Modèle de graphe pour l'analyse des structures de trajectoires maritimes

Maryam Maslek Elayam*, Cyril Ray*, Christophe Claramunt*

*Naval Academy Research Institute, F-29240 Brest, France {maryam.maslek_elayam, cyril.ray, christophe.claramunt}@ecole-navale.fr

Résumé. La représentation et l'analyse des données de mobilité peuvent être appréhendées à partir de différentes dimensions spatiales, temporelles et sémantiques inter-reliées. L'étude des structures émergentes devrait permettre de mieux comprendre les comportements et les tendances de mouvements d'entités de tout type et évoluant dans l'espace et le temps. Cet article propose une représentation des trajectoires maritimes basée sur un modèle de graphe, annoté avec une indexation hexagonale de l'espace. Ce modèle est implémenté dans une base de données de graphe en utilisant un référentiel de données historiques maritimes réelles. L'exploitation de requêtes exprimées sous le langage Cypher permet d'appliquer une série d'analyses caractéristiques conduisant à une meilleure compréhension de ces mobilités maritimes.

1 Introduction

Le volume mondial de production de données numériques continue de progresser à un rythme exponentiel. C'est particulièrement le cas pour l'espace maritime qui est le lieu d'une nette augmentation du volume de données issues de capteurs et de systèmes embarqués des navires, et qui permettent une meilleure gestion des trafics maritimes afin d'en assurer sa sécurité et une meilleure efficacité. L'acquisition de larges bases de données de positions de navires ouvre de nouvelles possibilités d'étude des flux, des mobilités et des structures des réseaux constituant ces espaces maritimes. De telles perspectives recouvrent l'analyse de patrons de déplacements, l'identification des propriétés des réseaux de transport et des flux maritimes et passent par la caractérisation des relations spatiales, temporelles et sémantiques entre les différents éléments qui constituent ces réseaux. Ces études de mobilité sont généralement menées à différentes échelles de l'espace et du temps, et permettent de générer des analyses selon différents angles de vue. Par exemple, les travaux effectués dans Renso et al. (2013) caractérisent des couloirs spatio-temporels à partir de données historiques de position, en permettant de définir les routes habituelles et d'identifier les comportements limites comme des navires en retard ou en avance, ou spatialement déviant en comparaison des routes principales.

Ces données de mobilité sont essentielles à la compréhension des activités maritimes, mais pas suffisantes dans la mesure où elles restent cependant essentiellement quantitatives. Il apparaît en effet que le déplacement d'un navire est conditionné par la nature de cet espace maritime (géographie, conditions en mer, situation météorologique, ...) et par des données sémantiques intrinsèques le caractérisant (type de navire, cargaison transportée, ...). Ces données

spatio-temporelles de mobilité doivent donc être complétées par des informations sémantiques pour favoriser les processus d'analyse et d'extraction intégrant cette dimension sémantique, et ce, afin de mieux comprendre les facteurs et motifs de déplacement, les comportements sousjacents et les différents types d'activités qui émergent dans ces environnements maritimes.

Afin d'analyser les patrons de déplacement dans un espace maritime, les données sémantiques associées aux navires doivent en être extraites pour donner un éventail plus complet des situations à analyser. Pour ce faire, il est indispensable de développer des outils de représentation sémantique et proposer de véritables solutions opérationnelles de stockage, de traitement et d'analyse des patrons émergent de ces larges volumes de données. Dans le but de mettre en œuvre une structure logique de représentation des trajectoires sémantiques, cet article propose une formalisation de ces trajectoires à partir d'un modèle de graphe pour définir les différents patrons et modalités des déplacements et activités sous-jacentes. Un tel modèle sémantique doit répondre à des principes de modélisation formelle et conceptuelle mais aussi à des principes de modélisation logiques permettant d'intégrer l'ensemble des données sémantiques et spatio-temporelles de ces navires en structures réseaux caractérisant les patrons de mobilité.

Le reste de cet article est structuré comme suit : la deuxième section présente l'état de l'art des travaux de recherche orientés vers la représentation et l'analyse de mobilité. La troisième section développe une formalisation d'un modèle de représentation des trajectoires maritimes. La quatrième section propose une implémentation du modèle. La dernière section conclut ce travail et présente quelques perspectives offertes par cette recherche.

