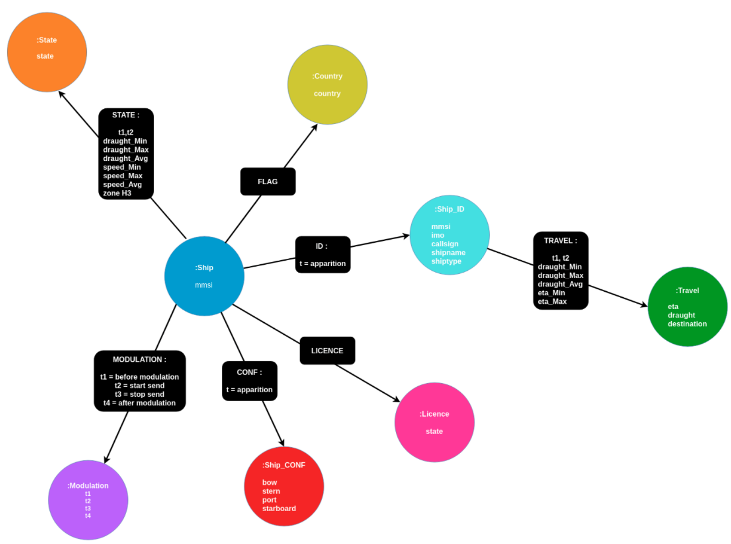

Fig. 1. Vessel's Identity Graph schema

To represent vessel's locations as sequence of statuses ordered by time, we used the Uber H3 spatial indexing [5]. This type of discretization uses hexagons with such resolutions as grid cell to divide the space. Hexagons have only one equal distance between the center of an hexagon and that of its six neighbours, versus two different distances for squares or three distances for triangles. This property simplifies the analysis and smoothing of gradients [3]. Uber H3 supports 16 hexagons resolutions. Each resolution has hexagonal cells with one-seventh of the area of the coarsest resolution. Hexagons cannot be perfectly subdivided into seven smaller hexagons, so the thinnest cells are only approximately contained in a parent cell, while the hexagons boundaries are correct for indexing locations at a specific resolution. As a summary all ship's trajectories are abstracted in the model by a sequence of hexagons (Uber H3 resolution 9) representing sequences of navigational statuses.

4 Implementation and results

Vessel's identity graph construction is based on AIS historical data. A preprocessing is necessary because AIS messages are always incomplete and contain erroneous samples. Preprocessing steps are performed by *Python* and the identity graph is implemented in the *Neo4j* graph database and explored by *Cypher* query language.

4.1 Data integration

The dataset used in this research relies on the representative heterogeneous maritime dataset produced under the aegis of the *H2020 datAcron* project [13]. It includes historical traces of maritime vessels at a regional scale with one major route crossing the Ushant traffic separation scheme. Data have been collected through the Automatic Identification System (AIS) and aggregated with correlated data spatially and temporally aligned[1]. Data are categorised into two types: static and dynamic data. The static part represents the various information entered by a human (e.g. ship origin, ship type...) while the dynamic part is the gathering of navigation messages (e.g. positions). It covers a time period of six months from October 1^{st} 2015 to March 31^{st} 2016 and provides around 18.6 million observations of ship positions collected from 4842 ships over the Celtic Sea, the North Atlantic ocean, the English channel, and the Bay of Biscay.

Locations of ports and stationary areas are essential for the analysis of patterns of life. We used several data sources referencing all the ports of the world and computed the spatial extent of these locations and surrounding stationary areas through the analysis of stationary AIS data [2]. Then by carrying out data processing and the use of Uber H3 indexing, these areas are associated with a combination of lists of H3 hexagons (with Uber H3 format at resolution 8 and 9). The figure 2 presents ports of Brest Bay. The database is composed of 25,924 port areas described by approximately 280,000 hexagons.

[1] Link to the dataset: https://zenodo.org/record/1167595

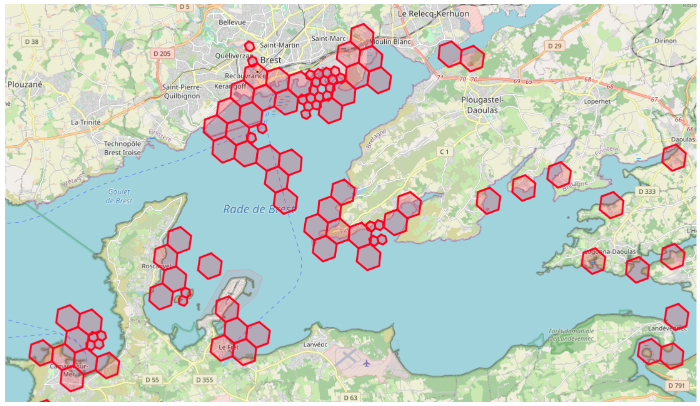

Fig. 2. Spatial representation, based on Uber H3, of Brest ports

4.2 Vessel's identity graph database

Graph model helps to manage semantic information across space and time. Nodes and edges are defined to structure all the data and in a relevant way in order to obtain useful information and analytic with simple queries. In addition, it allows a better representation of links and interactions between nodes. Moreover, this is a useful approach to avoid the use of complex queries (such as SQL queries with joins) when managing a large volume of interconnected data.

