

APPEL A PROJETS RECHERCHE 2023

FEDERATION DE RECHERCHE AGORANTIC

«CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES »

Attention :

- Ne pas dépasser 5 pages
- Transmettre le fichier au format PDF intitulé : ACRONYME-AAP-blc-Agorantic-2023
- Envoyer le fichier à agorantic@univ-avignon.fr **avant le 24 mars 2023.**

| | |
|---|--|
| Titre | De la source historique au graphe de confronts |
| Acronyme | HistoGraPhe - Histoire et Graphes de Confronts pour l'Exploitation de la Documentation Foncière |
| Nom du/des porteur(s) | Vincent Labatut / Guido Castelnuovo |
| Laboratoires associés | LIA / CIHAM |
| Budget demandé | 7500€ |
| Résumé Max. 1 000 caractères espaces compris | Le projet HistoGraPhe vise à exploiter les livres fonciers médiévaux en utilisant la théorie des graphes pour résoudre les problèmes de spatialisation des données socio-foncieres. Le projet se concentre sur la documentation avignonnaise et cherche à atteindre deux objectifs principaux. Le premier objectif est de spatialiser les informations détenues dans les sources historiques en utilisant la position absolue des objets localisables et le positionnement relatif que constituent les confronts. Le deuxième objectif consiste à analyser la documentation en utilisant les graphes pour détecter les manques dans les données, les combler, pour enfin réaliser une étude multiscale de la composition sociale de l'espace urbain. |

1. Contexte, positionnement, objectif(s)

À l'orée du XIII^e siècle, la documentation foncière s'accroît considérablement et constitue, à partir de là, un pan majeur de la documentation écrite (Fossier, 1978 ; Chastang, 2009). Qu'il s'agisse d'enquêtes domaniales, de terriers proprement dits, de censiers ou de livres de reconnaissances, il est question d'inscrire et de garder une trace écrite de ce qui est possédé. Ainsi, ce sont, avant tout, les déclarations des tenanciers qui énumèrent les biens qu'ils possèdent sur les terres d'un seigneur et pour lesquels ils doivent payer une redevance annuelle.

Ces livres fonciers de la fin du Moyen Âge ont rapidement attiré le regard des historiens et des géographes – et plus particulièrement des ruralistes suivis par les historiens des *Annales* – qui ont su déceler dans ces documents une source inépuisable d'analyse aux thématiques riches et variées. Les chercheurs se sont toutefois heurtés à plusieurs freins dans l'analyse de cette documentation : en premier, son caractère lacunaire et incomplet ; elle couvre rarement l'ensemble d'un territoire. Il faut donc, bien souvent, assembler différentes sources pour obtenir une vision des plus exhaustives d'un espace donné. En second, les informations contenues dans les sources, bien que de plus en plus précises à partir du XIV^e siècle, sont toujours relatives, notamment en ce qui concerne les

informations spatiales. La mention des *confronts*¹, c'est-à-dire du voisinage, tend à se généraliser à partir du XIV^e siècle. Cette approche consiste par exemple à situer un bien au nord d'un autre bien.

Longtemps, les historiens ont œuvré à la reconstitution du parcellaire² pensant qu'il s'agissait d'une étape obligatoire pour l'analyse de cette documentation. Mener à bien ces entreprises de reconstitution de façon fiable est quasiment impossible, tous s'accordent aujourd'hui sur ce point (Montpied, Rouault, 1989). Plusieurs approches, empruntées aussi bien aux sciences humaines et sociales qu'aux sciences exactes et expérimentales, peuvent aujourd'hui permettre de résoudre la question épineuse de la spatialisation des données socio-foncières sans passer par la case reconstitution. C'est ce que nous proposons de développer dans ce projet.

Il est question de se tourner vers la théorie des graphes, qui est particulièrement adaptée à la modélisation des relations, et de formaliser une méthode permettant une analyse plurielle de la documentation foncière. Nous partirons de la documentation avignonnaise et plus spécifiquement du corpus de sources construit dans la thèse de Margot Ferrand (Ferrand, 2022).