2 État de l'art

Nous considérons la trajectoire d'un objet mobile comme une trace importante de sa mobilité Renso et al. (2013) et est caractérisée spatio-temporellement. Une trajectoire sémantique est définie comme une trajectoire spatio-temporelle enrichie avec des données sémantiques de différents types Ribeiro de Almeida et al. (2020). La construction d'un réseau de trajectoires sémantiques permet implicitement d'étudier la mobilité dans un espace géographique à différentes échelles de l'espace et du temps, et d'en extraire des patrons voire des comportements particuliers. D'après l'état de l'art effectué par Siabato et al. Siabato et al. (2018), il existe plusieurs modèle de représentation de l'information géographique dans un contexte spatio-temporel. Un graphe spatio-temporel est typiquement construit à partir d'un ensemble d'entités dynamiques présentes à des instants t_1, t_2, t_3 qui constitueront les nœuds du graphe, et ayant des relations entre elles qui formeront les liens de ce graphe Del Mondo (2011). Plusieurs champs d'application profitent des avantages que présentent les modèles de graphes pour étudier les mobilités, allant du domaine maritime à la mobilité des individus « indoor » et « outdoor » Noureddine et al. (2020), aux réseaux de transport en ville Gleyze (2001).

Dans la plupart des cas, un graphe de trajectoires est construit à des échelles spatiales et temporelles bien définies et pour une sémantique prédéfinie. A une échelle globale, Peng et al. Peng et al. (2019) modélisent le réseau mondial de transport de pétrole brut à partir d'un modèle de graphe. Ce réseau permet le suivi et l'analyse structurelle de la distribution du pétrole brut entre les ports d'exportation et d'importation, en analysant les structures réseaux et les dépendances régionales. A une échelle intermédiaire, Varlamis et al. (2019) utilisent un modèle global et considèrent les points de retournement et les variations de vitesses obtenus à partir d'une analyse des « clusters » de trajectoires. Le réseau ainsi obtenu

modélise non seulement les principaux liens du réseau des trajectoires mais aussi certaines particularités de ces liens au sein de ce réseau. A un niveau régional, Laddada *et al.* Laddada et Ray (2020) et Itani *et al.* Itani et al. (2019) construisent un graphe de trajectoires spatiotemporelles port à port à l'échelle de la Bretagne. D'autres études enrichissent un graphe de trajectoires maritimes par des données sémantiques afin de mieux apprécier les caractéristiques intrinsèques et d'étudier son évolution. Dans ce cadre, Soares et *al.* Soares et al. (2019) utilisent le concept de trajectoire sémantique pour modéliser le réseau maritime à partir de données de positions et notamment de données météorologiques. L'alimentation de ce modèle à partir d'un flux de données en temps réel permet la détection d'anomalies de comportements des navires et l'émission d'alertes en considérant la détérioration des conditions météorologiques.

Ce court état de l'art permet de constater une prépondérance effective et naturelle des modèles de graphe pour la représentation des données inter-connectées sous forme de réseaux. Nous pouvons remarquer que les modèles existant sont complémentaires mais sont souvent étudiés séparément. Dans la plupart des cas, ces études sont menées à un niveau spatio-temporel défini au préalable (pas de multitude de niveaux spatiaux et/ou temporels) et si un modèle hiérarchique est proposé, il est défini à partir d'un processus de classification des données d'un niveau à un niveau d'abstraction supérieure.

3 Modèle de représentation des trajectoires maritimes

Cette section présente la formalisation d'un modèle de graphe pour la représentation des trajectoires maritimes et l'analyse des structures émergentes. Le modèle proposé est **un graphe orienté**, constitué de noeuds et d'arêtes représentants les relations entre ces noeuds. Afin de mieux représenter les trajectoires maritimes et leur environnement structurel, ce modèle comprend une diversité de noeuds de différents types : ports, zones de mouillage, zones d'attente, points de retournement, zones protégées et différents types d'informations sémantiques associées : données météorologiques, état de la mer...

Soit G=(V,E) un modèle de graphe spatio-temporel orienté avec V l'ensemble des noeuds et E l'ensemble des arêtes. V et E sont deux ensembles disjoints et non vides.

