

Projet HistoGraphe : Approximation de la distance spatiale au moyen d'un réseau de confronts

Application à la segmentation d'Avignon au Moyen-Âge

Margot Ferrand^{1,2} [Vincent Labatut](mailto:vincent.labatut@univ-avignon.fr)³

¹CIHAM UMR 5648 – margot.ferrand@alumni.univ-avignon.fr

²Ville d'Avignon, Maison du Patrimoine

³LIA UPR 4128 – vincent.labatut@univ-avignon.fr

Symposium de la FR Agorantic

3 décembre 2024



Plan de la présentation

- 1 Contexte historique
- 2 Méthodologie
- 3 Comparaison des méthodes
- 4 Partitionnement du réseau
- 5 Conclusion

Contexte historique

Notion de terrier

- Registre foncier décrivant les biens relevant d'un seigneur foncier, les tenanciers, obligations et redevances associés



- Une des nombreuses sources historiques utilisables pour étudier l'espace urbain
- Particulièrement pertinents pour la période du Moyen Âge
- Usage se généralisant à Avignon après l'arrivée des papes (XIII–XIV^e)

Contexte historique

Limites et difficultés

- ① Taille et **incomplétude** des données
 - Grand nombre de seigneuries, chacune décrite par un terrier dédié
 - Tous les terriers ne sont pas complets ou même disponibles
 - ② **Hétérogénéité** des données
 - Structure varie d'un terrier à l'autre, voire au sein du même terrier
 - Plusieurs langues utilisées (Latin, Provençal), orthographe non-normalisé → variabilité de la terminologie et même des noms propres
 - ③ **Incertitude** géographique
 - Pas de notion d'adresse, noms de rue informels et donc variables
 - Pas de cadastre à cette époque (pas avant le XVI^e)
 - Difficile, voire impossible d'appliquer l'approche standard consistant à positionner les biens spatiaux
- **Objectif** : proposer une méthode permettant de tirer parti de la localisation relative disponible, afin de partitionner l'espace urbain

Contexte historique

Limites et difficultés

- ① Taille et **incomplétude** des données
 - Grand nombre de seigneries, chacune décrite par un terrier dédié
 - Tous les terriers ne sont pas complets ou même disponibles
 - ② **Hétérogénéité** des données
 - Structure varie d'un terrier à l'autre, voire au sein du même terrier
 - Plusieurs langues utilisées (Latin, Provençal), orthographe non-normalisé → variabilité de la terminologie et même des noms propres
 - ③ **Incertitude** géographique
 - Pas de notion d'adresse, noms de rue informels et donc variables
 - Pas de cadastre à cette époque (pas avant le XVI^e)
 - Difficile, voire impossible d'appliquer l'approche standard consistant à positionner les biens spatiaux
- **Objectif** : proposer une méthode permettant de tirer parti de la localisation relative disponible, afin de partitionner l'espace urbain

Contexte historique

Limites et difficultés

- ① Taille et **incomplétude** des données
 - Grand nombre de seigneries, chacune décrite par un terrier dédié
 - Tous les terriers ne sont pas complets ou même disponibles
 - ② **Hétérogénéité** des données
 - Structure varie d'un terrier à l'autre, voire au sein du même terrier
 - Plusieurs langues utilisées (Latin, Provençal), orthographe non-normalisé → variabilité de la terminologie et même des noms propres
 - ③ **Incertitude** géographique
 - Pas de notion d'adresse, noms de rue informels et donc variables
 - Pas de cadastre à cette époque (pas avant le XVI^e)
 - Difficile, voire impossible d'appliquer l'approche standard consistant à positionner les biens spatiaux
- **Objectif** : proposer une méthode permettant de tirer parti de la localisation relative disponible, afin de partitionner l'espace urbain

Contexte historique

Limites et difficultés

- ① Taille et **incomplétude** des données
 - Grand nombre de seigneuries, chacune décrite par un terrier dédié
 - Tous les terriers ne sont pas complets ou même disponibles
 - ② **Hétérogénéité** des données
 - Structure varie d'un terrier à l'autre, voire au sein du même terrier
 - Plusieurs langues utilisées (Latin, Provençal), orthographe non-normalisé → variabilité de la terminologie et même des noms propres
 - ③ **Incertitude** géographique
 - Pas de notion d'adresse, noms de rue informels et donc variables
 - Pas de cadastre à cette époque (pas avant le XVI^e)
 - Difficile, voire impossible d'appliquer l'approche standard consistant à positionner les biens spatiaux
- **Objectif** : proposer une méthode permettant de tirer parti de la localisation relative disponible, afin de partitionner l'espace urbain

Méthodologie

Base de données géographique

- Extraction des informations [Fer22]
 - Sources : 8 terriers, comptes des clavares de la ville
 - Numérisation, OCR, NER, intervention manuelle significative
- Objets spatiaux
 - 9 types : biens (83 %), rues (9 %), édifices (4 %), bourgs (2 %), portes, livrées, paroisses/quartiers, remparts, repères géologiques
 - 3 dimensions : objets ponctuels, linéaires, surfaciques
 - 42 relations spatiales, normalisées en : près de, N/S/E/O de, intérieur de
- Données supplémentaires
 - Géolocalisation de 68 % des biens et 70 % des invariants
 - Relations spatiales rues–rues, rues–invariants
 - Segmentation spatiale des objets non-ponctuels

Méthodologie

Base de données géographique

- Extraction des informations [Fer22]
 - Sources : 8 terriers, comptes des clavaires de la ville
 - Numérisation, OCR, NER, intervention manuelle significative
- Objets spatiaux
 - 9 types : biens (83 %), rues (9 %), édifices (4 %), bourgs (2 %), portes, livrées, paroisses/quartiers, remparts, repères géologiques
 - 3 dimensions : objets ponctuels, linéaires, surfaciques
 - 42 relations spatiales, normalisées en : près de, N/S/E/O de, intérieur de
- Données supplémentaires
 - Géolocalisation de 68 % des biens et 70 % des invariants
 - Relations spatiales rues–rues, rues–invariants
 - Segmentation spatiale des objets non-ponctuels

Méthodologie

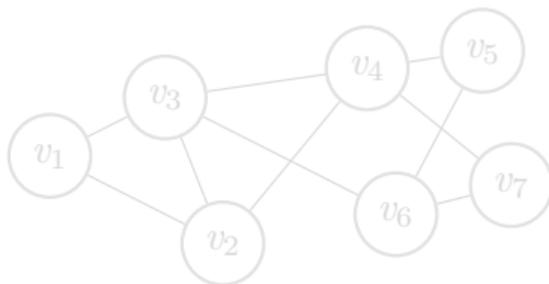
Base de données géographique

- Extraction des informations [Fer22]
 - Sources : 8 terriers, comptes des clavares de la ville
 - Numérisation, OCR, NER, intervention manuelle significative
- Objets spatiaux
 - 9 types : biens (83 %), rues (9 %), édifices (4 %), bourgs (2 %), portes, livrées, paroisses/quartiers, remparts, repères géologiques
 - 3 dimensions : objets ponctuels, linéaires, surfaciques
 - 42 relations spatiales, normalisées en : près de, N/S/E/O de, intérieur de
- Données supplémentaires
 - Géolocalisation de 68 % des biens et 70 % des invariants
 - Relations spatiales rues–rues, rues–invariants
 - Segmentation spatiale des objets non-ponctuels

