

**Rapport sur la thèse de Alexis Pister intitulée:**

**" Analyse Visuelle de Réseaux Sociaux Historiques:  
Traçabilité, Exploration et Analyse "**

Le manuscrit de Alexis Pister qui s'étend sur un total de 162 pages est organisé en 6 chapitres et rédigé en anglais. Il comprend une bibliographie rassemblant une liste exhaustive de références (254 au total) couvrant le domaine de la visualisation, et de l'analyse de réseaux (historiques, mais aussi plus généralement en sciences sociales). La facture du document est soignée et abondamment illustrée, son style rédactionnel le rend facile à lire, et la pédagogie de l'auteur contribue à donner au lecteur un discret fil rouge tissé entre sections et chapitres.

Dans les premiers chapitres, l'auteur s'attache à décrire comment le champ des sciences sociales, et par conséquent celui des sciences historiques, a évolué vers des approches quantitatives et plus particulièrement celle de l'analyse des réseaux. Il évoque combien celles-ci sont encore toutefois questionnées. J'aurai été curieux de savoir si les propos de Stone qui sont repris dans le texte (section 2.2.2) trouve encore un écho aujourd'hui, comme le fait Etienne Ollion en critiquant la place faite au big data en sciences sociales<sup>1</sup>.

Le ton emprunté marque clairement le souci de l'auteur à cerner les besoins de la communauté des historiens, tant du côté de la modélisation que de la visualisation de réseaux. Très tôt dans le manuscrit l'auteur souligne combien il lui semble indispensable de concevoir des approches et des outils garantissant une bonne utilisabilité des interfaces et assurant la traçabilité des manipulations. Il ajoute à ces deux prérequis la capacité, au travers de la modélisation, à capturer le plus fidèlement possible le sens attribué aux données par les historiens dans les phases amont de l'analyse. A son sens, les étapes de construction du réseau, et la capacité à retracer la provenance des données ayant mené à cette modélisation, est un trait caractéristique de la discipline HSNA (Historical Social Network Analysis).

De cette perspective irriguée par une observation attentive des besoins des utilisateurs de la communauté des sciences historiques émergent trois principes (chapitre 3) qui serviront de socle aux approches développées aux chapitres 4 (CombiNet) et 5 (PK Clustering). Sur la base de nombreux échanges au cours de collaborations avec des experts de domaines en histoire, l'auteur pose les principes de traçabilité, de simplicité et de réalité des documents historiques pour guider la conception de toute approche de visualisation analytique.

Les principes de simplicité et de réalité des documents l'amènent à conclure à la pertinence pour la modélisation et la visualisation de réseaux en histoire, de s'appuyer sur l'utilisation de graphes bipartis (2-mode). Ceux-ci permettent en effet d'embarquer en une même vue les acteurs identifiés dans les différents documents (et objets de l'analyse), et les documents qui y font référence. Les entités du réseau sont de plus enrichies de différents attributs, définissant ainsi le modèle de graphe biparti multivarié et dynamique comme apportant une réponse aux attentes des experts.

Cette intéressante discussion est documentée et argumentée sur la base de collaborations (d'où sont tirés les cas d'usage présentés dans le document). Cette réflexion et la proposition qui est faite sont justifiées d'autant que les graphes bipartis (souvent appelés des réseaux d'affinités) sont monnaie courante en sciences sociales. On aurait toutefois aimé que la conclusion s'appuie sur une validation plus « critique » du modèle proposé tant il existe d'autres approches qui semblent porter leur fruit.

---

<sup>1</sup> Ollion, É. and J. Boelaert (2015). "Au-delà des big data. Les sciences sociales et la multiplication des données numériques." Sociologie 6(3).

Everett et Borgatti<sup>2</sup> prétendent que l'approche alternative, utilisant la ou les projections 1-mode obtenues du graphe biparti, est toute aussi valable dès lors que sont conjuguées les deux projections — ce que proposent, par exemple, les travaux de Benjamin Renoust<sup>3</sup>.

Cela dit, le choix du modèle biparti multivarié fait sur la base d'échanges nourris avec les experts apporte un socle robuste pour les travaux décrits dans les chapitres suivants. CombiNet (chapitre 4) constitue la réponse aux attentes et besoins des experts. Les différents éléments de l'approche CombiNet suit d'une analyse des tâches que doivent mener les experts, s'appuyant sur le modèle biparti multivarié. Les questions des experts, à chaque fois ancrées dans le contexte du domaine étudié, sont traduites en des tâches conduites au niveau d'une interface de visualisation, et en des spécifications de l'interface tant des représentations graphiques des données que des manipulations à conduire.

