

CARACTÉRISATION DES DYNAMIQUES SPATIALES DANS LE CADRE DE SOURCES GÉOLOCALISÉES ET NON GÉOLOCALISÉE

Démarrage : septembre 2023

Contacts : geraldine.del_mondo@insa-rouen.fr, benoit.gauzere@insa-rouen.fr

Lieu : LITIS, INSA Rouen

Profil attendu : Master 2 en informatique/IA, des compétences minimales en théorie des graphes sont requises

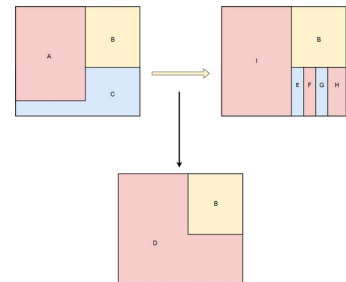
Salaire mensuel brut : 2044,12 euros

Mots clés : algorithme de graphe, extraction de patterns spatio-temporel, isomorphismes de graphes, application aux données historiques, modélisation par graphes, analyse spatio-temporelle.

OBJECTIFS ET CONTEXTE

Cette thèse en informatique se situe dans les domaines de la modélisation et du traitement de données spatiales et temporelles via des modèles fondés sur les graphes, et de l'analyse de données historiques. Son encadrement bénéficie de l'expertise de chercheurs en informatique¹ et en histoire².

L'objectif général se place dans un cadre plus large issu d'un besoin de chercheurs en histoire qui souhaitent développer une procédure d'analyse des dynamiques spatiales des peuplements à partir de sources de données de différents types (e.g. plans cadastraux, photographies aériennes, documents dépourvus de plan comme des registres fiscaux). Cette procédure doit rendre possible l'analyse des dynamiques spatiales (e.g. changements dans la structure des parcellaires, figure de droite) non seulement sur les 200 à 300 dernières années, mais aussi de remonter aux derniers siècles du Moyen Age lorsque la documentation le permet. Un des verrous à cette problématique est l'aspect incomplet et sporadique des données qui complique les approches fondées sur des modèles statistiques.



Dans ce but, l'idée générale est que la transformation des différents types de données dans un modèle commun (i.e. un graphe) peut permettre à la fois d'extraire plus facilement de l'information et de pouvoir les comparer. L'intérêt est notamment de détecter des patterns spécifiques (e.g. des forêts, qui ne sont pas inscrites dans les registres car non imposées ; connexions entre parcelles particulières) et des patterns de changement. Pour ces derniers, on s'intéresse particulièrement à des changements de haut niveau (e.g. périodes de division/fusion de parcelles ; identification de zone dynamique vs. stable en termes de changement).

Des outils pour aider les historiens

Les travaux réalisés dans cette thèse sont directement liés à ceux menés dans le cadre d'une ANR (Modelespace³, porteur Florent Hautefeuille), dans laquelle a été proposé un modèle de graphe d'adjacence (Figure 1) pour modéliser les données. Il s'avère que