Méthodologie

Modélisation à base de graphes

- Information disponible :
 - Objets spatiaux & leurs relations spatiales
 - Positions absolues renseignées/estimées de façon incomplète
- Modèle à base de graphes :
 - Particulièrement adaptés à la représentation de données relationnelles
 - Traitement possible même sans connaître toutes positions absolues

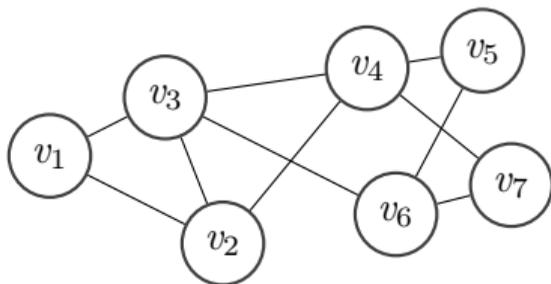


- Notre approche :
 - Sommets = objets spatiaux; arêtes = positionnements relatifs
 - Partitionner le graphe = segmenter l'espace urbain

Méthodologie

Modélisation à base de graphes

- Information disponible :
 - Objets spatiaux & leurs relations spatiales
 - Positions absolues renseignées/estimées de façon incomplète
- Modèle à base de graphes :
 - Particulièrement adaptés à la représentation de données relationnelles
 - Traitement possible même sans connaître toutes positions absolues

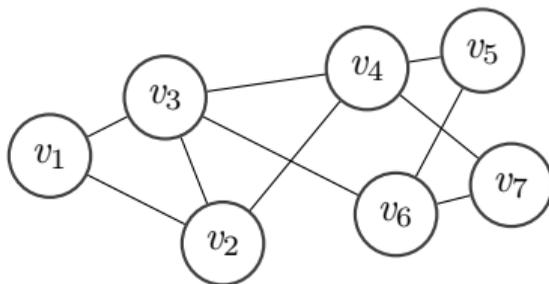


- Notre approche :
 - Sommets = objets spatiaux; arêtes = positionnements relatifs
 - Partitionner le graphe = segmenter l'espace urbain

Méthodologie

Modélisation à base de graphes

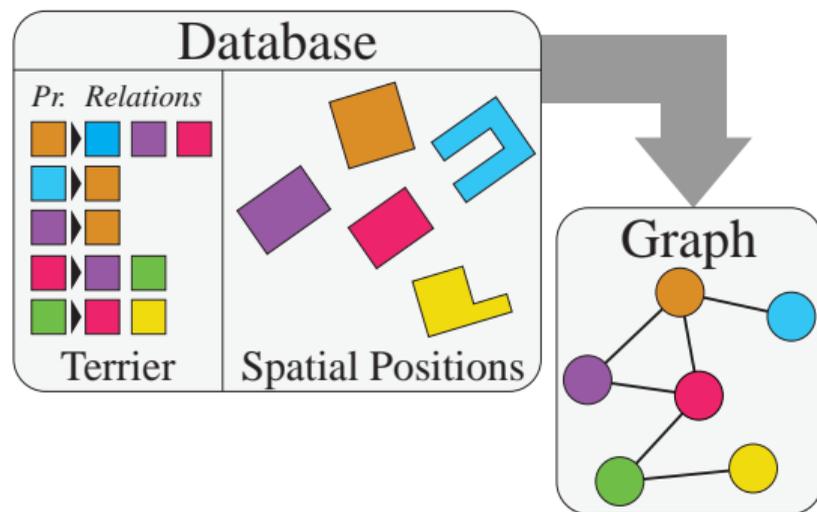
- Information disponible :
 - Objets spatiaux & leurs relations spatiales
 - Positions absolues renseignées/estimées de façon incomplète
- Modèle à base de graphes :
 - Particulièrement adaptés à la représentation de données relationnelles
 - Traitement possible même sans connaître toutes positions absolues



- Notre approche :
 - Sommets = objets spatiaux ; arêtes = positionnements relatifs
 - Partitionner le graphe = segmenter l'espace urbain

Modélisation à base de graphes

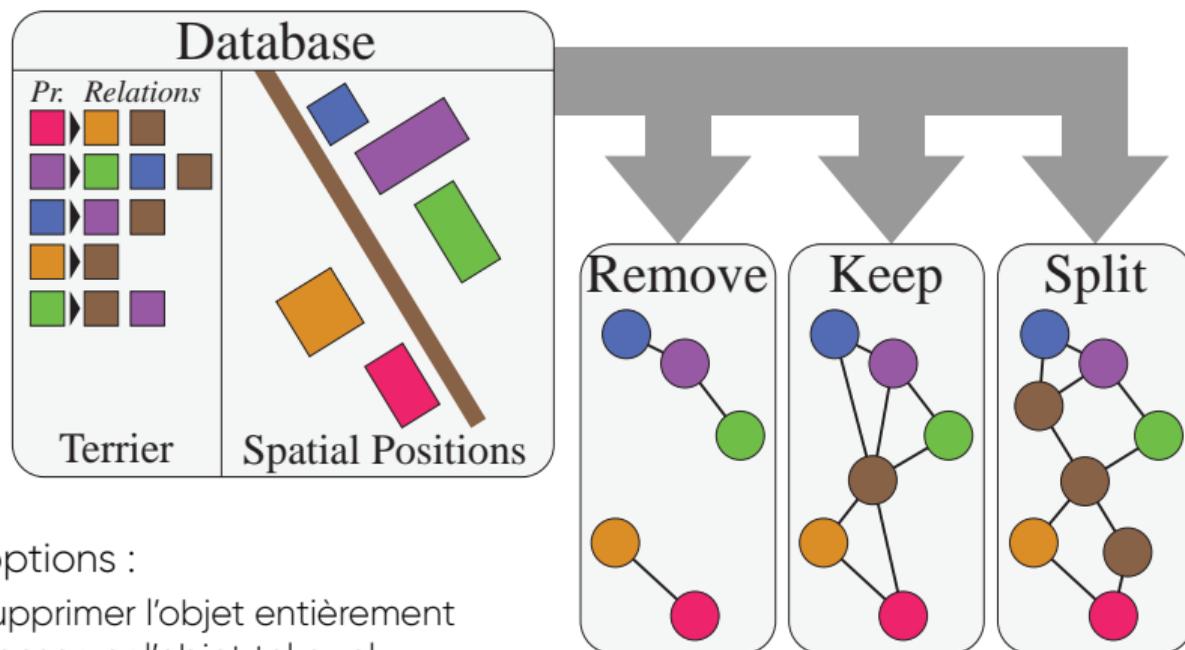
Extraction directe



- Chaque objet mentionné (bien ou confront) est représenté par un sommet
- Toutes les relations spatiales sont représentées par des arêtes
- Étape de nettoyage, incluant la suppression des petits composants