Deux objectifs sont définis :

Le premier concerne la spatialisation des informations détenues dans les sources historiques. Si certains objets historiques sont facilement localisables, on pense par exemple aux églises encore aujourd'hui en place ou dont les vestiges nous permettent d'estimer la localisation, d'autres objets, et notamment les « biens personnels » déclarés et non déclarés – maisons, jardins, cour, etc. – sont quasiment impossibles à spatialiser directement. Nous proposons d'exploiter conjointement la position absolue des objets qui ont pu être localisés et le positionnement relatif que constituent les confronts, pour estimer ces positions inconnues.

Le second objectif concerne plus spécifiquement l'analyse historique de la documentation rendue possible par l'utilisation des graphes. Il est, tout d'abord, question d'identifier les manques dans les données déjà collectées et organisées dans la base de données *Terrier* de Margot Ferrand (La base de données est hébergée sur un serveur externe Huma-Num). Nous nous focaliserons notamment sur les données sociales qui peuvent facilement être enrichies par des dépouillements archivistiques ciblés (Fonds Anne-Marie Hayez des archives départementales du Vaucluse). Une fois enrichie, la détection de communautés dans le graphe doit permettre de réaliser une étude multiscalaire de la composition sociale de l'espace urbain. Il sera notamment permis de prendre en compte les différents statuts socio-professionnels des tenanciers, statuts qui ont jusqu'ici été mis de côté.

2. Problématique, questionnement scientifique, axe(s) de la FR concerné(s)

La thèse de M. Ferrand propose une méthode d'extraction des graphes pour l'analyse d'un territoire donné à partir de la documentation foncière médiévale, ici Avignon. Elle est basée sur l'étude de la corrélation entre distance géographique et distance géodésique. Il est à présent question de réfléchir à l'exploitation à proprement parler du graphe d'un point de vue aussi bien informatique qu'historique.

Ce projet pose ainsi deux problématiques :

1. D'un point de vue informatique, il sera question de réfléchir à gérer l'incertitude géographique et de proposer une méthode visant à localiser les données non spatialisées en prenant en compte l'orientation géographique lorsque celle-ci est disponible dans les sources.

¹ Sur les confronts, nous renvoyons à la définition d'Hadrien PENET, « Le sens des limites. Construction et perception de l'espace dans les actes de la pratique : l'exemple sicilien (XII^e-XV^e siècle) », dans *Construction de l'espace au Moyen Âge : pratiques et représentation, Acte des congrès de la Société des historiens médiévistes de l'enseignement supérieur public*, Paris, Éditions de la Sorbonne, 2006, p. 405-411. « L'espace des confronts est, par définition, un espace de la contiguïté, littéralement « confiné » (...) L'espace du confront est donc aussi bien un espace de voisinage ».

² Ensemble de parcelles d'un terroir, d'une commune, ou manière dont le sol utilisable a été divisé.

2. D'un point de vue de l'analyse historique, la détection des communautés sera utilisée pour compléter notre connaissance de la topographie sociale de la cité à la fin du Moyen Âge. Pour ce faire, les informations relatives à chaque nœud du graphe de confronts doivent être complétées (notamment les informations relatives aux propriétaires des biens).

Axes de la FR concernés : Ce projet s'inscrit dans trois axes de la FR Agorantic : l'Axe 1 « Méthodologie et Interdisciplinarité », l'Axe 4 « Expériences et mises en récits de l'urbain », et l'Axe 5 « Structuration et exploitation de corpus ».

3. Méthodologie

Dans le graphe spatial proposé dans la thèse de M. Ferrand, les sommets représentent des biens, rues, et autres points de repère géographiques, et les arêtes correspondent à leurs positions relatives (i.e. v_1 est au nord/sud/est/ouest de v_2). Pour certains de ces sommets, la position GPS a été identifiée manuellement, mais pas pour d'autres (faute d'information). Pour y remédier et proposer une localisation de ces objets, nous interpolons ces coordonnées en moyennant celles des sommets voisins (à partir de la mention des confronts). Nous ne prenons donc pas en compte l'orientation des biens lorsqu'elle est renseignée dans les sources. Il serait ainsi préférable d'explorer des méthodes plus élaborées afin d'améliorer ces approximations. Nous proposons de **tester sur ces données historiques le pouvoir prédictif des réseaux de neurones convolutifs sur graphes**. Ces réseaux (Graph Neural Networks) constituent une généralisation des réseaux de neurones, destinée à être appliquée à des données prenant la forme de graphes (par opposition aux données numériques tabulaires classiques). Nous nous intéresserons à la version *convolutive* de cette méthode (*Graph Convolutional Networks*), qui permet notamment de prédire des valeurs manquantes dans un graphe incomplètement décrit par des attributs attachés à ses sommets ou à ses arêtes.