Définition 1 : Une trajectoire spatio-temporelle (TS) est un sous-graphe de G, orienté, défini par une séquence de noeuds et d'arêtes tel que :

$$TS(id) = (v_1, e_1, v_2, e_2,, e_{n-1}, v_n)$$
(1)

où tout $(v_i)_{i=1..n} \in V$ est spatialement référencé par une position $p_i = (x_i, y_i)$ et temporellement référencé par un intervalle temporel $t_i = [t_{i1}, t_{i2}]$ tel que $t_{i1} > t_{(i-1)2}$ et $t_{i2} < t_{(i+1)1}$; $(e_i)_{i=1..n-1} \in E$ dénote le lien orienté entre v_i et v_{i+1} , et est qualifié par une durée δ_i .

Par construction, une trajectoire a nécessairement un noeud de départ, un noeud d'arrivée et peut passer (ou pas) par un ou plusieurs noeuds intermédiaires.

Le graphe des trajectoires ainsi défini représente la dimension spatio-temporelle d'une trajectoire maritime. Dans notre modèle, l'adjonction d'une composante sémantique se traduit sur le graphe par l'étiquetage (ou l'annotation) des éléments du graphe par des données de différents types. Celle-ci permet d'enrichir l'information portée par le graphe et d'appliquer des analyses comportementales liées à l'environnement maritime. **Définition 2 :** Une trajectoire spatio-temporelle sémantique (TSS) est définie par le sousgraphe orienté de triplets :

$$TSS(id) = \{(s_1, v_1, q_1), e_1, (s_2, v_2, q_2), e_2, \dots, e_{n-1}, (s_n, v_n, q_n)\}$$
(2)

avec : $s_i \in S$ où $S = \{\text{port, zone d'attente, zone de pêche, zone de mouillage,...}\}$ est l'ensemble des types des éléments de V; v_i est la composante spatio-temporelle du noeud i et telle que $v_i = (p_i, t_i)$ avec $p_i = (x_i, y_i)$ une position et $t_i = [t_{i1}, t_{i2}]$ tel que $t_{i1} > t_{(i-1)2}$ et $t_{i2} < t_{(i+1)1}$; $q_i = (q_{i1}, ..., q_{ik}), k \in \mathbb{N}$ est la composante sémantique du noeud i introduite par les paramètres $(q_{i1}, ..., q_{ik})$ (sachant que cette composante sémantique est extensible en terme d'attributs encapsulés).

L'annotation des trajectoires en fonction de l'espace est un élément indispensable de localisation des différents éléments des trajectoires et permet de mettre en oeuvre des mesures et des requêtes spatiales. Une telle approche permet également de représenter un ensemble de données spatio-temporelles et sémantiques à priori volumineuses dans un espace discrétisé, ce qui permet notamment de réduire le temps de calcul par l'utilisation d'index physiques spatiaux. Il existe plusieurs méthodes de discrétisation d'un espace géographique autour de méthodes à grille régulière ou irrégulière Samet (1990).

Nous avons choisi une grille hexagonale pour la discrétisation de l'espace géographique et pour la représentation de la composante sémantique. Cette méthode permet une subdivision de l'espace en hexagones homogènes à une certaine résolution. Chaque hexagone a 6 voisins dont les centres sont équidistants à l'hexagone central Sahr et al. (2003). Cet avantage facilite la recherche optimale des voisins pour l'annotation des nœuds ainsi que l'analyse et le lissage des champs et des gradients des informations sémantiques. Ce type de grille a également montré son efficacité pour la prévision des routes optimales, sous plusieurs contraintes de l'environnement maritime Tsatcha (2014).

Définition 3 : H est une indexation spatiale, hexagonale et hiérarchique si et seulement si :

$$\forall (x,y) \exists \{I^{r_1}, I^{r_2}, ..., I^{r_m}; r_1 < r_2 < ... < r_m\} \in H$$

$$(x,y) \subset h_{r_1} \subset h_{r_2} \subset \subset h_{r_m}$$
(3)

où (x,y) sont des coordonnées géographiques, I^{r_i} est l'indice de la grille hexagonale h incluant (x,y) à la résolution r_i .