Modélisation à base de graphes

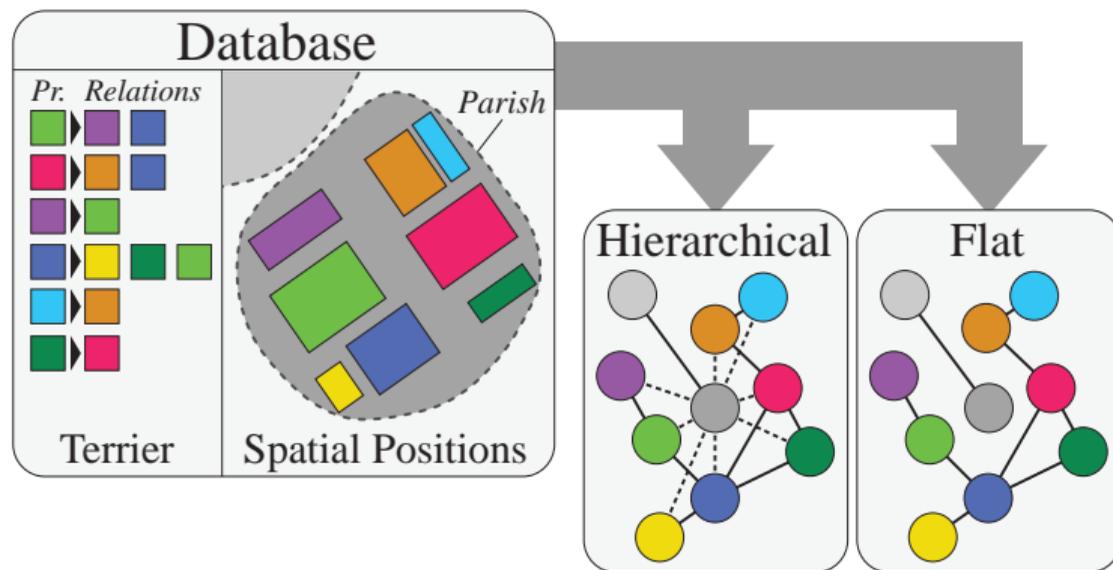
Objets non-ponctuels



- Trois options :
 - Supprimer l'objet entièrement
 - Conserver l'objet tel quel
 - Diviser l'objet

Modélisation à base de graphes

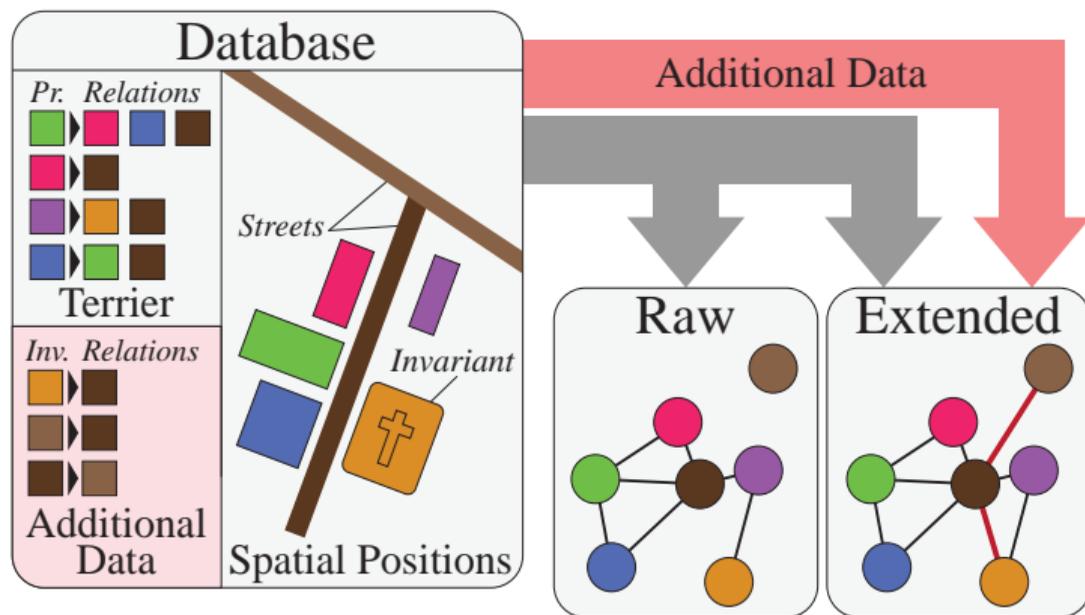
Relations hiérarchiques



- Deux options :
 - Conserver la relation telle quelle
 - Supprimer la relation intégralement

Modélisation à base de graphes

Relations supplémentaires



- Deux options :
 - N'utiliser que les données primaires
 - Exploiter en plus les données additionnelles

Comparaison des méthodes

Couverture vs. fidélité

- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



Comparaison des méthodes

Couverture vs. fidélité

- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



Comparaison des méthodes

Couverture vs. fidélité

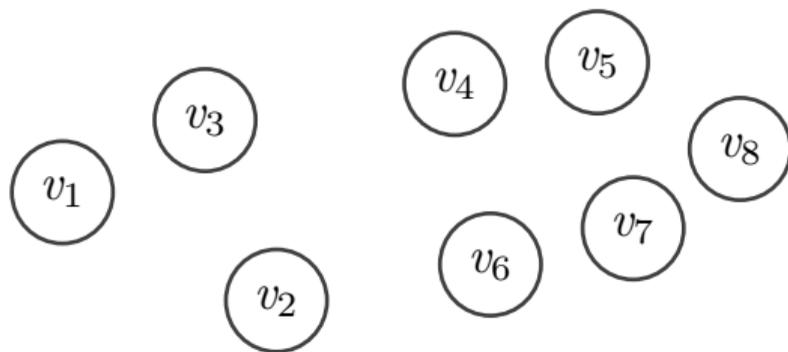
- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



Comparaison des méthodes

Couverture vs. fidélité

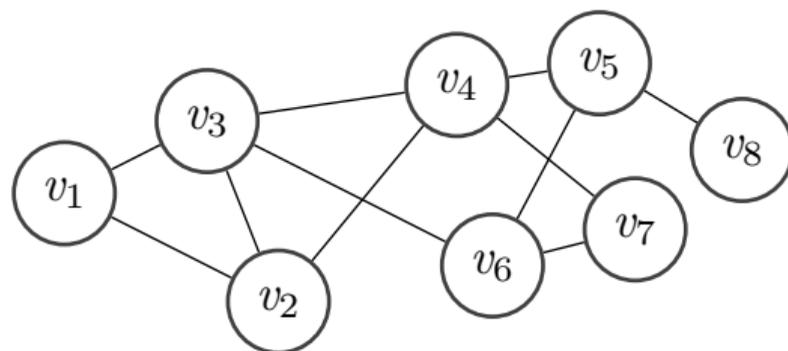
- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



