

Comment explorer des données relationnelles ou de réseaux

Jean-Daniel Fekete, Pascal Cristofoli

Atelier, mardi 13 octobre 9h30–12h15 et 13h45–16h30 (atelier répété 2 fois)

Présentation générale

En SHS, il est fréquent de vouloir comprendre un phénomène en étudiant les relations qui associent ses artefacts. L'analyse des données relationnelles permet de mettre en œuvre cette démarche pour laquelle les données potentiellement mobilisables sont pléthoriques. En général, elles sont créées à partir de documents ou de dispositifs techniques qui référencent des personnes, des lieux et des organisations, ou encore à partir de protocoles d'enquêtes spécifiques. Elles peuvent concerner de nombreux sujets et s'intéresser à des types de relations très variés.

Nous allons rappeler brièvement l'histoire de l'analyse de réseaux, en particulier sociaux, et montrer ensuite comment des réseaux peuvent être constitués et explorés, essentiellement visuellement. Nous mettrons l'accent sur la séparation entre l'exploration et la présentation à l'aide de visualisations.

Nous aborderons le cours en partant de plusieurs types de données classiques, stockées sous forme de tables de données, et nous montrerons comment les visualiser pour les explorer. Nous commencerons par la représentation à base de nœuds et liens, classique mais limitée, et montrerons les représentations alternatives qui permettent une meilleure exploration tout en étant souvent moins efficaces pour la présentation des résultats dans un article.

Nous montrerons comment visualiser et explorer des réseaux simples ou d'autres plus complexes: réseaux à un ou plusieurs modes, avec un nombre d'attributs et de relations plus ou moins grand, et/ou qui évoluent dans le temps (réseaux dynamiques ou longitudinaux).

L'atelier utilisera des logiciels faciles à utiliser et à s'approprier : en particulier les applications web The Vistorian et PAOVis, ainsi que le prototype PK-Clustering.

Environnement informatique

Les outils d'exploration des réseaux abordés dans l'atelier sont libres ou gratuits. Il s'agit d'applications web ne demandant aucune installation locale, mais nécessitant l'utilisation d'un navigateur web récent (nous conseillons d'installer la dernière version de Google Chrome : <https://www.google.com/chrome/>).

Pour interagir facilement avec les applications, il est aussi fortement conseillé de se munir d'une souris.

Le format de données en entrée de ces applications est le format tabulaire CSV. Les fichiers de ce type peuvent être créés dans un éditeur de texte ou un tableur (Libre office ou Open office de préférence, ou bien encore Excel).

Type de données traitées - droit d'accès

L'atelier se fonde sur l'exploitation de données relationnelles. Nous donnerons de nombreux exemples de types de réseaux et présenterons la façon de les collecter et de les coder dans un format tabulaire (type CSV). Les participants sont fortement encouragés à travailler avec leurs propres données. Ces données sont traitées par les logiciels que nous utiliserons sans être transmises à distance : elles restent sur le navigateur web de la machine et ne risquent pas de « fuiter » par erreur. L'application PK-Clustering les transfère à un serveur qui ne les garde pas après la session de travail.

Niveau requis

Aucune connaissance informatique n'est requise pour assister à l'atelier et utiliser les outils, il n'y aura pas de programmation ni d'utilisation de langage informatique. L'atelier suppose de comprendre le modèle relationnel et sa structuration en données tabulaires (ce qui sera rappelé en début de séance à l'aide d'exemples) ; le fait d'avoir préalablement constitué son propre corpus permettra aux auditeurs de mieux prendre en main les applications et s'approprier la démarche et les questions.

Objectifs

L'atelier vise à présenter les méthodes et outils dédiés à l'exploration de données relationnelles et de réseaux. Il propose d'utiliser quelques logiciels simples d'exploration interactive de réseaux, et de montrer comment les multiples représentations visuelles permettent d'explorer différents types de réseaux et, en diversifiant les points de vue, de faire progressivement connaissance avec un corpus de données.

Formule pédagogique

Les participants sont invités à utiliser leurs propres données pendant l'atelier. Le plus simple est de les saisir sous la forme de fichiers textes au format CSV ou dans un tableur en les exportant au format CSV (pour plus d'informations, consulter les pages de présentations des outils).

Références bibliographiques

- Edward Tufte, *The Visual Display of Quantitative Information*
- Alberto Cairo, *The Functional Art*
- Nathan Yau, *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*
- Tamara Munzner, *Visualization Analysis and Design*
- Wasserman, Stanley & Faust, Katherine (1994). *Social Networks Analysis: Methods and Applications*
- Di Battista, Giuseppe; Eades, Peter; Tamassia, Roberto; Tollis, Ioannis G. (1998), *Graph Drawing: Algorithms for the Visualization of Graphs*
- von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J., Fekete, J.-D. and Fellner, D. (2011), *Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges*. *Computer Graphics Forum*, 30: 1719-1749. doi:10.1111/j.1467-8659.2011.01898.x

Liens

Les applications web abordées dans l'atelier sont présentées sur les pages suivantes :

- The Vistorian : <https://vistorian.net/>
- PAOHVis: <https://www.aviz.fr/Research/Paohvis>
- PK-Clustering : <https://aviz.fr/pkclustering>