

Titre : Exploration relationnelle d'un corpus d'actes notariés médiévaux.

Résumé : Cette communication montre comment des outils mathématiques et informatiques, initialement développés dans le contexte de l'analyse des réseaux sociaux, peuvent aider à comprendre les principales caractéristiques d'une grande base de données documentaire et être une aide précieuse à l'extraction d'information et à la désambiguïsation. Nous travaillons sur un grand corpus de documents médiévaux, initialement rédigés entre le XIIIème et le XVème siècle et issus d'une seigneurie commune. Ces documents sont de nature homogène : ils décrivent des transactions relatives au système de rente de la seigneurie (baux à fief, héritages, transferts de baux...). Ils ont été retranscrits au XVIIème siècle par un feudiste mandé par le nouveau propriétaire qui cherchait à faire valoir ses droits sur les terres. Le corpus peut être considéré comme très représentatif pour la seigneurie concernée.

Les documents ont été retranscrits dans une base de données accessible librement à <http://graphcomp.univ-tlse2.fr>. Un grand réseau d'interaction est extrait de la base : il décrit l'implication active des personnes cités dans les transactions. Les sommets du réseau sont donc d'une part des personnes, d'autre part des transactions. Les outils de visualisation de réseaux, couplés à des méthodes de classification des sommets, permettent d'obtenir une vue globale de ces relations et d'identifier des individus importants associés à certaines structures locales du réseau. Ces outils interactifs aident l'historien à s'approprier la masse documentaire pour en comprendre les principales caractéristiques. En complément, nous montrons comment le calcul de caractéristiques numériques liées à la structure du réseau permet d'identifier de manière automatique des problèmes de retranscription des documents qui auraient été difficilement mis en évidence autrement ; en particulier, l'extraction de réseaux locaux, combinée à des informations supplémentaires obtenues par le biais des documents (dates / lieux ...) fournit des outils simples et rapide pour la désambiguïsation des noms de personnes.

Auteurs :

Nathalie Villa-Vialaneix est maîtresse de conférences en statistique, spécialisée en data mining. Plus précisément, son domaine d'expertise recouvre l'analyse de données de grande dimension et l'analyse de réseaux / graphes avec des aspects touchants à la classification et à la visualisation. Elle a publié 15 articles dans des revues à comité de lecture.

Fabrice Rossi est professeur en mathématiques appliquées à l'université Paris 1 Panthéon Sorbonne et chercheur de l'équipe SAMM. Ses recherches portent sur l'analyse des données et l'apprentissage statistique. Il travaille notamment sur la visualisation et la classification de données, en particulier pour les données relationnelles. Il a publié près de 100 articles donc 19 dans des revues à comité de lecture.

Florent Hautefeuille est maître de conférences en archéologie médiévale, membre de l'UMR 5604 TRACES, et directeur de la MSHS-Toulouse. Ses recherches portent sur l'histoire des sociétés rurales dans le Sud Ouest de la France à la fin du Moyen Age. Elles s'articulent autour de plusieurs projets et intègrent autant les données écrites que les sources archéologiques. Il a publié 38 articles sur ces thématiques.