Comparaison des méthodes

Couverture vs. fidélité

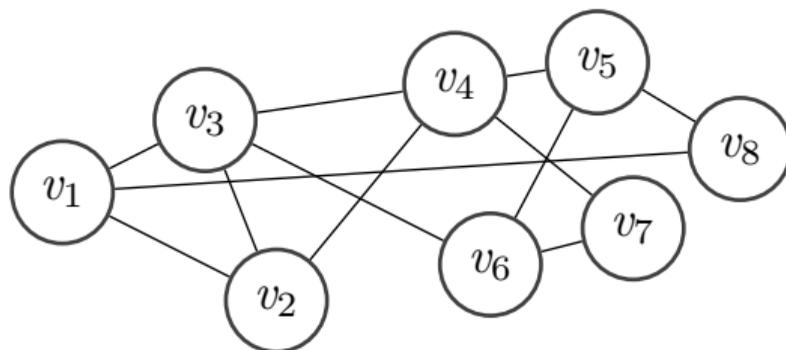
- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



Comparaison des méthodes

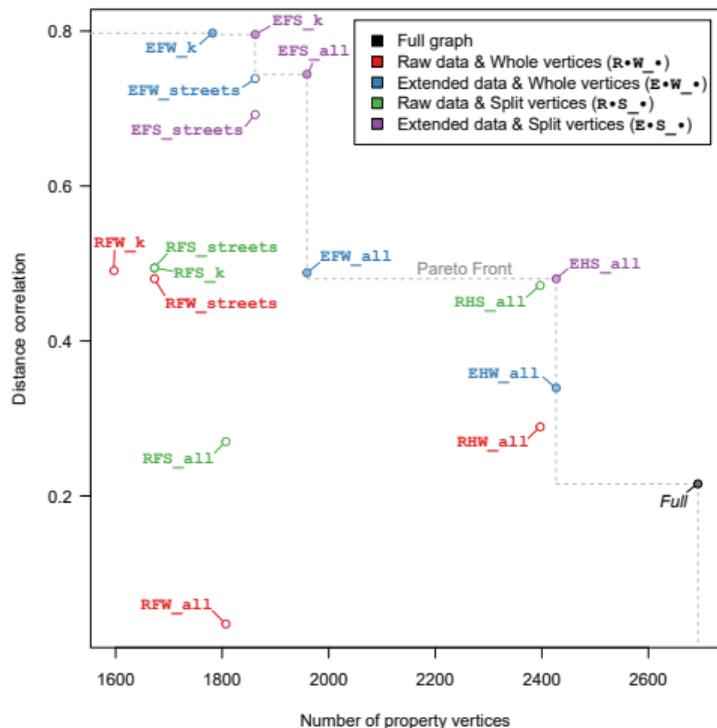
Couverture vs. fidélité

- 16 variantes de la méthode d'extraction des graphes
- Comparaison sur deux aspects :
 - Couverture : conserver autant que possible les biens décrits dans les données brutes
 - Fidélité : représenter aussi bien que possible les distances spatiales entre objets
- Formalisation : maximisation des deux critères suivants
 - Nombre de biens représentés dans le graphe
 - Corrélation entre distance spatiale et distance géodésique



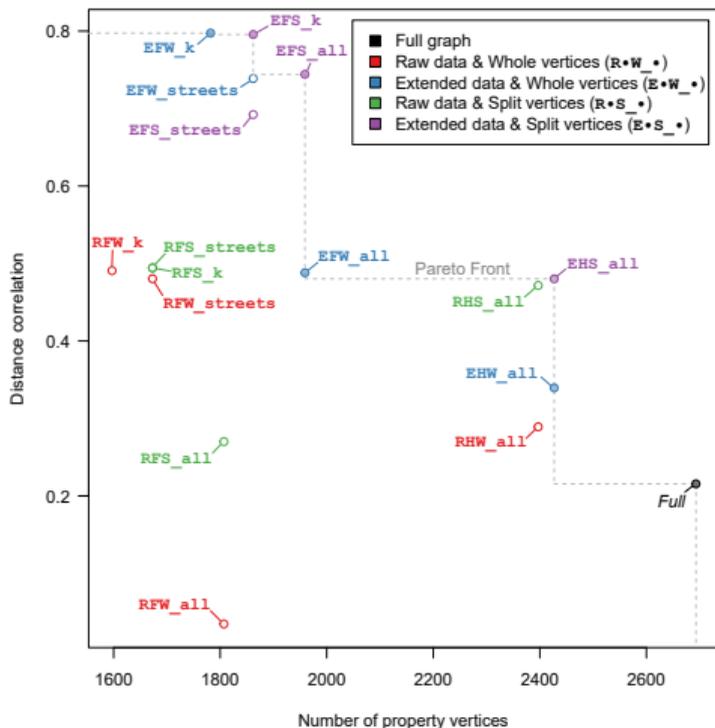
Comparaison des méthodes

Sélection de la méthode la plus appropriée



Comparaison des méthodes

Sélection de la méthode la plus appropriée



• Observations générales

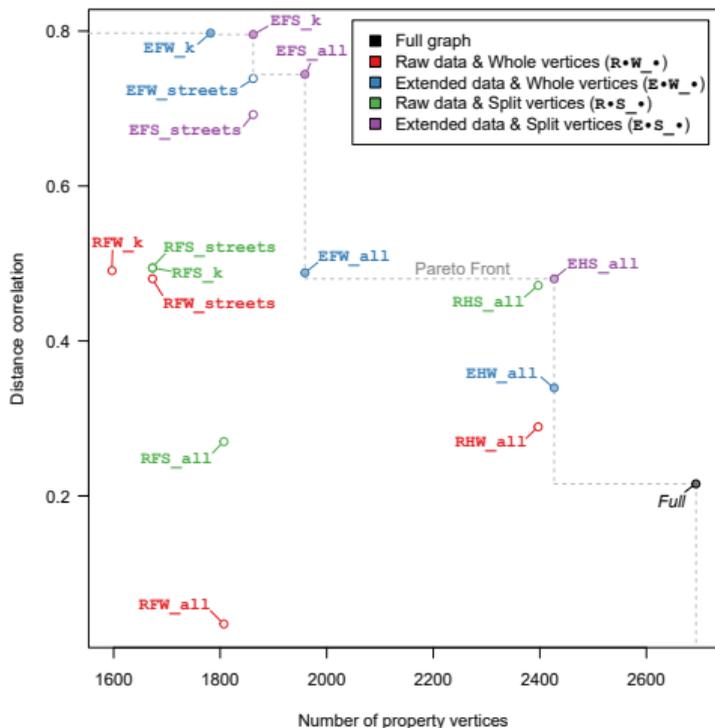
- 1 Relations hiérarchiques : $F \downarrow$
- 2 Relations supplémentaires $C \uparrow$ & $F \uparrow$
- 3 Suppr. non-ponctuels $F \uparrow$ & $C \sim$ (si rues)
- 4 Div. au lieu de suppr. : $F \uparrow$ & $C \sim$

• Meilleures méthodes

- Globalement : EFS_k
- Sans division : EFW_k
- Sans relations supplémentaires : RHS_all
- Données brutes : RFW_streets