The vessel's identity model (figure 1) has been implemented in a graph database (Neo4j). The model has been extended with new nodes representing several additional data sources; ships registers, geographic areas (e.g. fishing areas computed by [15]), and off course ports and stationary areas described in the previous section. The database includes 48,603 nodes and 62,492 edges for 5,055 ship (MMSI, where the MMSI is the international ship identifier, expected to be unique) and 6,123 ship identities.

4.3 Illustration of vessel's identity analysis

Querying and analysing the identity graph aims to extract useful information regarding the life cycle of ships (visited ports, encountered ships, mobility patterns, ..). In this section, the database is queried in order to answer such questions which make sense in the context of maritime surveillance and tied to the identity of ships. Cypher is Neo4j's graph query language[2]. It provides numerous operations and algorithms on functional graphs. In the following two examples are provided to illustrate the use and results of the model.

The first example has as objective to extract the identity graph (or a subset). It is well known that sometimes ships exhibit different parameters (name, length, type) along time but also at the same time. In the latter case, it mainly concerns ships using a false identity. Amongst these ships, military vessels tend to use false

[2] https://neo4j.com/developer/cypher

MMSI like 999999999 (or equivalent). The following query looks for such military vessels the 20^{th} of October (specifically around 5 pm as we know a joint use of the same identity occurred).

Query for question "Which military ship appeared on 10/20/2015 at 5:53:27 ?"

```
MATCH (shipID:Ship\_ID)<-[id:ID]-(ship:Ship)-
      [conf:CONF]->(confShip:CONF\_Ship)
          WHERE ship.mmsi = "999999999"
          AND toInteger(conf.t)>1445320407+3000
          AND toInteger(conf.t)<1445320407-3000
          AND id.t>1445320407+3000
          AND id.t<1445320407-3000
RETURN ship,confShip,conf,shipID,id0
```

Two ship configurations exist (blue nodes at the figure 3) for the ship of MMSI 999999999 (named FRENCH WARSHIP in orange node) and whose identity appeared since (20/10/2015 at 5:53:27). The two configurations are different as in the figure 3 because the size of the vessel has changed. Thus, the query allows to obtain a fast answer in order to identify ships having several identities and configurations, and to find which ship configuration belongs to which ship identity.

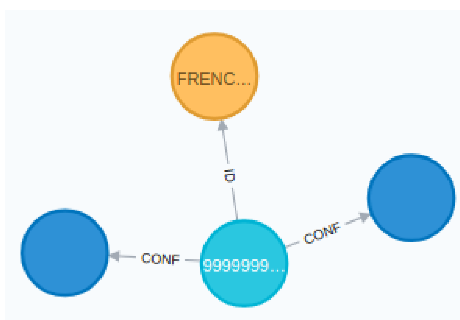

Fig. 3. Resulting identity sub-graph

The next example aims to analyse the navigational properties of a specific vessel. Figure 4 presents the usual destinations and statuses of the ship identified by the MMSI 305600000. The flow diagram shows that the ship makes usual trips between 5 ports that are: Brest, Le Havre, Bassens, Montoir de Bretagne, and La Rochelle. The arcs size and colours in the diagram are proportional to the number of trips that the ship performs between ports. It means that the most of its trips are between: Brest ⇔ Le Havre, Brest ⇔ La Rochelle, Brest ⇔ Bassens, Le Havre ⇔ La Rochelle, Le Havre ⇔ Montoir de Bretagne and Le Havre ⇔ Bassens, while its most destinations are Brest and Le Havre. When it comes to its usual statuses, it is most of the time "En transit" (navigating status) or "amarré" (anchored status).

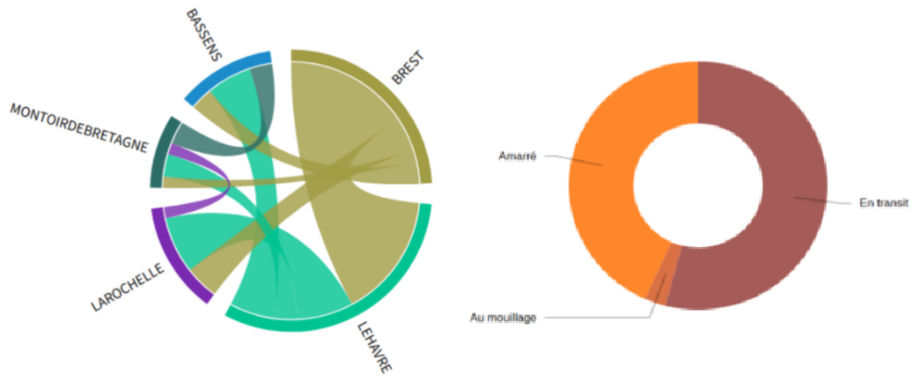

Fig. 4. Resulting chord (left) and donut diagram (right) showing flows between ports and distribution of statuses

Many other queries taking advantage of the graph-based model have been applied. Amongst them, ships life cycle, trends of the different flags, predictions using support vector machine of ship type, status, and next hexagons in the path, rendez-vous and co-occurrence of ships in stationary areas and ports.

