

# Comparaison des chemins de Hilbert adaptatif et des graphes de voisinage pour la caractérisation d'un parcellaire agricole

Thomas Guyet\*, Sébastien da Silva\*\*,\*\*\*, Claire Lavigne\*\*\*, Florence Le Ber\*\*\*\*

\* Agrocampus Ouest, Rennes

\*\* LORIA – INRIA Grand Est, Villers-lès-Nancy

\*\*\* INRA PSH, Avignon

\*\*\*\* ICUBE, Université de Strasbourg/ENGEES - CNRS, Strasbourg

**Résumé.** Cet article compare deux représentations de données spatiales, les graphes de voisinages et les chemins de Hilbert-Peano, utilisées par des algorithmes de fouille. Cette comparaison s'appuie sur la mise en œuvre d'une méthode d'énumération de « sacs de nœuds », qui permet d'obtenir des caractérisations homogènes à partir des deux représentations. La méthode est appliquée à la caractérisation de parcellaires agricoles et les résultats tendent à montrer que la linéarisation de l'espace capte la majorité de l'information, à l'exception des éléments rares, sur cet exemple particulier.

## 1 Introduction

Les méthodes de recherche de motifs spatiaux sont utilisées couramment pour construire des caractérisations de données spatiales (voir Selmaoui-Folcher et al. (2013)). Ces approches s'appuient sur des représentations de l'espace tels que des graphes de voisinage ou des chemins construits sur des courbes fractales (par ex. chemins de Hilbert-Peano, voir Mari et Le Ber (2006)). Les graphes de voisinages contiennent une information spatiale riche, mais ils sont plus complexes à fouiller, tandis que les chemins sont faciles à fouiller, mais ils réduisent l'information spatiale disponible.

Dans ce travail, nous confrontons ces deux représentations vis-à-vis de la caractérisation d'un parcellaire agricole. En particulier, nous cherchons à savoir si l'approximation par un chemin fractal permet de conserver une bonne caractérisation de l'organisation spatiale des parcelles agricoles en vue de l'application d'une méthode de recherche de motifs. Cette question soulève deux difficultés. La première est l'absence d'étiquetage des données qui permettrait d'évaluer une représentation sur une tâche de classification. Nous nous plaçons donc dans un contexte non-supervisé. La deuxième difficulté porte sur la comparaison des caractérisations obtenues, qui sont de natures différentes. Nous ne pouvons pas nous appuyer comme classiquement sur des calculs de corrélation (par ex. par une matrice de confusion) entre les localisations des motifs dans les chemins et dans les graphes.

Pour résoudre ces difficultés, nous utilisons des « sacs de nœuds », inspirés des sacs de mots introduits dans le contexte de l'analyse de texte (Salton et al. (1975)). Ce modèle permet de construire des caractérisations homogènes d'un même espace pour les graphes et les chemins.