Comparaison des méthodes

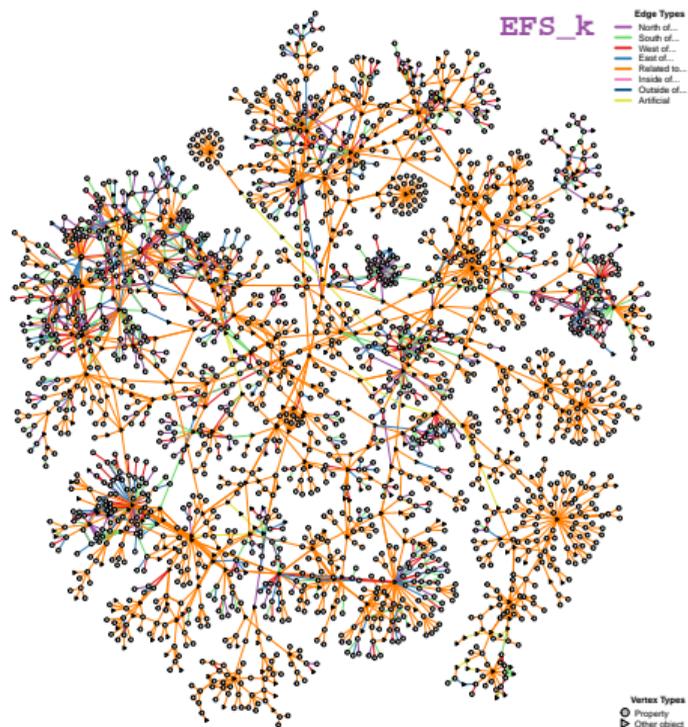
Sélection de la méthode la plus appropriée



- Observations générales
 - 1 Relations hiérarchiques : $F \downarrow$
 - 2 Relations supplémentaires $C \uparrow$ & $F \uparrow$
 - 3 Suppr. non-ponctuels $F \uparrow$ & $C \sim$ (si rues)
 - 4 Div. au lieu de suppr. : $F \uparrow$ & $C \sim$
- Meilleures méthodes
 - Globalement : EFS_k
 - Sans division : EFW_k
 - Sans relations supplémentaires : RHS_all
 - Données brutes : RFW_streets

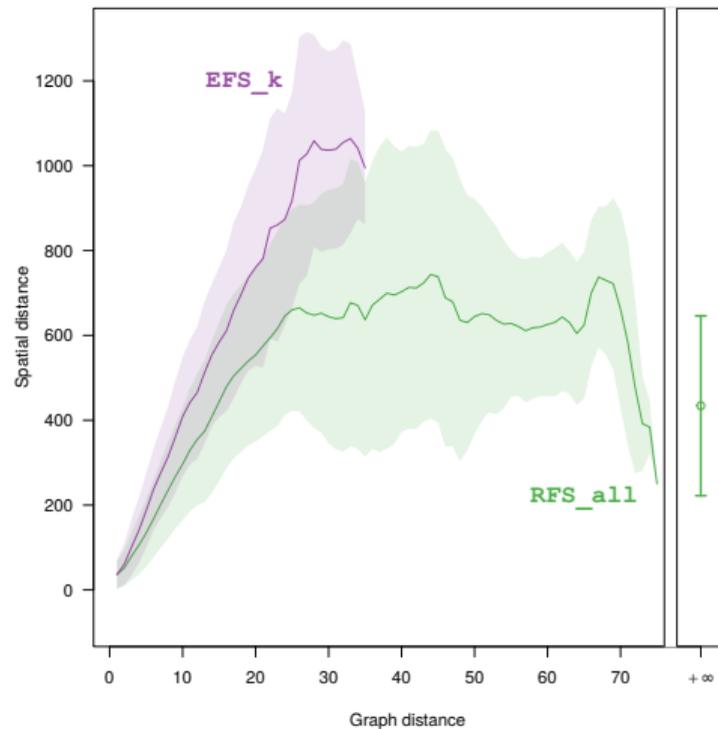
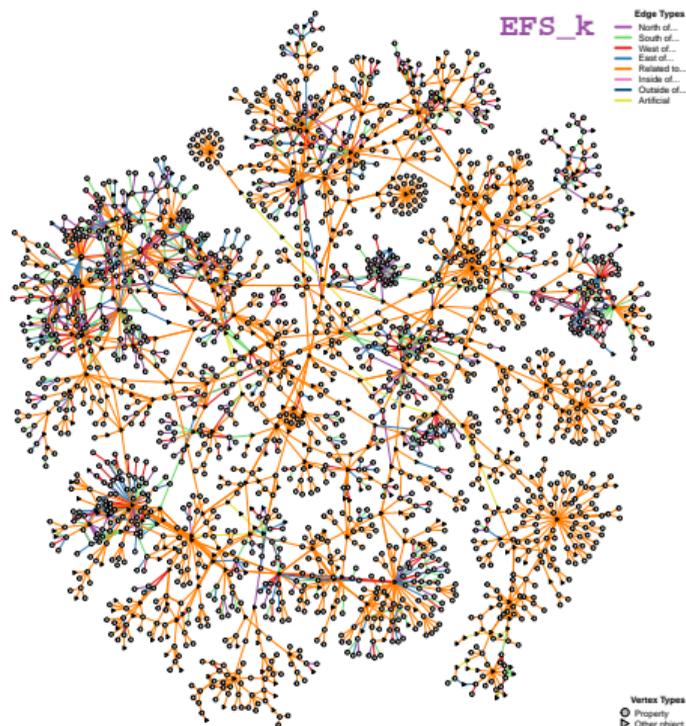
Comparaison des méthodes

Graphe retenu



Comparaison des méthodes

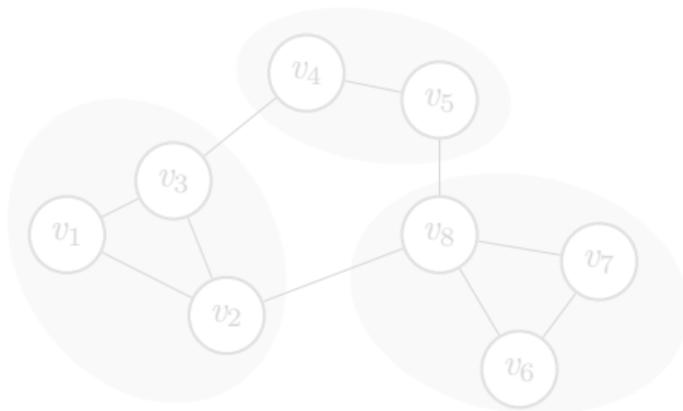
Grappe retenue



Partitionnement du réseau

Détection de communautés

- Communauté : sous-graphe relativement dense par rapport au reste du graphe
- Structure de communautés : partition de l'ensemble des sommets en communautés