Le résultat est convaincant. Les vues proposées sont élaborées en accord avec les principes et bonnes pratiques du domaine. La synchronisation des vues apporte une cohérence au niveau de l'utilisation de l'interface et répond bien aux principes de simplicité (et d'utilisabilité). Le retour aux informations contenues dans le document, et plus encore la matérialisation de la démarche suivie par l'utilisateur au cours de sons analyse (par un arbre encodant les différentes manipulations, retours en arrière et scénarios alternatifs), répondent aux principes de traçabilité et de réalité des documents. La proposition CombiNet renforce la validité des principes énoncés aux chapitre précédent.

L'approche PK Clustering (chapitre 5) vient en soutien d'une tâche de classification à laquelle se prête les experts dès lors que l'on souhaite regrouper les individus — par allégeance à l'une des personnes clés du réseau, par exemple. Les algorithmes de classification sont par nature des objets mathématiques difficiles à manœuvrer. Ils apparaissent la plupart du temps aux utilisateurs, même experts de domaine, comme des boîtes noires que l'on peut légitimement questionner. Aucun de ces algorithmes ne fait le lien, de manière intelligible, avec une connaissance du domaine d'origine — sauf peut-être à l'encapsuler sous la forme d'un artifice numérique (par des fonctions de poids, ou diverses contraintes mathématiques imposées à l'algorithme).

De plus, aucun algorithme ne peut prétendre à lui seul avoir réponse à toutes les tâches de classification. L'auteur fait donc le choix de rechercher un consensus parmi un ensemble d'algorithmes, la « convergence » des algorithmes vers une classification révélant en quelque sorte son caractère intrinsèque.

Face à cette situation, l'auteur propose une approche permettant à l'expert de décrire une connaissance a priori de la classification à trouver. Il est tout-à-fait vraisemblable, par exemple, qu'un expert sache de certains acteurs à qui ils ont fait allégeance. L'approche PK Clustering permet donc de l'inscrire comme une contrainte soumise aux algorithmes.

PK Clustering permet donc de positionner l'ensemble des connaissances a priori, et de faire ensuite une lecture d'un consensus émergeant des résultats des différents algorithmes. Il permet d'inspecter de plus près le résultat de l'un d'entre eux, ou de valider certains choix (label attribué à un acteur).

Mon intérêt pour ce sujet ne se tarit pas, et je dois avouer avoir lu ce chapitre avec grand intérêt. La proposition PK Clustering est à mon sens un travail très abouti d'une idée que plusieurs ont eu sans l'avoir menée à son terme. La facture soignée du tableau de bord doit encore ici être soulignée. Cela dit, on s'écarte peut-être du principe de simplicité, qu'il faut attribuer à la complexité de la tâche de classification et des algorithmes qu'elle sollicite. Les différents cas d'usage viennent encore une fois confirmer l'intérêt de cette approche pour le travail des experts.

---

<sup>2</sup> Everett, M. G. and S. P. Borgatti (2013). The dual-projection approach for two-mode networks. *Social Networks* 35(2): 204-210.

<sup>3</sup> Renoust, B., et al. (2015). "Detangler: Visual Analytics for Multiplex Networks." *Computer Graphics Forum* 34(3): 321-330.

La conclusion et les pistes de travaux futurs dessinées par l'auteur apportent des éléments de réponses aux quelques limitations que j'ai pu souligner précédemment. De l'aveu de l'auteur, le caractère dynamique des réseaux en histoire reste encore à exploiter. Cela pose encore des défis tant du point de vue de la modélisation (et de la classification !) que de la visualisation.

Pour conclure, mon expérience à travailler en contexte pluridisciplinaire depuis de nombreuses années et en particulier avec des chercheurs de sciences sociales, me permettent d'affirmer combien le travail d'Alexis Pister est pertinent et constitue une avancée dans le domaine. Clairement, les réflexions développées dans le manuscrit et leur matérialisation par les propositions CombiNet et PK Clustering marquent un jalon dans le développement de la visualisation analytique au service des sciences sociales.

A l'évidence, le manuscrit révèle la maturité du candidat, sa maîtrise de l'état de l'art, et sa capacité à faire avancer le domaine. J'apporte un soutien très favorable à la soutenance de la thèse.

Fait à Bordeaux, France, le 15 novembre 2022



Guy Melançon  
Professeur des Universités  
Université de Bordeaux