

Visualisation de données de prosopographie pour la reconstruction de carrières de personnages et de réseaux socio-professionnels

Nizar Messai, Thomas Devogele

LI, Université François Rabelais de Tours
{nizar.messai,thomas.devogele}@univ-tours.fr

Résumé. Dans cet article nous présentons deux approches de visualisation développées dans le cadre d'un projet collaboratif sur l'accès et l'exploitation des données prosopographiques de la Renaissance en France. L'objectif du projet est de modéliser et réaliser un portail sémantique assurant l'accès à différentes bases de données prosopographiques existantes afin de permettre une meilleure exploration et exploitation de ces données. Dans ce cadre, nous avons proposé deux interfaces de visualisation ProsoGraph et ProsoMap qui s'appuient respectivement sur la visualisation de graphes de réseaux sociaux et la visualisation de lieux géographiques et de trajectoires spatio-temporelles. Les deux interfaces communiquent avec le portail via une couche sémantique et lui offrent des fonctionnalités d'interrogation supplémentaires.

1 Introduction

La prosopographie est une science auxiliaire de l'histoire dont l'objectif est d'étudier les biographies des membres d'une catégorie spécifique de la société, en particulier leur origine, leur carrière, leurs déplacements et leur environnement familial. Il existe à nos jours plusieurs bases de données prosopographiques souvent peuplées à partir de fiches de description de personnes et à capacité d'interrogation réduite à l'extraction d'un ensemble de fiches. Le projet Personae propose d'intégrer un ensemble de bases de données prosopographiques de la Renaissance au sein d'un portail sémantique. Des mappings entre les ontologies locales de chaque base de données et l'ontologie du portail assurent l'accès à l'ensemble des concepts présents dans les bases reliées. L'interface d'interrogation principale de ce portail repose sur des formulaires d'interrogation classiques et les résultats retournés se présentent sous la forme de tableaux listant les réponses pertinentes dans les différentes bases. Malgré l'apport de cette solution sur le plan d'intégration sémantique et d'agrégation des données de différentes bases, elle demeure inefficace quand à la présentation de vue synthétique sur les données prosopographiques mettant en évidence les relations entre plusieurs personnages, les étapes "clés" dans la carrière d'un personnage, etc. Or, depuis plusieurs années le domaine du *Visual Analytics* n'a cessé d'apporter des réponses à ce genre de problème à travers différentes approches de visualisation de données permettant une exploration interactive et progressive de données Keim et al. (2008). Dans le cadre de ce projet, nous nous sommes appuyés sur des approches de

visualisation de graphes et de cartes géographiques pour proposer deux interfaces de visualisation adaptées à la nature des données prosopographiques étudiées et répondant aux besoins d'interrogation du portail mis en place. Les interfaces de visualisation proposées s'appuient d'une part sur des mesures de similarité et de proximité sémantique pour l'identification et la construction de réseaux personnels et professionnels à la Renaissance. D'autre part, elles exploitent les dimensions spatiales et temporelles très présentes en termes de prosopographie pour proposer des visualisations cartographiques retraçant la carrière d'une personne ou les flux de déplacements entre les différents lieux de l'époque. Le reste de l'article est organisé comme suit. Dans la section 2 nous présentons ProsoGraph, l'interface de visualisation de réseaux socio-professionnels. Ensuite, dans la section 3, nous présentons ProsoMap, l'interface de visualisation cartographique spatio-temporelle. Enfin, dans la section 4, nous développons les conclusions de ce travail et exposons les travaux en cours et futurs en lien avec notre projet.