Définition 4 : Une trajectoire spatio-temporelle sémantique représentée dans un système d'indexation spatiale, hexagonale et hiérarchique H s'écrit comme suit :

$$TSS_{H}(id) = \{(s_{1}, \tilde{v}_{1}, \tilde{q}_{1}), e_{1}, (s_{2}, \tilde{v}_{2}, \tilde{q}_{2}), e_{2}, ..., e_{n-1}, (s_{n}, \tilde{v}_{n}, \tilde{q}_{n})\}$$
(4)

où s_i est le type du noeud i; \tilde{v}_i est la composante spatio-temporelle du noeud i et telle que $v_i=(I_i^r,T_i),\,I_i^r\in H$ est l'indice de représentation spatiale, hexagonale et hiérarchique correspondant au point de coordonnées (x_i,y_i) à la résolution $r,t_i=[t_{i1},t_{i2}]$ est un intervalle de temps entre les instants t_{i1} et t_{i2} ; $\tilde{q}_i=(\tilde{q}_{i1},...,\tilde{q}_{ik}),k\in\mathbb{N}$ est une représentation de la composante sémantique du noeud i représentée dans l'espace discrétisé par l'indexation H durant l'intervalle du temps t_i .

Cette représentation spatiale hiérarchique permet une composition de différentes représentations selon différents niveaux d'abstraction dans l'espace et dans le temps. Cette approche

permet ainsi une représentation par graphe hiérarchique allant du graphe sémantique de trajectoires au graphe de flux maritimes. La construction de ce type de graphes est faite à partir des graphes de trajectoires et repose sur l'extraction et l'agrégation, par niveau d'abstraction, des différents flux qui transitent à l'échelle des trajectoires.

4 Implémentation et résultats

Notre modèle a été implémenté et testé sur un jeu de données réelles. Pour cette implémentation nous avons combiné Python, la librairie MovingPandas, l'index Uber H3, le système de gestion de base de données graphe Neo4j et son langage de requête Cypher.

Données et pré-traitements : Plusieurs systèmes de localisation de navires co-existent et transmettent des données nominatives et de position à fréquences plus ou moins élevées (quelques secondes à quelques heures). Parmi ces systèmes, l'*Automatic Identification System* (AIS) est le plus répandu Ducruet (2017). Un transpondeur AIS est un système d'anticollision collaboratif qui envoie et reçoit, en continu et à haute fréquence, les positions des navires vers les navires environnants et les centres de surveillance terrestres.

Nous avons utilisé un mois (Janvier 2016) de données AIS couvrant l'espace maritime Européen. Ce jeu de données contient 3,72 GO de données avec plus de 61 millions de positions et 74 949 navires. Les données sont structurées en 3 ensembles : les positions des navires, les informations nominatives et les localisations de 550 ports. Les positions des navires contiennent les paramètres suivants : un identifiant international (MMSI), son état de navigation (e.g. stop), sa vitesse de giration (° par min), sa vitesse de déplacement (knot), son cap (° au nord), la position du navire (longitude et latitude exprimées en WGS84), l'empreinte temporelle (t) indiquant le moment de la réception du message AIS. Les informations nominatives comprennent les identités et les données de voyages effectués par les navires et sont décrites par les champs suivants : le MMSI, l'indicatif international d'appel radio (callsign), le nom du navire, son type (e.g. cargo), sa destination, ses dimensions et l'estampille temporelle (t) correspondant au moment de réception du message AIS.

La construction du graphe de trajectoires sémantiques se base sur une étape de pré-traitement afin de représenter chaque trajectoire avec un minimum de points représentatifs et annotés. La première étape consiste à supprimer les positions avec des vitesses nulles (navires à l'arrêt, au mouillage, vitesse non renseignée), ce qui représente 60.6% des données. Nous avons extrait les trajectoires en utilisant *MovingPandas* qui permet de générer des trajectoires en respectant l'ordre temporel des positions associées à chaque navire. Elle présente aussi des fonctions de visualisation et d'analyse de trajectoires Graser (2019). A l'issu de ces étapes, le *jeu de données* contient 23 635 trajectoires.