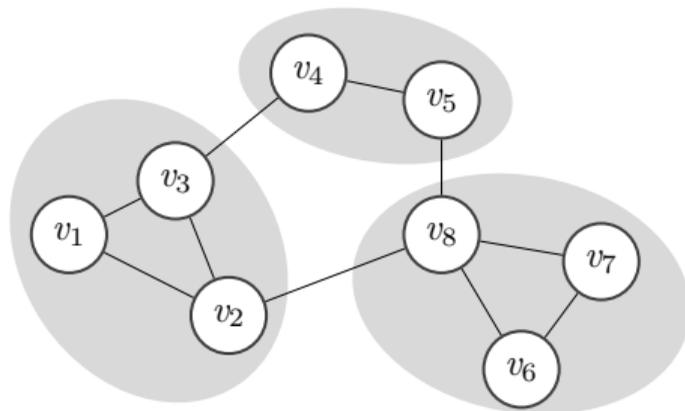


- Dans notre cas :
 - Le graphe encode une approximation de la distance spatiale
 - Sa structure de communautés approxime une partition de l'espace urbain
 - Zones spatiales représentatives de la perception de l'espace par les habitants
 - Méthode utilisée : Louvain [Blo+08]

Partitionnement du réseau

Détection de communautés

- Communauté : sous-graphe relativement dense par rapport au reste du graphe
- Structure de communautés : partition de l'ensemble des sommets en communautés

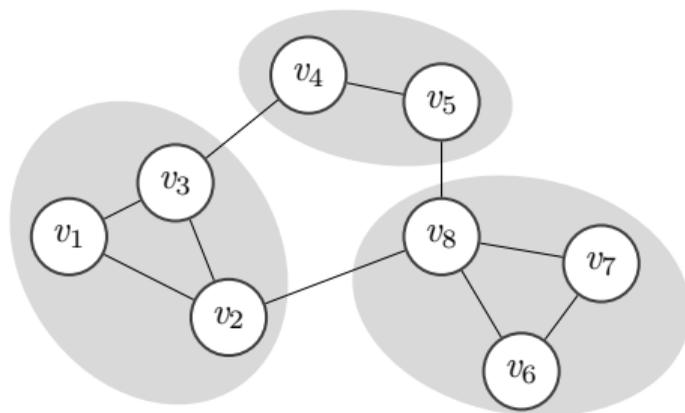


- Dans notre cas :
 - Le graphe encode une approximation de la distance spatiale
 - Sa structure de communautés approxime une partition de l'espace urbain
 - Zones spatiales représentatives de la perception de l'espace par les habitants
 - Méthode utilisée : Louvain [Blo+08]

Partitionnement du réseau

Détection de communautés

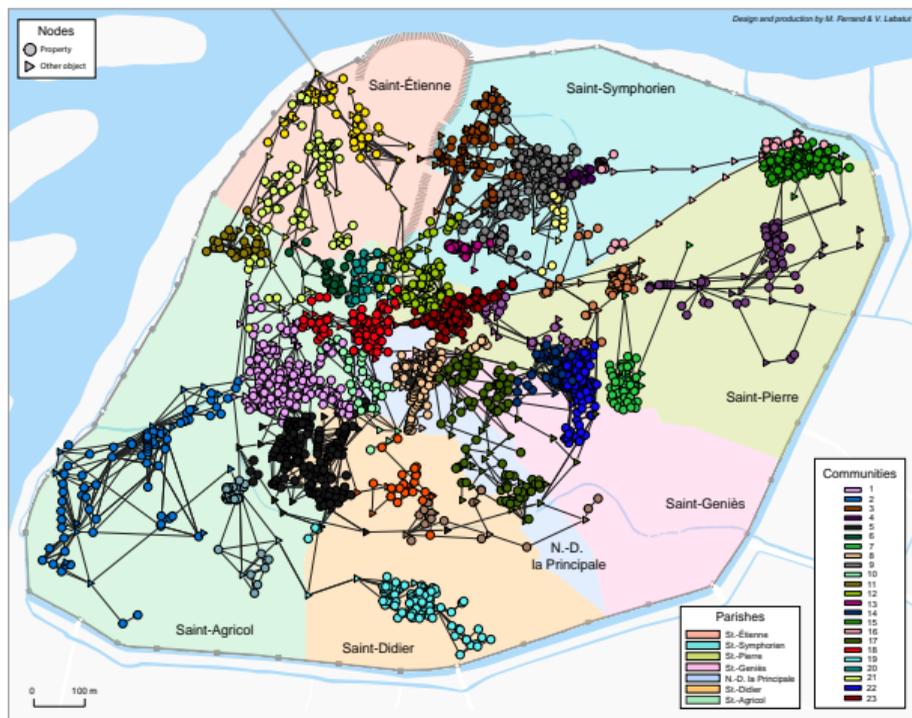
- Communauté : sous-graphe relativement dense par rapport au reste du graphe
- Structure de communautés : partition de l'ensemble des sommets en communautés



- Dans notre cas :
 - Le graphe encode une approximation de la distance spatiale
 - Sa structure de communautés approxime une partition de l'espace urbain
 - Zones spatiales représentatives de la perception de l'espace par les habitants
 - Méthode utilisée : Louvain [Blo+08]

Partitionnement du réseau

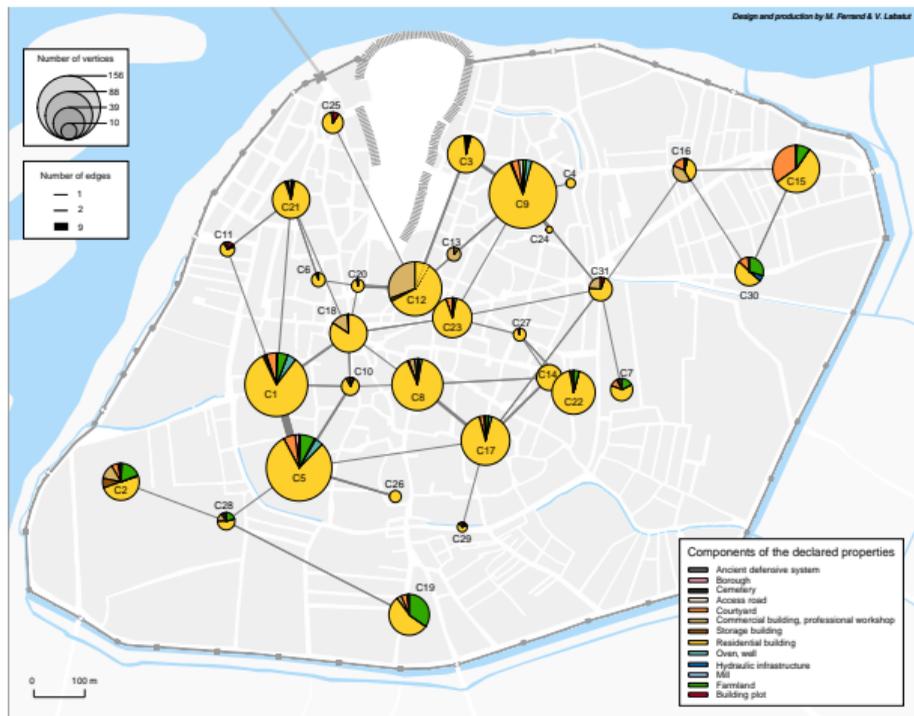
Communautés vs. paroisses



- Pas de correspondance biunivoque avec les paroisses
 - Plusieurs communautés par paroisses (toutes)
 - Plusieurs paroisses par communautés (ex. 5 pr C18)
- Justification :
 - Frontières mal connues des habitants
 - Pas de réalité physique
- Perception de l'espace urbain plutôt contrainte par
 - Obstacles : remparts, canaux, rocher des Doms...
 - Rues anciennes / structurantes