2 ProsoGraph : Visualisation de graphes de réseaux socio-professionnels de la Renaissance

Le portail Personae intègre plusieurs bases de données de prosopographies de la Renaissance et offre ainsi une vue agrégée sur une grande quantité de données bibliographiques de personnages permettant d'aboutir à des descriptions potentiellement complètes de la carrière d'un personnage ainsi que du réseau socio-professionnel qu'il ait pu développer. Ces données s'apparentent naturellement à la définition de réseaux sociaux du fait de l'existence de personnes avec des relations et interactions entre elles. Partant de ce fait et dans le but d'assurer une meilleure exploration des données du portail, nous avons développé l'interface ProsoGraph qui permet de visualiser les données prosopographiques sous la forme de graphes où les noeuds correspondent à des personnages, des lieux ou des professions de l'époque et les arcs correspondent aux liens entre ces différents objets. ProsoGraph s'appuie sur l'API de visualisation de graphes Sigma JS (<http://sigmaj.js.org/>). L'interface communique avec le portail à travers des requêtes SPARQL utilisant l'ontologie du portail et redistribuées ensuite sur l'ensemble des bases pour constituer une réponse globale. Les réponses, au format XML, sont par la suite parsées pour construire et visualiser le graphe correspondant. Selon la nature de la requête formulée et du graphe obtenu, ProsoGraph permet différents types d'exploration des données et d'interprétation des réseaux obtenus. La figure 1 montre une illustration dans le cas où la requête permet d'extraire le réseau socio-professionnel d'une personne (gauche) et dans le cas où la requête porte sur les interactions autour d'un lieu d'intérêt (droite).

3 ProsoMap : Visualisation cartographique spatio-temporelle de données prosopographiques

A l'aide de simples fiches, il s'avère impossible de représenter les phénomènes dynamiques (déplacement, changement de fonction, mutation, durée de séjour, etc.). Or, de nombreuses représentations des phénomènes dynamiques ont été proposées dans la littérature Andrienko et al. (2013). Pour le portail Personae, deux représentations spatio-temporelles ont été retenues et implémentées dans l'interface ProsoMap en utilisant la bibliothèque OpenLayers (<http://openlayers.org/>).

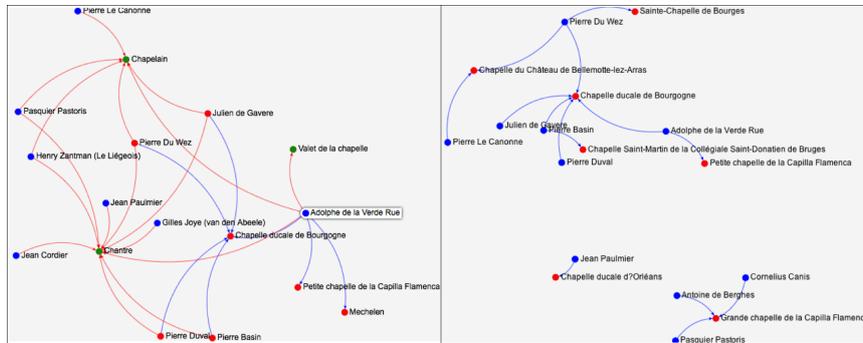


FIG. 1 – Un exemple de réseau socio-professionnel (gauche) et d'interactions autour d'un lieu d'intérêt (droite)

layers.org) et des cartes issues de différentes sources (GoogleMap, OpenStreetMap, GeoPortail, etc.). La première représentation se focalise sur la carrière d'un individu. La deuxième traite des flux de déplacements entre les différents lieux. Pour relier les lieux à des coordonnées géographiques, une URI Geonames(<http://www.geonames.org/>) est utilisée. Il faut noter que des lieux historiques disparus peuvent être ajoutés à la base Geonames. L'interface ProsoMap career retrace la trajectoire de vie d'un individu. Par rapport à la représentation classique de la *time geography*, une représentation en 2D a été retenue, la durée de résidence de l'individu dans un lieu est définie par la taille du cercle représentant ce lieu. Cette trajectoire est orientée par la nuance de gris (blanc pour le début de vie et noir pour la fin). L'interface ProsoMap flow, représente les flux de déplacements d'une population extraite du portail à l'aide de plusieurs critères obligatoires tels que le lieu et optionnels tels que la période, la profession, etc. Pour chaque lieu, deux cercles concentriques sont dessinés. Le plus petit représente le nombre d'individus de cette ville n'ayant pas changé de ville durant la période. Le plus grand décrit le nombre d'individus de cette ville ayant été dans la ville durant cette période. La ville étudiée est reliée aux villes d'où provient une partie de la population ou d'où part une autre partie durant cette période. La figure 2 donne une illustration des deux représentations implémentées dans ProsoMaps.