L'étape suivante est la compression des trajectoires en ne gardant que les Points Tournants Significatifs (PTS) : ce sont les points de changement de la direction du navire, portée par la pente entre k points, dont la variation est supérieure à 10° . Nous annotons les PTS selon plusieurs propriétés géographiques, de densité et de proximité.

Similairement à Itani et al. (2019), nous procédons à un *clustering* par densité des PTS de nos trajectoires. Pour ce faire, nous avons utilisé l'algorithme HDBSCAN. Les PTS impliqués dans les *clusters* sont étiquetés par le numéro du cluster auquel ils appartiennent. Les autres points sont supprimés.

La dernière étape consiste à réaliser une annotation spatiale des données, comme évoqué en section 3. Deux principaux systèmes de représentation spatiale à base d'hexagones existent : les indexes ISEA3 et Uber H3. Nous avons retenu l'index Open Source H3 Brodsky (2018). Chaque hexagone a un identifiant unique et permet de représenter des coordonnées géographiques en cellules homogènes à une certaine résolution. Ce système contient 15 résolutions différentes.

Base de données graphe: A l'issue de ces pré-traitements, nous obtenons des trajectoires compressées et annotées en fonction de la grille H3 de façon à reconnaître, pour chaque trajectoire, son port de départ, son port d'arrivée et les PTS. Ces données sont par la suite intégrées dans une base de données graphe Neo4j. La figure 1 présente le principe de notre modèle (cf. Section 3) implémenté dans Neo4j. Ce modèle contient plusieurs types de noeuds et d'arêtes, annotés par des attributs spatio-temporels. Le schéma permet de représenter chaque navire et sa trajectoire subdivisée en segments. Ces derniers sont connectés à leurs PTS (WayPoint) de départ et d'arrivée. Notre base de données est composée de 25 976 noeuds de trois types : navires, ports et PTS, et de 5 197 296 relations.

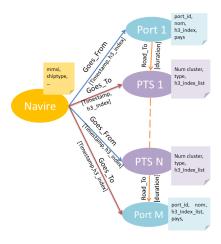


FIG. 1: Modèle de graphe pour la représentation des trajectoires maritimes sous Neo4j

Les requêtes sur la base de données sont effectuées en Python. Elles peuvent produire par exemple des résultats statistiques ou des sous-graphes de trajectoire. L'exemple suivant exploite le modèle proposé, sa dimension spatiale et temporelle. L'information recherchée étant : quels sont les navires qui partent du Maroc vers la France et traversant le Détroit de Gibraltar entre le 01 et le 15 Janvier?

```
MATCH (p1:Port {country:'MA'}) <-[r:GOES_FROM] - (s:Ship)
    -[g:GOES_TO{hexagon:$hexa}] -> (w:WayPoint)

MATCH p=(s) - [f:GOES_TO*] -> (p2:Port {country:'FR'})
WHERE r.time <= datetime('2016-01-15T00:00:00Z')
RETURN s.ship_id AS MMSI</pre>
```

Cette requête Cypher extrait ([MATCH]) tous les MMSI des navires au départ des ports Marocains ([GOES_FROM]) vers ([GOES_TO]) les ports Français en passant par le détroit de

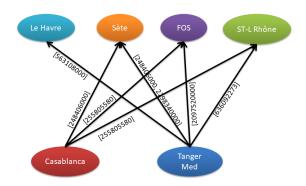


FIG. 2: Sous-graphe des trajectoires résultantes de la requête Cypher

Gibraltar sur un intervalle de deux semaines. Le détroit de Gibraltar est représenté par 3 781 hexagones (hexa) indicés par H3 à la résolution 9. La figure 2 illustre0 le graphe des trajectoires résultant de cette requête. Il s'agit d'une liste de 6 navires faisant les trajets : Casablanca \rightarrow Saint-Louis-du-Rhône, Casablanca \rightarrow Fos, Casablanca \rightarrow Sète, Tanger Méditerranée \rightarrow Saint-Louis-du-Rhône, Tanger Méditerranée \rightarrow Fos, Tanger Méditerranée \rightarrow port du Havre, Tanger Méditerranée \rightarrow Sète.