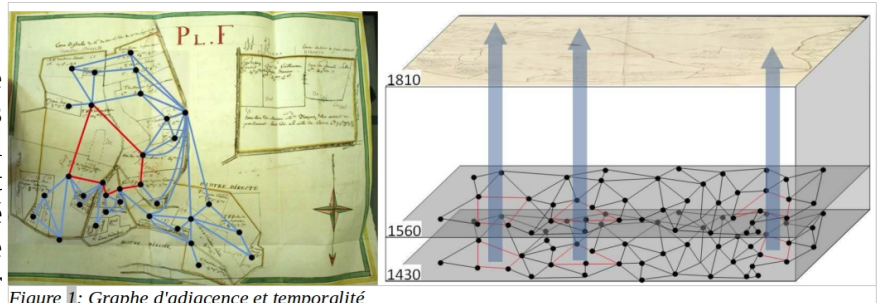


Figure 1: Graphe d'adjacence et temporalité

1 Clément Chatelain, Géraldine Del Mondo, Benoît Gauzère, LITIS, ROUEN

2 Florent Hautefeuille, TRACES, Toulouse

3 <http://modelespace.univ-tlse2.fr/>

ce modèle atteint rapidement ses limites en termes d'expressivité permettant de caractériser les patterns spatiaux et temporels, a fortiori sur des données non géolocalisées [1]. Il n'y a pas eu non plus d'automatisation d'extraction de patterns dans ce projet. Il s'agit donc dans cette thèse d'aller au-delà de ce modèle et de proposer les algorithmes d'extraction de patterns sur un nouveau modèle de graphe via des méthodes de modélisation et d'analyse de graphes [2, 4]. On se restreindra aux données parcellaires de type cadastre, géolocalisées ou non (et qui s'appuient ou non sur les supports planimétriques).

L'information spatiale dans les graphes dynamiques : une dimension à prendre en compte

Dans le cadre de ce travail, un pattern est déterminé non seulement par les contraintes proposées par les experts du domaine, mais aussi par les contraintes géographiques, ces dernières étant fournies dans une première approche par le biais de graphes d'adjacence. Un pattern spatio-temporel est une séquence de sous-graphes où chaque sous-graphe est extrait d'un graphe statique. Les relations temporelles entre les sous graphes sont rajoutées, créant ainsi un sous-graphe dynamique. Les temps associés à ce sous-graphe dynamique sont un sous-ensemble consécutif de temps disponible. Ce pattern spatio-temporel peut caractériser l'évolution particulière d'un ensemble de parcelles au cours du temps, pour permettre une étude plus approfondie par les historiens.

Une étude à plusieurs niveaux d'abstraction

Pour atteindre ces objectifs, il est potentiellement nécessaire d'utiliser une analyse multi-échelle (c'est-à-dire abstraire les détails dans différentes couches) en raison de la nature incomplète et potentiellement ambiguë des données. On peut imaginer représenter une série de partitions de l'espace liées entre elles à différents niveaux de détail, et la relier à une description qualitative du changement de granularité [3]. L'objectif étant de structurer le modèle spatio-temporel mis en œuvre selon plusieurs niveaux de granularité afin de pouvoir les exploiter lors de l'analyse des données.

RÉSUMÉ DES PRINCIPAUX OBJECTIFS DE LA THÈSE ET PERSPECTIVES

1. Proposer un nouveau modèle adapté permettant de dépasser les limites du modèle basé sur le graphe d'adjacence
2. Définir formellement les typologies des patterns spatiaux et temporels. Il y a *a minima* deux catégories de patterns, ceux au niveau local (une parcelle) et ceux au niveau global (groupe de parcelles).
3. Élaborer les algorithmes de recherche de ces patterns

Ces objectifs sont dans un premier temps à réaliser en utilisant des graphes déjà spatialisés. Mais le deuxième challenge de la thèse repose sur l'usage de données non géolocalisées qui nécessitent un pré-traitement pour faire correspondre (appariement) les parcelles d'un temps à un autre (figure de droite). Une technique envisagée en cas de changement localisé à petite échelle consiste à rechercher des correspondances entre des versions moins détaillées des graphes afin d'identifier les aspects inchangés (i.e. analyse multi-échelle). Dans tous les cas, les travaux menés sur les graphes spatialisés serviront de *baseline* pour cette extension du travail.



Une attention particulière sera apportée à l'importance de la généralité des approches développées. Des données immédiatement disponibles (voir détail ci-dessous) concernent les régions du sud de la France (Odars, Toulouse). Plusieurs autres corpus sont identifiés et facilement mobilisables dans la même aire (Verfeil, Saint Jean Lherm, Vallesvilles...) ; mais nous prévoyons de tester les approches sur de plus grands jeux de données concernant d'autres régions (données opendata Etalab), dont la région Normandie.