5 Conclusion

In this paper, we present the representation of a ship's life cycle as an annotated graph. Such representation of historical mobility data allows considering of ships mobility analysis along different dimensions including various space and time granularity. This favour insight on maritime routes, monitoring of activities at the sea through pattern recognition, anomaly detection, the centrality of ports, and much more. The pattern of life model was implemented on a regional dataset and the graph was successfully constructed in Neo4J. The life cycle model is driven by the evolution of nominative information correlated with navigational statuses along time. Hierarchical hexagonal indexation of ship's trajectories but also space (especially ports and stationary areas) allows to represents trajectories as a sequence of navigational status at different scales.

References

1. Alincourt, E., Ray, C., Ricordel, P.M., Dare-Emzivat, D., Boudraa, A.: Methodology for AIS signature identification through magnitude and temporal characterization. In: Proceedings of the OCEANS 2016 SHANGHAI Conference. Institute of Electrical and Electronics Engineers (IEEE) (apr 2016). https://doi.org/10.1109/oceansap.2016.7485420
2. Arnaud, M., Vincent, L.d.C., Cyril, R., Charles, G., Maryam, M.E., Maxime, D.: Construction d'une méta-base de ports à l'échelle mondiale. In: Conférence internationale de Géomatique et d'Analyse Spatiale. SAGEO'21, La Rochelle (May 2021 2021)

3. Birch, C.P., Oom, S.P., Beecham, J.A.: Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological modelling **206**(3-4), 347–359 (2007)
4. Borkowski, P.: The ship movement trajectory prediction algorithm using navigational data fusion. Sensors **17**(6), 1432 (2017)
5. Brodsky, I.: H3: Uber's hexagonal hierarchical spatial index. Available from Uber Engineering website: https://eng. uber. com/h3/[22 June 2019] (2018)
6. Do, T.M.T., Gatica-Perez, D.: The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. IEEE Transactions on Mobile Computing **13**(3), 638–648 (2014). https://doi.org/10.1109/TMC.2013.19
7. Johansen, T.A., Cristofaro, A., Perez, T.: Ship collision avoidance using scenario-based model predictive control. IFAC-PapersOnLine **49**(23), 14–21 (2016)
8. Jugović, A., Hess, S., Poletan Jugović, T.: Traffic demand forecasting for port services. Promet-Traffic&Transportation **23**(1), 59–69 (2011)
9. Kim, S., Lee, J.: Small infrared target detection by region-adaptive clutter rejection for sea-based infrared search and track. Sensors (Basel) **14**, 13210–13242 (2014). https://doi.org/10.3390/s140713210
10. Maslek Elayam, M., Ray, C., Claramunt, C.: Modèle de graphe pour l'analyse des structures de trajectoires maritimes. In: 21ème édition de la conférence Extraction et Gestion des Connaissances (EGC) (2021)
11. Petrossian, G.A.: Preventing illegal, unreported and unregulated (iuu) fishing: A situational approach. Biological Conservation **189**, 39–48 (2015)
12. Ray, C., Dréo, R., Camossi, E., Jousselme, A.L., Iphar, C.: Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance. Data in Brief **Vol. 25**, 104141 (2019). https://doi.org/https://doi.org/10.1016/j.dib.2019.104141, http://www.sciencedirect.com/science/article/pii/S2352340919304950
13. Ray, C., Dréo, R., Camossi, E., Jousselme, A.L., Iphar, C.: Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance. Data In Brief **25** (2019)
14. Riveiro, M., Pallotta, G., Vespe, M.: Maritime anomaly detection: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **0**(0), e1266 (2018). https://doi.org/10.1002/widm.1266, https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1266
15. Vespe, M., Gibin, M., Alessandrini, A., Natale, F., Mazzarella, F., Osio, G.C.: Mapping eu fishing activities using ship tracking data. Journal of Maps **12**(sup1), 520–525 (2016)
16. Yu, H., Fang, Z., Lu, F., Murray, A.T., Zhao, Z., Xu, Y., Yang, X.: Massive automatic identification system sensor trajectory data-based multilayer linkage network dynamics of maritime transport along 21st-century maritime silk road. Sensors **19**(19) (2019). https://doi.org/10.3390/s19194197, https://www.mdpi.com/1424-8220/19/19/4197