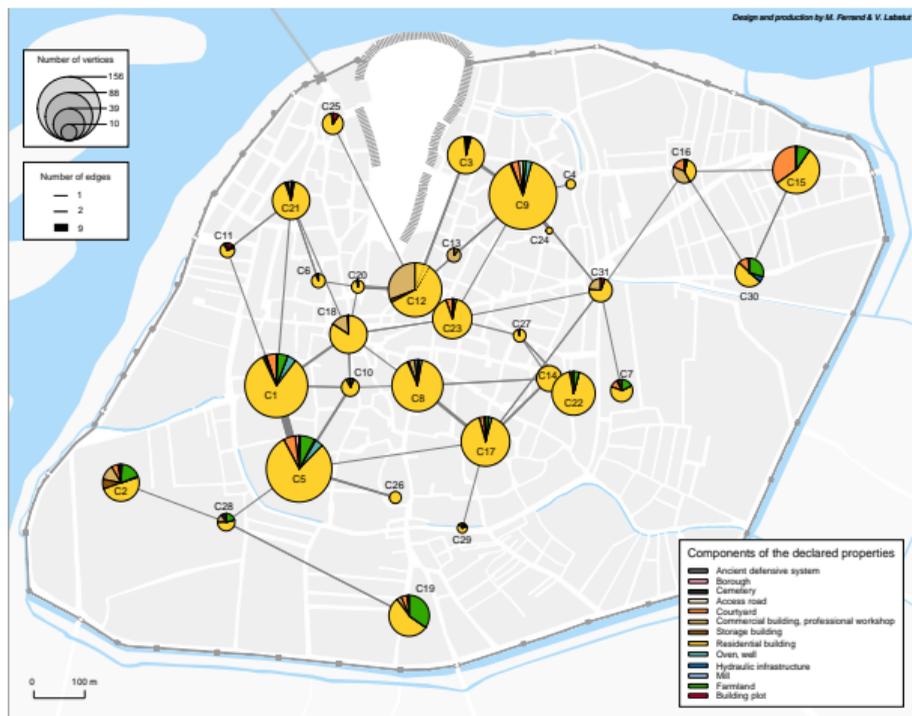
Partitionnement du réseau

Types de communautés



Partitionnement du réseau

Types de communautés

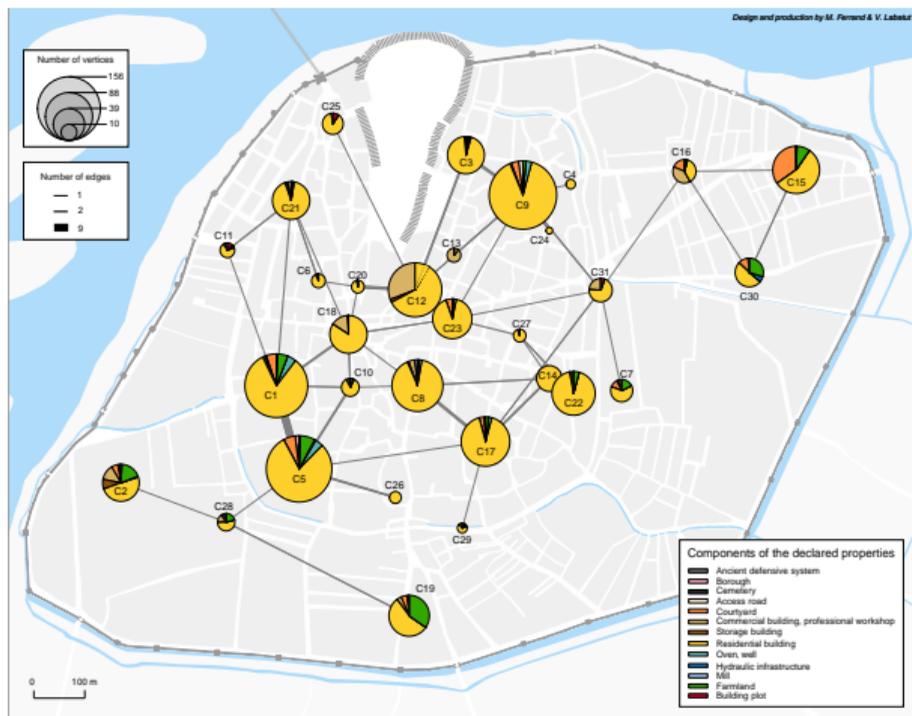


Dans les vieux murs

- Structurées autour de
 - Rues antiques, ex. C8
 - Résidence fameuse, ex. C1 (Tymburge Vayrane)
- Composées principalement de
 - Services, ex. C21 (hotellerie/bains)
 - Résidences, ex. C5 (vieilles familles)
 - Commerces, ex. C18 (marché, étals)
 - Sociabilisation, ex. C13 (cimetière)

Partitionnement du réseau

Types de communautés



Hors des vieux murs

- Coms semi-rurales, ex. C2
 - Localisées rel. bourgs
 - Compo. : jardins, dépendances de l'intra
- Bourgs, ex. C15
 - Struct. autour rue principale
 - Composition homogène, planifiée

Conclusion

- Contributions
 - Élaboration d'une base de données historique¹
 - Définition de plusieurs méthodes d'extraction de graphes²
 - Proposition d'une méthode d'évaluation / comparaison
 - Application à la segmentation urbaine d'Avignon au Moyen-Âge
- Travail publié : M. Ferrand et V. Labatut. « Approximating Spatial Distance Through Confront Networks : Application to the Segmentation of Medieval Avignon ». In : *Journal of Complex Networks* in press (2024). doi : [10.1093/comnet/cnae046](https://doi.org/10.1093/comnet/cnae046). ([hal-04786705](https://hal.archives-ouvertes.fr/hal-04786705))
- Perspectives
 - Approfondissement de l'étude des communautés identifiées
 - Adaptation des méthodes d'extraction à d'autres situations (Orléans)
 - Méthodes d'approximation des positions (absolues) manquantes (GNN)