En effet, si ce travail de thèse est focalisé sur la recherche de patterns historiques, caractériser des patterns dans d'autres domaines est possible [2]. Dans la mesure où les caractéristiques de ces patterns se fondent sur les dimensions spatiale et temporelle des données, la transposition des méthodes proposées dans un nouveau cadre est tout à fait envisageable. Dans ces conditions, l'étude de données environnementales en lien avec des parcelles agricoles ou la mobilité pourrait être une application à moyen terme des méthodes développées dans cette thèse.

DONNÉES À DISPOSITION

ODARS (Arch. dép. 31 - AMT 589) : Cet ensemble de données est dédié à l'analyse historique. C'est une base de données relationnelle relative à la région d'Odars avec six années d'évolution cadastrale (1476, 1497, 1551, 1598, 1759, 1811), deux sont géolocalisées et toutes ont déjà été transformées en graphes (respectivement 416 et 686 nœuds, 1158 et 1957 arêtes) ;

Les données non géolocalisées proviennent de sources fiscales pré-révolutionnaires. Elles sont comparables aux registres fonciers, dont la plupart sont dépourvus de cartes. L'identification et la localisation des biens sont basées sur une série de données variables. Elle contient généralement le nom du propriétaire, la nature du bien, sa valeur, un toponyme de localisation et la désignation des voisins de la parcelle concernée (avec ou sans orientation). C'est l'existence de cette dernière information (les confronts) qui permet de modéliser les contiguïtés et de tenter de reconstituer graphiquement l'espace décrit dans les registres. La base de données a été nettoyée et testée par un étudiant en Master en 2019 (dans le cadre d'un financement de la fédération NORMASTIC).

TOULOUSE (Arch. Mun. Toulouse,) : Toulouse constitue un deuxième jeu de données remarquable utilisé pour tester la généralité des approches développées. Contrairement à Odars, nous sommes ici dans un cadre urbain, même si 95 % du territoire communal est rural. Toulouse offre l'un des plus riches corpus disponibles à l'échelle française et européenne. Le premier document fiscal conservé, un livre d'estimes, date de 1335. Par la suite, les archives municipales ont conservé des cadastres sans plan en 1458, 1478, 1550, 1571. L'exceptionnel cadastre avec plan géométrique de 1680 est le plus ancien de France pour une juridiction de cette dimension. Un second document avec plans a été réalisé à partir de 1788. Le cadastre napoléonien réalisé dans les années 1830 n'est donc que la troisième couche historique avec données planimétriques systématiques. Celle-ci et le plan de 1680 sont déjà vectorisés et géoréférencés dans un SIG. L'autre intérêt de ce second corpus est son étendue, le document de 1830 compte près de 35 000 parcelles.

DGFIP / ETALAB : Cet ensemble sera utilisé pour tester nos algorithmes avec une plus grande quantité de données afin de vérifier le passage à l'échelle de nos approches. Il est composé d'un ensemble contemporain de 100 millions de parcelles actives, et à partir de 1990, environ 20 millions de filiations de parcelles sont également disponibles.

[1] S. Leturcq and R. Raveaux. Les graphes pour étudier les dynamiques spatiales à partir des séries fiscales médiévales et modernes, état des lieux de l'expérience modelespace. Bulletin du centre d'études médiévales d'Auxerre ; BUCEMA, 9, 2016.

[2] K. S. Oberoi and G. Del Mondo, « Graph-based Pattern Detection in Spatio-Temporal Phenomena, » in 16th Spatial Analysis and Geomatics Conference (SAGEO 2021), La Rochelle (virtuel), France, 2021

[3] J. G. Stell. Granular description of qualitative change. In 23rd International Joint Conference on Artificial Intelligence, pages 1111–1117, 2013

[4] K. S. Oberoi, G. Del Mondo, B. Gaüzère, Y. Dupuis, P. Vasseur Detecting Dynamic Patterns In Dynamic Graphs Using Subgraph Isomorphism, Pattern Analysis and Application, Accepté Janvier 2023