Il s'agit donc de définir, dans un premier temps, une méthode pour l'interpolation de la position géographique des objets historiques. Il est ensuite question d'**élaborer un modèle générateur de données**. Ce dernier permettra de fournir une vérité terrain permettant de valider la méthode de spatialisation proposée en amont.

En parallèle de cette recherche, un **travail de collecte et de dépouillement archivistique** sera mené (Archives départementales du Vaucluse, Archives municipales d'Avignon, Archives apostoliques du Vatican³) pour compléter certaines informations d'ores et déjà présentes dans la base de données *Terrier*. Dans ce projet, nous nous focaliserons avant tout sur le statut socio-économique des propriétaires fonciers. Les données relatives à chaque nœud du graphe seront notamment affinées à partir des informations présentes dans le fonds prosopographique des archives d'Anne-Marie Hayez conservées aux Archives départementales du Vaucluse. Une fois les dépouillements réalisés, il sera permis de réaliser une **analyse historique des données à partir de la détection des communautés du graphe**.

4. Résultats attendus et caractère innovant de la recherche

Comme expliqué précédemment, la documentation foncière fait partie des types de documents les plus importants à partir du XIII^e siècle (d'un point de vue quantitatif tout d'abord, mais également qualitatif : elle offre des informations cruciales sur la gestion domaniale d'un espace donné, sur l'organisation d'un territoire, sur la société). Du point de vue historique, la proposition et l'évaluation d'une approche permettant de les exploiter sans passer par l'étape difficile de reconstitution du parcellaire est donc extrêmement importante méthodologiquement parlant : elle pourrait être ensuite appliquée à des corpus similaires issus d'autres régions ou périodes.

³ Les registres qui nous intéressent sont numérisés et consultables au CIHAM.

Du point de vue informatique, les GNN constituent une méthode très récente pour l'apprentissage automatique de représentations pertinentes de graphes attribués, mais ils n'ont pour l'instant été que très peu appliqués à des données spatiales. Les travaux existants se concentrent sur la prédiction de la position d'objets en mouvement (Liu *et al.* 2023), et la classification d'objets possédant une structure spatiale, comme des images (Bera *et al.* 2022). Dans le premier cas, les méthodes proposées dans la littérature requièrent des séries temporelles régulières, tandis que dans le second, la tâche est complètement différente de la nôtre, et porte qui plus est sur des données complètes. En comparaison, l'approche que nous proposons vise à exploiter des données incomplètes et statiques⁴, et constitue donc une contribution méthodologique.

La formalisation de la méthode d'extraction du graphe de confronts définie dans la thèse de M. Ferrand est en train de faire l'objet de deux soumissions de papiers : la première porte sur la définition de la méthode d'extraction du graphe de confronts (dans un but de reproductibilité de la méthode) et la seconde sur les erreurs révélées par l'utilisation des graphes (erreurs présentes dans les sources historiques elles-mêmes ou provoquées par leur extraction des données ou de leur saisie).

Ce projet doit permettre de continuer le travail amorcé dans la thèse de M. Ferrand et de produire des résultats innovants sur l'exploitation du graphe en lui-même, d'un point de vue tant informatique qu'historique. Il pourrait aboutir à la soumission de deux nouveaux articles, l'un sur la méthode de spatialisation des données manquantes à partir des réseaux de neurones et l'autre sur l'analyse de la composition sociale des communautés du graphe.