5 Conclusion et perspectives

L'analyse des données de mobilité vise à comprendre les comportements, les tendances de mouvement des entités évoluant dans l'espace et le temps, ainsi que leur connectivité au travers des dimensions spatio-temporelles et sémantiques. L'abstraction de ces données à partir d'une structure de graphe offre de nouvelles opportunités pour une meilleure compréhension des mobilités. Cet article développe un modèle de graphe annoté d'informations spatio-temporelles et sémantiques, combinées à une indexation spatiale hexagonale. L'implémentation de ce modèle dans une base de données graphe permet une meilleure compréhension de la mobilité des navires. Dans la suite de ce travail, notre objectif sera de définir une représentation hiérarchique selon différents niveaux d'abstraction. Des fonctions caractéristiques associées au graphe hiérarchique (similarité, accessibilité,...) seront définies et permettront de décrire, comprendre et prédire les trajectoires, leurs motifs et leur dynamique temporelle.

Références

Brodsky, I. (2018). H3: Uber's hexagonal hierarchical spatial index. *Available from Uber Engineering website: https://eng. uber. com/h3/[22 June 2019].*

Del Mondo, G. (2011). Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques. Ph. D. thesis, University of Western Brittany, France.

Ducruet, C. (2017). Advances in Shipping Data Analysis and Modeling: Tracking and Mapping Maritime Flows in the Age of Big Data. Routledge.

- Gleyze, J.-F. (2001). Réseaux, territoires et accessibilité. *Institut Géographique National, Laboratoire COGIT*.
- Graser, A. (2019). Movingpandas: Efficient structures for movement data in python. GIForum 1, 54–68.
- Itani, A., C. Ray, A. El Falou, et J. Issa (2019). Mining ship motions and patterns of life for the eu common information sharing environment (cise). In *OCEANS 2019-Marseille*, pp. 1–6. IEEE
- Laddada, W. et C. Ray (2020). Graph-based analysis of maritime patterns of life. In *Atelier Gestion et Analyse de données Spatiales et Temporelles (GAST@ EGC)*.
- Noureddine, H., C. Ray, et C. Claramunt (2020). Semantic trajectory modelling in indoor and outdoor spaces. In 2020 21st IEEE International Conference on Mobile Data Management (MDM), pp. 131–136. IEEE.
- Peng, P., Y. Yang, S. Cheng, F. Lu, et Z. Yuan (2019). Hub-and-spoke structure: Characterizing the global crude oil transport network with mass vessel trajectories. *Energy 168*, 966–974.
- Renso, C., S. Spaccapietra, et E. Zimányi (2013). *Mobility Data*. Cambridge University Press.
- Ribeiro de Almeida, D., C. de Souza Baptista, F. Gomes de Andrade, et A. Soares (2020).
 A survey on big data for trajectory analytics. ISPRS International Journal of Geo-Information 9(2), 88.
- Sahr, K., D. White, et A. J. Kimerling (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science* 30(2), 121–134.
- Samet, H. (1990). Applications of spatial data structures.
- Siabato, W., C. Claramunt, S. Ilarri, et M. Á. Manso-Callejo (2018). A survey of modelling trends in temporal gis. *ACM Computing Surveys (CSUR)* 51(2), 30.
- Soares, A., R. Dividino, F. Abreu, M. Brousseau, A. W. Isenor, S. Webb, et S. Matwin (2019). Crisis: integrating ais and ocean data streams using semantic web standards for event detection. In 2019 International Conference on Military Communications and Information Systems (ICMCIS), pp. 1–7. IEEE.
- Tsatcha, D. (2014). Contribution à l'extraction et à la représentation des connaissances de l'environnement maritime : Proposition d'une architecture dédiée. aux applications de navigation. Ph. D. thesis.
- Varlamis, I., K. Tserpes, M. Etemad, A. S. Júnior, et S. Matwin (2019). A network abstraction of multi-vessel trajectory data for detecting anomalies. In *EDBT/ICDT Workshops*.

Summary

The study of the emerging structures of maritime networks should provide a better understanding of transportation patterns at spatial, temporal and semantic levels. This paper introduces a graph-based model for representing maritime trajectories combined with a spatial hexagonal indexing annotation. This model is implemented on top of a graph database and experimented with a sample of maritime historical data. A series of Cypher queries illustrate the potential of the approach for the extraction of some representative maritime network structures.