1. <https://doi.org/10.5281/zenodo.12804378>

2. <https://github.com/compNet/MedievalAvignon/>

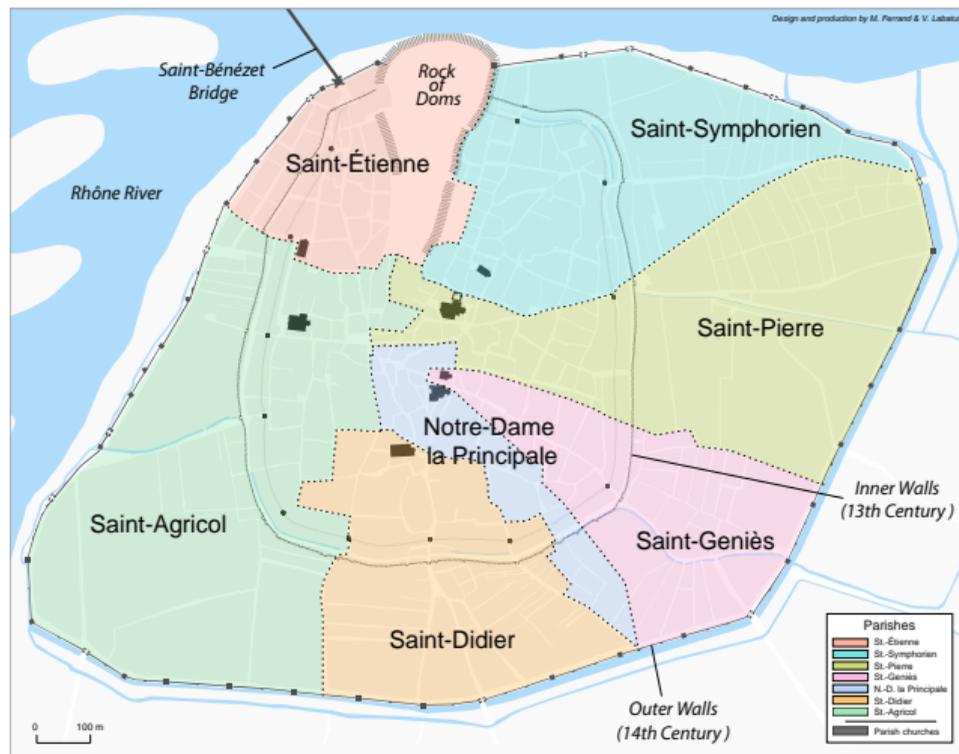
Questions ?

Section 6

Annexes

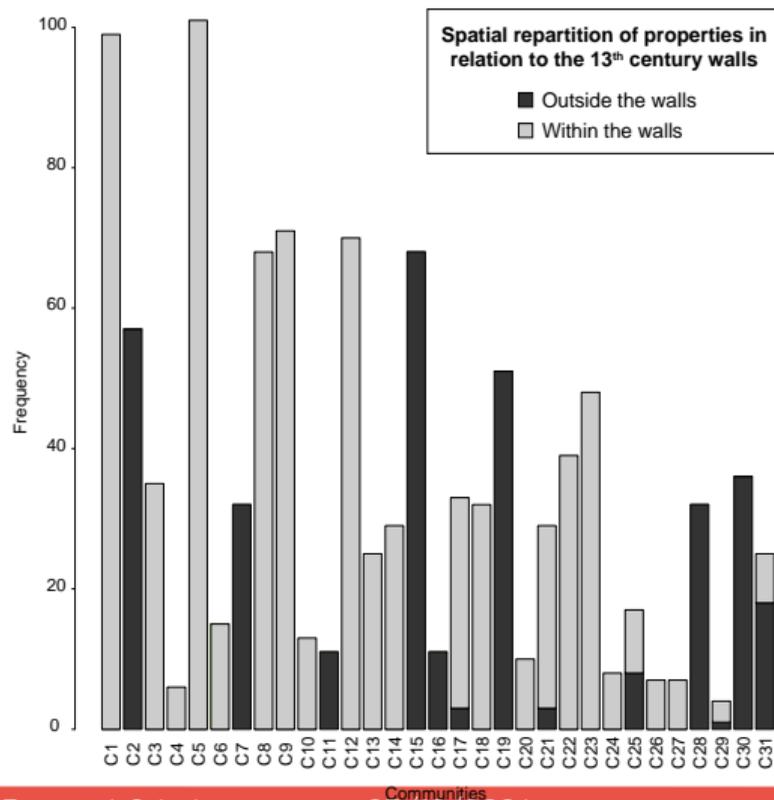
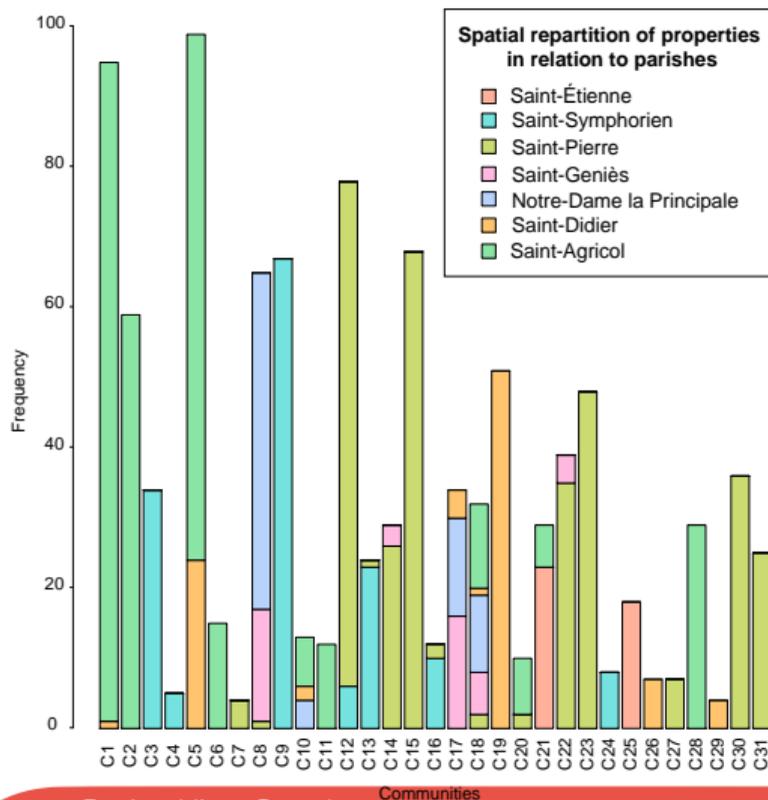
Carte d'Avignon

Paroisses historiques



Propriétés des communautés

Composition en termes de paroisses / anciens remparts



Section 7

Références bibliographiques

Références bibliographiques I

- [Blo+08] V. D. Blondel, J.-L. Guillaume, R. Lambiotte et E. Lefebvre. « Fast unfolding of communities in large networks ». In : *Journal of Statistical Mechanics* 2008.10 (2008), P10008. doi : [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008).
- [Fer22] M. Ferrand. « Usages et représentations de l'espace urbain médiéval : Approche interdisciplinaire et exploration de données géo-historiques d'Avignon à la fin du Moyen Âge ». PhD Thesis. Avignon Université, 2022. url : <https://www.theses.fr/2022AVIG1002>.
- [FL24] M. Ferrand et V. Labatut. « Approximating Spatial Distance Through Confront Networks : Application to the Segmentation of Medieval Avignon ». In : *Journal of Complex Networks* in press (2024). doi : [10.1093/comnet/cnae046](https://doi.org/10.1093/comnet/cnae046). [hal-04786705](https://hal.archives-ouvertes.fr/hal-04786705).