5. Dimension interdisciplinaire (champs disciplinaires associés) et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »

Ce projet fait appel à des méthodes et des outils empruntés aux sciences humaines et sociales, à la géomatique et à l'informatique. Il pose des problématiques plurielles, aussi bien disciplinaires qu'interdisciplinaires, en histoire et en informatique. Il s'inscrit dans plusieurs axes de la FR Agorantic et dans la continuité d'une thèse soutenue et financée par la FR. Le travail informatique reposera en grande partie sur Dariste Djimeli Tiona, étudiant de M2 à l'Université Nationale du Vietnam en double diplôme avec La Rochelle Université, et effectuant son stage de fin d'études au LIA.

6. Partenariats extérieurs envisagés

- Groupe de recherche Modelespace x Architerre (échange sur les différentes possibilités d'utiliser les graphes pour l'étude de la documentation foncière)
- Archives départementales du Vaucluse (Mise à disposition de la documentation foncière et des sources complémentaires)
- Archives municipales d'Avignon (Mise à disposition de la documentation foncière et des sources complémentaires)

⁴ Au sens où elles ne constituent pas des séries temporelles régulières et détaillées, à la différence des travaux cités.

Bibliographie

BERA A., WHARTON Z., LIU Y., BESSIS N., BEHERA A. « SR-GNN: Spatial Relation-Aware Graph Neural Network for Fine-Grained Image Categorization », *IEEE Transactions on Image Processing* 31:6017–6031, 2022.

FERRAND M., *Usages et représentations de l'espace urbain médiéval. Approche interdisciplinaire et exploration de données géo-historiques d'Avignon à la fin du Moyen Âge*, thèse de doctorat d'Histoire, Avignon Université, 2022.

FOSSIER R., *Polyptiques et censiers*, Turnhout, Brepols, 1978.

LIU J., CHEN Y., HUANG X., LI J., MIN G., « [GNN-based long and short term preference modeling for next-location prediction](#) », *Information Sciences* 629:1–14, 2023.

MONTPIED G., ROUAULT J., « Du texte au graphe : établissement d'une carte du parcellaire à partir des données de deux cadastres de la fin du Moyen Âge », dans *Les cadastres anciens des villes et leur traitement par l'informatique. Actes de la table ronde de Saint-Cloud, 31 janvier - 2 février 1985*, Rome, École Française de Rome, 120-1, 1989, p. 359-377.

MORICEAU J.-M., BRUNEL G., GUYOTJEANNIN O. (dirs.), *Terriers et plans- terriers du XIII au XVIII siècle, Actes du colloque de Paris (23-25 septembre 1998)*, Paris, École nationale des chartes, 2002.

PENET H., « Le sens des limites. Construction et perception de l'espace dans les actes de la pratique : l'exemple sicilien (XII-XVsiècle) », dans *Construction de l'espace au Moyen Âge : pratiques et représentation, Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public, 37 congrès*, Mulhouse, 2006, p. 405-411.

RODIER X., LE COUEDIC M., JOUVE B., HAUTEFEUILLE F., LETURCQ S., FIEUX E., « De l'espace aux graphes. Mesurer les dynamiques spatiales des terroirs villageois, dans *Mesure et histoire médiévale : XLII Congrès de la SHMESP (Tours, 31 mai-2 juin 2012)*, Paris, Éditions de la Sorbonne, 2013, p. 99-118.

| Budget (€)* | | |
|--|--|--------------|
| | Brève description | Montant |
| Missions | Pas de frais engagé pour les missions aux archives (dépôt à Avignon ou consultation des versions numérisées pour les Archives Vaticanes) | |
| Consommables, petits matériels** | Emprunt de l'appareil photo Agorantic pour certaines missions aux Archives | |
| Organisation de réunions | Réunion essentiellement sur place, pas de frais de déplacement à prévoir | |
| Stages***, vacations, prestations | - Stagiaire M2 en informatique (6 mois) | 3 480€ |
| | - Ingénieur de Recherche en histoire médiévale/humanités numériques (3 mois mi-temps) | 7 500€ |
| Budget total | | 10 980€ |
| Co financements le cas échéant | LIA (financement du stage de M2) | 3 480€ |
| Budget demandé à Agorantic | | 7500€ |
| Recettes extérieures | | |