4 Conclusion

Dans cet article, nous avons proposé deux interfaces complémentaires ProsoGraph et ProsoMap pour la visualisation de données prosopographiques. Ces interfaces offrent des approches génériques facilitant la navigation entre les concepts et l'analyse visuelle impossible avec un grand nombre de fiches prosopographiques. Les deux interfaces communiquent avec le portail à travers le requêtage des concepts ontologiques de haut niveau : personnes, lieux, fonctions et des relations temporelles entre ces concepts. De ce fait, il nous a été possible d'utiliser des mécanismes d'inférence de relations implicites entre les données afin de compléter les visualisations. Cet aspect est en cours d'approfondissement afin d'aboutir à la mise en place de mécanismes d'inférence capables de s'adresser aux problèmes de lacunes et imprécisions dans

Visualisation de données prosopographiques

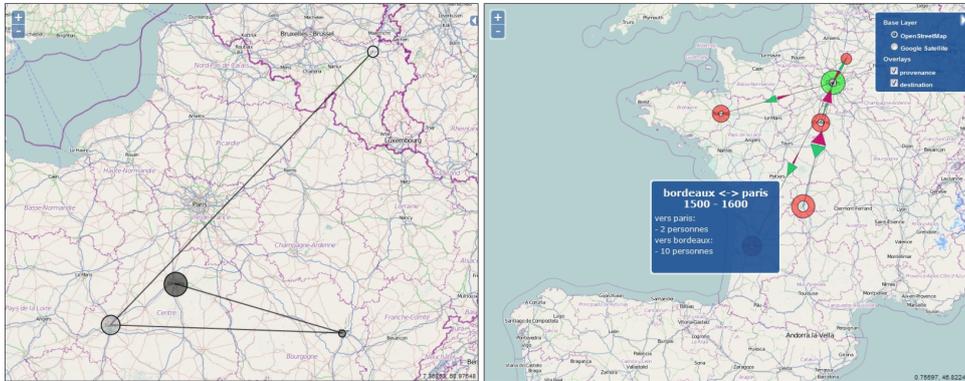


FIG. 2 – Un exemple de carte de carrière (gauche) et de flux (droite) obtenues par ProsoMap

les données prosopographiques. A terme, l'objectif est d'aboutir à un module de recommandation de nouvelles connaissances à l'expert historien pour compléter et enrichir les bases de données. Une fois le portail finalisé et l'ensemble des bases de données intégrées, il est prévu d'étudier les performances de ces interfaces. En présence de l'ensemble de données visées par le projet, ProsoGraph et ProsoMap auront pour défi de dégager des patterns de réseaux sociaux ou de parcours spatio-temporels génériques qui permettent d'identifier par exemple des écoles de pensée ou de mettre en évidence des concepts fortement connectés sur des populations de milliers d'individus.

Références

- Andrienko, G., N. Andrienko, P. Bak, D. Keim, et S. Wrobel (2013). *Visual Analytics of Movement*. Springer.
- Keim, D., F. Mansmann, J. Schneidewind, J. Thomas, et H. Ziegler (2008). Visual analytics: Scope and challenges. In *Visual Data Mining*, Volume 4404 of *LNCS*, pp. 76–90. Springer.

Summary

This paper presents two visualization tools ProsoGraph and ProsoMap developed within a collaborative project which focuses on the access and exploration of prosopographic datasets of the Renaissance period in France. The project aims at modeling and implementing a semantic portal ensuring access to existent prosopographic datasets. ProsoGraph and ProsoMap are developed based on social networks visualization techniques and spatio-temporal data visualization techniques, respectively. They interact with the portal through a semantic layer and allow additional advanced querying capabilities.